Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Aimee Lynn Rinas

Entitled
Advancing the Applicability of Fast Photochemical Oxidation of Proteins to Complex Systems

For the degree of Doctor of Philosophy ▾

Is approved by the final examining committee:

Lisa Jones
Chair

Millie Georgiadis

Eric Long

Nicholas Manicke

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Lisa Jones

Approved by: Eric Long 7/20/2016
Head of the Departmental Graduate Program Date

ADVANCING THE APPLICABILITY OF FAST PHOTOCHEMICAL OXIDATION OF PROTEINS TO COMPLEX SYSTEMS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Aimee Lynn Rinas

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2016

Purdue University

Indianapolis, Indiana

To My family and friends for your unending love and support, and for keeping my life balanced throughout this endeavor.

To Shilpi for walking in front of me when I needed guidance, with me when I needed support and encouragement, and behind me when I needed you to have my back. You are my favorite thing, my very favorite thing.

ACKNOWLEDGMENTS

First and foremost, I would like express my gratitude to my mentor, Dr. Lisa M. Jones, for accepting me into her group, and her guidance and dedication in my development as a scientist. It was an honor to your first PhD student. Furthermore, I am extremely grateful for her unending dedication to mentoring underprivileged and underrepresented minority students. The experience I gained working working with the high school and undergraduate students she brought into our group each summer is immeasurable. Having the opportunity to mentor these students has had a profound effect on my scientific and personal growth. Along with Dr. Jones, I want to acknowledge Elmer Sanders and his commitment to the Indiana CTSI Project SEED and Project STEM. I would also like to all of the members of the Jones group for being an excellent sounding board and their willingness to help. I truly enjoyed my time working with them.

In addition, I want to acknowledge my collaborators at IUPUI; Dr. Amber Mosley for her guidance with MudPIT and Seth Winfree for his assistance with the fluorescence imaging of the in cell FPOP flow cell.

Lastly, I gratefully acknowledge the members of my Ph.D. committee; Dr. Millie Georgiadis, Dr. Eric Long, and Dr. Nick Manicke. Thank you for your insightful comments and making my defense a gratifying experience.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES viii

ABSTRACT xi

1 INTRODUCTION 1
  1.1 Research Aims: An Upgraded FPOP Method 1
    1.1.1 Aim 1: Expedited Data Analysis 2
    1.1.2 Aim 2: Increased FPOP Identifications 2
    1.1.3 Aim 3: An Updated Flow System 2
  1.2 Principles of Protein Analysis by Mass Spectrometry 3
    1.2.1 Electrospray Ionization Mechanism 4
    1.2.2 Protein Identification by Mass Spectrometry 5
  1.3 Protein Structure Characterization 8
    1.3.1 High Resolution Structural Characterization 8
    1.3.2 Low Resolution Structural Characterization 9
  1.4 Protein Structural Characterization by Mass Spectrometry 10
    1.4.1 Chemical Cross-linking 11
    1.4.2 Protein Footprinting 11
  1.5 Hydroxyl Radical-Mediated Protein Footprinting 13
    1.5.1 Methods for Generating Hydroxyl Radical Labels for Footprinting 13
    1.5.2 Fast Photochemical Oxidation of Proteins 14
  1.6 References 17

2 AN EFFICIENT QUANTITATION STRATEGY FOR HYDROXYL RADICAL-MEDIATED PROTEIN FOOTPRINTING USING PROTEOME DISCOVERER 25
  2.1 Introduction 25
  2.2 Experimental Section 26
    2.2.1 Materials 26
    2.2.2 Protein Expression and Purification 26
    2.2.3 Oxidative Labeling 27
    2.2.4 Proteolysis 27

Page
2.2.5 LC-MS/MS Acquisition 28
2.2.6 Data Analysis 28
2.3 The Proteome Discoverer Analysis Strategy 30
2.4 Results 31
2.4.1 Peptide Level FPOP Oxidation of GCaMP2 32
2.4.2 Residue Level FPOP Oxidation of GCaMP2 38
2.5 Discussion 46
2.6 Conclusion 49
2.7 References 50

3 FAST PHOTOCHEMICAL OXIDATION OF PROTEINS COUPLED TO MUDPIT: EXPANDING FOOTPRINTING STRATEGIES TO COMPLEX SYSTEMS 53
3.1 Introduction 53
3.2 Experimental Section 54
3.2.1 Materials 54
3.2.2 Oxidative Labeling 55
3.2.3 Proteolysis 55
3.2.4 LC-MS/MS Acquisition 56
3.2.5 Data Analysis 57
3.3 Results 58
3.3.1 LC-MS Method Comparison 58
3.3.2 Increases in Identifications by MudPIT 59
3.3.3 Properties of Peptides Identified by MudPIT 62
3.3.4 MudPIT as a Method for Megadalton Protein Complexes . 64
3.4 Discussion 67
3.5 Conclusion 68
3.6 References 69

4 DEVELOPMENT OF A MICRO-FLOW SYSTEM FOR IN CELL PROTEIN FOOTPRINTING ANALYSIS 72
4.1 Introduction 72
4.2 Experimental Section 74
4.2.1 Materials 74
4.2.2 Flow Cell Construction 74
4.2.3 Oxidative Labeling 75
4.2.4 Cell Lysis and Proteolysis 76
4.2.5 LC-MS/MS 76
4.2.6 Data Analysis 77
4.2.7 Fluorescence Imaging 78
4.3 Results and Discussion 78
4.3.1 Flow Dynamics 78
4.3.2 IC-FPOP Using the Flow System 84

Page
4.4 Conclusion 91
4.5 References 92
4.6 Appendix: Flow Assembly for Cells Patent Application 95

VITA 119

PUBLICATIONS

An Efficient Quantitation Strategy for Hydroxyl Radical-Mediated Protein Footprinting using Proteome Discoverer 120
Fast Photochemical Oxidation of Proteins Coupled to Multidimensional Protein Identification Technology (MudPIT): Expanding Footprinting Strategies to Complex Systems 130

LIST OF TABLES

Table Page

1.1 Residues and Associated Masses 7

1.2 Calculating Fragment Ion Masses 8

2.1 Workflow Modification Distribution 32

2.2 GCaMP2 Oxidized Peptides 33

2.3 Rate Constants and Relative Intrinsic Activities 35

2.4 GCaMP2 Oxidized Residues 39

2.5 Comparison of CaM Domain PD Oxidation to Previously Published Data 45

3.1 Oxidatively Modified Residues Identified 61

LIST OF FIGURES

Figure Page

1.1 Diagram of Electrospray Ionization: 1. Liquid jet emitted from Taylor cone. 2. Solvent evaporation from droplets. 3. Coulomb repulsion exceeding the surface tension of the droplet triggers droplet fission. Reproduced with permission under the Creative Commons Attribution-Share Alike 4.0 International License.[22] 4

1.2 FPOP Schematic where UV laser pulses split H_2O_2 forming the ·OH that label the solvent exposed areas of the proteins. 14

2.1 PD workflow developed for FPOP analysis displaying the search hierarchy with Mascot and Sequest HT search levels (dark orange nodes), target-decoy validation (tan nodes), and precursor peak areas are calculation (pink node). 29

2.2 Comparison of PD method to PM method showing the sequence coverage (a) with PD total and footprinting coverage in blue and green, respectively, and PM total and footprinting coverage in red and yellow, respectively, as well as the agreement between the fold change (b) and correlation (c). 34

2.3 Correlation between SASA and log PF for both peptide level (a, b, and c) and residue level (d, e, and f) analysis for the CF state (a and d) and the CB monomeric (b and e) and dimeric (c and f) states. 37

2.4 Fold change between the CF and CB states for the 22 oxidized peptides identified by the PD method. 38

2.5 Conformational difference detected in the two residues in the M13-GFP linker with an overlay of the calcium free (light gray and blue, PDB ID: 3EKJ) and the calcium bound structures (dark gray and red, PDB ID: 3EK4)(a) and the oxidation and the SASA of the two residues for each conformation (b). 42

Figure Page

2.6 Inter-domain contacts between GFP (gray) and the N-terminal lobe of CaM (light purple) in the CF conformation (a) and structural changes in the CB conformation (b) with residues with higher oxidation in the CB conformation colored blue, and a bar graph (c) of the oxidation level of the highlighted residues. 44

2.7 MS/MS spectra of FPOP modifications on D194 of peptide 185-199 with an ambiguous assignment of a decarboxylation on D194 (a) that was removed by the strict filtering scheme, and the same peptide and location with a loss of CO (b) that pass the filters with fragment ions that validate the oxidation. 48

3.1 Visual comparison of IDs between MudPIT (red) and 1D-DDA (blue) methods by proteins (top), peptides (middle), and oxidatively modified peptides (bottom). 60

3.2 Distribution of the intensities of PSMs identified by MudPIT (red) and 1D-DDA (blue). (a) The spread of intensities is demonstrated in the box-and-whisker plot with the box lines marking the upper median and lower quartiles, and the whiskers marking the complete range. (b) The frequency of the distributions of intensities is displayed in a histogram. 63

3.3 Two perspectives of the structural location of MudPIT determined FPOP oxidation levels mapped to a yeast 80 s ribosomal crystal structure, 4V6I[20]. The lowest oxidation levels are in blue going to the highest in red. 65

3.4 Fractional SASA values versus Log PF values of MudPIT identified residues of RACK1, chain Aa of the 4V6I[20] ribosomal structure, illustrating the linear relationship between the values. 66

4.1 Hydrodynamic focusing of particles. 73

4.2 IC-FPOP Flow Cell Schematic: Blue arrows indicate flow direction, and amber lines represent polyimide coated fused capillary tubing (not to scale). The cellular solution and hydrogen peroxide are mixed in a tee to prevent breakdown by endogenous catalase. The capillary containing the cellular analyte is mounted so that the buffer surrounds the capillary, creating a sheath to centrally focus the cells. Optimal conditions were observed with a 10 to 1 sheath buffer to cellular analyte ratio. 75

4.3 Flow system average intensity projection showing the locations of the sheath buffer(green) and cellular analyte (red) as they flow through the flow cell. 79

Figure Page

4.4 3D average intensity heat map with the sheath on top and cellular analyte on the bottom. Lowest intensity is blue and highest is red. 81

4.5 Maximum intensity heat map showing the locations of all detected cells. A cutout of Figure 4.3 is shown on the bottom to illustrate overlap. . . 82

4.6 Cross-sectional YZ stack demonstrating 3D hydrodynamic focusing (a) and location of cells (b, red) with the sheath and cellular analyte buffer in gray for ease in viewing cells. 83

4.7 Visual comparison of the oxidized proteins identified with (blue) and without (red) using the flow system (a) and the flow system biological replicates (b). 85

4.8 Amino acid distribution of oxidation events relative to its overall abundance in all the proteins identified (a) and the average occurrence of the residue being solvent exposed. 87

4.9 Hierarchical network of oxidized proteins identified using the flow system mapped to the generic GO Slim ontologies for cellular components with the nodes sized by the percentage proteins. 88

4.10 Oxidation levels of 7 actin peptides (a) and correlation of residue level log protection factors to the relative SASA of open (1HLU) and tight (2BTF) states of actin (b). 91

ABSTRACT

Rinas, Aimee Lynn Ph.D., Purdue University, August 2016. Advancing the Applicability of Fast Photochemical Oxidation of Proteins to Complex Systems. Major Professor: Lisa M. Jones.

Hydroxyl radical protein footprinting coupled with mass spectrometry has become an invaluable technique for protein structural characterization. In this method, hydroxyl radicals react with solvent exposed amino acid side chains producing stable, covalently attached labels. Although this technique yields beneficial information, the extensive list of known oxidation products produced increases the complexity of identifying and quantifying oxidation products. The current methods available for quantifying the extent of oxidation either involve manual analysis steps, or limit the number of searchable modifications or the size of sequence database. This creates a bottleneck which can result in a long and arduous analysis process, which is further compounded in a complex sample. In addition to the data complexity, the peptides containing the oxidation products of hydroxyl radical-mediated protein footprinting experiments are typically much less abundant than their unoxidized counterparts. This is inherent to the design of the experiment as excessive oxidation may lead to undesired conformational changes or unfolding of the protein, skewing the results. Thus, as the complexity of the systems studied using this method expands, the detection and identification of these oxidized species can be increasingly difficult with the limitations of data-dependent acquisition (DDA) and one-dimensional chromatography. The recently published in cell FPOP method exemplifies where this field is headed - larger and more complex systems. This dissertation describes two new methodologies and one new technology for hydroxyl radical-mediated protein footprinting, expanding the applicability of the method. First is development of a new footprinting analysis method for both peptide and residue level analysis, allowing for faster quantification of results. This method utilizes a customized multilevel search workflow developed for an on-market search platform in conjunction with a quantitation platform developed using a free Excel add-in, expediting the analysis process. Second is the application of multidimensional protein identification technology (MudPIT) in combination with hydroxyl radical footprinting as a method to increase the identification of quantifiable peptides in these experiments. Last is the design and implementation of a flow system for in cell FPOP, which hydrodynamically focuses the cells, and when used yielded a 13-fold increase in oxidized proteins and 2 orders of magnitude increase in the dynamic range of the method.

1. INTRODUCTION

1.1 Research Aims: An Upgraded FPOP Method

Fast photochemical oxidation of proteins(FPOP)[1] is a mass spectrometry-based protein footprinting method that is used to probe the solvent accessible surface areas of proteins. Hydroxyl radicals, generated through the photolyzation of hydrogen peroxide, react with amino acid side chains forming stable, covalent labels that can be detected using mass spectrometry. As with other protein footprinting methods, FPOP is used to determine changes in solvent accessibility that arise from protein or ligand binding, the formation of aggregates, and other interactions that may perturb the structure of the protein or induce conformational changes.

Although hydroxyl radical footprinting and more specifically FPOP offers several advantages (discussed in more detail is section 1.3), data analysis for this method remains arduous, hindering the broad use and the capability of the method.[2] This is a result of the large number of possible oxidation products which makes their identification and quantification significantly more complex than a proteomics counterpart, such as SILAC, where stable isotope ratios are quantified.[3–5] While several efforts have been made to expedite data analysis, the available methods still have limitations; they are manually intensive, cannot provide a level of confidence in identifications, limit the number of modifications or protein sequences that the data can be searched against, limit the resolution in localizing the modification, or some combination of these.[2,6–13] These limitations are further compounded by sample complexity, forestalling the use of FPOP for complex systems.

My overall research goal was to give the FPOP method an upgrade, expanding its usability beyond single proteins or small interaction studies.

1.1.1 Aim 1: Expedited Data Analysis

Central to my overall goal was overcoming the hurdles of FPOP data analysis, requiring a new data analysis method. My objective was to develop an identification and quantitation platform that met the following design goals; it should be built-upon established proteomics standards, have minimal manual steps, be compatible with all levels of sample complexity, be adaptable for use any stable footprinting label, and be built on or in software packages already widely used in proteomics.

1.1.2 Aim 2: Increased FPOP Identifications

In addition to data analysis, the identification of FPOP oxidized species needed to be addressed. Typically, the oxidation products of FPOP have a lower abundance than their unoxidized counterparts as a result of the experimental design; excessive oxidation can induce protein unfolding.[14] Fundamentally, their identification becomes increasingly difficult as sample complexity increases. My objective here was to identify and adapt a better chromatographic separation method that was compatible with FPOP while maintaining chromatographic peak quality for quantitation.

1.1.3 Aim 3: An Updated Flow System

Complex samples also required an improved FPOP flow system. Adaptation of the FPOP method for use with live cells was occurring in our group concurrently with the analysis and identification upgrades. During this process, it was observed that the flow system currently in use would continually clog when used with cells, leading to sample-to-sample variability. As such, new flow system needed to be designed that would prevent this clogging. My objective here was to design a flow cell that would prevent clogging and cell aggregation from a material that was both low cost and transparent at 248 nm, the wavelength used to split hydrogen peroxide to for the radicals.

1.2 Principles of Protein Analysis by Mass Spectrometry

Mass spectrometry (MS) is a gas phase analytical technique used to identify and characterize the structure of molecules. The compound(s) in the analyte are ionized generating charged species which are fragmented and identified by the mass-to-charge ratios (*m/z*) of the ions.[15] While MS has been used for decades for characterizing small molecules, it was the development of “soft” ionization methods that expanded the technique to proteins and other macromolecules. “Soft” ionization techniques are capable of generating ions of nonvolatile compounds with little fragmentation, properties essential for protein characterization. As a charged species is required for detection, an additive is typically combined with the analyte to aid in generating intact protonated $(M+nH)^{n+}$ or deprotonated $(M-nH)^{n-}$ molecular ions.[16] The most frequently used ionization techniques for proteins are matrix assisted desorption/ionization (MALDI) and electrospray ionization (ESI).[16–18] In MALDI, a pulsed laser triggers desorption of a crystalized mixture of analyte and matrix, where the analyte is ionized in the gas formed by the laser ablation.[18] In ESI, the flowing liquid analyte transformed into gas phase ions by applying a high voltage to the solution.[16,17,19] Although each “soft” ionization method offers benefits over the others, ESI offers distinct advantages. ESI can be coupled to liquid-phase separation techniques, such as liquid chromatography (LC-MS), allowing analysis of more complex samples.[20] Additionally, ESI tends to produce multiply charged ions making the analysis of large molecular weight (MW) ions possible on instruments with a narrow *m/z* range.[21] This has made LC-MS is an indispensable tool in proteomics, the study of the functions, structures, and interactions of proteins.

1.2.1 Electrospray Ionization Mechanism

In ESI-MS, an electric potential is applied to the liquid analyte solution as it flows through a capillary tube, forming a large charge density on the liquid surface at the tip of the capillary.[17,19] The electric field generated by this potential distorts the liquid into a conical shape called a Taylor cone, and a jet of liquid emits from the tip of this cone (Figure 1.1[22]).[23] At the threshold voltage, this jet breaks down into a spray of fine, highly charge droplets, and electrospray is initiated. Solvent evaporation from the droplets increases the charge density until the Coulomb repulsion

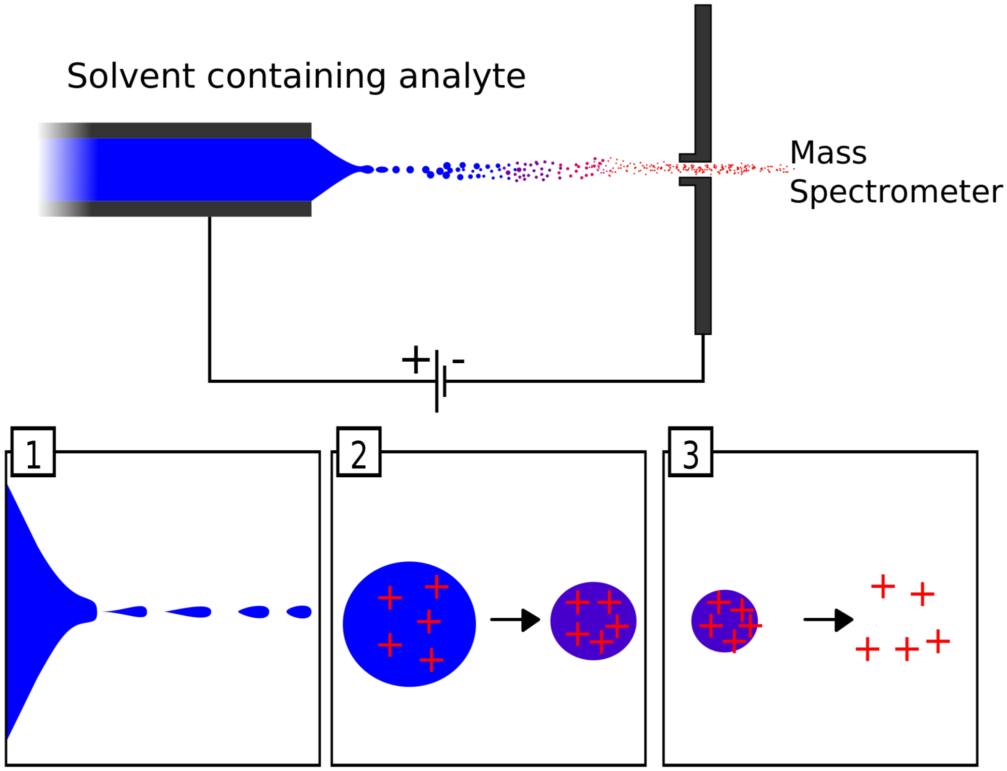


Figure 1.1. Diagram of Electrospray Ionization: 1. Liquid jet emitted from Taylor cone. 2. Solvent evaporation from droplets. 3. Coulomb repulsion exceeding the surface tension of the droplet triggers droplet fission. Reproduced with permission under the Creative Commons Attribution-Share Alike 4.0 International License.[22]

exceeds the surface tension of the droplet, triggering droplet fission; a cascading series of these evaporation and fission events occurs until nanoscale droplets are formed and the charged analyte ions are ejected into the gas phase.[24,25] Nano-ESI, also referred to as nanospray, sources generate smaller initial droplets, offering an increased ionization and sensitivity while requiring smaller sample sizes making them ideal for protein analysis.[26–28]

1.2.2 Protein Identification by Mass Spectrometry

Early analysis of proteins was completed through Edman degradation.[29] This method, which required a considerable amount of purified protein, identifies the amino acid (AA) sequence through stepwise chemical cleavages of the amino terminus (N-terminus) AA. The arduous process was often unsuccessful for long sequences or proteins with N-terminal modifications.[30] With the development of "soft" ionization techniques compatible with protein analysis, MS quickly replaced Edman degradation and has become a primary tool in proteomics.

Commensurate with the ionization sources used for proteomic analysis are the mass analyzers used in this field. There are four main types of mass analyzers, each with differing performance advantages, used in this field: ion trap, quadrupole, time of flight (TOF), Fourier transform ion cyclotron resonance (FT-ICR), and orbitrap.[16,20,30,31] Hybrid analyzers which combine these technologies are also used in this field; the objective is balancing mass accuracy, sensitivity, and resolution with acquisition time to yield a more accurate representation of the inherently complex samples in proteomics.

The key principle of orbitrap analyzers, like the Thermo Fisher Scientific Q Exactive used for this work, is an orbital method of trapping ions.[31] The ion trap consists of two axially symmetric types of electrodes, a central spindle-like electrode and barrel-like outer electrodes, that create an electric field. Ions are injected tangential to the electric field where they rotate around and move back and forth along the central electrode. These ions become electrostatically trapped due to the equilibrium between their attraction to the central electrode and the centrifugal force from rotation. This creates rings of specific *m/z* ions that oscillate harmonically along the widest part of the spindle.[31–33] These oscillations induce a current on the outer electrodes, which are used as a receiver to detect the image current of the oscillations. Similar to FT-ICR, all of the ions are detected simultaneously over a given period of time, and the time domains of the image currents are Fourier transformed; this data is used to generate the mass spectrum.[31–33]

Although intact proteins can be detected via MS, called "top-down" analysis, it is often necessary to break the proteins down into peptides for a more thorough "bottom-up" analysis. This is typically achieved via proteolytic digestion with trypsin. Trypsin is a protease that cleaves on the carboxy-terminal (C-terminal) side of lysine and arginine residues. This results in a basic residue on the C-terminal end of the peptides, ensuring that each peptide will be protonated, with doubly protonated $(M+2H)^{2+}$ peptides being the predominant species in ESI; protein identification by MS generally utilizes positive ions. Today, most *m/z* measurements are made in a two-step process known as tandem MS (MS/MS).[34,35] The first step, or MS1 scan, measures *m/z* of the intact peptide ion, referred to as the parent or precursor ion. In the second step, or MS2 scan, the parent ion is isolated and fragmented, and the *m/z* of the resulting fragment ions are measured.

Just as there are several types of mass analyzers used, a number of methods are used for gas-phase fragmentation of parent ions. Most frequently, fragment ions are produced by collisions with gas molecules, called collision induced dissociation (CID), and the process predominantly causes cleavage along the peptide backbone.[34–36] Fragment numbering extends from the terminal ends of the peptide; a-, b-, and c-ions appear to lengthen from and retain the charge on the N-terminus, while x-, y-, and z-ions appear to lengthen from and retain the charge on the C-terminus. It is the amide bond that is most likely to break in this process which produces the b- and y-ions.[30,36,37] While this is not an all-inclusive list of all possible fragments, these ion series are most frequently used for CID spectra interpretation. Table 1.1 lists the monoisotopic and commonly observed ions masses for each of the amino acids.[38,39] These masses along with the formulas in Table 1.2

Table 1.1
Residues and Associated Masses

Residue	Monoisotopic Mass	Immonium Ion[a]	Low Mass Ions[a]
Alanine	71.037114	44	
Arginine	156.101111	*129*	59,70,*73*,87,100,112
Asparagine	114.042927	*87*	70
Aspartic acid	115.026943	*88*	70
Cysteine	103.009185	76	
Glutamic acid	129.042593	102	
Glutamine	128.058578	101	56,84,129
Glycine	57.021464	30	
Histidine	137.058912	**110**	82,121,123,138,166
Isoleucine	113.084064	**86**	44,72
Leucine	113.084064	**86**	44,72
Lysine	128.094963	101	70,84,112,129
Methionine	131.040485	104	61
Phenylalanine	147.068414	**120**	91
Proline	97.052764	**70**	
Serine	87.032028	60	
Threonine	101.047679	74	
Tryptophan	186.079313	**159**	77,100,117,130,132,170
Tyrosine	163.06332	**136**	91,107
Valine	99.068414	72	41,55,69

[a] bold face indicates strong signals, italics indicates weak signals

are used to calculate the fragment *m/z* values.[40] The values are used to interpret the spectra and identify the sequence of the precursor peptide, which today is completed through software packages with search algorithms (discussed in Chapter 2).

Table 1.2
Calculating Fragment Ion Masses

Ion	Neutral Fragment Molecular Weight
a	[N]+[M]-CO-H
b	[N]+[M]-H
c	[N]+[M]+NH_2
x	[C]+[M]-CO-H
y	[C]+[M]-H
z	[C]+[M]+NH_2
a-, b-, y-NH_3	-17.0266
a-, b-, y-H_2O	-18.0106

1.3 Protein Structure Characterization

1.3.1 High Resolution Structural Characterization

Structural biology has played an essential role in advancing our functional understanding of biological systems. At present, approximately 11% of Swiss-Prot proteins have high resolution structures available, with roughly 90% of available structures solved via X-ray crystallography.[41–43] In this method, a beam of X-rays is diffracted by the atoms in a high purity protein crystal. The intensities of the resulting diffraction pattern are measured and phases calculated to create an electron density map, to which the primary sequence of the protein is modeled to fit. This process yields a structure with atomic level coordinates. However, the ability to obtain a high resolution structure by means of this method is limited by the ability to grow a high quality crystal.[42] This task is often difficult due to the inherent properties of the protein or protein assembly, or the ability to purify the macromolecule in quantities sufficient to obtain a quality crystal.[44] The difficulties are compounded when investigating the structures of large, megadalton sized molecular assemblies. At the outset, it is often challenging to purify all of the protein components of megadalton complexes, a necessity in obtaining a structure.[45] Even when this is accomplished, it can be equally difficult to crystallize these complexes, or the process may only yield crystals too small for analysis.[46]

Another method used for obtaining high resolution structures is nuclear magnetic resonance (NMR) spectroscopy. This method utilizes the magnetic properties of the ½ spin isotopes 1H, 13C, and 15N. One advantage of this method is that it generally completed on solution phase proteins, allowing for it to examine time-dependent fluctuations in the protein's structure versus the static structures that are obtained by X-ray crystallography; NMR is sensitive to local structural dynamics.[47] However, the method is not without limitations. Like X-ray crystallography, an ample quantity of high purity sample is required. And though crystal growth is not necessary, proteins and large assemblies that cannot have structures solved through X-ray crystallography may also not be amenable with this method either; molecules with subunits larger than about 40 kDa are not amenable to structure determination by NMR due to spectral overlap. Additionally, the acquired data does not always yield structural coordinates resulting in lower resolution characterization.

1.3.2 Low Resolution Structural Characterization

Since high resolution structures have only been obtained for a fraction of known proteins, researchers often turn to lower resolution methodologies to gain structural insight on their protein of interest. Circular dichroism spectroscopy, a solution phase method, can characterize the secondary structure of a protein. This method measures the absorption differences between left- and right-handed polarized light, and the resulting signals are used to determine what percent each of the secondary structure components exists in the protein.[48]

Fluorescence spectroscopy can also be used to investigate structural and dynamic properties of a protein. Phenylalanine (Phe), tyrosine (Tyr), and tryptophan

(Trp) residues are all fluorescent; however, Trp is the most sensitive and consequently, is most commonly used as an intrinsic fluorescent probe.[49] The emission spectra and changes in the emission spectra provide details on the local environment of the Trp residue. These changes are a result of unfolding, ligand or substrate binding, or subunit assembly, probing the different conformational states.[50] Similar to intrinsic fluorescence, Trp analogues, which are typically expressed recombinantly, can be used as pseudo-intrinsic probes to assess the aforementioned properties.[51] In addition, functional protein interactions can be quantified through techniques where fluorophores are conjugated to proteins, such as Förster resonance energy transfer (FRET) and fluorescence lifetime image microscopy (FLIM).[52]

Cryo-electron microscopy (cryo-EM) is a collection of techniques which are used to acquire both two- and three-dimensional protein structures, but at a lower resolution than crystallography. Here, a protein solution, frozen at cryogenic temperatures, is hit with an electron beam. Scattered electrons magnified by passing through a lens, creating a magnified image on the detector that are used to determine a structure.[53] Since thin films of protein solutions are used, these methodologies are particularly suitable for large protein assemblies, filling a gap left by high resolution techniques. Cryo-EM is also an emerging method for the determination of high resolution structures of large macromolecular complexes. While it suffers from the same limitations as X-ray crystallography for sample preparation, it is no longer limited to low resolution information. A single particle cryo-EM structure of human γ-secretase, an intermembrane protease, at 3.4 Å resolution was recently published; cryo-em can provide high resolution structures of proteins that cannot be solved by X-ray crystallography.[54]

1.4 Protein Structural Characterization by Mass Spectrometry

Structural MS-based methodologies, many of which utilize chemical reagents for investigation, have become an invaluable tool for evaluating protein structure and function.[55] These applications can provide structural information for membrane proteins, which make up less than 3% of known structures, or used to study conformational changes and folding dynamics.[41,56–60] As such, the use of MS based structural methodologies has become increasingly more popular in filling in the structural void as well as offering complementary information to high resolution methodologies.[55,61] Several strategies exist for this type of investigation, each yielding different structural information about the protein(s) in inquiry.

1.4.1 Chemical Cross-linking

Chemical cross-linking can divulge tertiary and quaternary information through both inter- and intramolecular covalent conjugation.[62] Here, two proteins or protein regions are covalently linked by a chemical cross-linker that reacts with the protein sidechain(s).[63] While chemical cross-linkers utilize numerous reactive groups, NHS ester cross-linkers are predominantly employed for this methodology. The popularity of these linkers emanates from both the highly efficient stable bonds formed with primary amine side chains, and prevalence of lysine residues in proteins.[64] Cross-linking experiments can also reveal global spatial relationships of large protein complexes; the use of various lengths of cross-linkers provides a series of distance constraints, allowing for an interactome network to be generated from the resulting data.[63,65] Although this method can yield quaternary structural information, there is a high level of complexity in identifying cross-linked species.[66]

1.4.2 Protein Footprinting

Protein footprinting utilizes chemical probes that covalently label proteins in solution.[55] Collectively, these methodologies are used to examine structural aspects of a protein assembly, and protein-protein or protein-ligand interactions. The primary aim of MS-based protein footprinting is to determine which sites on a protein are solvent accessible (SA), exploiting the fact that target site reactivities are dictated, in part, by the protein's conformation.[61] Strategies for chemical footprinting vary in whether they probe the amino acid side chains or the peptide backbone, or in the specificity of the chemical probe. One commonality between the strategies is that the chemical probe changes the mass of the peptide compared to the theoretical value by known amounts, a feature well suited for MS. Typically, the chemical probes are used to assess sites on a protein that show a change in SA consequent to being exposed to a ligand, or upon complexation with another protein(s).[3,67–69] The exception is hydrogen deuterium exchange (HDX), which also takes hydrogen bonding into account.[70]

HDX probes hydrogen bonding and solvent accessibility through monitoring the exchange of backbone amide hydrogen atoms yielding secondary structural properties. In HDX, deuterium is exchanged with backbone amide protons to probe the protein structure.[71] The rate of exchange is a function of both the local protein structure and the SA, and is determined through the increase in mass compared to the theoretical peptide mass. This protein footprinting method is widely used due to the non-specific nature and labeling efficiency of the deuterium probe. However, the instability of the deuterium label can lead to back exchange, requiring expeditious sample proteolysis and analysis via MS.[56,67,72]

Protein footprinting methods that are complementary to HDX utilize chemical probes to label side chains, revealing evidence of side chain solvent accessibility. The main advantage of these labels is that unlike HDX, these covalent labels are stable in solution.[73] A diverse set of labels are used which vary in their specificity from side chain specific labels. Highly specific label examples include organic anhydrides that can probe the SA of lysine residues, and carbodiimides that can probe the SA of carboxyl groups.[74] Conversely, hydroxyl radical (·OH) labels are indiscriminative in their side chain specificity, and can probe the SA of nearly all aminio acids.[14,74,75]

1.5 Hydroxyl Radical-Mediated Protein Footprinting

The use of ·OH for protein footprinting, first coupled with mass spectrometry by Chance and coworkers, has become more prevalent for probing sidechain SA due to the advantages this chemical probe offers.[76] Similar to deuterium in their specificity, ·OH are reactive with 19 of the 20 natural amino acids (excluding glycine), with varying rates.[75,77] Because of this property, these experiments can yield higher resolution information about the protein(s) SA regions than is achievable using other side chain labels. It should be noted, however, that detecting the oxidation products of serine and threonine can be problematic and are usually excluded from analysis.[75] Unlike hydrogen-deuterium exchange, ·OH-induced modifications do not undergo back-exchange and are stable, allowing for a higher, residue-level resolution.[3,55] Additionally, since ·OH have properties similar to water they can readily oxidize SA amino acid side chains without disrupting the structure when the exposure is limited.

1.5.1 Methods for Generating Hydroxyl Radical Labels for Footprinting

Multiple methods may be employed for generating the ·OH radicals for footprinting experiments, which was first used to identify protein binding sites on DNA. The ·OH non-specifically cleave surface exposed DNA along the backbone, which was blocked where the protein was bound. This work used a Fenton reagent, generally catalytic iron(II), to generate the ·OH from hydrogen peroxide (H_2O_2).[78] The benefit of using this method for ·OH radical generation is the low cost and wide availability of the required reagents, though the EDTA used to neutralize the Fenton reagent's charge is thought to bias reactivity.[79] Another method, pulse radiolysis of water, generates ·OH through interactions with high energy particles from synchrotron X-ray or γ-ray beams.[79,80] Since the ·OH are generated from water in the sample there is no need for added chemicals that may bias the results, but the need for high energy radiation sources severely limits the accessibility of this method. Photolysis employs UV radiation that homolytically cleaves H_2O_2 for ·OH generation. However, prolonged exposure to H_2O_2 may cause protein unfolding, biasing the results.[81]

1.5.2 Fast Photochemical Oxidation of Proteins

Fast photochemical oxidation of proteins (FPOP), used for the work presented here, generates ·OH through laser induced flash photolysis of H_2O_2 (Figure 1.2). This method minimizes the exposure time to H_2O_2 by several means. A KrF excimer laser is used to deliver high energy pulses of 248 nm UV light for H_2O_2 cleavage. A microliter flow system is used to expose the samples to the radiation source so that only a small bolus of sample is exposed at any given time and that each bolus is exposed to only one pulse; this ensures uniform sample exposure to the radiation source. The H_2O_2 is added to the sample immediately before irradiation, and controls that are not exposed to the radiation source are used to account for any oxidation caused by exposure to H_2O_2. Additionally, a radical scavenger, glutamine, is added to the sample to reduce the radical lifetime to less than 1 microsecond to safeguard against radical induced protein unfolding.[1,82] Moreover,

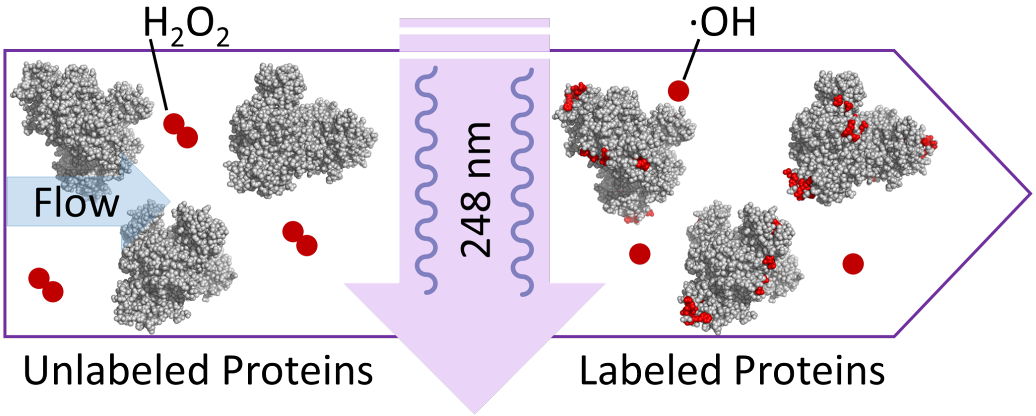


Figure 1.2. FPOP Schematic where UV laser pulses split H_2O_2 forming the ·OH that label the solvent exposed areas of the proteins.

the samples are immediately exposed to a second radical scavenger, methionine, and catalase as the exit the flow system to eliminate any remaining ·OH or H_2O_2 in the sample. This results in a method that generates a concentration of ·OH high enough to sufficiently oxidize SA areas of the protein in short bursts while any excess ·OH should be consumed before protein unfolding is induced. These properties have made FPOP increasingly popular in probing protein SA.

Although the method was published only a decade ago, the FPOP method[14] has been cited 166 times to date and used to evaluate several protein systems with a wide range of complexity. One popular use of FPOP is to probe protein conformational changes that occur upon ligand binding. Calcium-bound calmodulin (Cam), for example, was shown to undergo similar conformational changes when complexed with two separate peptides, Cam-Mel and CaM-Mas, as it does with M13, a complex with a resolved high resolution structure.[68] This method has also been used to characterize membrane proteins, with different labeling trends observed between the lipid embedded core and aqueous exposed regions of bacteriorhodopsin[58], and identifying a second heparin binding site on the Roundabout homolog 1 protein.[83]

While effective for characterizing protein-ligand interactions, FPOP has recently been used to evaluate more complex proteins and interactions. One such case is characterization of protein therapeutics. The advent of these types of pharmaceuticals has given rise to unique issues not present in small molecule therapeutics. Contrary to small molecule drugs, the safety and efficacy of protein therapeutics can vary between different conformations or if they undergo oligomerization.[84,85] Additionally, the analytical strategies used for evaluation of biosimilars (generics) have to demonstrate that differences between the branded and biosimilar protein therapeutic are not clinically relevant, as they cannot be characterized as identical to the branded form.[84] As such, FPOP has been established as a method that can provide higher order structural information for strengthening and protecting intellectual property as well as lot-to-lot quality assessment of protein therapeutics.[84–89] In addition to protein therapeutics, FPOP has recently been used for the structural characterization of proteins in live cells. This method, called in cell FPOP (IC-FPOP), was used to modify proteins in live African green monkey kidney (Vero) cells.[13] In this proof-of-concept paper, they were able to identify 105 FPOP modified proteins in a variety of cellular locations, demonstrating IC-FPOP as a novel application for the structural characterization of proteins within their native environment.

The work presented in the following chapters aims to expand the usability of FPOP, moving analysis by the method beyond single proteins or small interaction studies. The focus of these studies is to work toward faster data analysis, better identification of oxidized species, and bigger and more complex samples including IC-FPOP. A new data analysis platform, multidimensional chromatographic separation, and custom flow system are discussed, each of which can play a role in moving this field towards the next generation of protein footprinting.

1.6 References

1. David M. Hambly and Michael L. Gross. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *Journal of the American Society for Mass Spectrometry*, 16(12): 2057–2063, 2005. ISSN 1044-0305. doi: 10.1016/j.jasms.2005.09.008.
2. Brian C. Gau, Jiawei Chen, and Michael L. Gross. Fast photochemical oxidation of proteins for comparing solvent-accessibility changes accompanying protein folding: Data processing and application to barstar. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(6):1230–1238, 2013. ISSN 1570-9639. doi: 10.1016/j.bbapap.2013.02.023.
3. Guozhong Xu and Mark R. Chance. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews*, 107(8): 3514–3543, 2007. ISSN 0009-2665. doi: 10.1021/cr0682047.
4. Brian C. Gau, Hao Chen, Yun Zhang, and Michael L. Gross. Sulfate radical anion as a new reagent for fast photochemical oxidation of proteins. *Analytical Chemistry*, 82(18):7821–7827, 2010. ISSN 0003-2700. doi: 10.1021/ac101760y.
5. Shao-En Ong, Blagoy Blagoev, Irina Kratchmarova, Dan Bach Kristensen, Hanno Steen, Akhilesh Pandey, and Matthias Mann. Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, 1(5):376–386, 2002. ISSN 1535-9476.
6. Parminder Kaur, Janna G. Kiselar, and Mark R. Chance. Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. *Analytical Chemistry*, 81(19):8141–8149, 2009. ISSN 0003-2700. doi: 10.1021/ac9013644.
7. David N. Perkins, Darryl J. C. Pappin, David M. Creasy, and John S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS*, 20(18):3551–3567, 1999. ISSN 1522-2683. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.
8. M. Bern, Y. Cai, and D. Goldberg. Lookup peaks: A hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal Chem*, 79(4):1393–400, 2007. ISSN 0003-2700 (Print) 0003-2700. doi: 10.1021/ac0617013.
9. Stephen Tanner, Hongjun Shu, Ari Frank, Ling-Chi Wang, Ebrahim Zandi, Marc Mumby, Pavel A. Pevzner, and Vineet Bafna. Inspect:□ identification of posttranslationally modified peptides from tandem mass spectra. *Analytical Chemistry*, 77(14):4626–4639, 2005. ISSN 0003-2700. doi: 10.1021/ac050102d.
10. O. Charvatova, B. L. Foley, M. W. Bern, J. S. Sharp, R. Orlando, and R. J. Woods. Quantifying protein interface footprinting by hydroxyl radical oxidation and molecular dynamics simulation: Application to galectin-1. *J Am Soc Mass Spectrom*, 19(11):1692–705, 2008. ISSN 1044-0305 (Print) 1044-0305. doi: 10.1016/j.jasms.2008.07.013.

11. Carlee S. McClintock, Jerry M. Parks, Marshall Bern, Pavan K. GhattyVenkataKrishna, and Robert L. Hettich. Comparative informatics analysis to evaluate site-specific protein oxidation in multidimensional lc–ms/ms data. *Journal of Proteome Research*, 12(7):3307–3316, 2013. ISSN 1535-3893. doi: 10.1021/pr400141p.

12. Aimee Rinas and Lisa M Jones. Fast photochemical oxidation of proteins coupled to multidimensional protein identification technology (mudpit): Expanding footprinting strategies to complex systems. *Journal of The American Society for Mass Spectrometry*, 26(4):540–546, 2015. ISSN 1044-0305.

13. Jessica A. Espino, Vishaal S. Mali, and Lisa M. Jones. In cell footprinting coupled with mass spectrometry for the structural analysis of proteins in live cells. *Analytical Chemistry*, 87(15):7971–7978, 2015. ISSN 0003-2700. doi: 10.1021/acs.analchem.5b01888.

14. DavidM Hambly and MichaelL Gross. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *Journal of the American Society for Mass Spectrometry*, 16(12): 2057–2063, 2005. ISSN 1044-0305. doi: 10.1016/j.jasms.2005.09.008.

15. E. de Hoffmann and V. Stroobant. *Mass Spectrometry: Principles and Applications*. Wiley, 3rd edition, 2007. ISBN 9780470033104.

16. Ruedi Aebersold and Matthias Mann. Mass spectrometry-based proteomics. *Nature*, 422(6928):198–207, 2003. ISSN 0028-0836.

17. JB Fenn, M Mann, CK Meng, SF Wong, and CM Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926): 64–71, 1989. doi: 10.1126/science.2675315.

18. Franz Hillenkamp, Michael Karas, Ronald C. Beavis, and Brian T. Chait. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical Chemistry*, 63(24):1193A–1203A, 1991. ISSN 0003-2700. doi: 10.1021/ac00024a716.

19. Masamichi Yamashita and John B. Fenn. Electrospray ion source. another variation on the free-jet theme. *The Journal of Physical Chemistry*, 88(20): 4451–4459, 1984. ISSN 0022-3654. doi: 10.1021/j150664a002.

20. Guoan Zhang, Roland S Annan, Steven A Carr, and Thomas A Neubert. Overview of peptide and protein analysis by mass spectrometry. *Current Protocols in Molecular Biology*, pages 10.21. 1–10.21. 30, 2014. ISSN 0471142727.

21. Ruedi Aebersold and David R. Goodlett. Mass spectrometry in proteomics. *Chemical Reviews*, 101(2):269–296, 2001. ISSN 0009-2665. doi: 10.1021/cr990076h.

22. Evan Mason. Electrospray ionization spectroscopy, 21 January 2015 2015.

23. Geoffrey Taylor. Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 280(1382):383–397, 1964. doi: 10.1098/rspa.1964.0151.

24. Matthias Wilm. Principles of electrospray ionization. *Molecular & Cellular Proteomics*, 10(7), 2011. doi: 10.1074/mcp.M111.009407.

25. Lars Konermann, Elias Ahadi, Antony D. Rodriguez, and Siavash Vahidi. Unraveling the mechanism of electrospray ionization. *Analytical Chemistry*, 85(1): 2–9, 2013. ISSN 0003-2700. doi: 10.1021/ac302789c.

26. Matthias Wilm and Matthias Mann. Analytical properties of the nanoelectrospray ion source. *Analytical Chemistry*, 68(1):1–8, 1996. ISSN 0003-2700. doi: 10.1021/ac9509519.

27. Thomas R. Covey, Bruce A. Thomson, and Bradley B. Schneider. Atmospheric pressure ion sources. *Mass Spectrometry Reviews*, 28(6):870–897, 2009. ISSN 1098-2787. doi: 10.1002/mas.20246.

28. Anthonius A. M. Heemskerk, Jean-Marc Busnel, Bart Schoenmaker, Rico J. E. Derks, Oleg Klychnikov, Paul J. Hensbergen, André M. Deelder, and Oleg A. Mayboroda. Ultra-low flow electrospray ionization-mass spectrometry for improved ionization efficiency in phosphoproteomics. *Analytical Chemistry*, 84 (10):4552–4559, 2012. ISSN 0003-2700. doi: 10.1021/ac300641x.

29. P. Edman. A method for the determination of amino acid sequence in peptides. *Arch Biochem*, 22(3):475, 1949.

30. Hanno Steen and Matthias Mann. The abc's (and xyz's) of peptide sequencing. *Nat Rev Mol Cell Biol*, 5(9):699–711, 2004. ISSN 1471-0072. doi: doi:10.1038/nrm1468.

31. Alexander Makarov, Eduard Denisov, Alexander Kholomeev, Wilko Balschun, Oliver Lange, Kerstin Strupat, and Stevan Horning. Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Analytical Chemistry*, 78 (7):2113–2120, 2006. ISSN 0003-2700. doi: 10.1021/ac0518811.

32. Richard H. Perry, R. Graham Cooks, and Robert J. Noll. Orbitrap mass spectrometry: Instrumentation, ion motion and applications. *Mass Spectrometry Reviews*, 27(6):661–699, 2008. ISSN 1098-2787. doi: 10.1002/mas.20186.

33. Roman A. Zubarev and Alexander Makarov. Orbitrap mass spectrometry. *Analytical Chemistry*, 85(11):5288–5296, 2013. ISSN 0003-2700. doi: 10.1021/ac4001223.

34. FW McLafferty. Tandem mass spectrometry. *Science*, 214(4518):280–287, 1981. doi: 10.1126/science.7280693.

35. D F Hunt, J R Yates, J Shabanowitz, S Winston, and C R Hauer. Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences*, 83(17):6233–6237, 1986.

36. Klaus Biemann and Ioannis A. Papayannopoulos. Amino acid sequencing of proteins. *Accounts of Chemical Research*, 27(11):370–378, 1994. ISSN 0001-4842. doi: 10.1021/ar00047a008.

37. Vicki H. Wysocki, Katheryn A. Resing, Qingfen Zhang, and Guilong Cheng. Mass spectrometry of peptides and proteins. *Methods*, 35(3):211–222, 2005. ISSN 1046-2023. doi: 10.1016/j.ymeth.2004.08.013.

38. A. M. Falick, W. M. Hines, K. F. Medzihradszky, M. A. Baldwin, and B. W. Gibson. Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 4(11):882–893, 1993. ISSN 1044-0305. doi: 10.1016/1044-0305(93)87006-X.

39. Ioannis A Papayannopoulos. The interpretation of collision□induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews*, 14(1): 49–73, 1995. ISSN 1098-2787.

40. Simin D. Maleknia and Richard Johnson. *Mass Spectrometry of Amino Acids and Proteins*, pages 1–50. Wiley-VCH Verlag GmbH & Co. KGaA, 2011. ISBN 9783527631841. doi: 10.1002/9783527631841.ch1.

41. Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank www.rcsb.org. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235.

42. James M Holton and Kenneth A Frankel. The minimum crystal size needed for a complete diffraction data set. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):393–408, 2010. ISSN 0907-4449.

43. Amos Bairoch and Rolf Apweiler. The swiss-prot protein sequence database and its supplement trembl in 2000. *Nucleic Acids Research*, 28(1):45–48, 2000. ISSN 0305-1048 1362-4962.

44. Sébastien Boutet, Lukas Lomb, Garth J. Williams, Thomas R. M. Barends, Andrew Aquila, R. Bruce Doak, Uwe Weierstall, Daniel P. DePonte, Jan Steinbrener, Robert L. Shoeman, Marc Messerschmidt, Anton Barty, Thomas A. White, Stephan Kassemeyer, Richard A. Kirian, M. Marvin Seibert, Paul A. Montanez, Chris Kenney, Ryan Herbst, Philip Hart, Jack Pines, Gunther Haller, Sol M. Gruner, Hugh T. Philipp, Mark W. Tate, Marianne Hromalik, Lucas J. Koerner, Niels van Bakel, John Morse, Wilfred Ghonsalves, David Arnlund, Michael J. Bogan, Carl Caleman, Raimund Fromme, Christina Y. Hampton, Mark S. Hunter, Linda C. Johansson, Gergely Katona, Christopher Kupitz, Mengning Liang, Andrew V. Martin, Karol Nass, Lars Redecke, Francesco Stellato, Nicusor Timneanu, Dingjie Wang, Nadia A. Zatsepin, Donald Schafer, James Defever, Richard Neutze, Petra Fromme, John C. H. Spence, Henry N. Chapman, and Ilme Schlichting. High-resolution protein structure determination by serial femtosecond crystallography. *Science*, 337(6092):362–364, 2012. doi: 10.1126/science.1217737.

45. Alessandro Ori, Amparo Andrés-Pons, and Martin Beck. *Chapter 6 - the Use of Targeted Proteomics to Determine the Stoichiometry of Large Macromolecular Assemblies*, volume Volume 122, pages 117–146. Academic Press, 2014. ISBN 0091-679X. doi: 10.1016/B978-0-12-417160-2.00006-0.

46. Thomas R. M. Barends, Lutz Foucar, Sabine Botha, R. Bruce Doak, Robert L. Shoeman, Karol Nass, Jason E. Koglin, Garth J. Williams, Sebastien Boutet, Marc Messerschmidt, and Ilme Schlichting. De novo protein crystal structure determination from x-ray free-electron laser data. *Nature*, 505(7482):244–247, 2014. ISSN 0028-0836. doi: 10.1038/nature12773.

47. Ian R. Kleckner and Mark P. Foster. An introduction to nmr-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1814(8):942–968, 2011. ISSN 1570-9639. doi: 10.1016/j.bbapap.2010.10.012.

48. S. M. Kelly, T. J. Jess, and N. C. Price. How to study proteins by circular dichroism. *Biochim Biophys Acta*, 1751(2):119–39, 2005. ISSN 0006-3002 (Print) 0006-3002. doi: 10.1016/j.bbapap.2005.06.005.

49. Joseph R. Lakowicz. *Protein Fluorescence*, pages 529–575. Springer US, Boston, MA, 2006. ISBN 978-0-387-46312-4. doi: 10.1007/978-0-387-46312-4_16.

50. Maurice R. Eftink. *Intrinsic Fluorescence of Proteins*, volume 6, pages 1–15. Springer US, Boston, MA, 2000. ISBN 978-0-306-47102-5. doi: 10.1007/0-306-47102-7_1.

51. J. B. Alexander Ross, Elena Rusinova, Linda A. Luck, and Kenneth W. Rousslang. *Spectral Enhancement of Proteins by in Vivo Incorporation of Tryptophan Analogues*, pages 17–42. Springer US, Boston, MA, 2000. ISBN 978-0-306-47102-5. doi: 10.1007/0-306-47102-7_2.

52. Yuling Yan and Gerard Marriott. Analysis of protein interactions using fluorescence technologies. *Current Opinion in Chemical Biology*, 7(5):635–640, 2003. ISSN 1367-5931. doi: 10.1016/j.cbpa.2003.08.017.

53. Jacqueline LS Milne, Mario J Borgnia, Alberto Bartesaghi, Erin EH Tran, Lesley A Earl, David M Schauder, Jeffrey Lengyel, Jason Pierson, Ardan Patwardhan, and Sriram Subramaniam. Cryo□electron microscopy–a primer for the non□microscopist. *FEBS Journal*, 280(1):28–45, 2013. ISSN 1742-4658.

54. Xiao-chen Bai, Chuangye Yan, Guanghui Yang, Peilong Lu, Dan Ma, Linfeng Sun, Rui Zhou, Sjors H. W. Scheres, and Yigong Shi. An atomic structure of human [ggr]-secretase. *Nature*, 525(7568):212–217, 2015. ISSN 0028-0836. doi: 10.1038/nature14892.

55. Liwen Wang and Mark R. Chance. Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Analytical Chemistry*, 83(19): 7234–7241, 2011. ISSN 0003-2700. doi: 10.1021/ac200567u.

56. Andrew N. Hoofnagle, Katheryn A. Resing, and Natalie G. Ahn. Protein analysis by hydrogen exchange mass spectrometry. *Annual Review of Biophysics and Biomolecular Structure*, 32(1):1–25, 2003. doi: doi:10.1146/annurev.biophys.32.110601.142417.

57. Yan Pan, Leonid Brown, and Lars Konermann. Mapping the structure of an integral membrane protein under semi-denaturing conditions by laser-induced oxidative labeling and mass spectrometry. *Journal of Molecular Biology*, 394 (5):968–981, 2009. ISSN 0022-2836. doi: 10.1016/j.jmb.2009.09.063.

58. Yan Pan, Bradley B. Stocks, Leonid Brown, and Lars Konermann. Structural characterization of an integral membrane protein in its natural lipid environment by oxidative methionine labeling and mass spectrometry. *Analytical Chemistry*, 81(1):28–35, 2009. ISSN 0003-2700. doi: 10.1021/ac8020449.

59. Taylor A Poor, Lisa M Jones, Amika Sood, George P Leser, Manolo D Plasencia, Don L Rempel, Theodore S Jardetzky, Robert J Woods, Michael L Gross, and Robert A Lamb. Probing the paramyxovirus fusion (f) protein-refolding event from pre-to postfusion by oxidative footprinting. *Proceedings of the National Academy of Sciences*, 111(25):E2596–E2605, 2014. ISSN 0027-8424.

60. Antonio N. Calabrese, James R. Ault, Sheena E. Radford, and Alison E. Ashcroft. Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. *Methods*, 2015. ISSN 1046-2023. doi: 10.1016/j.ymeth.2015.02.018.

61. Lars Konermann, Bradley B. Stocks, Yan Pan, and Xin Tong. Mass spectrometry combined with oxidative labeling for exploring protein structure and folding. *Mass Spectrometry Reviews*, 29(4):651–667, 2010. ISSN 1098-2787. doi: 10.1002/mas.20256.

62. Xiaoduan Ye, Patrick K. O'Neil, Adrienne N. Foster, Michal J. Gajda, Jan Kosinski, Michal A. Kurowski, Janusz M. Bujnicki, Alan M. Friedman, and Chris Bailey-Kellogg. Probabilistic cross-link analysis and experiment planning for high-throughput elucidation of protein structure. *Protein Science*, 13(12):3298–3313, 2004. ISSN 1469-896X. doi: 10.1110/ps.04846604.

63. David Paramelle, Guillaume Miralles, Gilles Subra, and Jean Martinez. Chemical cross-linkers for protein structure studies by mass spectrometry. *PROTEOMICS*, 13(3-4):438–456, 2013. ISSN 1615-9861. doi: 10.1002/pmic.201200305.

64. Xiaoting Tang and James E. Bruce. *Chemical Cross-Linking for Protein–Protein Interaction Studies*, pages 283–293. Humana Press, Totowa, NJ, 2009. ISBN 978-1-59745-493-3. doi: 10.1007/978-1-59745-493-3_17.

65. C. Plaschka, L. Lariviere, L. Wenzeck, M. Seizl, M. Hemann, D. Tegunov, E. V. Petrotchenko, C. H. Borchers, W. Baumeister, F. Herzog, E. Villa, and P. Cramer. Architecture of the rna polymerase ii-mediator core initiation complex. *Nature*, 518(7539):376–380, 2015. ISSN 0028-0836. doi: 10.1038/nature14229.

66. Andrew N. Holding. Xl-ms: Protein cross-linking coupled with mass spectrometry. *Methods*, 89:54–63, 2015. ISSN 1046-2023. doi: 10.1016/j.ymeth.2015.06.010.

67. Yuko Tsutsui and Patrick L Wintrode. Hydrogen/deuterium exchange-mass spectrometry: A powerful tool for probing protein structure, dynamics and interactions. *Current medicinal chemistry*, 14(22):2344–2358, 2007. ISSN 0929-8673.

68. Hao Zhang, Brian C. Gau, Lisa M. Jones, Ilan Vidavsky, and Michael L. Gross. Fast photochemical oxidation of proteins for comparing structures of protein−ligand complexes: The calmodulin−peptide model system. *Analytical Chemistry*, 83(1):311–318, 2011. ISSN 0003-2700. doi: 10.1021/ac102426d.

69. Lisa M. Jones, Justin B. Sperry, James A. Carroll, and Michael L. Gross. Fast photochemical oxidation of proteins for epitope mapping. *Analytical Chemistry*, 83(20):7657–7661, 2011. ISSN 0003-2700. doi: 10.1021/ac2007366.

70. S. Walter Englander, Tobin R. Sosnick, Joan J. Englander, and Leland Mayne. Mechanisms and uses of hydrogen exchange. *Current Opinion in Structural Biology*, 6(1):18–23, 1996. ISSN 0959-440X. doi: 10.1016/S0959-440X(96)80090-X.

71. Thomas E. Wales and John R. Engen. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrometry Reviews*, 25(1):158–170, 2006. ISSN 1098-2787. doi: 10.1002/mas.20064.

72. Lars Konermann, Jingxi Pan, and Yu-Hong Liu. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chemical Society Reviews*, 40(3):1224–1234, 2011. ISSN 0306-0012. doi: 10.1039/C0CS00113A.

73. Elien Vandermarliere, Elisabeth Stes, Kris Gevaert, and Lennart Martens. Resolution of protein structure by mass spectrometry. *Mass Spectrometry Reviews*, pages n/a–n/a, 2015. ISSN 1098-2787. doi: 10.1002/mas.21450.

74. Vanessa Leah Mendoza and Richard W. Vachet. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrometry Reviews*, 28(5):785–815, 2009. ISSN 1098-2787. doi: 10.1002/mas.20203.

75. Guozhong Xu and Mark R. Chance. Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Analytical Chemistry*, 77(14):4549–4555, 2005. ISSN 0003-2700. doi: 10.1021/ac050299+.

76. Simin D. Maleknia, Michael Brenowitz, and Mark R. Chance. Millisecond radiolytic modification of peptides by synchrotron x-rays identified by mass spectrometry. *Analytical Chemistry*, 71(18):3965–3973, 1999. ISSN 0003-2700. doi: 10.1021/ac990500e.

77. Warren M. Garrison. Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chemical Reviews*, 87(2):381–398, 1987. ISSN 0009-2665. doi: 10.1021/cr00078a006.

78. H. J. H. Fenton. Lxxiii.-oxidation of tartaric acid in presence of iron. *Journal of the Chemical Society, Transactions*, 65(0):899–910, 1894. ISSN 0368-1645. doi: 10.1039/CT8946500899.

79. Keiji Takamoto and Mark R. Chance. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annual Review of Biophysics and Biomolecular Structure*, 35(1):251–276, 2006. doi: doi:10.1146/annurev.biophys.35.040405.102050.

80. Klaus-Dieter Asmus. Pulse radiolysis methodology. *Methods in enzymology*, 105:167–178, 1983. ISSN 0076-6879.

81. Joshua S. Sharp, Jeffrey M. Becker, and Robert L. Hettich. Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry. *Analytical Chemistry*, 76(3):672–683, 2004. ISSN 0003-2700. doi: 10.1021/ac0302004.

82. Brian C. Gau, Joshua S. Sharp, Don L. Rempel, and Michael L. Gross. Fast photochemical oxidation of protein footprints faster than protein unfolding. *Analytical Chemistry*, 81(16):6563–6571, 2009. ISSN 0003-2700. doi: 10.1021/ac901054w.

83. Zixuan Li, Heather Moniz, Shuo Wang, Annapoorani Ramiah, Fuming Zhang, Kelley W. Moremen, Robert J. Linhardt, and Joshua S. Sharp. High structural resolution hydroxyl radical protein footprinting reveals an extended robo1-heparin binding interface. *Journal of Biological Chemistry*, 290(17):10729–10740, 2015. doi: 10.1074/jbc.M115.648410.

84. Steven A. Berkowitz, John R. Engen, Jeffrey R. Mazzeo, and Graham B. Jones. Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *Nat Rev Drug Discov*, 11(7):527–540, 2012. ISSN 1474-1776.

85. Caroline Watson and Joshua S. Sharp. Conformational analysis of therapeutic proteins by hydroxyl radical protein footprinting. *The AAPS Journal*, 14(2):206–217, 2012. ISSN 1550-7416. doi: 10.1208/s12248-012-9336-7.

86. Galahad Deperalta, Melissa Alvarez, Charity Bechtel, Ken Dong, Ross McDonald, and Victor Ling. Structural analysis of a therapeutic monoclonal antibody dimer by hydroxyl radical footprinting. *mAbs*, 5(1):86–101, 2013. ISSN 1942-0862. doi: 10.4161/mabs.22964.

87. Lisa M. Jones, Hao Zhang, Weidong Cui, Sandeep Kumar, Justin B. Sperry, James A. Carroll, and Michael L. Gross. Complementary ms methods assist conformational characterization of antibodies with altered s–s bonding networks. *Journal of The American Society for Mass Spectrometry*, 24(6):835–845, 2013. ISSN 1879-1123. doi: 10.1007/s13361-013-0582-4.

88. Hao Zhang, Weidong Cui, and Michael L. Gross. Mass spectrometry for the biophysical characterization of therapeutic monoclonal antibodies. *FEBS Letters*, 588(2):308–317, 2014. ISSN 1873-3468. doi: 10.1016/j.febslet.2013.11.027.

89. Ying Zhang, Don L. Rempel, Hao Zhang, and Michael L. Gross. An improved fast photochemical oxidation of proteins (fpop) platform for protein therapeutics. *Journal of The American Society for Mass Spectrometry*, 26(3):526–529, 2014. ISSN 1879-1123. doi: 10.1007/s13361-014-1055-0.

2. AN EFFICIENT QUANTITATION STRATEGY FOR HYDROXYL RADICAL-MEDIATED PROTEIN FOOTPRINTING USING PROTEOME DISCOVERER

2.1 Introduction

Despite the advantages of ·OH footprinting, data analysis for this method remains arduous, limiting the potential it has in the field of structural proteomics. Analysis generally follows a typical proteomics workflow; proteins are proteolyzed, the resulting peptides are detected and identified using data-dependent LC-MS/MS, and precursor peak intensities or areas are used for quantitation.[1] However, the large number of possible oxidation products produced by this method makes identification and quantitation significantly more complex than a proteomics counterpart such as SILAC, where stable isotope ratios are quantified.[2–4] Furthermore, the vast quantity of data generated from these experiments makes manual analysis cumbersome, which can lead to errors in interpretation.[5]

Several efforts have been made to expedite analysis of these data sets. Gau et al. have developed a semiautomatic method that utilizes a Mascot error-tolerant search and an in house built Excel spreadsheet for identification and quantitation.[6,7] However, this method is manually intensive, and the use of an error-tolerant search prohibits the use of a decoy database search and, subsequently, a false discovery rate (FDR) or a level of confidence in identifications. Hybrid software platforms including ByOnic and InsPecT, that were developed for identifying post translational modifications, have been utilized for ·OH experiments.[8–11] While these platforms offer a better-quality database search, quantitation remains manually intensive. And although Kaur and coworkers[5] developed an automated software platform, ProtMapMS (PM), it was specifically designed for dose response studies from water radiolysis. Furthermore, the presupposition that the protein of interest is known excludes the platform from being used in highly complex samples, such as cell lysate and in cell labeling strategies that have been developed recently.[12,13] And while the resolving power of ProtMapMS has been improved in the most current version, the limitations that exclude its use on complex samples remains.[14,15] For each of these platforms, quantitation of the oxidation yield is performed using the ratio of the identified oxidized species to the sum of the oxidized and corresponding unoxidized species. However, the oxidation yield can also be determined indirectly by monitoring the decrease in the unoxidized species, whereby this decrease is attributed to the species being oxidized.[16] While this has been shown to be quantitatively accurate, this approach can only be used at the peptide level, limiting the resolution of the data.

2.2 Experimental Section

2.2.1 Materials

All chemicals were obtained from Thermo Fisher Scientific (Waltham, MA) unless otherwise noted.

2.2.2 Protein Expression and Purification

The pRSET vector containing GCamP2 was a generous gift from Michael Kotlikoff (Cornell University). The expression and purification of GCaMP2 was previously described.[17] Briefly, transformed BL21(DE3)pLysS (Promega, Madison, WI) competent E. coli cells were grown in lysogeny broth media. After induction with 1 mM isopropyl β-D-1-thiogalactopyranoside, proteins were expressed for 12 hrs at 18°C. The proteins were purified using HisPur Ni-NTA agarose resin.

2.2.3 Oxidative Labeling

Each 50 μL sample contained 10 mM phosphate buffered saline (PBS, Sigma Aldrich, St. Louis, MO) 10 mM -glutamine, 7.5 mM hydrogen peroxide, 10 mM ethylenediaminetetraacetic acid (calcium free, CF) or 10 mM calcium chloride (calcium bound, CB), and purified GCaMP2 at a concentration of 0.18 mg/mL. The hydrogen peroxide was added just prior to infusion. FPOP was performed similarly as described.[12,18,19] A 248 nm KrF excimer laser (GAM Laser Inc., Orlando, FL) was used to irradiate the sample solution at 135 mJ/pulse. The laser was focused through a 250 mm plano convex lens (Thorlabs, Inc., Newton, NJ) onto 150 μm i.d. fused silica tubing (Polymicro Technologies, Pheonix, AZ) with the polyimide coating removed, giving a 2.5 mm irradiation window. The flow rate, 33 μL/min, was set to allow for a 20 percent exclusion fraction. A total of 3 FPOP samples and 3 controls (no irradiation) for each condition were prepared.

2.2.4 Proteolysis

Post FPOP, the GCaMP2 samples were subjected to a two-step digestion process as previously described.[12,20] Each sample was dried in a vacuum centrifuge, and resuspended in 8 M urea 100 mM Tris-HCL pH 8.5 buffer. Proteins were reduced with tris(2-carboxyethyl)phosphine, alkylated with iodoacetamide (IAA) and quenched with dithiothreitol. Lys-C was added at a 100:1 substrate to protease ratio and incubated overnight at 37 °C. The samples were then diluted with 100 mM Tris buffer to bring the urea concentration down to 2 M. Trypsin was added at a 20:1 substrate to protease ratio and incubated overnight at 37 °C. Digestion was quenched with formic acid (Sigma Aldrich, St. Louis, MO) at a final concentration of 5%. Samples were de-salted using NuTipC18 media tips (Glygen Corporation, Columbia, MD), dried in a vacuum centrifuge, and resuspended in 20 μL of 2% acetonitrile 0.1% formic acid.

2.2.5 LC-MS/MS Acquisition

Analysis was completed using an UltiMate 3000 RSLC and a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA) as previously described.[12] For each experiment, 1 µg of the digest was loaded onto a 2 cm Acclaim Pepmap 100 C18 trap column (Thermo Fisher Scientific, Waltham, MA) and washed for 3 minutes with loading buffer (2% acetonitrile 0.1% formic acid) at a flow rate of 5 µ”/min. The samples were separated on a 75 µm inner diameter reverse phase analytical column packed in-house with a 30 cm bed of Magic 5 µm C18 particles (Michrom Bioresources Inc., Auburn, CA). Peptides were eluted with a 74 minute linear gradient at a flow rate of 300 nL/min to 45% acetonitrile 0.1% formic acid. The total run time was 97 minutes including loading, washing, and equilibration time. MS1 spectra were acquired over an *m/z* range of 300–1500 at a resolving power of 70,000 for 400 *m/z* ions, with a dynamic exclusion of 20 s. The 25 most abundant ions were selected for MS2 at a resolving power of 17,500 for 400 *m/z* ions. Ions with a charge-state of +1 and >+6 ions were rejected. AGC targets were set to 3e6 for MS1 and 1e5 for data-dependent MS2 with an underfill ratio of 2.5%, giving an intensity threshold of 5.0e4.

2.2.6 Data Analysis

All data files were searched using Proteome Discoverer (version 1.4; Thermo Fisher Scientific, San Jose, CA, USA) with Mascot (version 2.4; Matrix Science, London, UK) and Sequest HT (version 1.1.1.11; Thermo Fisher Scientific, San Jose, CA, USA) using a custom multi-search node workflow (Figure 2.1), and the ProtMapMS platform (version 2.5.0.30). For the PD method, files were searched against a FASTA database consisting of the protein of interest (GCaMP2 synthetic construct; NCBI GI 218681839) inserted into the cRAP (Common Repository of Adventitious Proteins) database (version 2012.01.01; http://www.thegpm.org/crap/), and extracted ion chromatogram (EIC) areas for each peptide spectrum match

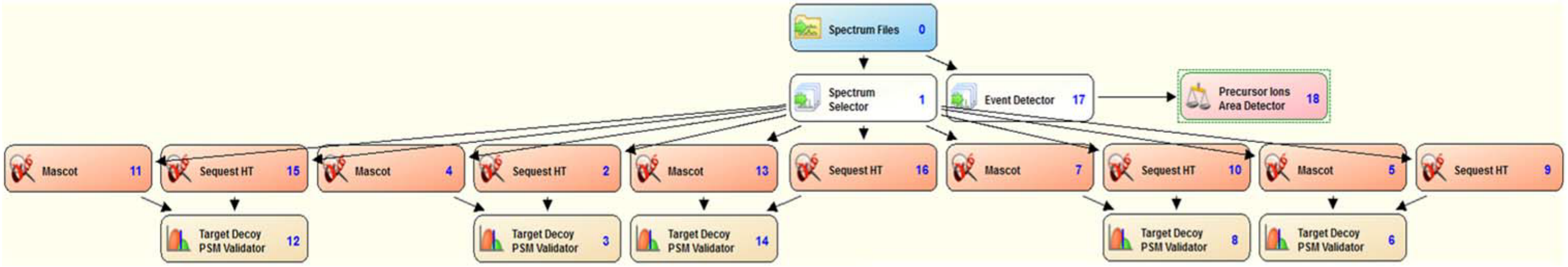

Spectrum Files 0
Spectrum Selector 1
Event Detector 17
Precursor Ions Area Detector 18
Mascot 11
Sequest HT 15
Mascot 4
Sequest HT 2
Mascot 13
Sequest HT 16
Mascot 7
Sequest HT 10
Mascot 5
Sequest HT 9
Target Decoy PSM Validator 12
Target Decoy PSM Validator 3
Target Decoy PSM Validator 14
Target Decoy PSM Validator 8
Target Decoy PSM Validator 6

Figure 2.1. PD workflow developed for FPOP analysis displaying the search hierarchy with Mascot and Sequest HT search levels (dark orange nodes), target-decoy validation (tan nodes), and precursor peak areas are calculation (pink node).

(PSM) were calculated. For the PM method, all files were searched against the GCaMP2 FASTA sequence only, restricted to tryptic peptides with up to one missed cleavage within a 350-5000 Da mass range. The precursor mass tolerance and the *m/z* peak integration width were set to ± 10 ppm, and the fragment mass tolerance to ± 0.02 Da.

Post search, peptides were ungrouped and filtered to a 1% FDR for both peptide and residue level analysis. The data were exported to Excel and summarized using the PowerPivot add-in that was customized for ·OH labeling experiments. The fractional oxidation per peptide or residue was determined according to the following equation:

$$\frac{\sum EIC\ area\ modified}{\sum EIC\ area} \tag{2.1}$$

where, for peptide level analysis, EIC area modified is the EIC area of a PSM containing an FPOP modification, and EIC area the EIC area of any PSM with a sequence identical to that containing the modification. For residue level analysis, EIC area modified is the EIC area of a PSM for a specific modified residue, and EIC area the EIC area of any PSMs with sequences identical to those containing the modification.

2.3 The Proteome Discoverer Analysis Strategy

The objective for the work presented in this chapter was to develop an identification and quantitation strategy, with minimal manual steps, for use in protein footprinting studies with four main features. First, we wanted to use an identification strategy that builds on the standards that have been established in the proteomics community. Second, we wanted the platform to be compatible with all levels of sample complexity. Third, we wanted a platform that would work with any stable footprinting label. Last, we wanted our platform to be built on software packages that are already widely used within the proteomics community. As follows, we created a workflow (Figure 2.1) in Thermo Scientific’s Proteome Discoverer (PD) that was customized for deep searching of ·OH modifications, as well as a rapid quantitation platform using the Excel PowerPivot add-in. Although our strategy was designed for ·OH labeling experiments, each half may be easily modified to accommodate any stable footprinting label.

The workflow (Figure 2.1) was constructed with five search algorithm levels, each restricted to tryptic peptides greater than 5 residues, with up to one missed cleavage, and within a 350-5000 Da mass range. The precursor mass tolerance was set to ± 10 ppm, and the fragment mass tolerance to ± 0.02 Da; these parameters are typical for a high resolution MS, such as the Thermo Scientific Q Exactive that was used for this work. Alkylation, a modification present on cysteine residues as a result of the addition of IAA during proteolysis, and all commonly observed ·OH modifications[2,3] were distributed across each search level (Table 2.1). Each raw file was searched against the FASTA database on each search level. After each search level, PSMs are subjected to target-decoy validation, with a 1% FDR setting for high confidence PSMs. The EICs for each precursor ion are calculated using the PD precursor ion area detection node.

2.4 Results

The synthetic construct, GCaMP2, was used to evaluate the efficacy of the PD method for both peptide and residue-level analysis. GCaMP2 belongs to a group of synthetic proteins known as genetically encoded indicators.[21] It is a calcium sensor assembled from circularly permutated enhanced green fluorescent protein (cpEGFP), the M13 helix of myosin light chain, and calmodulin (CaM).[22] The sensor undergoes conformational changes upon calcium binding that result in a change in solvent accessibility of the chromophore allowing for calcium dependent sensing in cells and tissues.[17,21] At low concentrations, comparable to those used in FPOP experiments, GCaMP2 exists as a monomer in both the calcium free (CF)

Table 2.1
Workflow Modification Distribution

Search Level	Amino Acid(s)	Mass Shift (Da)
1	C (Static)	57
2	A,D,E,F,H,I,K,L,M,N,P,Q,R,V,W,Y	16
3	E,I,K,L,P,Q,R,V	14
	C,F,M,W,Y	32
	H	-10
4	C,F,W,Y	48
	D,E	-28,-30
5	H	5
	R	-43
	D,E	-44
	H	-22,-23

and calcium bound (CB) states, with monomer crystal structures available for both states.[21] This makes GCaMP2 an ideal molecule for examining the ability of the PD method in discerning conformational changes via localized changes in oxidation, with the structures providing a means of validating the residue-level data.

2.4.1 Peptide Level FPOP Oxidation of GCaMP2

Peptide level FPOP oxidation of GCaMP2 was completed using both the PD method and the PM method. PD analysis resulted in the detection of 62 peptides, with 21 containing FPOP modifications (Table 2.2). There were a total of 443780 PSMs at a 1% FDR affiliated with the protein, with 11% corresponding to hydroxyl radical oxidation events. These peptides correspond to a sequence coverage of

Table 2.2
GCaMP2 Oxidized Peptides

Peptide	Calcium Free		Calcium Bound		Ri	SASA ($Å^2$)		
	Oxidation	Log PF	Oxidation	Log PF		3EKJ	3EK4	3EK7
32-41	0.091±0.028	5.676	0.109±0.048	5.675	35.08			
24-31	0.003±0.001	6.236	0.003±0.002	6.238	20.7			
53-69	0.700±0.145	5.090	0.521±0.351	5.093	61.77		585	1063
57-69	0.035±0.006	5.249	0.101±0.088	5.239	54.29		537	704
82-119	0.062±0.030	4.524	0.270±0.119	4.509	111.5	1543	1586	1686
128-151	0.055±0.041	4.537	0.398±0.142	4.517	110.2			
162-184	0.010±0.003	5.266	0.008±0.002	5.268	54.01	946	1023	1046
185-199	0.038±0.006	5.722	0.060±0.015	5.718	33.8	384	582	386
244-254	0.015±0.006	5.458	0.031±0.028	5.451	44.42	266	269	244
255-259	0.002±0.001	5.850	0.001±0.001	5.855	30.6	251	205	225
272-280	0.021±0.009	6.087	0.016±0.007	6.090	23.59	354	343	358
285-298	0.094±0.036	5.540	0.273±0.238	5.530	40.19	691	666	707
290-298	0.011±0.003	6.167	0.034±0.011	6.156	21.93	399	404	420
305-316	0.024±0.008	5.924	0.023±0.008	5.924	27.75	728		
305-324	0.032±0.023	5.160	0.183±0.160	5.143	59.4	1178		
317-324	0.014±0.003	5.797	0.008±0.002	5.802	31.65	450	302	353
379-389	0.007±0.005	5.678	0.021±0.011	5.667	35.91		777	890
381-393	0.020±0.011	5.911	0.001±0.001	5.942	28.14		940	923
394-409	0.139±0.083	5.475	0.006±0.005	5.506	42.73		1053	708
398-409	0.004±0.002	5.967	0.005±0.001	5.966	27.01		904	511
410-418	0.022±0.008	5.533	0.037±0.010	5.528	41.07			529
419-429	0.021±0.020	5.595	0.155±0.057	5.575	38.61			803

92% and a footprinting coverage of 66% of the protein (Figure 2.2a). By comparison, 83% of the sequence was covered using the PM method, with only 33% footprinting coverage (Figure 2.2a). There were 8 oxidized peptides identified in common between the two methods which were used to validate the accuracy of the PD method’s quantitation strategy. Recently, Chance and coworkers have in-

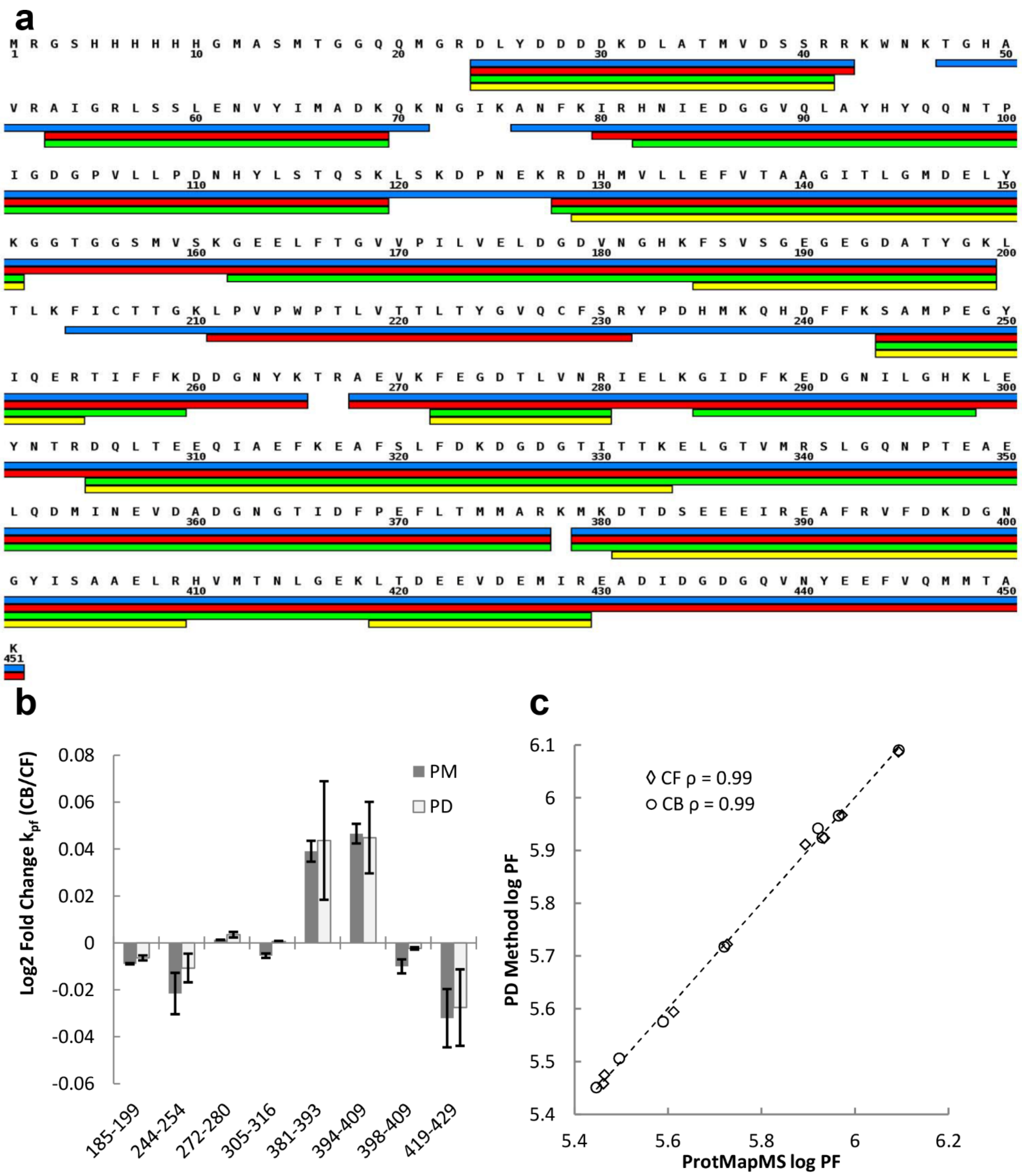


Figure 2.2. Comparison of PD method to PM method showing the sequence coverage (a) with PD total and footprinting coverage in blue and green, respectively, and PM total and footprinting coverage in red and yellow, respectively, as well as the agreement between the fold change (b) and correlation (c).

troduced an oxidation normalization factor, named protection factor (PF), where the relative intrinsic activity (Ri, Table 2.3) of the peptide (or residue for residue level analysis) is divided by the oxidation rate to account for the differing reactivities of each of the amino acids with ·OH, which we have applied to the data presented in this paper.[14,15] Comparison of the two methods shows a similar trend in fold change of oxidation (Figure 2.2b), and results in a correlation coefficient of 0.99 for both the CF and CB states (Figure 2.2c), with p-values of 3.59e-9 and 2.53e-9,

Table 2.3
Rate Constants and Relative Intrinsic Activities

Residue	k·OH (M^{-1} sec^{-1})	R_i
Cysteine	3.5e10	29.2
Methionine	8.5e9	20.5
Tryptophan	1.3e10	17.4
Tyrosine	1.3e10	12
Phenylalanine	6.9e9	11.2
Histidine	4.8e9	9.3
Isoleucine	1.8e9	4.4
Leucine	1.7e9	4.4
Arginine	3.5e9	2.9
Lysine	3.5e8	2.2
Valine	8.5e8	1.9
Threonine	5.1e8	1.6
Serine	3.2e8	1.4
Proline	6.5e8	1
Glutamine	5.4e8	0.69
Glutamic acid	2.3e8	0.66
Asparagine	4.9e7	0.44
Aspartic acid	7.5e7	0.42
Alanine	7.7e7	0.14
Glycine	1.7e7	0.04

respectively. These correlations demonstrate that the quantitative accuracy of the PD platform is comparable to the previously validated PM platform.

To further establish the efficacy of the PD method, the solvent accessible surface area (SASA) was calculated for each of the structures using VADAR[23], and compared to the natural log of the PFs for the 22 oxidized peptides identified (Figure 2.3), as previously demonstrated.[14,15] This resulted in a correlation of -0.84 for the CF state, and -0.59 to the CB monomeric structure (3EK4) and -0.75 to the CB dimeric structure (3EK7).[21] The higher correlation to the dimeric CB structure indicates that dimer was present in the CB samples, contrary to what was expected at the low concentration (0.18 µg/µL) of GCaMP2 in each sample prepared. This was most likely due to the method in which the samples were prepared. In order to minimize the variability between samples, calcium chloride was added to a concentrated stock of GCaMP2 which was later spiked into each sample tube. Studies completed by Akerboom and coworkers[21] on GCaMP2 demonstrated that dimeric CB GCaMP2 can be seen at concentrations of 10 µM and that the equilibrium kinetics between the monomeric and dimeric calcium-saturated forms was slow, with the solution remaining stable over several days. With the stock concentration of CB GCaMP2 for this study at 0.29 µM, nearly 3 times the concentration where dimer has been shown to exist, there is a strong likelihood that dimeric CB GCaMP2 was present in the samples at the time of oxidation.

Further discussion of GCaMP2 will be divided into two talking points: the GFP domain (which, for ease of discussion, also includes the M13 peptide and the domain linkers) and the CaM domain. There are two areas of the GFP domain that have unexpected differences in oxidation between the CF and CB states. The first occurs between the 2 peptides spanning residues 255-284, with a higher relative oxidation in the CF state. Alignment of the GFP domain (residues 62-301) of the two structures (PDB IDs 3EKJ and 3EK4) using Pymol results in an RMS deviation value of 0.35Å, indicating that there are only small structural changes in this domain between the two states.[21,24] To investigate further, the SASA was calculated for

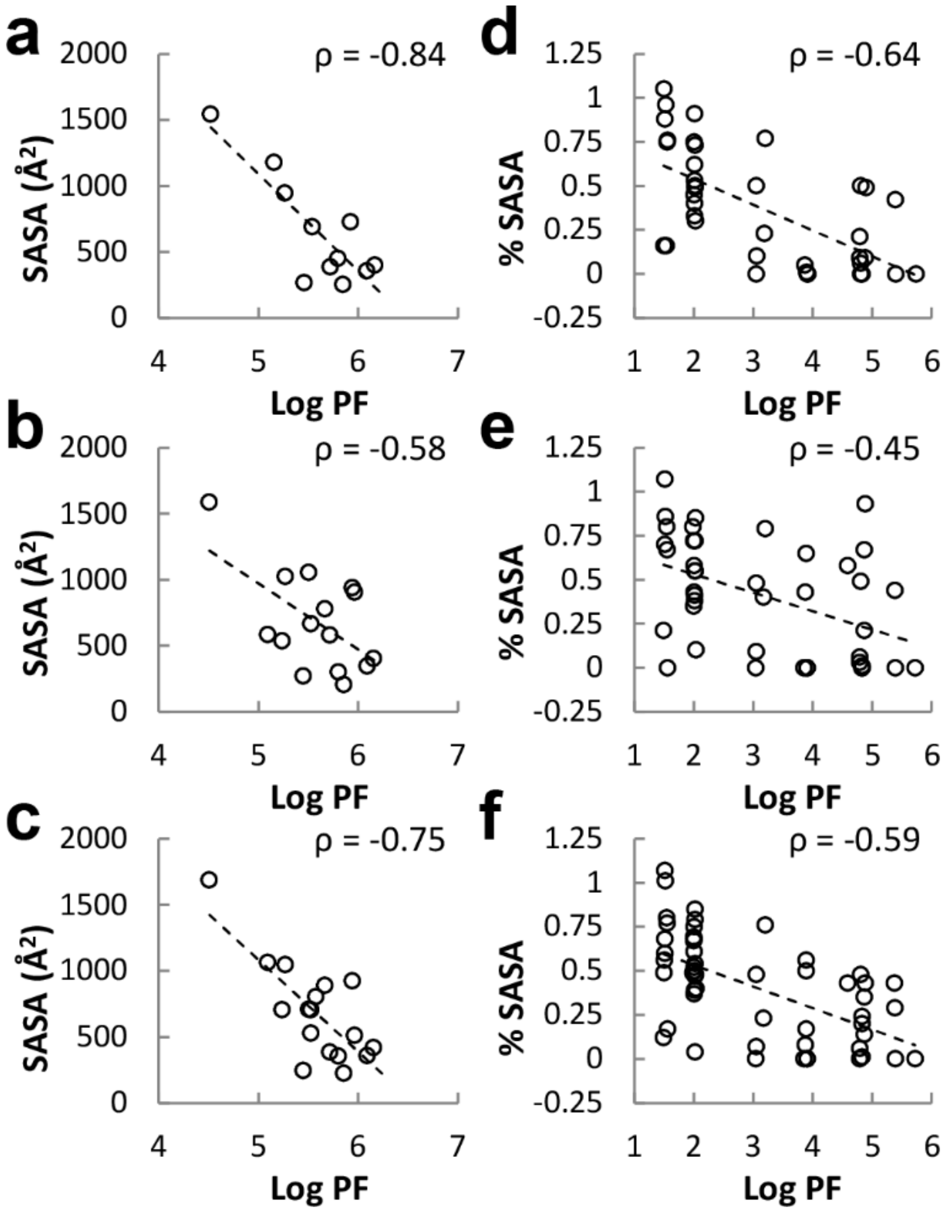


Figure 2.3. Correlation between SASA and log PF for both peptide level (a, b, and c) and residue level (d, e, and f) analysis for the CF state (a and d) and the CB monomeric (b and e) and dimeric (c and f) states.

each of the structures using VADAR.[23] Comparison of the SASA from residues 255-284 revealed that there is an increased SASA for the CF structure, in agreement with the FPOP oxidation analysis (Figure 2.4, Table 2.2). The 4% SASA difference

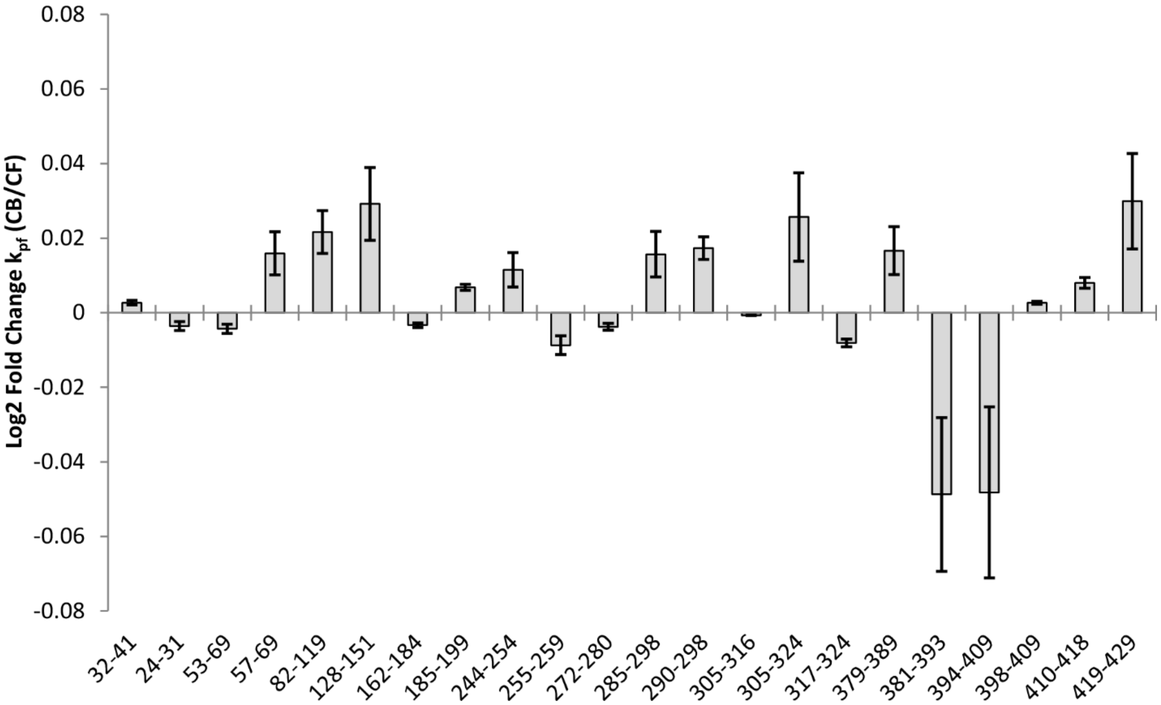


Figure 2.4. Fold change between the CF and CB states for the 22 oxidized peptides identified by the PD method.

between the states is mainly associated with 6 lysine and arginine residues (K259, K265, R267, K271, R280 and K284). Inspection of the PSMs for these peptides provides additional agreement with the SASA analysis, as there are considerably more oxidized PSMs associated with these residues in the CF state (105 vs 58 PSMs).

2.4.2 Residue Level FPOP Oxidation of GCaMP2

Residue level analysis of GCaMP2 using the strict filter settings resulted in a substantial decrease in PSMs with only 87298 passing all the criteria. An even more dramatic change can be seen when looking at oxidized PSMs, going from 443780 PSMs and 11% of the total, to 10384 PSMs and 7.4% of the total. This resulted in a total of 61 singly oxidized residues identified, with an additional 11 double oxidation identifications (Table 2.4). While drastic, these changes allow for a more discriminative view of the conformational changes that occur between the two states, as demonstrated by the M13-GFP linker.

Table 2.4: GCaMP2 Oxidized Residues

Residue	Calcium Free		Calcium Bound		Ri	Fractional SASA (Å^2)		
	Oxidation	LN PF	Oxidation	LN PF		3EKJ	3EK4	3EK7
K31	4.38E-5 ± 6.7E-6	3.197	1.37E-3 ± 7E-4	3.159	2.2			
D32	3.67E-4 ± 2.6E-5	1.517	1.07E-4 ± 3E-5	1.531	0.42			
L33	7.51E-6 ± 1.7E-6	3.910	3.29E-5 ± 8E-6	3.893	4.4			
M36	1.08E-2 ± 4.8E-3	5.369	1.20E-2 ± 3E-3	5.368	20.5			
D38	1.33E-3 ± 8.7E-4	1.504	5.30E-4 ± 5E-4	1.514	0.42			
L60	1.86E-5 ± 1.1E-5	3.900	5.14E-5 ± 3E-5	3.888	4.4	0.34	0.65	0.5
E61	8.99E-5 ± 2.1E-5	2.029	6.88E-5 ± 4E-5	2.032	0.69	0.3	0.1	0.4
Y64	4.28E-5 ± 1.4E-5	4.894	5.43E-4 ± 2E-5	4.866	12	0.49	0.21	0.35
I65	2.55E-5 ± 2.2E-5	3.896	1.93E-3 ± 9E-4	3.849	4.4	0.01	0	0
M66	2.17E-3 ± 7.2E-4	5.386	3.31E-3 ± 2E-3	5.382	20.5	0.42	0.44	0.29
D68	5.05E-5 ± 9.7E-6	1.539	9.98E-6 ± 5E-6	1.557	0.42	0.16	0	0.17
D103	1.68E-3 ± 1.1E-3	1.501	7.31E-4 ± 5E-4	1.510	0.42	1.05	1.07	1.07
D110	6.50E-6 ± 6.0E-6	1.562	2.79E-5 ± 2E-5	1.546	0.42	0.76	0.8	0.8
L144	7.68E-3 ± 5.5E-3	3.834	2.54E-2 ± 2E-2	3.822	4.4	1.03	0.92	0.82
F166	2.75E-5 ± 2.1E-5	4.829	2.23E-5 ± 2E-5	4.832	11.2	0	0.01	0.01
V169	5.46E-5 ± 1.9E-5	3.048	4.13E-5 ± 1E-5	3.051	1.9	0.5	0.48	0.48
V170	1.91E-5 ± 7.9E-6	3.059	4.43E-5 ± 1E-5	3.050	1.9	0.1	0.09	0.07
I172	7.43E-6 ± 1.3E-6	3.910	8.23E-6 ± 6E-6	3.909	4.4	0	0	0
E175	3.83E-4 ± 1.3E-4	2.013	3.00E-4 ± 2E-4	2.016	0.69	0.4	0.41	0.5
D177	2.33E-5 ± 1.1E-5	1.548	1.85E-5 ± 5E-6	1.550	0.42	0.75	0.67	0.77
K184	2.83E-5 ± 9.8E-6	3.202	3.51E-5 ± 1E-5	3.199	2.2	0.77	0.79	0.76
V187	4.29E-5 ± 3.8E-5	3.050	9.20E-5 ± 5E-5	3.042	1.9	0	0	0
E190	1.03E-3 ± 3.8E-4	2.003	1.97E-3 ± 9E-4	1.996	0.69	0.75	0.72	0.67
E192	4.73E-4 ± 3.5E-4	2.011	1.58E-4 ± 1E-4	2.023	0.69	0.53	0.55	0.54

Continued on next page

Table 2.4: Continued

Residue	Calcium Free		Calcium Bound		Ri	Fractional SASA (Å^2)		
	Oxidation	LN PF	Oxidation	LN PF		3EKJ	3EK4	3EK7
D194	2.38E-3 ± 7.4E-4	1.497	3.59E-3 ± 1E-3	1.493	0.42	0.16	0.21	0.12
Y197	1.53E-4 ± 7.3E-5	4.879	5.37E-4 ± 3E-4	4.866	12	0.09	0.67	0.14
K199	1.33E-4 ± 7.7E-5	3.185	4.93E-4 ± 3E-4	3.170	2.2	0.23	0.4	0.23
C206	3.11E-3 ± 2.4E-3	5.736	1.60E-2 ± 9E-3	5.719	29.2	0	0	0
Y224	1.00E-2 ± 9.4E-3	4.835	7.63E-2 ± 7E-2	4.814	12			
M246	9.45E-4 ± 4.2E-4	5.395	2.31E-3 ± 1E-3	5.386	20.5	0	0	0
E248	1.55E-4 ± 6.5E-5	2.023	1.73E-4 ± 4E-5	2.022	0.69	0.49	0.55	0.47
E253	6.04E-4 ± 9.9E-5	2.009	6.26E-4 ± 1E-4	2.008	0.69	0.45	0.43	0.37
F257	3.27E-4 ± 1.1E-4	4.802	2.53E-4 ± 5E-5	4.805	11.2	0.5	0.49	0.48
F272	9.82E-4 ± 1.7E-4	4.790	1.08E-3 ± 2E-4	4.789	11.2	0.09	0.06	0.06
E273	5.52E-4 ± 1.5E-4	2.010	4.43E-4 ± 1E-4	2.012	0.69	0.33	0.38	0.39
D275	3.99E-4 ± 4.8E-5	1.517	3.01E-4 ± 9E-5	1.520	0.42	0.88	0.86	1.01
L277	7.72E-6 ± 6.8E-7	3.909	1.18E-5 ± 1E-5	3.905	4.4	0	0	0
D305	1.74E-4 ± 6.8E-5	1.526	1.38E-4 ± 5E-5	1.528	0.42	0.96		
E309	2.03E-4 ± 7.5E-5	2.020	9.33E-5 ± 3E-5	2.029	0.69	0.5	0.85	0.48
E310	2.90E-4 ± 1.3E-5	2.016	1.47E-4 ± 8E-5	2.024	0.69	0.91	0.72	0.85
I312	2.87E-4 ± 2.5E-5	3.869	1.31E-4 ± 8E-5	3.878	4.4	0.05	0.43	0.08
E314	4.65E-4 ± 8.5E-5	2.011	1.14E-3 ± 2E-4	2.002	0.69	0.62	0.35	0.69
F315	2.77E-4 ± 9.4E-5	4.804	4.16E-5 ± 4E-5	4.825	11.2	0.06	0	0.2
E317	1.45E-4 ± 2.4E-6	2.024	5.91E-4 ± 5E-5	2.009	0.69	0.73	0.58	0.75
F319	2.47E-4 ± 6.8E-5	4.805	7.71E-4 ± 2E-4	4.793	11.2	0	0.03	0
F322	1.26E-3 ± 9.1E-5	4.788	3.74E-5 ± 4E-5	4.826	11.2	0.21	0	0.24
E385	7.89E-4 ± 4.6E-5	2.006	8.29E-3 ± 1E-3	1.981	0.69		0.8	0.49
D398	3.13E-4 ± 9.1E-5	1.519	7.08E-4 ± 2E-4	1.510	0.42		0.7	0.68
Y402	8.13E-5 ± 2.6E-5	4.886	1.30E-4 ± 7E-6	4.881	12		0.93	0.43
H410	2.50E-3 ± 1.1E-4	4.594	4.69E-3 ± 1E-4	4.588	9.3		0.58	0.43
L419	2.14E-5 ± 1.2E-5	3.898	3.33E-5 ± 1E-5	3.893	4.4			0.17
E422	7.63E-5 ± 2.7E-6	2.031	1.56E-4 ± 7E-5	2.023	0.69			0.79

Continued on next page

Table 2.4: Continued

Residue	Calcium Free		Calcium Bound		Ri	Fractional SASA (Å^2)		
	Oxidation	LN PF	Oxidation	LN PF		3EKJ	3EK4	3EK7
E423	1.27E-4 ± 4.3E-5	2.025	8.23E-4 ± 2E-4	2.005	0.69			0.52
D425	8.56E-5 ± 1.5E-5	1.533	1.04E-3 ± 2E-4	1.506	0.42			0.56
E426	2.32E-5 ± 5.7E-6	2.044	8.30E-4 ± 4E-4	2.005	0.69			0.61
M427	1.62E-3 ± 8.8E-4	5.389	9.27E-3 ± 1E-3	5.371	20.5			0.43
I433	9.39E-5 ± 5.6E-5	3.882	5.08E-5 ± 2E-5	3.888	4.4			0.56
D434	1.92E-4 ± 2.0E-5	1.525	5.91E-4 ± 9E-5	1.512	0.42			0.6
D436	1.21E-4 ± 3.1E-5	1.530	1.60E-3 ± 4E-4	1.502	0.42			0.49
E442	2.95E-4 ± 5.2E-5	2.016	6.81E-4 ± 2E-4	2.007	0.69			0.68
E443	2.34E-4 ± 1.1E-4	2.019	2.00E-4 ± 9E-5	2.021	0.69			0.04
L60,M66	7.34E-6 ± 9.5E-7		2.29E-5 ± 1E-5					
N62,M66	2.02E-5 ± 1.0E-5		3.72E-4 ± 3E-5					
Y64,M66	1.60E-5 ± 7.8E-6		3.10E-4 ± 5E-5					
H82,H94	5.94E-5 ± 2.7E-5		9.34E-5 ± 9E-5					
E190,E192	4.47E-6 ± 3.4E-6		9.84E-6 ± 9E-6					
E190,D194	1.28E-5 ± 2.5E-6		1.46E-5 ± 1E-5					
L295,K298	3.12E-2 ± 1.9E-2		4.84E-2 ± 5E-2					
H297,K298	1.08E-1 ± 7.5E-2		3.38E-2 ± 3E-2					
D434,D436	1.82E-6 ± 1.3E-6		8.04E-6 ± 2E-6					
N440,M448	1.78E-5 ± 6.3E-6		1.48E-5 ± 1E-5					
M447,M448	2.90E-4 ± 2.1E-4		8.66E-4 ± 3E-4					

At the residue level, the PD method was able to detect the conformational change between the two states on the L60-E61 linker between M13 and GFP. In the CF state, the linker is integrated into the leading GFP β-strand, with the structure stabilized through hydrogen bonding between E61 and R81. This integration decreases the SASA of L60, while exposing the E61 sidechain. However, the linker is forced into a different conformation in the CB state as a result of the strong interaction between the M13 peptide and the CaM domain (Figure 2.5a).[21] This interaction

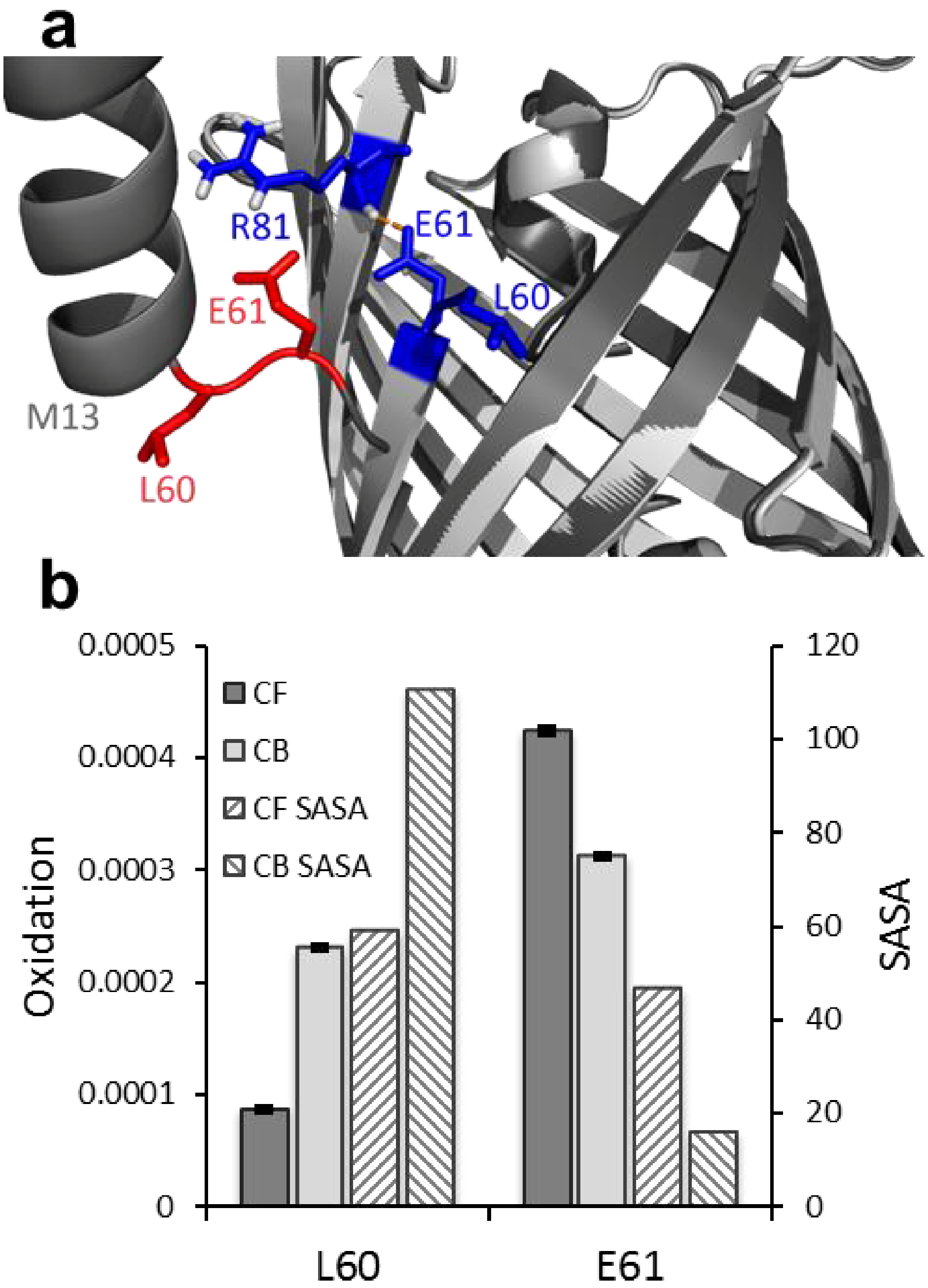


Figure 2.5. Conformational difference detected in the two residues in the M13-GFP linker with an overlay of the calcium free (light gray and blue, PDB ID: 3EKJ) and the calcium bound structures (dark gray and red, PDB ID: 3EK4)(a) and the oxidation and the SASA of the two residues for each conformation (b).

inverts the SASA for the linker residues, with L60 more exposed in the CB states. Residue level analysis using the PD method matches this trend, with a higher oxidation on L60 in the CB state and E61 in the CF state (Figure 2.5b). In addition to matching the trend for the linker, the correlations between % SASA (the fractional SASA of the residue) and log PF of the 61 oxidized residues are comparable to those that have been previously published, with -0.64 for the CF GCaMP2 and -0.45 and -0.59 for monomeric and dimeric CB GCamP2 (Figure 2.2).[14,15]

Another structural change between the CF and CB states stems from the architecture of the calcium sensor. In designing the cpEGFP moiety, four residues were removed from a GFP β strand. This creates an opening in the GFP barrel, resulting in an increased solvent accessibility in the barrel interior. Conformational changes induced by calcium binding cause the N-terminal domain of CaM to partially block this opening, decreasing the solvent accessibility of the chromophore.[21] In the CF state, the N-terminal domain of CaM is packed against the cpEGFP domain, creating a region of decreased SASA on the surface of the cpEGFP domain (Figure 2.6a). However, in the CB state the interaction of the CaM domain with the M13 peptide increases the exposed SA in this region of the GFP domain (Figure 2.6b, Table 2.3). Two oxidatively modified peptides were identified in this cpEGFP region. At the peptide level, peptides 128-151 and 185-199 both have an increase in oxidation in the CF state (Figure 2.4, Table 2.2). Residue level analysis also agrees with the conformational differences, identifying three residues on the occluded cpEGFP surface that have higher oxidation yields in the CB state (Figure 2.6c).

Although the surface of cpGFP has increased solvent exposure in the CB state, the same claim cannot be made for the CaM domain. Helices 8 and 11 of GCaMP2, which comprise the majority of the residues that are packed against the GFP domain in the CF state, are also complexed with the M13 peptide in the CB state.[21] Additionally, the 2 crystal structures are for this domain are incomplete, making it difficult to assess the CaM residue level data using the structures. However, the CB CaM domain is structurally similar to the CaM-M13 crystal structure, with only small

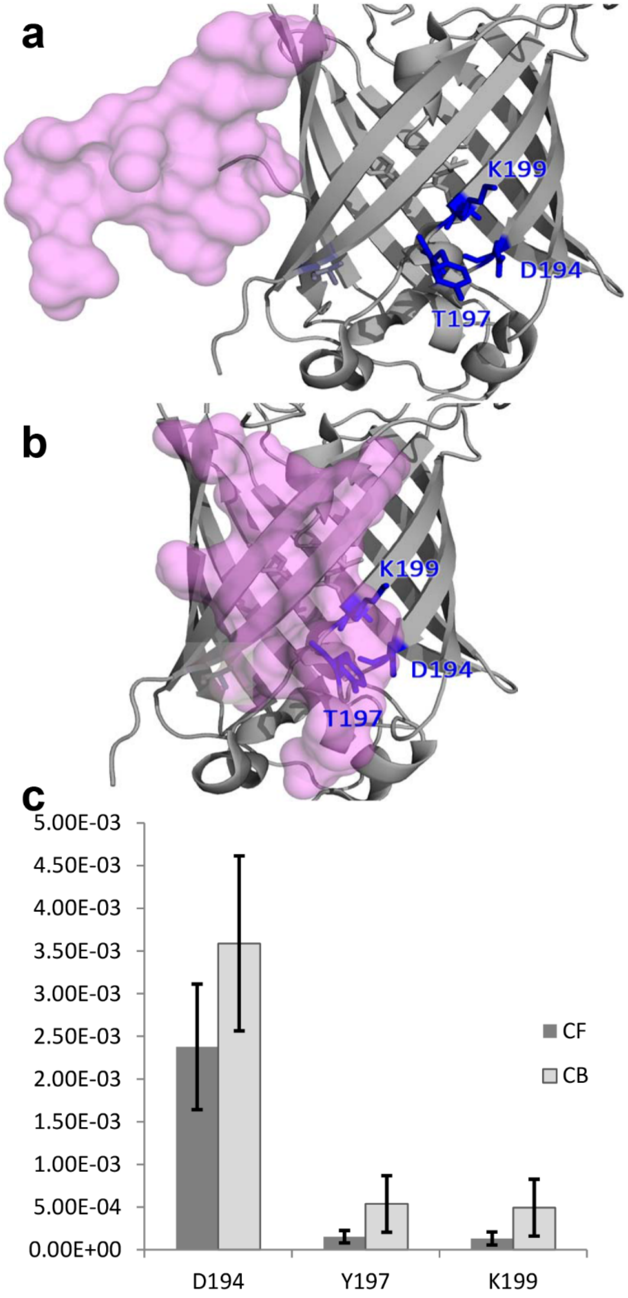


Figure 2.6. Inter-domain contacts between GFP (gray) and the N-terminal lobe of CaM (light purple) in the CF conformation (a) and structural changes in the CB conformation (b) with residues with higher oxidation in the CB conformation colored blue, and a bar graph (c) of the oxidation level of the highlighted residues.

differences associated with the CaM linker helix.[21,25] By reason of this structural similarity, a previously published FPOP study of calmodulin with the M13 peptide was used for comparison.[26] Since definitive oxidation values for calmodulin were not given, the trend of which state (CB or CF) yielded a higher oxidation was used for comparison (Table 2.4). There is an agreement between the previously published data and the data presented in this paper, with 2 of the residues in common between this and the previous study differing. One residue, I312 (I9 in isolated CaM), has nearly identical oxidation values in the CF and CB states in the previous study. Another residue, M427 (M124 in isolated CaM) is not significantly different in the previous study. This as well as the correlation between % SASA and log PF demonstrates the efficacy of the PD method in identifying and quantifying residue level oxidation.

Table 2.5
Comparison of CaM Domain PD Oxidation to Previously Published Data

GCaMP2 Res	CaM Res	GCamp2 Higher State	CaM-M13 Higher State[a]
I312	I9	CF	CB
F319	F16	CB	CB
L321	L18	CF	CF
F322	F19	CF	CF
Y402	Y99	CB	CB
M412	M109	CF	CF
M427	M124	CB	Same[b]

[a] CaM-M13 data taken from literature[26]

[b] Not significantly different

2.5 Discussion

Using protein footprinting methodologies with mass spectrometry for protein structural analysis can be advantageous, particularly when structural knowledge from traditional methodologies like X-ray crystallography is not available. Although these methods have the potential to fill the gap left by the traditional structural approaches, the potentially arduous task of analyzing the data can limit the size or complexity of the systems that can be studied using footprinting methods. Furthermore, any manual interpretation in the analysis steps can introduce a human bias, leading to errors in interpretation. As such, while footprinting can provide advantages over traditional methods, there is an inherent limitation in their ability to completely close the gap without a more comprehensive analysis strategy.

My goal was to develop a method that could rapidly analyze the raw data from any stable label footprinting experiment, regardless of the size or complexity of the data set. Developing a completely integrated software platform would have required an exhaustive search of every footprinting label currently in use, and would require an update for each newly published footprinting label. Additionally, this would have required restrictions on the raw data file format(s) or a strict requirement on file conversion.

To overcome these obstacles, I developed our strategy using a two pronged approach, separating database searching from quantitation. The PD workflow (Figure 2.1) used for this work can be adapted for any stable label by changing the variable modifications search to those used in the experiment, adding or deleting search levels as needed. Additionally, PD and the algorithms (Mascot and Sequest HT) allow the addition of new modifications, making it accessible to labels that have not yet been published. Nonetheless, PD is not required for quantitation in PowerPivot, as the quantitation method can be adapted to any data set formatted to a similar layout. By separating the search and quantitation strategies, our complete method can be adapted for any footprinting label and quantitation can be completed with or without the use of PD. Furthermore, post search quantitation takes under two hours, for one or thousands of proteins, regardless of the number or complexity of samples to analyze. Only the search time increases, as would be the case for any other platform.

The use of different filtering strategies for peptide and residue level analysis minimizes the amount of manual interpretation and validation required, while still providing a high level of confidence in the results obtained. For peptide level analysis, it is not necessary to know the exact location of the modification on a given peptide. Therefore, it is reasonable to conclude that corresponding modified PSMs identified within the mass tolerance are correct within a 95% confidence interval, and that the ambiguity of the location assignment is inconsequential to the overall outcome at this level of detail.

A more stringent approach must be taken to provide a high level of confidence in modification assignments for residue level analysis. Restricting PSMs to only those with one modification is necessary as there is not a reliable method to assign peak area proportions to each modification in multiply oxidized PSMs. The additional restrictions of a 1% FDR and limiting the search engine rank to 1 allow for a high level of confidence in per-residue analysis while limiting the amount of manual interpretation required. Comparing the MS/MS scans for 2 identified modifications on the same peptide in GCaMP2 best illustrates this point. A decarboxylation modification on residue D194 was identified with high confidence and a search engine rank of 3 (Figure 2.7a), but does not have fragment peaks to substantiate this assignment. Without the filters, this assignment would have had to be validated manually, increasing the amount of time required to complete analysis. For comparison, a loss of CO on this same residue was identified with high confidence and a search engine rank of 1 (Figure 2.7b), and has a more complete y-ion series coverage in the MS/MS spectrum with a fragment matching the modification. By using strict filters on the data up front, a high level of confidence in the assignments can be obtained without the need to manually remove PSMs individually which drastically

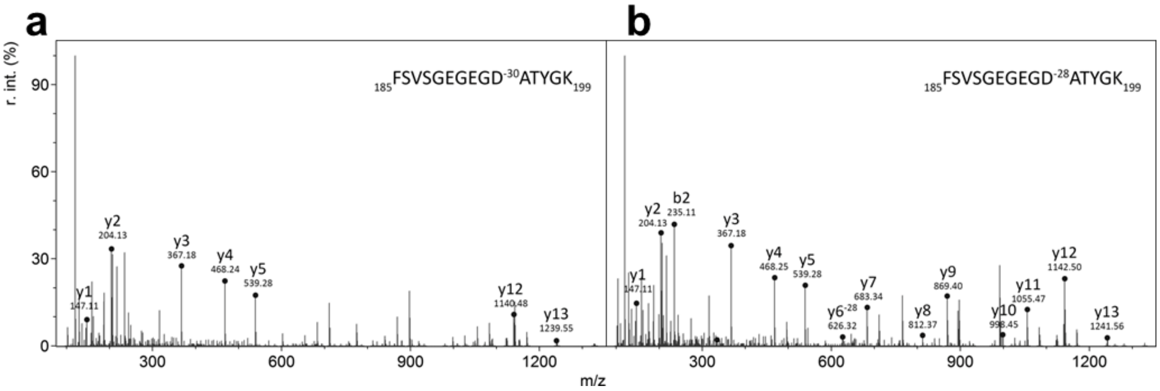


Figure 2.7. MS/MS spectra of FPOP modifications on D194 of peptide 185-199 with an ambiguous assignment of a decarboxylation on D194 (a) that was removed by the strict filtering scheme, and the same peptide and location with a loss of CO (b) that pass the filters with fragment ions that validate the oxidation.

decreases the time required for analysis. This is further demonstrated by the oxidation levels of the two M13-cpEGFP linker residues, L60-E61 (Figure 2.2). Glutamic acid is one of the least ·OH reactive amino acids, with a reactivity 2.3x108.[2] Manual interpretation may have been biased by the more highly reactive residues on the peptide; including L61 with a 6-fold higher reactivity, Y64 with a 17 fold higher reactivity, and M66 with a 30-fold higher reactivity. By letting the algorithms make the assignment, no human bias can by factored into the analysis, and the detection of oxidized residues with lower reactivities is not hindered. While there is still a small probability of incorrect assignments, the use of an n number of PSMs threshold decreases the likelihood of the incorrect assignments being included while interpreting the results. Furthermore, the upfront reduction of PSMs before analysis is completed also decreases the number of data points, and therefore the time required for manual verification of the results, if desired.

2.6 Conclusion

This chapter has demonstrated my PD method data analysis method, a complete identification and quantitation strategy for mass spectrometry-based protein footprinting studies. The data presented here has demonstrated the efficacy of the PD method in detecting and analyzing hydroxyl radical-mediated protein footprinting data. Use of this method on GCaMP2 has validated that this method not only provides significant footprinting coverage while maintaining a high level of confidence in the results obtained, but does so in less than two hours. Additionally, this method can be adapted for use with any stable covalent label and any level of experiment or sample complexity. The approach used when developing this method will allow for it to be modified as new technologies and footprinting strategies are developed.

2.7 References

1. P.R. Cutillas and J.F. Timms. *LC-MS/MS in Proteomics: Methods and Applications*. Humana Press, 2010. ISBN 9781607617792.

2. Guozhong Xu and Mark R. Chance. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chemical Reviews*, 107(8): 3514–3543, 2007. ISSN 0009-2665. doi: 10.1021/cr0682047.

3. Brian C. Gau, Hao Chen, Yun Zhang, and Michael L. Gross. Sulfate radical anion as a new reagent for fast photochemical oxidation of proteins. *Analytical Chemistry*, 82(18):7821–7827, 2010. ISSN 0003-2700. doi: 10.1021/ac101760y.

4. Shao-En Ong, Blagoy Blagoev, Irina Kratchmarova, Dan Bach Kristensen, Hanno Steen, Akhilesh Pandey, and Matthias Mann. Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, 1(5):376–386, 2002. ISSN 1535-9476.

5. Parminder Kaur, Janna G. Kiselar, and Mark R. Chance. Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. *Analytical Chemistry*, 81(19):8141–8149, 2009. ISSN 0003-2700. doi: 10.1021/ac9013644.

6. Brian C. Gau, Jiawei Chen, and Michael L. Gross. Fast photochemical oxidation of proteins for comparing solvent-accessibility changes accompanying protein folding: Data processing and application to barstar. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(6):1230–1238, 2013. ISSN 1570-9639. doi: 10.1016/j.bbapap.2013.02.023.

7. David N. Perkins, Darryl J. C. Pappin, David M. Creasy, and John S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *ELECTROPHORESIS*, 20(18):3551–3567, 1999. ISSN 1522-2683. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

8. M. Bern, Y. Cai, and D. Goldberg. Lookup peaks: A hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal Chem*, 79(4):1393–400, 2007. ISSN 0003-2700 (Print) 0003-2700. doi: 10.1021/ac0617013.

9. Stephen Tanner, Hongjun Shu, Ari Frank, Ling-Chi Wang, Ebrahim Zandi, Marc Mumby, Pavel A. Pevzner, and Vineet Bafna. Inspect:□ identification of posttranslationally modified peptides from tandem mass spectra. *Analytical Chemistry*, 77(14):4626–4639, 2005. ISSN 0003-2700. doi: 10.1021/ac050102d.

10. O. Charvatova, B. L. Foley, M. W. Bern, J. S. Sharp, R. Orlando, and R. J. Woods. Quantifying protein interface footprinting by hydroxyl radical oxidation and molecular dynamics simulation: Application to galectin-1. *J Am Soc Mass Spectrom*, 19(11):1692–705, 2008. ISSN 1044-0305 (Print) 1044-0305. doi: 10.1016/j.jasms.2008.07.013.

11. Carlee S. McClintock, Jerry M. Parks, Marshall Bern, Pavan K. GhattyVenkataKrishna, and Robert L. Hettich. Comparative informatics analysis to evaluate site-specific protein oxidation in multidimensional lc–ms/ms data. *Journal of Proteome Research*, 12(7):3307–3316, 2013. ISSN 1535-3893. doi: 10.1021/pr400141p.

12. Aimee Rinas and Lisa M Jones. Fast photochemical oxidation of proteins coupled to multidimensional protein identification technology (mudpit): Expanding footprinting strategies to complex systems. *Journal of The American Society for Mass Spectrometry*, 26(4):540–546, 2015. ISSN 1044-0305.

13. Jessica A. Espino, Vishaal S. Mali, and Lisa M. Jones. In cell footprinting coupled with mass spectrometry for the structural analysis of proteins in live cells. *Analytical Chemistry*, 87(15):7971–7978, 2015. ISSN 0003-2700. doi: 10.1021/acs.analchem.5b01888.

14. Parminder Kaur, Janna Kiselar, Sichun Yang, and Mark R. Chance. Quantitative protein topography analysis and high-resolution structure prediction using hydroxyl radical labeling and tandem-ion mass spectrometry (ms). *Molecular & Cellular Proteomics*, 14(4):1159–1168, 2015. doi: 10.1074/mcp.O114.044362.

15. Wei Huang, Krishnakumar M Ravikumar, Mark R Chance, and Sichun Yang. Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: A protection factor analysis. *Biophysical Journal*, 108(1):107–115, 2015. ISSN 0006-3495. doi: 10.1016/j.bpj.2014.11.013.

16. EduardoJ Pilau, AmadeuH Iglesias, and FabioC Gozzo. A new label-free approach for the determination of reaction rates in oxidative footprinting experiments. *Analytical and Bioanalytical Chemistry*, 405(24):7679–7686, 2013. ISSN 1618-2642. doi: 10.1007/s00216-013-7247-9.

17. Qi Wang, Bo Shui, Michael I. Kotlikoff, and Holger Sondermann. Structural basis for calcium sensing by gcamp2. *Structure*, 16(12):1817–1827, 2008. ISSN 0969-2126. doi: 10.1016/j.str.2008.10.008.

18. DavidM Hambly and MichaelL Gross. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *Journal of the American Society for Mass Spectrometry*, 16(12): 2057–2063, 2005. ISSN 1044-0305. doi: 10.1016/j.jasms.2005.09.008.

19. Brian C. Gau, Joshua S. Sharp, Don L. Rempel, and Michael L. Gross. Fast photochemical oxidation of protein footprints faster than protein unfolding. *Analytical Chemistry*, 81(16):6563–6571, 2009. ISSN 0003-2700. doi: 10.1021/ac901054w.

20. Lisa M. Jones, Justin B. Sperry, James A. Carroll, and Michael L. Gross. Fast photochemical oxidation of proteins for epitope mapping. *Analytical Chemistry*, 83(20):7657–7661, 2011. ISSN 0003-2700. doi: 10.1021/ac2007366.

21. J. Akerboom, J. D. Rivera, M. M. Guilbe, E. C. Malave, H. H. Hernandez, L. Tian, S. A. Hires, J. S. Marvin, L. L. Looger, and E. R. Schreiter. Crystal structures of the gcamp calcium sensor reveal the mechanism of fluorescence signal change and aid rational design. *J Biol Chem*, 284(10):6455–64, 2009. ISSN 0021-9258 (Print) 0021-9258. doi: 10.1074/jbc.M807657200.

22. Takeharu Nagai, Asako Sawano, Eun Sun Park, and Atsushi Miyawaki. Circularly permuted green fluorescent proteins engineered to sense ca2+. *Proceedings of the National Academy of Sciences*, 98(6):3197–3202, 2001. doi: 10.1073/pnas.051636098.

23. Leigh Willard, Anuj Ranjan, Haiyan Zhang, Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart. Vadar: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Research*, 31(13):3316–3319, 2003. ISSN 0305-1048 1362-4962.

24. LLC Schrodinger. The pymol molecular graphics system, version 1.3r1. 2010.

25. WE Meador, AR Means, and FA Quiocho. Target enzyme recognition by calmodulin: 2.4 a structure of a calmodulin-peptide complex. *Science*, 257 (5074):1251–1255, 1992. doi: 10.1126/science.1519061.

26. Hao Zhang, Brian C. Gau, Lisa M. Jones, Ilan Vidavsky, and Michael L. Gross. Fast photochemical oxidation of proteins for comparing structures of protein−ligand complexes: The calmodulin−peptide model system. *Analytical Chemistry*, 83(1):311–318, 2011. ISSN 0003-2700. doi: 10.1021/ac102426d.

3. FAST PHOTOCHEMICAL OXIDATION OF PROTEINS COUPLED TO MUDPIT: EXPANDING FOOTPRINTING STRATEGIES TO COMPLEX SYSTEMS

3.1 Introduction

Hydroxyl radical (·OH) based footprinting, first coupled with mass spectrometry by Chance and coworkers[1], is one of the most informative covalent labeling methods for a number of reasons. The ·OHs have similar properties to water and can freely oxidize solvent exposed side chains. Additionally, their reactivity is well known and researchers can capitalize on their low selectivity, increasing the amount of information obtained.[2] Furthermore, there are multiple methods available for generating ·OHs, increasing the accessibility of this method.[3–5] FPOP, used for the work in this chapter, generates ·OHs through laser induced photolysis of hydrogen peroxide.[6] This technique modifies proteins on a microsecond timescale, theoretically eliminating structural changes induced by labeling.[6,7]

A consequence of design features employed in FPOP experiments to eliminate radical induced unfolding is that oxidized species are present in lower abundance as compared to their unoxidized counterparts.[6] Therefore, the difficulty in detecting these species will grow concurrently with increasing sample complexity. Investigating the structures of large, megadalton sized molecular assemblies, has often proven difficult. While there several methods for obtaining protein structures, the majority come from X-ray crystallography.[8] At the outset, it is often challenging to purify all of the protein components of megadalton complexes, a necessity in obtaining a structure.[9] Even when this is accomplished, it can be equally difficult to crystallize these complexes, or the process may only yield crystals too small for analysis.[8] While FPOP has the potential to start to fill this gap in information, it is first necessary to overcome the hurdle of identifying the relatively low abundant oxidized species in a sea of higher abundant peptides. A major obstacle is using data-dependent acquisition (DDA) for MS/MS analysis. In this method, precursor ions are selected for fragmentation based on their signal intensities. Often, if chromatographic separation is not sufficient, peptides with higher abundance are identified while lower abundance peptides are not. A more proficient chromatographic separation could aid in increasing peptide identifications.

Multi-dimensional protein identification technology (MudPIT), is a method used to overcome the inability of single-dimensional separations to resolve complex biological samples.[10] The use of a biphasic analytical column increases the peak column capacity, and allows for online two-dimensional separations.[10,11] Coupling of FPOP labeling with MudPIT could provide an increase in identifications of oxidized peptides in complex systems. The use of this method to identify oxidatively modified peptides has been previously reported.[12] However, the study was mainly focused on comparison of informatics methods rather than as a method to be utilized to identify more oxidatively modified peptides for highly complex samples. Additionally, the researchers used a low complexity sample with a “mini-MudPIT” method consisting of only three salt steps. In this chapter, the combination of a full MudPIT method with FPOP on a highly complex sample, Saccharomyces cerevisiae yeast cell lysate is discussed. The objective was to improve the detection of FPOP labeled species and expand the application of FPOP to more complex systems.

3.2 Experimental Section

3.2.1 Materials

All chemicals were obtained from Thermo Fisher Scientific (Waltham, MA) unless otherwise noted.

3.2.2 Oxidative Labeling

Each 100 µL sample contained 10 mM PBS (Sigma Aldrich, St. Louis, MO) 10 mM L-glutamine, 7.5 mM hydrogen peroxide and yeast cell lysate (A gift from Dr. Amber Mosley and Whitney Smith-Kinnaman, Department of Biochemistry, Indiana University School of Medicine, Indianapolis, IN) at a concentration of 0.18 mg/mL. The hydrogen peroxide was added just prior to infusion. FPOP was performed similarly as described.[6,7] A 248 nm KrF excimer laser (GAM Laser Inc., Orlando, FL) was used to irradiate the sample solution at 135 mJ/pulse. The laser was focused through a 250 mm plano convex lens (Thorlabs, Inc., Newton, NJ) onto 150 µm i.d. fused silica tubing (Polymicro Technologies, Phoenix, AZ) with the polyimide coating removed, giving a 2.5 mm irradiation window. The flow rate, 33 µL/min, was set to allow for a 20 percent exclusion fraction. A total of 4 FPOP samples and 3 controls (no irradiation) were prepared.

3.2.3 Proteolysis

Post FPOP, the yeast lysate samples were subjected to a two-step digestion process as previously described.[13] Each sample was acetone precipitated[14] and resuspended in 8 M urea 150 mM Tris-HCl pH 8.5 buffer. Proteins were reduced with 10 mM tris(2-carboxyethyl) phosphine (TCEP) for 30 minutes at room temperature (RT). They were then alkylated with 20 mM iodoacetamide for 30 minutes at RT with a foil cover to protect the sample from light. The alkylation reaction was quenched with 10 mM dithiothreitol for 15 minutes at RT. Lys-C was added at a 100:1 substrate to protease ratio and incubated overnight at 37 °C. The samples were then diluted with 150 mM Tris buffer to bring the urea concentration to 2 M. Trypsin was added at a 50:1 substrate to protease ratio and incubated overnight at 37 °C. Digestion was quenched with formic acid (Sigma Aldrich, St. Louis, MO) at a final concentration of 5%.

3.2.4 LC-MS/MS Acquisition

Analysis was completed using an UltiMate 3000 RSLC and a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA). For each experiment, 1 µg of the digest was loaded onto a 2 cm Acclaim Pepmap 100 C18 trap column (Thermo Fisher Scientific, Waltham, MA). MS1 spectra were acquired over an *m/z* range of 350–2000 at a resolving power of 70 000. The 25 most abundant ions were selected for MS2 at a resolving power of 17500. Ions with a charge-state of +1 and >+8 ions were rejected.

One Dimensional LC-MS/MS

Samples were loaded onto a 100 µm x 2 cm Acclaim PepMap100 C18 nano trap column (5 µm, 100 Å) (Thermo Scientific, Waltham, MA) and washed for 10 minutes with loading buffer (LB, 2% acetonitrile 0.1% formic acid) with a flow rate of 5 µL/min. The samples were separated on a 75 µm inner diameter (ID) reverse phase (RP) analytical column packed in-house with a 30 cm bed of Magic 5 µm C18 particles (Michrom Bioresources Inc., Auburn, CA) with a 67 minute linear gradient at a flow rate of 300 nl/min to 40% acetonitrile 0.1% formic acid. The total run time was 105 minutes including loading, washing and equilibration. AGC targets were set to 3e6 for MS1 and 1e5 for data-dependent MS2 with an underfill ratio of 1.0%, giving an intensity threshold of 2.0e4.

MudPIT LC-MS/MS

Fully automated analysis was completed in a similar manner as previously described.[15,16] Each sample was loaded onto a trap column and washed for 10 minutes with LB at a flow rate of 5 µL/min. Samples were separated on a 75 µm ID RP analytical column packed in-house with a 26 cm bed of Magic 5 µm C18 particles (Michrom Bioresources Inc., Auburn, CA) followed by a 4 cm bed of Luna strong cation exchange (SCX) resin (Phenomenex, Torrence, CA). Peptide fractions were displaced from the SCX resin to the RP resin using the following salt pulses: (1) 0% (2) 5% (3) 10% (4) 15% (5) 20% (6) 30% (7) 40% (8) 50% (9) 60% (10) 80% of SCX buffer (SCXB, 500 mM ammonium acetate (Sigma Aldrich, St. Louis, MO) in 5% acetonitrile and 0.1% formic acid) mixed with LB by the loading pump mixer. The 0% fraction was used to displace the sample from the trap column to the analytical column. Each subsequent salt pulse was generated by increasing the SCXB percentage to the next concentration with a loading pump gradient during the previous salt step. A 2.6 µL aliquot of salt (roughly 15x the SCX bed volume) was collected by coupling a 30 cm 75 µm ID NanoViper line (Thermo Fisher Scientific, Waltham, MA) to the trap column with a stainless steel union, and delivered when the switching valve position was changed. Each salt pulse was pushed over the analytical column by the gradient pump for 20 minutes at a flow rate of 300 nL/min. Sample fractions were separated with a 67-minute linear gradient at a flow rate of 300 nL/min to 40% acetonitrile 0.1% formic acid. The total run time for each fraction was 105 minutes including loading, washing and equilibration time. AGC targets were set to 1e6 for MS1 and 5e4 for data-dependent MS2 with an underfill ratio of 1.0%, giving an intensity threshold of 1.0e4.

3.2.5 Data Analysis

All data files were searched using Proteome Discoverer version 1.4 (Thermo Fisher Scientific, Waltham, MA.) with Sequest HT and Mascot version 2.4 (Matrix Sciences Ltd., London, UK) against a Saccharomyces cerevisiae FASTA database (strain ATCC 204508 / S288c, downloaded from Uniprot February 2014), and EIC areas for each PSM were calculated using a custom multi-level workflow (as described in chapter 2). Peptides were ungrouped and filtered to a 1% FDR. Only PSMs identified as selected or unambiguous were used for analysis. The data was exported to Excel and summarized using the PowerPivot add-in. The fractional oxidation per residue on a given sequence was determined according to Equation 2.1.

3.3 Results

3.3.1 LC-MS Method Comparison

In order to make the most direct comparison between the one-dimensional chromatography data-dependent analysis (1D-DDA) and MudPIT, when possible, all processes and parameters were kept identical. However, some parameters were altered for the analysis. First, the typical FPOP sample size was doubled to ensure there was an adequate amount to analyze each sample by both methods. In addition, the total analytical column length was kept at 30 cm and the gradient for each sample or step was constant over the entire experiment.

For comparison of the two methods, the sample loading procedure for MudPIT had to be altered. In MudPIT analysis, samples are often pressure loaded or directly injected onto the analytical column.[10] In this experiment, the samples were loaded onto a trap column via the autosampler. There were several advantages to loading the samples in this manner. First, sample washing was identical for both methods. Any hydrophilic peptides that may have been washed off of the trap column should be the same over both methods. Directly loading the sample onto a three phase analytical column could have created a bias between the two methods. Second, trap column loading allowed both method analyses to be completed continuously, whereas pressure loading would require the MudPIT analysis to be completed discontinuously. Other parameters that were altered were automatic gain control (AGC) targets, for both MS1 and MS2, and dynamic exclusion times. These parameters were optimized for each method to provide peak performance.

3.3.2 Increases in Identifications by MudPIT

In agreement with previously reported results[10,11], using the MudPIT method to analyze the labeled yeast lysate samples gave a substantial increase in peptide spectrum matches (Figure 3.1). When comparing the two methods at the protein level, a 1.7-fold increase in protein group identifications (IDs), including 820 unique proteins, was observed with MudPIT (Figure 3.1 top). At the peptide level, a 1.3-fold increase in IDs with MudPIT was observed. Comparing unique peptides, MudPIT had a 1.7-fold increase in IDs with almost 1700 more unique peptides observed over 1D-DDA (Figure 3.1 middle). Although significant increases were observed with MudPIT on the protein and peptide levels, the true value of using the method in conjunction with protein footprinting is appreciated when looking at oxidatively modified peptides (Figure 3.1 bottom). Here, a 2.7-fold increase in oxidized peptide IDs was observed with MudPIT. Even more significant, MudPIT has a 4.6-fold increase in IDs of unique oxidized peptides over 1D-DDA. This demonstrates the efficacy for coupling MudPIT with FPOP. The higher sequence coverage of oxidatively modified peptides will provide a more complete description of the protein system.

To further evaluate the increased IDs achieved with the MudPIT method, we compared the identification of oxidatively modified residues on pyruvate kinase 1 (PK1, PDB ID: 1A3W)[17] and phosphoglycerate kinase (PGK1, PDB ID: 3PGK)[18]. These proteins were chosen as representative because both had high coverage with each method (greater than 75%) and each had oxidatively modified peptides identified by the search workflow. For both of these proteins, modifications were only included if they were identified more than once in the samples (PSM ≥ 2) and if the quantifiable oxidation levels were greater than the mean standard error. Table 3.1 shows residues that were identified by each method. For PK1, 1D-DDA identified only six of the fourteen oxidatively modified residues identified by MudPIT. For PGK1, the fourth most abundant protein found in S. cerevisiae[19], 1D-DDA only IDs

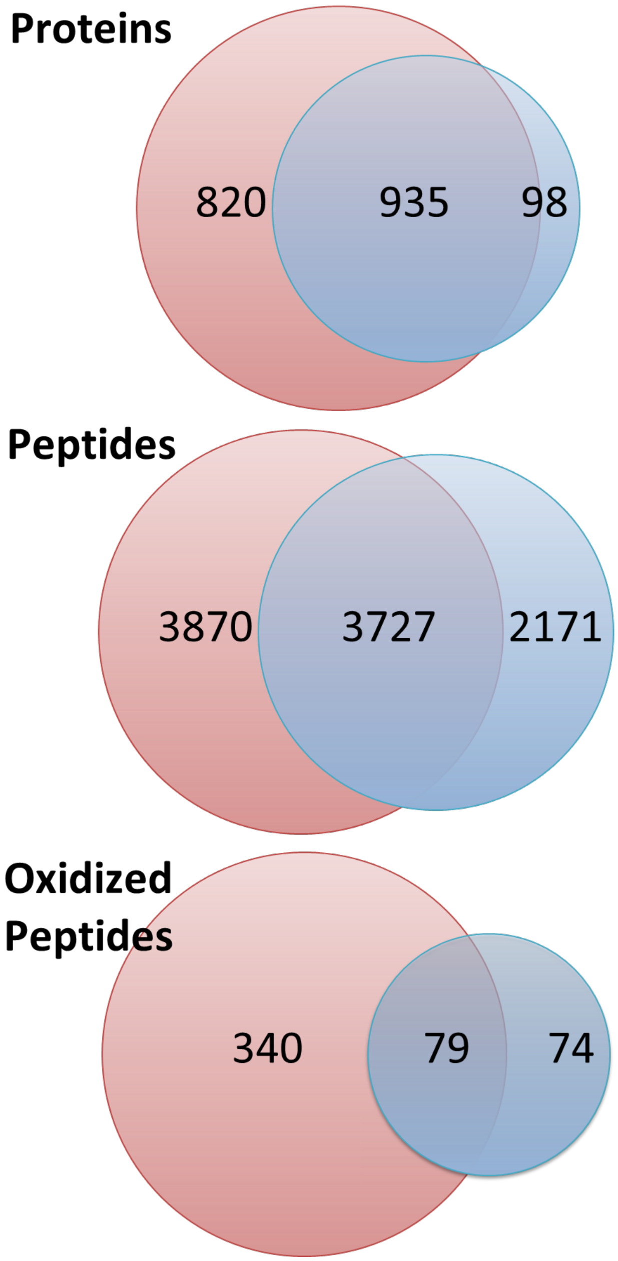


Figure 3.1. Visual comparison of IDs between MudPIT (red) and 1D-DDA (blue) methods by proteins (top), peptides (middle), and oxidatively modified peptides (bottom).

16 of the 41 residues that MudPIT identified. Since PKG1 is very abundant in yeast lysate, it can be assumed that this protein is oxidized more frequently than lower abundant proteins in the lysate. Consequently, the oxidized peptides from PKG1

may be relatively abundant. Despite that, 1D-DDA IDs less than half the number of oxidized residues as MudPIT.

Table 3.1
Oxidatively Modified Residues Identified

PK1 Residue	1D-DDA	MudPIT	PGK1 Residue	1D-DDA	MudPIT
D185	N	Y	M174	Y	Y
D187	N	Y	V175	Y	Y
N249	N	Y	E200	N	Y
F250	N	Y	N201	N	Y
D251	Y	Y	P202	N	Y
E252	N	Y	R204	N	Y
D266	N	Y	I234	N	Y
E270	N	Y	M238	N	Y
I399	N	Y	A239	N	Y
D451	Y	Y	E247	N	Y
W452	Y	Y	D253	N	Y
D454	Y	Y	E303	N	Y
D455	Y	Y	I305	N	Y
E457	Y	Y	P306	N	Y
			A307	N	Y
			W309	Y	Y
			Q310	N	Y
			D313	Y	Y
			I332	N	Y
			V333	N	Y

3.3.3 Properties of Peptides Identified by MudPIT

Wolters et al.[10] demonstrated that MudPIT has a high dynamic range with the ability to ID low abundance peptides. To determine whether the IDs from MudPIT are lower abundance than those from 1D-DDA, the intensities of the identified peptides were analyzed. Figure 3.2A compares the intensity of peptides identified from MudPIT and 1D-DDA. The average intensity of the peptides identified by both methods are similar. However, the minimum intensity of peptides identified by MudPIT is lower than for 1D-DDA. A histogram of frequency of identifications of peptides at varying intensities further demonstrates this (Figure 3.2B). Since the MudPIT method has more overall identifications, the histogram has been normalized to show the percent of total peptides. For both MudPIT and 1D-DDA the highest number of identifications were from peptides with intensities in the range of 1.00e6 (1e6 - 9e6), followed by intensities in the range 1.00e5 (1e5 - 9e5) and 1.00e7 (1e7 - 9e7). In the lowest intensity bin, 1.00e4 (1e4 - 9e4), MudPIT facilitated detection of three times as many peptides, 90 (1%) and 25 (0.4%) for MudPIT and 1D-DDA, respectively. This increase in lower intensity identifications is even more significant for oxidized peptides where 55 and 10 oxidatively modified peptides from MudPIT and 1D-DDA were identified, respectively.

Comparing MudPIT identifications to yeast lysate protein abundance further demonstrates the method can aid in identifying low abundance proteins. As mentioned previously, PGK1 is highly abundant in S. cerevisiae with an estimated abundance of 21,000 parts per million (ppm);[19] the sequence coverage for this protein is 75%. The ATP-dependent transporter protein YER036C is also identified by MudPIT with 25% sequence coverage. This protein has an abundance of 743 ppm 29-fold lower than PGK11 indicating the dynamic range of the MudPIT method.

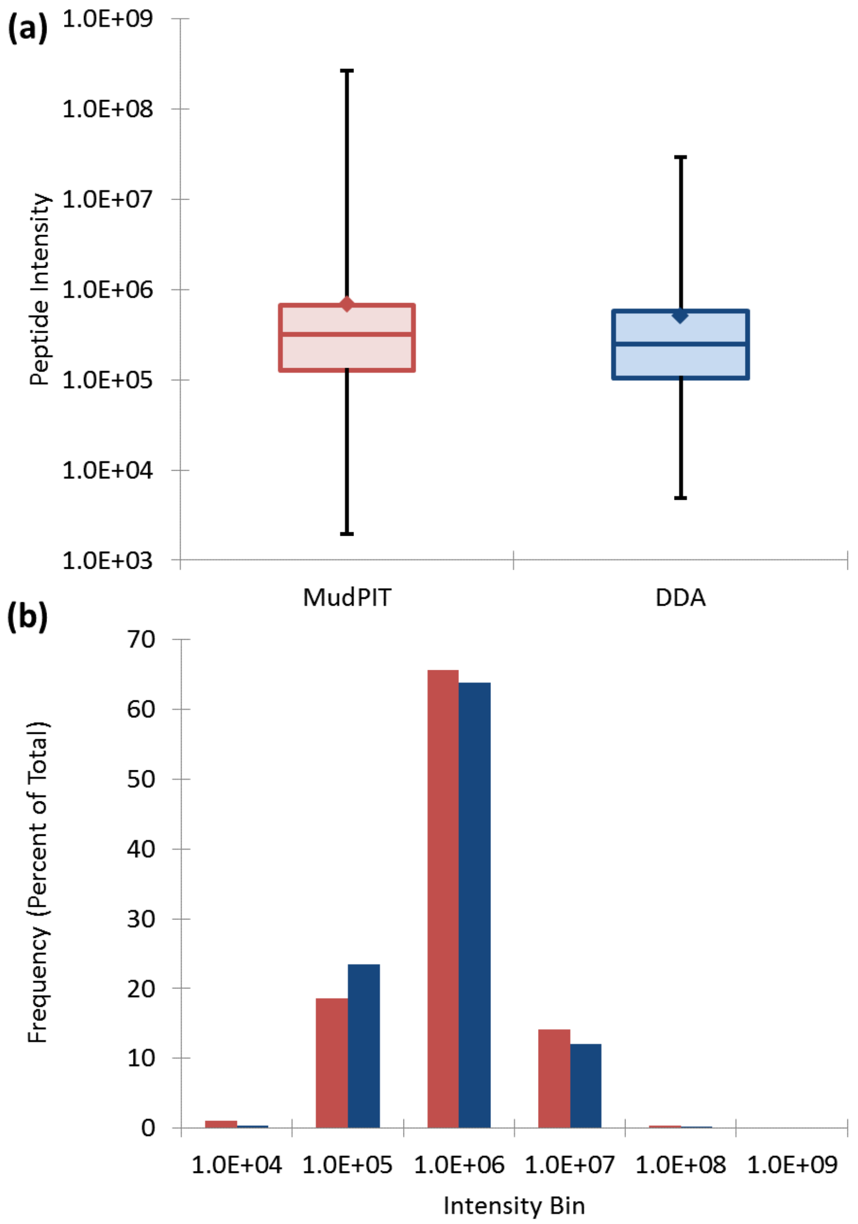


Figure 3.2. Distribution of the intensities of PSMs identified by MudPIT (red) and 1D-DDA (blue). (a) The spread of intensities is demonstrated in the box-and-whisker plot with the box lines marking the upper median and lower quartiles, and the whiskers marking the complete range. (b) The frequency of the distributions of intensities is displayed in a histogram.

3.3.4 MudPIT as a Method for Megadalton Protein Complexes

A major obstacle in oxidative labeling experiments is the ability to obtain residue level oxidation on large, macromolecular protein complexes. Given that the surface area to volume ratio decreases as a particle increases in size, it stands to reason that the proportion of oxidized species present when analyzing a MDa sized complex would also decrease, making the likelihood of detecting modifications even more difficult.

To demonstrate the power of using MudPIT analysis in oxidative footprinting experiments, residue level oxidation was calculated on a yeast 80S ribosome, which has a published structure (PDB ID 4V6I).[20] Ribosomes are cellular organelles, consisting of both protein and RNA, involved in protein assembly. The protein component of the structure is assembled in two subunits, 40S and 60S, and contains a total of 70 known proteins. We identified 52 of the 70 proteins in the MudPIT samples, with sequence coverage values ranging from 5 to 80 % (data not shown). A total of 86 residues were identified as oxidized and mapped to the crystal structure for a visual representation (Figure 3.3). Since RNase was not added to the sample at any time to remove the RNA, the structure is presented with the RNA present. The mapping of oxidized residues onto a surface representation of the crystal structure demonstrates that many solvent accessible residues are oxidized.

To further investigate the correlation between residue oxidation and solvent accessibility, the log protection factor (described in chapter 2) of residues identified by MudPIT was compared to the fractional SASA values, calculated using VADAR as described by Willard et al.[21] Since FPOP was performed on yeast lysate where various proteins could be interacting, we had to consider certain variables prior to the comparison. While a binary interactome of yeast has been published[22], it is unlikely that every interaction with this complex has been documented. With this in mind, it seemed unlikely that a comparison of SASA to oxidation over the complete complex would yield any reliable assessment of the method. As a consequence, a

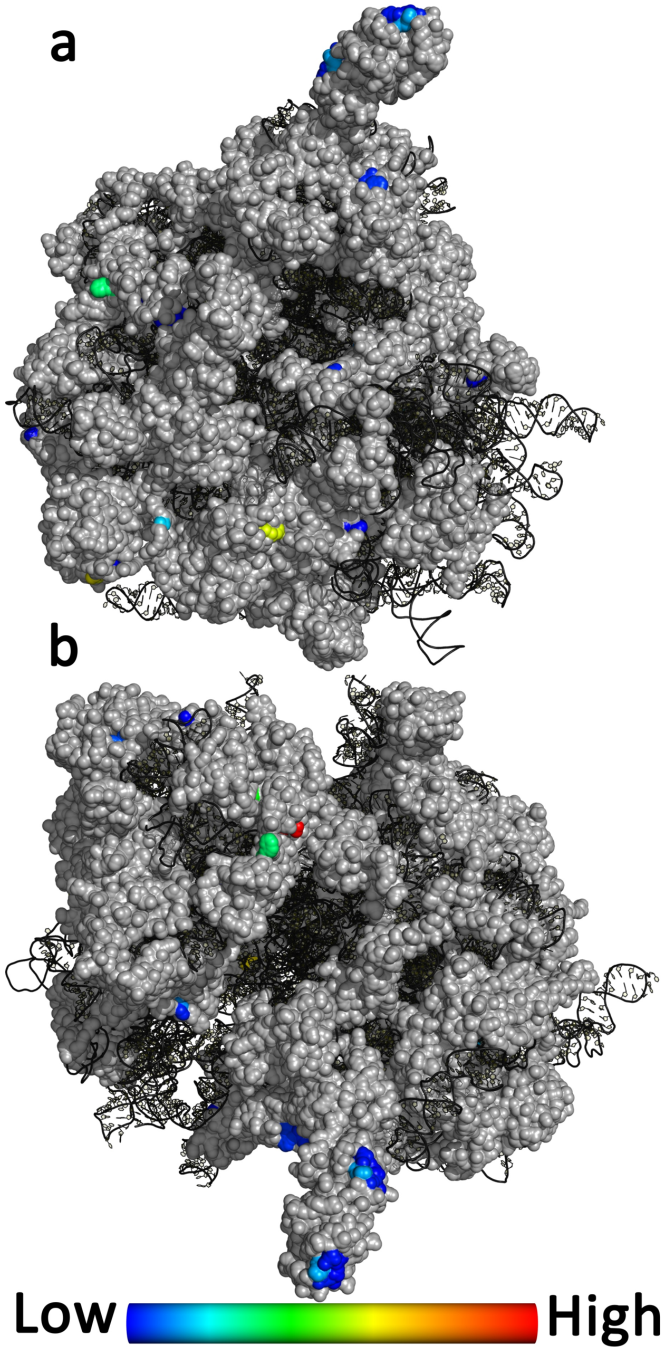


Figure 3.3. Two perspectives of the structural location of MudPIT determined FPOP oxidation levels mapped to a yeast 80 s ribosomal crystal structure, 4V6I[20]. The lowest oxidation levels are in blue going to the highest in red.

comparison was done on a single protein within the complex. The SASA was determined on the RACK1 protein of the 40s ribosomal subunit (pdb: 4V6I[20], UniProt ID P38011). A plot that correlates the residue relative SASA values to the residue protection factor (described in chapter 2), demonstrates a good correlation between the two parameters (Figure 3.4). The data fits well to a linear fit with an correlation coefficient of -0.75 with a p-value of 0.02. There is a possibility that protein-protein interactions are occurring that are not taken into account in the SASA calculations, which could by why the correlation value is not higher.

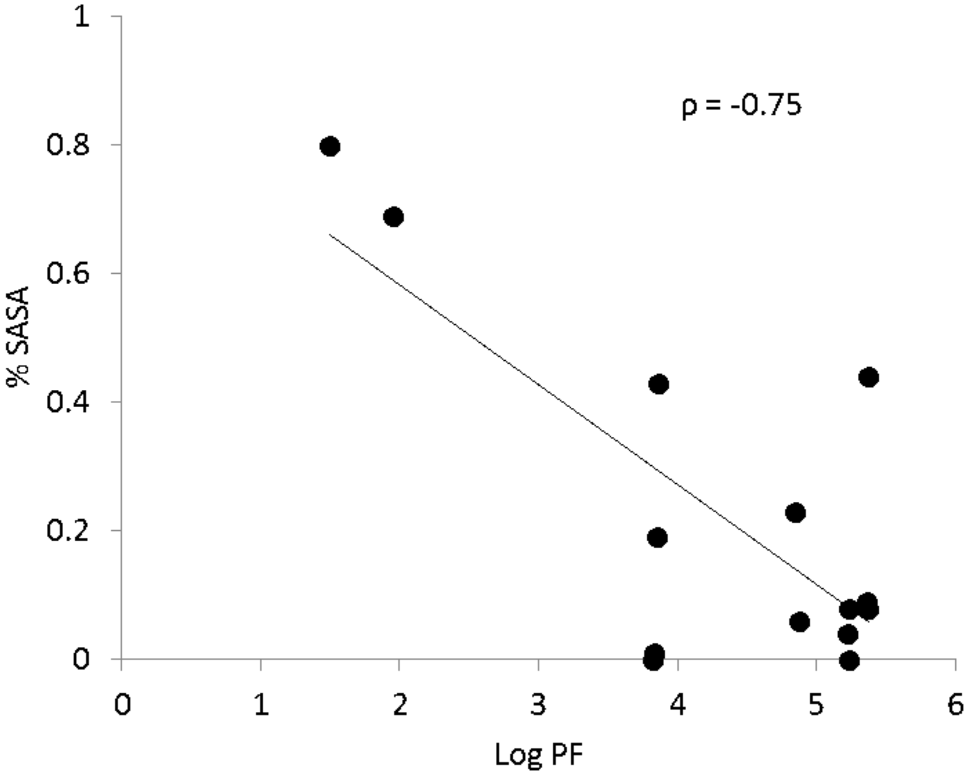


Figure 3.4. Fractional SASA values versus Log PF values of MudPIT identified residues of RACK1, chain Aa of the 4V6I[20] ribosomal structure, illustrating the linear relationship between the values.

3.4 Discussion

An advantage of using protein footprinting coupled with mass spectrometry for protein structural analysis is the ability to study large protein complexes. Analysis of these complexes is often hampered with other structural tools such as X-ray crystallography and NMR. Although MS analysis has the capability for analysis of complex proteins systems, the nature of data dependent acquisition limits the number of identifications achieved in analysis. Since DDA analysis focuses on the highest abundant proteins at a given time, it is often difficult to identify low abundant peptides with one-dimensional chromatography. This provides a challenge for oxidative labeling where it is advantageous to limit the levels of oxidation; thus, many oxidized peptides have a low abundance. The ability to carry out oxidative labeling on large protein complexes requires the capability to identify low abundant peptides.

The application of two-dimensional MudPIT chromatography to an oxidatively modified yeast lysate sample increased the number of identified proteins and peptides over one-dimensional chromatography. Yeast lysate (strain ATCC 204508 / S288c) contains 6,721 proteins and is indicative of a complex system. The increase in identification is most significant for oxidatively modified peptides where an almost 3-fold increase in identifications is observed (Figure 3.1 bottom). The higher abundance of identifications for oxidized peptides provides more detailed information on the proteins being analyzed. When investigating individual proteins, the benefit of MudPIT is further revealed. For both pyruvate kinase 1 and phosphoglycerate kinase, MudPIT identifies 5- and 2.6-fold higher numbers of oxidatively modified residues than 1D-DDA. There were peptides identified by 1D-DDA that were not observed with the MudPIT method, however. To gain as complete coverage as possible, it may be necessary to perform 1D-DDA and MudPIT in tandem.

Examining the intensity of peptides identified by MudPIT indicates this method is detecting lower abundant proteins. However, intensity alone does not account for the increased number of peptides identified by MudPIT. Another factor that may influence the number of IDs is ionization efficiency. Co-elution of peptides that compete for efficient ionization could lead to suppression of some peptides by higher abundant peptides. These suppressed peptides may be of lower abundance than their co-elution partners but are not low enough to be in the 1.00E+04 intensity range. Two dimensional chromatography could lead to better separation and reduction in co-elution and ionization suppression.

The identification of 52 of the 70 proteins in the ribosome complex demonstrates that coupling FPOP with MudPIT would be effective for studying large complexes in lysates. However, this approach could likely be improved by further enriching the protein complex with methods such as tandem affinity purification. Comparing extent of oxidation of residues to SASA calculations established a good correlation between the data. Since MudPIT analysis occurs over a longer time-scale than one-dimensional chromatography, there is an opportunity for spurious oxidation. Correlation of oxidative modification levels with solvent accessibility demonstrates that the sample is not adversely affected by the long MudPIT analysis.

3.5 Conclusion

This chapter has demonstrated the coupling of the MudPIT separation strategy with FPOP. The ability to obtain greater sequence coverage for oxidatively modified peptides increases the efficacy of FPOP for megadalton complexes. This is necessary to obtain more complete structural information on proteins using oxidative labeling. The data presented here has validated that MudPIT can provide this increased sequence coverage, yielding a more comprehensive understanding of the protein(s) or complex of interest.

3.6 References

1. Simin D. Maleknia, Michael Brenowitz, and Mark R. Chance. Millisecond radiolytic modification of peptides by synchrotron x-rays identified by mass spectrometry. *Analytical Chemistry*, 71(18):3965–3973, 1999. ISSN 0003-2700. doi: 10.1021/ac990500e.

2. Guozhong Xu and Mark R. Chance. Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Analytical Chemistry*, 77(14):4549–4555, 2005. ISSN 0003-2700. doi: 10.1021/ac050299+.

3. Liwen Wang and Mark R. Chance. Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Analytical Chemistry*, 83(19): 7234–7241, 2011. ISSN 0003-2700. doi: 10.1021/ac200567u.

4. Caroline Watson, Ireneusz Janik, Tiandi Zhuang, Olga Charvátová, Robert J. Woods, and Joshua S. Sharp. Pulsed electron beam water radiolysis for submicrosecond hydroxyl radical protein footprinting. *Analytical Chemistry*, 81(7): 2496–2505, 2009. ISSN 0003-2700. doi: 10.1021/ac802252y.

5. Jeffrey J. Hayes, Laurance Kam, and Thomas D. Tullius. *Footprinting Protein-DNA Complexes with Γ-Rays*, volume Volume 186, pages 545–549. Academic Press, 1990. ISBN 0076-6879. doi: 10.1016/0076-6879(90)86148-O.

6. DavidM Hambly and MichaelL Gross. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *Journal of the American Society for Mass Spectrometry*, 16(12): 2057–2063, 2005. ISSN 1044-0305. doi: 10.1016/j.jasms.2005.09.008.

7. Brian C. Gau, Joshua S. Sharp, Don L. Rempel, and Michael L. Gross. Fast photochemical oxidation of protein footprints faster than protein unfolding. *Analytical Chemistry*, 81(16):6563–6571, 2009. ISSN 0003-2700. doi: 10.1021/ac901054w.

8. Thomas R. M. Barends, Lutz Foucar, Sabine Botha, R. Bruce Doak, Robert L. Shoeman, Karol Nass, Jason E. Koglin, Garth J. Williams, Sebastien Boutet, Marc Messerschmidt, and Ilme Schlichting. De novo protein crystal structure determination from x-ray free-electron laser data. *Nature*, 505(7482):244–247, 2014. ISSN 0028-0836. doi: 10.1038/nature12773.

9. Alessandro Ori, Amparo Andrés-Pons, and Martin Beck. *Chapter 6 - the Use of Targeted Proteomics to Determine the Stoichiometry of Large Macromolecular Assemblies*, volume Volume 122, pages 117–146. Academic Press, 2014. ISBN 0091-679X. doi: 10.1016/B978-0-12-417160-2.00006-0.

10. Dirk A. Wolters, Michael P. Washburn, and John R. Yates. An automated multidimensional protein identification technology for shotgun proteomics. *Analytical Chemistry*, 73(23):5683–5690, 2001. ISSN 0003-2700. doi: 10.1021/ac010617e.

11. Michael P. Washburn, Dirk Wolters, and John R. Yates. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech*, 19(3):242–247, 2001. ISSN 1087-0156. doi: 10.1038/85686.

12. Carlee S. McClintock, Jerry M. Parks, Marshall Bern, Pavan K. GhattyVenkataKrishna, and Robert L. Hettich. Comparative informatics analysis to evaluate site-specific protein oxidation in multidimensional lc–ms/ms data. *Journal of Proteome Research*, 12(7):3307–3316, 2013. ISSN 1535-3893. doi: 10.1021/pr400141p.

13. Lisa M. Jones, Justin B. Sperry, James A. Carroll, and Michael L. Gross. Fast photochemical oxidation of proteins for epitope mapping. *Analytical Chemistry*, 83(20):7657–7661, 2011. ISSN 0003-2700. doi: 10.1021/ac2007366.

14. ThermoScientific. Tech tip # 49 acetone precipitation of proteins, 2009.

15. Laurence Florens and MichaelP Washburn. *Proteomic Analysis by Multidimensional Protein Identification Technology*, volume 328 of *Methods in Molecular Biology*™, book section 11, pages 159–175. Humana Press, 2006. ISBN 978-1-58829-519-4. doi: 10.1385/1-59745-026-X:159.

16. Paul Taylor, Peter A. Nielsen, Morten B. Trelle, Ole B. Hørning, Michael B. Andersen, Ole Vorm, Michael F. Moran, and Thomas Kislinger. Automated 2d peptide separation on a 1d nano-lc-ms system. *Journal of Proteome Research*, 8(3):1610–1616, 2009. ISSN 1535-3893. doi: 10.1021/pr800986c.

17. Melissa S. Jurica, Andrew Mesecar, Patrick J. Heath, Wuxian Shi, Thomas Nowak, and Barry L. Stoddard. The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure*, 6(2):195–210, 1998. ISSN 0969-2126. doi: 10.1016/S0969-2126(98)00021-5.

18. H. C. Watson, N. P. Walker, P. J. Shaw, T. N. Bryant, P. L. Wendell, L. A. Fothergill, R. E. Perkins, S. C. Conroy, M. J. Dobson, M. F. Tuite, and et al. Sequence and structure of yeast phosphoglycerate kinase. *Embo j*, 1(12): 1635–40, 1982. ISSN 0261-4189 (Print) 0261-4189.

19. L. M. de Godoy, J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Frohlich, T. C. Walther, and M. Mann. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 455(7217):1251–4, 2008. ISSN 1476-4687 (Electronic) 0028-0836 (Linking). doi: 10.1038/nature07341.

20. Jean-Paul Armache, Alexander Jarasch, Andreas M. Anger, Elizabeth Villa, Thomas Becker, Shashi Bhushan, Fabrice Jossinet, Michael Habeck, Gülcin Dindar, Sibylle Franckenberg, Viter Marquez, Thorsten Mielke, Michael Thomm, Otto Berninghausen, Birgitta Beatrix, Johannes Söding, Eric Westhof, Daniel N. Wilson, and Roland Beckmann. Cryo-em structure and rrna model of a translating eukaryotic 80s ribosome at 5.5-Å resolution. *Proceedings of the National Academy of Sciences*, 107(46):19748–19753, 2010. doi: 10.1073/pnas.1009999107.

21. Leigh Willard, Anuj Ranjan, Haiyan Zhang, Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart. Vadar: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Research*, 31(13):3316–3319, 2003. ISSN 0305-1048 1362-4962.

22. Haiyuan Yu, Pascal Braun, Muhammed A. Yıldırım, Irma Lemmens, Kavitha Venkatesan, Julie Sahalie, Tomoko Hirozane-Kishikawa, Fana Gebreab, Na Li, Nicolas Simonis, Tong Hao, Jean-François Rual, Amélie Dricot, Alexei Vazquez, Ryan R. Murray, Christophe Simon, Leah Tardivo, Stanley Tam, Nenad Švrzikapa, Changyu Fan, Anne-Sophie de Smet, Adriana Motyl, Michael E. Hudson, Juyong Park, Xiaofeng Xin, Michael E. Cusick, Troy Moore, Charlie Boone, Michael Snyder, Frederick P. Roth, Albert-László Barabási, Jan Tavernier, David E. Hill, and Marc Vidal. High-quality binary protein interaction map of the yeast interactome network. *Science*, 322(5898):104–110, 2008. doi: 10.1126/science.1158684.

4. DEVELOPMENT OF A MICRO-FLOW SYSTEM FOR IN CELL PROTEIN FOOTPRINTING ANALYSIS

4.1 Introduction

Hydroxyl radical (·OH) based footprinting coupled with mass spectrometry is increasingly gaining popularity for studying the structure of proteins for a number of reasons, including the similarity of ·OH to water, the various methods for generating radicals, and the well characterized ·OH reactivities with amino acid side chains.[1–3] FPOP, which was the method of ·OH generation used for the work in this chapter, generates ·OH through laser induced photolysis of hydrogen peroxide, a reaction that occurs on a microsecond timescale, theoretically eliminating structural changes induced by labeling.[4,5] FPOP, as well as other means of radical generation, are typically used for *in vitro* studies, which cannot fully account for the numerous interactions and other variables taking place within a cell.

Recently, a novel in cell FPOP (IC-FPOP) application was reported in which FPOP was used to label proteins in live cells, allowing for the complexity of the cellular environment to be accounted for.[6] While this initial report demonstrates a new approach for probing the SA of proteins, they were only able to identify 105 endogenous proteins that were oxidatively modified, making up a small fraction of the estimated 20,000 proteins (based on the number of human Swiss-Prot[7] verified proteins) in the Vero cells used in the study. The work in this chapter focuses on improving IC-FPOP identifications, and discusses a flow system prototype that was designed and built specifically for IC-FPOP.

When considering the requirements for IC-FPOP, flow cytometers have a key aspect in common; particles in solution are ordered into a centrally focused stream, a process known as hydrodynamic focusing.[8] Hydrodynamic focusing, an effect of fluid dynamics, arises when two fluids in contact in a straight microfluidic channel are introduced at different velocities (Figure 4.1).[9] The fluid moving at a higher velocity, referred to as the sheath, fills a larger portion of that channel which compresses the slower moving fluid.

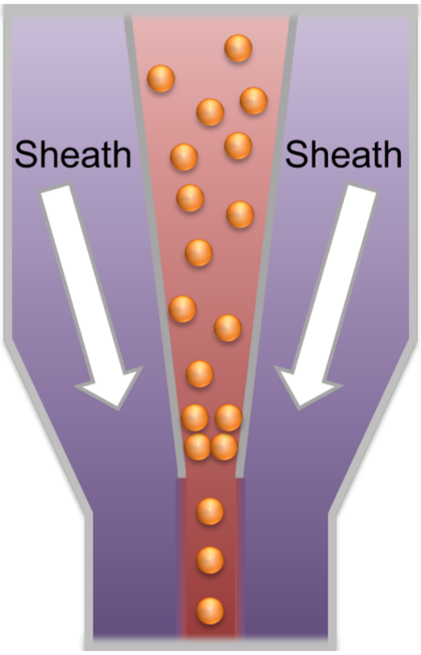


Figure 4.1. Hydrodynamic focusing of particles.

Like flow cytometry, this work requires that the cells be centrally focused so there is an even distribution of the cells as the cells are irradiated with the excimer laser used in our FPOP studies. Previous work has been done creating miniature microfluidic flow cells that hydrodynamically focus the cells. However, the low cost options to make these chips are not UV transparent at 248 nm, and UV transparent materials would be costly. Additionally, our lab is not equipped with the equipment that would be needed to manufacture a chip from these materials. My solution was to use larger ID fused silica tubing, which is UV transparent with the coating removed, and would fit 1/16 flat bottom HPLC fittings as the flow cell. The remainder of the parts used are common LC-MS fittings, with a design that makes it easy to change out any part when necessary. The efficacy of the flow system was evaluated for both its ability to centrally focus the cells as well as its performance when used for IC-FPOP, with the results compared to those previously published.

4.2 Experimental Section

4.2.1 Materials

All chemicals were obtained from Thermo Fisher Scientific (Waltham, MA) unless otherwise noted. Flow cell fittings were obtained from IDEX Health & Science (Oak Harbor, WA). Polyimide coated fused silica was obtained from Polymicro Technologies (Phoenix, AZ).

4.2.2 Flow Cell Construction

The flow cell (Figure 4.2) was constructed entirely from standard IDEX parts and Polymicro fused silica capillaries, with a modular design that can be adapted for differing experimental requirements. The cellular analyte solution and H_2O_2 enter the system through separate capillaries (360 µm OD, 75 µm ID) where they are mixed in a PEEK mixing tee to keep the exposure time to H_2O_2 constant throughout the experiment, and to prevent endogenous catalase from completely decomposing the H_2O_2. The capillary carrying the mixed cellular analyte and H_2O_2 (360 µm OD, 75 µm ID) is mounted so that it passes through the PEEK cross and is concentric with the larger diameter flow cell capillary (673 µm OD, 450 µm ID). The two capillaries (360 µm OD, 150 µm ID) carrying the sheath buffer are flush mounted with the screw which enables the sheath buffer to completely surround the cellular analyte/H_2O_2 capillary before any cells enter the capillary, centrally focusing the cells and compressing the diameter of the solution carrying the cells/H_2O_2. Cells

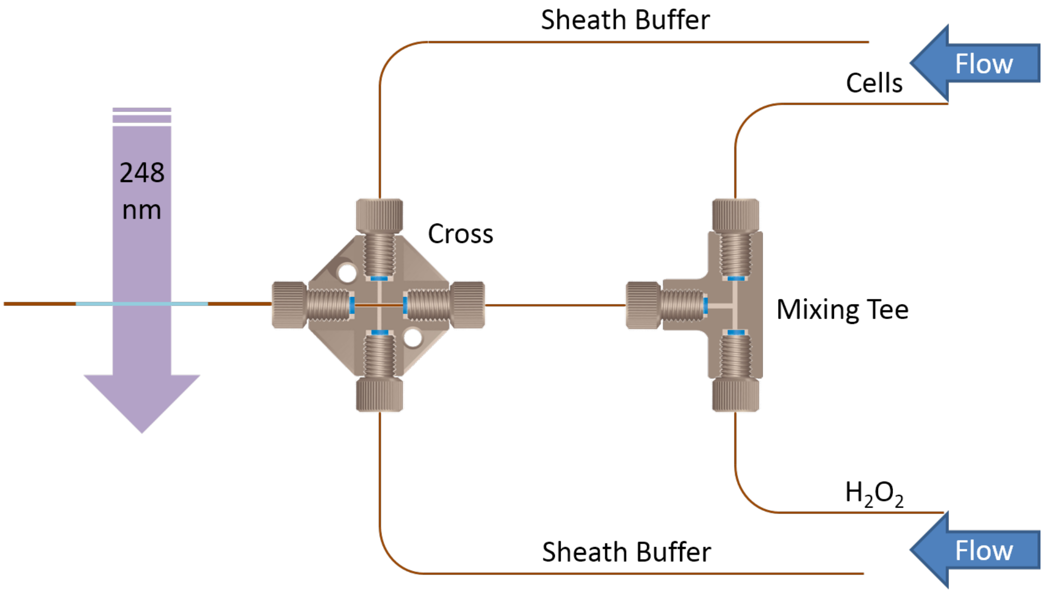


Figure 4.2. IC-FPOP Flow Cell Schematic: Blue arrows indicate flow direction, and amber lines represent polyimide coated fused capillary tubing (not to scale). The cellular solution and hydrogen peroxide are mixed in a tee to prevent breakdown by endogenous catalase. The capillary containing the cellular analyte is mounted so that the buffer surrounds the capillary, creating a sheath to centrally focus the cells. Optimal conditions were observed with a 10 to 1 sheath buffer to cellular analyte ratio.

are irradiated through a window in the flow cell capillary created by burning off a portion of the polyimide coating.

4.2.3 Oxidative Labeling

IC-FPOP was completed as previously described.[6] Vero cells (ATCC, Manassas, VA) were grown to 70% confluence in a T75 flask in Dulbecco’s Modified Eagle Medium (DMEM) with streptomycin/penicillin. The cells were trypsinized, centrifuged, and resuspended in 6 mL sterile PBS. Each sample was prepared with 450 µL of cells and 50 µL PBS, and drawn into the flow system. The cells were mixed in the syringe using a VP 710 tumble stirrer (V&P Scientific, San Diego, CA) with 6 VP 724F stir discs (V&P Scientific, San Diego, CA) to prevent cell settling. Samples were passed through the flow cell system at a total flow rate of 484 µL/min. The excimer laser was set to a pulse frequency of 20 Hz, with laser energy of 124 mJ, and a pulse width of 2.55 mm. Cells were collected in a tube containing 4 mL of a quench solution comprised of 100 mM N,N′-Dimethylthiourea (DMTU, Sigma-Aldrich, St. Louis, MO) and 100 mM N-tert-butyl-α-phenylnitrone (PBN). DMSO (1%) added to the quench solution just prior to infusion to inhibit methionine sulfoxide reductase. Cells were labeled in biological duplicate, each in technical triplicate with an equal number of controls (no laser irradiation).

4.2.4 Cell Lysis and Proteolysis

Post FPOP, cells were centrifuged, rinsed with PBS, resuspended in 100 µL RIPA Lysis and extraction buffer, frozen in liquid nitrogen, and stored at -80 °C overnight. Cells were thawed, incubated at 95 °C for 5 min, then cooled on ice for 5 min. After cooling, 2.5 units of nuclease were added to each sample, incubated at room temperature for 15 min, centrifuged, and the supernatant was transferred to a new tube. The cell lysate was reduced with DTT, alkylated with IAA and acetone precipitated overnight. The precipitate was resuspended in 25 mM triethylammonium bicarbonate buffer (TEABC, Sigma-Aldrich, St. Louis, MO) and digested with trypsin at 37 °C overnight. Digestion was quenched by adding formic acid to a final concentration of 5%.

4.2.5 LC-MS/MS

Analysis was completed using an UltiMate 3000 RSLC and a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA) as previously described.[6,10] For each experiment, the digest was loaded onto a 2 cm Acclaim Pepmap 100 C18 trap column (Thermo Fisher Scientific, Waltham, MA) and washed for 40 minutes with loading buffer (2% acetonitrile 0.1% formic acid) at a flow rate of 2.5 µL/min.

The samples were separated on a 75 µm inner diameter reverse phase analytical column packed in-house with a 30 cm bed of Magic 5 µm C18 particles (Michrom Bioresources Inc., Auburn, CA). Peptides were eluted with a 141 min stepped gradient at a flow rate of 300 nL/min from 4% to 10% acetonitrile 0.1% formic acid over 1 min and from 10% to 45% acetonitrile 0.1% formic acid over 140 min. The total run time was 207 minutes including loading, washing, and equilibration time. MS1 spectra were acquired over an *m/z* range of 300–2000 at a resolving power of 70,000 for 400 *m/z* ions, with a dynamic exclusion of 20 s. The 25 most abundant ions were selected for MS2 at a resolving power of 17,500 for 400 *m/z* ions. Ions with a charge-state of +7, +8, and >+8 ions were rejected. AGC targets were set to 3e6 for MS1 and 1e5 for data-dependent MS2 with an underfill ratio of 5%, giving an intensity threshold of 5.0e3.

4.2.6 Data Analysis

All data files were searched using Proteome Discoverer version 1.4 (Thermo Fisher Scientific, Waltham, MA.) with Sequest HT and Mascot version 2.4 (Matrix Sciences Ltd., London, UK) against the Swiss-Prot reviewed human FASTA database containing 20,193 proteins, and EIC areas for each PSM were calculated using a custom multi-level workflow (as described in chapter 2). The fragment ion tolerance was set at 0.02 Da and the parent ion tolerance at 10 ppm, and enzyme specificity was set to trypsin with 1 missed cleavage. Peptides were ungrouped and filtered to a 5% FDR. The data was exported to Excel and summarized using the PowerPivot add-in. Proteins were accepted if at least 2 distinct peptides were identified with 5% FDR filter. Quantitation of FPOP oxidation on actin was completed as described in chapter 2.

4.2.7 Fluorescence Imaging

Vero cells were grown and harvested as described above. Cells were incubated in 300 nM 4′,6-diamidino-2-phenylindole (DAPI) for 5 minutes and resuspended in PBS. Tetramethylrhodamine (TMRM) and fluorescein sodium salt (FITC, Sigma-Aldrich, St. Louis, MO) were added at a concentration of 5 µg/mL to the cells and sheath buffer, respectively. Flow system imaging was completed on a Leica SP8 confocal laser scanning microscope (Leica Microsystems Inc., Buffalo Grove, IL) with a Nikon Fluor 40x/0.80W DLL objective (Nikon Instruments Inc., Melville, NY)

4.3 Results and Discussion

4.3.1 Flow Dynamics

Similar to flow cytometry, the flow system has a central capillary through which the cells are injected, surrounded by a faster flowing sheath buffer; laminar flow prevents the cells and sheath from mixing. However, there is one significant design difference between this flow system and those used in flow cytometers. Typically, the central capillary is tapered, and this, along with the faster flowing sheath, hydrodynamically focuses the cells.[8] The IC-FPOP experimental parameters as well as the backpressure limitations of the syringe pump and fittings did not allow for the central capillary to be tapered. As such, the blunt end of the central capillary in the IC-FPOP flow system creates an area that may potentially cause turbulence and disrupt laminar flow.

Fluorescence imaging was performed to evaluate the IC-FPOP flow system's effectiveness in hydrodynamically focusing the cells. Different fluorophores were added to each solution to ensure laminar flow was maintained and to establish the location of each solution within the flow cell; FITC (green) was added to the sheath PBS and TMRM (red) was added to the cellular solution for imaging. PBS with TMRM was used in place of H_2O_2 for imaging, and collectively, will be referred to as cellular analyte. In addition, cell nuclei were stained with DAPI (blue) prior to imaging.

In order to establish the average position of each of the flowing solutions, an average intensity projection was generated using the Icy bioimaging platform.[11,12] This shows the average intensity of each pixel location over all of the 2666 frames (recorded at 26.6 fps) in the run (Figure 4.3). To generate the projection, the RGB

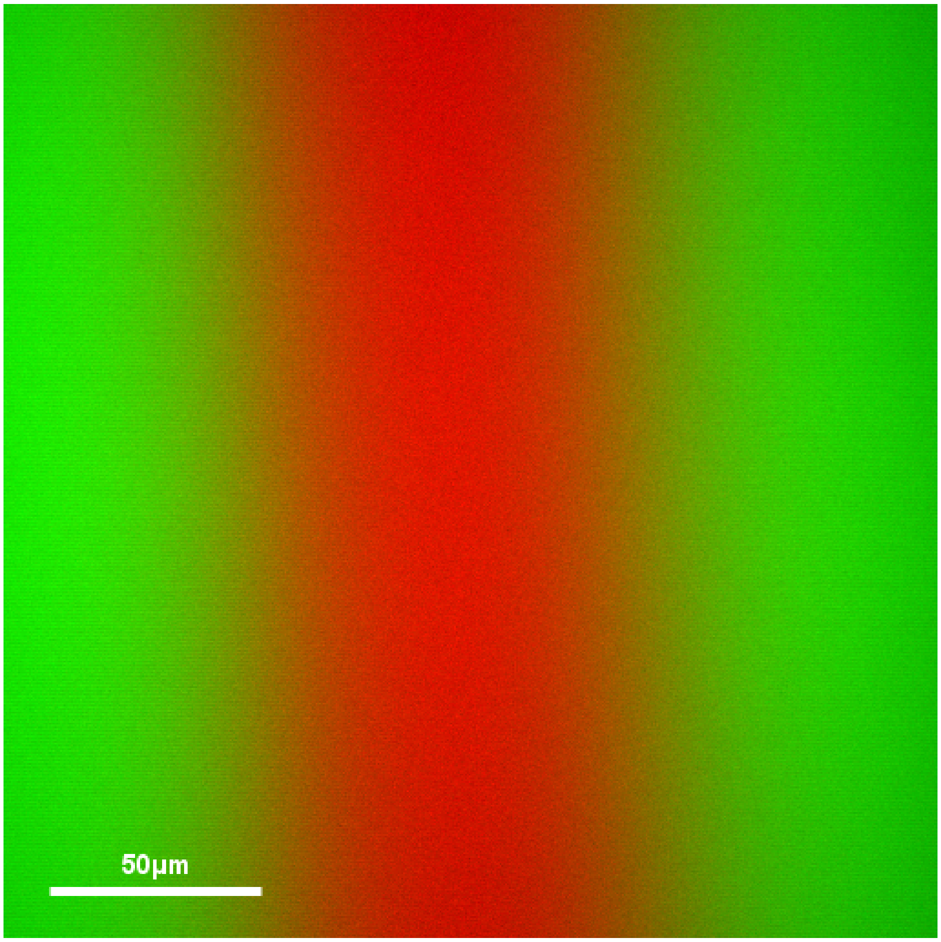


Figure 4.3. Flow system average intensity projection showing the locations of the sheath buffer(green) and cellular analyte (red) as they flow through the flow cell.

channels were separated from the multi-channel image stack, and an average intensity projection for the red (TMRM) and green (FITC) channels was performed;

the resulting projections were then merged. Analysis shows the cellular analyte was centrally compressed into a single stream approximately 50 μm in diameter. This is 25 μm smaller than the capillary ID, demonstrating a stream compression of 33%. The blurring between the sheath and cellular analyte is expected as it is an average over all of the frames; each individual frame is grainy due to the flow velocity and image capture speed. Other contributing factors may be vibrations causing slight movements of the capillary and crosstalk between the FITC and TMRM (Figure 4.3). To address this issue, a 3D average intensity heat map was generated (Figure 4.4) from the individual red and green projections. The map clarifies the blur between the channels and clearly shows the highest intensity of the sheath towards the sides of the field of view (left), and down the center for the cellular analyte (right). This illustrates that the flow system operating under laminar flow conditions as mixing due to turbulence would not display clear delineation in intensity.[13]

Fluorescence imaging was also used to determine the location of the cells as they pass through the flow cell capillary. In addition to the DAPI, the TMRM dye also stains the mitochondria of live cells as the accumulation of the dye is driven by the mitochondria membrane potential.[14,15] Again, an intensity projection was generated to show the location of the cells. However, the maximum intensity was used here, as each pixel location, on average, does not contain a cell. To generate this projection, a median filter was applied to remove many of the individual pixels. These pixels are likely from background noise or the TMRM that was in the cellular analyte that was not taken up by the cells. The maximum intensity projections for the filtered red (TMRM) and blue (DAPI) channels were merged and 2D heat map was generated from the merged projections (Figure 4.5). Nearly all of the pixels are from this map are located within the same dimensions of the cellular analyte from Figure 4.3 (a cropped portion of Figure 4.3 was added to the bottom of Figure 4.4 for ease of comparison. This demonstrates that the cells are kept in a narrow channel in the center of the flow cell capillary.

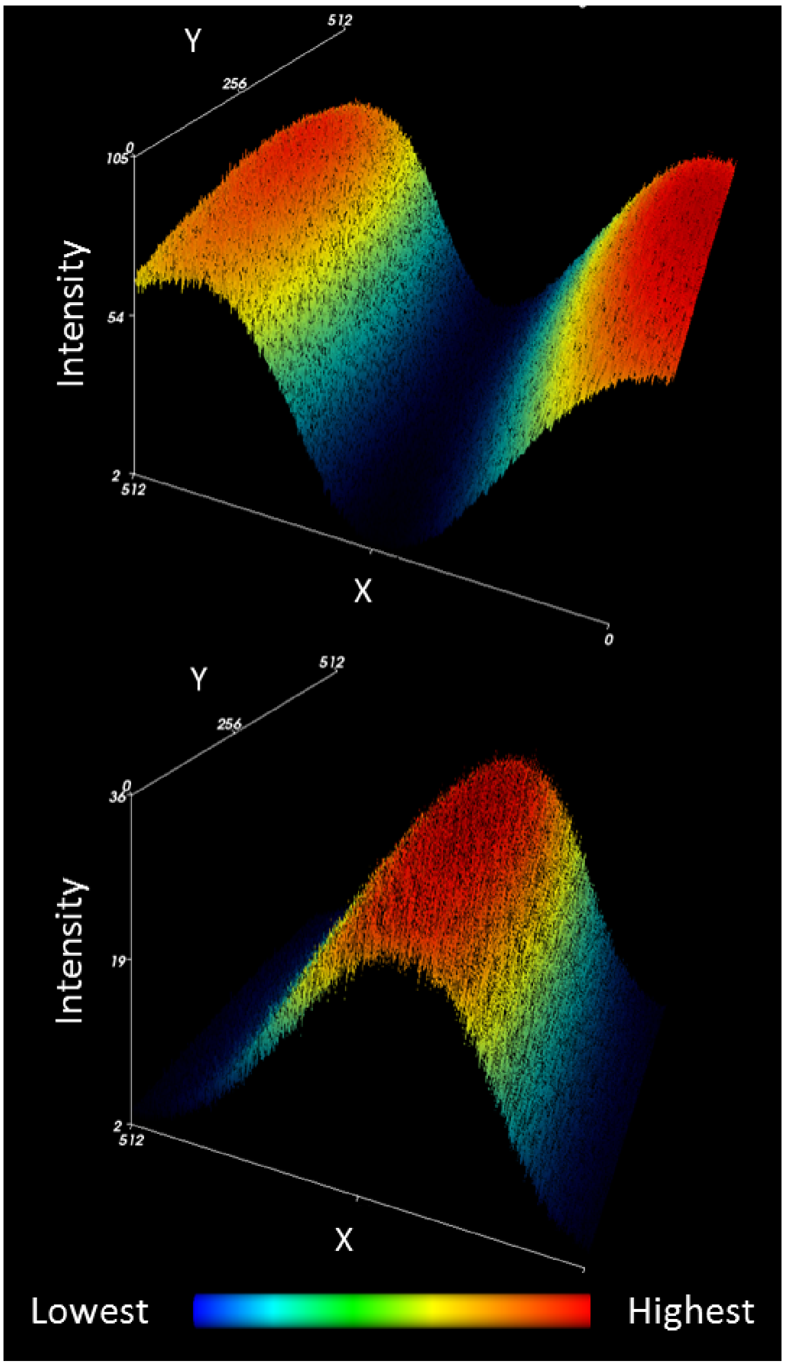


Figure 4.4. 3D average intensity heat map with the sheath on top and cellular analyte on the bottom. Lowest intensity is blue and highest is red.

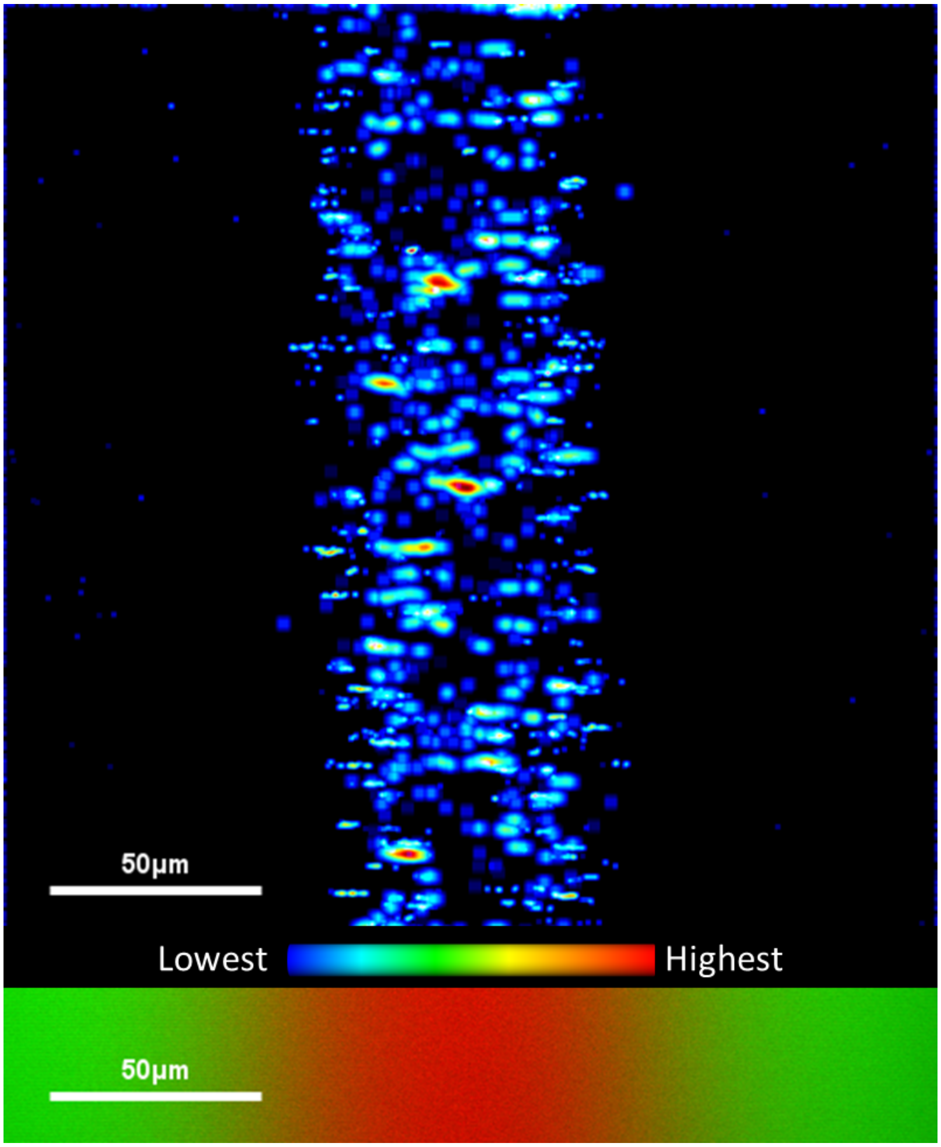


Figure 4.5. Maximum intensity heat map showing the locations of all detected cells. A cutout of Figure 4.3 is shown on the bottom to illustrate overlap.

As a confocal microscope was used for imaging, the images can only demonstrate that the cellular analyte is centrally focused in one XY plane of the flow cell capillary and cannot provide any evidence of 3D hydrodynamic focusing. To assess the 3D focusing, a cross-section YZ stack was collected (Figure 4.6). Although the focusing in the YZ plane is not as condensed as the XY plane, the cellular ana-

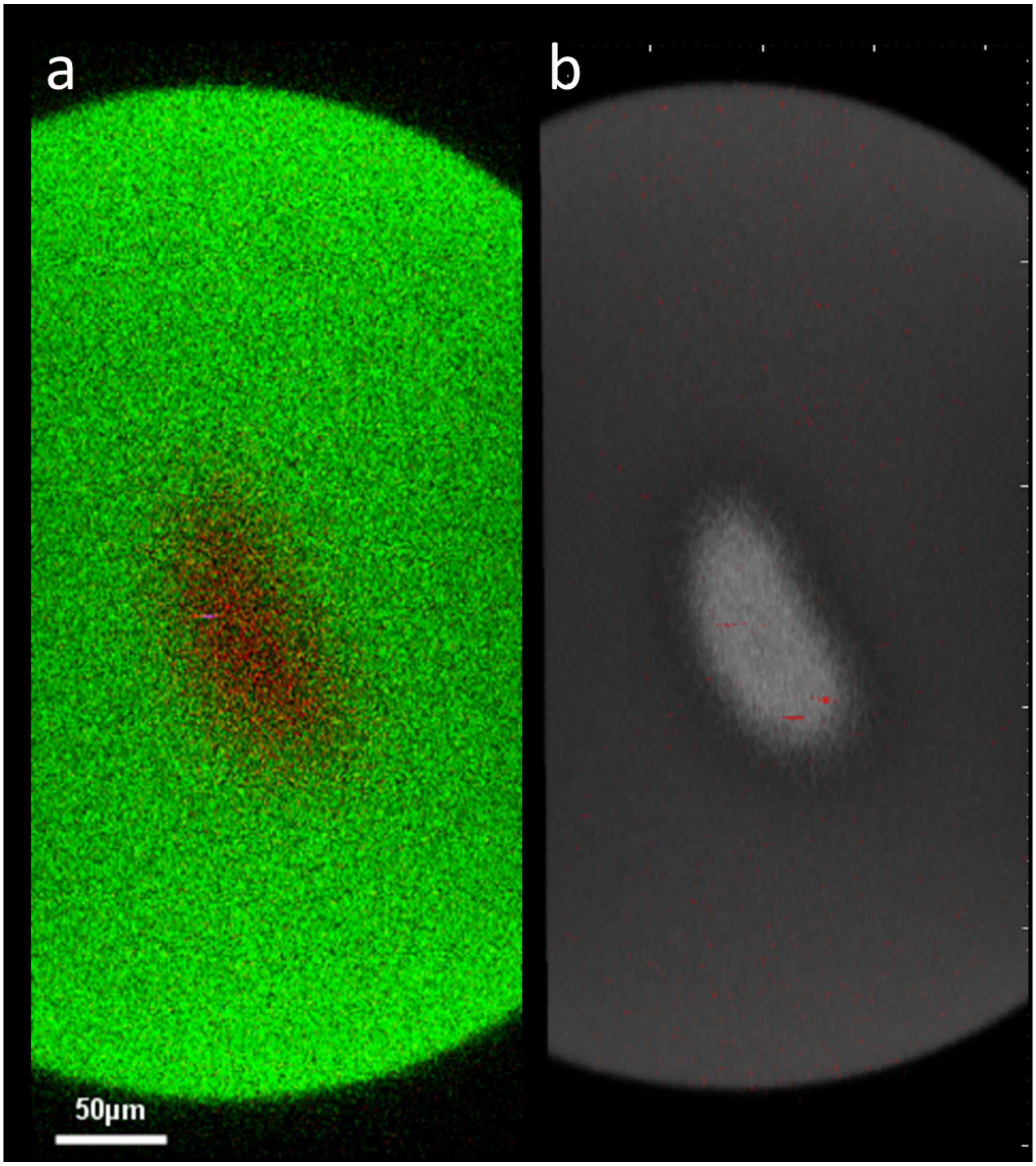


Figure 4.6. Cross-sectional YZ stack demonstrating 3D hydrodynamic focusing (a) and location of cells (b, red) with the sheath and cellular analyte buffer in gray for ease in viewing cells.

lyte is unquestionably centered through the flow cell capillary, confirming the flow system is able to centrally focus the cells. Additionally, the cross-sectional stack captured cells (Figure 4.6b) as they flowed through the capillary, verifying the cells stay centrally focused both vertically and horizontally. The discrepancy in focusing is likely due to the cross since the sheath buffer enters the flow system horizontally perpendicular to the cellular analyte capillary. While the focusing is adequate for IC-FPOP, 3D focusing could be improved by using a 6 port flow through manifold with the sheath buffer entering the flow system both horizontally and vertically perpendicular to the cellular analyte capillary.

4.3.2 IC-FPOP Using the Flow System

Method Comparison

IC-FPOP of Vero cells was completed as previously described using the flow system so the data could be compared to the previously published results.[6] However, there were 3 experimental design parameters, aside from the flow system, that deviated from those published. First, a flow rate of 484 µL/min was used. This was the fastest that could be used while keeping the 20 Hz pulse frequency with no exclusion volume. This is actually about half the published flow rate when scaled for the flow system. Second, a magnetic stirrer was used to prevent the cells from settling in the syringe. Lastly, there were minor deviations in the digestion protocol. The Thermo Scientific Pierce mass spectrometry sample kit protocol was used, but with reagent substitutions. RIPA and TEABC buffers were used, and only trypsin was used for digestion. All other parameters including cell prep, sample volume, quench and H_2O_2 final concentrations were kept consistent.

Increased Oxidized Protein Identifications

IC-FPOP using the flow system yielded a considerable number of oxidized proteins. The database search, using the PD method described in chapter 2, was completed against human FASTA sequences as the Vero cell proteome is incomplete with only 139 Swiss-Prot verified sequences. The search identified a total of 1391 endogenous proteins that were oxidatively modified over the 2 biological replicates. This corresponds to a 13.25-fold increase in oxidized proteins identified compared to the 105 identified in the previously published work (Figure 4.7a). Indi-

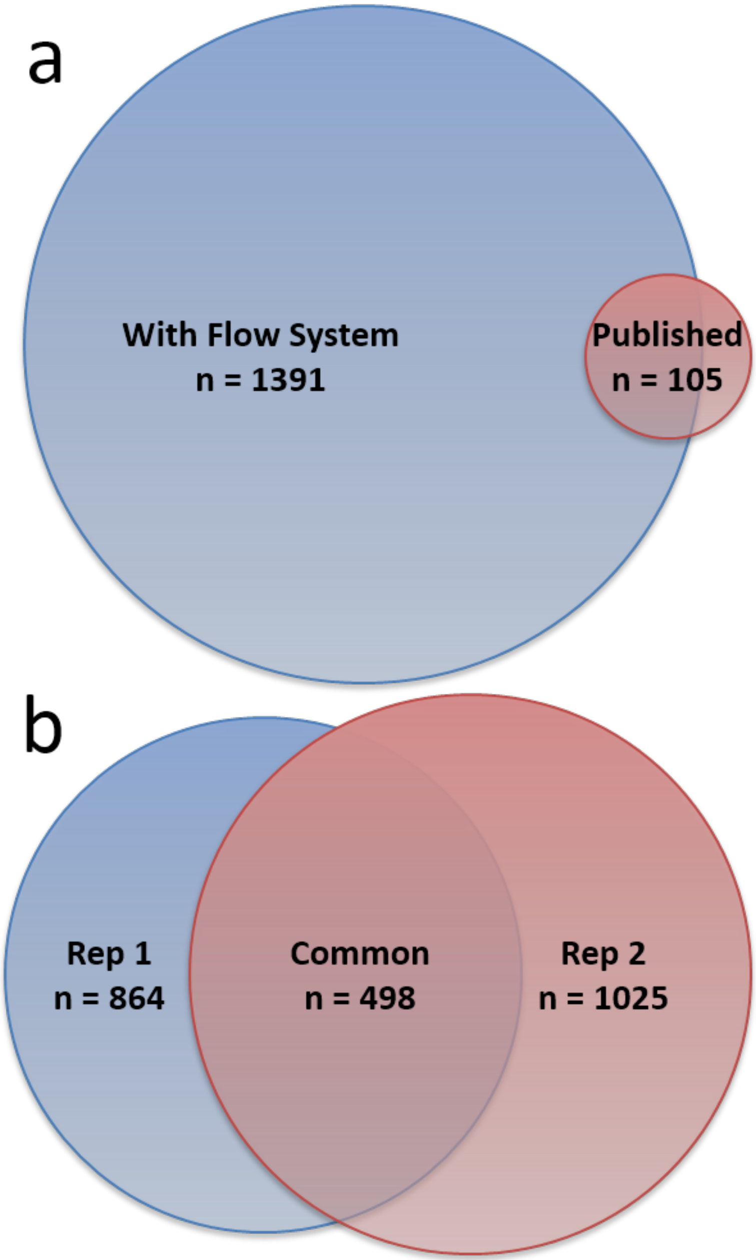


Figure 4.7. Visual comparison of the oxidized proteins identified with (blue) and without (red) using the flow system (a) and the flow system biological replicates (b).

vidually, the biological replicates also had sizeable gains, with an 8.23-fold increase for the first replicate (n = 864) and a 9.76-fold increase for the second (n = 1025) with 498 proteins in common between the replicates (Figure 4.7b).

The discernible advantage of using the flow system for IC-FPOP is likely attributed to several factors. First, the buffer sheath keeps the cells away from the walls of the flow cell capillary. As such, any cell aggregates that may begin to form will be pushed through the flow system and therefore cannot grow larger or accumulate on the walls. One can postulate that some of the cells exposed to the radiation will die and lyse, exposing the DNA which can lead to cell clumping.[16] These clumps may then stick to the wall of the capillary as the DNA would be attracted to the wall of the fused silica capillary.[17] Another factor is the method of adding the H_2O_2 to the sample. The H_2O_2 is in a separate syringe in the flow system, and mixed with the cells via a mixing tee just prior to entering the flow cell capillary. Based on the flow rate and sample volume in the published method, it takes just under 10 minutes for the entire sample to be labeled. Since the H_2O_2 is added as a single bolus, a significant portion of the H_2O_2 will be consumed prior to irradiation given that catalytic efficiency of the endogenous catalase is close to that of human catalase, 7.34e6 $s^{-1}M^{-1}$.[18] Finally, the smaller ID of the capillary (75 µm vs. 150 µm) used to deliver the cells in the flow system along with the compression from the hydrodynamic focusing may have an overall smaller dead volume than the published method depending on the length of the capillary used. The use of the stirrer to keep the cells suspended may also play a factor if the cells fell out of suspension before the end of the run in the published method.

Properties of Oxidized Proteins

There were a total of 1862 oxidized peptide sequences identified. The EIC peak areas from the controls (no irradiation) was subtracted from the sample peak areas to remove any background oxidation. Peptides were only included if the sample peak areas were greater than the controls. The number of oxidized sequences per protein ranged from 18 to 1, with an average of 1.8 sequences per protein. There were a total of 31229 PSMs corresponding to oxidation events, with oxidations occurring on 18 of the 20 amino acids (Figure 4.8a).

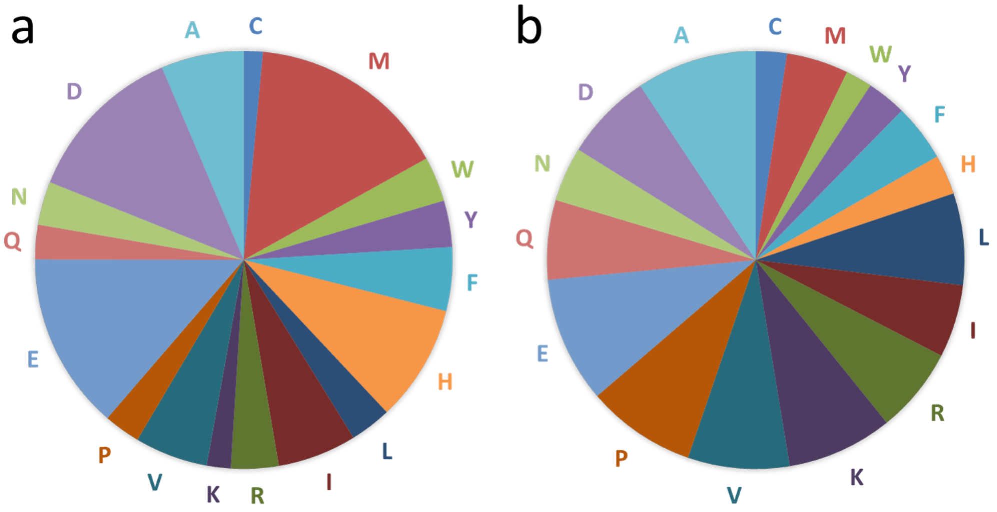


Figure 4.8. Amino acid distribution of oxidation events relative to its overall abundance in all the proteins identified (a) and the average occurrence of the residue being solvent exposed.

The oxidized proteins identified from IC-FPOP using the flow system were analyzed to identify from which cellular component they are found within the cell. Uniprot identifiers from the oxidized proteins were uploaded into the BiNGO[19] (version 3.03) network gene ontology plugin for Cytoscape[20] (version 3.3) to create a hierarchal network mapped to the generic GO Slim ontologies for cellular components (Figure 4.9). GO slim ontologies, which are a slimmed down version of GO terms that provide a higher level view, were used to provide a broader overview of the results. The proteins were found in 30 (out of 35) GO slim cellular component terms. Similar to the previous results, there were a large number of proteins located in the cytoplasm and membrane. However, there were also a significant number of proteins identified that are located within protein complexes, the cytoskeleton,

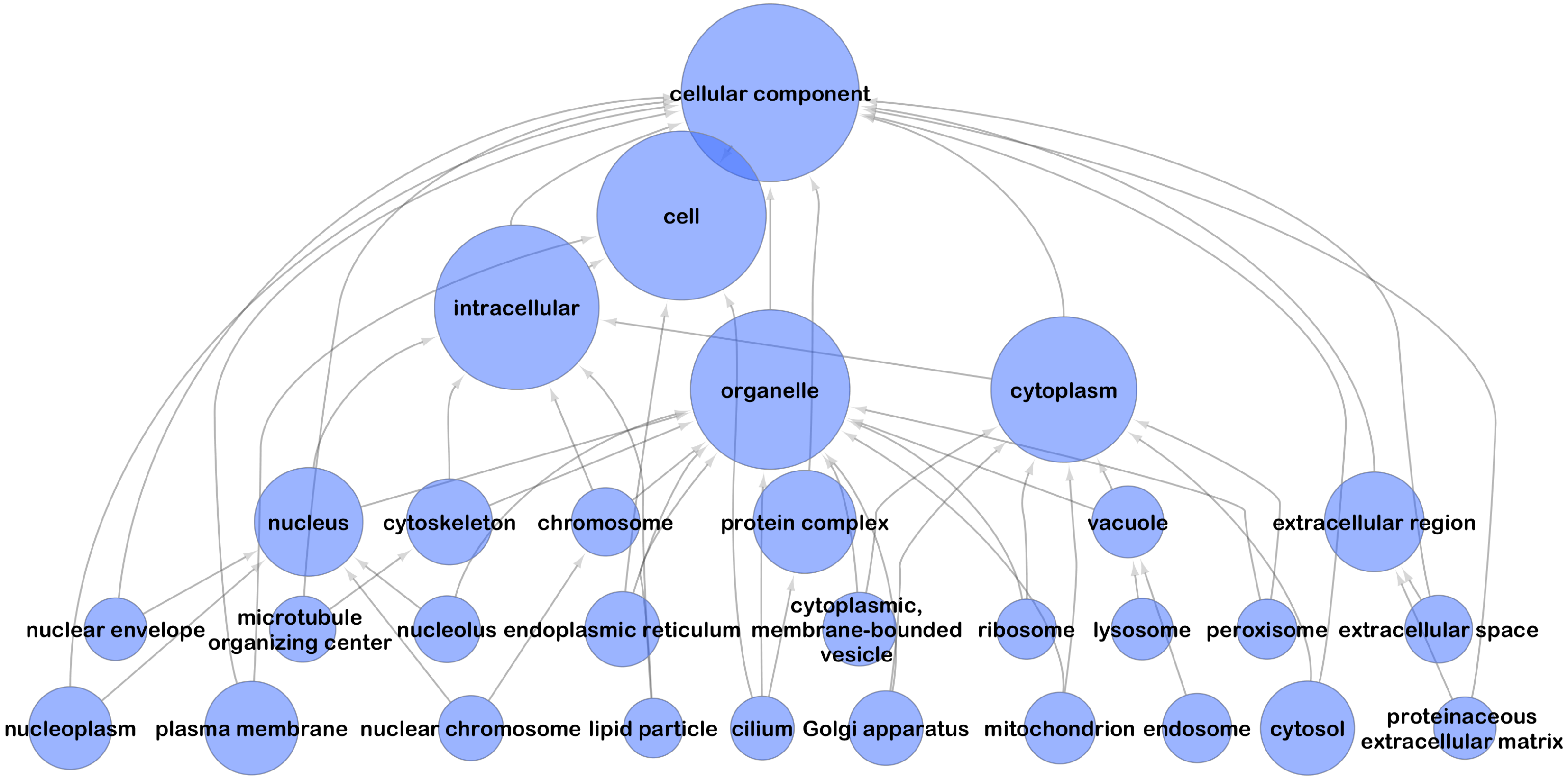

cellular component
cell
intracellular
organelle
cytoplasm
nucleus
cytoskeleton
chromosome
protein complex
vacuole
extracellular region
nuclear envelope
microtubule organizing center
nucleolus
endoplasmic reticulum
cytoplasmic, membrane-bounded vesicle
ribosome
lysosome
peroxisome
extracellular space
nucleoplasm
plasma membrane
nuclear chromosome
lipid particle
cilium
Golgi apparatus
mitochondrion
endosome
cytosol
proteinaceous extracellular matrix

Figure 4.9. Hierarchical network of oxidized proteins identified using the flow system mapped to the generic GO Slim ontologies for cellular components with the nodes sized by the percentage proteins.

and organelles, such as the nucleus. When compared to the 10 locations found in the previously published data, the results are fundamentally more comprehensive. Furthermore, the previous results were annotated Uniprot which uses 501 subcellular location terms In addition, identifying oxidized proteins in 86% of the GO slim locations provides evidence that using the mixing tee does indeed mix the cells with the H_2O_2, and does not affect the ability of H_2O_2 to fully permeate the cells before irradiation.

The dynamic range of the IC-FPOP method was evaluated in the previous work by comparing the oxidized proteins identified with their expression level (transcripts per million, TPM) in human kidney cells. From this, they estimated a dynamic range of 3 orders of magnitude. This was based on the fold change between the protein with the highest TPMs, heat shock cognate protein at 4635, and the lowest TPMs, protein shroom2 at 4 TPMs. Both of these proteins were identified as oxidized using the flow system, indicating that using the flow system does not decrease the dynamic range of IC-FPOP. The full dynamic range of IC-FPOP using the flow system was assessed using expression data acquired from the Human Protein Atlas. The database contains expression profiles on 83 normal human cell lines from 44 tissues, including the kidney. The maximum and minimum fragments per kilobase million (FKPM) for oxidized proteins was 1227 and 0.1, yielding a fold change of 1.25e4. The protein identified with the highest expression levels was elongation factor 1-alpha 1 (EEF1A1), and 137 different identified proteins with the lowest expression levels. The proteins identified in published report were also evaluated with this database for an equivalent comparison and resulted in a fold change of 350, with 278 FKPM for heat shock cognate protein and 0.8 FKPM for shroom2. Based on these expression profiles, using the flow system resulted in an approximately 2 orders of magnitude greater dynamic range.

Correlation of FPOP Modification to Solvent Accessibility

FPOP coupled with MS is used as a method to probe solvent accessibility in order to gain insight on protein structure. As previously reported, IC-FPOP, like *in vitro* FPOP, does correlate to solvent accessibility.[6] Oxidation events identified using the flow system were also investigated to assess the correlation to SA. The greatest number of oxidation events occurred on Glu and Asp, accounting for 22% and 13% of the oxidations, respectively. This is expected as less than 3% of these residues tend to be buried.[21,22] However, Met had the highest percentage of oxidation events relative to its overall abundance in all the proteins identified (Figure 4.8a). This is significantly higher than the average occurrence of this residue being solvent exposed (Figure 4.8b).[22] Met is highly reactive with ·OH, with a rate constant of 8.5e9, which is likely the reason relatively high number of oxidation events.[23]

Oxidation of actin, at both peptide and residue levels, was completed as this protein was used previously as a measure of probing SA. A total of 20 peptides were identified with 7 being oxidatively modified (Figure 4.10a). Generally, the overall trend is similar to the previously published results; however, peptide 40-50 is less oxidized with the flow system. This peptide has a total of 1467 PSMs for unoxidized precursors and only 28 PSMs for the oxidized precursors, only 1.8% of the total PSMs. Identification of oxidized species could be improved by increasing the dynamic exclusion and/or applying Mulitdimensional Protein Identification Technology (MudPIT) which, as demonstrated in chapter 3. Residue level analysis identified 11 residues that were oxidized in actin. Protection factor values were calculated (as described in chapter 2) and plotted against the relative SASA values (Figure 4.10b), obtained using VADAR[24], for the open and tight states of actin (PDB 1HLU and 2BTF). Correlations were in agreement with the previously published results, with a stronger correlation to the open state, demonstrating that data obtained using the flow system correlates to SA.

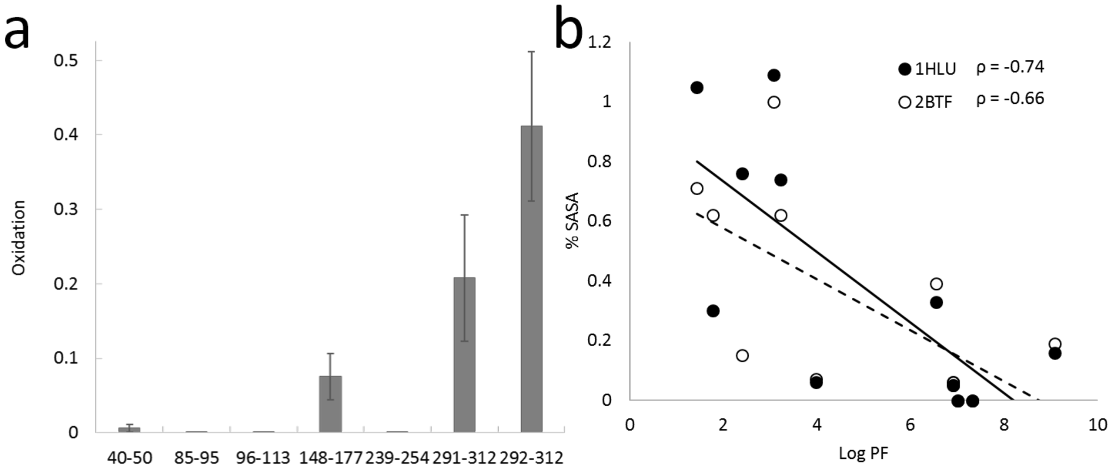


Figure 4.10. Oxidation levels of 7 actin peptides (a) and correlation of residue level log protection factors to the relative SASA of open (1HLU) and tight (2BTF) states of actin (b).

4.4 Conclusion

The work presented in this chapter has demonstrated a process improvement for IC-FPOP using a flow system prototype that was designed and built specifically for IC-FPOP. This low cost, modular design hydrodynamically focuses the cells in a small stream down the center of the flow cell capillary, which prevented cell aggregation and system clogging. Additionally, the focusing of the cells ensures that radiation exposure remains consistent throughout the experiment. The IC-FPOP data presented here has validated that using the flow system for IC-FPOP results in a dramatic increase in the identification of oxidized proteins, with a 13.25-fold increase observed, without compromising the dynamic range of the method or the ability to probe the SA of endogenous proteins.

4.5 References

1. Liwen Wang and Mark R. Chance. Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. *Analytical Chemistry*, 83(19): 7234–7241, 2011. ISSN 0003-2700. doi: 10.1021/ac200567u.
2. Caroline Watson, Ireneusz Janik, Tiandi Zhuang, Olga Charvátová, Robert J. Woods, and Joshua S. Sharp. Pulsed electron beam water radiolysis for sub-microsecond hydroxyl radical protein footprinting. *Analytical Chemistry*, 81(7): 2496–2505, 2009. ISSN 0003-2700. doi: 10.1021/ac802252y.
3. Jeffrey J. Hayes, Laurance Kam, and Thomas D. Tullius. *Footprinting Protein-DNA Complexes with Γ-Rays*, volume Volume 186, pages 545–549. Academic Press, 1990. ISBN 0076-6879. doi: 10.1016/0076-6879(90)86148-O.
4. DavidM Hambly and MichaelL Gross. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. *Journal of the American Society for Mass Spectrometry*, 16(12): 2057–2063, 2005. ISSN 1044-0305. doi: 10.1016/j.jasms.2005.09.008.
5. Brian C. Gau, Joshua S. Sharp, Don L. Rempel, and Michael L. Gross. Fast photochemical oxidation of protein footprints faster than protein unfolding. *Analytical Chemistry*, 81(16):6563–6571, 2009. ISSN 0003-2700. doi: 10.1021/ac901054w.
6. Jessica A. Espino, Vishaal S. Mali, and Lisa M. Jones. In cell footprinting coupled with mass spectrometry for the structural analysis of proteins in live cells. *Analytical Chemistry*, 87(15):7971–7978, 2015. ISSN 0003-2700. doi: 10.1021/acs.analchem.5b01888.
7. Amos Bairoch and Rolf Apweiler. The swiss-prot protein sequence database and its supplement trembl in 2000. *Nucleic Acids Research*, 28(1):45–48, 2000. ISSN 0305-1048 1362-4962.
8. Misha Rahman. Introduction to flow cytometry, 2006.
9. Joel P. Golden, Gusphyl A. Justin, Mansoor Nasir, and Frances S. Ligler. Hydrodynamic focusing – a versatile tool. *Analytical and bioanalytical chemistry*, 402(1):325–335, 2012. ISSN 1618-2642 1618-2650. doi: 10.1007/s00216-011-5415-3.
10. Aimee Rinas and Lisa M Jones. Fast photochemical oxidation of proteins coupled to multidimensional protein identification technology (mudpit): Expanding footprinting strategies to complex systems. *Journal of The American Society for Mass Spectrometry*, 26(4):540–546, 2015. ISSN 1044-0305.
11. Fabrice de Chaumont, Stephane Dallongeville, Nicolas Chenouard, Nicolas Herve, Sorin Pop, Thomas Provoost, Vannary Meas-Yedid, Praveen Pankajakshan, Timothee Lecomte, Yoann Le Montagner, Thibault Lagache, Alexandre Dufour, and Jean-Christophe Olivo-Marin. Icy: An open bioimage informatics platform for extended reproducible research. *Nat Meth*, 9(7):690–696, 2012. ISSN 1548-7091. doi: 10.1038/nmeth.2075.

12. F. de Chaumont, S. Dallongeville, and J. C. Olivo-Marin. Icy: A new open-source community image processing software. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 234–237. ISBN 1945-7928. doi: 10.1109/ISBI.2011.5872395.

13. Hye Yoon Park, Xiangyun Qiu, Elizabeth Rhoades, Jonas Korlach, Lisa W. Kwok, Warren R. Zipfel, Watt W. Webb, and Lois Pollack. Achieving uniform mixing in a microfluidic device:□ hydrodynamic focusing prior to mixing. *Analytical Chemistry*, 78(13):4465–4473, 2006. ISSN 0003-2700. doi: 10.1021/ac060572n.

14. B. Ehrenberg, V. Montana, M. D. Wei, J. P. Wuskell, and L. M. Loew. Membrane potential can be determined in individual cells from the nernstian distribution of cationic dyes. *Biophysical Journal*, 53(5):785–794, 1988. ISSN 0006-3495. doi: 10.1016/S0006-3495(88)83158-8.

15. D. L. Farkas, M. D. Wei, P. Febbroriello, J. H. Carson, and L. M. Loew. Simultaneous imaging of cell and mitochondrial membrane potentials. *Biophysical Journal*, 56(6):1053–1069, 1989. ISSN 0006-3495. doi: 10.1016/S0006-3495(89)82754-7.

16. H. A. Crissman, P. F. Mullaney, and J. A. Steinkamp. *Chapter 12 Methods and Applications of Flow Systems for Analysis and Sorting of Mammalian Cells1*, volume Volume 9, pages 179–246. Academic Press, 1975. ISBN 0091-679X. doi: 10.1016/S0091-679X(08)60076-X.

17. Peter E. Vandeventer, Jorge Mejia, Ali Nadim, Malkiat S. Johal, and Angelika Niemz. Dna adsorption to and elution from silica surfaces: Influence of amino acid buffers. *The Journal of Physical Chemistry B*, 117(37):10742–10749, 2013. ISSN 1520-6106. doi: 10.1021/jp405753m.

18. Jacek Switala and Peter C. Loewen. Diversity of properties among catalases. *Archives of Biochemistry and Biophysics*, 401(2):145–154, 2002. ISSN 0003-9861. doi: 10.1016/S0003-9861(02)00049-8.

19. Steven Maere, Karel Heymans, and Martin Kuiper. Bingo: A cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21(16):3448–3449, 2005. doi: 10.1093/bioinformatics/bti551.

20. Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003. doi: 10.1101/gr.1239303.

21. GD Rose, AR Geselowitz, GJ Lesser, RH Lee, and MH Zehfus. Hydrophobicity of amino acid residues in globular proteins. *Science*, 229(4716):834–838, 1985. doi: 10.1126/science.4023714.

22. Joel Janin. Surface and inside volumes in globular proteins. *Nature*, 277(5696): 491–492, 1979.

23. Guozhong Xu and Mark R. Chance. Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. *Analytical Chemistry*, 77(14):4549–4555, 2005. ISSN 0003-2700. doi: 10.1021/ac050299+.

24. Leigh Willard, Anuj Ranjan, Haiyan Zhang, Hassan Monzavi, Robert F. Boyko, Brian D. Sykes, and David S. Wishart. Vadar: A web server for quantitative evaluation of protein structure quality. *Nucleic Acids Research*, 31(13):3316–3319, 2003. ISSN 0305-1048 1362-4962.

4.6 Appendix: Flow Assembly for Cells Patent Application

IURTC-2015-123-02-WO-e

FLOW ASSEMBLY FOR CELLS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Patent Application Serial No. 62/144,542, filed on April 8, 2015, and entitled “FLOW ASSEMBLY FOR CELLS,” the complete disclosure of which is expressly incorporated by reference herein.

FIELD OF THE INVENTION

[0002] The present invention relates generally to an assembly configured for the flow of cells, and more particularly, to a micro-flow assembly configured for a single-file flow of cells.

BACKGROUND OF THE DISCLOSURE

[0003] In cell fast photochemical oxidation of proteins (“IC-FPOP”) is a tool for characterizing protein structure within a cell. Through IC-FPOP, hydrogen peroxide is photolyzed using an excimer laser to form hydroxyl radicals. The hydroxyl radicals covalently label side chains of amino acids exposed to solvent, thereby allowing for oxidative characterization of the protein structure within a cell.

[0004] However, current flow systems used for *in vitro* analysis of the protein structure within a cell may not be effective with IC-FPOP. More particularly, current flow systems may lead to cell aggregation which may clog the flow system and/or lead to inconsistent labeling and characterization of the cells because not every cell is exposed to the laser equally.

SUMMARY OF THE DISCLOSURE

[0005] In one embodiment, a flow assembly for cells comprises a first flow path configured to receive a plurality of cells and having an inner diameter of 50 – 100 µm, a second flow path configured to receive a buffer and having an inner diameter of 100-200 µm, and a third flow path configured to receive the plurality of cells and the buffer. The third flow path has an inner diameter greater than the inner diameters of the first and second flow paths. The plurality of cells are in a single-file orientation and the buffer generally surrounds the single-file orientation of the plurality of cells when in the third flow path.

[0006] In another embodiment, a flow assembly for cells comprises a first flow path configured to receive a plurality of cells, a second flow path configured to receive a buffer, and a third flow path configured to receive the plurality of cells and the buffer. The third flow path has an inner diameter of 0.3 – 2.0 mm. The plurality of cells are configured to flow in a single-file orientation and the buffer generally surrounds the single-file orientation of the plurality of cells when in the third flow path. At least one of the plurality of cells and the buffer has a flow rate of 25-40 μL/min.

[0007] In a further embodiment, a method of characterizing protein structure within a cell comprises providing a first flow path, flowing a plurality of cells through the first flow path, providing a second flow path spaced apart from the first flow path, and flowing a buffer through the second flow path. The method further comprises providing a third flow path with an inner diameter of 0.3 – 2.0 mm, surrounding the plurality of cells with the buffer, and flowing the plurality of cells in a single-file orientation through the third flow path. The method also comprises emitting a light source through at least a portion of the third flow path, passing the single-file orientation of the plurality of cells through the light source, and identifying a protein structure within each of the plurality of cells.

[0008] Additional features and advantages of the present invention will become apparent to those skilled in the art upon consideration of the following detailed description of the illustrative embodiment exemplifying the best mode of carrying out the invention as presently perceived.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The foregoing aspects and many of the intended advantages of this invention will become more readily appreciated as the same becomes better understood by reference to the following detailed description when taken in conjunction with the accompanying drawings.

[0010] Figure 1 is a schematic view of a flow system of the present disclosure;

[0011] Figure 2 is a schematic view of a portion of the flow system of Figure 1;

[0012] Figure 3 is a schematic view of an alternative embodiment flow system of the present disclosure;

[0013] Figure 4 is a schematic view of a portion of the alternative flow system of Figure 3;

[0014] Figure 5 is a schematic view of an alternative embodiment flow system of the present disclosure; and

[0015] Figure 6 is a schematic view of a further alternative embodiment flow system of the present disclosure.

[0016] Corresponding reference characters indicate corresponding parts throughout the several views. Although the drawings represent embodiments of various features and components according to the present disclosure, the drawings are not necessarily to scale and certain features may be exaggerated in order to better illustrate and explain the present disclosure. The exemplifications set out herein illustrate embodiments of the invention, and such exemplifications are not to be construed as limiting the scope of the invention in any manner.

DETAILED DESCRIPTION OF THE DRAWINGS

[0017] For the purposes of promoting an understanding of the principals of the invention, reference will now be made to the embodiments illustrated in the drawings, which are described below. The embodiments disclosed below are not intended to be exhaustive or limit the invention to the precise form disclosed in the following detailed description. Rather, the embodiments are chosen and described so that others skilled in the art may utilize their teachings. It will be understood that no limitation of the scope of the invention is thereby intended. The invention includes any alterations and further modifications in the illustrative devices and described methods and further applications of the principles of the invention which would normally occur to one skilled in the art to which the invention relates.

[0018] Referring to Figure 1, a flow assembly 10 for a plurality of cells 12 (Figure 2) is provided. Flow assembly 10 includes a first flow path or conduit 14, a second flow path 16 or conduit, a joining member 18, and a third flow path or conduit 20. Flow assembly 10 is configured to move cells 12 through third flow path 20 in a single-file orientation, as shown in Figure 2 and disclosed further herein.

[0019] In the illustrative embodiment of flow assembly 10, first flow path 14 is configured to receive cells 12 and provide a pathway to move cells 12 into joining member 18 along a flow direction A. Cells 12 may be provided within a solution or other material to facilitate flow through flow assembly 10. In one embodiment, cells 12 are provided within a phosphate buffer saline (“PBS”) solution and are configured to flow through flow assembly 10 at a flow rate of 25-40 µL/min. More particularly, cells 12 within the PBS material are configured to flow through flow assembly 10 at a flow rate of 33 µL/min.

[0020] Cells 12 also may be mixed with an oxygen-based compound, mixture, or solution, such as hydrogen peroxide (H_2O_2) for characterizing the proteins within cells. More particularly, cells 12 may be mixed with hydrogen peroxide prior to introducing cells 12 into flow assembly 10 such that the combination of cells 12 and hydrogen peroxide flow together through flow assembly 10. As disclosed further herein, the hydrogen peroxide is photolyzed by a laser or light source to form hydroxyl radicals which covalently label side chains of the amino acids of cells 12 through an in cell fast photochemical oxidation of proteins (“IC-FPOP”) process. In this way, the proteins within cells 12 are oxidatively characterized by the hydroxyl radicals when flowing through a portion of flow assembly 10.

[0021] First flow path 14 is comprised of any material configured to allow cells 12 to flow therethrough without adhering to the surface of first flow path 14. In one embodiment, first flow path 14 is comprised of fused silica capillary tubing coated with polyimide and/or quartz. Additionally, the length of first flow path 14 may vary to accommodate various configurations of flow assembly 10. An inner diameter of first flow path 14 may be 50 µm, 55 µm, 60 µm, 65 µm, 70 µm, 75 µm, 80 µm, 85 µm, 90 µm, 95 µm, 100 µm, or within any range delimited by any pair of the foregoing values. An outer diameter of first flow path 14 may be 300 µm, 310 µm, 320 µm, 330 µm, 340 µm, 350 µm, 360 µm, 370 µm, 380 µm, 390 µm, 400 µm, or within any range delimited by any pair of the foregoing values. Illustrative first flow path 14 has an inner diameter of 75 µm and an outer diameter of 360 µm.

[0022] Referring still to Figure 1, second flow path 16 is configured to receive a buffer material 22 and flow buffer material 22 into joining member 18 along a flow direction B. Illustratively, flow direction B is generally perpendicular to flow direction A, however the second flow path 16 and flow direction B may be in any orientation relative to first flow path 14 and flow direction A (e.g., parallel, perpendicular, or at any angle). Buffer 22 may be a PBS solution or any solution configured to flow through flow assembly 10 without mixing with cells 12, as disclosed further herein. In one embodiment, the flow rate of buffer 22 through flow assembly 10 is 25 – 40 µL/min and, more particularly, is 33 µL/min. For example, in one embodiment, the flow rate of buffer 22 may be the same as the flow rate of cells 12.

[0023] Second flow path 16 is comprised of any material configured to flow buffer 22. In one embodiment, second flow path 16 is comprised of fused silica capillary tubing coated with polyimide and/or quartz. Additionally, the length of second flow path 16 may vary to accommodate various configurations of flow assembly 10. An inner diameter of second flow path 16 may be greater than the inner diameter of first flow path 14, for example 100 μm, 110 μm, 120 μm, 130 μm, 140 μm, 150 μm, 160 μm, 170 μm, 180 μm, 190 μm, 200 μm or within any range delimited by any pair of the foregoing values. An outer diameter of second flow path 16 may be 300 μm, 310 μm, 320 μm, 330 μm, 340 μm, 350 μm, 360 μm, 370 μm, 380 μm, 390 μm, 400 μm, or within any range delimited by any pair of the foregoing values. Illustrative second flow path 16 has an inner diameter of 150 μm and an outer diameter of 360 μm.

[0024] Referring to Figures 1 and 2, joining member 18 is a tee assembly comprised of a polymeric material, for example polyether ether ketone ("PEEK"). Illustrative joining member 18 includes a plurality of 1/16-inch tubes 25a, 25b, ¼-28 flat-bottom nuts 24a, 24b, 24c, each with a sleeve extending therethrough (not shown). Each flat-bottom nut 24a, 24b, 24c has a through-hole with a diameter of 0.5 mm extending longitudinally along the length of nuts 24a, 24b, 24c. Because the diameter of the through-hole of each nut 24a, 24b, 24c is greater than the diameter of first and second flow paths 14, 16, the sleeves are included to control the size of the flow path for buffer 22 and cells 12. In particular, the inner diameter of the sleeves are configured to receive the outer diameter of first and second flow paths 14, 16. A plurality of respective ferrules 26a, 26b, 26c are provided to secure the sleeves to first and second flow paths 14, 16. Ferrules 26a, 26b, 26c may be comprised of the same material as joining member 18 and nuts 24a, 24b, 24c, however, illustrative ferrules 26a, 26b, 26c are comprised of ethylene tetrafluoroethylene ("ETFE"). In one embodiment, the sleeves (not shown) are comprised of fluorinated ethylene propylene ("FEP"). Flat-bottom nuts 24a, 24b, 24c may be comprised of the same material as joining member 18 or may be comprised of a different material. Illustratively, flat-bottom nuts 24a, 24b, 24c are comprised of PEEK.

[0025] Additionally, first flow path 14 extends between nuts 24a and 24c and is coupled to the sleeves therein with ferrules 26a and 26c. Alternatively, first flow path 14 may terminate in the sleeve of nut 24a and an additional portion of tubing having the same dimensions and characteristics as first flow path 14 may be provided between nuts 24a and 24c.

[0026] Referring still to Figures 1 and 2, third flow path 20 is coupled to nut 24c of joining member 18. Third flow path 20 is comprised of any material configured to flow buffer 22 and cells 12. In one embodiment, third flow path 20 is comprised of fused silica capillary tubing coated with polyimide. Alternatively, third flow path 20 may be comprised of a quartz capillary. Additionally, the length of third flow path 20 may vary to accommodate various configurations of flow assembly 10. An inner diameter of third flow path 20 may be as little as 0.3 mm, 0.35 mm, 0.4 mm, 0.45 mm, 0.5 mm, 0.55 mm, 0.6 mm, 0.65 mm, 0.7 mm, 0.75 mm, or as great as 1.0 mm, 1.25 mm, 1.5 mm, 1.75 mm, or 2.0 mm, or within any range delimited by any pair of the foregoing values. An outer diameter of third flow path 20 may be as little as 0.5 mm, 0.55 mm, 0.6 mm, 0.65 mm, 0.7 mm, 0.75 mm, 0.8 mm, 0.85 mm, or as great as 1.0 mm, 1.25 mm, 1.5 mm, 1.75 mm, 2.0 mm, or within any range delimited by any pair of the foregoing values. Illustrative third flow path 20 has an inner diameter of 0.45 mm and an outer diameter of 0.76 mm.

[0027] As shown in Figures 1 and 2, third flow path 20 includes a coated portion 20a and uncoated portion 20b. Coated portion 20a includes the polyimide or quartz coating, however, the coating has been removed in uncoated portion 20b such that the bare capillary tubing is exposed. As disclosed further herein, by removing a portion of the coating in uncoated portion 20b, an irradiation window is defined so that a laser or other light source can pulse light through third flow path 20 as cells 12 pass therethrough, thereby oxidatively modifying the proteins within cells 12 to understand the structure of the proteins therein.

[0028] In operation, cells 12 may be in a PBS solution and are mixed wtih hydrogen peroxide when flowing within first flow path 14 in flow direction A toward joining member 18. Simultaneously, buffer 22 flows within second flow path 16 in flow direction B toward joining member 18. Once at joining member 18, cells 12 flow through nut 24a and buffer 22 flows through nut 24b. Cells 12 continue to flow through first flow path 14 while buffer 22 flows around the outer diameter of first flow path 14. In this way, buffer 22 generally surrounds cells 12 but remains separated from and does not mix with cells 12.

[0029] After flowing through nut 24c, buffer 22 and cells 12 simultaneously enter third flow path 20, where, despite exiting first flow path 14, cells still 12 do not mix with buffer 22. Instead, buffer 22 generally defines a chamber surrounding cells 12 such that cells 12 flow through the middle of buffer 22 in a single-file orientation. As shown in Figure 2, when surrounded by buffer 22, the single-file orientation of cells 12 allows each cell 12 to pass one at a time through a given portion of third flow path 20 such that cells 12 are longitudinally aligned but do not vertically stack on top of each other. Both buffer 22 and cells 12 have the same flow rate, for example 33 μL/min, when flowing through flow assembly 10. The combination of this flow rate and the diameter of third flow path 20 allows cells 12 to flow in the single-file orientation through third flow path 20.

[0030] Once in third flow path 20, buffer 22 and cells 12 remain separate from each other and do not mix, as shown in Figure 2. Additionally, Figure 2 shows that a laser 50 or other light source is applied to uncoated portion 20b of third flow path 20. Illustratively, laser 50 is pulsed at 248 nm and 18 Hz frequency. Because cells 12 flow through third flow path 20 in a single-file orientation, laser 50 is applied equally to each cell 12. Laser 50 oxidatively modifies the proteins within each cell 12 so that the protein structure of each cell 12 can be identified and understood. More particularly, laser 50 is applied at a wavelength and frequency sufficient to photolyze the hydrogen peroxide mixed with cells 12 to form hydroxyl radicals which covalently label side chains of amino acids within cells 12. In this way, the interaction between the hydrogen peroxide and laser 50 characterize and/or identify the proteins within cells 12 through an IC-FPOP process which allows for analysis of the proteins within cells 12.

[0031] After flowing through third flow path 20 and experiencing laser 50, cells 12 and buffer 22 are collected. Cells 12 are separated from buffer 22 and/or any remaining hydrogen peroxide or other compounds, elements, or solutions in a centrifuge and cells 12 subsequently undergo mass spectrometry to identify the peptides (comprised of the covalently-labeled amino acids) of the proteins within cells 12, which allows the protein structure of each cell 12 to be understood. In one embodiment, cells 12 are spliced or otherwise cut into smaller pieces before undergoing mass spectrometry.

[0032] Referring to Figures 3 and 4, an alternative embodiment of flow assembly 10 is shown as flow assembly 10’, wherein like components having the same structure and functionality as those of the embodiment of Figures 1 and 2 are identified with like reference numbers. Flow assembly 10’ includes first flow path 14, a second flow path or conduit 16’, joining member 18, and a third flow path or conduit 20’. Additionally, flow assembly 10’ includes a first cross-member 28, a second cross-member 30, a dead volume portion 32, and an outlet capillary or conduit 34.

[0033] Second flow path 16’ is comprised of any material configured to flow buffer 22. In one embodiment, second flow path 16’ is comprised of fused silica capillary tubing coated with polyimide and/or quartz. Additionally, the length of second flow path 16’ may vary to accommodate various configurations of flow assembly 10’. An inner diameter of second flow path 16’ may be 100 μm, 110 μm, 120 μm, 130 μm, 140 μm, 150 μm, 160 μm, 170 μm, 180 μm, 190 μm, or 200 μm or within any range delimited by any pair of the foregoing values. An outer diameter of second flow path 16 may be 300 μm, 310 μm, 320 μm, 330 μm, 340 μm, 350 μm, 360 μm, 370 μm, 380 μm, 390 μm, or 400 μm, or within any range delimited by any pair of the foregoing values. Illustrative second flow path 16’ has an inner diameter of 150 μm and an outer diameter of 360 μm.

[0034] Second flow path 16’ has a first portion 16a’ configured to flow a first portion of buffer 22 into first cross-member 28, a second portion 16b’ configured to flow a second portion of buffer 22 into second cross-member 30, a third portion 16c’ configured to flow a third portion of buffer 22 into second cross-member 30, and a fourth portion 16d’ configured to flow a fourth portion of buffer 22 into joining member 18.

[0035] As shown in Figure 3, first cross-member 28 includes four 1/16-inch tubing, ¼-28 flat bottom nuts 28a, 28b, 28c, 28d, each with a sleeve extending therethrough (not shown). Each flat-bottom nut 28a, 28b, 28c, 28d has a through-hole with a diameter of 0.5 mm extending longitudinally along the length of nuts 28a, 28b, 28c, 28d. Because the diameter of the through-hole of each nut 28a, 28b, 28c, 28d is greater than the diameter of second flow path 16’, the sleeves are included to control the size of the flow path for buffer 22. In particular, the inner diameter of the sleeves are configured to receive the outer diameter of second flow path 16’. A plurality of ferrules 36a, 36b, 36c, 36d are provided to secure the sleeves to second flow path 16’. Ferrules 36a, 36b, 36c, 36d may be comprised of ETFE. Flat-bottom nuts 28a, 28b, 28c, 28d may be comprised PEEK.

[0036] Referring still to Figure 3, second cross-member 30 includes four 1/16-inch tubing, ¼-28 flat bottom nuts 30a, 30b, 30c, 30d, each with a sleeve extending therethrough (not shown). Each flat-bottom nut 30a, 30b, 30c, 30d has a through-hole with a diameter of 0.5 mm extending longitudinally along the length of nuts 30a, 30b, 30c, 30d. Because the diameter of the through-hole of each nut 30a, 30b, 30c, 30d is greater than the diameter of first and second flow paths 14, 16’, the sleeves are included to control the size of the flow path for cells 12 and buffer

22. In particular, the inner diameter of the sleeves are configured to receive the outer diameter of first and second flow paths 14, 16'. A plurality of ferrules 38a, 38b, 38c, 38d are provided to secure the sleeves to first and second flow paths 14, 16'. Ferrules 38a, 38b, 38c, 38d may be comprised of ETFE. Flat-bottom nuts 30a, 30b, 30c, 30d may be comprised PEEK.

[0037] As shown in Figure 3, nuts 28b and 30b are fluidly coupled together through second portion 16b' of second flow path 16' such that buffer 22 from first cross-member 28 enters second cross-member 30 through second portion 16b'. Similarly, nuts 28c and 30c are fluidly coupled together through third portion 16c' of second flow path 16' such that buffer 22 from first cross-member 28 enters second cross-member 30 through third portion 16c'. Nut 24c of joining member 18 is fluidly coupled to nut 30a of second cross-member 30 through first flow path 14 and nut 30d of second cross-member 30 is fluidly coupled to third flow path 20'.

[0038] Third flow path 20' is comprised of any material configured to flow buffer 22 and cells 12. In one embodiment, third flow path 20' is comprised of a quartz capillary. Additionally, the length of third flow path 20' may vary to accommodate various configurations of flow assembly 10'. An inner diameter of third flow path 20' may be 1.0 mm, 1.1 mm, 1.2 mm, 1.3 mm, 1.4 mm, 1.5 mm, 1.6 mm, 1.7 mm, 1.8 mm, 1.9 mm, 2.0 mm, or within any range delimited by any pair of the foregoing values. An outer diameter of third flow path 20' may be as little as 1.5 mm, 1.6 mm, 1.7 mm, 1.8 mm, 1.9 mm, 2.0 mm, 2.1 mm, 2.2 mm, 2.3 mm, 2.4 mm, 2.5 mm, or within any range delimited by any pair of the foregoing values. Illustrative third flow path 20' has an inner diameter of 1.5 mm and an outer diameter of 1.8 mm.

[0039] Dead volume portion 32 is coupled to third flow path 20' and includes two 1/16-inch tubing, ¼-28 flat bottom nuts 40a, 40b. Outlet capillary 34 is configured to be received through nuts 40a, 40b and secured with ferrules 42a, 42b, respectively. As such, outlet capillary 34 is fluidly coupled to third flow path 20' through dead volume portion 32. Outlet capillary 34 is comprised of PEEK and, in one embodiment, has an inner diameter of 0.3 mm and an outer diameter of 1.6 mm. Because the diameter of third flow path 20' is greater than the diameter of outlet capillary 34, dead volume portion 32 is provided to reduce air bubbles in buffer 22 and/or the solution of cells 12 to ensure that there is no delay or break in the flow of cells 12 and/or buffer 22 when going from the larger diameter of third flow path 20' to the smaller diameter of outlet capillary 34.

[0040] In operation, cells 12 may be in a PBS solution and mixed with hydrogen peroxide when flowing within first flow path 14 toward joining member 18. Once at joining member 18, cells 12 flow through nuts 24a, 24c and into nut 30a of second cross-member 30. Simultaneously, a portion of buffer 22 flows through nut 28d of first cross-member 28, into nut 24b of joining member 18, around first flow path 14, through nut 24c of joining member 18, and into nut 30a of second cross-member 30. Additionally, a portion of buffer 22 simultaneously flows through nut 28b of first cross-member 28, through second portion 16b' of second flow path 16', and into nut 30b of second cross-member 30. In this way, the portion of buffer 22 within second portion 16b' of second flow path 16' defines the lower portion of the buffer chamber or sheath that generally surrounds cells 12 within third flow path 20'. Additionally, another portion of buffer 22 simultaneously flows through nut 28c of first cross-member 28, through third portion 16c' of second flow path 16', and into nut 30c of second cross-member 30. In this way, the portion of buffer 22 within third portion 16c' of second flow path 16' defines the upper portion of the buffer chamber or sheath that generally surrounds cells 12 within third flow path 20'. The amount of buffer 22 flowing through second and third portions 16b', 16c' of second flow path 16' may be greater than the amount of buffer 22 flowing through fourth portion 16d' and into joining member 18.

[0041] Within joining member 18 and second cross-member 30, cells 12 (mixed with hydrogen peroxide) continue to flow through first flow path 14 while buffer 22 flows around the outer diameter of first flow path 14. In this way, buffer 22 generally surrounds cells 12 but does not mix with cells 12. After flowing through nut 30d of second cross-member, buffer 22 and cells 12 simultaneously enter third flow path 20, where, despite exiting first flow path 14, cells 12 do not mix with buffer 22. Instead, buffer 22 generally defines a chamber surrounding cells 12 such that cells 12 flow through the middle of buffer 22 in a single-file orientation. As shown in Figure 4, when surrounded by buffer 22, the single-file orientation of cells 12 allows each cell 12 to pass one at a time through a given portion of third flow path 20' such that cells 12 are longitudinally aligned but do not vertically stack on top of each other. Both buffer 22 and cells 12 have the same flow rate, for example 33 μL/min, when flowing through flow assembly 10'. The combination of the flow rate and the diameter of third flow path 20' allows for the single file orientation of cells 12 through third flow path 20'.

[0042] Once in third flow path 20’, buffer 22 and cells 12 (mixed with hydrogen peroxide) remain separate from each other and do not mix, as shown in Figure 4. Additionally, Figure 4 shows that laser 50 is applied to a portion of third flow path 20’. Illustratively, laser 50 is pulsed at 248 nm and 18 Hz frequency. Because cells 12 flow through third flow path 20’ in a single-file orientation, laser 50 is applied equally to each cell 12.

[0043] Laser 50 oxidatively modifies the proteins within each cell 12 through the IC-FPOP process so that the protein structure of each cell 12 can be identified and understood, as disclosed herein. After flowing through third flow path 20’ and experiencing laser 50, cells 12 and buffer 22 flow through dead volume portion 32, through outlet capillary 34, and are collected. Cells 12 are separated from buffer 22 and any remaining hydrogen peroxide or other compounds, solution, or elements in a centrifuge and cells 12 subsequently undergo mass spectrometry to identify the peptides of the proteins within cells 12, which allows the protein structure of each cell 12 to be understood. In one embodiment, cells 12 are spliced or otherwise cut into smaller pieces before undergoing mass spectrometry.

[0044] Referring to Figure 5, an alternative embodiment of flow assembly 10’ (Figure 3) is shown as flow assembly 110, wherein like components having the same structure and functionality as those of the embodiment of Figures 3 and 4 are identified with like reference numbers. Flow assembly 110 includes first flow path 14 (which has the same structure and functionality as flow path 14 of Figure 3), a second flow path 16 (which has the same structure and functionality as flow path 16 of Figure 3), a third flow path 112, joining member 18 (which has the same structure and functionality as joining member 18 of Figure 3), a fourth flow path 114 (which has the same structure and functionality as flow path 14 of Figure 3), a fifth flow path 20 (which has the same structure and functionality as flow path 20 of Figure 3), and cross-member 28 (which has the same structure and functionality as cross-member 28 of Figure 3).

[0045] As shown in Figure 5, joining member 18 is fluidly coupled to first flow path 14 to provide cells 12 to joining member 18 through nut 24a. Additionally, joining member 18 is fluidly coupled to third flow path 112 to provide an oxygen-based solution or compound, such as hydrogen peroxide, to joining member 18 through nut 24c. In this way, cells 12 are separated from hydrogen peroxide until cells 12 and hydrogen peroxide flow to joining member 18. Because hydrogen peroxide may be toxic or damaging to cells 12, flow assembly 110 allows for separation of cells 12 and hydrogen peroxide until both cells 12 and hydrogen peroxide flow to joining member 18, which decreases the length of time that cells 12 are in contact with hydrogen peroxide, thereby decreasing the likelihood that cells 12 may be damaged by hydrogen peroxide. In one embodiment, flow assembly 110 is provided for IC-FPOP processes.

[0046] Once cells 12 and hydrogen peroxide are combined at joining member 18, the combination of cells 12 and hydrogen peroxide flow together through fourth flow path 114 toward cross-member 28. Once at cross-member 28, cells 12 (mixed with hydrogen peroxide) flow into cross-member through nut 28a.

[0047] Buffer 22 also flows into cross-member 28 through at least one of the second flow paths 16 which are fluidly coupled to nuts 28b, 28d. At cross-member 28, buffer 22 remains separated from cells 12 such that buffer 22 flows in the same direction as cells 12 but flows along the outer circumference of the flow path or conduit of cells 12. In this way, cells 12 do not mix with buffer 22 when at cross-member 28.

[0048] In operation, cells 12 may be in a PBS solution (separate from hydrogen peroxide) and flow within first flow path 14 toward joining member 18 in direction A. Simultaneously, hydrogen peroxide flows within third flow path 112 toward joining member 18 in direction C. Once at joining member 18, cells 12 flow through nut 24a and hydrogen peroxide flows through nut 24c such that cells 12 mix with hydrogen peroxide once cells 12 and hydrogen peroxide are both within joining member 18. The mixture of hydrogen peroxide and cells 12 flows in direction D through fourth flow path 114 toward cross-member 28.

[0049] Simultaneously, at least a portion of buffer 22 flows through second flow path 16 and nut 28b of cross-member 28 and, in some embodiments, at least another portion of buffer 22 simultaneously flows through second flow path 16 and nut 28d of cross-member 28. In this way, the portion of buffer 22 entering nut 28b of cross-member 28 defines the upper portion of the buffer chamber or sheath that generally surrounds cells 12 within fifth flow path 20 and the portion of buffer 22 entering nut 28d of cross-member 28 defines the lower portion of the buffer chamber or sheath that generally surrounds cells 12 within fifth flow path 20. In this way, buffer 22 generally surrounds cells 12 but does not mix with cells 12.

[0050] Buffer 22 and cells 12 simultaneously enter fifth flow path 20, where, despite exiting fourth flow path 114, cells 12 do not mix with buffer 22. Instead, buffer 22 generally defines a chamber surrounding cells 12 such that cells 12 flow through the middle of buffer 22 in a single-file orientation. As shown in Figure 5, when surrounded by buffer 22, the single-file orientation of cells 12 allows each cell 12 to pass one at a time through a given portion of fifth flow path 20 such that cells 12 are longitudinally aligned but do not vertically stack on top of each other. Both buffer 22 and cells 12 have the same flow rate, for example 33 μL/min, when flowing through flow assembly 110. The combination of the flow rate and the diameter of fifth flow path 20 allows for the single file orientation of cells 12 through fifth flow path 20.

[0051] Once in fifth flow path 20, buffer 22 and cells 12 (mixed with hydrogen peroxide) remain separate from each other and do not mix, as shown in Figure 5. Additionally, Figure 5 shows that laser 50 is applied to uncoated portion 20b of fifth flow path 20. Illustratively, laser 50 is pulsed at 248 nm and 18 Hz frequency. Because cells 12 flow through fifth flow path 20 in a single-file orientation, laser 50 is applied equally to each cell 12.

[0052] Laser 50 oxidatively modifies the proteins within each cell 12 so that the protein structure of each cell 12 can be identified and understood, as disclosed herein. After flowing through fifth flow path 20 and experiencing laser 50, cells 12 are separated from buffer 22 and any remaining hydrogen peroxide or other compounds, solution, or elements in a centrifuge and cells 12 subsequently undergo mass spectrometry to identify the peptides of the proteins within cells 12, which allows the protein structure of each cell 12 to be understood. In one embodiment, cells 12 are spliced or otherwise cut into smaller pieces before undergoing mass spectrometry.

[0053] Referring to Figure 6, an alternative embodiment of flow assembly 110 is shown as flow assembly 210, wherein like components having the same structure and functionality as those of the embodiment of Figure 5 are identified with like reference numbers. Flow assembly 210 includes first flow path 14 (which has the same structure and functionality as flow path 14 of Figure 5), a second flow path 16 (which has the same structure and functionality as flow path 16 of Figure 5), a third flow path 20 (which has the same structure and functionality as flow path 20 of Figure 5), and cross-member 28 (which has the same structure and functionality as cross-member 28 of Figure 5).

[0054] As shown in Figure 6, joining member 18 of Figure 5 is removed such that cells 12 are combined with hydrogen peroxide prior to entering first flow path 14. As such, cells 12 flow together with hydrogen peroxide through first flow path 14 toward cross-member 28. Cells 12 and hydrogen peroxide flow into cross-member 28 through nut 28a.

[0055] Buffer 22 also flows into cross-member 28 through at least one of the second flow paths 16 (in direction B) which are fluidly coupled to nuts 28b, 28d. At cross-member 28, buffer

22 remains separated from cells 12 such that buffer 22 flows in the same direction as cells 12 but flows along the outer circumference of the flow path or conduit provided for cells 12. In this way, cells 12 do not mix with buffer 22 when at cross-member 28.

[0056] In operation, cells 12 (mixed with hydrogen peroxide) may be in a PBS solution and flow within first flow path 14 in direction A toward cross-member 28. Simultaneously, at least a portion of buffer 22 flows through nut 28b of cross-member 28 and, in some embodiments, at least another portion of buffer 22 simultaneously flows through nut 28d of cross-member 28. In this way, the portion of buffer 22 entering nut 28b of cross-member 28 defines the upper portion of the buffer chamber or sheath that generally surrounds cells 12 within third flow path 20 and the portion of buffer 22 entering nut 28d of cross-member 28 defines the lower portion of the buffer chamber or sheath that generally surrounds cells 12 within third flow path 20. In this way, buffer 22 generally surrounds cells 12 but does not mix with cells 12.

[0057] Buffer 22 and cells 12 simultaneously enter third flow path 20, where, despite exiting first flow path 14, cells 12 do not mix with buffer 22. Instead, buffer 22 generally defines a chamber surrounding cells 12 such that cells 12 flow through the middle of buffer 22 in a single-file orientation. As shown in Figure 6, when surrounded by buffer 22, the single-file orientation of cells 12 allows each cell 12 to pass one at a time through a given portion of third flow path 20 such that cells 12 are longitudinally aligned but do not vertically stack on top of each other. Both buffer 22 and cells 12 have the same flow rate, for example 33 μL/min, when flowing through flow assembly 110. The combination of the flow rate and the diameter of third flow path 20 allows for the single file orientation of cells 12 through third flow path 20.

[0058] Once in third flow path 20, buffer 22 and cells 12 (mixed with hydrogen peroxide) remain separate from each other and do not mix, as shown in Figure 6. Additionally, Figure 6 shows that laser 50 is applied to uncoated portion 20b of third flow path 20. Illustratively, laser 50 is pulsed at 248 nm and 18 Hz frequency. Because cells 12 flow through third flow path 20 in a single-file orientation, laser 50 is applied equally to each cell 12.

[0059] Laser 50 oxidatively modifies the proteins within each cell 12 so that the protein structure of each cell 12 can be identified and understood, as disclosed herein. After flowing through third flow path 20 and experiencing laser 50, cells 12 are separated from buffer 22 and any remaining hydrogen peroxide or other compounds, solution, or elements in a centrifuge and cells 12 subsequently undergo mass spectrometry to identify the peptides of the proteins within cells 12, which allows the protein structure of each cell 12 to be understood. In one embodiment, cells 12 are spliced or otherwise cut into smaller pieces before undergoing mass spectrometry.

[0060] While this invention has been described as having an exemplary design, the present invention may be further modified within the spirit and scope of this disclosure. This application is therefore intended to cover any variations, uses, or adaptations of the invention using its general principles. Further, this application is intended to cover such departures from the present disclosure as come within known or customary practices in the art to which this invention pertains.

WHAT IS CLAIMED IS:

1. A flow assembly for cells, comprising:

a first flow path configured to receive a plurality of cells and having an inner diameter of 50 – 100 µm;

a second flow path configured to receive a buffer and having an inner diameter of 100-200 µm; and

a third flow path configured to receive the plurality of cells and the buffer and having an inner diameter greater than the inner diameters of the first and second flow paths, the plurality of cells being in a single-file orientation and the buffer generally surrounding the single-file orientation of the plurality of cells when in the third flow path.

2. The flow assembly of claim 1, wherein the inner diameter of the first flow path is 75 µm.

3. The flow assembly of claim 1, wherein the inner diameter of the second flow path is 150 µm.

4. The flow assembly of claim 1, wherein the inner diameter of the third flow path is 0.3 mm – 2.0 mm.

5. The flow assembly of claim 1, wherein the buffer is phosphate buffered saline.

6. A flow assembly for cells, comprising:

a first flow path configured to receive a plurality of cells;

a second flow path configured to receive a buffer; and

a third flow path configured to receive the plurality of cells and the buffer and having an inner diameter of 0.3 – 2.0 mm, the plurality of cells being configured to flow in a single-file orientation and the buffer generally surrounding the single-file orientation of the plurality of cells when in the third flow path, and at least one of the plurality of cells and the buffer having a flow rate of 25-40 µL/min.

7. The flow assembly of claim 6, wherein the plurality of cells has a flow rate of 33 µL/min.

8. The flow assembly of claim 6, wherein the buffer has a flow rate of 33 µL/min.

9. The flow assembly of claim 6, wherein the inner diameter of the third flow path is 0.45-1.5 mm.

10. The flow of assembly of claim 6, wherein an inner diameter of the first flow path is less than an inner diameter of the second flow path, and the inner diameters of the first and second flow paths are less than the inner diameter of the third flow path.

11. The flow assembly of claim 6, wherein the buffer is phosphate buffered saline.

12. A method of characterizing protein structure within a cell, comprising:

providing a first flow path;

flowing a plurality of cells through the first flow path;

providing a second flow path spaced apart from the first flow path;

flowing a buffer through the second flow path;

providing a third flow path with an inner diameter of 0.3 – 2.0 mm;

surrounding the plurality of cells with the buffer;

flowing the plurality of cells in a single-file orientation through the third flow path;

emitting a light source through at least a portion of the third flow path;

passing the single-file orientation of the plurality of cells through the light source; and

identifying a protein structure within each of the plurality of cells.

13. The method of claim 12, wherein the light source has a wavelength of 248 nm.

14. The method of claim 12, wherein the first flow path has an inner diameter of 50 – 100 μm.

15. The method of claim 12, wherein the second flow path has an inner diameter of 100 – 200 μm.

16. The method of claim 12, wherein the inner diameter of the third flow path is 0.45 – 1.5 mm.

17. The method of claim 12, wherein flowing the plurality of cells in a single-file orientation including flowing the plurality of cells at a flow rate of 25-40 μL/min.

18. The method of claim 17, wherein the flow rate is 33 μL/min.

19. The method of claim 12, wherein identifying the protein structure within each of the plurality of cells includes performing mass spectrometry on the plurality of cells.

20. The method of claim 12, further comprising providing a fourth flow path, flowing an oxygen-based compound through the fourth flow path, and mixing the oxygen-based compound with the plurality of cells prior to passing the single-file orientation of the plurality of cells through the light source.

ABSTRACT

In one embodiment, a flow assembly for cells comprises a first flow path configured to receive a plurality of cells, a second flow path configured to receive a buffer, and a third flow path configured to receive the plurality of cells and the buffer. The plurality of cells are in a single-file orientation and the buffer generally surrounds the single-file orientation of the plurality of cells when in the third flow path.

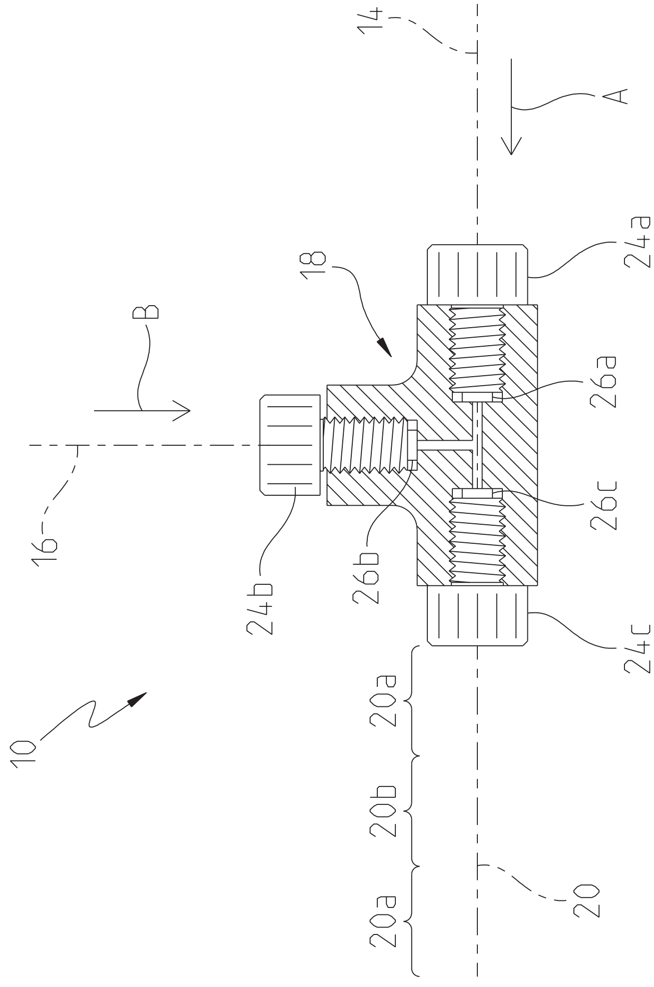

10
14
A
16
B
18
24a
24b
24c
26a
26b
26c
20a
20b
20a
20

Figure 1

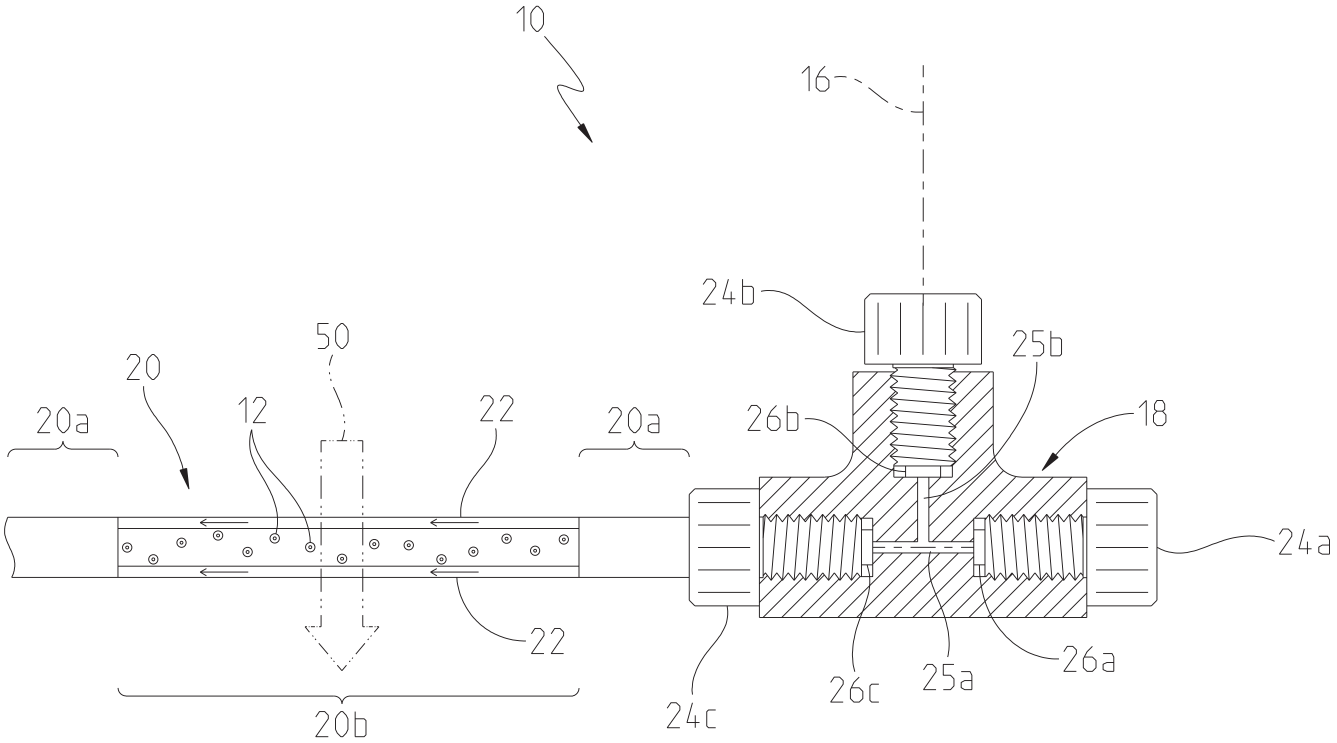

10
16
24b
25b
18
50
20
12
22
20a
20a
26b
24a
22
20b
24c
26c
25a
26a

Figure 2

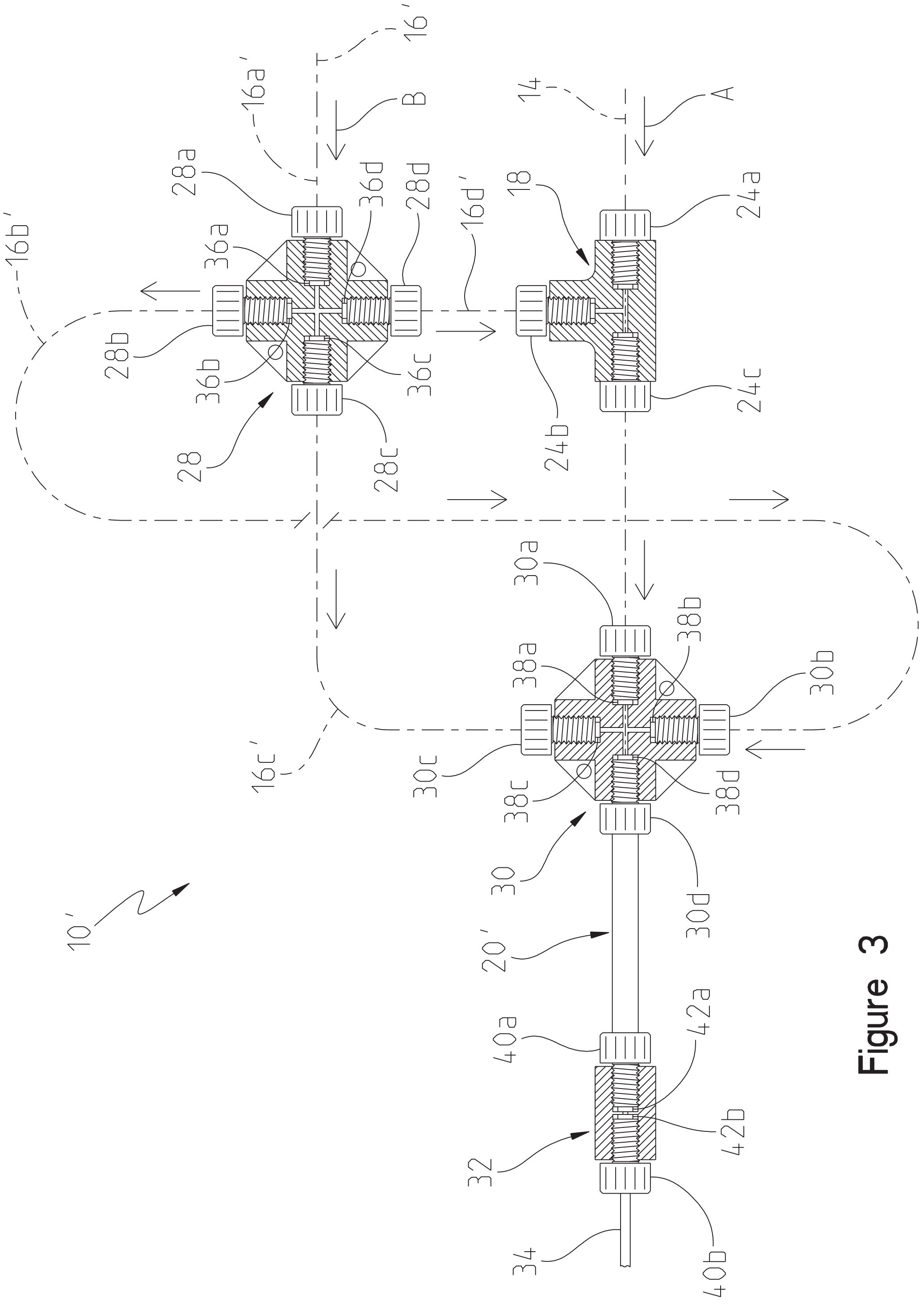

B
A

Figure 3

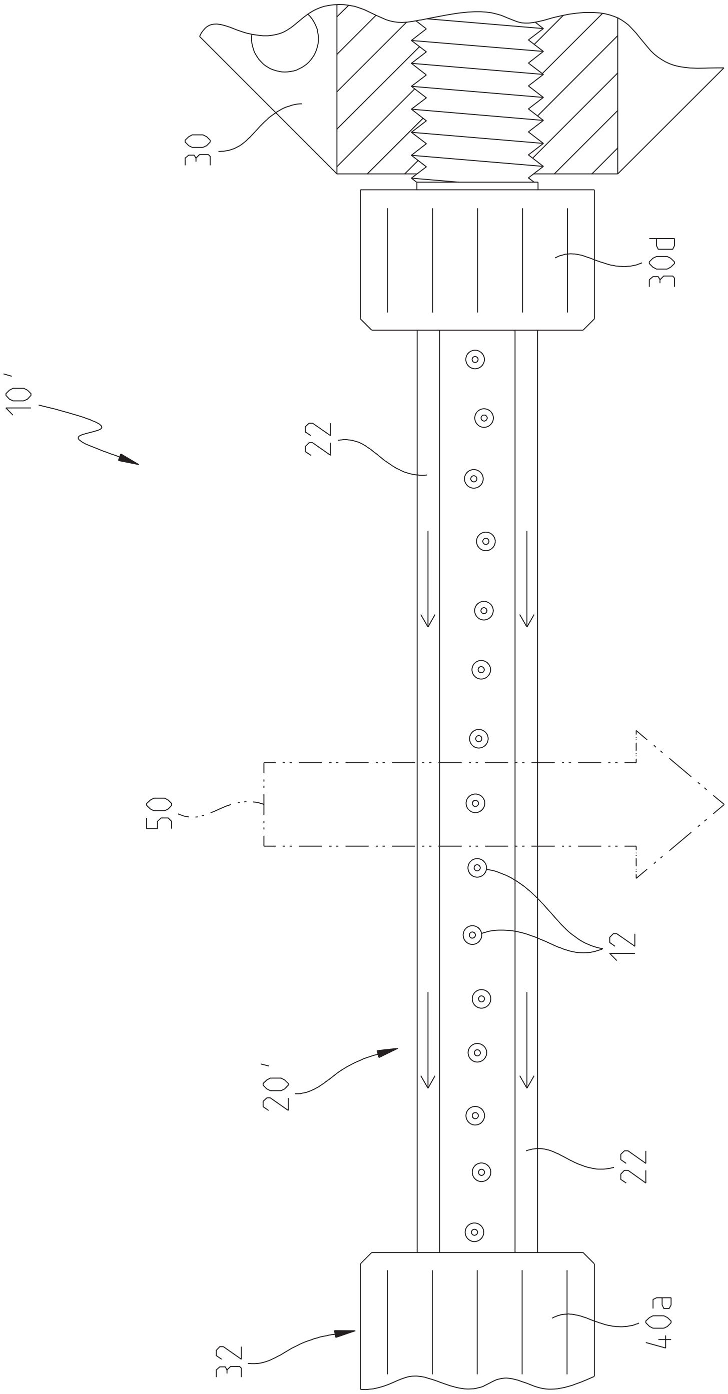

30
30d
10'
22
50
12
20'
22
32
40a

Figure 4

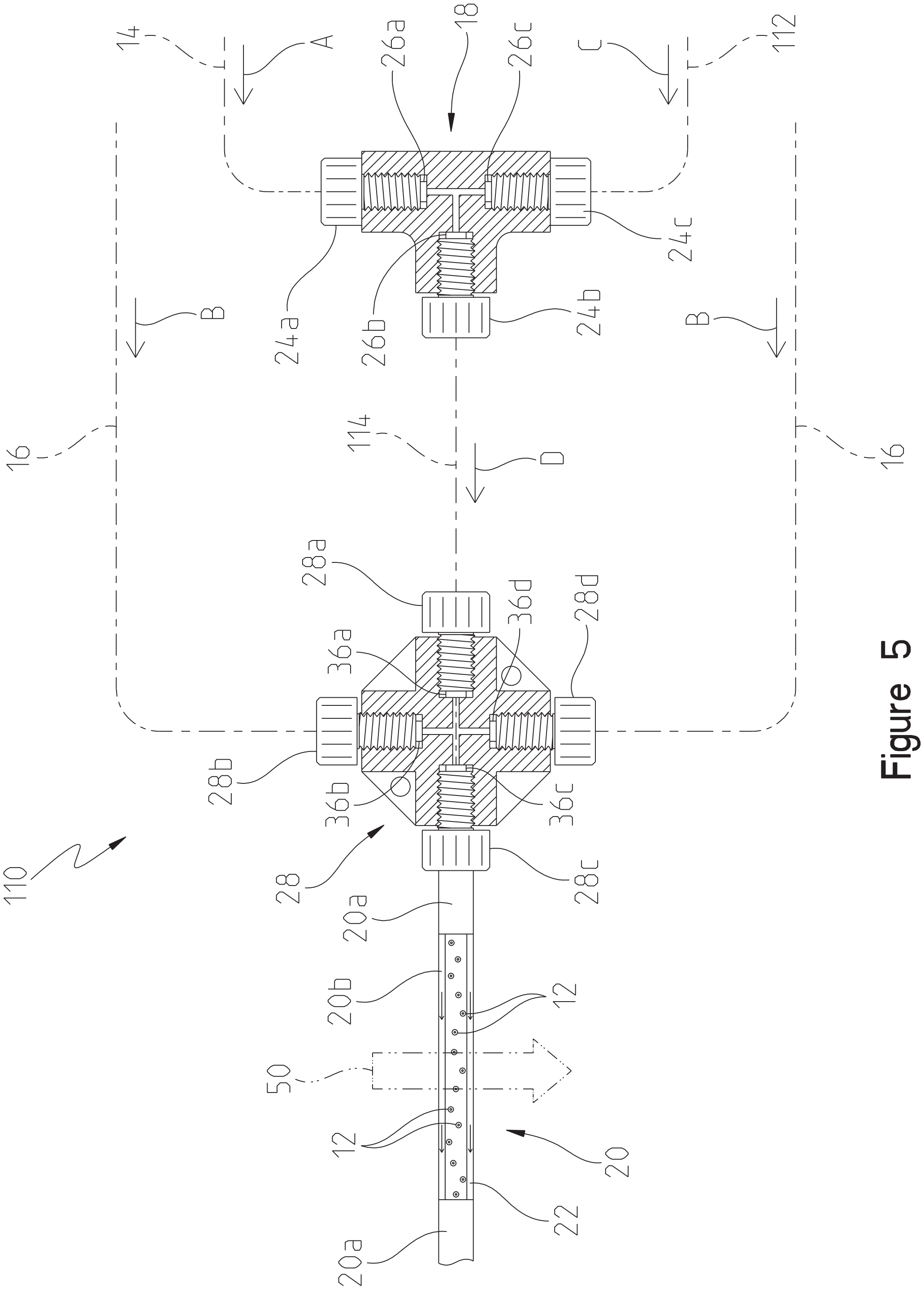

A
C
B
B
D

Figure 5

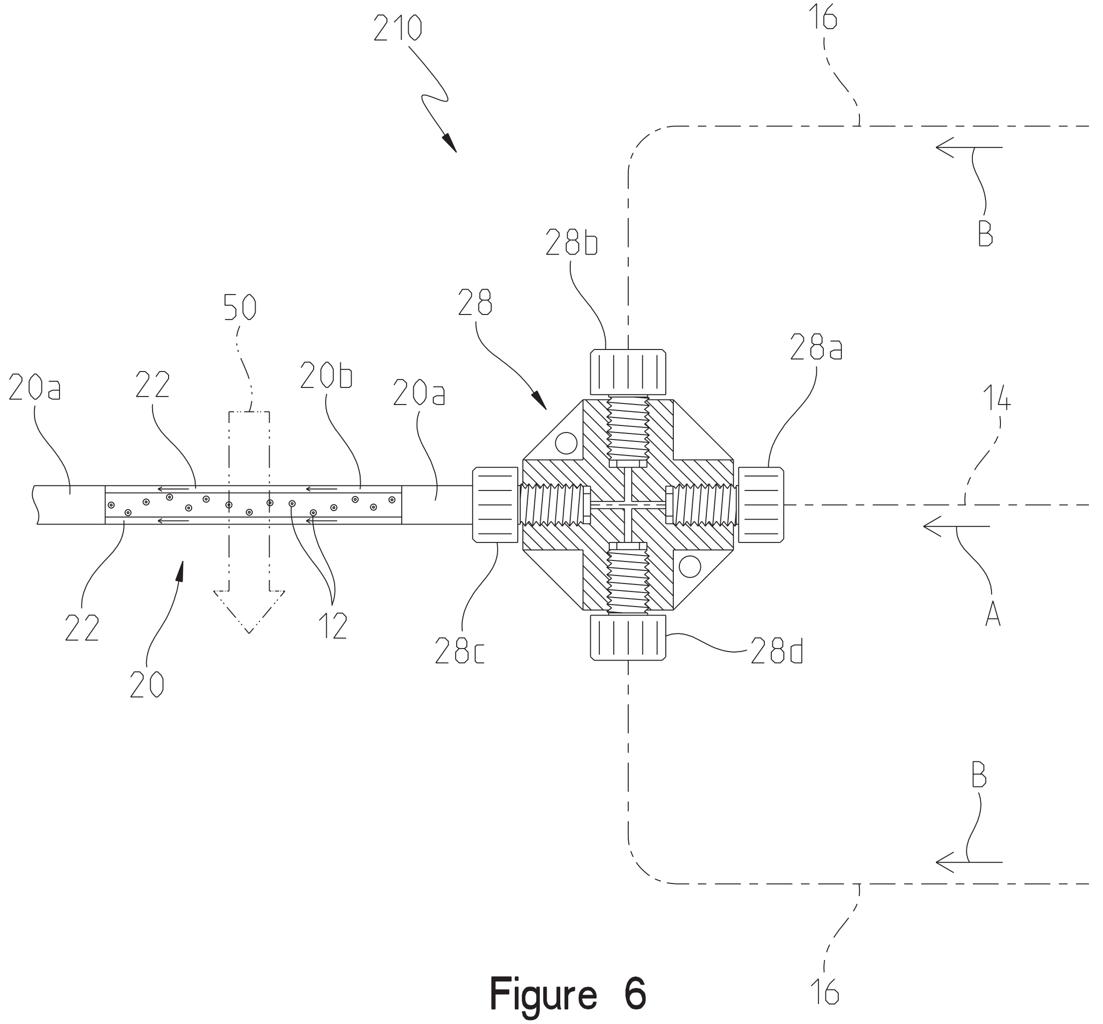

210
16
B
28b
28
50
28a
20a
22
20b
20a
14
22
12
28c
28d
A
20
B
16

Figure 6

VITA

VITA

Aimee Rinas received her B.S. in Chemistry from Purdue University in 2006, M.S. in Biology from Purdue University in 2007, and returned to Purdue in 2011 to to pursue her doctoral studies. In the time between receiving her M.S. and returning for her PhD, she worked in industry at Roche Diagnostics and Thermo Fisher Scientific, both in Indianapolis, in clinical immunoassay development and support.

Upon returning to Purdue, Aimee joined the research lab of Dr. Lisa Jones in December of 2012 after switching focus from computational chemistry with Dr. Jingzhi Pu to bioanalytical mass spectrometry. With Dr. Jones, her research focused on upgrading the mass spectrometry-based protein footprinting method Fast Photochemical Oxidation of Proteins (FPOP) to the next generation, expanding the applicability of the method. Her research work includes both method and device development allowing for complex biological samples, including live cells, to be studied using FPOP. She defended her dissertation in April 2016, and will be returning to industry in as a proteomics application specialist at Thermo Fisher Scintific.

PUBLICATIONS

Author's personal copy

Anal Bioanal Chem
DOI 10.1007/s00216-016-9369-3

RESEARCH PAPER

An efficient quantitation strategy for hydroxyl radical-mediated protein footprinting using Proteome Discoverer

Aimee Rinas[1] · **Jessica A. Espino**[1] · **Lisa M. Jones**[1]

Received: 1 November 2015 / Revised: 17 January 2016 / Accepted: 27 January 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Hydroxyl radical protein footprinting coupled with mass spectrometry has become an invaluable technique for protein structural characterization. In this method, hydroxyl radicals react with solvent exposed amino acid side chains producing stable, covalently attached labels. Although this technique yields beneficial information, the extensive list of known oxidation products produced make the identification and quantitation process considerably complex. Currently, the methods available for analysis either involve manual analysis steps, or limit the amount of searchable modifications or the size of sequence database. This creates a bottleneck which can result in a long and arduous analysis process, which is further compounded in a complex sample. Here, we report the use of a new footprinting analysis method for both peptide and residue-level analysis, demonstrated on the GCaMP2 synthetic construct in calcium free and calcium bound states. This method utilizes a customized multi-search node workflow developed for an on-market search platform in conjunction with a quantitation platform developed using a free Excel add-in. Moreover, the method expedites the analysis process, requiring only two post-search hours to complete quantitation, regardless of the size of the experiment or the sample complexity.

Keywords Oxidative modification · Mass spectrometry · Data analysis · Protein footprinting

Introduction

Structural biology has played an essential role in advancing our functional understanding of biological systems. At present, approximately 11 % of Swiss-Prot proteins have high-resolution structures available [1, 2], with roughly 90 % of available structures solved via X-ray crystallography [1]. However, the ability to obtain a high-resolution structure by means of this method is limited by the ability to obtain a high-quality crystal [3]. This task is often difficult due to the inherent properties of the protein or protein assembly, or the ability to purify the macromolecule in quantities sufficient to obtain a quality crystal [4]. As such, the use of mass spectrometry (MS)-based structural methodologies has become increasingly more popular in filling the structural void as well as offering complementary information to high-resolution methodologies [5, 6]. Additionally, these applications can provide structural information for membrane proteins [7, 8], which make up less than 3 % of known structures [1], or to study conformational changes and folding dynamics [9–11].

One type of MS-based structural method, protein footprinting, utilizes a chemical probe that covalently labels the protein in solution [6]. These approaches are used to probe structural features of the protein. The primary aim of MS-based protein footprinting is to determine which sites on a protein are solvent accessible (SA), exploiting the fact that target site reactivities are dictated, in part, by the protein's conformation [5]. The exception is hydrogen-deuterium

---

Published in the topical collection featuring *Young Investigators in Analytical and Bioanalytical Science* with guest editors S. Daunert, A. Baeumner, S. Deo, J. Ruiz Encinar, and L. Zhang.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-016-9369-3) contains supplementary material, which is available to authorized users.

✉ Lisa M. Jones
joneslis@iupui.edu

[1] Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, 402 N. Blackford St LD326, Indianapolis, IN 46202, USA

Published online: 12 February 2016

Author's personal copy

exchange (HDX), which also takes hydrogen bonding into account [12]. Typically, chemical probes are used to assess sites on a protein that show a change in SA consequent to being exposed to a ligand [13, 14], or upon complexation with another protein(s) [15, 16].

There are several different strategies for chemical footprinting that vary in whether they probe the amino acid side chains or the peptide backbone, or in the specificity of the chemical probe. In HDX, the non-specific label, deuterium, is exchanged with backbone amide protons to probe the protein structure [17]. In contrast, amino acid-specific labels may be used, such as using organic anhydrides to probe the SA of lysine residues [18].

The use of hydroxyl radicals (·OH) for protein footprinting, first coupled with mass spectrometry by Chance and coworkers [19], has become more prevalent for probing sidechain SA due to the advantages this chemical probe offers. Similar to deuterium in their specificity, ·OH are reactive with 19 of the 20 natural amino acids (excluding glycine), with varying rates [20, 21]. It should be noted, however, that detecting the oxidation products of serine and threonine can be problematic and are usually excluded from analysis [21]. Unlike hydrogen-deuterium exchange, ·OH-induced modifications do not undergo back-exchange and are stable, allowing for a higher, residue-level resolution [6, 16]. Additionally, since ·OH have properties similar to water, they can readily oxidize SA amino acid side chains without disrupting the structure when the exposure is limited [6]. Furthermore, multiple methods may be employed for generating the ·OH radicals, including pulse radiolysis of water, Fenton chemistry, and flash photolysis of hydrogen peroxide, making the ·OH labeling method available to a wide array of researchers [22–24].

Despite the advantages of·OH footprinting, data analysis for this method remains arduous, limiting the potential it has in the field of structural proteomics. Analysis generally follows a typical proteomics workflow; proteins are proteolyzed, the resulting peptides are detected and identified using data-dependent LC-MS/MS, and precursor peak intensities or areas are used for quantitation [25]. However, the abundant amount of possible oxidation products produced by this method makes identification and quantitation significantly more complex than a proteomics counterpart such as SILAC, where stable isotope ratios are quantified [16, 26, 27]. Furthermore, the vast quantity of data generated from these experiments makes manual analysis cumbersome, which can lead to errors in interpretation [28].

Several efforts have been made to expedite analysis of these data sets. Gau et al. have developed a semiautomatic method that utilizes a Mascot error-tolerant search and a house-built Excel spreadsheet for identification and quantitation [29, 30]. However, this method is manually intensive, and the use of an error-tolerant search prohibits the use of a decoy database search and, subsequently, a false discovery rate (FDR) or a level of confidence in identifications. Hybrid software platforms including ByOnic and InsPecT, that were developed for identifying post-translational modifications, have been utilized for ·OH experiments [31–34]. While these platforms offer a better-quality database search, quantitation remains manually intensive. And although Kaur and coworkers [28] developed an automated software platform, ProtMapMS (PM), it was specifically designed for dose response studies from water radiolysis. Furthermore, the presupposition that the protein of interest is known excludes the platform from being used in highly complex samples, such as cell lysate and in cell labeling strategies that have been developed recently [35, 36]. And while the resolving power of ProtMapMS has been improved in the most current version, the limitations that exclude its use on complex samples remain [37, 38]. For each of these platforms, quantitation of the oxidation yield is performed using the ratio of the identified oxidized species to the sum of the oxidized and corresponding unoxidized species. However, the oxidation yield can also be determined indirectly by monitoring the decrease in the unoxidized species, whereby this decrease is attributed to the species being oxidized [39]. While this has been shown to be quantitatively accurate, this approach can only be used at a peptide level, limiting the resolution of identifying the oxidized locations.

In this paper, we present our Proteome Discoverer (PD) analysis strategy using fast photochemical oxidation of proteins (FPOP) as the footprinting method. To illustrate the efficacy of the strategy, both peptide and residue-level analysis of the GCaMP2 synthetic construct in the apo and holo conformations is provided, comparing the results between the conformations to the available crystal structures and results from ProtMapMS [28, 40].

Materials and methods

Materials

All chemicals were obtained from Thermo Fisher Scientific (Waltham, MA) unless otherwise noted.

Protein expression and purification

The pRSET vector containing GCamP2 was a kind gift from Michael Kotlikoff (Cornell University). The expression and purification of GCaMP2 was previously described [41]. Briefly, transformed BL21(DE3)pLysS (Promega, Madison, WI) *E. coli* cells were grown in LB media. After induction with 1 mM IPTG, proteins were expressed for 12 h at 18 °C. Proteins were purified using the HisPur Ni-NTA agarose resin.

Oxidative labeling

Each 50-μL sample contained 10 mM phosphate-buffered saline (PBS, Sigma Aldrich, St. Louis, MO) 10 mM

Author's personal copy

L-glutamine, 7.5 mM hydrogen peroxide, 10 mM ethylenediaminetetraacetic acid (calcium free, CF) or 10 mM calcium chloride (calcium bound, CB), and purified GCaMP2 at a concentration of 0.18 mg/ml. The hydrogen peroxide was added just prior to infusion. FPOP was performed similarly as described [24, 35, 42]. A 248-nm KrF excimer laser (GAM Laser Inc., Orlando, FL) was used to irradiate the sample solution at 135 mJ/pulse. The laser was focused through a 250-mm plano convex lens (Thorlabs, Inc., Newton, NJ) onto 150 μm i.d. fused silica tubing (Polymicro Technologies, Pheonix, AZ) with the polyimide coating removed, giving a 2.5-mm irradiation window. The flow rate, 33 μl/min, was set to allow for a 20 % exclusion fraction. A total of three FPOP samples and three controls (no irradiation) were prepared for each condition were prepared.

Proteolysis

Post FPOP, the GCaMP2 samples were subjected to a two-step digestion process as previously described [15, 35]. Each sample dried in a vacuum centrifuge, and resuspended in 8 M urea 100 mM Tris–HCl pH 8.5 buffer. Proteins were reduced with tris(2-carboxyethyl)phosphine, alkylated with iodoacetamide and quenched with dithiothreitol. Lys-C was added at a 100:1 substrate to protease ratio and incubated overnight at 37 °C. The samples were then diluted with 100 mM Tris buffer to bring the urea concentration down to 2 M. Trypsin was added at a 20:1 substrate to protease ratio and incubated overnight at 37 °C. Digestion was quenched with formic acid (Sigma Aldrich, St. Louis, MO) at a final concentration of 5 %. Samples were de-salted using NuTip$_{C18}$ media tips (Glygen Corporation, Columbia, MD), dried in a vacuum centrifuge, and resuspended in 20 μl 2 % acetonitrile 0.1 % formic acid.

LC-MS/MS acquisition

Analysis was completed using an UltiMate 3000 RSLC and a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA) as previously described [35]. For each experiment, 1 μg of the digest was loaded onto a 2 cm Acclaim Pepmap 100 C18 trap column (Thermo Fisher Scientific, Waltham, MA) and washed for 3 min with loading buffer (2 % acetonitrile 0.1 % formic acid) at a flow rate of 5 μl/min. The samples were separated on a 75-μm inner diameter reverse phase analytical column packed in-house with a 30 cm bed of Magic 5 μm C18 particles (Michrom Bioresources Inc., Auburn, CA). Peptides were eluted with a 74-min linear gradient at a flow rate of 300 nl/min to 45 % acetonitrile 0.1 % formic acid. The total run time was 97 min including loading, washing, and equilibration time. MS1 spectra were acquired over an m/z range of 300–1500 at a resolving power of 70,000 for 400 m/z ions, with a dynamic exclusion of 20 s. The 25 most abundant ions were selected for MS2 at a resolving power of 17,500 for 400 m/z ions. Ions with a charge-state of +1 and > +6 ions were rejected. AGC targets were set to 3e6 for MS1 and 1e5 for data-dependent MS2 with an underfill ratio of 2.5 %, giving an intensity threshold of 5.0e4.

Data analysis

All data files were searched using Proteome Discoverer (version 1.4; Thermo Fisher Scientific, San Jose, CA, USA) with Mascot (version 2.4; Matrix Science, London, UK) and Sequest HT (version 1.1.1.11; Thermo Fisher Scientific, San Jose, CA, USA) using a custom multi-search node workflow (Fig. 1), and the ProtMapMS platform (version 2.5.0.30). For the PD method, files were searched against a FASTA database consisting of the protein of interest (GCaMP2 synthetic construct; NCBI GI 218681839) inserted into the cRAP (Common Repository of Adventitious Proteins) database (version 2012.01.01; http://www.thegpm.org/crap/), and extracted ion chromatogram (EIC) areas for each peptide spectrum match (PSM) were calculated. For the PM method, files were searched against the GCaMP2 FASTA sequence only, restricted to tryptic peptides with up to one missed cleavage within a 350–5000 Da mass range. The precursor mass tolerance and the m/z peak integration width were set to ± 10 ppm, and the fragment mass tolerance to ±0.02 Da.

Post search, peptides were ungrouped and filtered to a 1 % FDR for both peptide and residue-level analysis. The data were exported to Excel and summarized using the PowerPivot add-in that was customized for ·OH labeling experiments. The fractional oxidation per peptide or residue was determined according to the following equation:

$$\frac{\sum \text{EIC area modified}}{\sum \text{EIC area}} \tag{1}$$

where, for peptide level analysis, EIC area modified is the EIC area of a PSM containing an FPOP modification, and EIC area the EIC area of any PSM with a sequence identical to that containing the modification. For residue-level analysis, EIC area modified is the EIC area of a PSM for a specific modified residue, and EIC area the EIC area of any PSMs with sequences identical to those containing the modification.

Results

Analysis platform

Our objective for the work presented here was to develop an identification and quantitation strategy, with minimal manual steps, for use in protein footprinting studies and with four main features. First, we wanted to use an identification

Author's personal copy

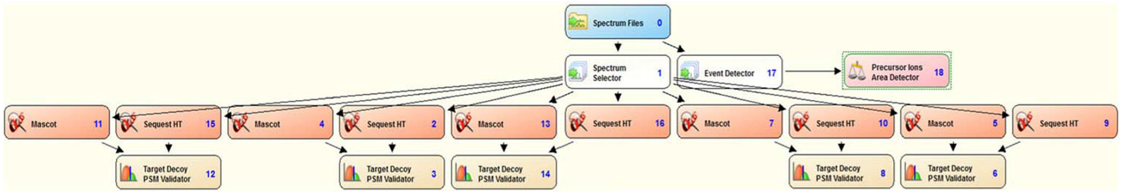

Spectrum Files 0
Spectrum Selector 1
Event Detector 17
Precursor Ions Area Detector 18
Mascot 11
Sequest HT 15
Mascot 4
Sequest HT 2
Mascot 13
Sequest HT 16
Mascot 7
Sequest HT 10
Mascot 5
Sequest HT 9
Target Decoy PSM Validator 12
Target Decoy PSM Validator 3
Target Decoy PSM Validator 14
Target Decoy PSM Validator 8
Target Decoy PSM Validator 6

Fig. 1 PD workflow developed for FPOP analysis displaying the search hierarchy with Mascot and Sequest HT search levels (*dark orange nodes*), target-decoy validation (*tan nodes*), and precursor peak areas are calculation (*pink node*)

strategy that builds on the standards that have been established in the proteomics community. Second, we wanted the platform to be compatible with all levels of sample complexity. Third, we wanted a platform that would work with any stable footprinting label. Last, we wanted our platform to be built on software packages that are already widely used within the proteomics community. As follows, we created a workflow (Fig. 1) in Thermo Scientific's Proteome Discoverer (PD) that was customized for deep searching of hydroxyl radical modifications, as well as a rapid quantitation platform using the Excel PowerPivot add-in. Although our strategy was designed for hydroxyl radical labeling experiments, each half may be easily modified to accommodate any stable footprinting label.

The workflow (Fig. 1) was constructed with five search algorithm levels, each restricted to tryptic peptides greater than 5 residues, with up to one missed cleavage, and within a 350–5000 Da mass range. The precursor mass tolerance was set to ± 10 ppm, and the fragment mass tolerance to ± 0.02 Da. Alkylation and all commonly observed · OH modifications [16, 26] were distributed across each search level (Table 1), and each data set was searched against the FASTA database. After each search level, PSMs are subjected to target-decoy validation, with a 1 % FDR setting for high confidence PSMs. Extracted ion chromatograms (EICs) for each precursor are calculated using the PD precursor ion area detection node.

After the search is completed, the ungrouped (all visible PSMs) peptide PSMs matching the protein(s) of interest are filtered appropriately (1 % FDR for peptide level analysis; 1 % FDR and search engine rank of 1 for residue level analysis), removing PSMs that do not have a calculated peak area; peak areas are only calculated by PD when the expected precursor pattern is detected. The resulting data is then exported to an Excel spreadsheet, where it is prepared for import into PowerPivot. Once imported, a series of PowerPivot measures that were formulated specifically for hydroxyl radical labeling, quantitate the extent of oxidation for each peptide or residue (Eq. 1). The time required for all post search steps, including filtering, exporting, preparation, and quantitation, generally takes under 2 h; this is typical regardless of the number of samples, proteins, or conditions for the given experiment.

Table 1 Workflow modification distribution

Search Level	Amino Acid(s)	Mass Shift (Da)
1	C (Static)	+57
2	A,D,E,F,H,I,K,L,M,N,P,Q,R,V,W,Y	+16
3	E,I,K,L,P,Q,R,V	+14
	C,F,M,W,Y	+32
	H	−10
4	C,F,W,Y	+48
	D,E	−28,-30
	H	+5
	R	−43
5	D,E	−44
	H	−22,-23

In order to complete residue-level analysis with confidence, changes were made to how the data was pre-filtered before tabulation. First, all PSMs with more than one oxidation location are calculated independent from the singly oxidized counterparts as there is no way to discern what percentage of the area associated with that PSM can be attributed to each oxidation location. Second, PSMs are restricted to those assigned a search engine rank of 1 to reduce the number of PSMs that have a low probability of having a b or y ion specifically identifying the labeled residue, and to ensure that each MS/MS spectrum has at most one match per search algorithm (Sequest HT and Mascot) node. These strict filters are applied up front to minimize the amount of manual validation of MS/MS spectra required post analysis.

Analysis of GCaMP2 in calcium-free and calcium-bound states

The synthetic construct, GCaMP2, was used to evaluate the efficacy of the PD method for both peptide and residue-level analysis. GCaMP2 belongs to a group of synthetic proteins known as genetically encoded indicators [40]. It is a calcium sensor assembled from circularly permutated enhanced green fluorescent protein (cpEGFP), the M13 helix of myosin light chain, and calmodulin (CaM) [43]. The sensor undergoes conformational changes upon calcium binding that result in a change in solvent accessibility of the chromophore allowing for calcium dependent sensing in cells and tissues [40, 41]. At low concentrations, comparable to those used in FPOP experiments, GCaMP2 exists as a monomer in both the calcium-

Author's personal copy

free (CF) and calcium-bound (CB) states, with monomer crystal structures available for both states [40]. This makes GCaMP2 an ideal molecule for examining the ability of the PD method in discerning conformational changes via localized changes in oxidation, with the structures providing a means of validating the residue-level data.

Peptide level FPOP oxidation of GCaMP2 was completed using both the PD method and the PM method. PD analysis resulted in the detection of 62 peptides, with 21 containing FPOP modifications which are presented in the Electronic Supplementary Material (ESM, Table S1). There were a total of 443,780 PSMs at a 1 % FDR affiliated with the protein, with 11 % corresponding to hydroxyl radical oxidation events. These peptides correspond to a sequence coverage of 92 % and a footprinting coverage of 66 % of the protein (Fig. 2a). By comparison, 83 % of the sequence was covered using the PM method with only 33 % footprinting coverage (Fig. 2a). There were eight oxidized peptides identified in common between the two methods which were used to validate the accuracy of the PD method's quantitation strategy. Recently, Chance and coworkers have introduced an oxidation normalization factor, named protection factor (PF), where the relative intrinsic activity (R_i) of the peptide (or residue for residue-level analysis) is divided by the oxidation rate to account for the differing reactivities of each of the amino acids with ·OH, which we have applied to the data presented in this paper [37, 38]. Comparison of the two methods shows a similar trend in fold change of oxidation (Fig. 2b), and results in a correlation coefficient of 0.99 for both the CF and CB states (Fig. 2c), with *p* values of 3.59e−9 and 2.53e−9, respectively. These correlations demonstrate that the quantitative accuracy of the PD platform is comparable to the previously validated PM platform. To further establish the efficacy of the PD method, the solvent accessible surface area (SASA) was calculated for each of the structures using VADAR [44], and compared to the natural log of the PFs for the 22 oxidized peptides identified (Fig. 3), as previously demonstrated [37, 38]. This resulted in a correlation of −0.84 for the CF state, and −0.59 to the CB monomeric structure (3EK4) and −0.75 to the CB dimeric structure [40]. The higher correlation to the dimeric CB structure indicates that dimer was present in the CB samples, contrary to what was expected at the low concentration (0.18 μg/μl) of GCaMP2 in each sample prepared. This was most likely due to the method in which the samples were prepared. In order to minimize the variability between samples, calcium chloride was added to a concentrated stock of GCaMP2 which was later spiked into each sample tube. Studies completed by Akerboom and coworkers [40] on GCaMP2 demonstrated that dimeric CB GCaMP2 can be seen at concentrations of 10 μM and that the equilibrium kinetics between the monomeric and dimeric calcium-saturated forms was slow, with the solution remaining stable over several days. With the stock concentration of CB GCaMP2 for this study at 0.29 μM, nearly three times the concentration where dimer has been shown to exist, there is a strong likelihood that dimeric CB GCaMP2 was present in the samples at the time of oxidation.

Further discussion of GCaMP2 will be divided into two talking points: the GFP domain (which, for ease of discussion, also includes the M13 peptide and the domain linkers) and the CaM domain. There are two areas of the GFP domain that have unexpected differences in oxidation between the CF and CB states. The first occurs between the two peptides spanning residues 255–284, with a higher relative oxidation in the CF state. Alignment of the GFP domain (residues 62–301) of the two structures (PDB IDs 3EKJ and 3EK4) using Pymol results in an RMS deviation value of 0.347 Å, indicating that there are only small structural changes in this domain between the two states [40, 45]. To investigate further, the solvent-accessible surface area (SASA) was calculated for each of the structures using VADAR [44]. Comparison of the SASA from residues 255–284 revealed that there is an increased SASA for the CF structure, in agreement with the FPOP oxidation analysis (Fig. 4, ESM Table S1). The 4 % SASA difference between the states is mainly associated with six lysine and arginine residues (K259, K265, R267, K271, R280, and K284). Inspection of the PSMs for these peptides provides additional agreement with the SASA analysis, as there are considerably more oxidized PSMs associated with these residues in the CF state (105 vs 58 PSMs).

Another point of interest in the GFP domain is with peptides 53–69 and 57–69 (Fig. 4, ESM Table S1). These two peptides contain a portion of the M13 peptide, the linker joining M13 to GFP, and the beginning of the C-terminal half of GFP. At first it would seem unlikely that in overlapping peptides there would be a change in which state would have a greater relative oxidation. This is especially true considering there is only a four-residue difference between the two peptides. However, this change can be explained by the structures. The CF structure (PDB ID 3EKJ) is lacking the M13 peptide due to its flexibility, while in the CB state (PDB ID 3EK4) M13 is complexed with the CaM domain [40]. This would correlate to a higher oxidation on the M13 peptide in the CF state, as is seen in peptide 53–69. Peptide 57–69 has fewer M13 residues, so it is likely that higher oxidation in the CB state is associated with the subtle structural changes in the M13-GFP linker between the two states. These differences, however, may be better assessed on a per-residue basis.

Residue-level analysis of GCaMP2 using the strict filter settings resulted in a substantial decrease in PSMs with only 87,298 passing all the criteria. An even more dramatic change can be seen when looking at oxidized PSMs, going from 443,780 PSMs and 11 % of the total, to 10,384 PSMs and 7.4 % of the total. This resulted in a total of 61 singly oxidized residues identified, with an additional 11 double oxidation identifications (ESM Table S2). While drastic, these changes allow for a more discriminative view of the conformational changes that occur between the two states, as demonstrated by the M13-GFP linker.

Author's personal copy

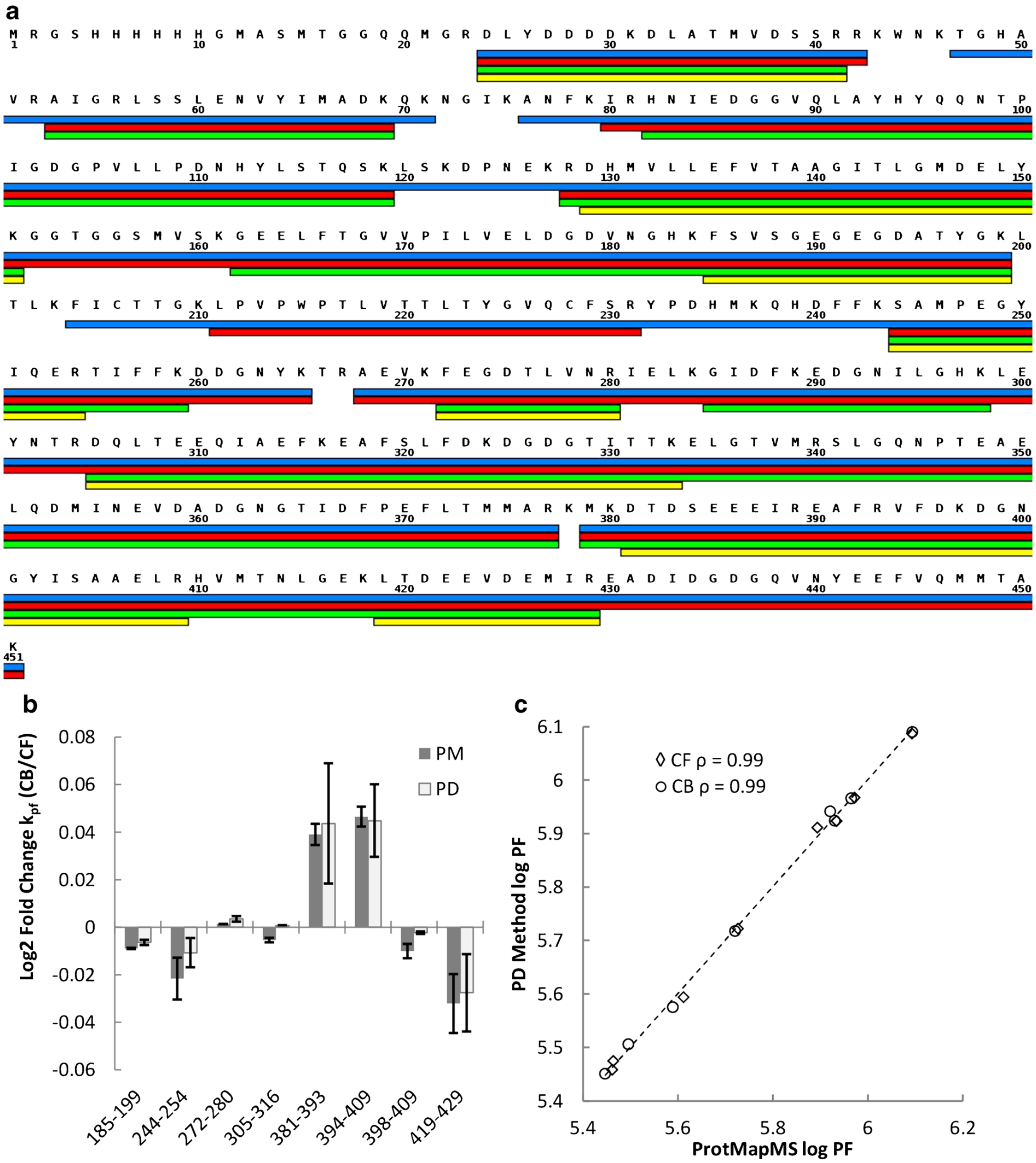


Fig. 2 Comparison of PD method to PM method showing the sequence coverage (**a**) with PD total and footprinting coverage in *blue* and *green*, respectively, and PM total and footprinting coverage in *red* and *yellow*, respectively, as well as the agreement between the fold change (**b**) and correlation (**c**)

At a residue level, the PD method was able to detect the conformational change between the two states on the L60-E61 linker between M13 and GFP. In the CF state, the linker is integrated into the leading GFP β-strand, with the structure stabilized through hydrogen bonding between E61 and R81. This integration decreases the SASA of L60, while exposing the E61 sidechain. However, the linker is forced into a different conformation in the CB state as a result of the strong

Author's personal copy

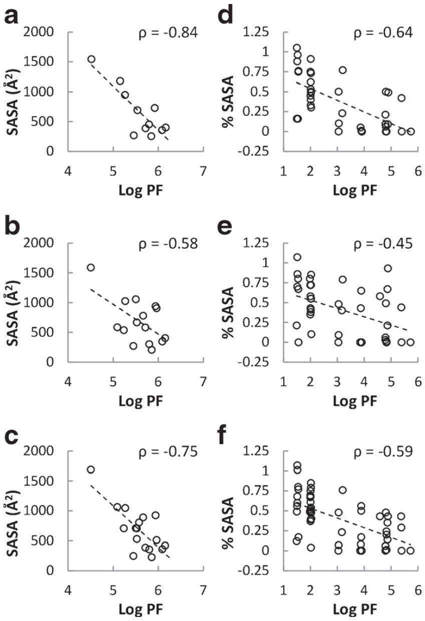


Fig. 3 Correlation between SASA and log PF for both peptide level (**a**, **b**, and **c**) and residue level (**d**, **e**, and **f**) analysis for the CF state (**a** and **d**) and the CB monomeric (**b** and **e**) and dimeric (**c** and **f**) states

interaction between the M13 peptide and the CaM domain (Fig. 5a) [40]. This interaction inverts the SASA for the linker residues, with L60 more exposed in the CB states. Residue-level analysis using the PD method matches this trend, with a higher oxidation on L60 in the CB state and E61 in the CF state (Fig. 5b). In addition to matching the trend for the linker, the correlations between %SASA (the fractional SASA of the residue) and log PF of the 61 oxidized residues are comparable to those that have been previously published, with −0.64 for the CF GCaMP2 and −0.45 and −0.59 for monomeric and dimeric CB GCamP2 (Fig. 2) [37, 38].

Another structural change between the CF and CB states stems from the architecture of the calcium sensor. In designing the cpEGFP moiety, four residues were removed from a GFP β strand. This creates an opening in the GFP barrel, resulting in an increased solvent accessibility in the barrel interior. Conformational changes induced by calcium binding cause the N-terminal domain of CaM to partially block this opening, decreasing the solvent accessibility of the chromophore [40]. In the CF state, the N-terminal domain of CaM is packed against the cpEGFP domain, creating a region of decreased SASA on the surface of the cpEGFP domain (Fig. 6a). However, in the CB state, the interaction of the CaM domain with the M13 peptide increases the exposed SA in this region of the GFP domain (Fig. 6b, ESM Table S2). Two oxidatively modified peptides were identified in this cpEGFP region. At the peptide level, peptides 128–151 and 185–199 both have an increase in oxidation in the CF state (Fig. 4, ESM Table S1). Residue-level analysis also agrees with the conformational differences, identifying three residues on the occluded cpEGFP surface that have higher oxidation yields in the CB state (Fig. 6c).

Although the surface of cpGFP has increased solvent exposure in the CB state, the same claim cannot be made for the CaM domain. Helices 8 and 11 of GCaMP2, which comprise the majority of the residues that are packed against the GFP domain in the CF state, are also complexed with the M13 peptide in the CB state [40]. Additionally, the two crystal structures for this domain are incomplete, making it difficult to assess the CaM residue level data using the structures. However, the CB CaM domain is structurally similar to the CaM-M13 crystal structure, with only small differences associated with the CaM linker helix [40, 46]. By reason of this structural similarity, a previously published FPOP study of calmodulin with the M13 peptide was used for comparison [14].

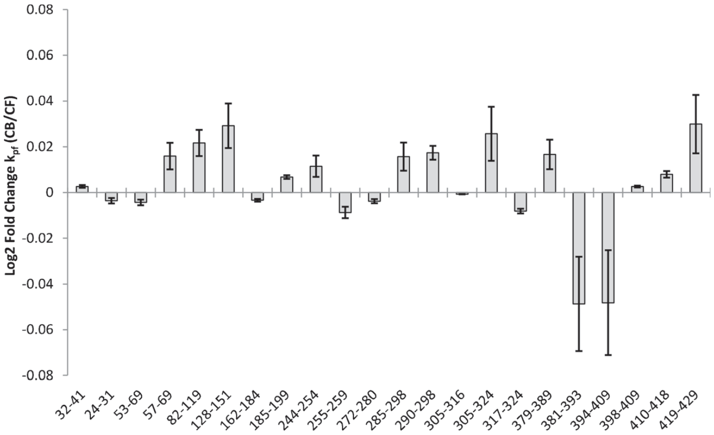


Fig. 4 Fold change between the CF and CB states for the 22 oxidized peptides identified by the PD method

Author's personal copy

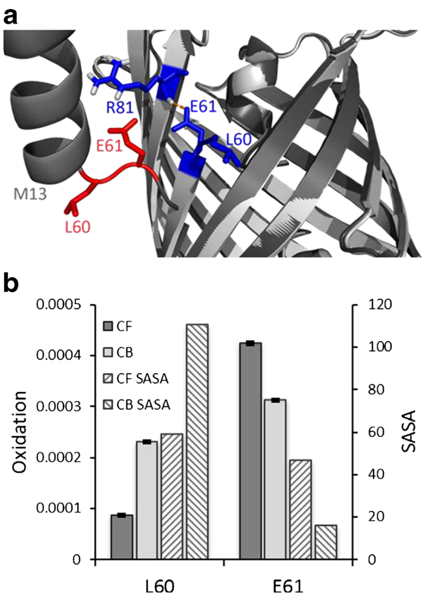


Fig. 5 Conformational difference detected in the two residues in the M13-GFP linker with an overlay of the calcium free (*light gray* and *blue*, PDB ID: 3EKJ) and the calcium bound structures (*dark gray* and *red*, PDB ID: 3EK4) (**a**) and the oxidation and the SASA of the two residues for each conformation (**b**)

Since definitive oxidation values for calmodulin were not given, the trend of which state (CB or CF) yielded a higher oxidation was used for comparison (Table 2). There is an agreement between the previously published data and the data presented in this paper, with two of the residues in common between this and the previous study differing. One residue, I312 (I9 in isolated CaM), has nearly identical oxidation values in the CF and CB states in the previous study. Another residue, M427 (M124 in isolated CaM) is not significantly different in the previous study. This as well as the correlation between %SASA and log PF demonstrates the efficacy of the PD method in identifying and quantifying residue level oxidation.

Discussion

Using protein footprinting methodologies with mass spectrometry for protein structural analysis can be advantageous, particularly when structural knowledge from traditional methodologies like X-ray crystallography is not available. Although these methods have the potential to fill the gap left by the traditional structural approaches, the potentially arduous task of analyzing the data can limit the size or complexity of the systems that can be studied using footprinting methods. Furthermore, any manual interpretation in the analysis steps can introduce a human bias, leading to errors in interpretation. As such, while footprinting can provide advantages over

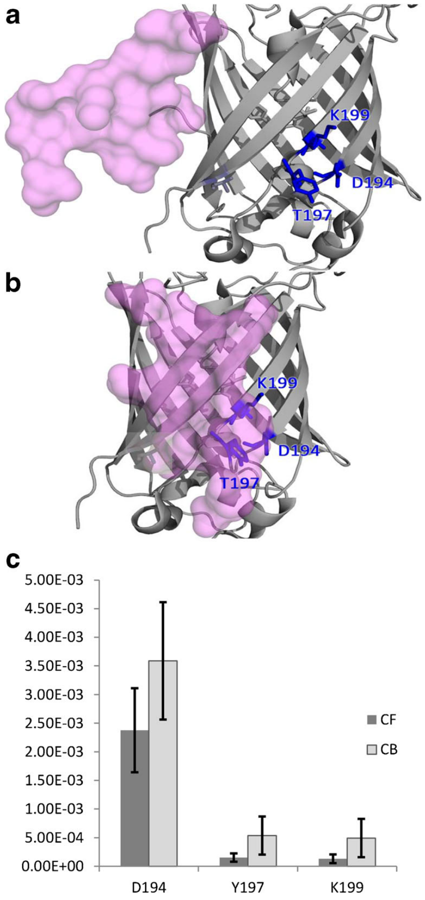


Fig. 6 Inter-domain contacts between GFP (*gray*) and the N-terminal lobe of CaM (*light purple*) in the CF conformation (**a**) and structural changes in the CB conformation (**b**) with residues with higher oxidation in the CB conformation colored *blue*, and a bar graph (**c**) of the oxidation level of the highlighted residues

Table 2 Comparison of CaM domain PD oxidation to previously published data

GCaMP2 Residue	CaM Residue	GCamp2 Higher State	CaM-M13 Higher State
I312	I9	CF	CB
F319	F16	CB	CB
L321	L18	CF	CF
F322	F19	CF	CF
Y402	Y99	CB	CB
M412	M109	CF	CF
M427	M124	CB	Same[a]

CaM-M13 data taken from literature [14]

[a] Not significantly different

Author's personal copy

traditional methods, there is an inherent limitation in their ability to completely close the gap without a more comprehensive analysis strategy.

Our goal was to develop a method that could rapidly analyze the raw data from any stable label footprinting experiment, regardless of the size or complexity of the data set. Developing a completely integrated software platform would have required an exhaustive search of every footprinting label currently in use, and would require an update for each newly published footprinting label. Additionally, this would have required restrictions on the raw data file format(s) or a strict requirement on file conversion.

To overcome these obstacles, we developed our strategy using a two-pronged approach, separating database searching from quantitation. The PD workflow (Fig. 1) used for this work can be adapted for any stable label by changing the variable modifications search to those used in the experiment, adding or deleting search levels as needed. Additionally, PD and the algorithms (Mascot and Sequest HT) allow the addition of new modifications, making it accessible to labels that have not yet been published. Nonetheless, PD is not required for quantitation in PowerPivot, as the quantitation method can be adapted to any data set formatted to a similar layout. By separating the search and quantitation strategies, our complete method can be adapted for any footprinting label and quantitation can be completed with or without the use of PD. Furthermore, post-search quantitation takes under 2 h, for one or thousands of proteins, regardless of the number of samples to analyze. Only the search time increases, as would be the case for any other platform.

The use of different filtering strategies for peptide and residue-level analysis minimizes the amount of manual interpretation and validation required, while still providing a high level of confidence in the results obtained. For peptide level analysis, it is not necessary to know the exact location of the modification on a given peptide. Therefore, it is reasonable to conclude that corresponding modified PSMs identified within the mass tolerance are correct within a 95 % confidence interval, and that the ambiguity of the location assignment is inconsequential to the overall outcome at this level of detail.

A more stringent approach must be taken to provide a high level of confidence in modification assignments for residue-level analysis. Restricting PSMs to only those with one modification is necessary as there is not a reliable method to assign peak area proportions to each modification multiply oxidized PSMs. The additional restrictions of a 1 % FDR and limiting the search engine rank to 1 allow for a high level of confidence in per-residue analysis while limiting the amount of manual interpretation required. Comparing the MS/MS scans for two identified modifications on the same peptide in GCaMP2 best illustrates this point. A decarboxylation modification on residue D194 was identified with high confidence and a search engine rank of 3 (ESM Fig. S1a), but does not have fragment peaks to substantiate this assignment. Without the filters, this assignment would have had to be validated manually, increasing the amount of time required to complete analysis. For comparison, a loss of CO on this same residue was identified with high confidence and a search engine rank of 1 (ESM Fig.S1b), and has a more complete y ion series coverage in the MS/MS spectrum with a fragment matching the modification. By using strict filters on the data up front, a high level of confidence in the assignments can be obtained without the need to manually remove PSMs individually which drastically decreases the time required for analysis. This is further demonstrated by the oxidation levels of the two M13-cpEGFP linker residues, L60-E61 (Fig. 2). Glutamic acid is one of the least ·OH-reactive amino acids, with a reactivity 2.3×10^8 [16]. Manual interpretation may have been biased by the more highly reactive residues on the peptide; including L61 with a 6-fold higher reactivity, Y64 with a 17-fold higher reactivity, and M66 with a 30 fold higher reactivity. By letting the algorithms make the assignment, no human bias can by factored into the analysis, and the detection of oxidized residues with lower reactivities is not hindered. While there is still a small probability of incorrect assignments, the use of an *n* number of PSMs threshold decreases the likelihood of the incorrect assignments being included while interpreting the results. Furthermore, the upfront reduction of PSMs before analysis is completed also decreases the number of data points, and therefore the time required for manual verification of the results, if desired.

Conclusions

In this work, we have demonstrated our PD method is a complete identification and quantitation strategy for mass spectrometry-based protein footprinting studies. The data presented here demonstrates the efficacy of the PD method in detecting and analyzing protein footprinting data. Use of this method on GCaMP2 validates that this method not only provides significant footprinting coverage while maintaining a high level of confidence in the results obtained, but does so in less than 2 h. Additionally, our method can be adapted for use with any stable covalent label and any level of experiment or sample complexity. The approach we used when developing our method will allow for it to be modified as new technologies and footprinting strategies are developed.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank www.rcsb.org. Nucleic Acids Res. 2000;28(1):235–42. doi:10.1093/nar/28.1.235.

Author's personal copy

2. Consortium TU. UniProt: a hub for protein information. Nucleic Acids Res. 2015;43(D1):D204–12. doi:10.1093/nar/gku989.
3. Holton JM, Frankel KA. The minimum crystal size needed for a complete diffraction data set. Acta Crystallogr Sect D: Biol Crystallogr. 2010;66(4):393–408.
4. Boutet S, Lomb L, Williams GJ, Barends TRM, Aquila A, Doak RB, et al. High-resolution protein structure determination by serial femtosecond crystallography. Science. 2012;337(6092):362–4. doi:10.1126/science.1217737.
5. Konermann L, Stocks BB, Pan Y, Tong X. Mass spectrometry combined with oxidative labeling for exploring protein structure and folding. Mass Spectrom Rev. 2010;29(4):651–67. doi:10.1002/mas.20256.
6. Wang L, Chance MR. Structural mass spectrometry of proteins using hydroxyl radical based protein footprinting. Anal Chem. 2011;83(19):7234–41. doi:10.1021/ac200567u.
7. Pan Y, Stocks BB, Brown L, Konermann L. Structural characterization of an integral membrane protein in its natural lipid environment by oxidative methionine labeling and mass spectrometry. Anal Chem. 2009;81(1):28–35. doi:10.1021/ac8020449.
8. Pan Y, Brown L, Konermann L. Mapping the structure of an integral membrane protein under semi-denaturing conditions by laser-induced oxidative labeling and mass spectrometry. J Mol Biol. 2009;394(5):968–81. doi:10.1016/j.jmb.2009.09.063.
9. Hoofnagle AN, Resing KA, Ahn NG. Protein analysis by hydrogen exchange mass spectrometry. Annu Rev Biophys Biomol Struct. 2003;32(1):1–25. doi:10.1146/annurev.biophys.32.110601.142417.
10. Poor TA, Jones LM, Sood A, Leser GP, Plasencia MD, Rempel DL, et al. Probing the paramyxovirus fusion (F) protein-refolding event from pre-to postfusion by oxidative footprinting. Proc Natl Acad Sci. 2014;111(25):E2596–605.
11. Calabrese AN, Ault JR, Radford SE, Ashcroft AE. Using hydroxyl radical footprinting to explore the free energy landscape of protein folding. Methods. 2015. doi:10.1016/j.ymeth.2015.02.018.
12. Englander SW, Sosnick TR, Englander JJ, Mayne L. Mechanisms and uses of hydrogen exchange. Curr Opin Struct Biol. 1996;6(1):18–23. doi:10.1016/S0959-440X(96)80090-X.
13. Tsutsui Y, Wintrode PL. Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions. Curr Med Chem. 2007;14(22):2344–58.
14. Zhang H, Gau BC, Jones LM, Vidavsky I, Gross ML. Fast photochemical oxidation of proteins for comparing structures of protein–ligand complexes: the calmodulin−peptide model system. Anal Chem. 2011;83(1):311–8. doi:10.1021/ac102426d.
15. Jones LM, Sperry JB, Carroll JA, Gross ML. Fast photochemical oxidation of proteins for epitope mapping. Anal Chem. 2011;83(20):7657–61. doi:10.1021/ac2007366.
16. Xu G, Chance MR. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. Chem Rev. 2007;107(8):3514–43. doi:10.1021/cr0682047.
17. Wales TE, Engen JR. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. Mass Spectrom Rev. 2006;25(1):158–70. doi:10.1002/mas.20064.
18. Mendoza VL, Vachet RW. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. Mass Spectrom Rev. 2009;28(5):785–815. doi:10.1002/mas.20203.
19. Maleknia SD, Brenowitz M, Chance MR. Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. Anal Chem. 1999;71(18):3965–73. doi:10.1021/ac990500e.
20. Garrison WM. Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. Chem Rev. 1987;87(2):381–98. doi:10.1021/cr00078a006.
21. Xu G, Chance MR. Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. Anal Chem. 2005;77(14):4549–55. doi:10.1021/ac050299+.
22. Asmus K-D (1984) [20] Pulse radiolysis methodology. In: Methods in Enzymology, vol Volume 105. Academic Press, pp 167–178. doi:10.1016/S0076-6879(84)05023-0
23. Tullius TD, Dombroski BA. Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. Proc Natl Acad Sci. 1986;83(15):5469–73.
24. Hambly D, Gross M. Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. J Am Soc Mass Spectrom. 2005;16(12):2057–63. doi:10.1016/j.jasms.2005.09.008.
25. Cutillas PR, Timms JF (2010) LC-MS/MS in proteomics: methods and applications. Humana Press
26. Gau BC, Chen H, Zhang Y, Gross ML. Sulfate radical anion as a new reagent for fast photochemical oxidation of proteins. Anal Chem. 2010;82(18):7821–7. doi:10.1021/ac101760y.
27. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics. 2002;1(5):376–86.
28. Kaur P, Kiselar JG, Chance MR. Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. Anal Chem. 2009;81(19):8141–9. doi:10.1021/ac9013644.
29. Gau BC, Chen J, Gross ML. Fast photochemical oxidation of proteins for comparing solvent-accessibility changes accompanying protein folding: data processing and application to barstar. Biochim Biophys Acta (BBA) - Proteins Proteome. 2013;1834(6):1230–8. doi:10.1016/j.bbapap.2013.02.023.
30. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999;20(18):3551–67. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.
31. Bern M, Cai Y, Goldberg D. Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. Anal Chem. 2007;79(4):1393–400. doi:10.1021/ac0617013.
32. Tanner S, Shu H, Frank A, Wang L-C, Zandi E, Mumby M, et al. InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem. 2005;77(14):4626–39. doi:10.1021/ac050102d.
33. Charvatova O, Foley BL, Bern MW, Sharp JS, Orlando R, Woods RJ. Quantifying protein interface footprinting by hydroxyl radical oxidation and molecular dynamics simulation: application to galectin-1. J Am Soc Mass Spectrom. 2008;19(11):1692–705. doi:10.1016/j.jasms.2008.07.013.
34. McClintock CS, Parks JM, Bern M, GhattyVenkataKrishna PK, Hettich RL. Comparative informatics analysis to evaluate site-specific protein oxidation in multidimensional LC–MS/MS data. J Proteome Res. 2013;12(7):3307–16. doi:10.1021/pr400141p.
35. Rinas A, Jones LM. Fast photochemical oxidation of proteins coupled to multidimensional protein identification technology (MudPIT): expanding footprinting strategies to complex systems. J Am Soc Mass Spectrom. 2015;26(4):540–6.
36. Espino JA, Mali VS, Jones LM. In cell footprinting coupled with mass spectrometry for the structural analysis of proteins in live cells. Anal Chem. 2015;87(15):7971–8. doi:10.1021/acs.analchem.5b01888.
37. Kaur P, Kiselar J, Yang S, Chance MR. Quantitative protein topography analysis and high-resolution structure prediction using hydroxyl radical labeling and tandem-Ion mass spectrometry (MS).

Author's personal copy

Mol Cell Proteomics. 2015;14(4):1159–68. doi:10.1074/mcp.O114.044362.

38. Huang W, Ravikumar Krishnakumar M, Chance Mark R, Yang S. Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: a protection factor analysis. Biophys J. 2015;108(1):107–15. doi:10.1016/j.bpj.2014.11.013.
39. Pilau E, Iglesias A, Gozzo F. A new label-free approach for the determination of reaction rates in oxidative footprinting experiments. Anal Bioanal Chem. 2013;405(24):7679–86. doi:10.1007/s00216-013-7247-9.
40. Akerboom J, Rivera JD, Guilbe MM, Malave EC, Hernandez HH, Tian L, et al. Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design. J Biol Chem. 2009;284(10):6455–64. doi:10.1074/jbc.M807657200.
41. Wang Q, Shui B, Kotlikoff MI, Sondermann H. Structural basis for calcium sensing by GCaMP2. Structure. 2008;16(12):1817–27. doi:10.1016/j.str.2008.10.008.
42. Gau BC, Sharp JS, Rempel DL, Gross ML. Fast photochemical oxidation of protein footprints faster than protein unfolding. Anal Chem. 2009;81(16):6563–71. doi:10.1021/ac901054w.
43. Nagai T, Sawano A, Park ES, Miyawaki A. Circularly permuted green fluorescent proteins engineered to sense Ca2+. Proc Natl Acad Sci. 2001;98(6):3197–202. doi:10.1073/pnas.051636098.
44. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, et al. VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res. 2003;31(13):3316–9.
45. Schrodinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1
46. Meador W, Means A, Quiocho F. Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. Science. 1992;257(5074):1251–5. doi:10.1126/science.1519061.

© American Society for Mass Spectrometry, 2014

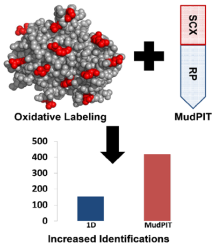
J. Am. Soc. Mass Spectrom. (2015) 26:540–546
DOI: 10.1007/s13361-014-1017-6

FOCUS: EMERGING INVESTIGATORS: RESEARCH ARTICLE

Fast Photochemical Oxidation of Proteins Coupled to Multidimensional Protein Identification Technology (MudPIT): Expanding Footprinting Strategies to Complex Systems

Aimee Rinas, Lisa M. Jones

Department of Chemistry and Chemical Biology, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202, USA



Abstract. Peptides containing the oxidation products of hydroxyl radical-mediated protein footprinting experiments are typically much less abundant than their unoxidized counterparts. This is inherent to the design of the experiment as excessive oxidation may lead to undesired conformational changes or unfolding of the protein, skewing the results. Thus, as the complexity of the systems studied using this method expands, the detection and identification of these oxidized species can be increasingly difficult with the limitations of data-dependent acquisition (DDA) and one-dimensional chromatography. Here we report the application of multidimensional protein identification technology (MudPIT) in combination with hydroxyl radical footprinting as a method to increase the identification of quantifiable peptides in these experiments. Using this method led to a 37% increase in unique peptide identifications as well as a 70% increase in protein group identifications over one-dimensional data-dependent acquisition on the same samples. Furthermore, we demonstrate the combination of these methods as a means to investigate megadalton complexes.

Key words: Protein footprinting, MudPIT

Received: 9 September 2014/Revised: 4 October 2014/Accepted: 4 October 2014/Published Online: 20 November 2014

Introduction

Structural mass spectrometry-based methodologies, many of which utilize chemical reagents for investigation, have become an invaluable tool for evaluating protein structure and function [1]. Several strategies exist for this type of investigation, each yielding different structural information about the protein(s) in inquiry. One such method is chemical crosslinking, which can divulge tertiary and quaternary information through both inter- and intramolecular covalent conjugation [2]. Hydrogen-deuterium exchange probes hydrogen bonding and solvent accessibility through monitoring the exchange of backbone amide hydrogen atoms yielding secondary structural properties [3]. Another method complementary to hydrogen-deuterium exchange is protein footprinting. Here, chemical probes are used to label side chains, revealing evidence of side chain solvent accessibility [4].

Hydroxyl radical (·OH) based footprinting, first coupled with mass spectrometry by Chance and coworkers [5], is one of the most informative covalent labeling methods for a number of reasons. The · OHs have similar properties to water and can freely oxidize solvent exposed side chains. Additionally, their reactivity is well known and researchers can capitalize on their low selectivity [6], increasing the amount of information obtained. Furthermore, there are multiple methods [1, 7, 8] available for generating · OHs, increasing the accessibility of this method. Fast photochemical oxidation of proteins (FPOP), used for this work, generates · OHs through laser induced photolysis of hydrogen peroxide [9]. This technique modifies proteins on a microsecond timescale [9], theoretically eliminating structural changes induced by labeling [10].

A consequence of design features employed in FPOP experiments to eliminate radical induced unfolding [9] is that oxidized species are present in lower abundance compared with their unoxidized counterparts. Therefore, the difficulty in detecting these species will grow concurrently with increasing sample complexity. Investigating the structures of large, megadalton-sized molecular assemblies has often proven difficult. Although there are several methods for obtaining protein structures, the majority come from X-ray crystallography [11]. At the outset, it is often challenging to purify all of the protein

---

Correspondence to: Lisa M. Jones; *e-mail:* joneslis@iupui.edu

components of megadalton complexes [12], a necessity in obtaining a structure. Even when this is accomplished, it can be equally difficult to crystallize these complexes, or the process may only yield crystals too small for analysis [11]. Although FPOP has the potential to start to fill this gap in information, it is first necessary to overcome the hurdle of identifying the relatively low abundant oxidized species in a sea of higher abundant peptides. A major obstacle is using data-dependent acquisition (DDA) for MS/MS analysis. In this method, precursor ions are selected for fragmentation based on their signal intensities. Often, if chromatographic separation is not sufficient, peptides with higher abundance are identified whereas lower abundance peptides are not. A more proficient chromatographic separation could aid in increased peptide identifications.

Multidimensional protein identification technology (MudPIT), is a method used to overcome the inability of single-dimensional separations to resolve complex biological samples [13]. The use of a biphasic analytical column increases the peak column capacity and allows for online two-dimensional separations [13, 14]. The coupling of FPOP labeling with MudPIT could provide an increase in identifications of oxidized peptides in complex systems. The use of this method to identify oxidatively modified peptides has been previously reported [15]. However, the study was mainly focused on comparison of informatics methods rather than as a method to be utilized to identify more oxidatively modified peptides for highly complex samples. Additionally, the researchers used a low complexity sample with a “mini-MudPIT” method consisting of only three salt steps. In this paper, we describe the combination of a full MudPIT method with FPOP on a highly complex sample, *Saccharomyces cerevisiae* yeast cell lysate. Our objective is to improve the detection of FPOP labeled species and expand the application of FPOP to more complex systems.

Materials and Methods

All chemicals were obtained from Thermo Fisher Scientific (Waltham, MA, USA) unless otherwise noted.

Oxidative Labeling

Each 100 µL sample contained 10 mM phosphate buffered saline (PBS; Sigma Aldrich, St. Louis, MO, USA) 10 mM -glutamine, 7.5 mM hydrogen peroxide, and yeast cell lysate (a gift from Dr. Amber Mosley and Whitney Smith-Kinnaman, Department of Biochemistry, Indiana University School of Medicine, Indianapolis, IN) at a concentration of 0.18 mg/mL. The hydrogen peroxide was added just prior to infusion. FPOP was performed similarly as described [9, 10]. A 248 nm KrF excimer laser (GAM Laser Inc., Orlando, FL, USA) was used to irradiate the sample solution at 135 mJ/pulse. The laser was focused through a 250 mm plano convex lens (Thorlabs, Inc., Newton, NJ, USA) onto 150 µm i.d. fused silica tubing (Polymicro Technologies, Phoenix, AZ, USA) with the polyimide coating removed, giving a 2.5 mm irradiation window. The flow rate, 33 µL/min, was set to allow for a 20% exclusion fraction. A total of four FPOP samples and three controls (no irradiation) were prepared.

Proteolysis

Post-FPOP, the yeast lysate samples were subjected to a two-step digestion process as previously described [16]. Each sample was acetone precipitated [17] and resuspended in 8 M urea 150 mM tris-HCL pH 8.5 buffer. Proteins were reduced with 10 mM tris(2-carboxyethyl) phosphine (TCEP) for 30 min at room temperature (RT). They were then alkylated with 20 mM iodoacetamide for 30 min at RT with a foil cover to protect the sample from light. The alkylation reaction was quenched with 10 mM dithiothreitol (DTT) for 15 min at RT. Lys-C was added at a 100:1 substrate to protease ratio and incubated overnight at 37°C. The samples were then diluted with 150 mM Tris buffer to bring the urea concentration to 2 M. Trypsin was added at a 50:1 substrate to protease ratio and incubated overnight at 37°C. Digestion was quenched with formic acid (Sigma Aldrich) at a final concentration of 5%.

LC-MS

Analysis was completed using an UltiMate 3000 RSLC and a Q Exactive mass spectrometer (Thermo Fisher Scientific). For each experiment, 1 µg of the digest was loaded onto a 2 cm Acclaim Pepmap 100 C18 trap column (Thermo Fisher Scientific). MS1 spectra were acquired over an *m/z* range of 350–2000 at a resolving power of 70,000. The 25 most abundant ions were selected for MS2 at a resolving power of 17,500. Ions with a charge-state of +1 and > +8 ions were rejected.

One-Dimensional LC-MS

Samples were loaded onto a 100 µm × 2 cm Acclaim PepMap100 C18 nano trap column (5 µm, 100 Å) (Thermo Scientific, Waltham, MA, USA) and washed for 10 min with loading buffer (LB, 2% acetonitrile 0.1% formic acid) with a flow rate of 5 µL/min. The samples were separated on a 75 µm i.d. reverse phase (RP) analytical column packed in-house with a 30 cm bed of Magic 5 µm C18 particles (Michrom Bioresources Inc., Auburn, CA, USA) with a 67 min linear gradient at a flow rate of 300 nL/min to 40% acetonitrile 0.1% formic acid. The total run time was 105 min including loading, washing, and equilibration. AGC targets were set to 3e6 for MS1 and 1e5 for data-dependent MS2 with an underfill ratio of 1.0%, giving an intensity threshold of 2.0 e4

MudPIT LC-MS

Fully automated analysis was completed in a similar manner as previously described [18, 19]. Each sample was loaded onto a trap column and washed for 10 min with LB at a flow rate of 5 µL/min. Samples were separated on a 75 µm i.d. RP analytical column packed in-house with a 26-cm bed of Magic 5 µm C18 particles (Michrom Bioresources Inc.) followed by a 4-cm bed of Luna strong cation exchange (SCX) resin (Phenomenex, Torrence, CA, USA). Peptide fractions were displaced from the SCX resin to the RP resin using the following salt pulses: (1) 0% (2) 5% (3) 10% (4) 15% (5) 20% (6) 30% (7) 40% (8) 50% (9) 60% (10) 80% of SCX buffer (SCXB, 500 mM ammonium acetate (Sigma Aldrich) in 5% acetonitrile and 0.1% formic acid) mixed with LB by the loading pump mixer. The 0% fraction was used to displace the sample from the trap column to the analytical column. Each subsequent salt pulse was generated by increasing the SCXB percentage to the next concentration with a loading pump gradient during the previous salt step. A 2.6 µL aliquot of salt (roughly 15× the SCX bed volume) was collected by coupling a 30 cm 75 µm i.d. NanoViper line (Thermo Fisher Scientific) to the trap column with a stainless steel union, and delivered when the switching valve position was changed. Each salt pulse was pushed over the analytical column by the gradient pump for 20 min at a flow rate of 300 nL/min. Sample fractions were separated with a 67-min linear gradient at a flow rate of 300 nL/min to 40% acetonitrile 0.1% formic acid. The total run time for each fraction was 105 min including loading, washing and equilibration time. AGC targets were set to 1e6 for MS1 and 5e4 for data-dependent MS2 with an underfill ratio of 1.0%, giving an intensity threshold of 1.0e4.

Analysis of MS/MS Data

All data files were searched using Proteome Discoverer version 1.4 (Thermo Fisher Scientific) with Sequest HT and Mascot ver. 2.4 (Matrix Sciences Ltd., London, UK) against a *Saccharomyces cerevisiae* FASTA database (strain ATCC 204508/S288c, downloaded from Uniprot February 2014), and extracted ion chromatogram (EIC) areas for each peptide spectrum match (PSM) were calculated using a custom multi-level workflow. Peptides were ungrouped and filtered to a 1% false discovery rate (FDR). Only PSMs identified as selected or unambiguous were used for analysis. The data was exported to Excel and summarized using the PowerPivot add-in. The fractional oxidation per residue on a given sequence was determined according to Equation 1:

$$\frac{\sum EIC\ area\ modified}{\sum EIC\ area} \tag{1}$$

where EIC area modified is the EIC area of a specific modified residue and EIC area the EIC area of any PSMs with a peptide sequence identical to that containing the modification.

Results and Discussion

Method Comparison

In order to make the most direct comparison between the one-dimensional chromatography data-dependent analysis (1D-DDA) and MudPIT, when possible, all processes and parameters were kept identical. However, some parameters were altered for the analysis. First, we doubled typical FPOP sample size to ensure there was an adequate amount to analyze each sample by both methods. In addition, the total analytical column length was kept at 30 cm and the gradient for each sample or step was constant over the entire experiment.

For comparison of the two methods, the sample loading procedure for MudPIT had to be altered. In MudPIT analysis, samples are often pressure loaded or directly injected onto the analytical column [13]. In this experiment, the samples were loaded onto a trap column via the autosampler. There were several advantages to loading the samples in this manner. First, sample washing was identical for both methods. Any hydrophilic peptides that may have been washed off of the trap column should be the same over both methods. Directly loading the sample onto a three phase analytical column could have created a bias between the two methods. Second, trap column loading allowed both method analyses to be completed continuously, whereas pressure loading would require the MudPIT analysis to be completed discontinuously. Other parameters that were altered were automatic gain control (AGC) targets, for both MS1 and MS2 (see section 2), and dynamic exclusion times. These parameters were optimized for each method to provide peak performance.

Increases in Identifications by MudPIT

In agreement with previously reported results [13, 14], using the MudPIT method to analyze the labeled yeast lysate samples gave a substantial increase in peptide spectrum matches (Figure 1). Comparing the two methods at the protein level, a 1.7-fold increase in protein group identifications (IDs), including 820 unique proteins, was observed with MudPIT (Figure 1 top). At the peptide level, a 1.3-fold increase in IDs with MudPIT was observed. Comparing unique peptides, MudPIT had a 1.7-fold increase in IDs with almost 1700 more unique peptides observed over 1D-DDA (Figure 1 middle). Although significant increases were observed with MudPIT on the protein and peptide levels, the true value of using the method in conjunction with protein footprinting is appreciated when looking at oxidatively modified peptides (Figure 1 bottom). Here, a 2.7-fold increase in oxidized peptide IDs was observed with MudPIT. Even more significant, MudPIT has a 4.6-fold increase in IDs of unique oxidized peptides over 1D-DDA. This demonstrates the efficacy for coupling MudPIT with FPOP.

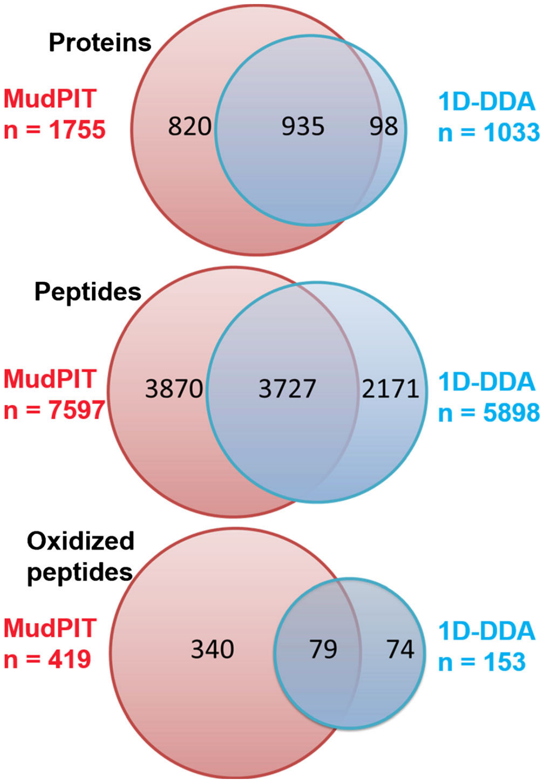


Figure 1. Visual comparison of IDs between MudPIT (red) and 1D-DDA (blue) methods by proteins **(a)**, peptides **(b)**, and oxidatively modified peptides **(c)**

The higher sequence coverage of oxidatively modified peptides will provide a more complete description of the protein system.

To further evaluate the increased IDs achieved with the MudPIT method, we compared the identification of oxidatively modified residues on pyruvate kinase 1 (PK1, PDB ID: 1A3W [20]) and phosphoglycerate kinase (PGK1, PDB ID: 3PGK [21]). These proteins were chosen as a representation because both had high coverage with each method (greater than 75%) and each had oxidatively modified peptides identified by the search workflow. For both of these proteins, modifications were only included if they were identified more than once in the samples (PSM ≥2) and if the quantifiable oxidation levels were greater than the mean standard error. Table 1 shows residues that were identified by each method. For PK1, 1D-DDA identified only six of the 14 oxidatively modified residues identified by MudPIT. For PGK1, the fourth most abundant protein found in *S. cerevisiae* [22], 1D-DDA only IDs 16 of the 41 residues that MudPIT identified. Since PKG1 is very abundant in yeast lysate, it can be assumed that this protein is oxidized more frequently than lower abundant proteins in the lysate. Consequently, the oxidized peptides from PKG1 may be relatively abundant. Despite that, 1D-DDA only IDs less than half the number of oxidized residues as MudPIT.

Table 1. Identified Oxidatively Modified Residues

	Residue	1D-DDA oxidation	MudPIT oxidation
PK1	D185	N	Y
	D187	N	Y
	N249	N	Y
	F250	N	Y
	D251	Y	Y
	E252	N	Y
	D266	N	Y
	E270	N	Y
	I399	N	Y
	D451	Y	Y
	W452	Y	Y
	D454	Y	Y
	D455	Y	Y
	E457	Y	Y
PGK1	M174	Y	Y
	V175	Y	Y
	E200	N	Y
	N201	N	Y
	P202	N	Y
	R204	N	Y
	I234	N	Y
	M238	N	Y
	A239	N	Y
	E247	N	Y
	D253	N	Y
	E303	N	Y
	I305	N	Y
	P306	N	Y
	A307	N	Y
	W309	Y	Y
	Q310	N	Y
	D313	Y	Y
	I332	N	Y
	V333	N	Y

Properties of Peptides Identified by MudPIT

Wolters et al. [13] demonstrated that MudPIT has a high dynamic range with the ability to ID low abundant peptides. To determine whether the IDs from MudPIT are lower abundance than those from 1D-DDA, the intensities of the identified peptides were analyzed. Figure 2a compares the intensity of peptides identified from MudPIT and 1D-DDA. The average intensity of the peptides identified by both methods is similar. However, the minimum intensity of peptides identified by MudPIT is lower than for 1D-DDA. A histogram of frequency of identifications of peptides at varying intensities further demonstrates this (Figure 2b). Since the MudPIT method has more overall identifications, the histogram has been normalized to show the percent of total peptides. For both MudPIT and 1D-DDA, the highest number of identifications were from peptides with intensities in the range of 1.00E + 06 (1E + 06–9E + 06), followed by intensities in the range 1.00E + 05 (1E + 05–9E + 05) and 1.00E + 07 (1E + 07–9E + 07). At the lowest intensity bin, 1.00e + 04 (1E + 04–9E + 04), MudPIT facilitated detection of three times as many peptides, 90 (1%) and 25 (0.4%) for

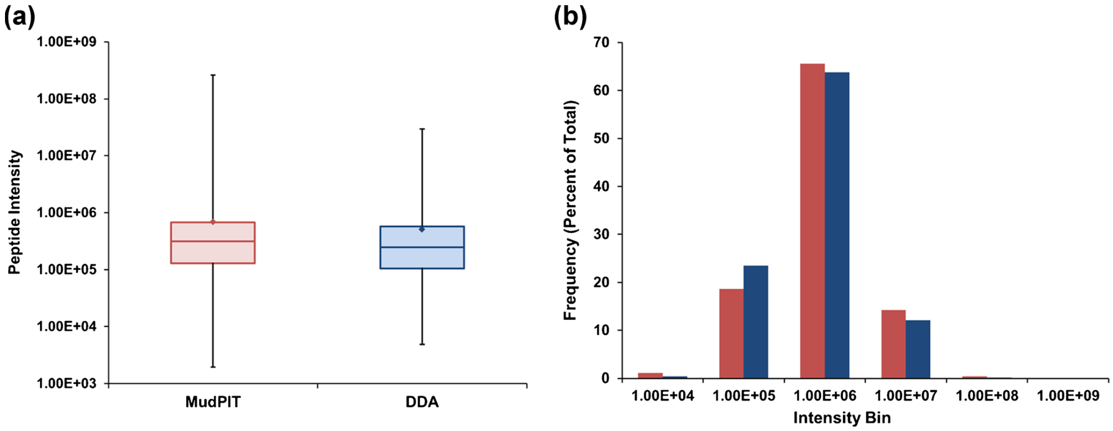


Figure 2. Distribution of the intensities of PSMs identified by MudPIT (red) and 1D-DDA (blue). **(a)** The spread of intensities is demonstrated in the box-and-whisker plot with the box lines marking the upper median and lower quartiles, and the whiskers marking the complete range. **(b)** The frequency of the distributions of intensities is displayed in a histogram

MudPIT and 1D-DDA, respectively. This increase in lower intensity identifications is even more significant for oxidized peptides where 55 and 10 oxidatively modified peptides from MudPIT and 1D-DDA were identified, respectively.

Comparing MudPIT identifications to yeast lysate protein abundance further demonstrates that the method can aid in identifying low abundance proteins. As mentioned previously, PGK1 is highly abundant in *S. cerevisiae* with an estimated abundance of 21,000 parts per million (ppm) [22]. The sequence coverage for this protein is 75%. The ATP-dependent transporter protein YER036C is also identified by MudPIT with 25% sequence coverage. This protein has an abundance of 743 ppm, 29-fold lower than PGK11 indicating the dynamic range of the MudPIT method.

MudPIT as a Method for Megadalton Protein Complexes

A major obstacle in oxidative labeling experiments is the ability to obtain residue level oxidation on large, macromolecular protein complexes. Given that the surface area to volume ratio decreases as a particle increases in size, it stands to reason that the proportion of oxidized species present when analyzing a MDa sized complex would also decrease, making the likelihood of detecting modifications even more difficult.

To demonstrate the power of using MudPIT analysis in oxidative footprinting experiments, residue level oxidation was calculated on a yeast 80S ribosome, which has a published structure (PDB ID 4V6I) [23]. Ribosomes are cellular organelles, consisting of both protein and RNA, involved in protein assembly. The protein component of the structure is assembled in two subunits, 40S and 60S, and contains a total of 70 known proteins. We identified 52 of the 70 proteins in the MudPIT samples, with sequence coverage values ranging from 5% to 80% (data not shown). A total of 86 residues were identified as oxidized and mapped to the crystal structure for a visual representation (Figure 3). Since RNase was not added to the sample at any time to remove the RNA, the structure is presented with the RNA present. The mapping of oxidized residues onto a surface representation of the crystal structure demonstrates that many solvent-accessible residues are oxidized.

To further investigate the correlation between residue oxidation and solvent accessibility, the extent of oxidation of residues identified by MudPIT was compared with solvent accessibility surface area (SASA) calculations. Since FPOP was performed on yeast lysate where various proteins could be interacting, we had to consider certain variables prior to the comparison. While a binary interactome of yeast has been published [24], it is unlikely that every interaction with this complex has been documented. With this in mind, it seemed unlikely that a comparison of SASA to oxidation over the complete complex would yield any reliable assessment of the method. As a consequence, we chose to do this comparison on a single protein within the complex. The SASA was determined on an asymmetric unit of the 40s ribosome (pdb: 31ZB). A plot that correlates the extent of oxidation compared to the residue SASA, demonstrates a good correlation between the two parameters (Figure 4). The data fits well to a linear fit with an R^2 of 0.7. There is a possibility that protein–protein interactions are occurring that are not taken into account in the SASA calculations, which could explain why the R^2 value is not higher.

Discussion

An advantage of using protein footprinting coupled with mass spectrometry for protein structural analysis is the ability to study large protein complexes. Analysis of these complexes

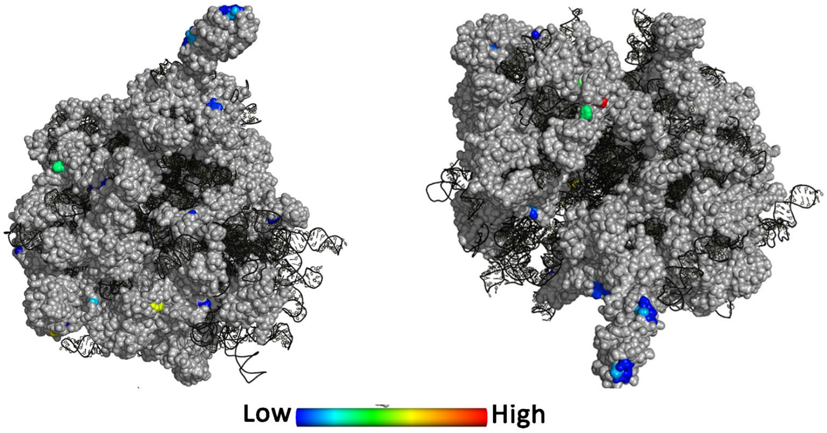


Figure 3. Two perspectives of the structural location of MudPIT determined FPOP oxidation levels mapped to a yeast 80 s ribosomal crystal structure, 4V6I [23]. The lowest oxidation levels are in blue going to the highest in red

is often hampered with other structural tools such as X-ray crystallography and NMR. Although MS analysis has the capability for analysis of complex protein systems, the nature of data-dependent acquisition limits the number of identifications achieved in analysis. Since DDA analysis focuses on the highest abundant peptides at a given time, it is often difficult to identify low abundant peptides with one-dimensional chromatography. This provides a challenge for oxidative labeling where it is advantageous to limit the levels of oxidation; thus, many oxidized peptides are of low abundance. Therefore, the ability to carry out oxidative labeling on large protein complexes hinges upon the capability to identify low abundant peptides.

The application of two-dimensional MudPIT chromatography to an oxidatively modified yeast lysate sample increased the number of identified proteins and peptides over one-dimensional chromatography. Yeast lysate contains thousands of proteins and is indicative of a complex system. The increase in identification is most significant for oxidatively modified peptides where an almost 3-fold increase in identifications is observed (Figure 1c). The higher abundance of identifications for oxidized peptides provides more detailed information on the proteins being analyzed. When investigating individual proteins, the benefit of MudPIT is further revealed. For both pyruvate kinase 1 and phosphoglycerate kinase, MudPIT identifies 5- and 2.6-fold higher numbers of oxidatively modified residues than 1D-DDA. There were peptides identified by 1D-DDA that were not observed with the MudPIT method, however. To gain as complete a coverage as possible, it may be necessary to perform 1D-DDA and MudPIT in tandem.

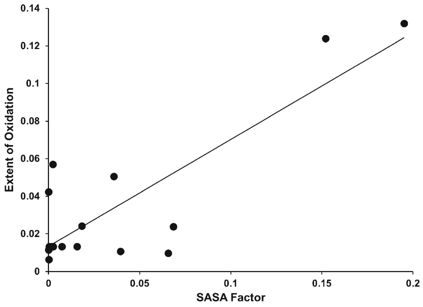


Figure 4. Extent of oxidation on MudPIT identified residues versus the calculated SASA factor for ASC1, chain a of the 3IBZ portion of the complete ribosomal structure, illustrating the linear relationship between them

Examining the intensity of peptides identified by MudPIT indicates this method is detecting lower abundant proteins. However, intensity alone does not account for the increased number of peptides identified by MudPIT. Another factor that may influence the number of IDs is ionization efficiency. Co-elution of peptides that compete for efficient ionization could lead to suppression of some peptides by higher abundant peptides. These suppressed peptides may be of lower abundance than their co-elution partners but are not low enough to be in the 1.00E + 04 intensity range. Two-dimensional chromatography could lead to better separation and reduction in co-elution and ionization suppression.

The identification of 52 of the 70 proteins in the ribosome complex demonstrates that coupling FPOP with MudPIT would be effective for studying large complexes in lysates. However, this approach could likely be improved by further enriching the protein complex with methods such as tandem affinity purification. Comparing extent of oxidation of residues to SASA calculations established a good correlation between the data. Since MudPIT analysis occurs over a longer timescale than one-dimensional chromatography, there is an opportunity for spurious oxidation. Correlation of oxidative modification levels with solvent accessibility demonstrates that the sample is not adversely affected by the long MudPIT analysis.

The ability to obtain greater sequence coverage for oxidatively modified peptides increases the efficacy of FPOP for megadalton complexes. In order to obtain structural information on proteins using oxidative labeling, it is imperative to have good sequence coverage of your oxidative modified peptides. The data presented here demonstrates that MudPIT can provide this increased sequence coverage.

Conclusions

A hallmark of scientific progress is the unceasing march of new technological frontiers and solutions. This holds true in the field of structural mass spectrometry. In order for the use and application of hydroxyl radical-mediated covalent labeling to continue to expand, we must look for new approaches in analysis. In this work, we have demonstrated the use of MudPIT in conjunction with FPOP as a means for increased detection of modified species, and expansion of protein footprinting for complex systems.

Acknowledgments

This work was funded in part by an Indiana University DRIVE grant and a GAANN fellowship for A.R. The authors thank Amber Mosley for helpful discussions.

References

1. Wang, L., Chance, M.R.: Structural Mass spectrometry of proteins using hydroxyl radical based protein footprinting. Anal. Chem. **83**(19), 7234–7241 (2011)
2. Ye, X., O'Neil, P.K., Foster, A.N., Gajda, M.J., Kosinski, J., Kurowski, M.A., Bujnicki, J.M., Friedman, A.M., Bailey-Kellogg, C.: Probabilistic cross-link analysis and experiment planning for high-throughput elucidation of protein structure. Protein Sci. **13**(12), 3298–3313 (2004)
3. Wales, T.E., Engen, J.R.: Hydrogen exchange mass spectrometry for the analysis of protein dynamics. Mass Spectrom. Rev. **25**(1), 158–170 (2006)
4. Mendoza, V.L., Vachet, R.W.: Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. Mass Spectrom. Rev. **28**(5), 785–815 (2009)
5. Maleknia, S.D., Brenowitz, M., Chance, M.R.: Millisecond radiolytic modification of peptides by synchrotron X-rays identified by mass spectrometry. Anal. Chem. **71**(18), 3965–3973 (1999)
6. Xu, G., Chance, M.R.: Radiolytic modification and reactivity of amino acid residues serving as structural probes for protein footprinting. Anal. Chem. **77**(14), 4549–4555 (2005)
7. Watson, C., Janik, I., Zhuang, T., Charvátová, O., Woods, R.J., Sharp, J.S.: Pulsed electron beam water radiolysis for submicrosecond hydroxyl radical protein footprinting. Anal. Chem. **81**(7), 2496–2505 (2009)
8. Hayes, J.J., Kam, L., Tullius, T.D.: Footprinting Protein-DNA Complexes with γ-Rays. In: Lester Packer, A.N.G. (Ed.) Methods in Enzymology, Vol. 186, pp. 545–549. Academic Press, San Diego, California (1990)
9. Hambly, D., Gross, M.: Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale. J. Am. Soc. Mass Spectrom. **16**(12), 2057–2063 (2005)
10. Gau, B.C., Sharp, J.S., Rempel, D.L., Gross, M.L.: Fast photochemical oxidation of protein footprints faster than protein unfolding. Anal. Chem. **81**(16), 6563–6571 (2009)
11. Barends, T.R.M., Foucar, L., Botha, S., Doak, R.B., Shoeman, R.L., Nass, K., Koglin, J.E., Williams, G.J., Boutet, S., Messerschmidt, M., Schlichting, I.: De novo protein crystal structure determination from X-ray free-electron laser data. Nature **505**(7482), 244–247 (2014)
12. Ori, A., Andrés-Pons, A., Beck, M.: The Use of Targeted Proteomics to Determine the Stoichiometry of Large Macromolecular Assemblies. In: Valérie, D. (Ed.) Methods in Cell Biology, vol. Vol. 122, Chap. 6, pp. 117–146. Academic Press, San Diego, California (2014)
13. Wolters, D.A., Washburn, M.P., Yates, J.R.: An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. **73**(23), 5683–5690 (2001)
14. Washburn, M.P., Wolters, D., Yates, J.R.: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotech. **19**(3), 242–247 (2001)
15. McClintock, C.S., Parks, J.M., Bern, M., GhattyVenkataKrishna, P.K., Hettich, R.L.: Comparative informatics analysis to evaluate site-specific protein oxidation in multidimensional LC-MS/MS data. J. Proteome Res. **12**(7), 3307–3316 (2013)
16. Jones, L.M., Sperry, J.B., Carroll, J.A., Gross, M.L.: Fast photochemical oxidation of proteins for epitope mapping. Anal. Chem. **83**(20), 7657–7661 (2011)
17. ThermoScientific: TECH TIP # 49 Acetone precipitation of proteins. In: Pierce Protein Biology Products Tech Tips, Vol. 2014. Thermo Scientific (2009)
18. Florens, L., Washburn, M.: Proteomic Analysis by Multidimensional Protein Identification Technology. In: New and Emerging Proteomic Techniques. Vol. 328. Methods in Molecular Biology™, pp. 159–175. Humana Press, Totowa, New Jersey (2006)
19. Taylor, P., Nielsen, P.A., Trelle, M.B., Hørning, O.B., Andersen, M.B., Vorm, O., Moran, M.F., Kislinger, T.: Automated 2D peptide separation on a 1D nano-LC-MS system. J. Proteome Res. **8**(3), 1610–1616 (2009)
20. Jurica, M.S., Mesecar, A., Heath, P.J., Shi, W., Nowak, T., Stoddard, B.L.: The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. Structure **6**(2), 195–210 (1998)
21. Watson, H.C., Walker, N.P., Shaw, P.J., Bryant, T.N., Wendell, P.L., Fothergill, L.A., Perkins, R.E., Conroy, S.C., Dobson, M.J., Tuite, M.F.: Sequence and structure of yeast phosphoglycerate kinase. EMBO J. **1**(12), 1635–1640 (1982)
22. de Godoy, L.M., Olsen, J.V., Cox, J., Nielsen, M.L., Hubner, N.C., Frohlich, F., Walther, T.C., Mann, M.: Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature **455**(7217), 1251–1254 (2008)
23. Armache, J.-P., Jarasch, A., Anger, A.M., Villa, E., Becker, T., Bhushan, S., Jossinet, F., Habeck, M., Dindar, G., Franckenberg, S., Marquez, V., Mielke, T., Thomm, M., Berninghausen, O., Beatrix, B., Söding, J., Westhof, E., Wilson, D.N., Beckmann, R.: Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5-Å resolution. Proc. Natl. Acad. Sci. U. S. A. **107**(46), 19748–19753 (2010)
24. Yu, H., Braun, P., Yıldırım, M.A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.-F., Dricot, A., Vazquez, A., Murray, R.R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A.-S., Motyl, A., Hudson, M.E., Park, J., Xin, X., Cusick, M.E., Moore, T., Boone, C., Snyder, M., Roth, F.P., Barabási, A.-L., Tavernier, J., Hill, D.E., Vidal, M.: High-quality binary protein interaction map of the yeast interactome network. Science **322**(5898), 104–110 (2008)