1-1-2019

Implementing mass spectrometry for the structural and compositional analysis of proteins

Mohammad Riaz

IMPLEMENTING MASS SPECTROMETRY FOR THE STRUCTURAL AND COMPOSITIONAL ANALYSIS OF PROTEINS

A Thesis

presented in partial fulfillment of requirement

for the degree of Master in Pharmaceutical Sciences

in the BioMolecular Sciences Graduate Program

The University of Mississippi

By

MOHAMMAD RIAZ

August 2019

ABSTRACT

To understand the functions of proteins at a molecular level, it is often necessary to determine their three-dimensional structure. Fast Photochemical Oxidation of Protein (FPOP) is a hydroxyl-radical-based protein footprinting (HRPF) technique that utilizes a pulsed KrF laser (248 nm) to trigger photolysis of hydrogen peroxide to produce hydroxyl radicals, which subsequently modify the solvent exposed surface area of proteins. However, this technique has some disadvantages: being time-consuming, especially when dealing with a large sample size, adjustment of flow rate and difficulty with membrane protein oxidation. To address these issues, we developed a platform to perform FPOP in microtiter plates instead of the traditional capillary set up. To ensure reliability and reproducibility of microtiter FPOP and evaluate microtiter FPOP against traditional flow FPOP, we used three systems: adenine-based hydroxyl radical dosimetry; oxidation of the model peptide [Glu]1-Fibrinopeptide B (GluB); and HRPF analysis of the model protein myoglobin. We demonstrated that microtiter based FPOP can provide comparable oxidation of model peptide and model protein as compared to traditional capillary FPOP. Automation of the system substantially reduces experimental time and minimize human errors.

Lectins are sugar-binding proteins that perform various biological functions such as cellular recognition, attachments, etc. Pulses, the dried edible seeds of certain plants, are great sources of lectins, and a proteomics approach directed at lectins identification would bring substantial improvement in the lectin identification process. Lectins from nine different plant species were

analyzed through SDS-PAGE and proteomics analysis. After LC-MS/MS analysis, proteins were identified with protein database searches using Uniprot. Functionally uncharacterized proteins were identified with database searches, were annotated with the Pfam and NCBI-CCD databases. In-vitro pharmacological screening will be carried out to assess the pharmacological effects of these lectins.

DEDICATION

This work is dedicated to my family and friends who made lots of contributions for me in pursuing higher studies in USA.

LIST OF ABBREVIATIONS AND SYMBOLS

Aqueous Two-Phase System (ATPS)

Bovine Serum Albumin (BSA)

Carbohydrate Recognition Domain (CRD)

Circular Dichroism (CD)

Collision-Induced Dissociation (CID)

Differential Scanning Calorimetry (DSC)

Dithiothreitol (DTT)

Electron Transfer Dissociation (ETD)

False Discovery Rate (FDR)

Fast Photochemical Oxidation of Proteins (FPOP)

Fourier Transform Infrared Spectroscopy (FTIR)

Granulocyte Colony-Stimulating Factor (GCSF)

[Glu1]-Fibrinopeptide B (Glub)

High Pressure Liquid Chromatography (HPLC)

Hydrogen-Deuterium Exchange (HDX)

Hydroxyl Radical Protein Footprinting (HRPF)

Ion Mobility Spectroscopy (IMS)

Liquid Chromatography–Mass Spectrometry (LC-MS)

Mass Spectrometry (MS)

Nuclear Magnetic Resonance (NMR)

Polymerase Chain Reaction (PCR)

Reactive Oxygen Species (ROS)

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Vascular Endothelial Growth Factor (VEGF)

ACKNOWLEDGEMENT

I want to thank Dr. Joshua Sharp for giving me the opportunity to be a part of Department of BioMolecular Sciences at Ole Miss and introducing me to the realm of mass spectrometry. He is an excellent mentor and guided me from day one. Through-out my time here, he was supportive and caring, I am forever grateful to Dr. Sharp.

I want to thank my committee members, Dr. Kristie Willett and Dr. John Rimoldi for their valuable comments and guidance for my thesis.

I specially thank Dr. Sandeep Misra and Dr. Surrendar Tadi who relentlessly helped me during my time in OleMiss. They are always ready to help, teach the basic and help me in troubleshooting the problems. They are very humble, friendly and hardworking individuals.

I appreciate Dr. Charles Mobley, Dr. Quntao Liang, Hao Liu, Niloofar Abolhasani Khaje, Selina Cheng, Addison Roush and all of the undergraduates for all the help, friendship and technical support.

Finally, I am grateful to my family and friends who have provided immense support in my difficult times. I would like to thank National Institute of General Medical Sciences (R01GM127267), the NIH Center for Research Excellence in Natural Products Neuroscience (P20GM104932), and National Center for Natural Products Research for research grant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Introduction

Identification and determination of protein structures remain a central part of life science research (Domon & Aebersold, 2006). Proteins can form a myriad of shapes that correspond to functional domains of biological significance due to the presence of noncovalent interactions (Wang & Chance, 2011). The number of diseases associated with protein misfolding are grouped together under the name of protein conformational disorders (Soto, 2001). Although proteins' conformational changes are vital for normal physiological functions, an error in the process could be lethal (Ellis & Pinheiro, 2002). Moreover, the three dimensional structures of protein therapeutics are a key component in understanding the molecular structure, biological function and pharmacological function of a protein (Li & Chen, 2019). Deviation from the native structure can have significant impact on quality, stability, safety and efficacy of protein therapeutics with increased risk of immunogenicity and loss of biological function.

Mass spectrometry (MS) is a powerful analytical technique employed to assess the structure of proteins, conformational changes in the presence and/or absence of ligands, and proteins-protein/protein-ligand interactions. Footprinting is a powerful method for mapping the interactions between the domains of macromolecules and mapping conformational changes due to ligand binding. During the footprinting experiments, the solvent accessible surface of the proteins are examined through chemical modification or cleavage with the use of specific

chemicals or enzymes (Takamoto & Chance, 2006). Footprinting based approaches were originally developed to study nucleic acids structure and dynamics (Wang & Chance, 2017). Three different types of approaches are available for MS-based structural analysis (Wang & Chance, 2011): covalent labeling of macromolecules, chemical cross-linking, and hydrogen–deuterium exchange (HDX). Of the three approaches, covalent modification/labeling represents a robust method to probe protein and nucleic acid macromolecular structures in solution (Brenowitz, Chance, Dhavan, & Takamoto, 2002). Hydroxyl radical-based approaches are highly popular due to the high reactivity of hydroxyl radicals that react with most of the amino acids with different reaction kinetics and comparable size to a water molecule (Kiselar & Chance, 2010).

With the instrumental and software advancement of mass spectrometry as applied to probe proteins, a remarkable progress took place in the MS-based protein footprinting. In the early 1990s, chemical reagent-based studies were carried out to probe the site of modification of proteins based on mass shift (Hanai & Wang, 1994; Steiner, Albaugh, Fenselau, Murphy, & Vestling, 1991). Meanwhile, the HDX method coupled with MS was used to probe intact proteins and conformational changes within the proteins (Katta, Chait, & Carr, 1991; Z. Zhang & Smith, 1993). Later, hydroxyl radical-based footprinting technology was introduced to probe the solvent accessible sites of proteins (Maleknia, Chance, & Downard, 1999).

Hydroxyl Radical Protein Footprinting

Hydroxyl Radical Protein Footprinting (HRPF) coupled with mass spectrometry is used to comprehensively study protein structure and conformational changes of proteins (Huang, Ravikumar, Chance, & Yang, 2015). Folding of a one-dimensional amino acid sequence into a

three-dimensional structure creates protein topography. In this folded conformation, some residues are buried, whereas others are more exposed and solvent-accessible on the protein surface. In HRPF, hydroxyl radicals are used to monitor the solvent accessibility of the proteins, as shown in **Figure 1**. Due to the hydroxyl radical's similarity in size to water molecules, hydroxyl radicals can easily oxidize side chains of surface reactive amino acid residues in aqueous solution. Moreover, hydroxyl radicals react with protein functional groups depending on their specific chemistry and solvent accessibility (Wang & Chance, 2011). The hydroxyl radicals react with amino acids with inherent reactivity while backbone C-α and amide bonds are much less reactive targets of radical attack compared to the preferred side chain sites. The relative reactivity of the 20 side chains follows the order: Cys > Met > Trp > Tyr > Phe > Cystine > His > Leu ∼ Ile > Arg ∼ Lys∼ Val > Ser ∼ Thr ∼ Pro > Gln ∼ Glu > Asp ∼ Asn > Ala > Gly (Takamoto & Chance, 2006). Hydroxyl radicals (•OH) are generated within microseconds to minutes depending on the methods used to generate the radicals, using photochemical or discharge sources from water in aqueous solution (Gupta, Celestre, Petzold, Chance, & Ralston, 2014). Moreover, the time frame for protein folding/unfolding kinetics ranges from microseconds to milliseconds (Dai, Gardner, & Fitzgerald, 2005). With the oxidation of amino acids, the conformation of proteins changes depending on the states of oxidation, and further oxidation might provide false structural information. While some small proteins can undergo global unfolding in microseconds (Naganathan & Muñoz, 2005), changes in secondary structures within the proteins are even faster than this timescale, and a higher lifespan of radicals lead to false positive data (Hambly & Gross, 2005). An ideal compound should label the proteins below the proteins folding kinetics timescale. In laser induced photochemical oxidation of proteins, in the presence of radical

scavengers, the active radical is removed in 1 μs, which is faster than the overall protein unfolding timescale (Gau, Sharp, Rempel, & Gross, 2009; Hambly & Gross, 2005). The controlled radical-based methods differ from the naturally occurring oxidation from reactive oxygen species (ROS) in the proteins because the experimental procedure is tuned to control the exposure time minimizing the double oxidation of proteins. There are several ways to produce hydroxyl radicals. Each method has its own shortcomings, and previous studies of lysozyme footprinting with hydroxyl radicals generated from two different methods provide similar oxidation (Maleknia Chance, & Downard, 1999; Maleknia, Ralston, Brenowitz, Downard, & Chance, 2001). There are six tryptophan residues in lysozyme, and the oxidation of these residues, induced with hydroxyl radicals generated with two methods, were measured. Out of six tryptophans in the lysozymes, two tryptophans at positions W62 and W123 showed oxidation in both methods, and the level of oxidation is closely related. Level of oxidation also demonstrated that W62 is 1.4 times more solvent accessible than W123. Another study directed to map the surface of the protein solvent accessible sites of the myoglobin demonstrated that protein oxidation by chemically generated radicals yields oxidation comparable to the NMR based study (Sharp, Becker, & Hettich, 2003).

Apomyoglobin, a monomeric protein with eight helices, was used to study the folding of proteins using the radical footprinting method (Maleknia and Downard 2001). Radicals generated from X-ray radiolysis of water were used after treating the proteins with varying concentrations of urea. Apomyoglobin is an intermediate product in the biosynthesis of myoglobin and possesses similar structural features with the heme-containing native state. LC-MS analysis of apomyoglobin after radical exposure provided information regarding the folding pattern of the protein. Out of eight helices, analysis of three of the helices showed that helix A and B/C follow

the cooperative global unfolding pattern where helix G showed local unfolding behavior, as shown in **Figure 2** (Maleknia & Downard, 2001). Previous NMR and Raman studies yielded a similar result, indicating a complex pattern of helix G unfolding showing some of helix G undergoes local unfolding where others helices are resistant to local unfolding. Radical footprinting approaches have been employed to study protein stability, protein-protein interactions, and binding site determinations (Wang & Chance, 2017).

Fast Photochemical Oxidation of Proteins (FPOP)

There are multiples ways to generate hydroxyl radicals, including (but not limited to) the Fenton reaction (Sharp, Becker, & Hettich, 2003), UV irradiation of hydrogen peroxide (Sharp, Becker, & Hettich, 2004) and the radiolysis of water with a high-energy X-ray synchrotron beam (Kiselar, Janmey, Almo, & Chance, 2003a, 2003b). Due to the characteristic folding and unfolding behavior of proteins, it is desirable to have the hydroxyl radical's lifetime very short, and in fast photochemical oxidation of proteins (FPOP) method oxidation takes place faster than protein unfolding (Gau, Sharp, Rempel, & Gross, 2009).

FPOP is a protein footprinting technology where hydroxyl radicals are generated from hydrogen per oxide through pulsed laser photolysis in a flow system. The FPOP method limits the sample exposure time to hydroxyl radicals to a sub-millisecond timescale by scavenging the excess radicals and exposing the samples to the laser system in a flow system (Zhang, Cheng, Rempel, & Gross, 2018). Gross and coworkers are credited for introducing the FPOP method to probe protein structure and conformational changes (Hambly & Gross, 2005). The major components of the FPOP instrumentations include: i) a pulsed laser source, ii) a focusing lens,

and iii) a syringe pump connected to a fused silica capillary tubing to form a flow cell for illumination, as depicted in **Figure 3**. The pulsed KrF excimer laser at 248nm photolyzes hydrogen peroxide to generate hydroxyl radicals which subsequently label the solvent accessible amino acids. In the FPOP method, generally 15-20% protein samples remained unexposed to the radicals; this volume is referred to as the exclusion volume, which minimizes the double exposure of the samples to the radicals. During the experiment, the sample of interest is mixed with a radical dosimeter and a radical scavenger in a suitable buffer. Just before the sample exposure to the laser hydrogen peroxide is added. Using a syringe pump, samples are introduced to capillary and exposed to a pulsed laser, which photolyzes the hydrogen peroxide to generate hydroxyl radicals. The spectrophotometer is used to measure the hydroxyl radical generated in real time (Sharp, Misra, Persoff, Egan, & Weinberger, 2018), and subsequently the samples are quenched with catalase and methionine amide to quench extra radicals or peroxide. The proteins are subsequently denatured and digested using the proteolytic enzyme trypsin and analyzed with LC-MS. It is absolutely vital to maintain the calculated flow rate during the experiment to minimize the double oxidation of protein and to minimize sample-to-sample variation. However, it was recently shown that injection of higher amount of FPOP samples greatly reduce the sample-to- sample variation and improve the data quality (Abolhasani et al., 2018). Moreover, the capillary needs to be cleaned with water between sample injections to minimize cross contamination. To prevent the breakdown of the capillary, the capillary needs to be changed after every 2-3 samples. Otherwise it might be broken during the experiment, resulting in the loss of samples. Any kind of vibration in the capillary platform during the experiment might change the laser exposure to the sample, leading to poor results. The position of the lens is crucial

to focus the laser beam to the specific small regions of capillary, and attention must be given to examine whether the lens is focusing all the laser beam in the process.

FPOP was previously employed to determine the binding site of the antigen-antibody complex. The binding sites of VEGF to a monoclonal antibody were determined by the FPOP method (Zhang, Wecksler, Molina, Deperalta, & Gross, 2017). Peptide level oxidation analysis after an LC-MS/MS experiment revealed three peptides showed significant difference in bound form compared to ligand free form as shown in **Figure 4A** and subsequent residue level analysis confirms the binding sites of the VEGF to the antibody, as shown in **Figure 4B**. The binding sites of the Robo 1-heparin were determined through the FPOP footprinting method, and two binding sites were identified. Out of two sites, one site was newly identified and confirmed by X-ray crystallography and mutagenesis techniques (Li et al., 2015). The heme binding site of myoglobin (Hambly & Gross, 2005) and Ca/peptides binding sites of calmodulin (Zhang, Gau, Jones, Vidavsky, & Gross, 2011) were determined using FPOP methods. FPOP can also be used to study the aggregation of proteins in different pathological conditions, especially Alzheimer's or Parkinson’s. Li and coworkers employed FPOP to study the oligomerization of amyloid beta protein associated with Alzheimer's and showed that solvent accessibility changes during different stages of protein aggregation (Li, Rempel, & Gross, 2016). Another study examined an isoform of oligomeric apolipoprotein, a protein that regulates lipid metabolism and controls lipid redistribution in tissue and cells. This study showed that isoform 4 of the protein ApoE, an isoform associated with Alzheimer's disease, significantly differs in structure at the M108 position from the other two isoforms, ApoE2 and ApoE3, although the rest of the structure is highly similar (Gau, Garai, Frieden, & Gross, 2011). Recently, footprinting of solvent accessible sites of living

cells has been introduced. **Figure 5** demonstrates that FPOP can modify different proteins in various subcellular locations in living cells (Chea & Jones, 2018a; Espino, Mali, & Jones, 2015).

With the development and USFDA approval of new protein-based therapeutics, MS-based structural interrogation techniques have bigger roles to play to ensure the safety and efficacy of these products. Improper conformation in the protein-based therapeutics results in loss of efficacy and can lead to toxicity issues (Manning, Chou, Murphy, Payne, & Katayama, 2010). FPOP was previously employed to study conformations of Neupogen®, a patented granulocyte colony-stimulating factor (GCSF), with several expired samples of recombinant GCSF, as well as heat-treated Neupogen® (Watson & Sharp, 2012). The results revealed the conformational differences among differently treated samples of Neupogen® (**Figure 6)**. Recently, the FPOP method was used to study membrane proteins in solution. Study of solvent accessibility of the outer membrane protein OmpT in the presence of two surfactants showed one surfactant stabilized the protein more than the other one (Watkinson, Calabrese, Ault, Radford, & Ashcroft, 2017). These studies indicate the widespread application of FPOP and the major roles it might play in the biologics and biosimilar characterization as well as to ensure the safety and efficacy of these products. This is a much faster approach to analyze the protein based therapeutics compared to the traditional NMR or X-ray crystallography-based approaches.

Current capillary-based FPOP has been extensively used with great success to study the protein higher order structure; however, there are some challenges associated with the flow-based capillary FPOP setup. Exclusion volume, the amount of sample unirradiated with the laser, needs to be considered to control the flow rate. Moreover, the FPOP experiment is time consuming and involves lots of physical work, which increases the chance of unwanted human

errors. Furthermore, living cell-based FPOP holds great promises to probe the protein in the native conformations, and microtiter-based FPOP would be a great addition to perform the living cell FPOP experiment. A practically implementable FPOP instrumental platform would address these issues and bring a positive change to FPOP-based research. An ideal instrumental set-up would be simple to design, inexpensive, easy to handle, and minimize all the difficulties associated with capillary-based FPOP experiment. Such an innovative set up with automation technology would attract scientists or pharmaceutical companies to integrate and implement FPOP-based technology to probe the protein structure.

Identification of lectins from pulse proteomes

Proteomics is a rapidly evolving field within the omics sphere and it, by definition, is the study of the proteome, including expression, structure, function, interaction and modification of proteins at any stage (Domon & Aebersold, 2006). For MS-based proteomics, the extracted protein samples are digested using one or several enzymes to produce a collection of peptides (Wiśniewski & Mann, 2012). Based on the pattern of peptides detected by mass and MS/MS fragmentation patterns, different bio-informatics tools compare the MS data with existing database records for protein ID. Many bioinformatics tools are currently employed for protein domain and motif analysis, and rapid analysis of protein–protein interaction, as well as data analysis of MS (Aslam, Basit, Nisar, Rasool, & Khurshid, 2017). Various bioinformatics tools in proteomics are advancing the field while computational algorithms, database repositories, and new software platforms continue to be developed (Deutsch, Lam, & Aebersold, 2008; Matthiesen, Azevedo, Amorim, & Carvalho, 2011). In combination with protein and peptide separation methods, mass spectrometry offers the efficient qualitative identification of proteins

in complex mixtures (Ishihama et al., 2008). Despite the exceptional impact of mass spectrometry and peptide separation techniques on proteomics, the identification and quantification of all of the proteins in a biological system are still unmet technical challenges (Bantscheff, Schirle, Sweetman, Rick, & Kuster, 2007). Pulses are nutritious seeds providing many nutritional and physiological benefits to human health and are very rich in protein, carbohydrates, dietary fiber, and other bioactive components (Selvaraj et al. 2019). They are the dried seeds of leguminous plants, commonly dry peas, lentils, and chickpeas. Pulses have very high protein contents (Ofuya and Akhidue 2005), and many of these are carbohydrate binding proteins. The nutritional profiles of pulses are depicted in **Figure 7**.

Lectins are a group of proteins that specifically and reversibly bind to carbohydrates and are present in microorganisms, plants, and animals (Lakhtin, Lakhtin, & Alyoshkin, 2011; Zeng, Andrade, Oliveira, & Sun, 2012). Research associated with lectins have attracted different fields of science and technology due to the diverse application of lectins. Lectins perform various functions and control biological process including cellular adhesion during glycoprotein synthesis and cell attachment and tissue colonization in microorganisms (He et al., 2018). Lectins bind to intestinal carbohydrates and are resistant against proteolytic digestion, which leads to toxicity (Menard, Cerf-Bensussan, & Heyman, 2010). Plant lectins specifically are effective as antifungal agents (Hoff, Brill, & Hirsch, 2009), and lectins play a vital role in the study of monitoring cell-surface oligosaccharides (He, Shi, Walid, Ma, & Xue, 2013). Lectins play a major role in recognition processes of the carbohydrate expression on living cells in response to drugs (Mun et al., 2012). Additionally, lectins have been used in lectin affinity chromatography to identify glycoproteins and possess various biological functions such as anti-tumor (Xia & Ng, 2006), immunomodulatory

(Rubinstein, Ilarregui, Toscano, & Rabinovich, 2004), anti-human immunodeficiency virus (HIV) (Barrientos & Gronenborn, 2005), and anti-insect (Macedo et al., 2003). Previously, a study aimed to identify the lectins from black turtle bean, an important variety of common beans, integrated both proteomics and PCR methods and was able to identify a lectin from this specie (He, Shi, Li, Ma, & Xue, 2015). The identified lectin exhibited a single band of 31 kDa in SDS PAGE analysis, and the sequence homology matching revealed the protein has a high degree of homology with the previously characterized lectins. A mass spectrometry based approach was employed to determine the isoforms of banana lectins (Gnanesh & Surolia, 2018), and three isoforms of lectins were identified. One lectin presents as non-acetylated form where the rest of the two lectins were acetylated. A similar study aimed at determining the different isoforms of *Arachis hypogaea* stem lectin (SL-I) (Agrawal, Kumar, & Das, 2010). With the combination of proteomics, gel electrophoresis and western blot analysis, six isoforms of the lectin were characterized. Lectin from *Amaranthus leucocarpus* syn hypocondriacus (ALL), which has specificity for the N-acetyl-D-galactosamine, was previously characterized using LC-MS based proteomics approach (Hernandez et al., 2001). The amino acid sequence of Cucurbitaceae phloem lectin, a specialized group of chitin-binding proteins, was determined to study the role of tryptophan in its activity with the combination of chitin affinity chromatography and mass spectrometry analysis (Kumar, Mishra, Anantharam, & Surolia, 2015). A combination of ion exchange, chitin affinity, and size exclusion chromatography coupled with LC/-MS/MS was employed for the isolation and identification of a novel lectin from the seeds of *Bauhinia forficata* (Silva et al., 2012).

Lectins perform wide range of functions in plants and animals including cellular recognition, cellular attachment, etc., and considering the effects of lectins on human and animal

health, investigation to identify new lectins and assess the effects of the newly identified lectins holds great promise. There is a shortage in comprehensive proteomic approaches to identify lectins from different pulses. Implementing the LC-MS/MS approach to identify new lectins would eliminate the need to fully purify the extract for identification. Trace amount of lectins can be identified, and annotation of uncharacterized lectins can be made using different database searches. The proteomics approach would be a great addition in lectin identification due to the sensitivity of the mass spectrometer and the continually improving protein database. This would minimize the time required to identify the lectins through different biochemical methods and improve the overall lectins identification process.

Research Aims

Considering the discussion in the FPOP section, it is evident that a high-throughput method will drastically eliminate the overall duration of FPOP experiments. Such innovative methods with the integration of automation will minimize human errors and will widen the FPOP application in different sectors. On the other hands, the integration of proteomics-based approach in the identification of plant lectins holds a great promise due to the high sensitivity of the MS instruments. There are two major aims of this research study

1) Development of a high-throughput hydroxyl radical protein footprinting in a microtiter plate.
2) Identification plant lectins from different pulses with the integration of LC-MS/MS and proteomics approaches.

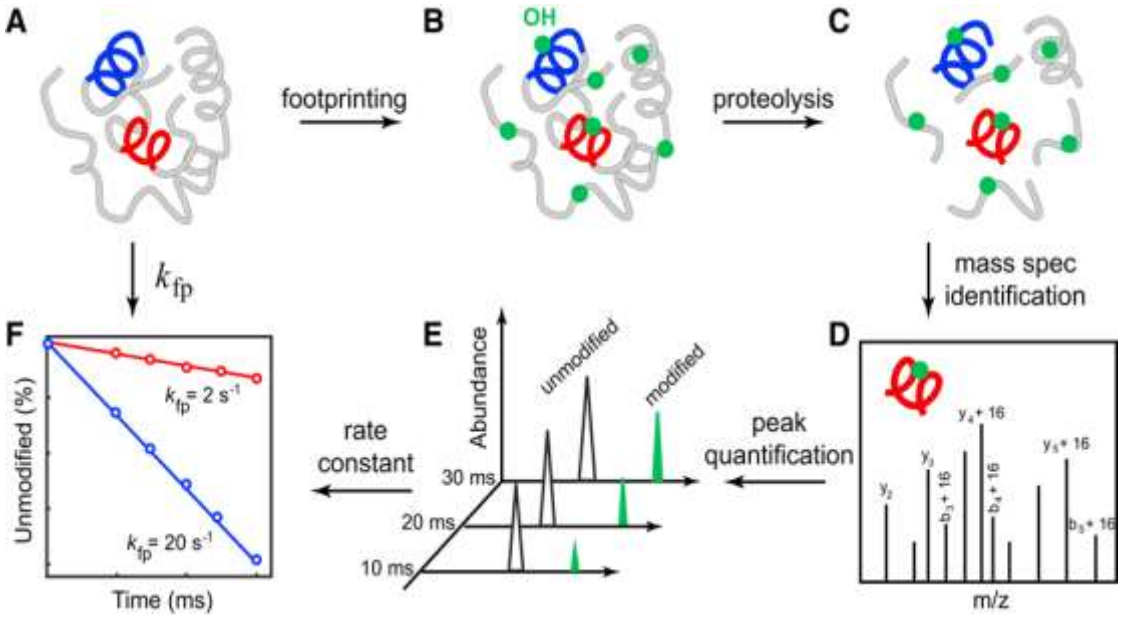
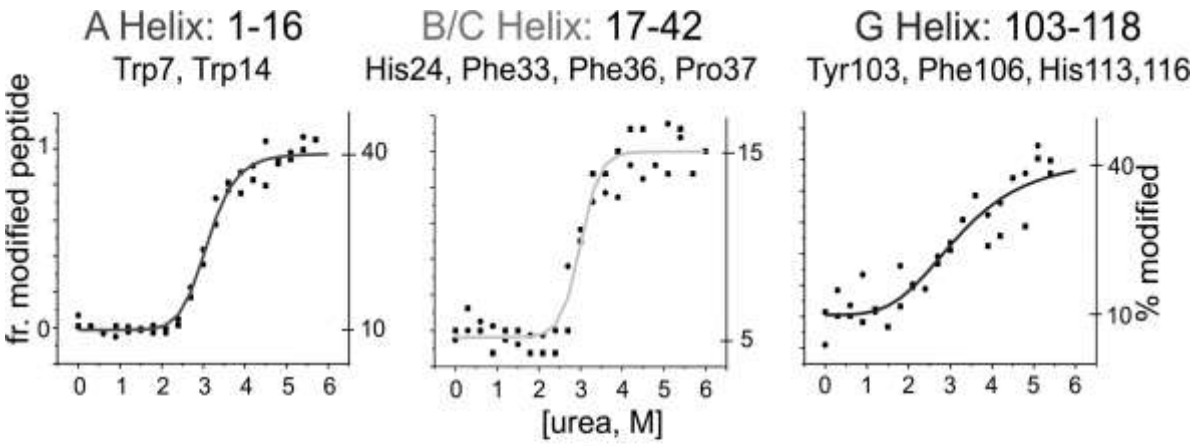


Figure 1: Schematic of RP-MS based footprinting. The protein of interest is labelled with hydroxyl radicals and later process by proteolytic digestion. The digested protein analyzed by LC-MS/MS analysis. Used with permission (Huang, et al. 2015).



Protein				
2.9 ± 0.1	3.1 ± 0.1	3.0 ± 0.1	3.2 ± 0.1	$k_{midpoint}$ (M)
4.7 ± 0.6	5.4 ± 0.7	7.6 ± 1.6	2.5 ± 0.5	ΔG^{o}_{N-U} (kcal/mol)
1.7 ± 0.2	1.7 ± 0.2	2.0 ± 0.2	0.6 ± 0.1	m (kcal/mol/ M)

Figure 2: Plots of the oxidation levels within peptide segments of apomyoglobin comprising helices A, B/C and G as a function of urea concentration. Used with permission (Maleknia & Downard, 2001)

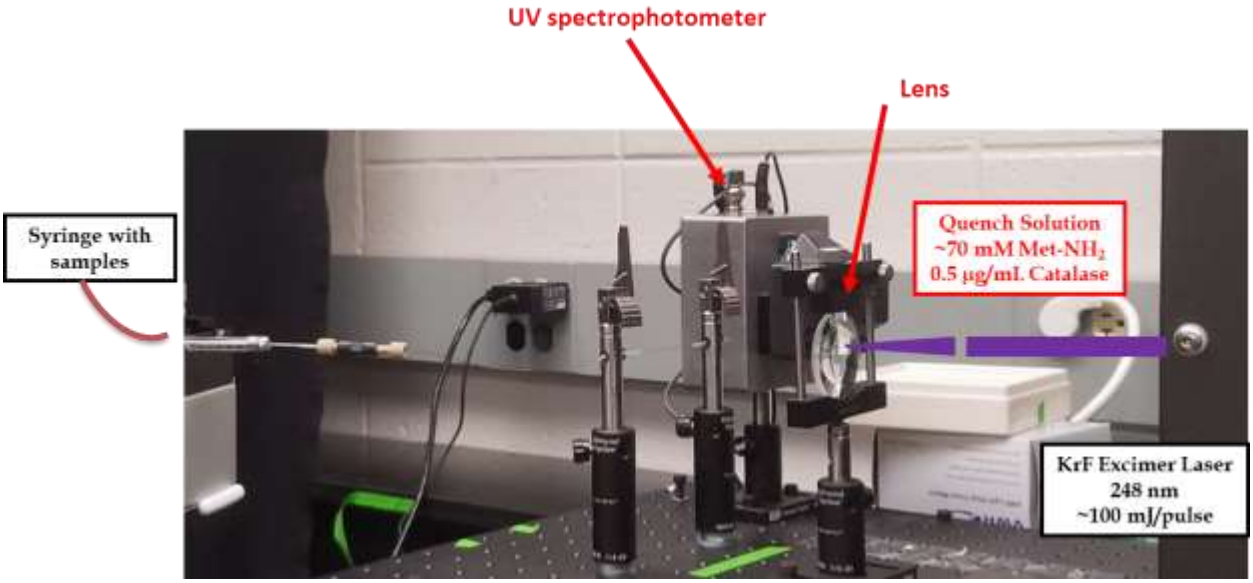


Figure 3: Schematic of FPOP procedure. The samples are flown through capillary with the syringe pump. Laser beam is focused to capillary using lens and UV spectrometer used as online dosimeter. After sample oxidation, samples collected in eppendorf with quenching solution.

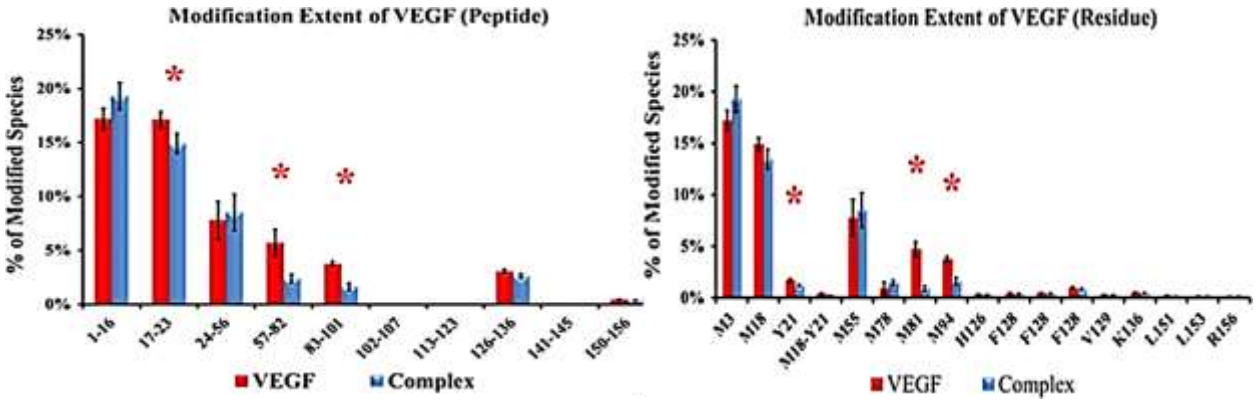


Figure 4A: Peptide level analysis of VEGF indicates there are significant changes of oxidation in three peptides based on the presence of ligands. 4B: Residue level analysis of VEGF shows us possible site of interaction between VEGF and ligands. Three amino acids show significant difference in oxidation with the presence of absence of ligands. Used with permission (Zhang, et al. 2017)

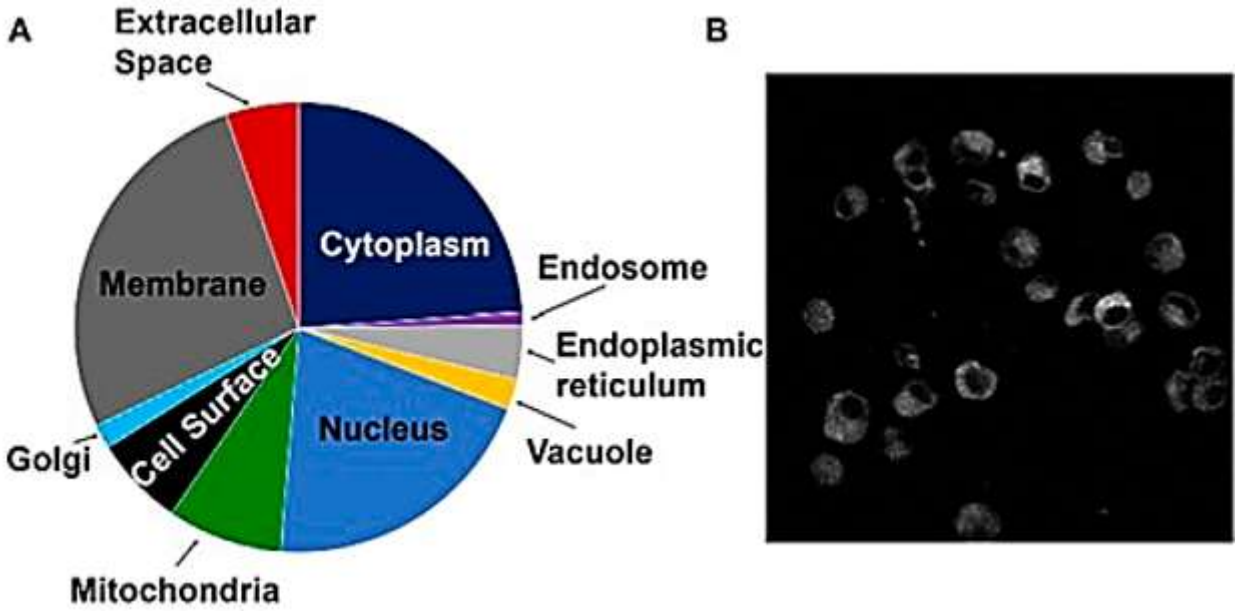


Figure 5: FPOP of living cell. A. Subcellular location of proteins oxidized in FPOP experiment. B. Imaging of oxidized cells indicates cells are still active. Used with permission (Espino, et al. 2015)

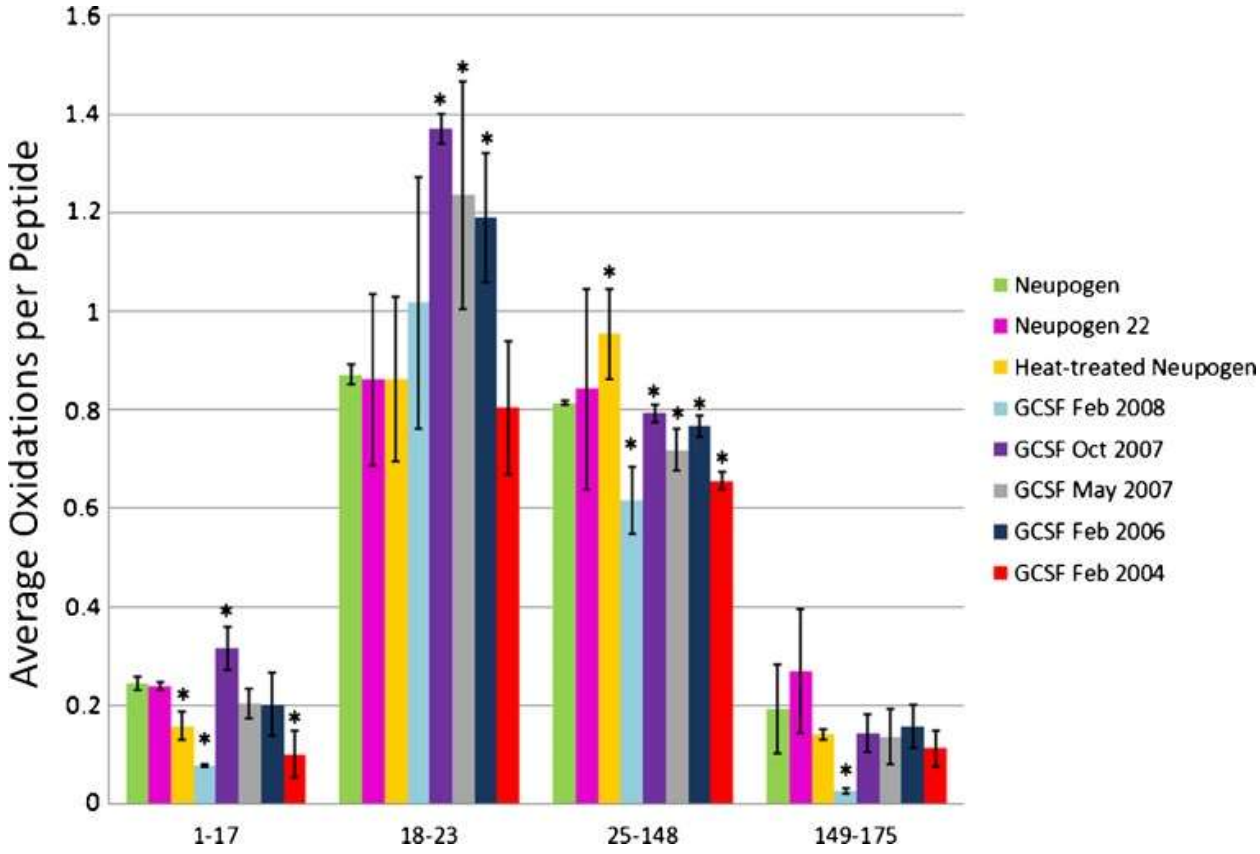


Figure 6: FPOP data for Neupogen therapeutics and Granulocyte Colony-Stimulating Factor (GCSF) in different conditions. Four different peptides of the protein in various conditions are shown here. Asterik in bar indicates statistically significant difference in oxidation compared to Neupogen indicating differences in the protein conformation (n=3). Used with permission (Watson and Sharp 2012)

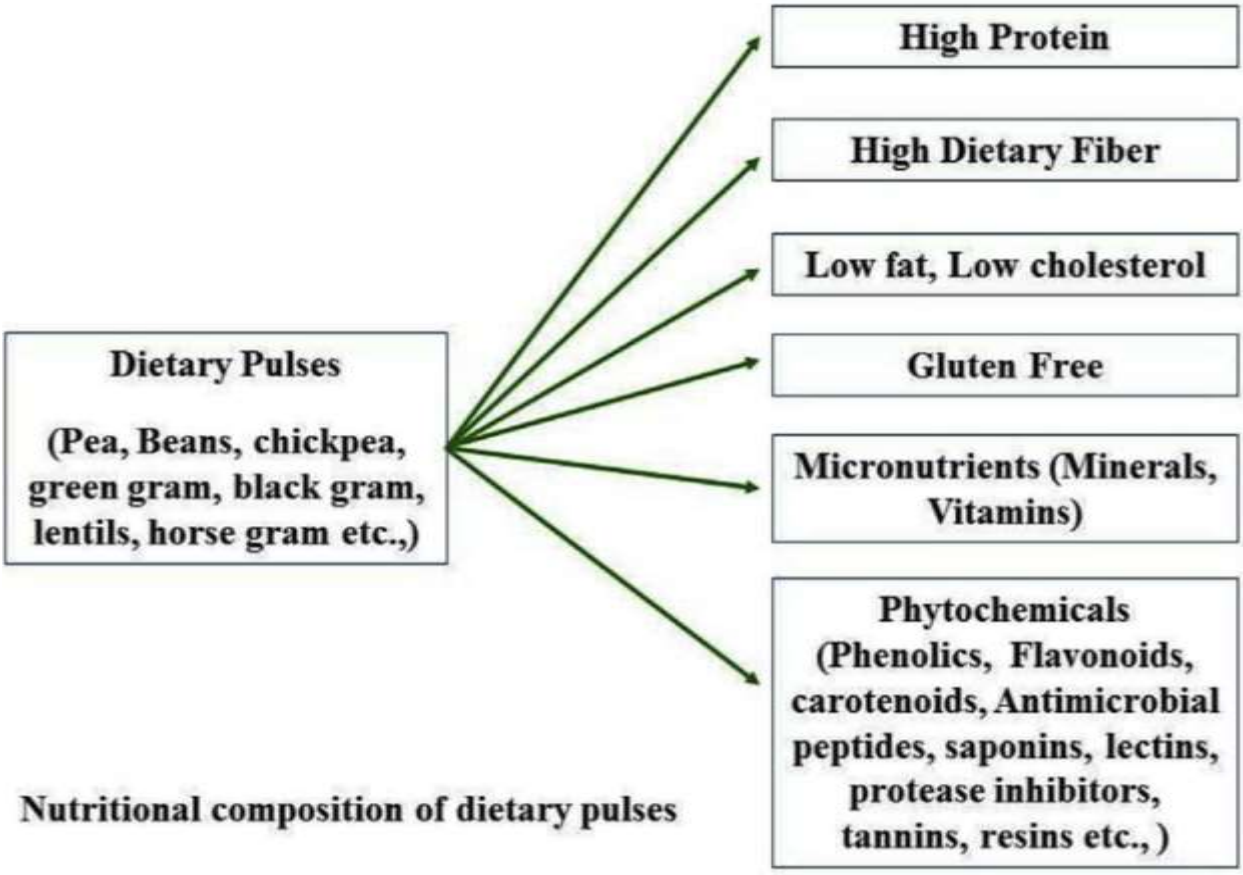


Figure 7: Pictorial representation of the nutritional components present in the pulses. Used with permission (Venkidasamy, et al. 2019)

CHAPTER 2: DEVELOPING HYDROXYL RADICAL PROTEIN FOOTPRINTING IN MICROTITER PLATE

Introduction

Mass spectrometry (MS) is rapidly emerging as an analytical tool in structural biology research and able to provide insights into protein high order structure (Zhang et al., 2018). Many biophysical techniques such as circular dichroism (CD) (Li et al., 2011), Fourier transform infrared spectroscopy (FTIR) (Kong & Yu, 2007), electron microscopy (Plath et al., 2016), and differential scanning calorimetry (DSC) (Durowoju, Bhandal, Hu, Carpick, & Kirkitadze, 2017) have been developed to study the protein three dimensional structure. Compared to the traditional higher resolution techniques such as nuclear magnetic resonance (NMR), and X-ray crystallography, MS is gaining its popularity because MS requires lower sample amounts and assures faster analysis turnaround (Takamoto & Chance, 2006; Xu & Chance, 2007).

Protein footprinting is a way to study the higher order structure of protein and protein interaction within macromolecular assembly. Significant improvement is achieved for MS resolving power, speed, and the application of reversible labeling and irreversible labeling have seen increase, now providing more structure resolution, depth, and accuracy of footprinting (Wang & Chance, 2011). Hydroxyl radical footprinting (HRF) coupled with mass spectrometry is the most widely used irreversible covalent labeling method to characterize protein structure and

dynamics (Wang & Chance, 2011). Complementary to other MS-based methods, such as hydrogen-deuterium exchange (HDX) and ion mobility spectroscopy (IMS), HRPF measures changes in the protein topography using a stable and broadly reactive labeling group (Riaz, Misra, & Sharp, 2018). Several methods such as Fenton chemistry, UV irradiation H_2O_2, radiolysis of water with X-ray synchrotron beam, etc. have been developed to produce a labelling group. UV irradiation requires minute time scale and is not compatible with a protein folding and unfolding time frame. Additionally, X-ray synchrotron maintenance and operation is expensive for most researchers. To minimize the drawbacks related to radicals generation, two groups independently introduced two laser-induced methods for pulsed UV photolysis of hydrogen peroxide: Fast Photochemical Oxidation of Proteins (FPOP) (Hambly and Gross 2005) and batch photolysis (Aye, Low et al. 2005).

First developed by Hambly and Gross in 2005(Hambly & Gross, 2005), FPOP is an approach to footprint proteins by radicals generated by laser photolysis in a flow system. During the experiment, a mixture of the analyte protein, hydrogen peroxide, and a hydroxyl radical scavenger is pushed through a capillary past a focused KrF excimer laser (248 nm wavelength). The pulsed laser is used to photolysis the hydrogen peroxide (H_2O_2) to generate hydroxyl radicals, which have a comparable size to water and high reactivity with a majority of protein amino acid side-chains to produce detectable oxidized products (Huang et al., 2018). Hydrogen peroxide (H_2O_2) photolysis produces a sharp and short-lived burst of hydroxyl radicals that are consumed in under a microsecond (Gau, Chen, & Gross, 2013; Gau et al., 2009). Ideally, with the introduction of exclusion volume, each sample is expected to be exposed to only a single laser pulse eliminating the double exposure of protein to labelling compounds (Aye, Low, & Sze, 2005;

Watson et al., 2009). However, due to the characteristic parabolic flow of the solution within the capillary system under laminar flow conditions, a small volume of sample may be double-exposed (Riaz et al., 2018). Previous analysis of the labeling kinetics of intact proteins by FPOP suggest that the double exposed sample is limited to an undetectable amount if the appropriate amount of exclusion volume is included (Gau et al., 2009).

FPOP is being extensively employed in structural biology; however, FPOP is a time consuming method especially when the sample size is large. Furthermore, it requires significant manual handling of the samples, increasing the possibility of human errors. It requires 13s of mid flow time to irradiate 4µl of sample in a conventional capillary FPOP set up (Watkinson et al., 2017), excluding sample handling. Moreover, conventional capillary FPOP set up is not compatible to study the living cell as living cell imparts a challenge to maintain the appropriate flow rate within the capillary (Chea & Jones, 2018b; Rinas, Mali, Espino, & Jones, 2016). Modification of capillary setup that removes sample reloading, cleaning the capillary between the samples, and procedures like flow adjustment and addition of exclusion volume could substantially improve the FPOP application.

Previously, protein oxidation was carried out in a microfuge tube in a nanosecond timescale with a single laser pulse (Aye et al., 2005). Later, researchers used 12 well plates to conduct protein footprinting of EGFR (Zhu et al., 2017). More recently, calmodulin footprinting was carried out using photoleucine as a labeling agent in both static (Jumper, Bomgarden, Rogers, Etienne, & Schriemer, 2012; Jumper & Schriemer, 2011) and capillary set ups (Zhang, Rempel, & Gross, 2016). Although the two results were grossly similar, there were significant differences in the peptide-level footprint that might arise due to variables other than exposure

technique. The correlation between labeling by FPOP and amino acid side chain solvent accessibility has been established (Huang et al., 2015; Kaur, Kiselar, Yang, & Chance, 2015; Xie, Sood, Woods, & Sharp, 2017) . However, there is no systematic comparison of single pulse batch photolysis and FPOP results that establish the equivalency of the two technologies.

Inspired by work pioneered by Sze and co-workers, we report our first investigations into the technologies required to develop FPOP and related technologies into a high throughput format: the development and evaluation of a 96-well microtiter plate-based FPOP exposure system. We describe efforts involved in ensuring reliability and reproducibility in microtiter FPOP and evaluate microtiter FPOP against traditional flow FPOP using three systems: adenine-based hydroxyl radical dosimetry; oxidation of the model peptide [Glu]1-Fibrinopeptide B (GluB); and HRPF analysis of the model protein myoglobin. With low sample requirements (4 µL or less), very fast exposure times (<5 seconds per sample, which could be automated to <1 second per sample), no sample dilution from unirradiated volumes, and compatibility with common 96-well microtiter-based sample handling workflows already in place in biochemistry and proteomics, our results demonstrate the considerable promise of microtiter-based single pulse batch photolysis for increasing the throughput of HRPF, an emerging technique for the analysis of protein pharmaceuticals (Storek et al., 2018; Watson & Sharp, 2012).

Materials & Methods

Myoglobin, GluB, and catalase were purchased from Sigma Aldrich Corp (St. Louis, MO). Sequencing grade modified trypsin was purchased from Promega (Madison, WI). Adenine and glutamine were purchased from Acros Organics (Fair Lawn, NJ) and methionine amide was from

Bachem (Torrance, CA). Hydrogen peroxide (30%), LC-MS grade water, acetonitrile, formic acid were purchased from Fisher Scientific (Fair Lawn, NJ). Corning™ Costar™ flat bottom 96 well plates were from Fisher Scientific (Fair Lawn, NJ) and V-shaped 96 well microtiter plates were purchased from Greiner Bio-One (Monroe, NC).

Sample oxidation

Fast photochemical oxidation of protein (FPOP) was performed with a Compex Pro 102 excimer laser (Coherent, Germany) at 248 nm wavelength as described previously (Li et al., 2015). Adenine dosimetry was used to ensure the consistency of free radicals generated during FPOP and microtiter plate single pulse photolysis among different samples illuminated with different laser fluences (Xie & Sharp, 2015). For microtiter plate single pulse photolysis, a TECHSPEC® Excimer laser line mirror from Edmund Optics (Barrington, NJ) was placed at 45° of the laser beam at ~15 cm distance of the source as shown in **Figure 8**. The laser beam was deflected by the mirror and subsequently focused by an FL uncoated, UV plano-convex lens (Edmund Optics). The focused beam was allowed to score an index card target mounted on the optical bench to assist with alignment of the plate with the beam. Another card positioned at the height of the surface of the sample was used to position the focusing lens to ensure the laser spot covered the entire surface area of the sample. A clear V-shaped well polystyrene microtiter plate was placed beneath the lens that contained the sample to be irradiated with the laser.

The sample was mixed to contain 5 µM of GluB peptide, 5 µM myoglobin, 1mM adenine, 17 mM glutamine, 100 mM hydrogen peroxide, and 50 mM sodium phosphate buffer (pH 7.4). Immediately after mixing, 4 µl of sample was placed in each microtiter well. The samples were

oxidized with a single shot laser pulse and immediately quenched by addition of 5 µl of quenching solution consisting for a final concentration of catalase (0.5 µg/µl) and methionine amide (0.5 µg/µl) to remove the residual hydrogen peroxide and other secondary oxidants. Samples were illuminated in triplicate at each laser fluence tested. The laser fluence was calculated from the laser energy and laser beam. The estimated fluence was calculated based on the spot size of the beam at the approximate site of contact with the sample, assuming no loss of light from the mirror and focusing. After quenching for one hour, the absorbance of adenine was measured by Nanodrop UV-Vis spectrophotometer from Thermo Fisher Scientific (Waltham, MA).

Traditional capillary flow FPOP was performed as previously described (Xie et al., 2017). Briefly, 20µl of the sample containing 5 µM of GluB peptide, 5 µM myoglobin, 1mM adenine, 17 mM glutamine, 100 mM hydrogen peroxide, and 50 mM sodium phosphate buffer (pH 7.4) for the dosimetry measurement, were flowed through a 100 µm ID capillary through the path of a focused laser. The laser spot width and laser pulse rate were calculated to give an exclusion volume of 15%. Samples were immediately collected in a 1.5 mL microfuge tube containing 25µl quenching solution (final concentration of (0.5 µg/µl) catalase and (0.5 µg/µl) methionine amide) to remove extra hydrogen peroxide and secondary oxidants, and adenine dosimetry was performed as described above.

Peptide/Protein digestion

After measuring the adenine absorbance, Tris HCl (50mM) and DTT (5mM) were added to each sample and final volume was made to 40µl. Samples were heated at 95C for 15min for protein denaturation. Samples were immediately cooled down at room temperatures. Protein

digestion was carried out by adding sequencing grade trypsin in 1:5 ratio trypsin/protein. Samples were incubated at 37C for overnight and reaction stopped by adding 0.1% formic acid. The samples were stored at -20C.

LC-MS/MS analysis

The protein samples were analyzed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to a Dionex Ultimate nano 3000 (Dionex, Sunnyvale, CA). Samples were loaded onto an Acclaim PepMap 100 C18 nanocolumn (0.75 mm × 150 mm, 2 µm, Thermo Fisher Scientific). The peptides were eluted with a gradient consisting of 2% to 35% solvent B (0.1% formic acid in acetonitrile) with a balance of solvent A (0.1% formic acid in water) over 28 min, ramped to 95% B over 5 min, held for 2 min, and then returned to 2% B over 1 min and held for 9 min. Peptides were eluted directly into the nanospray source using a conductive nanospray emitter. All data were acquired in positive ion mode at a spray voltage of 2500 V and a heated capillary temperature of 300° C. Both collision-induced dissociation (CID) and electron transfer dissociation (ETD) were used to fragment peptides, with an isolation width of 3 m/z units. In CID mode, full MS scans were acquired from m/z 350 to 2000 followed by eight subsequent MS2 scans on the top eight most abundant peptide ions. In ETD mode of fragmentation, calibrated charge dependent ETD parameters were used with 5% supplementation activation of collision energy.

Data analysis

Initial data analysis was performed using ByOnic (Protein Metrics, San Carlos CA) with major oxidation products inserted as variable modification. Data analysis was completed

manually, using Xcalibur version 4.1.31.9. Peak intensities of the unoxidized peptides and their corresponding oxidation products observed in LC-MS were used to calculate the average oxidation events per peptide in the sample with maximum 7 ppm mass error. Peptide level oxidation was calculated by summing the ion intensities of all the oxidized peptides multiplied by the number of oxidation events required for the mass shift (e.g., one event for +16, two events for +32) and then divided by the sum of the ion intensities of all unoxidized and oxidized peptide masses as represented by equation 1.

$$P = [I(+16)\text{oxidized} \times 1 + I(+32)\text{oxidized} \times 2 + I(+48)\text{oxidized} \times 3 + \ldots / [I\text{unoxidized} + I(+16)\text{oxidized} + I(+32)\text{oxidized} + I(+48)\text{oxidized} \ldots] \quad (1).$$

where P is the average oxidation events per peptide, and I values are the peak intensities of corresponding oxidized and unoxidized peptides.

Results

Initial measurements were made to determine the efficiency with which we generate hydroxyl radicals in microtiter FPOP. To test this, we measured the loss of absorbance at 260 nm of 1 mM adenine after exposure to hydroxyl radicals. As we previously demonstrated, this loss of adenine absorbance is linear with effective radical concentration (Xie & Sharp, 2015). Our first efforts were performed using traditional flat-bottomed microtiter plates. However, the resulting data were highly irreproducible and showed very poor correlation, as shown in a representative data set in **Figure 9A**. Based on the success with batch laser photolysis reported both by our group and Aye and coworkers (Aye et al., 2005; Charvatova et al., 2008; Zhu et al., 2017), we

hypothesized that irreproducibility in sample illumination was the cause. Polystyrene is largely opaque at 248 nm; we hypothesized that UV shadowing of the sample near the wall of the well resulting from an imperfectly perpendicular incident laser beam and/or an imperfectly aligned well may be leading to the poor correlation and reproducibility observed. To test this hypothesis, we performed similar adenine dosimetry experiments by placing the sample at the center of the well in a V-shaped well plate, where the edges of the sample are far from the sides of the well. The results are shown in **Figure 9B**. Adenine dosimetry indicated an effective hydroxyl radical dose comparable to that of traditional flow FPOP, and the radical dose increased linearly with estimated laser fluence. All further experiments described were carried out in a V-shaped 96-well microtiter plate, with the sample centered in the well away from the walls to prevent shadowing.

Model peptide oxidation

In order to ensure that the microtiter format was not causing spurious readings in our adenine dosimetry, we also tested the efficiency at which we could oxidize a model peptide, GluB, by microtiter FPOP. We examined the average oxidation events per molecule for 5 µM GluB in both microtiter FPOP and traditional flow FPOP at four different laser fluences. The results are shown in **Figure 10**. For GluB, the results of microtiter FPOP and traditional flow FPOP were indistinguishable. Differences in the optical interface and the lack of an exclusion volume in FPOP could contribute to measured differences between capillary and microtiter FPOP oxidation efficiency, but both GluB and adenine dosimetry show that oxidation is comparable between the two methods. We are uncertain why adenine showed a small difference in exposure efficiency for microtiter FPOP compared to flow FPOP, while GluB did not. However, in both cases, the change in oxidation was linearly related to estimated laser fluence with strong reproducibility.

This suggests that perhaps insufficient oxidized adenine recovery, especially at lower hydroxyl radical doses, may contribute to observed differences in adenine dosimetry. A separate set of adenine dosimetry triplicate analyses in microtiter plates gave identical results as the microtiter dosimetry results shown in Figure 2B (data not shown) indicating that, whatever the cause of differential slopes of adenine dosimetry response between FPOP and microtiter plate, the dosimetry is reproducible.

HRPF of a model protein

As the primary use of FPOP lies in measuring changes in protein topography, we tested whether performing the oxidation in the microtiter plate environment gave similar topography results as samples tested by traditional flow FPOP. We used myoglobin as our model protein due to its long history of HRPF studies as a model system (Chance, 2001; Sharp et al., 2003). After the exposure of myoglobin at four different laser fluences, the protein was digested, run on LC-MS and the oxidation level of the peptides was calculated. The sequence coverage of myoglobin was 88.3% on the ByOnic search for microtiter setup. On peptide level analysis of oxidation in microtiter experiment, we found that out of ten peptides observed, seven peptides showed oxidation in a linear fashion while the other three peptides did not show any oxidation. The data analysis of oxidation event of seven peptides showed a linear response to increasing laser fluence for all peptides as shown in **Figure 11**. With the exception of peptide 17-31, all peptides also showed the good precision of measurement.

Finally, we performed the FPOP experiment for myoglobin in a capillary setting to compare the results obtained with the microtiter plate. Two relatively similar laser fluences (8.5

mJ/mm2 in capillary compared to 8.3 mJ/mm2 in the plate) were used to compare the peptide level oxidation between two methods. After tryptic digestion and LC-MS run of the capillary set up experiment, we found the sequence coverage to be 99.35%. We found eleven peptides in capillary set up and eight of them showed detectable oxidation. Upon comparing the oxidation of peptides between two methods, no peptide showed statistically significant differences providing the evidence that microtiter plate oxidation of protein produces comparable oxidation to the capillary method as presented in **Figure 12**.

Myoglobin oxidized by both capillary FPOP and microtiter single pulse photolysis methods yielded equivalent results, indicating that the oxidation methods are equivalent if exposure is appropriately controlled. We observed that the plate set up provided us similar oxidation although it had experienced slightly less laser fluence (8.5mJ/mm2 in capillary compared to 8.3mJ/mm2 in the plate). In FPOP, a 15% exclusion volume was used to prevent illumination of any significant volume of sample with two laser pulses; therefore, approximately 15% of the sample remains unoxidized, lowering the average oxidation measured. However, in microtiter single pulse photolysis, no exclusion volume is necessary; the entire sample volume is illuminated by a single laser pulse. Therefore, assuming equivalent optical properties (e.g. light scatter from the liquid-air interface vs. light scatter from the capillary), microtiter single pulse photolysis should yield more efficient oxidation than FPOP for equivalent laser fluence. The relatively small change in oxidation suggests that differential optical properties of the capillary versus the open microtiter well compensates largely for the exclusion volume.

Discussion

Here, we demonstrate that microtiter single pulse photolysis can achieve comparable oxidation as traditional FPOP, yielding identical topographical measurements of proteins. With the simple and low-cost design of experimental setup, the newly developed technique reduces the experiment time substantially and provides the opportunity to conduct the experiment with low sample volume. Current robotic options for microtiter plate-based assays can be leveraged to minimize human error during the sample handling, especially when dealing with a large number of samples. Automation of the technique will bring a drastic change in experimental time and reduce sample variation due to human errors. This is a key step in the adaptation of HRPF for sample screening based on topographical changes (e.g. screening for allosteric binding). Similarly, the lack of a flow system adds flexibility to HRPF, allowing for the analysis of viscous samples such as some poorly soluble membrane protein preparations or other samples that present problems in a standard flow system. Additionally, as the whole sample is exposed to the laser pulse, the exclusion volume of the capillary set up can be eliminated, alleviating concerns with laminar flow and improving oxidation efficiency.

Microtiter single pulse photolysis has several drawbacks compared to traditional FPOP that should be clearly acknowledged, as well. The area of incidence between the laser light and the sample is considerably larger for microtiter single pulse photolysis. The use of cylindrical lenses to asymmetrically compress the beam to the shape of the capillary allows traditional FPOP to reach higher laser fluences at more modest pulse energies (Zhang et al., 2018). Because the area illuminated is much larger in the microtiter plate, the laser energy per pulse required is considerably higher, potentially requiring a higher power laser. Online FPOP offers online mixing

of reagents, allowing for interesting technologies including fluidic focusing of living cells for footprinting (Rinas et al., 2016) and rapid mixing for time-resolved footprinting experiments (Vahidi, Stocks, Liaghati-Mobarhan, & Konermann, 2013). Additionally, traditional FPOP offers flexibility by examining theoretically unlimited volumes of sample without reloading, limited only by the amount of time the investigator wishes to spend on the experiment and the volume of the sample reservoir used. Additionally, if each laser illumination event is viewed as a separate experiment on the same sample, it becomes apparent that FPOP data actually represents an average of many technical replicates of the photochemical event. While our data showed no systematic advantage in oxidation precision to FPOP over microtiter photolysis, one can imagine such a reproducibility advantage when dealing with light sources with high pulse-to-pulse variability.

With UV-opaque microtiter plates, the volume must be kept from the edges of the well unless the researcher is willing to perform highly precise laser/sample alignment to ensure complete and reproducible illumination of all samples. This limits the volume of sample that can be illuminated in a V-shaped 96-well microtiter plate. The use of UV-transparent quartz microtiter plates and automated plate alignment may alleviate this problem, but the sample volume will still be limited by the well volume and the desired path length for the incident UV light. Microtiter single pulse photolysis has a potential role to play as an extension of current HRPF technologies towards particular high throughput needs or specialty sample analysis, rather than a wholesale replacement of FPOP.

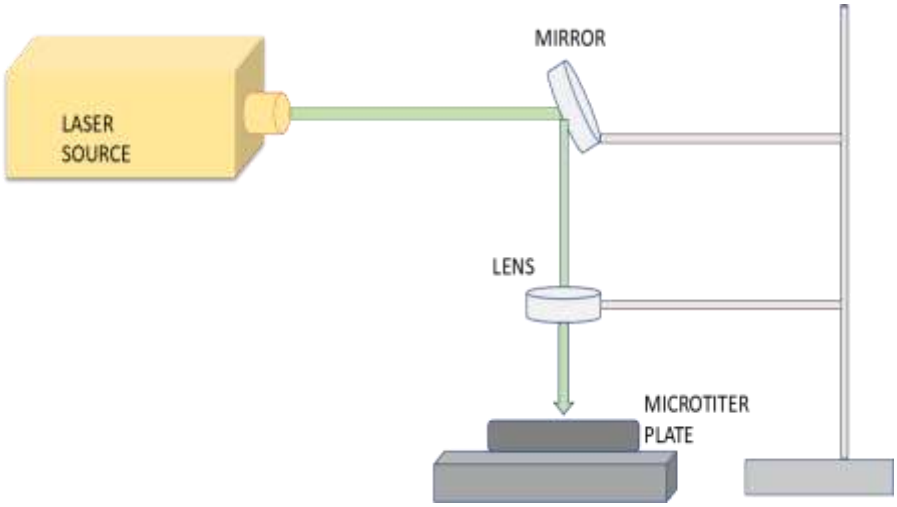


Figure 8: Schematic cartoon for microtiter FPOP. A mirror placed at 45^0 angles of laser path deflects the laser beam onto a lens, which then focuses the light onto the well of the microtiter plate.

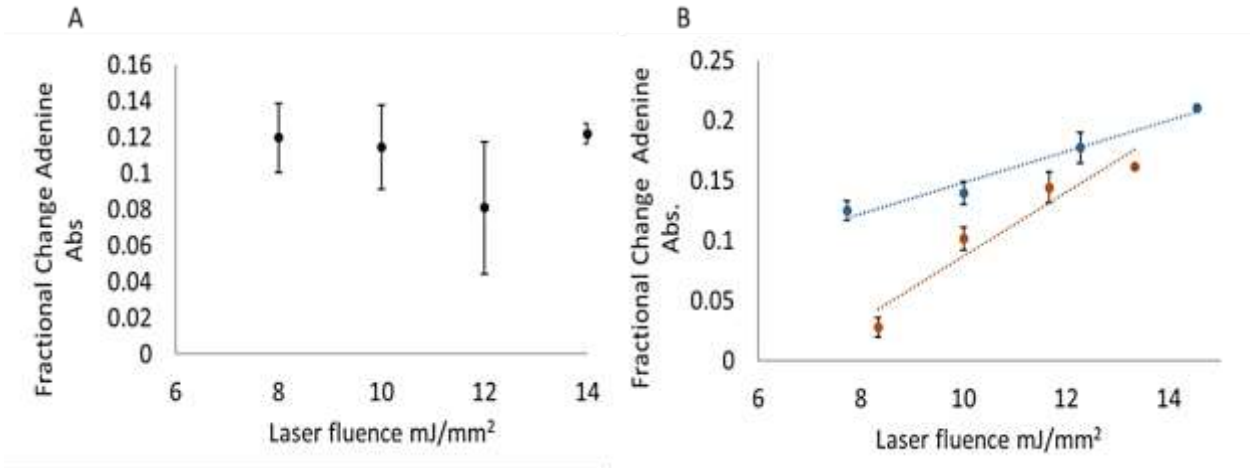


Figure 9. Adenine dosimetry of microtiter FPOP. (A): Representative data set of microtiter single pulse photolysis carried out in a flat-bottomed 96-well microtiter plate. (B) Microtiter single pulse photolysis carried out in a (orange) V-shaped 96-well microtiter plate or (blue) traditional flow FPOP in a fused silica capillary. Error bars represent one standard deviation from a triplicate data set.

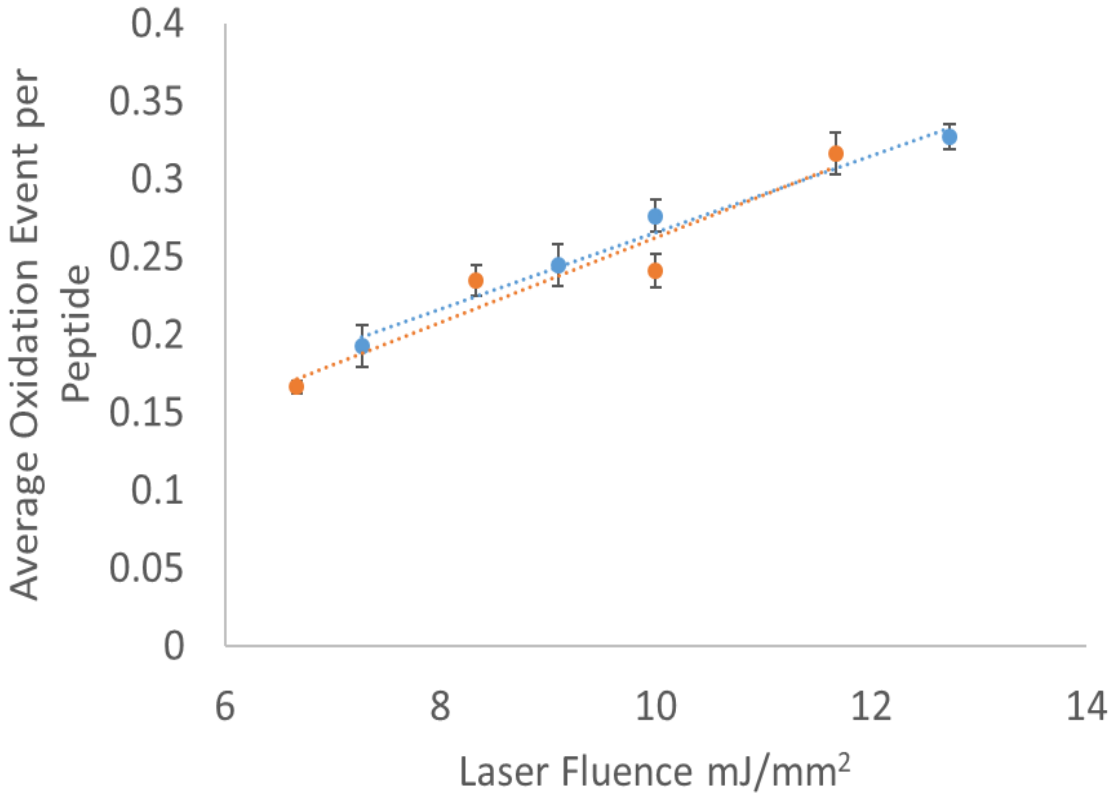


Figure 10. Comparison of GluB peptide oxidation by (orange) microtiter single pulse photolysis versus (blue) traditional capillary FPOP**.** Error bars represent one standard deviation from a triplicate data set.

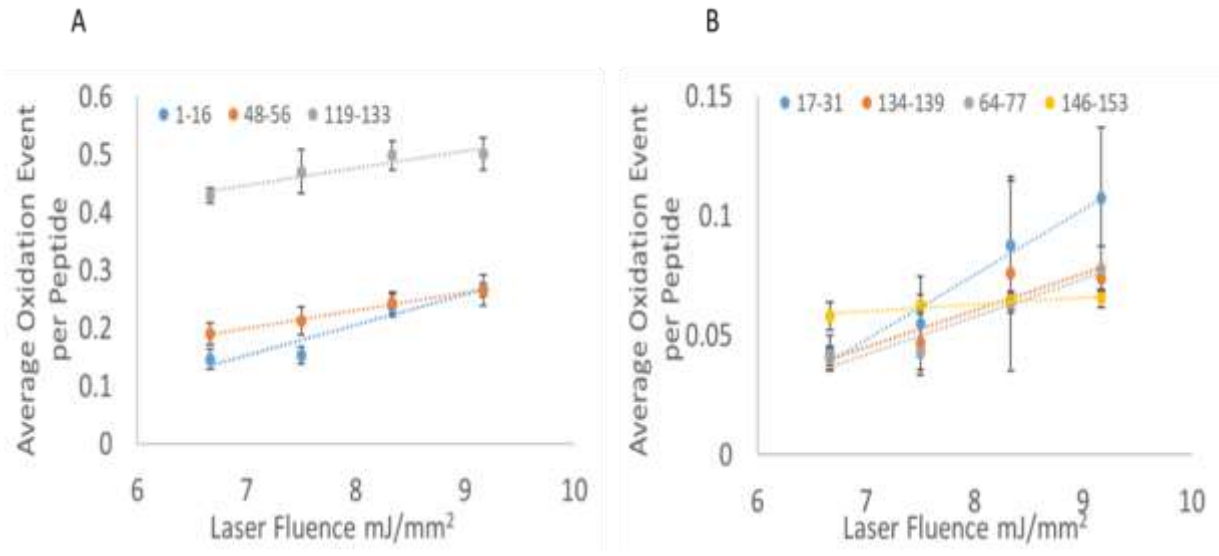


Figure 11. Correlation among oxidation event for seven peptides of myoglobin. 5 µM myoglobin exposed to four different laser fluences keeping reagent concentration constant. Error bars represent one standard deviation from a triplicate data set. Peptides are grouped according to level of oxidation for improved clarity of data; both left and right panels represent peptides from the same set of experiments.

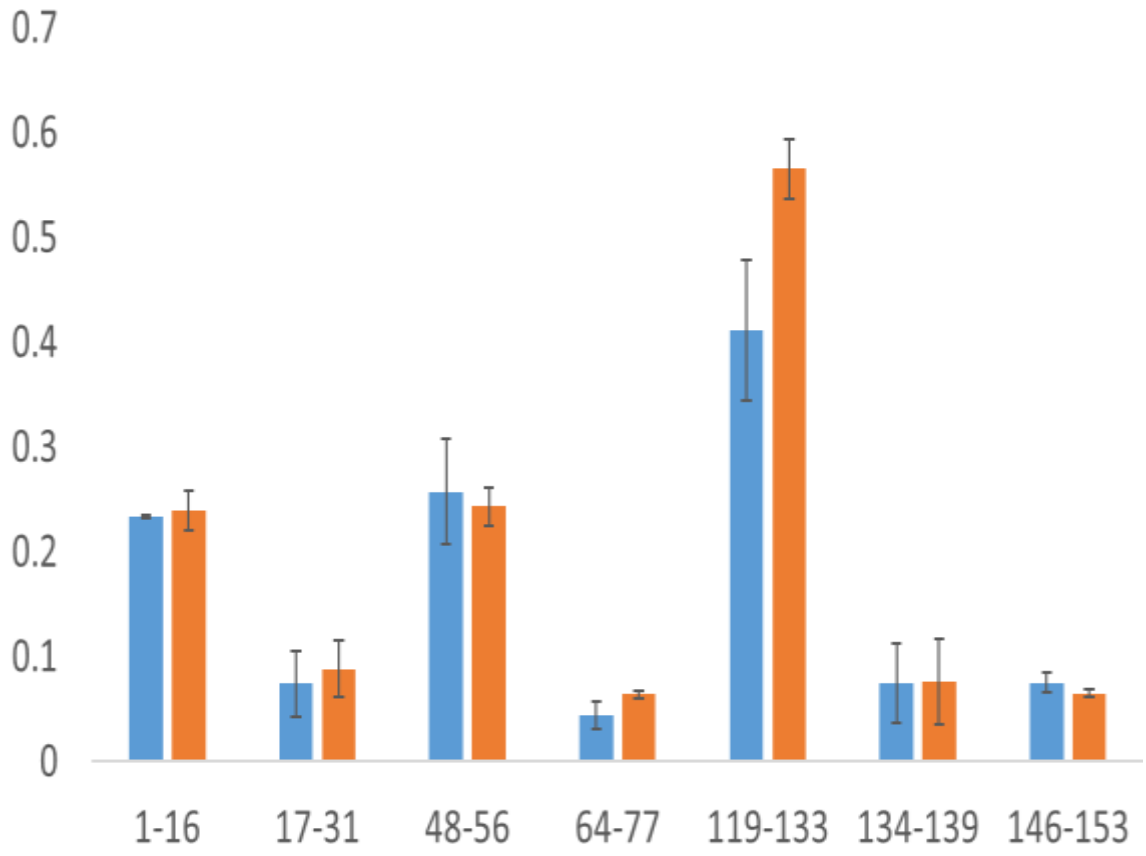


Figure 12. Comparison of myoglobin peptide oxidation between (blue) capillary FPOP and (orange) microtiter single pulse photolysis. Error bars represent one standard deviation from a triplicate data set. No statistically significant differences in oxidation were detected ($\alpha = 0.05$).

CHAPTER 3: IDENTIFICATION OF LECTINS FROM PULSES THROUGH A PROTEOMIC APPROACH

Introduction

Lectins are a diverse group of proteins that bind glycans of glycoproteins, glycolipids, or polysaccharides with high affinity (Goldstein & Hayes, 1978). They are non-enzymatic proteins containing at least one non-catalytic domain and bind specifically and reversibly to different types of glycoproteins, mono- or oligosaccharides (He et al., 2013). Lectins serve as recognition molecules within a cell, between cells, or between organisms due to their binding specificity to different carbohydrates (Chrispeels & Raikhel, 1991). Lectins perform various basic biological functions, including regulation, adhesion, protection against pathogens, and many others (Wiederschain, 2009). Different classes of lectins are usually present in an organism in various isoforms known as isolectins (Damme, Peumans, Pusztai, & Bardocz, 1998). The isoforms differ slightly in terms of structural orientation and sugar binding specificity but have significant difference in their biological activities (Leavitt, Felsted, & Bachur, 1977). Lectins can be classified according to their structural composition such as merolectins that contain a single carbohydrate-binding domain, hololectins that are composed of two or more such domains, and chimerolectins that contain additional, non-lectin domains, usually catalytic in function (Peumans & Damme, 1995). Most of the characterized plant lectins are secretory proteins, meaning that they enter

the secretory system and subsequently accumulate either in vacuoles or in the cell wall and intercellular spaces (Chrispeels & Raikhel, 1991).

The scientific discipline called "lectinology" originated back in 1888 when scientist Stillmark observed the toxicity of castor bean extracts linked to the presence of a proteinaceous hemagglutinating factor called ricin (Kader & Delseny, 2011). The term hemagglutinin was introduced by Elfstrand in 1898 as a common name for all plant proteins that clump blood cells. The discovery of nontoxic lectins in early 1900 from the legumes *Phaseolus vulgaris*, *Pisum sativum*, *Lens culinaris*, and *Vicia sativa* led to the idea of the beneficial effects of plant lectins. Subsequently many nontoxic lectins were discovered, and it became clear that lectins are widespread in the plant kingdom. During that period, it was discovered that plant lectins are capable of selectively agglutinating particular human blood groups within the ABO system, and this finding led to the introduction of the term "lectin," which is derived from "legere," the Latin verb for "to select" (Boyd & Reguera, 1949; Renkonen, 1948). However, because most hemagglutinins are also capable of agglutinating other cells, they were also referred to as agglutinins. Though the term lectin is actually most commonly used, the terms agglutinin and hemagglutinin still persist as synonyms.

Among plant lectins, legume lectins have been the most widely found (Hamid, Masood, Wani, & Rafiq, 2013) and belong to a group of highly homologous proteins that are abundant in the seeds (Lagarda, Guzman, & Vazquez, 2017). The three-dimensional structure of plant lectins is characterized by β-sheets that are connected by α turns, β turns and bends **Figure 13** (Lagarda-Diaz et al., 2017). Additionally, the quaternary interfaces are formed between β-sheets, and β-sheets are connected by loops forming antiparallel chains usually devoid of α helices (Sharma &

Surolia, 1997; Nathan & Lis, 1990; Sharon & Lis, 2002). Structural analysis of monomers indicate a sandwich of 25–30 kDa containing a carbohydrate recognition domain (CRD) and metal binding sites for divalent cations (Ca2+ and Mn2+) (Lagarda-Diaz et al., 2017). Despite the similarities in primary, secondary and tertiary structures, the quaternary structures show considerable variations and have functional implications as they dictate the specificity of multivalent glycan binding (Bouckaert, Hamelryck, Wyns, & Loris, 1999; Vijayan & Chandra, 1999).

In this study, we employed bottom-up proteomics to identify lectins from various pulses that might have different pharmacological functions especially antimicrobial and antimalarial effects. We analyzed eleven different samples from nine different species, namely, *Phaseolus vulgaris, Cajanus cajan, Vigna radiata, Lens culinaris, Vigna mungo, Cicer arietinum, Vigna unguiculata, Phaseolus lunatus,* and *Pisum sativum.* These are the most widely produced and consumed pulses with significant nutritional values. Pulses have provided us with different nutritional compounds that perform various physiological functions. Some pulses are effective as antimicrobial (Breitenbach et al., 2018) or anticancer agents (Lam & Ng, 2011). In this study, we aimed to identify new lectins from the pulses and to perform in-vitro experiments to determine their pharmacological activities.

Materials & Methods

Preparation of Crude Extract

Pulses were purchased from local shops and grinded in a Retsch mixer mill 400, at 30 Hz for 2-4 minutes using two large metal balls. Later, 1 g of fine lentil powder was added to 10 ml PBS (1X, pH- 7.0) and mixed by agitation overnight at 4°C. The extract was filtered through two

layers of cheese cloth and centrifuged at 10,000 rpm for 30 minutes at 4°C. The supernatant was used as crude extract and stored at -20 °C.

Preparation of Aqueous Two-Phase System (ATPS)

The ATPS was prepared by adding 15% (NH4)2SO4 (w/w), 18% PEG 600 (w/w), 0.4 g/5 g NaCl, and 1 mL of pulse seed crude extract. Ultrapure water was used to bring the final weight of the system to 5 g. The mixture was vortexed, and the pH of the system was adjusted to 7.5 using NaOH (1 mol/L). The complete phase separation was achieved by centrifuging at 2000 g for 10 minutes at 25°C. The top phases were collected to measure protein concentration and protein separation by SDS-PAGE.

Electrophoresis

SDS-PAGE was carried out in order to evaluate the purity of protein in the phases, using 12% Criterion™ TGX™ Precast Midi Protein Gel. 10 µl of crude extract and 20 µl of semi-purified upper fraction were added to 6X loading buffer. Later the buffer was boiled for 10 minutes and 15 µl of each sample was loaded on the 12% SDS gel. The gel was run for approximately 95 minutes using tris-glycine–SDS running buffer at constant voltages of 80V for stacking gel and 120V for separating gel. After the end of the electrophoresis, the gel was stained with Coomassie brilliant blue R-250 for 30 minutes and decolorized with destaining solution.

In gel digestion

Protein was denatured and digested in gel following the procedures published previously (Shevchenko, Tomas, Havlis, Olsen, & Mann, 2006). The gel bands were cut, transferred to

eppendorf tubes, and the gel was grinded using tips. Later 500 µl of acetonitrile was added to each sample and kept for 10 minutes. Acetonitrile was then removed and proteins were denatured by incubating the samples in 40 µl of 10 mM DTT for 30 minutes at 56^0C. Samples were cooled down at room temperature, DTT was removed, and 500 µl of acetonitrile was added to each sample. After 10 minutes, acetonitrile was removed and 40 µl of 55 mM iodoacetamide was added and incubated in the dark for 20 minutes at room temperature. Later iodoacetamide was removed and 500 µl acetonitrile was added and incubated for 10 minutes at room temperature. Protein was digested by incubating the proteins with sequencing grade trypsin in 10 mM of ammonium bicarbonate buffer overnight at 37 ^{0}C. Proteins were extracted using an extraction buffer, and the extraction buffer later evaporated through vacuum centrifuge. Proteins were re-dissolved in 20 µl of 0.1% TFA, centrifuged for 15 minutes and stored at -20^0C.

LC-MS/MS analysis

The protein samples were analyzed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to a Dionex Ultimate nano 3000 (Dionex, Sunnyvale, CA). Samples were loaded onto an Acclaim PepMap 100 C18 nanocolumn (0.75 mm × 150 mm, 2 µm, Thermo Fisher Scientific). The peptides were eluted with a gradient consisting of 2%–35% solvent B (0.1% formic acid in acetonitrile) with a balance of solvent A (0.1% formic acid in water) over 18 minutes, ramped to 95% B over 1 minute, held for 4 minutes, and then returned to 2% B over 1 minute and held for 6 min with the flow of 0.3 µl/min. Peptides were eluted directly into the nanospray source using a conductive nanospray emitter. All data were acquired in positive ion mode at a spray voltage of 2400 V and a heated capillary temperature of 300 °C. Collision-induced dissociation (CID) was used to fragment peptides, with an isolation width of 3 m/z units. In CID

mode, full MS scans were acquired from m/z 350 to 2000 followed by eight subsequent MS2 scans on the top eight most abundant peptide ions.

Database searching

The target database was formed by using the Non-redundant reviewed database (Swiss-Prot), species specific reviewed database (Swiss-Prot) and species specific un-reviewed database (TrEMBL) from Uniport (https://www.uniprot.org/). Swiss-Prot is a database for manually annotated and reviewed proteins, and TrEMBL is a computationally analyzed un-reviewed database. The decoy database was created by reversing the sequence of the target database. Protein search was performed using ByOnic (Protein Metrics, San Carlos CA) software with the following search parameters: precursor mass tolerance 100 ppm, fragment mass tolerance 0.4 Da, enzymatic cleavage with two possible missed cleavages, and fixed modification with iodoacetamide (57.021464Da). The proteins were identified with 2% FDR and only those identifications which showed minimum of three peptides in protein search were considered for future analysis in the Swiss-Prot database. However, for the analysis of the TrEMBL database, 2% FDR and no peptide cut off were employed, and all the proteins identified through Byonic search were subjected to further analysis.

Pfam (Bateman et al., 2009) and the NCBI-CCD database (Marchler-Bauer et al., 2015) were used to annotate the uncharacterized proteins identified in the TrEMBL search.

Results

SDS gel analysis indicated the presence of lectins in three different molecular weight bands, as shown in **Figure 14**. Nine different species were analyzed, and mass spectrometric

analysis was performed for three molecular bands (45,35,25 kDa) for each sample from the gel. A total of seven lectin proteins were identified for *P. vulgaris*, of which three were characterized in the Swiss-Prot database. The other four proteins were identified in the TrEMBL database search and later annotated using Pfam, as shown in **Table 1**. Homology analysis of these seven proteins was performed using Clustal Omega (Sievers et al., 2011), and multiple sequence alignments of the lectin from different databases were made. As shown in **Figure 15**, lectins from TrEMBL database annotated with Pfam, had a higher degree of consensus with the Swiss-Prot lectins, indicating the positive identification. Alpha-amylase inhibitor protein had a higher degree of amino acid sequence similarity with legume lectin domain protein (V7C3R6), however, phytohemagglutinin leucoagglutinating (PHA L) and erythroagglutinating (PHA E) had similarity with V7C787, V7C654, V7BC16 proteins identified from the TrEMBL database search, indicating that the genes for PHA-E, PHA-L, and alpha-amylase inhibitor are all linked, and occur at the same locus based on the gene mapping research (Chrispeels & Raikhel, 1991). No Swiss-Prot reviewed protein was identified from the species *C. cajan*; however, one lectin was identified in the TrEMBL database source, as shown in **Table 2**. As the number of identified peptides and sequence coverage were very low, we assume it might not be the true identification. Moreover, very little similarity among the lectins of *C. cajan* and *P. vulgaris* were found when the sequences were aligned. The species, *V. radiate*, also produced similar results as *C. cajan* with no lectin identification in the Swiss-Prot search, but the TrEMBL database search identified lectins with very few peptides and sequence coverage, as shown in **Table 3**. For both *C. cajan* and *V. radiate* species, PHA L and E were identified from *P. vulgaris*. Swiss-Prot and TrEMBL database searches of the *L. culinaris* identified lectins in both searches (**Table 4),** and lectin identified in the TrEMBL

database search has shown similarity with Swiss-Prot identified lectin when the amino acid sequence aligned using Clustal Omega (**Figure 16)**. For *V. mungo*, lectin was identified in TrEMBL search, but no specific lectin from this specie was identified in the Swiss-Prot search. PHA E/L from *P. vulgaris* were identified with a non-redundant database search, as shown in **Table 5**. Lectin form *C. arietinum* was identified in both non-redundant and Swiss-Prot database searches (**Table 6)**. Swiss-Prot search for the species *V. unguiculata* did not find any lectin for this specie, but TrEMBL identified one seed lectin with only one peptide matching and very low sequence coverage (**Table 7)**. For *P. lunatus,* no lectin was identified in the Swiss-Prot search; however, four different lectins were identified with the TrEMBL search, as shown in **Table 8,** and sequence alignment analysis indicates a high degree of similarity (**Figure 17)**. In the case of *P. sativum,* lectin proteins were identified in all the database searches along with PHA E lectin in the non-redundant database, as shown in **Table 9**. The number of proteins identified using different databases significantly vary from each other, as shown in **Appendix 1**.

Discussion

In 1960, Nowell made the discovery that lectin of the red kidney bean (*P. vulgaris*), known as phytohemagglutinin (PHA), is mitogenic (Nowell, 1960). This finding brought lectins into the limelight, and since then, there have been great strides in the knowledge of lectin and lectinology. Lectins are generally found between 35-45 kDa in SDS gel analysis, and during this experiment, three bands (25, 35, 45 kDa) were cut and analyzed. We detected three major lectins, PHA E, PHA L and alpha amylase inhibitor, in all three bands. We reasoned the lectins were detected in all three bands due to the high sensitivity of the MS instrument. Moreover, during the LC analysis there might be some carryover of lectins that might be detected in the MS analysis. Furthermore,

the proteins also travel within the gel rather than being static, and, as we analyzed very close bands, it was possible to detect the same proteins in these three close bands. These types of lectins show remarkable homology sequences with other legume lectins (He et al., 2015), and the structures are well characterized. PHA consists of two types of polypeptide chains called E and L, reflecting their preferential binding to erythrocytes and leukocytes, respectively (He et al., 2018). For *P. vulgaris* and two of its varieties, Pinto beans and Black turtle, we positively identified the two dominants lectins, PHA E & L in the Swiss-Prot database search, indicating presence of the lectins and reliability of the methods. Moreover, PHA E and L were identified for almost all the pulses analyzed maybe due to their extreme similarity with other legume lectins. Additionally, four lectin domains were identified in a TrEMBL search and annotated using the Pfam database.

Globulins, albumins, and glutelins are the major proteins found in the seeds of *C. cajan,* also known as Pigeon pea, (Roy, Boye, & Simpson, 2010) with globulins making up about 65% of the total protein in pigeon pea (Singh & Jambunathan, 1982). Although lectins might be present as minor bioactive proteins (Campos, Loarca, & Oomah, 2010) in *C. cajan,* our proteomics analysis showed no lectins present in the Swiss-Prot search. This might be due to the fact that the Uniprot reviewed database has no reviewed lectin entries for this species and consequently no lectin identification for the *C. cajan*. *V. radiata,* also known as mung beans, are composed of about 20–24% protein. Globulin and albumin are the main storage proteins found in mung bean seeds and make up over 60% and 25% of the total mung bean protein, respectively (Tang, Dong, Ren, Li, & He, 2014). Like *C. cajan*, the Uniprot reviewed database has no lectins specified for *V. radiate,* which is why we were not able to find any lectins in a reviewed database search, although a TrEMBL search identified two lectin domains with very little sequence coverage and might be

a false identification. The identification of proteins was made with 2% FDR for the TrEMBL database, and functionally non-assigned proteins were annotated using the Pfam database. However, for both *C. cajan* and *V. radiate,* lectins were identified from *P. vulgaris* with the non-redundant database search and this might be due to the fact that both of these species seed lectins show extreme homology with PHA E and L lectins when aligned in Clustar Omega software, as shown in **Figure 18**.

L. culinaris, also known as lentils, are abundant sources of protein storage, providing essential and non-essential amino acids to the human body (Ganesan & Xu, 2017). The predominant proteins in lentils are globulin (47% of the total seed proteins) and an adequate quantity of albumin (Lombardi, Ruggeri, Aguzzi, & Cappelloni, 2003). Lectins are present as major bioactive components in the lentils and are well-characterized proteins. Bottom-up proteomics analysis identified different lectins from *L. culinaris* in all three database searches. Homology analysis indicated sequence agreement of *L. culinaris* and *P. vulgaris* lectins; however, no significant homology was observed with lectins identified from *P. sativum* and another species. Again, there is no reviewed Swiss-Prot entry for *V. mungo,* and no lectin was identified with a reviewed database searches. The homology analysis of identified lectins showed moderate agreement with lectins of *V. mungo*. *C. arietinum*, commonly known as the chickpea, is a good source of carbohydrates and protein, together constituting about 80 % of the total dry seed mass (Chibbar, Ambigaipalan, & Hoover, 2010).

Lectins were identified with three database searches for *C. arietinum*; however, we identified lectins from *P. vulgaris* during the analysis of *C. arietinum*. Homology analysis showed poor sequence alignment, but we are not sure why PHA E and L lectins were identified during the

analysis. *V. unguiculata,* or cowpeas seeds, provide a rich source of proteins and calories, as well as minerals and vitamins (Goncalves et al., 2016) and can consist of 25% protein with very low fat content (Rangel, Domont, Pedrosa, & Ferreira, 2003). There are few lectin entries in the uniprot reviewed database; however, our proteomics analysis was not able to identify any of the lectins from cowpeas with the search parameters. Proteomics analysis of *P. lunatus* showed no lectin was detected during the analysis with reviewed database searches although there is a lectin entry in the Swiss-Prot database. However, a TrEMBL search found four different lectins with high degree of homology in their sequences indicating the presence of lectins. *P. sativum,* or peas, presents interesting pharmacological activities, and recent studies demonstrated that pea lectin inhibited Ehrlich ascites carcinoma cells (Dias et al., 2015). All the lectins identified from different database searches showed significant homology while their sequences aligned using Clustal Omega **Figure 19**.

Conclusion

We employed proteomics strategy to identify new lectins from various pulses. With LC-MS and protein database analysis, we were able to identify different lectins from various pulses. Moreover, we annotate some lectins proteins identified in protein database searches using the Pfam database. Currently, all the pulses samples are under *in vitro* investigation to find pharmacological (antimicrobial and anti-malarial) effects of the pulses components. Separation of the lectins from these pulses using affinity chromatography and subsequent pharmacological investigation holds promise to diversify the use of these pulses.

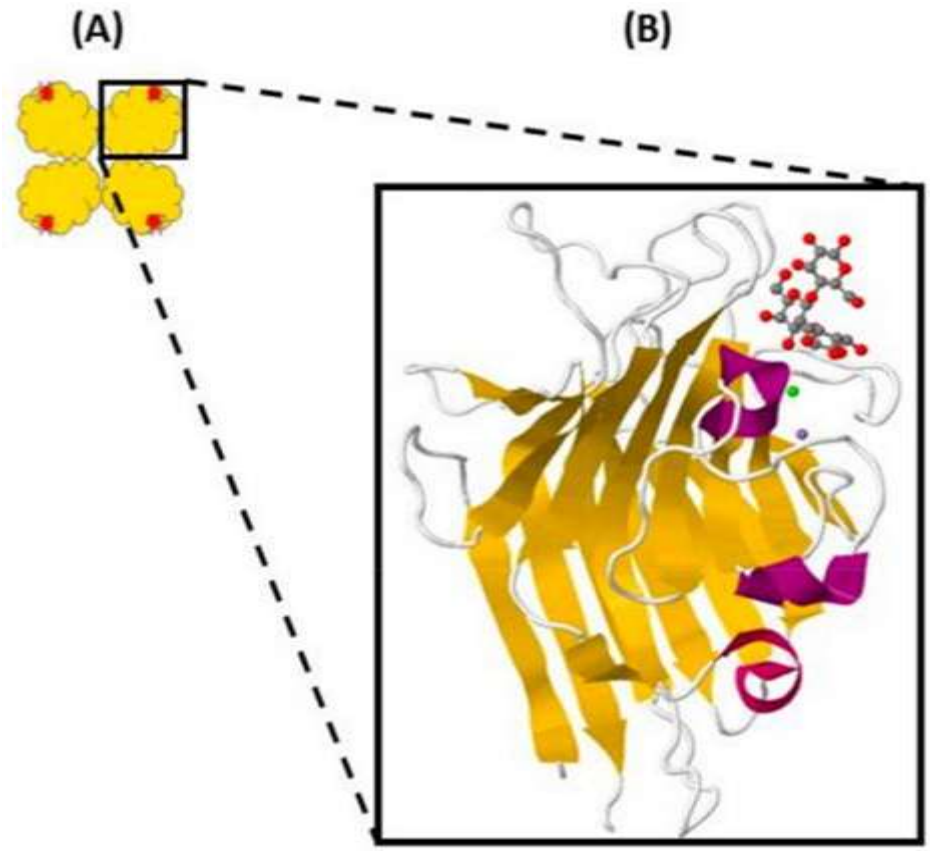


Figure 13: Three dimensional structure of lectin (A) tetramer with carbohydrate recognition domains (red); (B) amplified image of monomer with β-sheets (yellow), α-turns (purple), metal binding sites (area with green and gray spheres) and carbohydrate recognition domain (area occupied by the grey and red molecule). Open access and permitted to use (Lagarda-Diaz, Guzman-Partida et al. 2017)

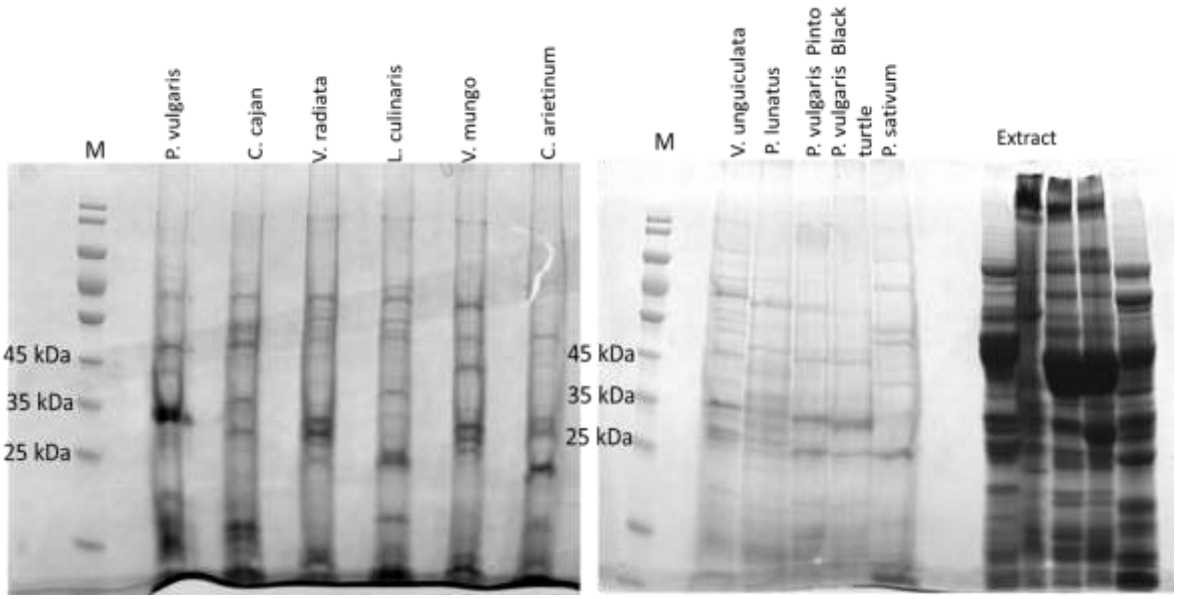


Figure 14: Gel electrophoresis analysis of different pulses to identify lectin.

Protein name	Accession number	Species	Log probability	Database
PHA L	P05087	*P. vulgaris*	129.03	Non-redundant
PHA E	P05088	*P. vulgaris*	76.91	Non-redundant
Alpha-amylase inhibitor	P02873	*P. vulgaris*	41.20	Non-redundant
Legume lectin domain	V7C787	*P. vulgaris*	230.51	Pfam
Legume lectin domain	V7C654	*P. vulgaris*	142.51	Pfam
Legume lectin domain	V7C3R6	*P. vulgaris*	86.92	Pfam
Legume lectin domain	V7BC16	*P. vulgaris*	6.07	Pfam

Table 1: Lectin identified from P. vulgaris with protein database search. Swiss-Prot search identified lectin where Pfam database used to annotate the uncharacterized protein in TrEMBL search

Protein name	Accession number	Species	Log probability	Database
PHA L	P05087	*P. vulgaris*	42.12	Non-redundant
PHA E	P05088	*P. vulgaris*	31.36	Non-redundant
Lectin-domain containing receptor kinase	A0A151SKL1	*C. cajan*	0.58	TrEMBL

Table 2: Lectin identified from *C. cajan* with protein database search. Swiss-Prot or non-redundant search did not identify lectin where TrEMBL found of lectin domain

Protein name	Accession number	Species	Log probability	Database
PHA L	P05087	*P. vulgaris*	21.90	Non redundant
PHA E	P05088	*P. vulgaris*	6.20	Non redundant
Mannose-specific lectin	A0A1S3V2P1	*V. radiata*	5.10	TrEMBL
Probable L-type lectin-domain	A0A1S3VR88	*V. radiat*a	0.40	TrEMBL

Table 3: Lectin identified from *V. radiata* with protein database search. Swiss-Prot or non-redundant search did not identify lectin where TrEMBL found of lectin domain

Protein name	Accession number	Species	Log probability	Database
Lectin	Q93WH6	*L. culinaris*	22.82	Non redundant
PHA E	P05088	*P. vulgaris*	15.78	Non redundant
Legumin	P15838	*P. sativum*	7.87	Non redundant
Alpha-amylase inhibitor	P02873	*P. vulgaris*	15.35	Non redundant
Lectin	A0A173G7D5	*L. culinaris*	61.44	TrEMBL
PHA L	P05087	*P. vulgaris*	28.82	Non redundant

Table 4: Lectin identified from *L. culinaris* with protein database search. Both reviewed and un-reviewed database were able to identify lectin.

Protein name	Accession number	Species	Log probability	Database
Lectin (Fragment)	A2BCP2	*V. mungo*	1.35	TrEMBL
PHA E	P05088	*P. vulgaris*	15.56	Non redundant
PHA L	P05087	*P. vulgaris*	15.83	Non redundant
Alpha-amylase inhibitor	P02873	*P. vulgaris*	5.50	Non redundant
Lectin	Q93WH6	*L. culinaris*	10.93	Non redundant

Table 5: Lectin identified from *V. mungo* with protein database search. No lectin identified for *V. mungo* with Swiss-Prot search, however, lectin identified in non-redundant search from different species.

Protein name	Accession number	Species	Log probability	Database
Legumin	Q9SMJ4	*C. arietinum*	18.77	Non redundant
PHA E	P05088	*P. vulgaris*	6.56	Non redundant
PHA L	P05087	*P. vulgaris*	16.07	Non redundant
legumin A-like isoform	A0A1S2XSB9	*C. arietinum*	68.93	TrEMBL
legumin J-like isoform	A0A1S2XVG1	*C. arietinum*	33.07	TrEMBL
legumain	A0A1S2Z637	*C. arietinum*	10.91	TrEMBL
legumin like	A0A1S2XTK6	*C. arietinum*	8.57	TrEMBL

Table 6: Lectin identified from *C. arietinum* with protein database search.

Protein name	Accession number	Species	Log probability	Database
Lectin (Fragment)	A4FS28	*V. unguiculata*	5.44	TrEMBL
PHA E	P05088	*P. vulgaris*	10.48	Non redundant

Table 7: Lectin identified from *V. unguiculata* with protein database search. No lectin identified for with Swiss-Prot search but lectin found with TrEMBL with minimum peptide and sequence coverage.

Protein name	Accession number	Species	Log probability	Database
Alpha-amylase inhibitor	Q9LED6	*P. lunatus*	139.51	TrEMBL
Lectin	Q9LED8	*P. lunatus*	96.31	TrEMBL
Lectin 2	P93456	*P. lunatus*	11.68	TrEMBL
Lectin 1	P93455	*P. lunatus*	5.05	TrEMBL
PHA E	P05088	*P. vulgaris*	3.96	Non-redundant

Table 8: Lectin identified from *P. lunatus* with protein database search. No lectin identified for with Swiss-Prot search but lectin found with TrEMBL with minimum peptide and sequence coverage.

Protein name	Accession number	Species	Log probability	Database
Alpha-amylase inhibitor	P02873	*P. vulgaris*	3.48	Non-redundant
Lectin	P02867	*P. sativum*	12.40	Non-redundant
PHA L	P05087	*P. vulgaris*	20.46	Non-redundant
Lectin	B5A8N6	*P. sativum*	43.93	TrEMBL
PHA E	P05088	*P. vulgaris*	46.39	Non-redundant

Table 9: Lectin identified from *P. sativum* with protein database search. No lectin identified for with Swiss-Prot search but lectin found with TrEMBL with minimum peptide and sequence coverage.

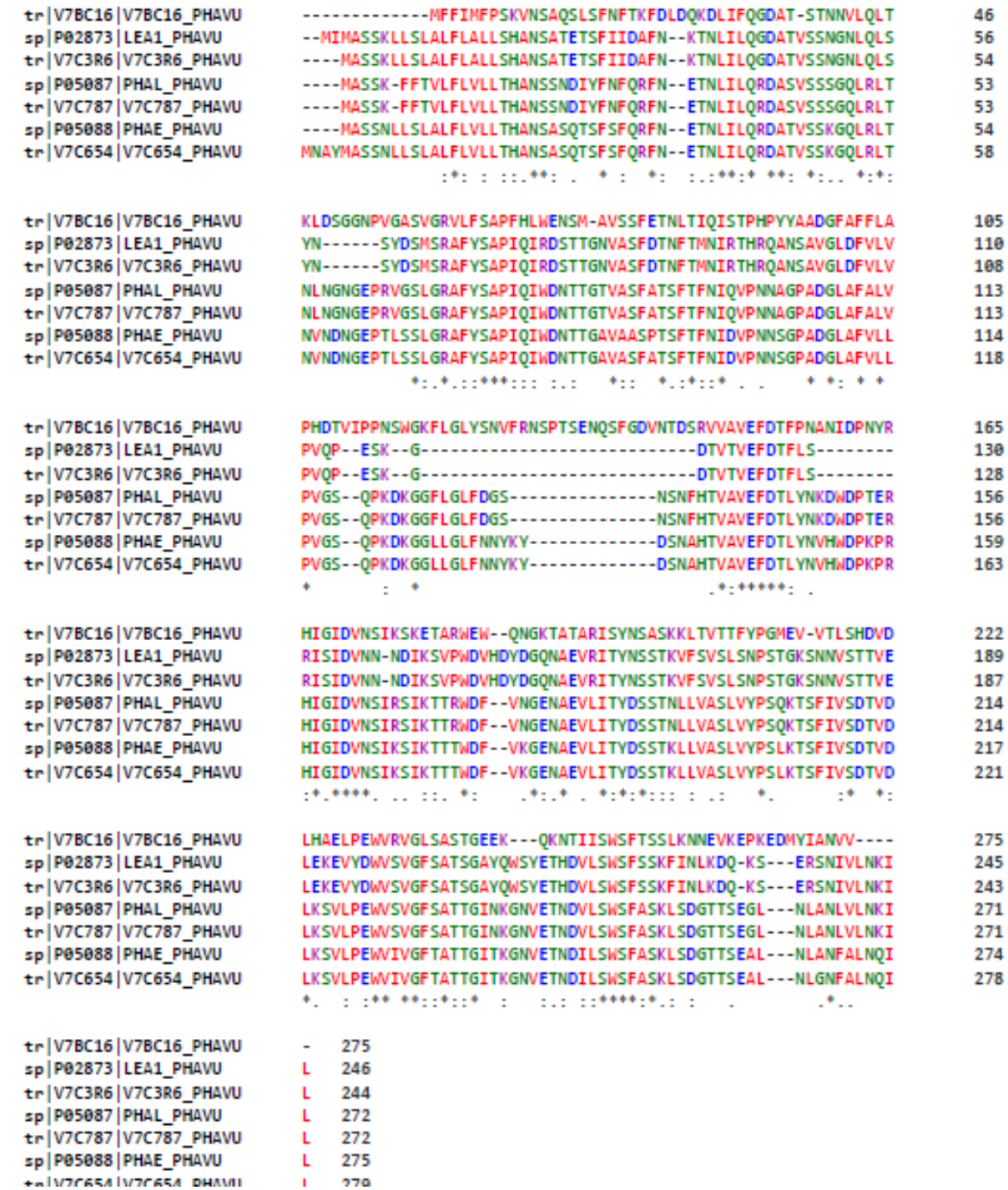


Figure 15: Sequence Homology analysis of lectins identified with different database search for *P. vulgaris*

Figure 16: Sequence Homology analysis of lectins identified with different database search for *L. culinaris*

Figure 17: Sequence Homology analysis of lectins identified with different database search for *P. lunatus*

Figure 18: Sequence alignment for lectin protein from three different species *P. vulgaris, C. cajan, V. radiata*

Figure 19: Sequence Homology analysis of lectins identified with different database search for *P. sativum*

LIST OF REFERENCES

Abolhasani Khaje, N., C. K. Mobley, S. K. Misra, L. Miller, Z. Li, E. Nudler, and J. S. Sharp. "Variation in Fpop Measurements Is Primarily Caused by Poor Peptide Signal Intensity." *J Am Soc Mass Spectrom* 29, no. 9 (2018): 1901-07.

Agrawal, P., S. Kumar, and H. R. Das. "Mass Spectrometric Characterization of Isoform Variants of Peanut (Arachis Hypogaea) Stem Lectin (Sl-I)." *J Proteomics* 73, no. 8 (2010): 1573-86.

Aslam, Bilal, Madiha Basit, Muhammad Atif Nisar, Muhammad Hidayat Rasool, and Mohsin Khurshid. "Proteomics: Technologies and Their Applications." *Journal of Chromatographic Science* 55, no. 2 (2017): 182-96.

Aye, T. T., T. Y. Low, and S. K. Sze. "Nanosecond Laser-Induced Photochemical Oxidation Method for Protein Surface Mapping with Mass Spectrometry." *Anal Chem* 77, no. 18 (2005): 5814-22.

Bantscheff, Marcus, Markus Schirle, Gavain Sweetman, Jens Rick, and Bernhard Kuster. "Quantitative Mass Spectrometry in Proteomics: A Critical Review." *Analytical and Bioanalytical Chemistry* 389, no. 4 (October 01 2007): 1017-31.

Barrientos, Laura G, and Angela M Gronenborn. "The Highly Specific Carbohydrate-Binding Protein Cyanovirin-N: Structure, Anti-Hiv/Ebola Activity and Possibilities for Therapy." *Mini Rev Med Chem* 5, no. 1 (2005): 21-31.

Bateman, Alex, Andreas Heger, Erik L. L. Sonnhammer, Goran Ceric, Jaina Mistry, Joanne E. Pollington, John Tate*, et al.* "The Pfam Protein Families Database." *Nucleic Acids Res* 38, no. suppl_1 (2009): D211-D22.

Bouckaert, J., T. Hamelryck, L. Wyns, and R. Loris. "Novel Structures of Plant Lectins and Their Complexes with Carbohydrates." *Curr Opin Struct Biol* 9, no. 5 (1999): 572-7.

Boyd, WC, and RM Reguera. "Studies on Haemagglutinins Present in Seeds of Some Representatives of the Family Leguminoseae." *J Immunol* 62 (1949): 333-39.

Breitenbach Barroso Coelho, L. C., P. Marcelino Dos Santos Silva, W. Felix de Oliveira, M. C. de Moura, E. Viana Pontual, F. Soares Gomes, P. M. Guedes Paiva, T. H. Napoleao, and M. T. Dos Santos Correia. "Lectins as Antimicrobial Agents." *J Appl Microbiol* 125, no. 5 (2018): 1238-52.

Brenowitz, M., M. R. Chance, G. Dhavan, and K. Takamoto. "Probing the Structural Dynamics of Nucleic Acids by Quantitative Time-Resolved and Equilibrium Hydroxyl Radical "Footprinting"." *Curr Opin Struct Biol* 12, no. 5 (2002): 648-53.

Campos-Vega, Rocio, Guadalupe Loarca-Piña, and B Dave Oomah. "Minor Components of Pulses and Their Potential Impact on Human Health." *Food research international* 43, no. 2 (2010): 461-82.

Chance, M. R. "Unfolding of Apomyoglobin Examined by Synchrotron Footprinting." *Biochem Biophys Res Commun* 287, no. 3 (2001): 614-21.

Charvatova, O., 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, no. 11 (2008): 1692-705.

Chea, E. E., and L. M. Jones. "Analyzing the Structure of Macromolecules in Their Native Cellular Environment Using Hydroxyl Radical Footprinting." *Analyst* 143, no. 4 (2018): 798-807.

———. "Modifications Generated by Fast Photochemical Oxidation of Proteins Reflect the Native Conformations of Proteins." *Protein Sci* 27, no. 6 (2018): 1047-56.

Chibbar, Ravindra N, Priyatharini Ambigaipalan, and Ratnajothi Hoover. "Molecular Diversity in Pulse Seed Starch and Complex Carbohydrates and Its Role in Human Nutrition and Health." *Cereal chemistry* 87, no. 4 (2010): 342-52.

Chrispeels, M. J., and N. V. Raikhel. "Lectins, Lectin Genes, and Their Role in Plant Defense." *Plant Cell* 3, no. 1 (1991): 1-9.

Dai, S. Y., M. W. Gardner, and M. C. Fitzgerald. "Protocol for the Thermodynamic Analysis of Some Proteins Using an H/D Exchange- and Mass Spectrometry-Based Technique." *Anal Chem* 77, no. 2 (2005): 693-7.

De Hoff, Peter L, Laurence M Brill, and Ann M Hirsch. "Plant Lectins: The Ties That Bind in Root Symbiosis and Plant Defense." *Molecular genetics and genomics* 282, no. 1 (2009): 1-15.

Deutsch, Eric W, Henry Lam, and Ruedi Aebersold. "Data Analysis and Bioinformatics Tools for Tandem Mass Spectrometry in Proteomics." *Physiological genomics* 33, no. 1 (2008): 18-25.

Dias Rde, O., S. Machado Ldos, L. Migliolo, and O. L. Franco. "Insights into Animal and Plant Lectins with Antimicrobial Activities." *Molecules* 20, no. 1 (2015): 519-41.

Domon, B., and R. Aebersold. "Mass Spectrometry and Protein Analysis." *Science* 312, no. 5771 (2006): 212-7.

Durowoju, I. B., K. S. Bhandal, J. Hu, B. Carpick, and M. Kirkitadze. "Differential Scanning Calorimetry - a Method for Assessing the Thermal Stability and Conformation of Protein Antigen." *J Vis Exp* 4, no. 121 (2017): 55262.

Ellis, R John, and Teresa JT Pinheiro. "Medicine: Danger—Misfolding Proteins." *Nature* 416, no. 6880 (2002): 483.

Espino, J. A., V. S. Mali, and L. M. Jones. "In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells." *Anal Chem* 87, no. 15 (2015): 7971-8.

Ganesan, K., and B. Xu. "Polyphenol-Rich Lentils and Their Health Promoting Effects." *Int J Mol Sci* 18, no. 11 (2017).

Gau, B. C., J. Chen, and M. L. Gross. "Fast Photochemical Oxidation of Proteins for Comparing Solvent-Accessibility Changes Accompanying Protein Folding: Data Processing and Application to Barstar." *Biochim Biophys Acta* 6, no. 8 (2013): 26.

Gau, B. C., J. S. Sharp, D. L. Rempel, and M. L. Gross. "Fast Photochemical Oxidation of Protein Footprints Faster Than Protein Unfolding." *Anal Chem* 81, no. 16 (2009): 6563-71.

Gau, B., K. Garai, C. Frieden, and M. L. Gross. "Mass Spectrometry-Based Protein Footprinting Characterizes the Structures of Oligomeric Apolipoprotein E2, E3, and E4." *Biochemistry* 50, no. 38 (2011): 8117-26.

Gnanesh Kumar, B. S., and Avadhesha Surolia. "Identification of Banana Lectin Isoforms and Differential Acetylation through Mass Spectrometry Approaches." *The Protein Journal* 37, no. 1 (February 01 2018): 38-46.

Goldstein, I. J., and C. E. Hayes. "The Lectins: Carbohydrate-Binding Proteins of Plants and Animals." *Adv Carbohydr Chem Biochem* 35 (1978): 127-340.

Goncalves, A., P. Goufo, A. Barros, R. Dominguez-Perles, H. Trindade, E. A. Rosa, L. Ferreira, and M. Rodrigues. "Cowpea (Vigna Unguiculata L. Walp), a Renewed Multipurpose Crop for a More Sustainable Agri-Food System: Nutritional Advantages and Constraints." *J Sci Food Agric* 96, no. 9 (2016): 2941-51.

Gupta, S., R. Celestre, C. J. Petzold, M. R. Chance, and C. Ralston. "Development of a Microsecond X-Ray Protein Footprinting Facility at the Advanced Light Source." *J Synchrotron Radiat* 21, no. Pt 4 (2014): 690-9.

Hambly, D. M., and M. L. Gross. "Laser Flash Photolysis of Hydrogen Peroxide to Oxidize Protein Solvent-Accessible Residues on the Microsecond Timescale." *J Am Soc Mass Spectrom* 16, no. 12 (2005): 2057-63.

Hamid, Rabia, Akbar Masood, Ishfak H Wani, and Shaista Rafiq. "Lectins: Proteins with Diverse Applications." *J. Appl. Pharm. Sci* 3, no. 4.1 (2013): 93-103.

Hanai, Ryo, and James C Wang. "Protein Footprinting by the Combined Use of Reversible and Irreversible Lysine Modifications." *Proceedings of the National Academy of Sciences* 91, no. 25 (1994): 11904-08.

He, S., B. K. Simpson, H. Sun, M. O. Ngadi, Y. Ma, and T. Huang. "Phaseolus Vulgaris Lectins: A Systematic Review of Characteristics and Health Implications." *Crit Rev Food Sci Nutr* 58, no. 1 (2018): 70-83.

He, Shudong, John Shi, Xisheng Li, Ying Ma, and Sophia Jun Xue. "Identification of a Lectin Protein from Black Turtle Bean (Phaseolus Vulgaris) Using Lc-Ms/Ms and Pcr Method." *LWT-Food Science and Technology* 60, no. 2 (2015): 1074-79.

He, Shudong, John Shi, Elfalleh Walid, Ying Ma, and Sophia Jun Xue. "Extraction and Purification of a Lectin from Small Black Kidney Bean (Phaseolus Vulgaris) Using a Reversed Micellar System." *Process Biochemistry* 48, no. 4 (2013): 746-52.

He, Shudong, Benjamin K Simpson, Hanju Sun, Michael O Ngadi, Ying Ma, and Tiemin Huang. "Phaseolus Vulgaris Lectins: A Systematic Review of Characteristics and Health Implications." *Critical reviews in food science and nutrition* 58, no. 1 (2018): 70-83.

Hernandez, P., H. Debray, H. Jaekel, Y. Garfias, C. Jimenez Md Mdel, S. Martinez-Cairo, and E. Zenteno. "Chemical Characterization of the Lectin from Amaranthus Leucocarpus Syn. Hypocondriacus by 2-D Proteome Analysis." *Glycoconj J* 18, no. 4 (2001): 321-9.

Huang, R. Y., S. R. Krystek, Jr., N. Felix, R. F. Graziano, M. Srinivasan, A. Pashine, and G. Chen. "Hydrogen/Deuterium Exchange Mass Spectrometry and Computational Modeling Reveal a Discontinuous Epitope of an Antibody/Tl1a Interaction." *MAbs* 10, no. 1 (2018): 95-103.

Huang, W., K. M. Ravikumar, M. R. Chance, and S. Yang. "Quantitative Mapping of Protein Structure by Hydroxyl Radical Footprinting-Mediated Structural Mass Spectrometry: A Protection Factor Analysis." *Biophys J* 108, no. 1 (2015): 107-15.

Ishihama, Y., T. Schmidt, J. Rappsilber, M. Mann, F. U. Hartl, M. J. Kerner, and D. Frishman. "Protein Abundance Profiling of the Escherichia Coli Cytosol." *BMC Genomics* 9, no. 102 (2008): 1471-2164.

Jumper, C. C., R. Bomgarden, J. Rogers, C. Etienne, and D. C. Schriemer. "High-Resolution Mapping of Carbene-Based Protein Footprints." *Anal Chem* 84, no. 10 (2012): 4411-8.

Jumper, C. C., and D. C. Schriemer. "Mass Spectrometry of Laser-Initiated Carbene Reactions for Protein Topographic Analysis." *Anal Chem* 83, no. 8 (2011): 2913-20.

Kader, Jean-Claude, and Michel Delseny. *Advances in Botanical Research*. Vol. 60: Academic Press, 2011.

Katta, Viswanatham, Brian T Chait, and Steven Carr. "Conformational Changes in Proteins Probed by Hydrogen-Exchange Electrospray-Ionization Mass Spectrometry." *Rapid Communications in Mass Spectrometry* 5, no. 4 (1991): 214-17.

Kaur, P., J. Kiselar, S. Yang, and M. R. Chance. "Quantitative Protein Topography Analysis and High-Resolution Structure Prediction Using Hydroxyl Radical Labeling and Tandem-Ion Mass Spectrometry (Ms)." *Mol Cell Proteomics* 14, no. 4 (2015): 1159-68.

Kiselar, J. G., and M. R. Chance. "Future Directions of Structural Mass Spectrometry Using Hydroxyl Radical Footprinting." *J Mass Spectrom* 45, no. 12 (2010): 1373-82.

Kiselar, J. G., P. A. Janmey, S. C. Almo, and M. R. Chance. "Structural Analysis of Gelsolin Using Synchrotron Protein Footprinting." *Mol Cell Proteomics* 2, no. 10 (2003): 1120-32.

———. "Visualizing the Ca2+-Dependent Activation of Gelsolin by Using Synchrotron Footprinting." *Proc Natl Acad Sci U S A* 100, no. 7 (2003): 3942-7.

Kong, J., and S. Yu. "Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures." *Acta Biochim Biophys Sin* 39, no. 8 (2007): 549-59.

Kumar, Gnanesh, Padmanabh Mishra, Vellareddy Anantharam, and Avadhesha Surolia. "Luffa Acutangula Agglutinin: Primary Structure Determination and Identification of a Tryptophan Residue Involved in Its Carbohydrate-Binding Activity Using Mass Spectrometry." *IUBMB Life* 67, no. 12 (2015): 943-53.

Lagarda-Diaz, I., A. M. Guzman-Partida, and L. Vazquez-Moreno. "Legume Lectins: Proteins with Diverse Applications." *Int J Mol Sci* 18, no. 6 (2017).

Lakhtin, Vladimir, Mikhail Lakhtin, and Vladimir Alyoshkin. "Lectins of Living Organisms. The Overview." *Anaerobe* 17, no. 6 (2011): 452-55.

Lam, Sze Kwan, and Tzi Bun Ng. "Lectins: Production and Practical Applications." *Applied microbiology and biotechnology* 89, no. 1 (2011): 45-55.

Leavitt, R. D., R. L. Felsted, and N. R. Bachur. "Biological and Biochemical Properties of Phaseolus Vulgaris Isolectins." *J Biol Chem* 252, no. 9 (1977): 2961-6.

Li, C. H., X. Nguyen, L. Narhi, L. Chemmalil, E. Towers, S. Muzammil, J. Gabrielson, and Y. Jiang. "Applications of Circular Dichroism (Cd) for Structural Analysis of Proteins: Qualification

of near- and Far-Uv Cd for Protein Higher Order Structural Analysis." *J Pharm Sci* 100, no. 11 (2011): 4642-54.

Li, J., and G. Chen. "The Use of Fast Photochemical Oxidation of Proteins Coupled with Mass Spectrometry in Protein Therapeutics Discovery and Development." *Drug Discov Today* 24, no. 3 (2019): 829-34.

Li, K. S., D. L. Rempel, and M. L. Gross. "Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1-42 Aggregation." *J Am Chem Soc* 138, no. 37 (2016): 12090-8.

Li, Z., H. Moniz, S. Wang, A. Ramiah, F. Zhang, K. W. Moremen, R. J. Linhardt, and J. S. Sharp. "High Structural Resolution Hydroxyl Radical Protein Footprinting Reveals an Extended Robo1-Heparin Binding Interface." *J Biol Chem* 290, no. 17 (2015): 10729-40.

Lombardi-Boccia, G., S. Ruggeri, A. Aguzzi, and M. Cappelloni. "Globulins Enhance in Vitro Iron but Not Zinc Dialysability: A Study on Six Legume Species." *J Trace Elem Med Biol* 17, no. 1 (2003): 1-5.

Macedo, Maria Lígia Rodrigues, Daniela Carla S Damico, Maria das Graças Machado Freire, Marcos H Toyama, Sérgio Marangoni, and José C Novello. "Purification and Characterization of an N-Acetylglucosamine-Binding Lectin from Koelreuteria Paniculata Seeds and Its Effect on the Larval Development of Callosobruchus Maculatus (Coleoptera: Bruchidae) and Anagasta Kuehniella (Lepidoptera: Pyralidae)." *Journal of agricultural and food chemistry* 51, no. 10 (2003): 2980-86.

Maleknia, S. D., M. R. Chance, and K. M. Downard. "Electrospray-Assisted Modification of Proteins: A Radical Probe of Protein Structure." *Rapid Commun Mass Spectrom* 13, no. 23 (1999): 2352-8.

Maleknia, S. D., and K. M. Downard. "Unfolding of Apomyoglobin Helices by Synchrotron Radiolysis and Mass Spectrometry." *Eur J Biochem* 268, no. 21 (2001): 5578-88.

Maleknia, S. D., C. Y. Ralston, M. D. Brenowitz, K. M. Downard, and M. R. Chance. "Determination of Macromolecular Folding and Structure by Synchrotron X-Ray Radiolysis Techniques." *Anal Biochem* 289, no. 2 (2001): 103-15.

Manning, M. C., D. K. Chou, B. M. Murphy, R. W. Payne, and D. S. Katayama. "Stability of Protein Pharmaceuticals: An Update." *Pharm Res* 27, no. 4 (2010): 544-75.

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer*, et al.* "Cdd: Ncbi's Conserved Domain Database." *Nucleic Acids Res* 43, no. Database issue (2015): 20.

Matthiesen, Rune, Luisa Azevedo, Antonio Amorim, and Ana Sofia Carvalho. "Discussion on Common Data Analysis Strategies Used in Ms-Based Proteomics." *Proteomics* 11, no. 4 (2011): 604-19.

Menard, S, N Cerf-Bensussan, and M Heyman. "Multiple Facets of Intestinal Permeability and Epithelial Handling of Dietary Antigens." *Mucosal immunology* 3, no. 3 (2010): 247.

Mun, Ji-Young, Kyung Jin Lee, Yu Jin Kim, Ohsuk Kwon, Su-Jin Kim, Seung-Goo Lee, Wei Sun Park, Won Do Heo, and Doo-Byoung Oh. "Development of Fluorescent Probes for the Detection

of Fucosylated N-Glycans Using an Aspergillus Oryzae Lectin." *Applied microbiology and biotechnology* 93, no. 1 (2012): 251-60.

Naganathan, Athi N, and Victor Muñoz. "Scaling of Folding Times with Protein Size." *J Am Chem Soc* 127, no. 2 (2005): 480-81.

Nowell, P. C. "Phytohemagglutinin: An Initiator of Mitosis in Cultures of Normal Human Leukocytes." *Cancer Res* 20 (1960): 462-6.

Ofuya, ZM, and V Akhidue. "The Role of Pulses in Human Nutrition: A Review." *Journal of Applied Sciences and Environmental Management* 9, no. 3 (2005): 99-104.

Peumans, Willy J, and EJ Van Damme. "Lectins as Plant Defense Proteins." *Plant physiology* 109, no. 2 (1995): 347.

Plath, F., P. Ringler, A. Graff-Meyer, H. Stahlberg, M. E. Lauer, A. C. Rufer, M. A. Graewert*, et al.* "Characterization of Mab Dimers Reveals Predominant Dimer Forms Common in Therapeutic Mabs." *MAbs* 8, no. 5 (2016): 928-40.

Rangel, A., G. B. Domont, C. Pedrosa, and S. T. Ferreira. "Functional Properties of Purified Vicilins from Cowpea (Vigna Unguiculata) and Pea (Pisum Sativum) and Cowpea Protein Isolate." *J Agric Food Chem* 51, no. 19 (2003): 5792-7.

Renkonen, KO. "Studies on Hemagglutinins Present in Seeds of Some Representatives of the Family of Leguminoseae." Paper presented at the Annales medicinae experimentalis et biologiae fenniae, 1948.

Riaz, M., S. K. Misra, and J. S. Sharp. "Towards High-Throughput Fast Photochemical Oxidation of Proteins: Quantifying Exposure in High Fluence Microtiter Plate Photolysis." *Anal Biochem* 562 (2018): 32-36.

Rinas, A., V. S. Mali, J. A. Espino, and L. M. Jones. "Development of a Microflow System for in-Cell Footprinting Coupled with Mass Spectrometry." *Anal Chem* 88, no. 20 (2016): 10052-58.

Roy, F, JI Boye, and BK Simpson. "Bioactive Proteins and Peptides in Pulse Crops: Pea, Chickpea and Lentil." *Food research international* 43, no. 2 (2010): 432-42.

Rubinstein, Natalia, Juan Martin Ilarregui, Marta Alicia Toscano, and Gabriel Adrián Rabinovich. "The Role of Galectins in the Initiation, Amplification and Resolution of the Inflammatory Response." *Tissue antigens* 64, no. 1 (2004): 1-12.

Sharma, V., and A. Surolia. "Analyses of Carbohydrate Recognition by Legume Lectins: Size of the Combining Site Loops and Their Primary Specificity." *J Mol Biol* 267, no. 2 (1997): 433-45.

Sharon, N., and H. Lis. "How Proteins Bind Carbohydrates: Lessons from Legume Lectins." *J Agric Food Chem* 50, no. 22 (2002): 6586-91.

Sharon, Nathan, and Halina Lis. "Legume Lectins--a Large Family of Homologous Proteins." *The FASEB Journal* 4, no. 14 (1990): 3198-208.

Sharp, J. S., J. M. Becker, and R. L. Hettich. "Analysis of Protein Solvent Accessible Surfaces by Photochemical Oxidation and Mass Spectrometry." *Anal Chem* 76, no. 3 (2004): 672-83.

———. "Protein Surface Mapping by Chemical Oxidation: Structural Analysis by Mass Spectrometry." *Anal Biochem* 313, no. 2 (2003): 216-25.

Sharp, J. S., S. K. Misra, J. J. Persoff, R. W. Egan, and S. R. Weinberger. "Real Time Normalization of Fast Photochemical Oxidation of Proteins Experiments by Inline Adenine Radical Dosimetry." *Anal Chem* 90, no. 21 (2018): 12625-30.

Shevchenko, A., H. Tomas, J. Havlis, J. V. Olsen, and M. Mann. "In-Gel Digestion for Mass Spectrometric Characterization of Proteins and Proteomes." *Nat Protoc* 1, no. 6 (2006): 2856-60.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez*, et al.* "Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega." *Mol Syst Biol* 7, no. 539 (2011): 75.

Silva, Mariana CC, Lucimeire A Santana, Reinhard Mentele, Rodrigo S Ferreira, Antônio de Miranda, Rosemeire A Silva-Lucca, Misako U Sampaio, Maria TS Correia, and Maria LV Oliva. "Purification, Primary Structure and Potential Functions of a Novel Lectin from Bauhinia Forficata Seeds." *Process Biochemistry* 47, no. 7 (2012): 1049-59.

Singh, U, and R Jambunathan. "Distribution of Seed Protein Fractions and Amino Acids in Different Anatomical Parts of Chickpea (Cicer Arietinum L.) and Pigeonpea (Cajanus Cajan L.)." *Plant Foods for Human Nutrition* 31, no. 4 (1982): 347-54.

Soto, C. "Protein Misfolding and Disease; Protein Refolding and Therapy." *FEBS Lett* 498, no. 2-3 (2001): 204-7.

Steiner, Robert F, Sharon Albaugh, Catherine Fenselau, Constance Murphy, and Martha Vestling. "A Mass Spectrometry Method for Mapping the Interface Topography of Interacting

Proteins, Illustrated by the Melittin-Calmodulin System." *Analytical biochemistry* 196, no. 1 (1991): 120-25.

Storek, K. M., M. R. Auerbach, H. Shi, N. K. Garcia, D. Sun, N. N. Nickerson, R. Vij*, et al.* "Monoclonal Antibody Targeting the Beta-Barrel Assembly Machine of Escherichia Coli Is Bactericidal." *Proc Natl Acad Sci U S A* 115, no. 14 (2018): 3692-97.

Takamoto, K., and M. R. Chance. "Radiolytic Protein Footprinting with Mass Spectrometry to Probe the Structure of Macromolecular Complexes." *Annu Rev Biophys Biomol Struct* 35 (2006): 251-76.

Tang, Dongyan, Yinmao Dong, Hankun Ren, Li Li, and Congfen He. "A Review of Phytochemistry, Metabolite Changes, and Medicinal Uses of the Common Food Mung Bean and Its Sprouts (Vigna Radiata)." *Chemistry Central Journal* 8, no. 1 (2014): 4.

Vahidi, S., B. B. Stocks, Y. Liaghati-Mobarhan, and L. Konermann. "Submillisecond Protein Folding Events Monitored by Rapid Mixing and Mass Spectrometry-Based Oxidative Labeling." *Anal Chem* 85, no. 18 (2013): 8618-25.

Van Damme, Els JM, Willy J Peumans, Arpad Pusztai, and Susan Bardocz. *Handbook of Plant Lectins: Properties and Biomedical Applications*. John Wiley & Sons, 1998.

Venkidasamy, Baskar, Dhivya Selvaraj, Arti Shivraj Nile, Sathishkumar Ramalingam, Guoyin Kai, and Shivraj Hariram Nile. "Indian Pulses: A Review on Nutritional, Functional and Biochemical Properties with Future Perspectives." *Trends in Food Science & Technology* (2019).

Vijayan, M., and N. Chandra. "Lectins." *Curr Opin Struct Biol* 9, no. 6 (1999): 707-14.

Wang, L., and M. R. Chance. "Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment." *Mol Cell Proteomics* 16, no. 5 (2017): 706-16.

———. "Structural Mass Spectrometry of Proteins Using Hydroxyl Radical Based Protein Footprinting." *Anal Chem* 83, no. 19 (2011): 7234-41.

Watkinson, T. G., A. N. Calabrese, J. R. Ault, S. E. Radford, and A. E. Ashcroft. "Fpop-Lc-Ms/Ms Suggests Differences in Interaction Sites of Amphipols and Detergents with Outer Membrane Proteins." *J Am Soc Mass Spectrom* 28, no. 1 (2017): 50-55.

Watson, C., I. Janik, T. Zhuang, O. Charvatova, R. J. Woods, and J. S. Sharp. "Pulsed Electron Beam Water Radiolysis for Submicrosecond Hydroxyl Radical Protein Footprinting." *Anal Chem* 81, no. 7 (2009): 2496-505.

Watson, C., and J. S. Sharp. "Conformational Analysis of Therapeutic Proteins by Hydroxyl Radical Protein Footprinting." *Aaps J* 14, no. 2 (2012): 206-17.

Wiederschain, G. Ya. "Essentials of Glycobiology." *Biochemistry (Moscow)* 74, no. 9 (September 01 2009): 1056-56.

Wiśniewski, Jacek R, and Matthias Mann. "Consecutive Proteolytic Digestion in an Enzyme Reactor Increases Depth of Proteomic and Phosphoproteomic Analysis." *Analytical chemistry* 84, no. 6 (2012): 2631-37.

Xia, Lixin, and TB Ng. "A Hemagglutinin with Mitogenic Activity from Dark Red Kidney Beans." *Journal of Chromatography B* 844, no. 2 (2006): 213-16.

Xie, B., and J. S. Sharp. "Hydroxyl Radical Dosimetry for High Flux Hydroxyl Radical Protein Footprinting Applications Using a Simple Optical Detection Method." *Anal Chem* 87, no. 21 (2015): 10719-23.

Xie, B., A. Sood, R. J. Woods, and J. S. Sharp. "Quantitative Protein Topography Measurements by High Resolution Hydroxyl Radical Protein Footprinting Enable Accurate Molecular Model Selection." *Sci Rep* 7, no. 1 (2017): 017-04689.

Xu, G., and M. R. Chance. "Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics." *Chem Rev* 107, no. 8 (2007): 3514-43.

Zeng, Xiangqun, Cesar AS Andrade, Maria DL Oliveira, and Xue-Long Sun. "Carbohydrate–Protein Interactions and Their Biosensing Applications." *Analytical and Bioanalytical Chemistry* 402, no. 10 (2012): 3161-76.

Zhang, B., M. Cheng, D. Rempel, and M. L. Gross. "Implementing Fast Photochemical Oxidation of Proteins (Fpop) as a Footprinting Approach to Solve Diverse Problems in Structural Biology." *Methods* 144 (2018): 94-103.

Zhang, B., D. L. Rempel, and M. L. Gross. "Protein Footprinting by Carbenes on a Fast Photochemical Oxidation of Proteins (Fpop) Platform." *J Am Soc Mass Spectrom* 27, no. 3 (2016): 552-5.

Zhang, H., B. C. Gau, L. M. Jones, I. Vidavsky, and M. L. Gross. "Fast Photochemical Oxidation of Proteins for Comparing Structures of Protein-Ligand Complexes: The Calmodulin-Peptide Model System." *Anal Chem* 83, no. 1 (2011): 311-8.

Zhang, Y., A. T. Wecksler, P. Molina, G. Deperalta, and M. L. Gross. "Mapping the Binding Interface of Vegf and a Monoclonal Antibody Fab-1 Fragment with Fast Photochemical Oxidation of Proteins (Fpop) and Mass Spectrometry." *J Am Soc Mass Spectrom* 28, no. 5 (2017): 850-58.

Zhang, Zhongqi, and David L Smith. "Determination of Amide Hydrogen Exchange by Mass Spectrometry: A New Tool for Protein Structure Elucidation." *Protein Science* 2, no. 4 (1993): 522-31.

Zhu, Y., A. Serra, T. Guo, J. E. Park, Q. Zhong, and S. K. Sze. "Application of Nanosecond Laser Photolysis Protein Footprinting to Study Egfr Activation by Egf in Cells." *J Proteome Res* 16, no. 6 (2017): 2282-93.

APPENDIX

Species Name	Molecular Weight (kDa)	Database	Number of Proteins
Phaseolus vulgaris	25	Non-redundant Swiss-Prot	41
		Specie Specific Swiss Prot	7
		TrEMBL	257
	35	Non-redundant Swiss-Prot	18
		Specie Specific Swiss Prot	24
		TrEMBL	155
	45	Non-redundant Swiss-Prot	47
		Specie Specific Swiss Prot	6
		TrEMBL	233
Cajanus cajan	25	Non-redundant Swiss-Prot	28
		Specie Specific Swiss Prot	30
		TrEMBL	196
	35	Non-redundant Swiss-Prot	20

		Specie Specific Swiss Prot	16
		TrEMBL	177
	45	Non-redundant Swiss-Prot	43
		Specie Specific Swiss Prot	5
		TrEMBL	201
Vigna radiata	25	Non-redundant Swiss-Prot	33
		Specie Specific Swiss Prot	0
		TrEMBL	272
	35	Non-redundant Swiss-Prot	22
		Specie Specific Swiss Prot	0
		TrEMBL	213
	45	Non-redundant Swiss-Prot	44
		Specie Specific Swiss Prot	7
		TrEMBL	293
Lens culinaris	25	Non-redundant Swiss-Prot	38
		Specie Specific Swiss Prot	1

		TrEMBL	76
	35	Non-redundant Swiss-Prot	38
		Specie Specific Swiss Prot	1
		TrEMBL	77
	45	Non-redundant Swiss-Prot	42
		Specie Specific Swiss Prot	1
		TrEMBL	106
Vigna mungo	25	Non-redundant Swiss-Prot	34
		Specie Specific Swiss Prot	1
		TrEMBL	39
	35	Non-redundant Swiss-Prot	25
		Specie Specific Swiss Prot	0
		TrEMBL	52
	45	Non-redundant Swiss-Prot	7
		Specie Specific Swiss Prot	0

		TrEMBL	1
Cicer arietinum	25	Non-redundant Swiss-Prot	33
		Specie Specific Swiss Prot	4
		TrEMBL	189
	35	Non-redundant Swiss-Prot	26
		Specie Specific Swiss Prot	14
		TrEMBL	144
	45	Non-redundant Swiss-Prot	33
		Specie Specific Swiss Prot	8
		TrEMBL	205
Vigna unguiculata	25	Non-redundant Swiss-Prot	25
		Specie Specific Swiss Prot	18
		TrEMBL	53
	35	Non-redundant Swiss-Prot	24
		Specie Specific Swiss Prot	0

		TrEMBL	98
	45	Non-redundant Swiss-Prot	24
		Specie Specific Swiss Prot	4
		TrEMBL	66
Phaseolus lunatus	25	Non-redundant Swiss-Prot	21
		Specie Specific Swiss Prot	2
		TrEMBL	59
	35	Non-redundant Swiss-Prot	18
		Specie Specific Swiss Prot	2
		TrEMBL	58
	45	Non-redundant Swiss-Prot	28
		Specie Specific Swiss Prot	2
		TrEMBL	64
	25	Non-redundant Swiss-Prot	24

Phaseolus vulgaris var Pinto		Specie Specific Swiss Prot	17
		TrEMBL	193
	35	Non-redundant Swiss-Prot	19
		Specie Specific Swiss Prot	20
		TrEMBL	163
	45	Non-redundant Swiss-Prot	29
		Specie Specific Swiss Prot	4
		TrEMBL	173
Phaseolus vulgaris var Black turtle	25	Non-redundant Swiss-Prot	24
		Specie Specific Swiss Prot	27
		TrEMBL	191
	35	Non-redundant Swiss-Prot	22
		Specie Specific Swiss Prot	6
		TrEMBL	140
	45	Non-redundant Swiss-Prot	30

		Specie Specific Swiss Prot	30
		TrEMBL	172
Pisum sativum	25	Non-redundant Swiss-Prot	43
		Specie Specific Swiss Prot	37
		TrEMBL	122
	35	Non-redundant Swiss-Prot	32
		Specie Specific Swiss Prot	17
		TrEMBL	118
	45	Non-redundant Swiss-Prot	39
		Specie Specific Swiss Prot	33
		TrEMBL	107

Appendix 1: Total number of proteins identified with three different databases for various pulses.

VITA

Name Mohammad Riaz

Phone (740) 249-7686

E-mail riazlnam@gmail.com

Education

MS in Pharmaceutical Sciences 2019
University of Mississippi, University, MS 38677

MS in Clinical Pharmacy & Pharmacology 2013
University of Dhaka, Dhaka, Bangladesh

BS in Pharmacy 2012
University of Dhaka, Dhaka, Bangladesh

Professional Membership

2016- American Society of Mass Spectrometry

2019- American Association of Pharmaceutical Scientist

Presentations:

Title: Characterizing the commercially available whey proteins through mass spectrometry methods
Type: Poster
Conference: 67th ASMS Conference on Mass Spectrometry

Title: Towards High-Throughput Fast Photochemical Oxidation of Proteins: Quantifying Exposure in High Fluence Microtiter Plate Photolysis
Type: Poster
Conference: 66th ASMS Conference on Mass Spectrometry

Title: Development of high-throughput hydroxyl radical protein footprinting in a microtiter plate
Type: Poster
Conference: IdEA conference