January 2015

High Resolution Mass Spectrometric Approaches To Study Protein Structure and Environment in Lyophilized Solids

Lavanya Iyer
Purdue University

Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Lavanya Kannan Iyer

Entitled

High Resolution Mass Spectrometric Approaches To Study Protein Structure And Environment In Lyophilized Solids

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Elizabeth M. Topp
Chair

Gregory T. Knipp

Lynne S. Taylor

Steven L. Nail

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

Approved by Major Professor(s): Elizabeth M. Topp

Approved by: Elizabeth M. Topp 8/28/2015

Head of the Departmental Graduate Program Date

HIGH-RESOLUTION MASS SPECTROMETRIC APPROACHES TO STUDY PROTEIN STRUCTURE AND ENVIRONMENT IN LYOPHILIZED SOLIDS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Lavanya K Iyer

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

For my parents and grandparents

ACKNOWLEDGEMENTS

I would like to thank my major advisor, Dr. Elizabeth M. Topp. She has been an excellent mentor. I appreciate the freedom and encouragement she gave me to explore different ideas. I will always value the time I spent in her group, learning about great science. I thank my committee members, Dr. Greg Knipp and Dr. Lynne Taylor for their scientific insight and am indebted to Dr. Steve Nail, who gave me the opportunity to intern at Baxter Biosciences. It was a wonderful learning experience.

I thank Dr. Andreas Sophocleous, who was a great source of inspiration when I started my graduate work. His scientific expertise, strong work ethic and willingness to share knowledge are qualities that I try to incorporate in my life. Many thanks to Dr. Shenbaga Moorthy Balakrishnan for being a great teacher and for interesting discussions about science and cricket. I am grateful to Dr. Saradha Chandrasekhar, Dr. Jun Zhang, Ehab Moussa and Anshul Mishra for being excellent lab mates. I would also like to thank Mary Ellen Hurt, Nancy Cramer and Janine Mott, for making graduate life smoother.

I will forever be grateful to my parents, Geetha and Kannan and my sister Madhuri for their unconditional love and support throughout life. They inspire me to work hard, be honest and be brave enough to try new things. I owe a great deal to my grandparents, V. Ramaswamy and Kalyani Ramaswamy for reminding me to have faith, continue learning and be a good person. Special thanks to my aunt, Lalitha Shankar, for being a pillar of strength. I want to express my gratitude to my wonderful parents-in-law, Minaxi and Piyushkumar Panchal, whose

encouragement kept me going. Finally, I would like to thank my husband, Jainik Panchal, for his endless patience, love and support.

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF ABBREVIATIONS

ApoMb, Apomyoglobin

CaM, Calmodulin

CID, Collision Induced Dissociation

ESI, Electrospray Ionization

FA, Formic Acid

FTIR spectroscopy, Fourier-Transform Infrared Spectroscopy

GCG, Glucagon

HDX-MS, Hydrogen-Deuterium Exchange-Mass Spectrometry

HoloMb, Holomyoglobin

HPLC, High Performance Liquid Chromatography

Mb, Myoglobin

MS, Mass Spectrometry

MW, Molecular Weight

MWCO, Molecular Weight Cutoff

PAA, Photoactive Amino acid Analog

pBpA, p-benzoyl-L-phenylalanine

pLeu, L-2 amino 4,4' azipentanoate (Photo-leucine)

PL-MS, Photolytic Labeling- Mass Spectrometry

SDA, succinimidyl 4,4' azipentanoate

SS, Solid State

UV, Ultraviolet

ABSTRACT

Iyer, Lavanya K. PhD, Purdue University, December 2015. High-Resolution Mass Spectrometric Approaches to Study Protein Structure and Environment in Lyophilized Solids. Major Professor: Elizabeth Topp.

Proteins comprise a growing class of therapeutics that is used to treat various diseases such as diabetes and cancer. However, intrinsic structural features such as the primary sequence and extrinsic factors such as pH, temperature, agitation and metal ions can promote instability that manifests as chemical degradation (e.g. oxidation, deamidation, hydrolysis) and/or physical degradation (aggregation, phase separation). Since several degradation pathways are accelerated by diffusion in solution, proteins are lyophilized to improve stability. The lyophilized formulation may still undergo degradation during manufacture and/or storage. The mechanism of protein aggregation in lyophilized solids is not well understood or predictable by conventional analytical methods such as solid-state Fourier-transform infrared spectroscopy (ssFTIR) and differential scanning calorimetry (DSC) and this poses challenges in rational formulation design.

This dissertation is aimed at understanding local protein structure and environment in the solid state using high-resolution mass spectrometric methods. Chapter 2 examines protein side-chain matrix accessibility using solid-state photolytic labeling- mass spectrometry (ssPL-MS). The use of a photoactive probe, photo-leucine (pLeu) enabled side-chain labeling in lyophilized formulations, reported by our group for the first time. High-resolution information at the peptide level was obtained using bottom-up tandem mass spectrometry. Differences in labeling patterns and side-chain matrix accessibility were observed when sucrose or guanidine hydrochloride was used as an excipient. This work also used a photoactive probe incorporated within the amino acid sequence of a

glucagon-derived peptide to detect interactions with excipients and peptides in the solid state. Residue-level information about the preferred site of peptide-peptide crosslinking was obtained using tandem mass spectrometry.

Although peptide-matrix interactions could be visualized using a photoactive amino acid (PAA) derivative within the primary sequence, incorporating an unnatural amino acid into larger proteins is fairly difficult and may alter higher order structure by disturbing intra-protein contacts. Therefore, a novel photo-crosslinking method was developed to further examine the solid-state environment of lyophilized proteins, described in Chapter 3. A heterobifunctional crosslinking reagent was used to crosslink the protein with the matrix in the solid state. Some loop regions showed increased peptide-peptide adducts, while helix E showed more hydration compared to other regions. In the presence of raffinose, water replacement was not detected in the solid state; instead there was some evidence of micro-phase separation without crystallization in the solid state. Thus local protein environment in the solid state could be probed without the need for PAA incorporation within the protein sequence.

Lyophilization is an effective, yet expensive stabilization strategy, since conservative freeze-drying cycles often require long hours of drying. The stochastic nature of ice nucleation and lack of control over freezing can result in vial-to-vial heterogeneity due to differences in the degree of supercooling and ice crystal size. The research described in Chapter 4 focuses on using a variety of analytical methods to characterize lyophilized protein formulations to determine the effect of excipient and freezing step on protein structure. Myoglobin in the presence or absence of sucrose was lyophilized with or without controlled ice nucleation in a pilot-scale LyoStar freeze dryer. Ice nucleation occurred over a range of temperatures and times with uncontrolled nucleation, while controlled ice nucleation with rapid depressurization resulted in near-simultaneous ice nucleation. The sucrose-containing formulation showed greater retention of protein structure by ssFTIR and solid-state hydrogen-deuterium exchange mass spectrometry (ssHDX-MS). Greater conformational

homogeneity was observed in the sucrose-containing formulation by ssHDX-MS peak width analysis. No significant differences in secondary structure were detected between controlled and uncontrolled nucleation using ssFTIR and ssHDX-MS. Myoglobin lyophilized with controlled nucleation in the presence of sucrose showed the greatest side-chain labeling, as determined by ssPL-MS. The results show that high-resolution mass spectrometric methods can be used to study process- and excipient effects on protein structure.

This thesis addresses limitations in current analytical methods used to characterize protein structure in the solid state. Whereas ssFTIR and DSC have lower sensitivity and provide information averaged over the entire sample, mass spectrometric methods can provide peptide-level information about conformational changes occurring in a small subpopulation of protein. High-resolution mass spectrometric methods have the potential to provide reliable and predictable protein formulation screening and facilitate rational drug design.

CHAPTER 1. INTRODUCTION

1.1 PROTEIN AGGREGATION IN THE SOLID STATE

Protein drugs are an increasingly important sector in the pharmaceutical market. In 2009, the ten best-selling protein drugs had a combined sales value of close to $50 billion [1]. The number of protein drugs on the market is expected to rise in the next few years, given the expiration of patents and growth of generic protein drugs or “biosimilars”. Protein instability is a major issue hampering formulation development, especially since each protein behaves uniquely in different environments. Formulation development for proteins is largely based on trial and error, making drug development very expensive and time-consuming. Many proteins are lyophilized to improve formulation stability. Although lyophilization confers greater stability on formulations compared to solution, degradation is known to occur in the solid state and during each step of the freeze-drying process [2-4]. Protein aggregation is a serious problem in the clinical setting because it can reduce efficacy and compromise safety. Aggregation is also critical for the pharmaceutical industry, because it complicates the manufacturing and formulation process. Therefore it is important to characterize these lyophilized proteins at the conformational level to ensure integrity of protein structure, especially for biosimilars.

Protein-side chains play an important role in the aggregation pathway. Intermolecular backbone and side-chain interactions facilitate the formation of amorphous aggregates [5, 6]. Site-directed mutagenesis and molecular dynamic simulation (MDS) studies on peptides have implicated hydrophobic interactions between aromatic side-chains and electrostatic interaction through salt-bridges in the formation and stabilization of amyloid fibrils [7-9]. Hence it becomes important to characterize the side-chain environment. The aggregation pathway is not well understood for

proteins in solution and lyophilized formulations. It is especially difficult to follow aggregation in amorphous solids because of the inherent structural and spatial heterogeneity. Our lack of understanding of solid-state protein aggregation is compounded by the absence of robust, high-resolution analytical methods, which makes it difficult to study protein stability at the molecular level. Traditional analytical techniques such as Fourier-Transform Infrared Spectroscopy (FTIR), Raman Spectroscopy and Differential Scanning Calorimetry (DSC) are semi-quantitative at best, suffer from low sensitivity and provide low-resolution information at the global level [10-12]. Solid state Nuclear Magnetic Resonance spectroscopy (ssNMR) can provide site-specific information about conformational changes [13]; however it requires extensive sample preparation using isotopic labeling and is also less sensitive to amorphous systems. Solid-state hydrogen-deuterium exchange in combination with mass spectrometry (ssHDX-MS) is an analytical tool that allows protein backbone environment to be probed with higher resolution. This technique has been used previously by our group to characterize formulations based on the amount of protection against exchange afforded by carbohydrate excipients [14].

The aim of this dissertation is to characterize lyophilized protein structure and environment with high resolution. In our ongoing research program, we have developed two novel analytical techniques: ssPL-MS and ssPC-MS to probe protein side-chain environment in lyophilized formulations. These techniques have been used in molecular biology to map the interactome within cells [15-17]. PL-MS has also been described for solution-state studies to probe protein topography and ligand binding [18, 19], while PC-MS is typically used to map the protein-protein interaction (PPI) interface in solutions [20, 21]. To our knowledge, our research is the first application of PL-MS and PC-MS to study molecular interactions in lyophilized solids. The research is significant to both industry and patients in several ways. These two high-resolution analytical methods can potentially identify local reactive sites participating in aggregation. Knowledge gained from these experiments can be used to design formulations rationally, by using excipients or chaperones that block reactive sites on the protein structure. Thus, protein aggregation can be controlled by designing an ideal solid-

state environment that promotes protein stability and minimizes aggregation. Our novel technique is expected to detect aggregate-prone regions earlier than conventional analytical methods, which is important in mitigating patient risk and improving formulations in the early stages of development. This will also reduce the cost of formulation development and time to reach market, thus lowering the burden of healthcare on consumers.

1.2 COVALENT LABELING OF PROTEINS

Covalent labeling of proteins refers to the modification of amino acids by reaction with side-chain groups. The labeling agent contains a functional group that is reactive towards specific or non-specific amino acids under certain conditions (e.g. alkaline pH or UV irradiation). Covalent labeling combined with mass spectrometry is a useful proteomics tool. It allows the side-chain environment to be mapped with high resolution and provides information about solvent (or matrix) accessibility of surface amino acids to the probe. The effect of excipients on protein tertiary structure can be determined by changes in labeling pattern. Labeling reagents may be classified as chemical or photolytic agents.

1.2.1 Chemical Labeling Agents

These probes undergo activation over a certain pH range and form covalent bonds with amino acid side chains in their proximity. The ratio of labeling agent to protein must be optimized to obtain a sufficient fraction of labeled protein without significantly perturbing protein structure. Previously used chemical agents include N-hydroxysuccinimide esters (NHS esters; amine specific), 2,3-butanedione (BD; Arg specific), N-alkylmaleimides (Cys specific) and diethylpyrocarbonate (DEPC; His, amine, hydroxyl specific). Although pH control ensures that the dominant reaction of the agent is with its target amino acid, sometimes side-reactions may occur and reduce the yield of the desired labeled amino acid. For example, BD is reactive towards Arg at pH 7-10; however it can also undergo photoactivation and react with Lys and His [22]. Hence, reaction conditions may need to be modified to minimize side-reactions.

Chemical labeling agents have been used to probe the structure of pre-aggregate species in solution. Mendoza et al. used three complementary labeling agents (NHSA, BD and DEPC) and examined the effect of β2 microglobulin (β2m) dimer formation on extent of amino acid modification in solution [19]. About one-third of the surface amino acids and about one-half of the amino acids in the dimer interface were probed. The change in reactivity and extent of labeling of amino acids with increasing dimer formation was indicative of a change in the side-chain environment. Covalent labeling combined with molecular dynamic simulations suggested that residues with the greatest change in modification are likely present at or near the dimer interface.

1.2.2 Photolytic Labeling Agents

These probes undergo activation of certain functional groups on exposure to UV light. Activation leads to the formation of short-lived, unstable radicals or neutral molecules with unpaired electrons. These species readily participate in insertion or addition reactions with neighboring molecules with the formation of a new covalent bond. Photolytic analogs of amino acids have been synthesized to study protein-protein interactions. These photolytic amino acids (PAAs) can be inserted into protein and peptide sequences through mutagenesis, translational incorporation during protein expression or solid-phase synthesis. The most common photoactive moieties in PAAs are arylazides, diazirines and benzophenones.

1.2.2.1 Arylazides

Arylazides are activated when exposed to UV light below 310 nm, forming reactive singlet nitrenes (lifetime ~ 1 ns) with expulsion of molecular nitrogen. Nitrenes can undergo ring expansion to form dehydroazepines that are particularly reactive towards nucleophilic amines and form covalent adducts. Nitrenes can also add to unsaturated bonds or insert into C-H and N-H bonds. The main disadvantage of crosslinking proteins with arylazides is possible damage to proteins at the activation wavelength (254 nm).

1.2.2.2 Diazirines

Diazirine-containing PAAs such as 2-amino 4,4' azipentanoic acid (photo-leucine, pLeu) and L-2-amino-5,5-azi-hexanoic acid (photo-methionine, pMet) undergo activation at higher wavelengths than arylazides. At 350-365 nm, the diazirine ring loses molecular nitrogen and forms an active carbene species. The lifetime of the carbene is very short (on the order of nanoseconds) and it undergoes insertion into any C-C and X-H bond (X= C, H, N, O, S) or addition on to a C=C bond in its immediate molecular cage (Fig. 1.1). Thus carbenes do not show preference for any particular amino acid and are expected to label any surface residue indiscriminately. Other reactions of carbenes include quenching by water to form a hydroxy derivative and self-interaction to form an alkene.

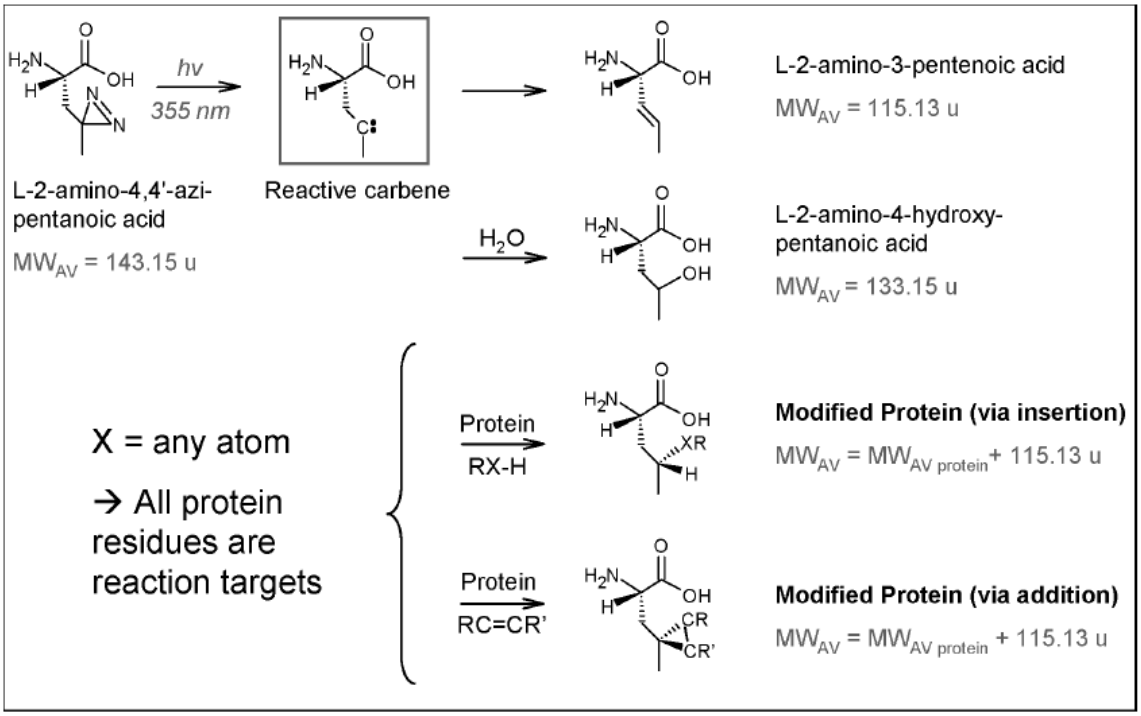


Figure 1.1. Schematic showing reactions of carbene formed upon activation of pLeu (reprinted with permission from 'Mass Spectrometry of Laser-Initiated Carbene Reactions for Protein Topographic Analysis [18] Copyright (2011) ACS).

Photolytic labeling- mass spectrometry (PL-MS) with pLeu has been reported in solution for myoglobin (Mb) and calmodulin (CaM) using a 1:10,000 molar ratio of protein to pLeu and a pulsed laser for irradiation [18]. CaM was detected carrying up to 4 labels, while Mb showed up to 2 labels. PL-MS was sensitive to changes in CaM conformation upon ligand binding, with ~ 39 % reduction in labeling for the ligand-bound protein compared to free CaM.

1.2.2.3 Benzophenones

In contrast to diazirines, benzophenone-containing PAAs (e.g. p-benzoyl L-phenylalanine; pBpA) appear to have greater affinity for electron-rich residues. Upon exposure to UV-A light, the carbonyl group on the benzophenone is activated to a diradicaloid triplet state (with a lifetime of 80-100 μs in the absence of an H-donor) [23]. The oxygen radical abstracts a hydrogen atom from a suitably oriented C-H group in its vicinity and forms a ketyl radical (Fig. 1.2). The hydrogen-deficient alkyl radical and the ketyl radical recombine to form a covalently bonded adduct. Effective H-donors include C-H bonds in Leu and Val and CH_2 groups adjacent to heteroatom containing amino acids like Met, Arg and Lys [23]. Unlike diazirines, activation of benzophenones is reversible. The diradicaloid species relaxes to its ground state in the absence of a suitably oriented H-donor. Thus, adduct formation may take a long time with several excitation-relaxation cycles until a favorable geometry is achieved.

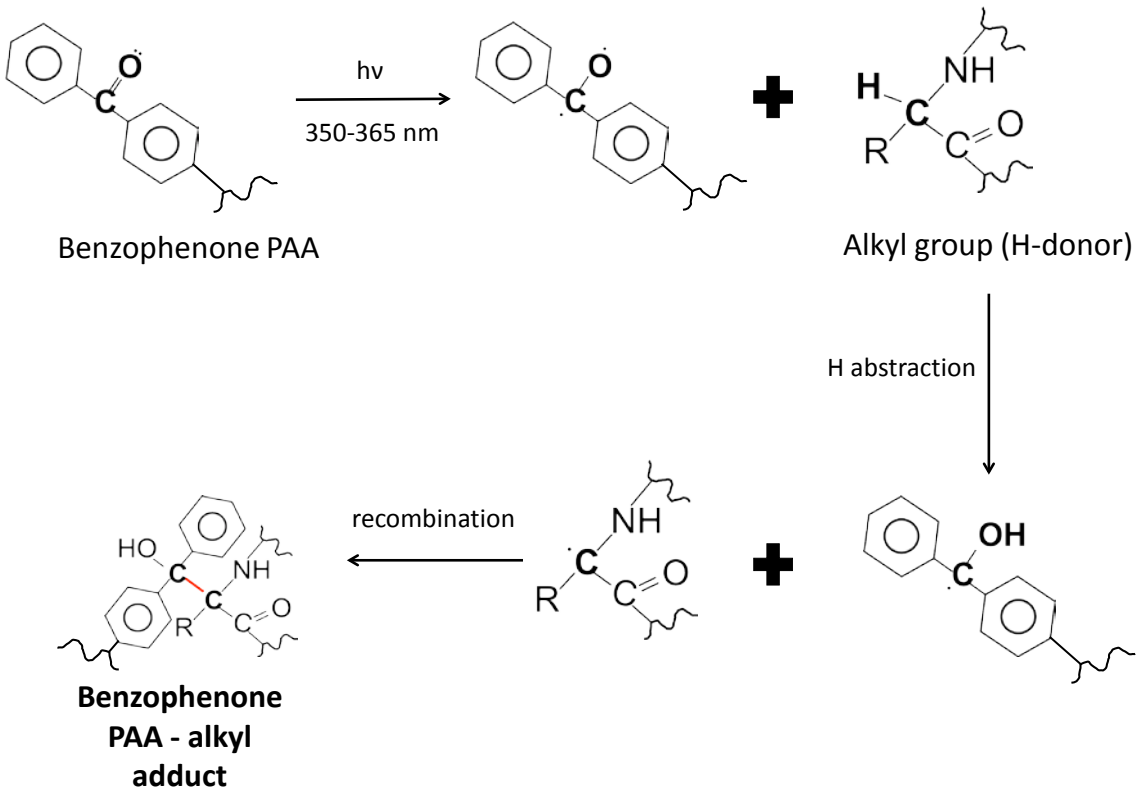


Figure 1.2. Schematic depicting covalent bond formation between a benzophenone- containing PAA and an amino acid on exposure to UV-A light (adapted from [23]).

1.3 CROSSLINKING OF PROTEINS

Crosslinking refers to the formation of a new intramolecular or intermolecular covalent bond between two amino acid side chains. This may be achieved by chemical and/or photolytic means. Crosslinkers may contain one functional group (e.g. 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride; EDC), but usually are homobifunctional or heterobifunctional. Homobifunctional crosslinkers (e.g. disuccinimidyl suberate; DSS) contain identical functional groups at each end of a spacer arm. They must be used in a single-step reaction since they react identically with their target group (e.g. amine to amine crosslinking), hence they are not very precise. Heterobifunctional crosslinkers (e.g. succinimidyl 4,4'- azipentanoate, SDA) have two different reactive groups at each end of a spacer arm and hence can be used to crosslink two specific functional groups. Heterobifunctional agents are used in a two-step process and offer better

precision than homobifunctional agents. Step 1 involves reaction of one end of the crosslinker with its target side chain (e.g. amine-reactive succinimidyl ester). After the reaction is complete, the excess unreacted crosslinker is removed by dialysis or desalting. Step 2 involves activation of the other end of the crosslinker (e.g. sulfhydryl-reactive N-alkylmaleimide) which results in formation of covalently linked adducts. Semi-specific labeling can be achieved by using a heterobifunctional crosslinker with a chemically reactive functionality at one end of the spacer and a photoactive functionality at the other end. For example, SDA contains an amine-reactive NHS-ester moiety at one end and a non-specific diazirine ring at the other end (Fig. 1.3).

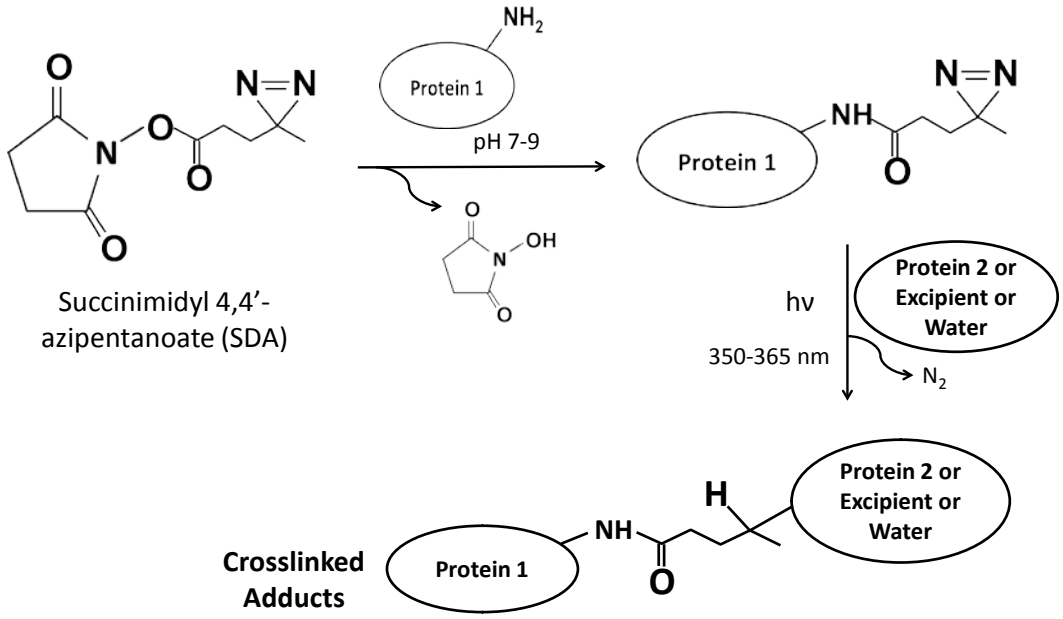


Figure 1.3. Schematic illustrating mechanism of crosslinking two proteins using the heterobifunctional photoactive crosslinker succinimidyl 4,4'- azipentanoate (SDA) (adapted from Life Technologies https://www.lifetechnologies.com/order/catalog/product/26167).

Spacer arms are usually carbon chains, but may also include reducible disulfide bonds. The length of the spacer arm determines steric effects and limits the number of crosslinked adducts formed. Cross-linkers may be classified based on the length of the spacer arm. Zero-length cross-linkers have no spacer arm and form a direct covalent bond between two molecules without themselves participating in the crosslink. The zero-length carbodiimide crosslinkers EDC is commonly used to

conjugate carboxyl groups and amine groups via an amide bond. In general, short spacer arms (4-8 Å) are better suited to study intramolecular crosslinks, while long spacer arms (~ 12 Å) favor intermolecular crosslinks. The spacer arm may also be cleaved to facilitate separation of the crosslinked species and reduce the complexity of MS analysis.

PAAs with a single active functionality can also be used for crosslinking. Peptides and proteins with a PAA incorporated within the amino acid sequence can be irradiated to produce cross-linked molecules. Kolbel et al observed differences in crosslinking patterns as a function of secondary structure using peptides containing pLeu within the primary sequence [20]. Similarly, changes in the conformation of the receptor PPAR-α upon agonist and antagonist binding were observed using genetically encoded pBpA incorporated within the sequence of PPAR-α [24]. In addition, PAAs incorporated within the protein sequence using the translational machinery of cells have been used to map the intracellular interactome [15, 25, 26].

Crosslinking with mass spectrometry can provide useful, high-resolution information about interacting partners and binding interfaces. Gomes and Gozzo used an HPP (succinimidyl 2-[(4,4´-azipentanamido)ethyl]-1,3´-dithioproprionate (SDAD) to crosslink Mb [27]. SDAD has an NHS ester at one end and a diazirine ring at the other, separated by a 13.6 Å spacer arm with a cleavable disulfide bond. The NHS moiety was reacted first at alkaline pH, followed by removal of excess unreacted SDAD. The labeled protein was irradiated to activate the diazirine end, resulting in formation of crosslinked products. The spacer arm was cleaved by reduction to facilitate analysis. Both intra- and intermolecular crosslinks were detected by MS analysis. Furthermore, the sites of crosslinking could be localized using MS/MS.

Although crosslinking improves resolution of side-chain environment, it also increases the complexity of data, especially when there are multiple crosslinks within the same peptide. In order to simplify analysis of crosslinked peptides, Schilling et al have classified modifications based on

the type of product formed [28]. Internal rearrangement or quenching of the activated labeling agent result in Type 0 'deadend' modifications. The reactive probe may also label an amino acid within the same polypeptide chain, forming a Type 1 intrapeptide crosslink. Reaction of the label with an amino acid belonging to another protein molecule results in a Type 2 interpeptide crosslink between a longer peptide (α) and a shorter peptide (β). Additionally, various combinations of these modifications are also possible, making data analysis complicated.

1.4 ADVANTAGES OF MODIFICATION USING PAAs

PAAs offer a number of advantages over chemical labeling agents. For example, since labeling is not biased towards a particular amino acid, the entire protein surface can be probed. In addition, small PAAs like pLeu and pMet can be incorporated into the sequence of the protein through metabolic labeling. The PAAs simply need to be added to cell culture media instead of their wild-type counterpart, and the cells' translational machinery will incorporate the PAA into the protein sequence. Thus the modified proteins themselves can be used as labeling agents. Furthermore, UV irradiation allows PAAs to be added to the reaction mixture prior to activation. It also provides better control of reaction than chemical agents such as NHS that require pH control and quenching of excess reagent.

1.5 ADVANTAGES OF CROSSLINKING

Thus far, indirect evidence of solid-state protein-excipient and protein-water interactions has been reported using FTIR spectroscopic data[29, 30]. These inferences are based on band areas for carboxylate hydrogen-bonding interactions. A disadvantage of FTIR is that band resolution depends on arbitrary deconvolution input parameters such as half-bandwidth, resulting in altered peak position and intensity for the same spectrum. Solid-state crosslinking allows for direct interrogation of protein-matrix interactions using mass spectrometry. High-resolution qualitative information about the presence of protein-protein, protein-excipient and protein-water adducts in

different formulations can be obtained. An additional advantage is lack of interference from water vapor.

1.6 SPECIFIC AIMS

The inherent instability of proteins makes the formulation of biologics challenging. Formulation development is often done by a trial and error approach, which can be time-consuming and expensive. In addition, analytical methods currently used to characterize protein structure lack sufficient resolution, making it difficult to draw meaningful conclusions about conformation. In this research, two novel analytical methods, ***solid-state photolytic labeling-mass spectrometry (ssPL-MS)*** and ***solid-state crosslinking-mass spectrometry (ssPC-MS)*** are being developed. The overall objective of this research is ***to develop ssPL-MS and ssPC-MS as tools for rational protein formulation***, supplanting the current paradigm of trial and error. PL-MS using photoreactive amino acid analogs (PAAs) has been described for solution samples, to study protein-protein interaction in vitro as well as in vivo [15, 18, 20]. Similarly, PC-MS using heterobifunctional photoactive probes (HPPs) has been used to elucidate the three-dimensional structure and molecular interactions of proteins [27, 31, 32]. This proposal aims to adapt PL-MS and PC-MS to the solid state (ssPL-MS and ssPC-MS) for lyophilized formulations. A photoactive labeling or crosslinking reagent (PAA/HPP) will be used to probe the protein side chain environment in three different ways; *(a) PAA incorporated in the lyophilized formulation matrix as an excipient and then irradiated (external labeling), (b) PAA incorporated within a protein/peptide sequence and then lyophilized (internal labeling)* and *(c) Protein side-chains derivatized with HPP, lyophilized and irradiated (crosslinking).* When the PAA/HPP is irradiated with UV light (350-365 nm), the photoreactive functional group on the probe is activated and reacts with protein molecules in its vicinity (within a certain distance). As a result, a covalent bond is formed between the PAA/HPP and protein side chain. The labeled/crosslinked protein is analyzed by MS at the intact level and after enzymatic digestion.

Labeling and crosslinking can provide ***direct information about the local environment of amino acid side-chains in protein formulations***. This makes the method complementary to hydrogen-deuterium exchange mass spectrometry (HDX-MS), which probes secondary structure. The location of the label can identify side-chains on the protein that are accessible to the PAA/HPP. The three dimensional structure of a protein or protein complex can also be elucidated, since the crosslinking reaction is constrained by distance. An added advantage is the absence of back-exchange of the label, which is a limitation of HDX-MS.

SPECIFIC AIM 1. To probe protein-protein and protein-matrix interactions in lyophilized protein formulations using external photolytic labeling and crosslinking.

In order to rationally design a lyophilized protein formulation, high-resolution, molecular-level information about the protein and surrounding matrix is required. To obtain this information, two approaches will be used (i) a PAA probe will be added to the excipient matrix in lyophilized formulations and irradiated with UV-A light (365 nm) to form covalent bonds between the probe and neighboring protein molecules. The protein-PAA adducts will be analyzed by LC-MS at the intact protein level and at the peptide level (ii) a PAA probe will be incorporated within a peptide sequence and lyophilized with excipients. Crosslinking will be initiated by UV irradiation and peptide-peptide and peptide-excipient adducts will be analyzed by LC-MS at the intact protein level and at the peptide level. The studies test the hypothesis that ssPL-MS and ssPC-MS can detect protein-protein and protein-matrix interactions.

SPECIFIC AIM 2. To probe protein conformation, protein-protein and protein-matrix interactions in solids and solutions using crosslinking.

The goal of Specific Aim 2 is to crosslink a model protein with matrix components in lyophilized formulations. Complementary to Specific Aim 1, this research allows detection of protein-protein, protein-excipient and protein-water adducts in the solid state. Myoglobin will be derivatized with a heterobifunctional crosslinking reagent in solution and lyophilized with different excipients. The

crosslinker will be irradiated in the solid state to produce crosslinked adducts. ESI-HPLC-MS will be used to identify these adducts after trypsin digestion. The studies test the hypothesis that changes in protein-protein and protein-matrix interactions can be detected with high resolution in different lyophilized formulations using ssPC-MS.

SPECIFIC AIM 3. To study the effect of process and excipient on lyophilized protein conformation using mass spectrometric methods.

Processing conditions can affect protein structure and result in instability during lyophilization, storage or reconstitution. Specific Aim 3 focuses on detection of conformational changes in the solid-state using high-resolution analytical methods. Myoglobin will be lyophilized with and without controlled ice nucleation in a pilot-scale freeze-dryer. Product temperature will be monitored during lyophilization and conventional product characterization techniques such as FTIR spectroscopy and moisture content analysis will be performed. Backbone conformational changes will be monitored using solid-state hydrogen deuterium exchange-mass spectrometry (ssHDX-MS) and side-chain matrix accessibility will be assessed using solid-state photolytic labeling- mass spectrometry (ssPL-MS). The hypothesis is that these high-resolution methods are more sensitive to structural changes than conventional solid-state FTIR spectroscopy.

1.7 OVERALL APPROACH

Model Proteins and Peptides: Myoglobin will be used as a model protein for Specific Aim 1, 2 and 3, as it is a fairly small molecule with no cysteines and has also been used for ssHDX studies in our lab previously. Hence enzymatic digestion and MS analysis will be relatively straightforward. A glucagon-derived peptide (GDP) obtained from the N- terminus of glucagon (1-HSQGTFTS-8; hereafter referred to as GCG (1-8)*) will be used for Specific Aim 1 for internal labeling experiments. The PAA probe p-benzoyl-L-phenylalanine (pBpA) will replace Phe in GCG (1-8)*. This peptide was selected because the N- and C-termini of glucagon have been implicated in aggregation[33, 34].

Photo amino acid analogs (PAAs) as probes: The PAA probes to be used are L-photo-leucine (L-2-amino-4,4-azipentanoic acid; pLeu), L-photo-methionine (L-2-amino-5,5-azi-hexanoic acid; pMet) ((Pierce Biotechnology, Rockford, IL) and p-benzoyl-L-phenylalanine (pBpA) (Bachem, Torrance, CA). These PAAs have different reaction mechanisms upon exposure to UV light at 350-365 nm; the diazirine functional group of pLeu and pMet forms a reactive carbene intermediate that inserts non-specifically into any C-C and X-H bond (X= C, H, N, O, S) or adds on to a C=C bond in its immediate molecular cage [18]. Labeling with pLeu and pMet is quite promiscuous, as the carbene intermediate does not favor a particular amino acid. On the other hand, the benzophenone group in pBpA forms a reactive ketyl radical that reacts preferentially with C-H bonds and forms new C-C covalent linkages [23]. Besides these PAAs, an HPP succinimidyl 4,4' azipentanoate (NHS-diazirine; SDA) will also be used. SDA contains an N-hydroxysuccinimide (NHS) ester that reacts with primary amines at alkaline pH to form an amide bond. It also contains a second functional group, a photoactive diazirine ring that forms a carbene upon exposure to UV-A light and reacts with any amino acid side-chain. The two groups are connected by a short (3.9 Å) carbon chain spacer arm.

Data analysis: In silico digestion of labeled proteins can be performed in MassHunter software as well as others such as the FindPept tool (ExPASy). This theoretical list can be matched with the observed masses using MassHunter. Analysis of cross-linked peptides is more challenging because the fragment ions obtained by MS/MS are also cross-linked. This greatly increases the complexity of the data and manual assignment of masses often must be made. The software GPMAW can compute the mass of possible crosslinked peptides after data is provided for the primary sequence of the crosslinked proteins, the type of crosslinker and the enzyme used for digestion. XQuest and XLink assign m/z values to MS/MS fragment peaks and can be used to analyze fragmentation of cross-linked or modified peptides.

1.8 REFERENCES

1. Walsh, G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol* **2010,** *28*, (9), 917-924.
2. Kasper, J. C.; Friess, W. The freezing step in lyophilization: physico-chemical fundamentals, freezing methods and consequences on process performance and quality attributes of biopharmaceuticals. *Eur J Pharm Biopharm* **2011,** *78*, (2), 248-263.
3. Luthra, S.; Obert, J. P.; Kalonia, D. S.; Pikal, M. J. Investigation of drying stresses on proteins during lyophilization: differentiation between primary and secondary-drying stresses on lactate dehydrogenase using a humidity controlled mini freeze-dryer. *J Pharm Sci* **2007,** *96*, (1), 61-70.
4. Lai, M. C.; Topp, E. M. Solid-state chemical stability of proteins and peptides. *J Pharm Sci* **1999,** *88*, (5), 489-500.
5. Gsponer, J.; Haberthur, U.; Caflisch, A. The role of side-chain interactions in the early steps of aggregation: Molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35. *Proc Natl Acad Sci U S A* **2003,** *100*, (9), 5154-5159.
6. Marchut, A. J.; Hall, C. K. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. *Biophys J* **2006,** *90*, (12), 4574-4584.
7. Gazit, E. A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J* **2002,** *16*, (1), 77-83.
8. de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Ventura, S. Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. *FEBS J* **2006,** *273*, (3), 658-668.
9. Berhanu, W. M.; Hansmann, U. H. Structure and dynamics of amyloid-beta segmental polymorphisms. *PLoS One* **2012,** *7*, (7), e41479.
10. Carpenter, J. F.; Prestrelski, S. J.; Dong, A. Application of infrared spectroscopy to development of stable lyophilized protein formulations. *Eur J Pharm Biopharm* **1998,** *45*, (3), 231-238.
11. Sane, S. U.; Wong, R.; Hsu, C. C. Raman spectroscopic characterization of drying-induced structural changes in a therapeutic antibody: correlating structural changes with long-term stability. *J Pharm Sci* **2004,** *93*, (4), 1005-1018.
12. Elkordy, A. A.; Forbes, R. T.; Barry, B. W. Study of protein conformational stability and integrity using calorimetry and FT-Raman spectroscopy correlated with enzymatic activity. *Eur J Pharm Sci* **2008,** *33*, (2), 177-190.
13. Hu, K. N.; Tycko, R. What can solid state NMR contribute to our understanding of protein folding? *Biophys Chem* **2010,** *151*, (1-2), 10-21.
14. Sophocleous, A. M.; Zhang, J.; Topp, E. M. Localized hydration in lyophilized myoglobin by hydrogen-deuterium exchange mass spectrometry. 1. Exchange mapping. *Mol Pharm* **2012,** *9*, (4), 718-726.
15. Suchanek, M.; Radzikowska, A.; Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* **2005,** *2*, (4), 261-267.
16. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2005,** *2*, (3), 201-206.
17. Soutourina, J.; Wydau, S.; Ambroise, Y.; Boschiero, C.; Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **2011,** *331*, (6023), 1451-1454.
18. Jumper, C. C.; Schriemer, D. C. Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Anal Chem* **2011,** *83*, (8), 2913-2920.
19. Mendoza, V. L.; Antwi, K.; Baron-Rodriguez, M. A.; Blanco, C.; Vachet, R. W. Structure of the preamyloid dimer of beta-2-microglobulin from covalent labeling and mass spectrometry. *Biochemistry* **2010,** *49*, (7), 1522-1532.

20. Kolbel, K.; Ihling, C. H.; Sinz, A. Analysis of peptide secondary structures by photoactivatable amino acid analogues. *Angew Chem Int Ed Engl* **2012,** *51*, (50), 12602-12605.
21. O'Neil, K. T.; Erickson-Viitanen, S.; DeGrado, W. F. Photolabeling of calmodulin with basic, amphiphilic alpha-helical peptides containing p-benzoylphenylalanine. *J Biol Chem* **1989,** *264*, (24), 14571-14578.
22. Mendoza, V. L.; Vachet, R. W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom Rev* **2009,** *28*, (5), 785-815.
23. Dorman, G.; Prestwich, G. D. Benzophenone photophores in biochemistry. *Biochemistry* **1994,** *33*, (19), 5661-5673.
24. Schwarz, R.; Tanzler, D.; Ihling, C. H.; Muller, M. Q.; Kolbel, K.; Sinz, A. Monitoring conformational changes in peroxisome proliferator-activated receptor alpha by a genetically encoded photoamino acid, cross-linking, and mass spectrometry. *J Med Chem* **2013,** *56*, (11), 4252-4263.
25. Weaver, M. S.; Workman, G.; Sage, E. H. The copper binding domain of SPARC mediates cell survival in vitro via interaction with integrin beta1 and activation of integrin-linked kinase. *J Biol Chem* **2008,** *283*, (33), 22826-22837.
26. Zhao, W. Q.; Santini, F.; Breese, R.; Ross, D.; Zhang, X. D.; Stone, D. J.; Ferrer, M.; Townsend, M.; Wolfe, A. L.; Seager, M. A.; Kinney, G. G.; Shughrue, P. J.; Ray, W. J. Inhibition of calcineurin-mediated endocytosis and alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors prevents amyloid beta oligomer-induced synaptic disruption. *J Biol Chem* **2010,** *285*, (10), 7619-7632.
27. Gomes, A. F.; Gozzo, F. C. Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. *J Mass Spectrom* **2010,** *45*, (8), 892-899.
28. Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom* **2003,** *14*, (8), 834-850.
29. Allison, S. D.; Chang, B.; Randolph, T. W.; Carpenter, J. F. Hydrogen bonding between sugar and protein is responsible for inhibition of dehydration-induced protein unfolding. *Arch Biochem Biophys* **1999,** *365*, (2), 289-298.
30. Remmele, R. L., Jr.; Stushnoff, C.; Carpenter, J. F. Real-time in situ monitoring of lysozyme during lyophilization using infrared spectroscopy: dehydration stress in the presence of sucrose. *Pharm Res* **1997,** *14*, (11), 1548-1555.
31. Chong, P. C.; Hodges, R. S. A new heterobifunctional cross-linking reagent for the study of biological interactions between proteins. I. Design, synthesis, and characterization. *J Biol Chem* **1981,** *256*, (10), 5064-5070.
32. Lewis, R. V.; Roberts, M. F.; Dennis, E. A.; Allison, W. S. Photoactivated heterobifunctional cross-linking reagents which demonstrate the aggregation state of phospholipase A2. *Biochemistry* **1977,** *16*, (25), 5650-5654.
33. Pedersen, J. S.; Dikov, D.; Otzen, D. E. N- and C-terminal hydrophobic patches are involved in fibrillation of glucagon. *Biochemistry* **2006,** *45*, (48), 14503-14512.
34. Svane, A. S.; Jahn, K.; Deva, T.; Malmendal, A.; Otzen, D. E.; Dittmer, J.; Nielsen, N. C. Early stages of amyloid fibril formation studied by liquid-state NMR: the peptide hormone glucagon. *Biophys J* **2008,** *95*, (1), 366-377.

CHAPTER 2. PHOTOLYTIC LABELING TO PROBE MOLECULAR INTERACTIONS IN LYOPHILIZED POWDERS

This chapter was published as a research article in Molecular Pharmaceutics (2013) and can be found online at http://pubs.acs.org/doi/abs/10.1021/mp4004332

2.1 ABSTRACT

Local side-chain interactions in lyophilized protein formulations were mapped using solid-state photolytic labeling-mass spectrometry (ssPL-MS). Photoactive amino acid analogs (PAAs) were used as probes and either added to the lyophilized matrix or incorporated within the amino acid sequence of a peptide. In the first approach, apomyoglobin was lyophilized with sucrose and varying concentrations of photo-leucine (L-2-amino-4, 4'-azipentanoic acid; pLeu). The lyophilized solid was irradiated at 365 nm to initiate photolabeling. The rate and extent of labeling were measured using ESI-HPLC-MS, with labeling reaching a plateau at ~ 30 min, forming up to 6 labeled populations. Bottom-up MS/MS analysis was able to provide peptide-level resolution of the location of pLeu. ssPL-MS was also able to detect differences in side-chain environment between sucrose and guanidine hydrochloride formulations. In the second approach, peptide GCG (1-8)* containing p-benzoyl-L-phenylalanine (pBpA) in the amino acid sequence was lyophilized with various excipients and irradiated. Peptide-peptide and peptide-excipient adducts were detected using MS. Top-down MS/MS on the peptide dimer provided amino acid-level resolution regarding interactions and the cross-linking partner for pBpA in the solid state. The results show that ssPL-MS can provide high-resolution information about protein interactions in the lyophilized environment.

2.2 INTRODUCTION

Protein drugs are an increasingly important part of the global pharmaceuticals market. In 2009, the ten best-selling protein drugs had a combined sales value of close to $50 billion[1]. The number of approved protein drugs is expected to increase in the next few years, particularly given the expiration of patents and the growth of biosimilars. According to a report by Global Industry Analysts, Inc., biosimilars are expected to be valued at $17.9 million by 2017[2]. However the inherent instability of proteins and their tendency to aggregate is an obstacle to the development of these life-saving medicines. In an attempt to maintain stability and provide adequate shelf life, many proteins are lyophilized. In addition to those products marketed as lyophilized powders, the protein itself may be lyophilized for storage prior to final formulation in either solution or solid forms. Although lyophilized formulations usually confer greater stability when compared to solution, degradation may still occur in the solid state and during the freeze-drying process[3-6]. Retention of native protein structure in the lyophilized solid has generally been associated with improved stability during shelf-storage and a decreased propensity for aggregate formation[7-9]. Ensuring the retention of native conformation would benefit from analytical methods that could identify subtle protein structural perturbations in lyophilized solids with high resolution. Such information could be used to design formulations rationally and to screen candidate formulations efficiently.

Most of the current analytical techniques used to characterize proteins in the solid state lack sufficient resolution to serve as design tools, however. Methods such as Fourier transform-infrared spectroscopy (FTIR), Raman spectroscopy and differential scanning calorimetry (DSC) have been used to study structural changes in lyophilized proteins[10-13]. These methods are semi-quantitative at best, suffer from low sensitivity and can provide only low-resolution information on protein structure. Solid-state nuclear magnetic resonance spectroscopy (ssNMR) can provide site-specific information about conformational changes[14, 15], but requires extensive sample preparation and isotopic labeling, and is less sensitive in amorphous samples than in those that are crystalline. Thus, ssNMR is not always useful for lyophilized protein formulations, which are usually amorphous.

Recently, our group has developed solid-state hydrogen-deuterium exchange with mass spectrometric analysis (ssHDX-MS) to allow the protein environment in amorphous solids to be probed with higher resolution. ssHDX-MS provides structural information with peptide level resolution, and has been used previously by our group to characterize protein conformations in lyophilized solids containing various excipients[16, 17].

In the work reported here, we have developed a complementary analytical technique, solid-state photolytic labeling with mass spectrometric analysis (ssPL-MS), to probe protein structure and matrix interactions in lyophilized formulations. In solution, PL-MS with photoreactive amino acid analogs (PAAs) has been used to study protein/peptide conformation and protein-protein interactions (PPIs)[18-20]. The approach has also been used in living cells to map the interactome[21-23]. In solution PL-MS, a solution containing protein and PAA is irradiated with UV light (350-365 nm), activating the PAA photoreactive functional group, which then forms a covalent bond between the PAA and protein in its immediate vicinity. The labeled protein is analyzed by MS at the intact protein level and by MS/MS fragmentation after enzymatic digestion (bottom-up) or direct fragmentation (top-down). The location of the label identifies sites on the protein that are accessible to the photoreactive probe, providing information about the side-chain environment. This makes the method complementary to HDX-MS, which probes backbone environment and secondary structure. Moreover, the covalently attached label is permanent and does not undergo back-exchange, a limitation of HDX. Solution state PL-MS has also been carried out by incorporating the PAA within a protein or peptide sequence[19, 24]. Exposure to UV light generates photoadducts of the PAA-containing protein/peptide with interacting molecules (e.g. ligand) in the microenvironment. These photoadducts can then be digested enzymatically and analyzed to identify the reactive sites at the interface of the complex.

In the current work, we have adapted PL-MS for proteins in lyophilized solids. PAAs were used to probe the side chain protein environment in two different ways: (i) by incorporating a PAA into the

lyophilized solid as an excipient and (ii) by incorporating a PAA into the sequence of a model peptide. In studies using a PAA probe as an excipient (i), apomyoglobin (ApoMb) was selected as a model protein and L-photo-leucine (L-2-amino-4, 4'-azipentanoic acid; pLeu) was used as an excipient. In studies with the PAA incorporated into the protein sequence (ii), an octapeptide derived from the N-terminus of human glucagon (1-HSQGTFTS-8) with the phenylalanine residue (F6) replaced by the PAA p-benzoyl-L-phenylalanine (pBpA) was used. The two PAAs have different reaction mechanisms upon exposure to UV light. The diazirine functional group of pLeu forms a reactive carbene intermediate that inserts non-specifically into any C-C or X-H bond (X= C, N, O, S), or adds to a C=C bond in its immediate molecular cage. The benzophenone group in pBpA forms a reactive ketyl radical that reacts preferentially with C-H bonds and forms new C-C covalent linkages[25, 26]. The results demonstrate that photolytic labeling occurs in lyophilized solids when the label is either incorporated into the matrix (i) or into a model peptide (ii). The results also show that the extent of labeling varies with position in the protein sequence and with solid composition. To our knowledge, this is the first reported use of photolytic labeling to map the protein environment in lyophilized solids. The findings support further development of the method to probe the amorphous solid state and in formulation development.

2.3 MATERIALS AND METHODS

Apomyoglobin (apoMb) from equine skeletal muscle, monobasic and dibasic potassium hydrogen phosphate, L-methionine (Met), L-leucine (Leu), sucrose, trehalose, urea and guanidine hydrochloride were purchased from Sigma Aldrich (St. Louis, MO). L-photo-leucine (L-2-amino-4, 4-azipentanoic acid; pLeu) was obtained from Thermo Scientific (Rockford, IL). An octapeptide derived from the N-terminus of glucagon (HSQGT-pBpA-TS; henceforth referred to as GCG (1-8)*) containing the photoreactive amino acid p-benzoyl-L-phenylalanine (pBpA) within its sequence was synthesized by American Peptide Company (Sunnyvale, CA) and received as a lyophilized powder. Trypsin and chymotrypsin were purchased from Promega (Madison, WI) and mass spectrometry-grade water, acetonitrile and formic acid from Fisher Scientific (Fair Lawn, NJ).

2.3.1 Sample Preparation

ApoMb was dissolved in potassium phosphate buffer (20 mM, pH 7.4) to produce a stock solution of 200 µM protein and the solution was dialyzed using Biotech Cellulose Ester dialysis tubing (MWCO 8,000-10,000 Da, Spectrum Laboratories, Rancho Dominguez, CA) for 24 h into the same buffer. After dialysis, the solution was filtered through a 0.2 µm syringe filter (Gelman Nylon Acrodisc 13) and used for further experiments. Sucrose stock solution (33.9 mg/mL) was prepared by dissolving sucrose in potassium phosphate buffer and filtering through a 0.2 µm syringe filter. The resulting solution was stored at 4 °C until use. A stock solution of pLeu (30 mM) was prepared similarly. Lyophilization was carried out with different ratios of protein to pLeu using a VirTis Plus AdVantage freeze dryer (SP Industries Inc., Gardiner, NY). ApoMb, sucrose and pLeu stock solutions were mixed such that the final protein concentration was 100 µM, the protein to sucrose ratio was 1:2 w/w and the protein to pLeu molar ratio was 1:20, 1:50 or 1:100. Control samples contained apoMb and sucrose without pLeu. The final volume for lyophilization was 80 µL. In order to produce a pharmaceutically relevant formulation, ~ 50 % or more of the solid matrix consisted of sucrose, and buffer was less than 10 % of the formulation by weight (Table 2.1).

All samples were lyophilized in vials made of borosilicate clear glass using an established conservative freeze-drying cycle. During the lyophilization cycle, the shelves were precooled to -2 °C. Freezing was carried out at -40 °C for 50 min, followed by drying under vacuum (70 mTorr) over 5 steps (-35 °C for 10 h, -20 °C for 8 h, -5 °C for 6 h, 10 °C for 6 h and 25 °C for 6 h). The lyophilized samples were stored at -20 °C until use. Solution controls were prepared at each composition.

Table 2.1. Composition of lyophilized formulations containing apomyoglobin (apoMb)[a] and photo-leucine (pLeu)[b]

apoMb:pLeu molar ratio	**Composition** (%w/w)			
	ApoMb	**Sucrose**	**pLeu**	**Buffer**
No pLeu	31.0	61.9	0.0	7.1
1:20	29.4	58.5	5.0	6.8
1:50	27.4	54.8	11.6	6.3
1:100	24.5	49.1	20.7	5.6

2.3.2 Photolytic Labeling and MS Analysis of Intact Protein

Photolytic labeling was carried out using a UV Stratalinker 2400 equipped with five 365 nm UV lamps (Stratagene Corp., La Jolla, CA). The lamps were allowed to warm up for 5 min. Vials containing lyophilized samples and solution controls were uncapped and placed inside the UV chamber. The distance between the lamps and the cake at the bottom of the vial was approximately 15 cm. All samples were irradiated with UV light for 40 min. After irradiation, the solid was reconstituted with 800 µL of Solution A (A= 0.1 % formic acid in MS water) to give a final protein concentration of 10 nmol/mL. The solution formulation was diluted similarly. The samples were diluted further with Solution A and 20 pmol of protein was injected into the HPLC-MS system. Intact labeled protein was analyzed using HPLC-MS equipped with an ESI source (1200 series LC, 6520 qTOF; Agilent Technologies, Santa Clara, CA). Mass spectra were processed and deconvoluted using MassHunter software (Agilent Technologies).

Percentages of protein populations with 1 through 6 labels were calculated using peak heights from extracted ion chromatograms:

$$\% L_i = PH_i / (PH_i + PH_u) \times 100 \qquad \textbf{Equation 2.1}$$

where *i* denotes the number of labels (1-6), PH_i denotes the peak height for labeled protein L_i and PH_u denotes the peak height of the unlabeled protein as observed by mass spectrometry. Hereinafter, the term 'unlabeled' will refer to a protein/peptide that has been exposed to pLeu and irradiation, but was not labeled, while the term 'native' will refer to a protein/peptide that has not been exposed to pLeu and irradiation.

2.3.3 Effect of Irradiation Time and pLeu concentration on Labeling Efficiency

ApoMb lyophilized with sucrose and pLeu (1:100 molar ratio protein: pLeu, which is equivalent to 20.7 % w/w pLeu) was used to study the kinetics of photolytic labeling. Lyophilized samples were subjected to photolysis for different periods of time (0, 0.5, 2, 4, 10, 20, 30, 40 and 60 min). The samples were reconstituted and analyzed as described above. In a separate study, apoMb was lyophilized with sucrose and varying pLeu concentrations (0, 0.3, 1.3, 1.5, 2.5, 11.6 and 20.7 % w/w). The solid was irradiated at 365 nm for 40 min, reconstituted and analyzed for labeled protein. The fraction of labeled protein (F_L) was calculated using peak heights from extracted ion chromatograms:

$$F_L = 1 - [PH_u/(PH_u + PH_L)] \qquad \textbf{Equation 2.2}$$

where PH_L denotes the peak height for labeled protein and PH_u denotes the peak height of the unlabeled protein as observed by mass spectrometry. F_L represents the sum of populations of apoMb with 1-6 labels.

2.3.4 MS- and MS/MS- Analysis of Labeled apoMb Peptides

To identify the sites of photolytic labeling, apoMb was lyophilized with sucrose and pLeu as described above using 0, 0.3, 1.3, 1.5, 2.5, 11.6 and 20.7 % w/w pLeu. The solid was irradiated at 365 nm for 40 min and then reconstituted in ammonium bicarbonate buffer (100 µM, pH 8.0) to give a protein concentration of 10 nmol/mL. Enzymatic digestion of labeled apoMb was performed for 24 h at 60 ^{o}C using a combination of trypsin and chymotrypsin (1:1 molar ratio) at a total enzyme to protein molar ratio of 1:10. The reaction was then quenched with solution A and 20 pmol was injected into the LC-MS system. The proteolytic fragments were separated on a ZORBAX 300SB-C18 column (Agilent Technologies; 1.0 x 50 mm, particle size 3.5 µm). The column was equilibrated with 5% Mobile Phase B (B= 0.1 % formic acid in acetonitrile) and peptides were eluted at 50 µL/min using a gradient that increased from 5 to 45% B over 22 min and then from 45 to 95% B over 0.5 min. Mass spectra were processed using MassHunter and a theoretical digest map (with known sites of enzymatic cleavage, and allowing for up to 8 missed cleavages) was used to create a mass list for peptides carrying 0 through 7 labels. This theoretical list was matched against mass values obtained experimentally.

Based on this analysis, up to 15 labeled peptides were detected that carried one, two or four labels. One of these peptides, L32-K42 (LFTGHPETLEK) with one label was selected for MS/MS analysis. This precursor peptide had m/z = 462.9133 (z = +3) and was subjected to fragmentation using low-energy CID (Agilent Technologies), which predominantly produces *b*- and *y*-ions. Product ions were identified using MassHunter software.

2.3.5 Formulation Effects

In order to study the effect of excipients on side-chain environment, apoMb was lyophilized with 100x molar excess of pLeu in two formulations: the first with sucrose (1:2 w/w ratio of protein to sucrose) and the second with guanidine hydrochloride (Gdn HCl; 1.5 M final concentration). The final protein concentration was 100 µM and the concentration of pLeu was 20.7 % w/w (sucrose

formulation) or ~ 1 % w/w (Gdn HCl formulation). The lyophilized formulations were subjected to photolysis at 365 nm for 40 min. After reconstitution with ammonium bicarbonate buffer, the samples were digested with trypsin and chymotrypsin and peptide-level MS analysis was carried out as described above.

2.3.6 Photolytic Labeling with p-Benzoyl-L-Phenylalanine (pBpA)

GCG (1-8)* was dissolved in water to give a final concentration of 1 mM. The peptide was lyophilized alone or with one of the following excipients: sucrose, trehalose, urea, L-methionine and L-leucine (1:2 w/w ratio of peptide to excipient). After lyophilization, the formulations were irradiated with UV light (365 nm) for 30 min. The irradiated samples were then reconstituted in 200 µL of MS water containing 0.1 % formic acid. Solution controls prepared with or without excipients were lyophilized and reconstituted before irradiation. The samples were further diluted to 20 pmol of peptide for injection into the LC-MS system. MassHunter software was used to detect peptide-peptide and peptide-excipient adducts.

Photolytic labeling with GCG (1-8)* was also carried out with pLeu in the matrix. Two formulations were prepared. The first contained GCG (1-8)* and pLeu at a 1:1 molar ratio, while the second contained GCG (1-8)* and pLeu at a 1:1 molar ratio, together with sucrose (1:2 w/w ratio of GCG (1-8)* to sucrose). Both formulations were lyophilized as described above, irradiated with UV light for 30 min, reconstituted and analyzed by ESI-LC-MS. Solution controls were prepared and analyzed as described above.

2.3.7 MS/MS Analysis of GCG (1-8)* Dimer

Both lyophilized and solution formulations showed the presence of GCG (1-8)* dimer. LC-MS/MS was carried out on GCG (1-8)* monomer (m/z 968.41, z= +1) and GCG (1-8)* dimer (m/z 646.28, z= +3) in the peptide-Leu formulation. CID fragmentation was performed at 10 V and the resultant

b- and *y*-ions were monitored. The dimer from solution controls (unlyophilized solution and lyophilized-rehydrated solution) was also analyzed by MS/MS.

2.4 RESULTS

2.4.1 Intact Protein Labeling

The mechanisms of photolytic labeling with pLeu in solution are well understood[18, 25]. Briefly, photolysis of pLeu at 365 nm results in the loss of N_2 with the generation of a reactive carbene. The carbene labels any C-C, C=C or X-H group (X=C, O, N, S) in its proximity without bias towards a particular amino acid or functional group. Photolabeling of myoglobin with pLeu has been carried out in solution, with successful labeling at a 1:10,000 molar ratio of protein to pLeu[18]. Here, we investigated the covalent labeling of apoMb with pLeu in lyophilized solids. Intact protein was co-lyophilized with sucrose and pLeu as excipients in weight fractions that are pharmaceutically relevant.

Mass spectrometric analysis of lyophilized solids containing apoMb and pLeu showed that carbene labeling also occurs in the solid state. Peaks corresponding to labeled protein were observed, with masses differing by multiples of ~115 amu (Fig. 2.1). The extent of labeling depended on the amount of pLeu in the matrix. Peaks corresponding to singly- and doubly-labeled apoMb populations were observed when apoMb was lyophilized with a 20-fold molar excess of pLeu (Fig. 2.1). Similarly, peaks corresponding to up to 4 and 6 labels per protein molecule were observed for the 1:50 and 1:100 formulations, respectively. ApoMb lyophilized without pLeu showed no adduct formation after irradiation (data not shown), confirming that UV light did not cause protein cross-linking. Also, protein lyophilized with pLeu showed no labeling in the absence of UV light (data not shown). Moreover, solution controls showed no labeling of apoMb with 20-, 50-, 100- or 1000-fold molar excess of pLeu (data not shown), suggesting that the reactive carbene species was consumed by reaction with water rather than reacting with protein.

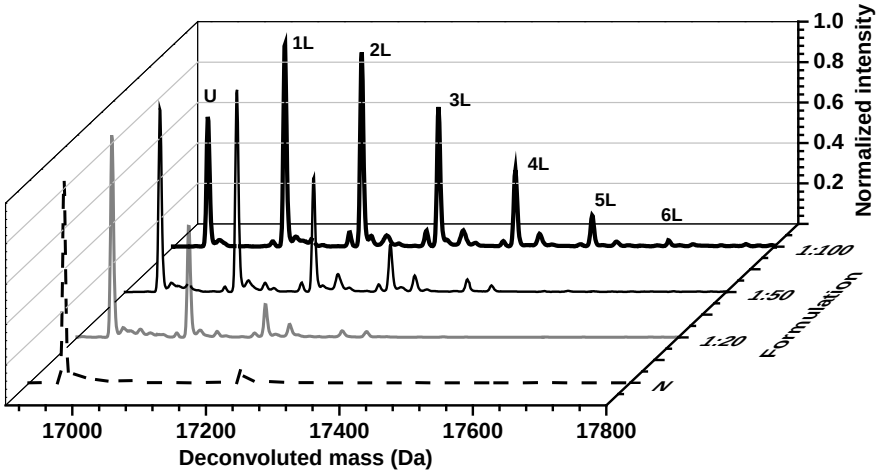


Figure 2.1 Deconvoluted mass spectra of native ApoMb (N) and ApoMb co-lyophilized with sucrose (1:2 w/w ratio of protein to sucrose) and pLeu in molar ratios 1:20, 1:50 and 1:100. Mass spectra show peaks for unlabeled apoMb (*U*) and labeled apoMb (*nL*) (n=1-6). The peaks differ by ~ 115 amu, corresponding to the mass of one pLeu label.

2.4.2 Labeling Kinetics

During exposure of solid samples to UV irradiation, the fraction of labeled protein increased with time (Fig. 2.2A). The rate of formation of labeled protein was rapid initially and plateaued at ~30 min with ~20% of the protein remaining unlabeled (Fig. 2.2A). Labeling followed monoexponential kinetics as a function of irradiation time. To determine the effect of pLeu concentration on the plateau value, the extent of labeling was measured at different initial concentrations of pLeu with 40 min of irradiation (Fig. 2.2B). At 0 % w/w pLeu, no labeling occurred. As pLeu concentration was increased, the fraction of labeled protein increased until at 20.7 % w/w pLeu, ~35 % unlabeled protein remained after 40 min of irradiation. The dependence of the extent of modification on pLeu concentration also followed monoexponential behavior.

An exponential model was used to simultaneously fit the rate and extent of labeling:

$$F_L(C, t) = A(1-e^{-k_1 t})(1-e^{-k_2 C}) \qquad \textbf{Equation 2.3}$$

where $F_L(C, t)$ is the fraction of labeled protein as a function of pLeu concentration (C) and irradiation time (t), k_1 and k_2 are apparent first-order rate constants for the rate and extent of labeling, respectively, and A is the fraction of protein labeled at plateau. Nonlinear regression (Origin Pro v.8.6, OriginLab, Northampton, MA; n = 48) returned values of the regression parameters of A = 0.82 (±0.03), k_1= 0.22 min^{-1} (±0.02) and k_2= 0.12 mM^{-1} (±0.01).

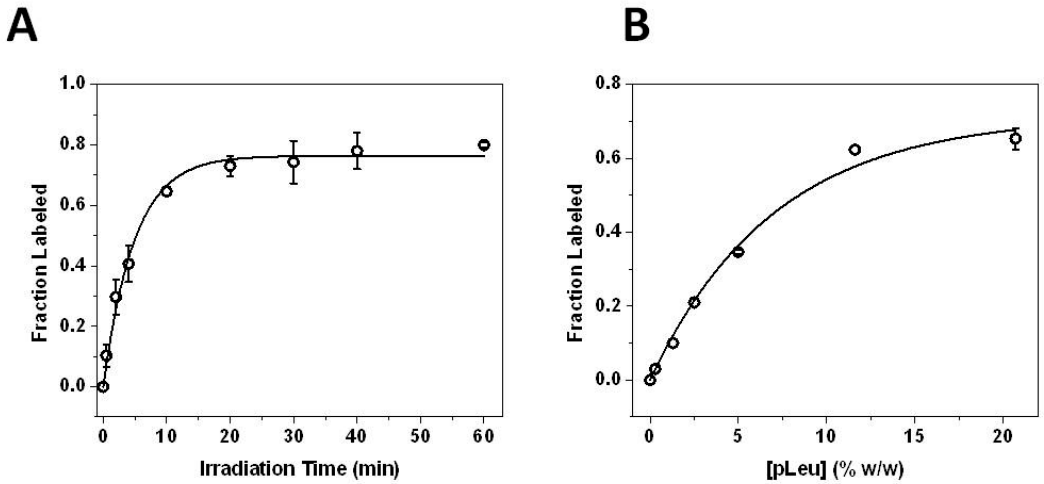


Figure 2.2. (A) Kinetics of photolytic labeling of apoMb in lyophilized solids containing 20.7 % w/w pLeu in the matrix, 365 nm irradiation. The solid line is fit to Eqn. 2.3. n = 3 ± SD. **(**B) Dependence of ApoMb photolytic labeling on the concentration of pLeu after 40 min irradiation at 365 nm. The solid line is fit to Eqn. 2.3. n = 3 ± SD.

2.4.3 Peptide Labeling

In order to investigate the specificity of labeling, sites of labeling were probed using bottom-up mass spectrometry. Digestion of native apoMb with trypsin/chymotrypsin produced 36 peptides, of which 13 were selected to provide 100 % sequence coverage (Appendix, Fig. A1). Labeled apoMb showed ~ 96 % sequence coverage and a maximum of fifteen peptide fragments (obtained with 20.7 % w/w pLeu) with one, two or four labels (Fig. 2.3). As expected, the signal intensity of labeled peptides was less than that of unlabeled peptides, supporting the incomplete labeling observed at the intact protein level. Proteolytic digestion was influenced by the presence of pLeu labels: both

the labeled and unlabeled peptides obtained after digestion of labeled apoMb differed from those in the native protein. This suggests that the label interferes with digestion by obstructing access by the enzyme.

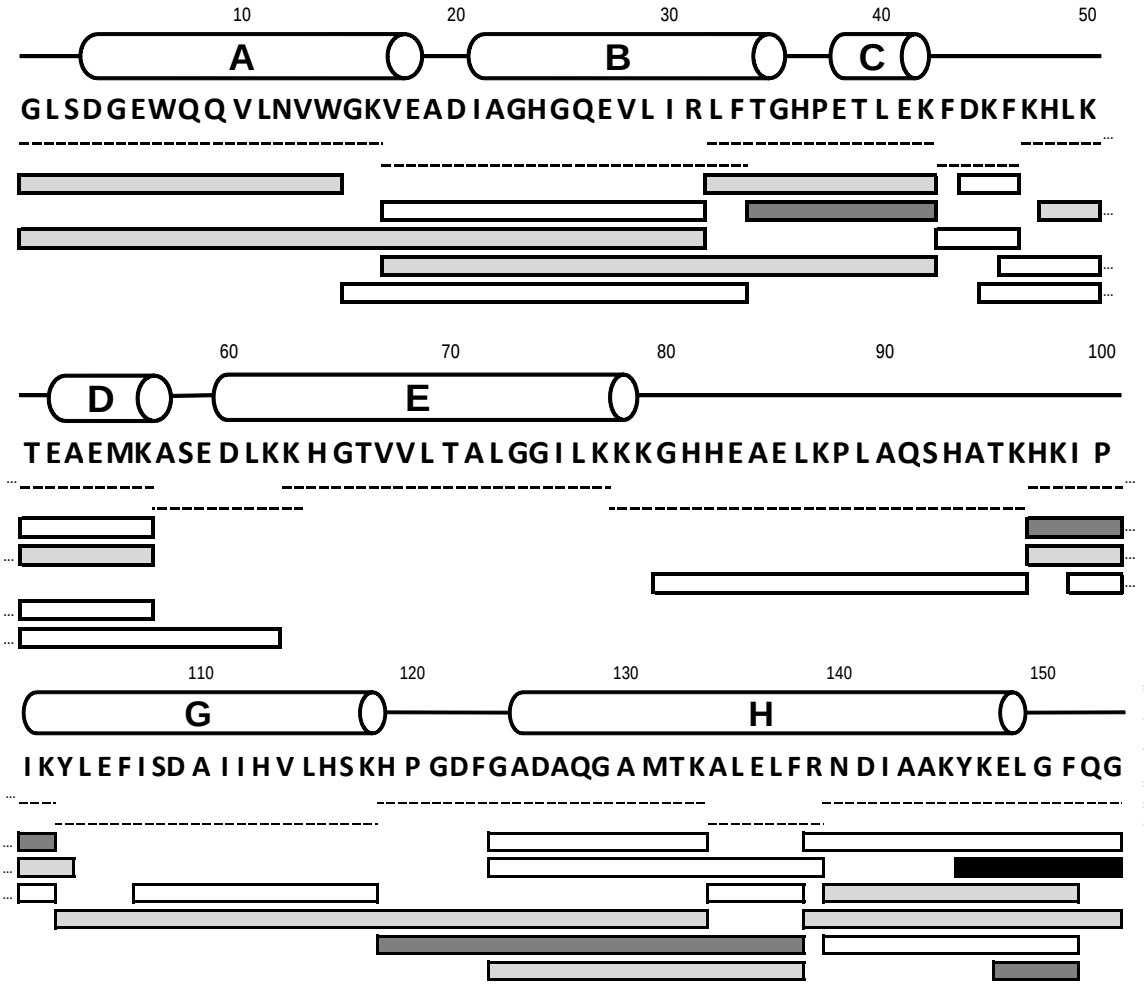


Figure 2.3. Digest map of apoMb labeled with 10 mM pLeu. Labeled apoMb was digested with a combination of trypsin and chymotrypsin. White bars represent unlabeled peptides, while labeled peptides are shown in gray (light gray bars carry one label; dark gray bars carry two labels and the black bar carries four labels). Dashed lines represent native peptides. Helical secondary structure is represented by cylinders labeled A-E, G and H. Helix F of holomyoglobin (H82-H97) is disordered in native apoMb at neutral pH[27].

As the concentration of pLeu increased, labeling was detected in different regions of the protein (Fig. 2.4). Labeling at the peptide level was obtained using MS analysis of digested labeled apoMb at various pLeu concentrations (Fig. 2.4(b-h)). At 0 % w/w pLeu, no labeling was observed (Fig. 2.4b). At 0.3 % w/w pLeu, peptides L32-K42 and T34-K42 were labeled (Fig. 2.4c). This region forms helix C and part of helix B. At 1.3 % w/w pLeu, an additional peptide HKIPIKY (H97-Y103; located on a loop and part of helix G) was labeled (Fig. 2.4d). As pLeu concentration was increased to 2.5 % w/w, labeling was detected in peptide H119-F138 (helix H) in addition to L32-K42, T34-K42 and H97-Y103 (Fig. 2.4e). At 5 % w/w and 11.6 % w/w pLeu, Y103-K133 was labeled as well (Fig. 2.4f, g). At 20.7 % w/w pLeu, label was detected in G1-W14, G1-R31, V17-K42 (helices A, B and C), H48-K56 (helix D), H97-K102 (helix G), G124-F138, N140-F151, Y146-G153, E148-F151 and R139-G153 (helix H), in addition to the previously mentioned sequences (Fig. 2.4h). Increased label uptake at the C terminus is consistent with our previous solid-state hydrogen-deuterium exchange (ssHDX-MS) results for myoglobin, which showed greater deuterium uptake in this region even in the solid state[16]. Overall, labeling was observed across helices A, B, C, D, G and H. No labeling was observed on amino acids A57-K96, which form helices E and F. These two helices are involved in heme binding in holomyoglobin (holoMb), but are considerably disordered in apoMb[27, 28]. The absence of label suggests that this region is protected from matrix exposure at the tertiary structure level in the solid state.

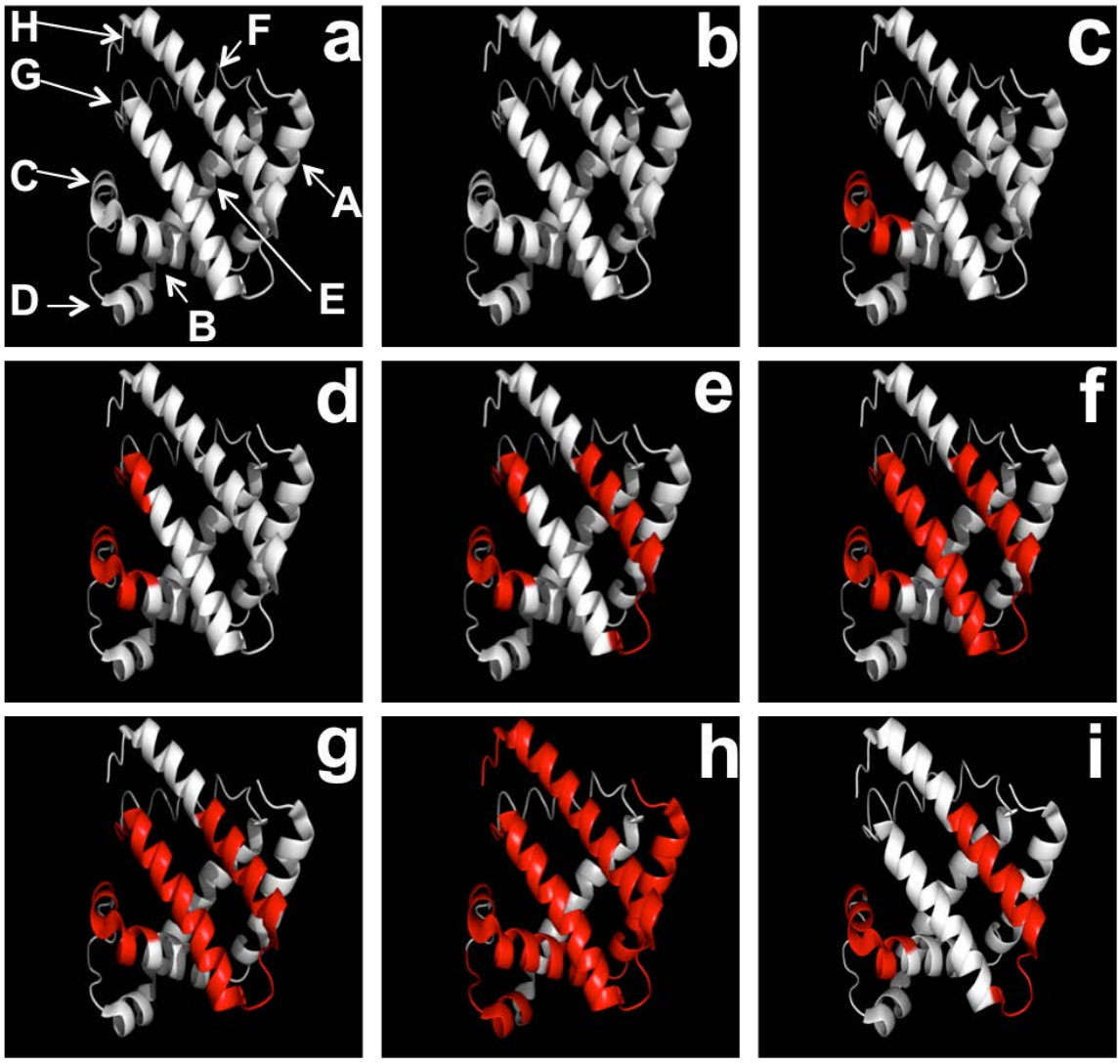


Figure 2.4. (a) Ribbon diagram of apoMb showing helices A-E, G and H. (b-h) Ribbon diagram of apoMb showing covalent labeling with increasing amounts of pLeu in the matrix in the presence of sucrose. (b) 0 (c) 0.3 (d) 1.3 (e) 2.5 (f) 5.0 (g) 11.6 (h) 20.7 %w/w pLeu. (i) Ribbon diagram of apoMb showing covalent labeling with 20.7 % w/w pLeu in the presence of Gdn HCl (1.5 M). The ribbon diagrams were generated using PyMOL (PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC) and the crystal structure of myoglobin (PDB ID 1WLA; www.rcsb.org). Helix F (H82-H97) in the myoglobin structure was modified to an unstructured region, which is observed for native apoMb at neutral pH[27].

2.4.4 MS/MS Analysis of Peptide L32-K42 (LFTGHPETLEK)

To obtain additional information on the sites of photolytic labeling, tandem MS analysis was carried out on the singly-labeled peptide L32-K42, both in the labeled and native form. After fragmenting the native peptide, almost all *b*- and *y*-ions were observed (Appendix, Table A1, I and II). However, no *b*-ions were observed in the labeled peptide product ion mass spectrum. Six *y*-ions (y_6, y_7, y_8, y_9, y_{10} and y_{11}) with z = +2 and three *y*-ions (y_4, y_5 and y_6) with z = +1 were identified by fragmenting the labeled peptide at 13 V (Fig. 2.5; Appendix Table A1, III and IV). Unlabeled *y*-ions (y_1-y_{10}) were also observed upon fragmentation of the labeled peptide (Appendix, Table A1, III and IV; dotted arrows in Fig. 2.5). Assuming that the ionization and fragmentation efficiencies of the labeled and native peptides are similar, and that the instrument is sensitive toward all possible labeled and unlabeled ions, the results suggest two possible reasons for the differences in fragmentation patterns: (1) Labeling is site-specific at Thr (peptide TLEK), since unlabeled y_1-y_3 and labeled y_4-y_{11} were observed. The presence of unlabeled y_4-y_{10} could indicate loss of label from Thr during fragmentation. (2) Labeling is heterogeneous, with multiple sites of modification ranging from Leu to Thr (peptide LFTGHPET), since labeled y_4-y_{11} and unlabeled y_1-y_{10} were observed. The presence of unlabeled y_1-y_{10} may be due to neutral loss of label from any of the labeled amino acids in peptide LFTGHPET. The absence of *b*-ions in the product ion spectrum of the labeled peptide makes it difficult to establish the cause of the differences in labeling pattern.

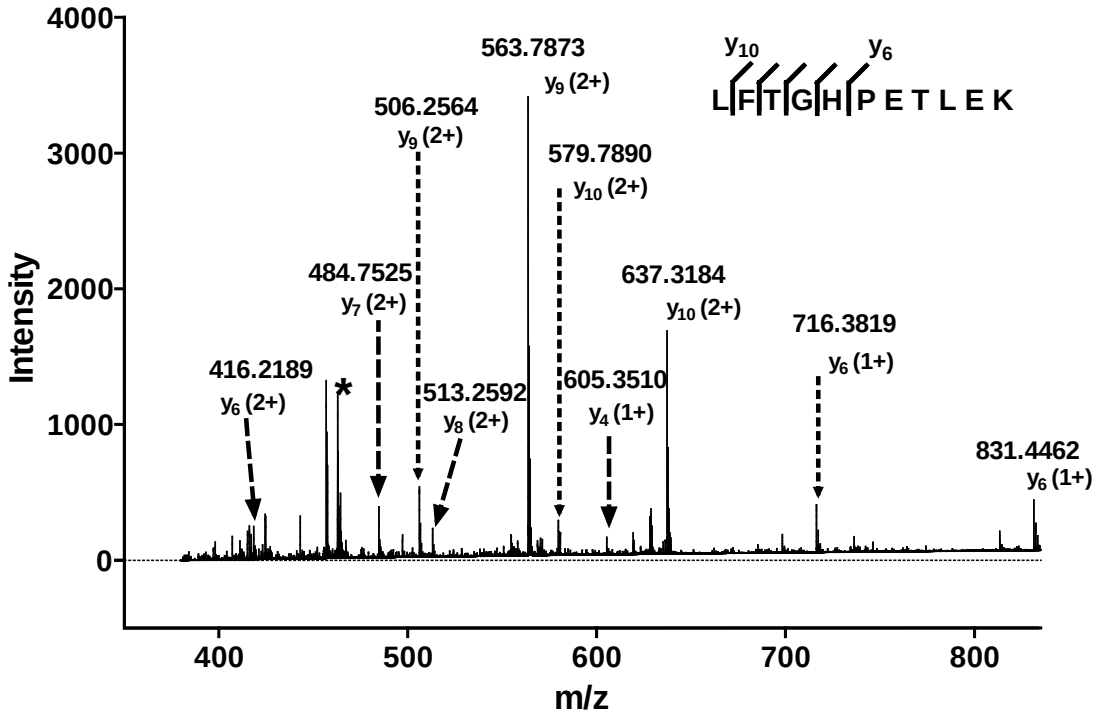


Figure 2.5. MS/MS spectrum of labeled peptide L32-K42 showing *y*-ion products obtained by CID fragmentation. The asterisk indicates the precursor peptide peak (*m/z* = 462.91). Dashed arrows represent *y*-ion peaks produced from the labeled peptide and dotted arrows represent *y*-ion peaks produced by possible loss of the label from the corresponding labeled *y*-ions. Labeled *y*-ions y_{11} (*z*=+2) and y_5 (*z*=+1) are not shown due to low abundance.

2.4.5 Formulation Effects

ApoMb lyophilized with Gdn HCl and 100x molar excess of pLeu (~ 1% w/w pLeu) was analyzed for label uptake at the peptide level. MS analysis after enzymatic digestion showed that labeling occurred at peptides L32-K42, T34-K42 and H119-F138 (Fig. 2.4(i)). This is similar to the labeling observed with sucrose at 2.5 % w/w pLeu, but with no labeling on the G helix (at the BG contact interface). Gdn HCl is expected to have a chaotropic effect on protein structure and to cause increased label uptake due to protein unfolding and higher solvent exposure. In contrast, sucrose

is expected to preserve the native structure of the protein through preferential exclusion and show lower labeling. These differences between expected and observed labeling patterns may be attributed to changes in protein side-chain environment caused by Gdn HCl.

2.4.6 Photolytic Labeling with p-benzoyl-L-phenylalanine (pBpA)

GCG (1-8)* lyophilized with and without excipients and irradiated in the solid state showed the formation of peptide-peptide adducts. Peptide dimers and trimers were observed by ESI-LC-MS; these adducts were also present in solution controls. Peptide-excipient adducts were also observed, but not for all formulations. Only two formulations (peptide-Met and peptide-Leu) showed peptide-excipient adducts in both solid and solution, while the sucrose, trehalose and urea formulations showed no peptide-excipient adducts in either solid or solution state (Table 2.2).

The first formulation containing GCG (1-8)* and pLeu showed several adducts in the lyophilized formulation. Cross-linking occurred between GCG (1-8)* and itself (dimer and trimer) and between GCG (1-8)* and pLeu (with and without the loss of N_2). The solution control showed GCG (1-8)* adducts (dimer and trimer) and GCG (1-8)*-pLeu adducts with the loss of N_2. The second formulation containing GCG (1-8)*, pLeu and sucrose produced peptide adducts (dimer and trimer), peptide-pLeu adducts (with and without the loss of N_2), pLeu-sucrose adducts with the loss of N_2 and peptide-pLeu-sucrose adducts with the loss of N_2. The solution control showed GCG (1-8)* adducts (dimer and trimer) and peptide-pLeu adducts (with the loss of N_2), but no adducts with sucrose.

Table 2.2. Cross-linked products formed after irradiation of GCG (1-8)* in various lyophilized formulations

		Cross-linked Products			
Formulation	**State**	GCG (1-8)* **Monomer**	GCG (1-8)* **Dimer**	GCG (1-8)* **Trimer**	GCG (1-8)* **+ Excipient Adduct**
GCG (1-8)* alone	Solid	+	+	+	N/A
	Solution	+	+	+	N/A
GCG (1-8)* + Sucrose[a]	Solid	+	+	+	-
	Solution	+	+	+	-
GCG (1-8)* + Trehalose[a]	Solid	+	+	+	-
	Solution	+	+	+	-
GCG (1-8)* + L-methionine[a]	Solid	+	+	+	+
	Solution	+	+	+	+
GCG (1-8)* + L-leucine[a]	Solid	+	+	+	+
	Solution	+	+	+	+
GCG (1-8)* + Urea[a]	Solid	+	+	+	-
	Solution	+	+	+	-
Formulation A[b]	Solid	+	+	+	+[d,e]
	Solution	+	+	+	+[d]
Formulation B[c]	Solid	+	+	+	+[d,e,f,g]
	Solution	+	+	+	+[d]

[a] Excipients were added in a 2:1 w/w ratio with GCG (1-8)*.

[b] Formulation A = GCG (1-8)* and pLeu in a 1:1 molar ratio.

[c] Formulation B = GCG (1-8)*, pLeu and sucrose, with a 1:1 molar ratio of GCG (1-8)* and pLeu and 1:2 w/w ratio of GCG (1-8)* and sucrose.

[d] GCG (1-8)* + pLeu adduct with loss of N_2.

[e] GCG (1-8)* + pLeu adduct without loss of N_2.

[f] pLeu + sucrose adduct with loss of N_2.

[g] GCG (1-8)* + pLeu + sucrose adduct with loss of N_2.

2.4.7 MS/MS Analysis of GCG (1-8)* Monomer and Dimer

All *b*-ions (b_1-b_8; z=+2) and several *y*-ions were detected after CID fragmentation of GCG (1-8)* monomer (data not shown). Fragmentation of the dimer from lyophilized formulations produced cross-linked product ions in addition to internal fragment (non-cross-linked) *b*- and *y*-ions (Fig. 2.6; Appendix Table A2, I and II). In order to assign product ions to cross-linked sequences, the nomenclature proposed by Schilling and coworkers was used[29]. GCG (1-8)* monomer was designated as α, while *b*- and *y*-ions (from the second monomer unit in the dimer) cross-linked with α were designated as *b*~α- and α~*y*-ions. The following cross-linked ions were detected: b_4~α, b_5~α, b_6~α, b_7~α, α~y_5 and α~y_6. Internal fragment product ions b_1, b_2, b_3, b_5, y_1, y_2, y_3 and y_4 were also detected. The evidence suggests that, for lyophilized GCG (1-8)*, peptide-peptide cross-linking occurs preferentially between pBpA and Gly residues.

In solution controls, the fragmentation of the GCG (1-8)* dimer also produced internal fragment ions and cross-linked product ions (data not shown). An unambiguous assignment of the site of cross-linking could not be made, however, suggesting multiple sites of cross-linking in solution.

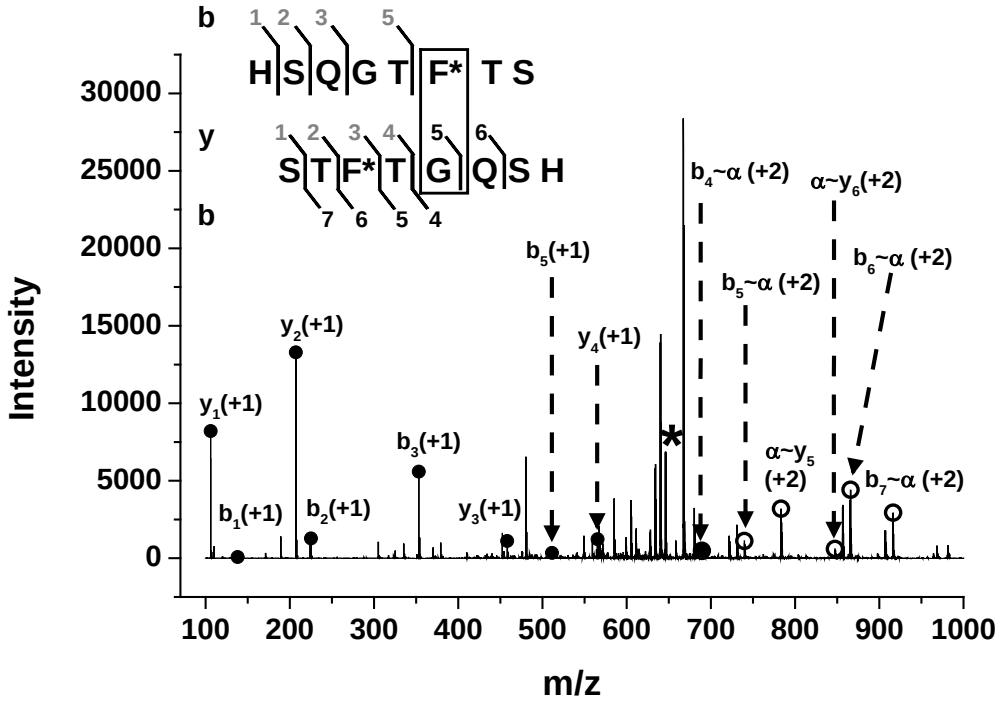


Figure 2.6. MS/MS spectrum of GCG (1-8)* dimer in the lyophilized formulation with L-leucine showing *b*- and *y*-ion products obtained by CID fragmentation. The asterisk denotes the precursor dimer peak (*m/z* = 646.28). Closed circles represent simple (non-cross-linked) *b*- and *y*-ions. Open circles represent cross-linked *b*- and *y*-ions, labeled as *b*~α and α~*y*. Inset shows *b*- and *y*-ion sequences for internal fragment ions (numbered in grey) and cross-linked ions (numbered in black) detected.

2.5 DISCUSSION

The studies presented here demonstrate successful photolytic labeling with pLeu and pBpA in lyophilized powders. To our knowledge, this is the first reported use of PAAs to study protein-protein and protein-matrix interactions in amorphous solids, though previous studies have employed PL-MS in solutions in liquid and frozen states. For example, PL-MS using pLeu has been reported in solution for myoglobin and calmodulin[18] using a 1:10,000 molar ratio of protein to pLeu and a pulsed laser for irradiation. Calmodulin was detected carrying up to 4 labels, while myoglobin showed up to 2 labels. Our studies with apoMb were unable to detect covalent labeling in solution at a 1000x

molar excess of pLeu. This may be due to differences in irradiation energy in the two studies. However, solid state labeling with 100x molar excess of pLeu showed up to 6 labeled populations in our studies, suggesting that labeling with pLeu is more efficient in the solid state than in solution, perhaps due to greater proximity of protein and pLeu, low water content and/or reduced mobility in the solid state.

PL-MS with pLeu has also been used previously to study the effect of carbene diffusion and solvent accessibility in frozen calmodulin solutions[30]. In frozen solutions, Jumper et al observed labeling at multiple sites, with higher labeling yields at Glu and Asp and no correlation with solvent accessibility. They proposed that pre-concentration of pLeu at the protein surface prior to freezing (driven by electrostatic interaction) and carbene diffusion (driven by temperature) dictated preferential labeling at carboxylate groups. In our studies in lyophilized solids, site-specific labeling such as this was not detected, though we were able to localize the label to the peptide level. Jumper et al used high-energy collisional dissociation (HCD) and electron-transfer dissociation (ETD), rather than the CID fragmentation used here. It is possible that CID fragmentation may have caused some loss of label, as has been reported previously[30, 31]. Alternatively, there could be multiple sites of labeling in the lyophilized samples, as expected given the non-specific nature of carbene reactivity. Kolbel and coworkers observed multiple cross-links between pLeu and Gly, Leu and Tyr when pLeu was incorporated within a peptide sequence and irradiated in solution[19]. Their results indicated preferential labeling based on secondary structural constraints, rather than chemical reactivity.

Though our results do not support preferential labeling of specific functional groups, preferential labeling was observed at the peptide level in lyophilized samples, which varied with the pLeu content of the solid (Fig. 2.4 (b-h)). Interestingly, the peptides labeled preferentially (i.e., labeled at the lowest pLeu concentrations and at higher concentrations) correspond to those in the molten globule of apoMb in solution. In solution, apoMb has molten globule characteristics at neutral pH, with helices A, G and H forming its core[32]. The most commonly accepted folding pathway is AGH

→ ABGH → ABCDEGH[33]. Interactions between the BG helix pair are critical in maintaining the stability of the AGH core and promoting favorable interactions between the GH helix pair[34, 35]. BG and GH interactions cause the largest decrease in solvent accessible surface area upon folding[36]. These interactions are also thought to destabilize helices E and F, which are less stable than helices A, B, G and H[35, 37]. Our ssPL-MS data showed that helices B, G and H are among the first to be labeled at lower pLeu concentrations, while helices E and F show no label uptake even at pLeu higher concentrations. This suggests that the molten globule is intact in lyophilized solids and is preferentially labeled, perhaps because amino acid side chains are exposed to pLeu in the matrix when the helices are intact. Interaction of pLeu with these regions prior to lyophilization cannot be ruled out, however.

ssPL-MS was also used to examine formulation effects on the side-chain environment, with peptide level resolution. In the presence of sucrose and 100x molar excess of pLeu (20.7 % w/w pLeu), apoMb showed labeling on all helices except E and F. When Gdn HCl was included as an excipient, CD spectroscopy of the solution prior to lyophilization showed loss of signal at 222 nm and 208 nm (data not shown), confirming that the protein had lost helicity. We expected that the Gdn HCl unfolded protein would remain unfolded after lyophilization and would be labeled to a greater extent than folded protein (e.g., in sucrose, as in our previous ssHDX studies[17, 38]). Instead, photolytic labeling was *less* in solids containing Gdn HCl than in those containing sucrose. This may be due in part to the high mass fraction of Gdn HCl in the lyophilized solid (~0.97), limiting interaction between the protein and pLeu by simple dilution. The high Gdn HCl fraction in the solid is the result of the high molar concentration used to unfold apoMb in solution, and is greater than the mass fraction of sucrose (~0.50) in the sucrose formulation. Preferential interaction of guanidinium ions with apoMb may also contribute, blocking protein-pLeu interactions and thereby inhibiting pLeu labeling[39]. Similarly, the high ionic strength of the Gdn HCl solutions prior to lyophilization may inhibit ionic interactions between pLeu and apoMb, so that labeling is reduced.

To complement studies with pLeu incorporated into the matrix as an excipient, studies were also performed with a PAA incorporated into the peptide sequence. This approach has been used to map the interactome in cells and to study PPIs in vitro[21, 40, 41]. The studies used an octapeptide derived from the N-terminal sequence of glucagon (GCG (1-8)*), with pBpA at the F6 position. Glucagon is a 29 amino-acid peptide used to treat insulin-induced hypoglycemia. The monomeric peptide is relatively unstructured in solution, but forms fibrils in acidic and alkaline pH[42-44]. Previous experimental and computational reports have assigned higher aggregation tendency to glucagon's N- and C-termini[44-46]. For example, Pedersen et al used experimental Ala mutation to study glucagon aggregation in solution and observed that mutations at residues F6, Y10, V23 and M27 decreased the rate of fibrillation at acidic pH[44]. Their results indicated that regions 6-10 and 23-27 are involved in fibrillation. Solution-state HDX-NMR studies have also indicated involvement of the N-terminus in aggregation[45].

We used GCG (1-8)* to study peptide-peptide and peptide-matrix interactions of the N-terminal sequence in solution and in the solid state, in the presence of various excipients. In solid samples, adducts of pBpA with L-Met and L-Leu excipients were observed. Adducts were not detected in lyophilized solids containing sucrose, trehalose or urea. The formation of adducts with L-Met and L-Leu may be attributable to their free, electron-rich C-H groups, which are known to react with the ketyl radical of pBpA[26, 47]. Preferential exclusion of sucrose and trehalose from the vicinity of the peptide in the pre-lyophilized solution[48, 49] may contribute to the lack of adduct formation with these excipients. Urea was selected as a negative control, since it has no C-H groups and hence is not expected to form adducts with pBpA, as was observed. In addition to protein-matrix interactions, studies with GCG (1-8)* were able to capture PPIs at the interface of dimers in the solid state, with amino acid-level resolution. The studies showed that the pBpA label interacts preferentially with G6 in forming the dimer (Fig. 2.6). In contrast, there appear to be multiple cross-linking sites in solution, perhaps due to greater mobility of the peptide in solution and/or multiple alignments of two monomer units.

Formulations containing both GCG (1-8)* and pLeu, with and without sucrose, were used to examine interactions in the solid state and in solution. In these studies, photolytic labels are present in both the peptide sequence and in the matrix. Following photoirradiation, peptide dimers and trimers, binary adducts of GCG (1-8)* with pLeu, binary adducts of pLeu with sucrose, and ternary adducts containing GCG (1-8)*, pLeu and sucrose were detected (Table 2.2). The two PAAs (i.e., pLeu, pBpA) are activated at the same wavelength, but have different mechanisms of labeling. The formation of a peptide-pLeu complex with the loss of N_2 is consistent with the mechanism of carbene labeling through pLeu activation, while adducts formed without the loss of N_2 are consistent with labeling through pBpA activation. In solids containing binary mixtures of GCG (1-8)* and pLeu, both types of adducts were detected, indicating activation and labeling via both pLeu and pBpA (Formulation A, Table 2.2). In solution, products were detected only with loss of N_2 indicating adduct formation via pLeu and not via pBpA, perhaps due to reaction of activated pBpA with water. In solids containing GCG (1-8)*, pLeu and sucrose (Formulation B, Table 2.2), ternary adducts were detected with the loss of N_2, indicating participation of both PAAs in the formation of the adduct. It is unlikely that GCG (1-8)* interacts with sucrose directly in these ternary adducts, since it did not form adducts with sucrose in the binary formulation. Together, these studies with samples containing both GCG (1-8)* and pLeu show that peptide-peptide and peptide-matrix interactions can be detected, and that reactivity of the two PAAs differs in solution and in the solid state.

The irreversible nature of photolytic labeling and complementarity to ssHDX-MS makes ssPL-MS a useful tool to study the protein environment in lyophilized powders. The primary advantage of using a PAA in the excipient matrix is the ease of labeling; the PAA simply needs to be added in an appropriate concentration to the pre-lyophilized solution. Moreover, since the PAA is only activated at a certain wavelength range and has a very short lifetime (nanosecond scale for singlet state carbene in solution, 80-120 μs for ketyl radicals in the triplet state in solution[26, 50]), the photolabeling reaction can be better controlled than with other labeling reagents such as sulfo-N-hydroxysuccinimide acetate (NHSA) that require quenching and removal of excess unreacted

reagent. Another benefit of the non-specific diazirine chemistry is that the entire protein structure can be probed, as opposed to reagents such as NHSA and 2,3- butanedione that target only lysine, N-terminal amino acids and arginine. However, non-specific labeling with diazirine-based probes poses analytical challenges. Our results showed that ssPL-MS with apoMb and pLeu could identify the location of the label at the peptide level, but MS/MS using CID failed to provide amino acid-level resolution. Labeling with PAAs incorporated in the protein sequence overcomes this hurdle by localizing the site of labeling to particular amino acid(s), with the attendant disadvantage that the PAA-labeled peptide/protein must first be synthesized. Incorporating the label in the protein sequence provided residue-level information about the sites of interaction, as shown with GCG (1-8)*.

The results have implications for formulation design and stability testing in the biopharmaceutical industry. The high resolution of ssPL-MS can facilitate rational design of formulations by allowing excipients to be selected and created based on their interactions with the protein side-chain. The information can also be used to improve protein drugs themselves through protein engineering. Ongoing work in our laboratory is developing alternate approaches to incorporate photolytic label into the protein sequence, including the use of auxotrophic cell lines and site-directed mutagenesis [51, 52]. The use of heterobifunctional cross-linkers such as succinimidyl 4,4'- azipentanoate, which contains a primary amine-specific NHS functional group and a non-specific diazirine functional group, is also being explored as an alternative approach to label incorporation.

2.6 CONCLUSIONS

Peptide-level information about protein structure and environment in lyophilized formulations was obtained using ssPL-MS. Photoactive probes can be used externally in the matrix or incorporated within the protein/peptide sequence to study side-chain accessibility or visualize protein-matrix interactions respectively.

2.7 REFERENCES

1. Walsh, G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol* **2010,** *28*, (9), 917-924.
2. Biosimilars- A global strategic business report. 2012.
3. Kasper, J. C.; Friess, W. The freezing step in lyophilization: physico-chemical fundamentals, freezing methods and consequences on process performance and quality attributes of biopharmaceuticals. *Eur J Pharm Biopharm* **2011,** *78*, (2), 248-263.
4. Luthra, S.; Obert, J. P.; Kalonia, D. S.; Pikal, M. J. Investigation of drying stresses on proteins during lyophilization: differentiation between primary and secondary-drying stresses on lactate dehydrogenase using a humidity controlled mini freeze-dryer. *J Pharm Sci* **2007,** *96*, (1), 61-70.
5. Chang, L. L.; Pikal, M. J. Mechanisms of protein stabilization in the solid state. *J Pharm Sci* **2009,** *98*, (9), 2886-2908.
6. Lai, M. C.; Topp, E. M. Solid-state chemical stability of proteins and peptides. *J Pharm Sci* **1999,** *88*, (5), 489-500.
7. Carpenter, J. F.; Pikal, M. J.; Chang, B. S.; Randolph, T. W. Rational design of stable lyophilized protein formulations: some practical advice. *Pharm Res* **1997,** *14*, (8), 969-975.
8. Tang, X.; Pikal, M. J. Design of freeze-drying processes for pharmaceuticals: practical advice. *Pharm Res* **2004,** *21*, (2), 191-200.
9. Costantino, H. R.; Langer, R.; Klibanov, A. M. Aggregation of a lyophilized pharmaceutical protein, recombinant human albumin: effect of moisture and stabilization by excipients. *Biotechnology (N Y)* **1995,** *13*, (5), 493-496.
10. Carpenter, J. F.; Prestrelski, S. J.; Dong, A. Application of infrared spectroscopy to development of stable lyophilized protein formulations. *Eur J Pharm Biopharm* **1998,** *45*, (3), 231-238.
11. Costantino, H. R.; Carrasquillo, K. G.; Cordero, R. A.; Mumenthaler, M.; Hsu, C. C.; Griebenow, K. Effect of excipients on the stability and structure of lyophilized recombinant human growth hormone. *J Pharm Sci* **1998,** *87*, (11), 1412-1420.
12. Sane, S. U.; Wong, R.; Hsu, C. C. Raman spectroscopic characterization of drying-induced structural changes in a therapeutic antibody: correlating structural changes with long-term stability. *J Pharm Sci* **2004,** *93*, (4), 1005-1018.
13. Elkordy, A. A.; Forbes, R. T.; Barry, B. W. Study of protein conformational stability and integrity using calorimetry and FT-Raman spectroscopy correlated with enzymatic activity. *Eur J Pharm Sci* **2008,** *33*, (2), 177-190.
14. Hu, K. N.; Tycko, R. What can solid state NMR contribute to our understanding of protein folding? *Biophys Chem* **2010,** *151*, (1-2), 10-21.
15. Yoshioka, S.; Miyazaki, T.; Aso, Y. Beta-relaxation of insulin molecule in lyophilized formulations containing trehalose or dextran as a determinant of chemical reactivity. *Pharm Res* **2006,** *23*, (5), 961-966.
16. Sophocleous, A. M.; Zhang, J.; Topp, E. M. Localized hydration in lyophilized myoglobin by hydrogen-deuterium exchange mass spectrometry. 1. Exchange mapping. *Mol Pharm* **2012,** *9*, (4), 718-726.
17. Sinha, S.; Li, Y.; Williams, T. D.; Topp, E. M. Protein conformation in amorphous solids by FTIR and by hydrogen/deuterium exchange with mass spectrometry. *Biophys J* **2008,** *95*, (12), 5951-5961.
18. Jumper, C. C.; Schriemer, D. C. Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Anal Chem* **2011,** *83*, (8), 2913-2920.
19. Kolbel, K.; Ihling, C. H.; Sinz, A. Analysis of peptide secondary structures by photoactivatable amino acid analogues. *Angew Chem Int Ed Engl* **2012,** *51*, (50), 12602-12605.
20. Kauer, J. C.; Erickson-Viitanen, S.; Wolfe, H. R., Jr.; DeGrado, W. F. p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide. *J Biol Chem* **1986,** *261*, (23), 10695-10700.

21. Suchanek, M.; Radzikowska, A.; Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* **2005,** *2*, (4), 261-267.
22. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2005,** *2*, (3), 201-206.
23. Soutourina, J.; Wydau, S.; Ambroise, Y.; Boschiero, C.; Werner, M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* **2011,** *331*, (6023), 1451-1454.
24. Janz, J. M.; Ren, Y.; Looby, R.; Kazmi, M. A.; Sachdev, P.; Grunbeck, A.; Haggis, L.; Chinnapen, D.; Lin, A. Y.; Seibert, C.; McMurry, T.; Carlson, K. E.; Muir, T. W.; Hunt, S., 3rd; Sakmar, T. P. Direct interaction between an allosteric agonist pepducin and the chemokine receptor CXCR4. *J Am Chem Soc* **2011,** *133*, (40), 15878-15881.
25. Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg Med Chem* **2012,** *20*, (2), 554-570.
26. Dorman, G.; Prestwich, G. D. Benzophenone photophores in biochemistry. *Biochemistry* **1994,** *33*, (19), 5661-5673.
27. Fontana, A.; Zambonin, M.; Polverino de Laureto, P.; De Filippis, V.; Clementi, A.; Scaramella, E. Probing the conformational state of apomyoglobin by limited proteolysis. *J Mol Biol* **1997,** *266*, (2), 223-230.
28. Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. *Nat Struct Biol* **1998,** *5*, (2), 148-155.
29. Schilling, B.; Row, R. H.; Gibson, B. W.; Guo, X.; Young, M. M. MS2Assign, automated assignment and nomenclature of tandem mass spectra of chemically crosslinked peptides. *J Am Soc Mass Spectrom* **2003,** *14*, (8), 834-850.
30. Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. High-resolution mapping of carbene-based protein footprints. *Anal Chem* **2012,** *84*, (10), 4411-4418.
31. Mendoza, V. L.; Antwi, K.; Baron-Rodriguez, M. A.; Blanco, C.; Vachet, R. W. Structure of the preamyloid dimer of beta-2-microglobulin from covalent labeling and mass spectrometry. *Biochemistry* **2010,** *49*, (7), 1522-1532.
32. Lin, L.; Pinker, R. J.; Forde, K.; Rose, G. D.; Kallenbach, N. R. Molten globular characteristics of the native state of apomyoglobin. *Nat Struct Biol* **1994,** *1*, (7), 447-452.
33. Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993,** *262*, (5135), 892-896.
34. Hughson, F. M.; Baldwin, R. L. Use of site-directed mutagenesis to destabilize native apomyoglobin relative to folding intermediates. *Biochemistry* **1989,** *28*, (10), 4415-4422.
35. Nishimura, C.; Dyson, H. J.; Wright, P. E. Energetic frustration of apomyoglobin folding: role of the B helix. *J Mol Biol* **2010,** *396*, (5), 1319-1328.
36. Weaver, D. L. Hydrophobic interaction between globin helices. *Biopolymers* **1992,** *32*, (5), 477-490.
37. Pappu, R. V.; Weaver, D. L. The early folding kinetics of apomyoglobin. *Protein Sci* **1998,** *7*, (2), 480-490.
38. Sophocleous, A. M.; Topp, E. M. Localized hydration in lyophilized myoglobin by hydrogen-deuterium exchange mass spectrometry. 2. Exchange kinetics. *Mol Pharm* **2012,** *9*, (4), 727-733.
39. Monera, O. D.; Kay, C. M.; Hodges, R. S. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci* **1994,** *3*, (11), 1984-1991.
40. MacKinnon, A. L.; Garrison, J. L.; Hegde, R. S.; Taunton, J. Photo-leucine incorporation reveals the target of a cyclodepsipeptide inhibitor of cotranslational translocation. *J Am Chem Soc* **2007,** *129*, (47), 14560-14561.

41. Pettelkau, J.; Schroder, T.; Ihling, C. H.; Olausson, B. E.; Kolbel, K.; Lange, C.; Sinz, A. Structural Insights into retinal guanylylcyclase-GCAP-2 interaction determined by cross-linking and mass spectrometry. *Biochemistry* **2012,** *51*, (24), 4932-4949.
42. Panijpan, B.; Gratzer, W. B. Conformational nature of monomeric glucagon. *Eur J Biochem* **1974,** *45*, (2), 547-553.
43. Onoue, S.; Ohshima, K.; Debari, K.; Koh, K.; Shioda, S.; Iwasa, S.; Kashimoto, K.; Yajima, T. Mishandling of the therapeutic peptide glucagon generates cytotoxic amyloidogenic fibrils. *Pharm Res* **2004,** *21*, (7), 1274-1283.
44. Pedersen, J. S.; Dikov, D.; Otzen, D. E. N- and C-terminal hydrophobic patches are involved in fibrillation of glucagon. *Biochemistry* **2006,** *45*, (48), 14503-14512.
45. Svane, A. S.; Jahn, K.; Deva, T.; Malmendal, A.; Otzen, D. E.; Dittmer, J.; Nielsen, N. C. Early stages of amyloid fibril formation studied by liquid-state NMR: the peptide hormone glucagon. *Biophys J* **2008,** *95*, (1), 366-377.
46. Pande, V. S. A universal TANGO? *Nat Biotechnol* **2004,** *22*, (10), 1240-1241.
47. Wittelsberger, A.; Thomas, B. E.; Mierke, D. F.; Rosenblatt, M. Methionine acts as a "magnet" in photoaffinity crosslinking experiments. *FEBS Lett* **2006,** *580*, (7), 1872-1876.
48. Lee, J. C.; Timasheff, S. N. The stabilization of proteins by sucrose. *J Biol Chem* **1981,** *256*, (14), 7193-7201.
49. Xie, G.; Timasheff, S. N. The thermodynamic mechanism of protein stabilization by trehalose. *Biophys Chem* **1997,** *64*, (1-3), 25-43.
50. Ford, F.; Yuzawa, T.; Platz, M. S.; Matzinger, S.; Fülscher, M. Rearrangement of Dimethylcarbene to Propene: Study by Laser Flash Photolysis and ab Initio Molecular Orbital Theory. *Journal of the American Chemical Society* **1998,** *120*, (18), 4430-4438.
51. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. Identification of an expanded set of translationally active methionine analogues in Escherichia coli. *FEBS Lett* **2001,** *502*, (1-2), 25-30.
52. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc Natl Acad Sci U S A* **2002,** *99*, (17), 11020-11024.

CHAPTER 3. PHOTOLYTIC CROSSLINKING TO PROBE MOLECULAR INTERACTIONS IN LYOPHILIZED SOLIDS

This chapter was published as a research article in Molecular Pharmaceutics (2013) and can be found online at http://pubs.acs.org/doi/abs/10.1021/acs.molpharmaceut.5b00183

3.1 ABSTRACT

Protein structure and local environment in lyophilized formulations were probed using high-resolution solid-state photolytic crosslinking with mass spectrometric analysis (ssPC-MS). In order to characterize structure and microenvironment, protein-protein, protein-excipient and protein-water interactions in lyophilized powders were identified. Myoglobin (Mb) was derivatized in solution with the heterobifunctional probe succinimidyl 4,4'-azipentanoate (SDA), and the structural integrity of the labeled protein (Mb-SDA) confirmed using CD spectroscopy and liquid chromatography / mass spectrometry (LC-MS). Mb-SDA was then formulated with and without excipients (raffinose, guanidine hydrochloride (Gdn HCl)) and lyophilized. The freeze-dried powder was irradiated with ultraviolet light at 365 nm for 30 min to produce crosslinked adducts that were analyzed at the intact protein level and after trypsin digestion. SDA-labeling produced Mb carrying up to 5 labels, as detected by LC-MS. Following lyophilization and irradiation, crosslinked peptide-peptide, peptide-water and peptide-raffinose adducts were detected. The exposure of Mb side chains to the matrix was quantified based on the number of different peptide-peptide, peptide-water and peptide-excipient adducts detected. In the absence of excipients, peptide-peptide adducts involving the CD, DE and EF loops and helix H were common. In the raffinose formulation, peptide-peptide adducts were more distributed throughout the molecule. The Gdn HCl formulation showed more protein-protein and protein-water adducts than the other formulations, consistent with protein unfolding and increased matrix interactions. The results demonstrate that ssPC-MS can be used to distinguish

excipient effects and characterize the local protein environment in lyophilized formulations with high resolution

3.2 INTRODUCTION

Protein drugs are the fastest growing sector of the pharmaceutical industry, a trend likely to continue given multiple impending patent expirations and a crowded biosimilars pipeline [1]. A distinguishing feature of protein drugs is the relationship between conformation, dynamics and biological function. The three-dimensional structure of proteins is the result of hydrophobic, covalent and electrostatic interactions and hydrogen bonding, and can be disrupted during manufacture, formulation and storage. It is generally accepted that maintaining a near-native-conformation in the formulation is essential for both efficacy and safety. Misfolded or partially unfolded species are often more prone to degradation and/or aggregation, complicating manufacturing and increasing the potential for adverse immunogenic reactions in patients. With the emergence of biosimilars, extensive characterization of protein structure is required to demonstrate that the product is “highly similar” to the reference product; hence, it is even more essential to reliably characterize protein structure in both solid and solution formulations with sufficient resolution.

Though proteins are often lyophilized to preserve structure during API storage and/or in the final formulation, degradation and aggregation can occur during the freeze-drying process, storage and reconstitution [2-4]. Stabilizers such as disaccharides offer some protection, but are not always effective. As a result, formulation is often a largely trial-and-error process, and can be time-consuming and expensive. Moreover, the structure of proteins in lyophilized solids is not well studied by conventional techniques, further hindering the formulation process.

Lyophilization typically produces an amorphous solid powder, unless crystallizing excipients such as mannitol are used. Formulations containing cryoprotective disaccharides such as sucrose and

trehalose have demonstrated the ability to retain native protein structure and activity [5-7]. Two hypotheses have been proposed to explain this stabilization: (1) the water replacement theory, which asserts that carbohydrates substitute for water and form hydrogen bonds with the protein and (2) the vitrification theory, which claims that the formation of a glassy solid reduces protein mobility and so preserves structure and stability. While support for each of these hypotheses has been presented by a number of groups, to date it has not been possible to probe protein-water interactions in amorphous solids directly, and so only indirect evidence regarding water replacement has been available [8-10]. To understand the interactions that control protein conformation and stability in amorphous solids, a method to directly detect both protein-matrix and protein-water interactions in lyophilized solids is needed.

Current methods used to characterize protein structure in lyophilized solids cannot detect these interactions and lack structural resolution. For example, differential scanning calorimetry (DSC) is used to study the thermal stability of lyophilized protein formulations based on the glass transition temperature (T_g), while Fourier-transform infrared (FTIR) spectroscopy has been used to determine protein secondary structure. Although these methods are used to compare formulations, their low-resolution and lack of detailed structural information are inherent limitations. Moreover, T_g is a bulk measure and does not always correlate with protein stability, since degradation mediated by local fluctuations and residual water can occur at temperatures below T_g [11, 12]. High-resolution methods such as nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography are not generally applicable to amorphous samples, since they require large amounts of sample with some long-range order and/or isotopic labeling. In addition, FTIR and NMR generate ensemble-averaged spectra that usually cannot distinguish sub-populations containing different protein conformers. Thermogravimetric analysis (TGA) and Karl-Fischer titration have been used to determine the bulk water content of the formulation, but cannot identify the local distribution of hydration within a protein molecule or spatial differences in this distribution in the sample as a whole.

To provide higher resolution structural information on proteins in lyophilized solids, our group has developed solid-state hydrogen deuterium exchange with mass spectrometric analysis (ssHDX-MS) and applied it successfully to analyze protein conformation in lyophilized powders, achieving peptide-level resolution. ssHDX-MS is able to distinguish the effects of different formulation excipients on structure in lyophilized solids [13, 14], and, in a recent study of lyophilized myoglobin formulations, provided significantly higher correlation with aggregation during storage than FTIR [15]. ssHDX-MS is not without its limitations, however. As in solution HDX, loss of the deuterium label due to back-exchange occurs rapidly for side-chain functional groups, so that only the exposure of the peptide backbone can be probed. Back exchange also necessitates rapid analysis of deuterated samples.

To address these limitations, we have developed a complementary approach to ssHDX-MS called solid-state photolytic labeling- mass spectrometry (ssPL-MS) [16]. This method utilizes a photoactive reagent such as photo-leucine (pLeu; L-2-amino-4, 4'-azipentanoic acid) as an excipient and an external probe. UV irradiation of the freeze-dried solid activates the probe, leading to covalent labeling of matrix-accessible protein side-chains. Unlike ssHDX-MS, there are no constraints with respect to experimental conditions (pH, temperature) as the pLeu label is stable and does not undergo back-exchange. Using this method, we studied excipient effects on protein side-chain environment with peptide-level resolution[16].

Building on those findings, the studies reported here present a new approach to interrogating protein interactions in amorphous solids based on photolytic crosslinking. Photolytic crosslinking has been widely used in molecular biology to study protein-protein interactions in living cells [17-19], and is adapted here to a condensed phase. In this approach, termed solid-state photolytic cross-linking with mass spectrometric analysis (ssPC-MS), a heterobifunctional crosslinking reagent (e.g. succinimidyl 4, 4'- azipentanoate; SDA) is first used to derivatize reactive side chains in the protein of interest (Appendix, Fig. A2). Following lyophilization and exposure of the powder to UV light of

a certain wavelength, a covalent bond is created between the derivatized side chain and another nearby molecule in the solid matrix. After reconstitution, the crosslinked protein is analyzed by LC-MS at the intact level, or digested enzymatically prior to LC-MS analysis to assess the number and type of adducts formed and to identify the different interactions experienced by particular proteolytic fragments. Alternatively, the reactive side chain may be engineered into the protein sequence, e.g., using photoactive amino acid derivatives such as pLeu. The length of the crosslinker can be varied by changing the length of the spacer arm, allowing the environment at different distances from the protein side chain to be probed.

ssPC-MS is similar to ssPL-MS in that both use photolytic reactions and hence are amenable to the solid state, in contrast to solution-state labeling reagents that are pH-sensitive. In ssPL-MS, the photoactive functional group is part of an excipient in the solid matrix, while in ssPC-MS the photoreactive functional group is incorporated onto protein side chains (Appendix, Fig. A3). ssPL-MS reactions are carried out in a single step while crosslinking with a heterobifunctional reagent requires two-step activation. Matrix-accessible side-chains are derivatized by covalent labeling, whereas crosslinking results in covalent linking of a side-chain with any matrix component such as protein, water or excipient. Thus labeling provides information about structural changes and matrix accessibility at the side-chain level whereas crosslinking advances this method by providing direct information about the microenvironment of a side-chain. The labeling reagent photo-leucine and the crosslinker SDA both contain a photoactive diazirine ring that is activated at 350-365 nm and forms a reactive singlet carbene (Appendix, Fig. A1). The carbene can undergo internal conversion, insert into any X-H bond (X= C, O, N, S) or add on to a C=C bond, forming covalent adducts with species within the distance of the spacer arm, including water, formulation additives (e.g. raffinose) and other protein molecules [20]. ssPC-MS and ssPL-MS are similar to ssHDX-MS in that all three techniques label the protein and reflect protein conformation in the solid state. The methods differ in that ssPC-MS and ssPL-MS map the interactions of protein side-chains with the surrounding matrix, while ssHDX-MS probes protein backbone conformation and dynamics. Unlike ssHDX-MS,

the labeling reactions of ssPC-MS and ssPL-MS are irreversible and so are not subject to the back-exchange that occurs in ssHDX-MS and other hydrogen-deuterium exchange methods.

To evaluate the utility of ssPC-MS, we used the heterobifunctional crosslinker SDA (spacer arm length 3.9 Å) to derivatize equine myoglobin (Mb) in various formulations. Crosslinking with SDA is a two-step process. In the first step, a succinimidyl ester is activated in solution at pH 6-9 and reacts with available primary amines in the protein, usually Lys side chains and the N-terminus. Following lyophilization, the photoactive diazirine group of SDA is activated by exposing the solid powder to UV-A light at 365 nm, resulting in the loss of N_2 and the formation of a reactive carbene. The carbene inserts into any X-H bond (X= C, O, N, S) or adds on to a C=C bond, forming covalent adducts with species within the distance of the spacer arm, including water, formulation additives (e.g. raffinose) and other protein molecules. Based on the number of peptide-peptide, peptide-water and peptide-excipient adducts, the microenvironment of derivatized protein side chains was characterized with high resolution. Importantly, SDA labeling and ssPC-MS provided direct evidence for the perturbation of protein structure in the solid state and provided support for regional water-replacement in lyophilized protein-carbohydrate systems.

3.3 MATERIALS AND METHODS

Holo-myoglobin from equine skeletal muscle (Mb), potassium phosphate monobasic and dibasic, Tris base, D-(+)-raffinose pentahydrate, guanidine hydrochloride (Gdn) and anhydrous dimethyl sulfoxide (DMSO) were purchased from Sigma Aldrich (St. Louis, MO). The heterobifunctional crosslinker succinimidyl 4,4'-azipentanoate (SDA) was obtained from Thermo Scientific (Rockford, IL). Trypsin was obtained from Promega (Madison, WI) and mass spectrometry-grade water, acetonitrile and formic acid from Fisher Scientific (Fair Lawn, NJ).

3.3.1 Sample Preparation

Mb was dissolved in potassium phosphate buffer (2.5 mM, pH 7.4) and dialyzed using cellulose ester tubing (MWCO 8-10 kDa, Spectrum Laboratories, Rancho Dominguez, CA) against the same buffer for 24 h. The dialyzed protein stock solution was filtered through a 0.22 µm syringe filter (Gelman Nylon Acrodisc 13) and the protein concentration measured by visible spectroscopy (extinction coefficient ε_{555nm} = 12.92 $mM^{-1}cm^{-1}$). This stock solution was used for further experiments. Stock solutions for raffinose and Gdn (3 M) in potassium phosphate buffer (2.5 mM, pH 7.4) were prepared, filtered through a 0.22 µm syringe filter and stored at 4 °C until use. A 10 mM stock solution of SDA in DMSO was prepared and stored away from light at room temperature.

3.3.2 Labeling Mb with SDA in Solution

To covalently link the SDA label to Mb via the NHS group, stock solutions of Mb and SDA were mixed such that the protein: SDA molar ratio was 1:10 (final SDA concentration 0.39 mM). The mixture was allowed to stand at room temperature in the dark for 15 min followed by quenching with Tris HCl (100 mM final concentration, pH 8.0). The labeled protein sample (hereinafter referred to as Mb-SDA) was desalted using a spin desalting column (MWCO 7 kDa; Thermo Scientific, Rockford, IL) to remove excess unreacted SDA. The desalted Mb-SDA solution was stored at 4 °C and used for crosslinking experiments.

3.3.3 Structural Integrity of Labeled Protein

Far-UV CD spectroscopy was used to determine the effect of SDA labeling on protein secondary structure. Unlabeled and SDA-labeled Mb samples (Mb labeled with 10x molar excess of SDA; 0.39 mM SDA) were diluted to 3.6 µM and molar ellipticity measured on a JASCO J-815 spectrometer (JASCO Analytical Instruments, Easton, MD) in a 0.1 cm path length quartz cuvette. Spectra were acquired from 180 nm to 260 nm at a scanning speed of 50 nm/min. Structural integrity was also monitored by measuring the extent of protein modification as a function of SDA concentration [21]. Mb was labeled with varying concentrations of SDA (0.05, 0.1, 0.26, 0.51, 0.77

and 1.02 mM) for 15 min. The reaction was quenched with Tris HCl as above and the samples diluted to 20 pmol protein for LC-MS analysis. The fraction of each labeled species was calculated from the respective peak heights in the extracted ion chromatogram (EIC):

$$F_{L,\,i} = \frac{(\text{Peak height})_{L,\,i}}{\sum_{i=0}^{10}(\text{Peak height})_{L,i}} \qquad \textbf{Equation 3.1}$$

where $F_{L,\,i}$ is the fraction of protein containing *i* SDA labels (*i* = 0,1,…,10), the numerator is the peak height for protein containing *i* SDA labels and the denominator is the sum of peak heights for unlabeled protein (protein remaining unlabeled after quenching the labeling reaction; *i* = 0) and labeled protein (*i* = 2,…,10). The concentrations of each labeled species ($P_{L,\,i}$) were calculated by multiplying $F_{L,\,i}$ by the initial protein concentration (P_0).

$$P_{L,\,i} = P_0 F_{L,\,i} \qquad \textbf{Equation 3.2}$$

The concentrations of unlabeled protein (P) and unused SDA remaining after quenching the labeling reaction (X) were calculated as follows:

$$P = P_0 F_{L,\,i=0} \qquad \textbf{Equation 3.3}$$

$$X = X_0 - \sum_{i=1}^{10} i P_{L,\,i} \qquad \textbf{Equation 3.4}$$

where P_0 is the initial protein concentration and X_0 is the initial SDA concentration. To test whether the labeling reaction is second order, the natural logarithm of the ratio (PX_0/P_0X) was plotted against X_0 to detect any deviation of the slope (second order rate constant) from linearity.

3.3.4 Lyophilization and Crosslinking in the Solid State

Stock solutions of Mb-SDA and raffinose were mixed such that the protein: raffinose ratio was 1:3 w/w (Table 3.1). A second formulation containing Mb-SDA and Gdn was prepared with a final concentration of 1.5 M Gdn (Table 3.1). The formulations were lyophilized as described previously[16]. Briefly, samples were lyophilized in borosilicate clear glass vials according to the following cycle: loading samples on shelves precooled to -2 °C, freezing at -40 °C for 50 min (shelves precooled to -2 °C), followed by drying under vacuum (70 mTorr) over 5 steps (-35 °C for 10 h, -20 °C for 8 h, -5 °C for 6 h, 10 °C for 6 h and 25 °C for 6 h). Lyophilized samples were stored at -20 °C until use. Unlabeled Mb (Mb without SDA labeling) and Mb-SDA were formulated and lyophilized separately and used as controls.

Table 3.1. Composition of lyophilized formulations.

Lyophilized Formulation	% w/w			
	Mb [a]	SDA[a]	Buffer	Excipient
Mb-SDA (10x) [b]	60.6	0.7	38.7	N/A
Mb-SDA (10x) + Raffinose (1:3 w/w)	21.5	0.2	13.7	64.5
Mb-SDA (10x) + Gdn [a] (1.5 M)	0.18	0.03	0.27	99.53

[a] Mb, myoglobin; Gdn, guanidine hydrochloride; SDA, succinimidyl 4,4'-azipentanoate.

[b] Mb-SDA (10x) denotes Mb labeled with 10x molar excess of SDA in solution.

The samples were tested for SDA-labeling-induced structural perturbations of Mb secondary structure in lyophilized powders. Solid-state Fourier transform infrared (ssFTIR) spectroscopy was carried out for the unlabeled and SDA-labeled samples using a Tensor 37 spectrometer (Bruker Optics, Billerica, MA) as described previously[13]. The moisture content of the SDA-labeled Mb formulations was determined using a gravimetric analyzer (Q5000SA; TA Instruments, New Castle, DE). The humidity chamber was equilibrated to 0 % RH at 50 ˚C. Approximately 1-2 mg of the

lyophilized powder was loaded onto the platinum sample pan and exposed at 50 ˚C, 0 % RH for 2 h, with data acquisition at 4 s intervals.

Crosslinking was initiated by irradiating the freeze-dried samples at 365 nm for 30 min using a UV Stratalinker 2400 (Stratagene Corp., La Jolla, CA) as described previously[16]. The irradiated samples were reconstituted in 200 µL ammonium bicarbonate (100 mM, pH 8.0) and stored at 4 °C until further use. For intact protein analysis using LC-MS, the reconstituted samples were diluted to 20 pmol protein with MS water containing 0.1 % formic acid.

3.3.5 Digestion of Crosslinked Protein

Mb-SDA crosslinked in the presence or absence of excipients in the solid state was reconstituted with 200 µL ammonium bicarbonate (100 mM, pH 8.0) and digested with trypsin (1:10 molar ratio of trypsin to protein) at 60 °C for 16 h, then quenched with MS water containing 0.1 % formic acid. Solution controls were prepared for all three formulations and were digested similarly after crosslinking in solution.

3.3.6 Mass Spectrometry

Labeled and crosslinked solid- and solution-state samples were analyzed using an HPLC-MS system equipped with an ESI source (1200 series HPLC, 6520 qTOF; Agilent Technologies, Santa Clara, CA). Tryptic peptides (SDA-labeled and unlabeled) and peptide adducts were separated on a ZORBAX SB-C18 column (Agilent Technologies; 1.0 x 50 mm, particle size 3.5 µm) using a gradient, as described previously[16]. MS/MS was performed on selected peptides labeled with SDA (Appendix, Table A3). The peptides were fragmented using CID (13 V) and the product ions analyzed using MassHunter software.

3.3.7 Data Analysis

The software package GPMAW (Version 9.21b3, Lighthouse Data, Odense, Denmark) was used to generate a list of theoretical masses for peptide-peptide adducts. Information regarding the protein (amino acid sequence from UniProtKB P68082), enzyme (trypsin; up to 4 missed cleavages) and crosslinker SDA (heterobifunctional; MW of the crosslinking spacer arm (C_5H_6O) 82.042 Da, amine to carboxylic acid specificity) was created in the software. Two other lists were prepared manually for peptide-raffinose and peptide-water adducts. Up to four missed cleavages with trypsin and up to four SDA labels per peptide (with up to four raffinose or water adducts, correspondingly) were considered, along with dead-end modifications (SDA-N_2), in which N_2 is lost without the formation of an adduct. The theoretical masses were compared with observed masses using MassHunter software (Agilent Technologies, Santa Clara, CA) to detect peptide-peptide, peptide-excipient and peptide-water adducts. To compare excipient effects quantitatively, peptide-peptide, peptide-water and peptide-excipient adducts were counted for each formulation. Local changes in protein-matrix interactions were quantified by calculating peptide 'crosslinking numbers', described in detail below.

3.3.7.1 Data Analysis for Crosslinking Numbers (X_{1n})

45 overlapping peptides were found to be involved in peptide-peptide adducts (refer below, section 'Data Analysis for Qualitative Matrices). To compare local excipient effects quantitatively, overlapping peptides were assigned to 8 groups, roughly corresponding to their position in the amino acid sequence (Table 3.2). The peptide-peptide adducts obtained for peptides in each group were counted and summed together to obtain a 'crosslinking number' for that group. The identity of the crosslinking partner peptide was not considered; instead all partner peptides obtained after crosslinking were considered in counting the number of adducts. Statistical analyses were performed using OriginPro (Version 8.6, OriginLab, Northampton, MA) to compare crosslinking numbers (1) between groups within the same formulation and (2) for the same group across

formulations. Crosslinking numbers were also obtained for peptide-water and peptide-excipient adducts for all 8 groups and compared statistically.

The number of SDA labels varied from 1 to 4 for tryptic peptides (see Results). This may give rise to artifacts in crosslinking numbers, since peptides carrying more SDA labels are expected to crosslink with multiple partners, giving artificially high crosslinking numbers for some groups. To avoid this bias, crosslinking numbers for each group were normalized by dividing by the average number of SDA labels (*n*) in the group (Table 3.2), calculated as follows:

$$n = \frac{[(1)(\text{number of peptides with 1 SDA}) + (2)(\text{number of peptides with 2 SDA}) + (3)(\text{number of peptides with 3 SDA}) + (4)(\text{number of peptides with 4 SDA})]}{(\text{Total number of labeled peptides (with } 1-4 \text{ SDA) in the group})}$$

A second bias may arise due to the number of tryptic peptides in each group. Groups with more peptides will generate larger crosslinking numbers, as the crosslinking number is a sum of the number of adducts obtained for all peptides within a group. For example, Group 1 has 1 tryptic peptide while Group 4 has 15 peptides (Table 3.2). Hence a larger crosslinking number for Group 4 may not necessarily reflect more interactions, but may simply be a result of the number of peptides within the group. Crosslinking numbers were therefore normalized a second time by dividing by the number of tryptic peptides in each group (Table 3.2).

The “peptide crosslinking number” is defined as the number of chemically distinct adducts detected between that peptide and another peptide fragment, water and/or raffinose. Peptide crosslinking numbers are specific to a particular peptide fragment and do not represent the total number of adducts in the protein as a whole. For a particular group (Table 3.2), the number of unique adducts detected (i.e. sum of peptide crosslinking numbers for all peptides included within the group) is summarized in the “group crosslinking number”, X_{1n}. Here, the subscripts n=1, 2, 3 indicate peptide-peptide (X_{11}), peptide-water (X_{12}) and peptide-raffinose (X_{13}) adducts, respectively. Peptides

crosslinked by up to 4 SDA labels are included, as described above. The parameter X_{11} encompasses all peptide-peptide adducts formed for peptides within a group without regard to the identity of the binding partner. For example, the following crosslinked peptide-peptide adducts were detected for Group (3) (Table 3.2) in the absence of excipients: (Leu^{32}-Lys^{45}) x (Lys^{63}-Lys^{77}), (Leu^{32}-Lys^{47}) x (Ala^{57}-Lys^{62}) and (Leu^{32}-Lys^{47}) x (Ala^{57}-Lys^{63}). Hence the peptide-peptide crosslinking number (X_{11}) for Group (3) is 3. This group also formed the following peptide-water adducts in the absence of excipients: Leu^{32}-Lys^{47} + 3SDA + $2H_2O$ and Leu^{32}-Lys^{45} + 4SDA + H_2O (NB: The reaction of diazirine with water occurs with loss of nitrogen ($-N_2$) and may or may not include dead-end modifications, so product masses are reduced accordingly). Hence the peptide-water crosslinking number (X_{12}) for Group (3) is 2. X_{1n} values were normalized by dividing by the average number of SDA labels in each group and by the number of tryptic peptides in each group, as described above. This normalized value is denoted as X_{1n}^{*}.

Table 3.2. Classification of peptides based on trypsin digestion pattern.

Group	Amino Acids Included	Number of SDA labels [a]	Number of trypsin cleavage sites	Tryptic Peptides included	Secondary Structure Elements
1	1-16	1	1	Gly^{1}-Lys^{16}	N-terminus, Helix A
2	17-31	2.5	1	Val^{17}-Arg^{31}	AB loop, Helix B
3	32-47	1.5	3	Leu^{32}-Lys^{42}, Leu^{32}-Lys^{45}, Leu^{32}-Lys^{47}	Helix B, BC loop, Helix C, CD loop
4	43-63	2	6	Phe^{43}-Lys^{45}, Phe^{43}-Lys^{47}, Phe^{43}-Lys^{50}, Phe^{43}-Lys^{56}, Phe^{43}-Lys^{62}, Phe^{46}-Lys^{50}, Phe^{46}-Lys^{56}, Phe^{46}-Lys^{62}, His^{48}-Lys^{56}, His^{48}-Lys^{63}, Thr^{51}-Lys^{56}, Thr^{51}-Lys^{62}, Thr^{51}-Lys^{63}, Ala^{57}-Lys^{62}, Ala^{57}-Lys^{63}	CD loop, Helix D, DE loop, Helix E

Table 3.2. Classification of peptides based on trypsin digestion pattern (continued).

5	63-79	1.3	4	Lys^{63}-Lys^{77}, Lys^{63}-Lys^{78}, Lys^{63}-Lys^{79}	Helix E, EF loop
6	78-98	1.4	5	Lys^{78}-Lys^{79}, Lys^{78}-Lys^{87}, Lys^{78}-Lys^{98}, Lys^{79}-Lys^{87}, Lys^{79}-Lys^{96}, Gly^{80}-Lys^{87}, Gly^{80}-Lys^{96}, Pro^{88}-Lys^{98}	Helix E, EF loop, Helix F, FG loop
7	97-118	1.2	3	His^{97}-Lys^{102}, His^{97}-Lys^{118}, Ile^{99}-Lys^{118}, Tyr^{103}-Lys^{118}	FG loop, Helix G
8	119-153	2	4	His^{119}-Lys^{133}, His^{119}-Arg^{139}, Ala^{134}-Arg^{139}, Ala^{134}-Lys^{145}, Ala^{134}-Lys^{147}, Asn^{140}-Lys^{145}, Asn^{140}-Lys^{147}, Asn^{140}-Gly^{153}, Tyr^{146}-Gly^{153}, Gln^{148}-Gly^{153}	GH loop, Helix H, C-terminus

[a] Average number of SDA labels per group (n) was calculated as described in Materials and Methods.

3.3.7.2 Data Analysis for Qualitative Matrices

At the Mb-SDA digest level, 100 non-redundant overlapping peptides labeled with 0-4 labels were detected (72 labeled and 28 unlabeled). Due to matrix heterogeneity arising from variable SDA-labeled populations, promiscuity of the reactive carbene and the amorphous nature of lyophilized solids, considering only non-overlapping peptides may result in loss of information regarding the adducts present. Hence, all overlapping peptides were included in the analysis. Since the crosslinked species differ in abundance and ionization efficiencies, and since authentic standards of the more than 100 crosslinked species produced were not available, the crosslinked adducts formed in the solid state were not quantified. Instead, a qualitative approach was used to describe the detectable interactions of the protein in lyophilized formulations. Theoretically, each of the 72 labeled peptides can crosslink with any of the 100 non-redundant overlapping peptides upon irradiation. Additional combinations are possible due to multiple dead-end modifications, and a crosslinked adduct may contain more than 2 peptides if they are crosslinked by more than one SDA molecule. The list for all such possible combinations is $> 2.2 \times 10^7$ compounds. For simplicity, only those adducts consisting of 2 peptides linked by 1 to 4 SDA were considered. For all three formulations, 69-80 % of the peptide-peptide interactions involved crosslinking through 1 SDA, with fewer adducts detected containing ≥ 2 SDA molecules. 45 overlapping tryptic peptides were found to be involved in such peptide-peptide adducts and 44 overlapping tryptic peptides labeled with 1 to 4 SDA molecules were found to interact with water and raffinose and were selected for the matrix (Fig. 4 and 5).

Peptide-peptide interactions for each formulation were mapped qualitatively as a symmetric matrix showing the interactions detected in three replicate LC-MS injections (main text, Fig. 4). In the map, color intensity indicates the number of injections (1, 2 or 3) in which a particular interaction was

detected. An interaction was considered “detected” if one or more masses corresponding to the 2 peptides linked by 1 to 4 SDA was observed. Adducts detected in a single injection represent crosslinking between multiple pairs of proteins. Repeat injections of the same sample did not always give the same adducts, perhaps due to matrix heterogeneity, variable number of SDA labels and/or low concentration of crosslinked species (Appendix, Fig. A7). For example, for Mb-SDA crosslinked in the absence of excipients, the first injection produced 41 total peptide-peptide adducts (including 16 adducts absent in the second and third injections), the second injection produced 40 adducts (including 11 adducts absent in the first and third injections) while the third injection produced 37 adducts (including 6 adducts absent in the first and second injections). An average of 54-67 % of the adducts were observed in all three injections for all formulations. Similar maps showing the maximum number of SDA linkages (1, 2, 3 or 4), maximum number of water molecules (1, 2, 3 or 4) and maximum number of raffinose molecules (1) in each adduct after a single injection are provided in SI (Appendix, Fig. A5 and A6).

45 overlapping peptides were found to be involved in peptide-peptide adducts (see Results). To compare local excipient effects quantitatively, overlapping peptides were assigned to 8 groups, roughly corresponding to their position in the amino acid sequence (Table 3.2). The peptide-peptide adducts obtained for peptides in each group were counted and summed together to obtain a ‘crosslinking number’ for that group. The identity of the crosslinking partner peptide was not considered; instead all partner peptides obtained after crosslinking were considered in counting the number of adducts. Statistical analyses were performed using OriginPro (Version 8.6, OriginLab, Northampton, MA) to compare crosslinking numbers (1) between groups within the same formulation and (2) for the same group across formulations. Crosslinking numbers were also obtained for peptide-water and peptide-excipient adducts for all 8 groups and compared statistically.

The number of SDA labels varied from 1 to 4 for tryptic peptides (see Results). This may give rise to artifacts in crosslinking numbers, since peptides carrying more SDA labels are expected to

crosslink with multiple partners, giving artificially high crosslinking numbers for some groups. To avoid this bias, crosslinking numbers for each group were normalized by dividing by the average number of SDA labels (*n*) in the group (Table 3.2), calculated as follows:

$$n = \frac{[(1)(\text{number of peptides with 1 SDA}) + (2)(\text{number of peptides with 2 SDA}) + (3)(\text{number of peptides with 3 SDA}) + (4)(\text{number of peptides with 4 SDA})]}{(\text{Total number of labeled peptides (with } 1-4 \text{ SDA) in the group})}$$

A second bias may arise due to the number of tryptic peptides in each group. Groups with more peptides will generate larger crosslinking numbers, as the crosslinking number is a sum of the number of adducts obtained for all peptides within a group. For example, Group 1 has 1 tryptic peptide while Group 4 has 15 peptides (Table 3.2). Hence a larger crosslinking number for Group 4 may not necessarily reflect more interactions, but may simply be a result of the number of peptides within the group. Crosslinking numbers were therefore normalized a second time by dividing by the number of tryptic peptides in each group (Table 3.2).

3.4 RESULTS

3.4.1 Intact Protein Labeling with SDA

Following initial succinimidyl derivatization, Mb-SDA carrying up to five labels was detected by LC-MS (Fig. 3.1). No significant secondary structural changes after SDA-labeling were detected using CD spectroscopy (Fig. 3.2A) and solid-state FTIR spectroscopy (Appendix, Fig. A4). However, this does not preclude any tertiary structure changes that may have occurred but were undetected by CD and FTIR. The relationship between the ratio (PX_0/P_0X) and SDA concentration (X) was consistent with second-order kinetics up to 0.51 mM SDA (Fig. 3.2B), further indication that minimal structural perturbation is induced by SDA labeling below this value. All further experiments were performed using a 10:1 ratio of SDA to protein with SDA concentrations below 0.51 mM to minimize effects of labeling on protein structure.

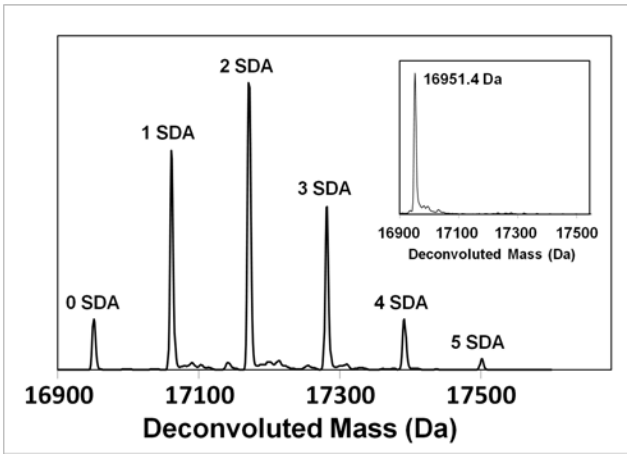


Figure. 3.1. Deconvoluted mass spectrum of Mb labeled with 10x molar excess of SDA (0.39 mM SDA). Up to 5 labeled species were detected. *Inset:* Deconvoluted mass spectrum of Mb without SDA labeling.

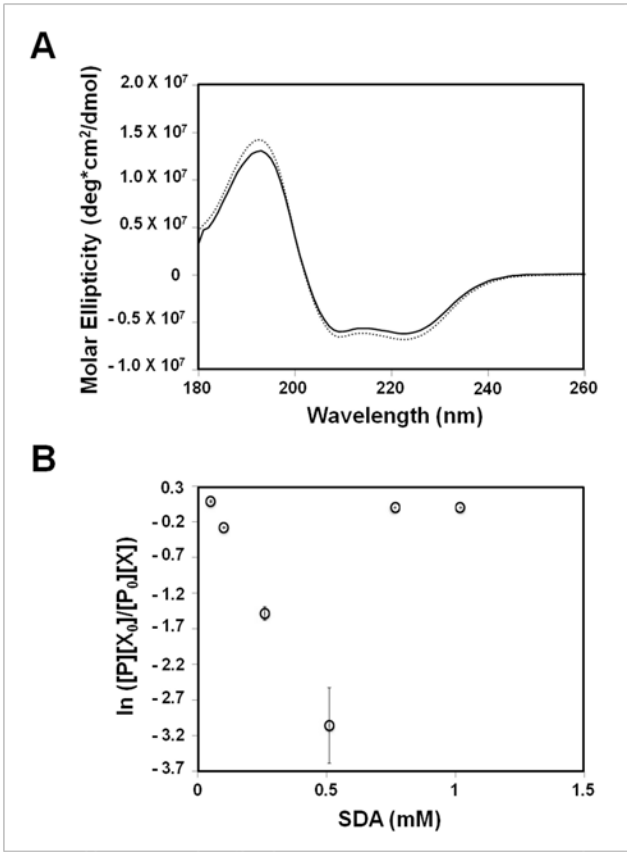


Figure 3.2. (A) Far-UV CD spectra of Mb without SDA labeling (dotted line) and Mb labeled with 10x molar excess of SDA (solid line) (B) Dose-response curve for Mb labeled with varying

concentrations of SDA. [P], protein remaining unlabeled after quenching the labeling reaction; [P_0], initial protein concentration; [X], SDA remaining unused after quenching the labeling reaction; [X_0], initial SDA concentration. The plot shows linearity up to 0.51 mM SDA (no deviation of the second order rate constant) indicating minimal perturbation of tertiary structure.

3.4.2 Peptide-Level Labeling with SDA

LC/MS analysis with proteolytic digestion was conducted to identify the sites of attachment of the SDA to Mb via an NHS-linkage. Digestion of Mb-SDA yielded a total of 72 overlapping labeled tryptic fragments that provided complete sequence coverage (Fig. 3.3). LC-MS/MS analysis conclusively established that labeling occurred on the N-terminal Gly^{1}, Lys^{42}, Lys^{50}, Lys^{56}, Lys^{87} and Lys^{147}, consistent with the accepted reaction mechanism and with preferential labeling at primary amines by NHS esters at pH 7.4. In the peptides selected for MS/MS analysis, labeling was not detected on Lys^{16}, Lys^{77}, Lys^{78}, Lys^{79}, Lys^{96} and Lys^{118}. For the other labeled peptides, the site of labeling could not be identified definitively at the amino-acid level due to low abundance and insufficient *b*- and *y*-ions. Interestingly, the peptide Asn^{140}-Lys^{147} showed 4 SDA labels, although it contains only two Lys. Similarly, peptides Val^{17}-Arg^{31} (containing no Lys), His^{119}-Lys^{133} (one Lys) and Ala^{57}-Lys^{63} (two Lys) each carried up to four SDA labels. This suggests that SDA does not label primary amines exclusively, but shows some reactivity towards other residues, as reported previously for Ser and Tyr with NHS esters [22, 23].

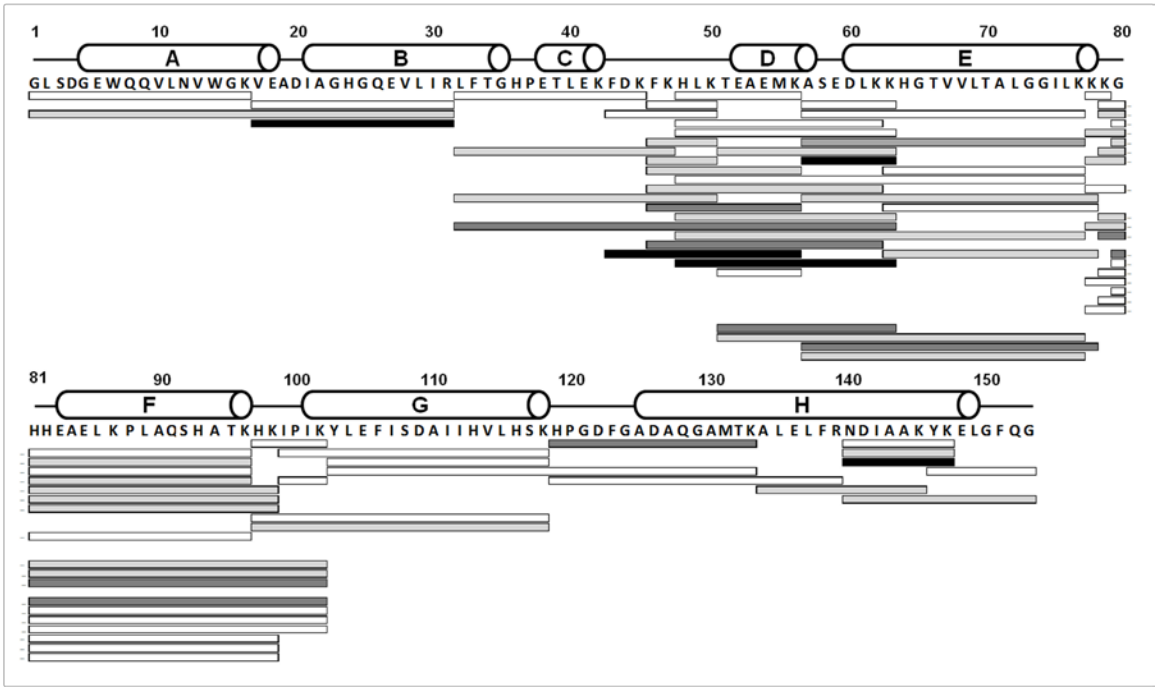


Figure 3.3. Amino acid sequence of Mb showing the domain organization with white cylinders representing the α-helices. Solid bars represent the tryptic peptides labeled with one SDA (white); two SDA (light grey); three SDA (dark grey) and four SDA (black).

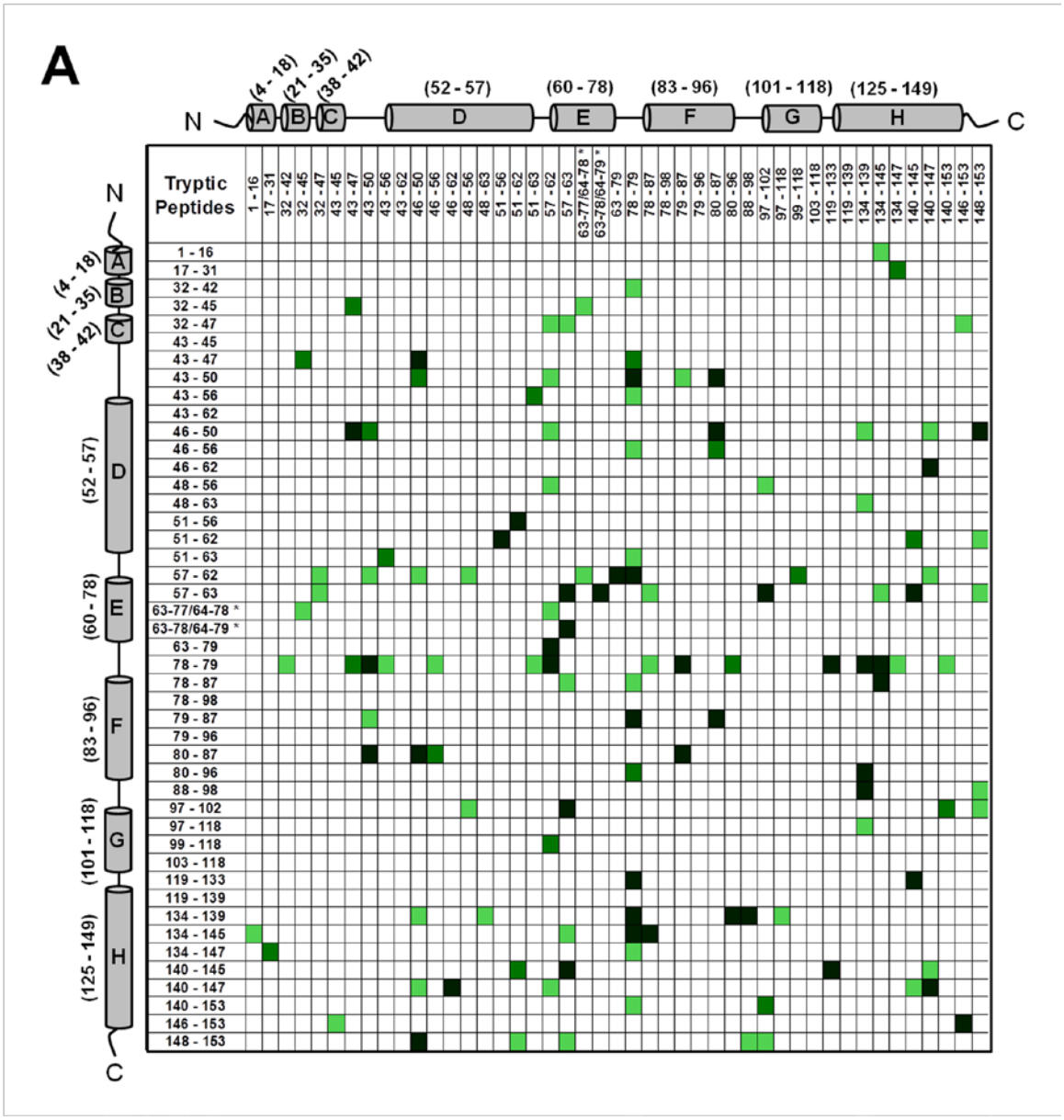
3.4.3 Crosslinking in the Solid State

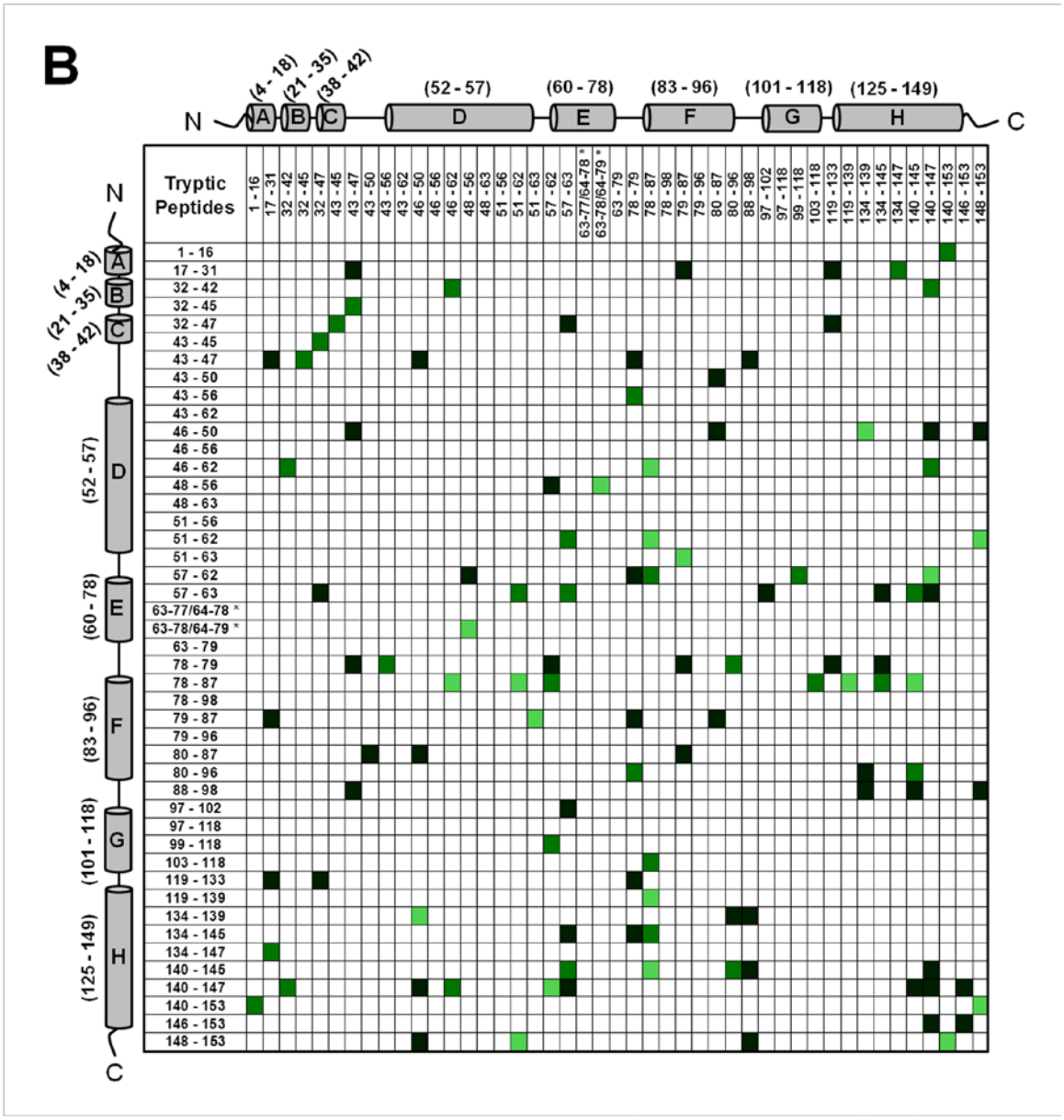
Mb-SDA irradiated in the solid state (with and without excipients) and digested with trypsin showed peptide-peptide, peptide-water and peptide-excipient adducts, as indicated by comparing the theoretical masses with the masses observed on LC-MS. The theoretically possible peptide-water and peptide-excipient adducts are listed in Table A4 (Appendix), allowing for a maximum of four SDA labels per tryptic peptide and up to four missed cleavages. A qualitative approach was used to describe the detectable interactions of the protein in lyophilized formulations. The criteria used for peptide selection and associated variability are described in Materials and Methods (refer section ‘Data Analysis For Qualitative Matrices’). Peptide-peptide adducts linked by up to 4 SDA for each formulation were mapped qualitatively as a symmetric matrix showing the interactions detected in three replicate LC-MS injections (Fig. 3.4). In the map, color intensity indicates the number of injections (1, 2 or 3) in which a particular interaction was detected. An interaction was

considered “detected” if one or more masses corresponding to the 2 peptides linked by 1 to 4 SDA was observed. The adducts detected in a single injection represent crosslinking between many pairs of protein molecules.

Intermolecular peptide-peptide adducts were detected throughout the Mb sequence in all formulations (Fig. 3.4). The crosslinking reaction is not expected to favor a particular amino acid, since the photoactive diazirine generates a singlet alkyl carbene that reacts non-specifically with X-H groups (X = C, N, O, S) or C=C bonds on exposure to UV-A light [24]. In the absence of excipient (‘control formulation’), adducts involving the CD, DE and EF loops and helix H were common, as shown in horizontal and vertical bands near the center and edge of the map (Fig. 3.4A). In formulations containing raffinose, adducts were more distributed than in the control formulation as shown by the spread of colored boxes in the matrix (Fig. 3.4B). In the Gdn HCl formulation, the map shows a number of interactions not detected in the control and raffinose formulations (Fig. 3.4C), consistent with unfolding and increased molecular contacts.

We infer that the peptide-peptide adducts for the control and raffinose formulations are intermolecular, since the calculated distance between the peptides in the crystal structure is greater than the length of the NHS spacer arm (3.9 Å) (PDB ID 1WLA; PyMOL Molecular Graphics System, Version 1.3, Schrödinger LLC). Although secondary structure changes in the control and raffinose formulations were not detected by CD and FTIR spectroscopy, it is possible that some intramolecular crosslinking may also have occurred as the result of tertiary structure perturbation. For the Gdn HCl formulation where the protein concentration (< 1% w/w) was low relative to the amount of Gdn HCl (~99% w/w) in the solid-state, the protein is considered to be fully denatured. At such a high excipient-to-protein ratio, it is likely that peptide-peptide adducts are the result of intramolecular interactions. However, intramolecular and intermolecular adducts cannot be definitively distinguished in the present work.


A
N
(4 - 18)
(21 - 35)
(38 - 42)
(52 - 57)
(60 - 78)
(83 - 96)
(101 - 118)
(125 - 149)
A
B
C
D
E
F
G
H
C
Tryptic Peptides
1 - 16
17 - 31
32 - 42
32 - 45
32 - 47
43 - 45
43 - 47
43 - 50
43 - 56
43 - 62
46 - 50
46 - 56
46 - 62
48 - 56
48 - 63
51 - 56
51 - 62
51 - 63
57 - 62
57 - 63
63-77/64-78 *
63-78/64-79 *
63 - 79
78 - 79
78 - 87
78 - 98
79 - 87
79 - 96
80 - 87
80 - 96
88 - 98
97 - 102
97 - 118
99 - 118
103 - 118
119 - 133
119 - 139
134 - 139
134 - 145
134 - 147
140 - 145
140 - 147
140 - 153
146 - 153
148 - 153


B
N
(4 - 18)
(21 - 35)
(38 - 42)
(52 - 57)
(60 - 78)
(83 - 96)
(101 - 118)
(125 - 149)
A
B
C
D
E
F
G
H
C
Tryptic Peptides
1 - 16
17 - 31
32 - 42
32 - 45
32 - 47
43 - 45
43 - 47
43 - 50
43 - 56
43 - 62
46 - 50
46 - 56
46 - 62
48 - 56
48 - 63
51 - 56
51 - 62
51 - 63
57 - 62
57 - 63
63-77/64-78 *
63-78/64-79 *
63 - 79
78 - 79
78 - 87
78 - 98
79 - 87
79 - 96
80 - 87
80 - 96
88 - 98
97 - 102
97 - 118
99 - 118
103 - 118
119 - 133
119 - 139
134 - 139
134 - 145
134 - 147
140 - 145
140 - 147
140 - 153
146 - 153
148 - 153

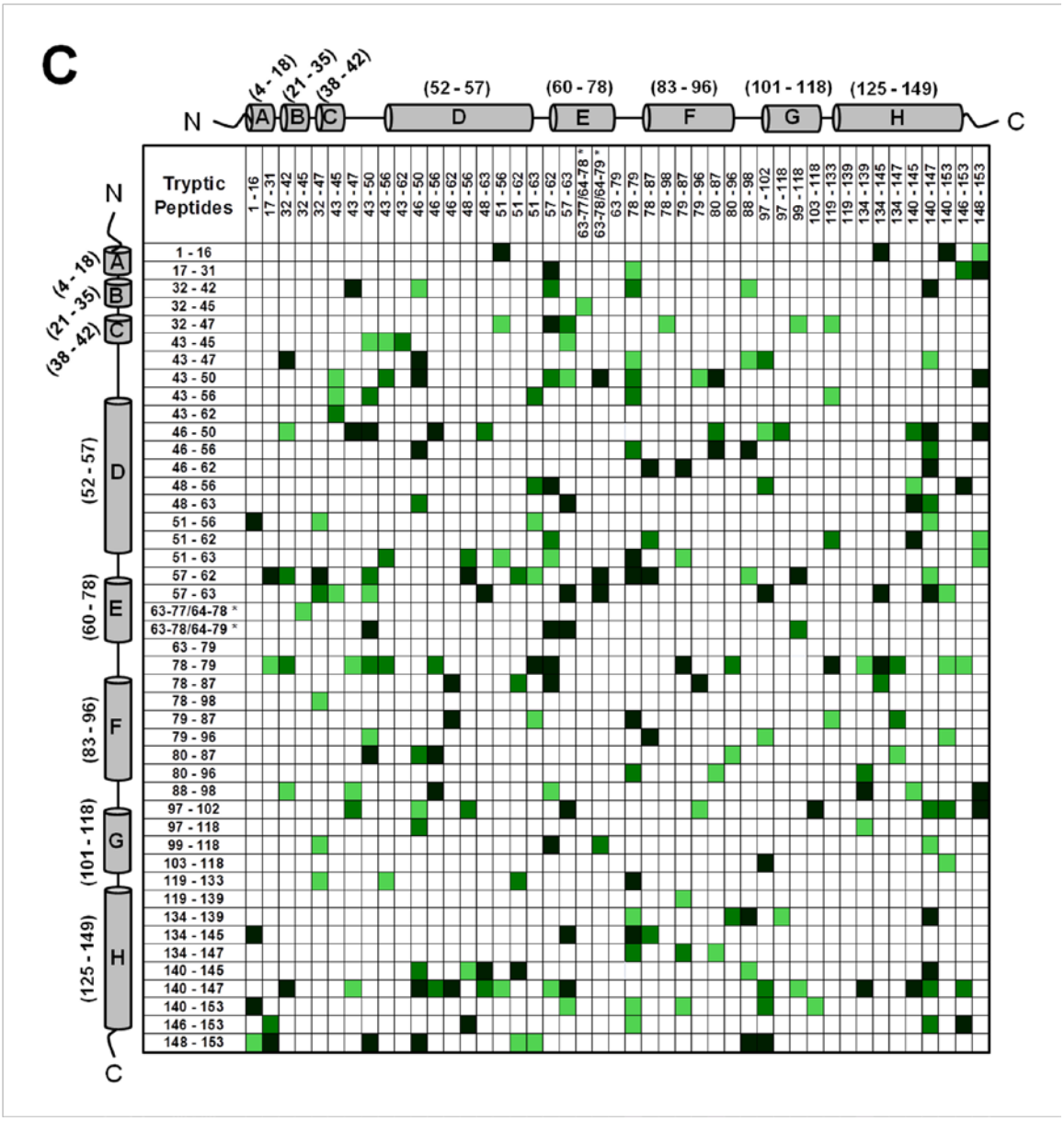


Figure 3.4. Peptide-peptide adducts formed in (A) Mb-SDA alone (control), (B) Mb-SDA with raffinose and (C) Mb-SDA with Gdn HCl formulations. Peptide-peptide adducts detected in single (■), duplicate (■) and triplicate (■) sample injection(s) are plotted. Adducts were mapped irrespective of the number of SDA linkages (1-4 SDA). The α-helices from N-terminus to C-terminus in Mb are represented by cylinders labeled A to H respectively.

*The molecular mass of tryptic fragments 63-77 and 64-78, 63-78 and 64-79 are identical and cannot be differentiated. The molecular mass for peptide-peptide adducts (32-45 x 43-47) and

(32-47 x 43-45); (79-87 x 51-63) and (78-87 x 51-62); (63-78 x 57-63) and (63-79 x 57-62) are identical and cannot be differentiated.

Peptide-water (and peptide-raffinose) adducts were mapped similarly for each formulation (both in solid- and solution state), by considering up to 4 water or raffinose molecules crosslinked with a peptide via up to 4 SDA (Fig. 3.5 and Fig. A6 in Appendix). 40 overlapping tryptic peptides labeled with 1 to 4 SDA molecules were found to interact with water and raffinose and were selected for the matrix. Peptide-water adducts were distributed across the entire molecule for all three formulations. Qualitative differences were observed, with several adducts detected only in the Gdn HCl formulation (Fig. 3.5, columns E and F). Peptide-water adducts across helices D and E were fewer in the raffinose formulation (both solid and solution-state) than in the control and Gdn HCl formulations (columns C and D). Fewer peptide-raffinose adducts were detected for the solid-state formulation than in solution (columns G and H). Only raffinose adducts, and not raffinose pentahydrate, were detected. Peptide-Gdn adducts, although detected, are not reported since their masses could not be distinguished from those of some unlabeled peptides and their abundance was not sufficient to provide definitive MS/MS fragmentation patterns.

3.4.4 Total Number of Adducts

The total numbers of chemically distinct peptide-peptide, peptide-water and peptide-excipient adducts detected in lyophilized and solution-state formulations were counted and averaged across triplicate LC-MS injections (Table 3.3). The solid-state formulations showed significantly more peptide-peptide adducts than in solution ($p < 0.05$), with the maximum number observed in the presence of Gdn HCl. The number of peptide-water adducts was significantly greater ($p < 0.05$) in the solid state than in solution for the control and Gdn HCl formulations, but was less than in solution for the lyophilized raffinose formulation. The number of peptide-raffinose adducts in the solid state was also significantly lower than in solution. Comparing the number of peptide-peptide adducts across the three lyophilized formulations, the control and raffinose formulations were not

significantly different from one another, whereas the numbers of peptide-water adducts across the three lyophilized formulations were significantly different ($p < 0.05$).

Within the control formulation, the number of peptide-peptide adducts was similar to the number of peptide-water adducts in both solution- and solid state. In the presence of raffinose, more peptide-peptide interactions were formed than peptide-water and peptide-raffinose interactions in the solid state, whereas more peptide-raffinose adducts were formed in solution. In the presence of Gdn HCl, the number of peptide-water adducts was slightly greater than peptide-peptide adducts in solution, but decreased in the solid state.

Table 3.3. Total number of peptide-peptide, peptide-water and peptide-excipient adducts detected by LC-MS in solid- and solution-state Mb-SDA formulations without excipients, with raffinose and with Gdn HCl. The numbers represent the average number of adducts (± SD) from three LC-MS injections.

Type of Adducts	Number of Adducts Detected					
	Mb-SDA		Mb-SDA + Raffinose		Mb-SDA + Gdn HCl	
	Solid	Solution	Solid	Solution	Solid	Solution
Peptide-Peptide	44.7 ± 3.5	30.3 ± 0.6	50.7 ± 3.2	31.3 ± 1.2	105.0 ± 12.3	28.3 ± 2.1
Peptide-Water	42.7 ± 3.8	30.0 ± 1.7	11.3 ± 1.2	19.7 ± 1.2	74.3 ± 10.2	34.3 ± 1.5
Peptide-Excipient	N/A	N/A	11.3 ± 0.6	41.3 ± 0.6	N/A [a]	N/A [a]

[a] Peptide-excipient adducts for the Gdn HCl formulation could not be identified unambiguously by LC-MS and are not reported.

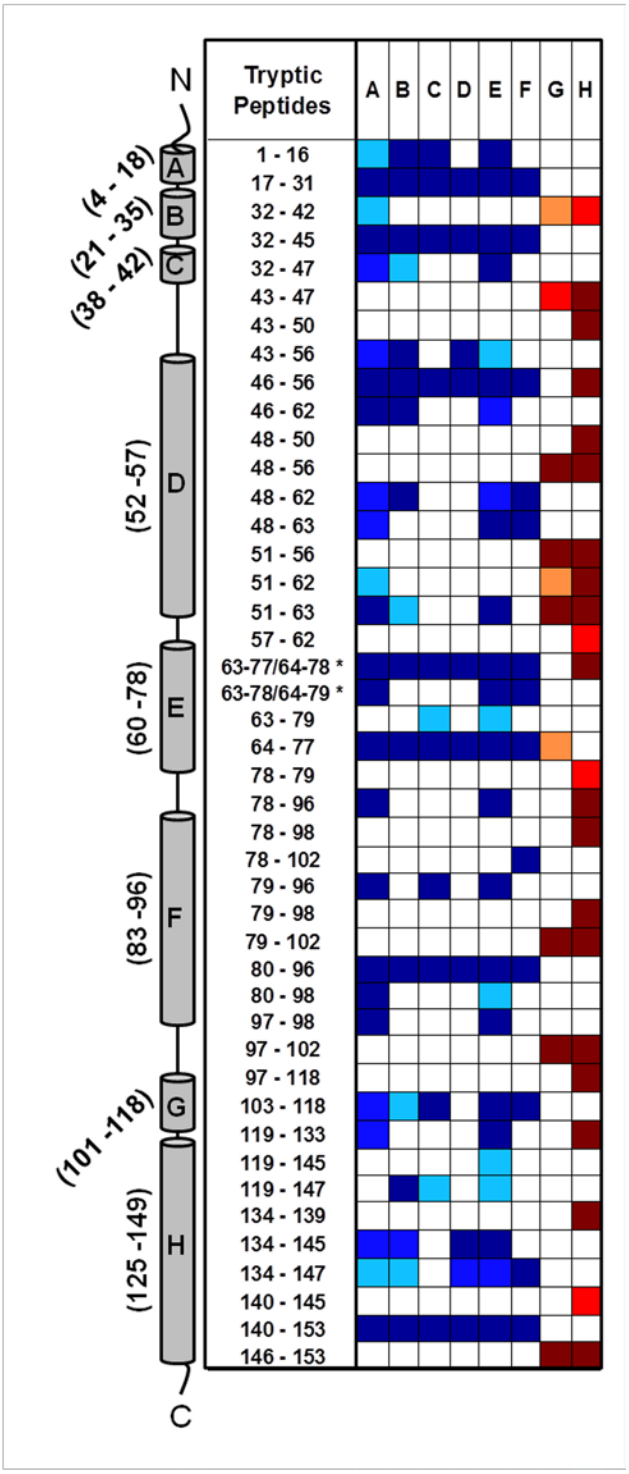


Figure 3.5. Tryptic peptides of Mb detected as peptide-water adducts in **(A)** lyophilized Mb-SDA, **(B)** Mb-SDA solution, **(C)** Mb-SDA lyophilized with raffinose, **(D)** Mb-SDA solution with raffinose, **(E)** Mb-SDA lyophilized with Gdn HCl and **(F)** Mb-SDA solution with Gdn HCl formulations. Tryptic

peptides of Mb detected as peptide-raffinose adducts in **(G)** Mb-SDA lyophilized with raffinose and **(H)** Mb-SDA solution with raffinose formulations. Peptide-water adducts detected in single (■), duplicate (■) and triplicate (■) sample injection(s) and peptide-raffinose adducts detected in single (■), duplicate (■) and triplicate (■) sample injection(s) are plotted. Adducts were mapped irrespective of the number of water or raffinose molecules linked. The α-helices from N-terminus to C-terminus in Mb are represented by cylinders labeled A to H respectively.

***** The molecular mass of tryptic fragments 63-77 and 64-78; 63-78 and 64-79 are identical and cannot be differentiated.

3.4.5 Peptide Crosslinking Numbers (X_{1n}) and Formulation Effects

To summarize the data and allow meaningful inferences about formulation differences at the local level, crosslinked peptides were assigned to 8 groups according to the overlapping tryptic fragments obtained (Table 3.2). Peptide crosslinking numbers (X_{1n}) were calculated as described in Materials and Methods. Normalized X_{1n} values (denoted X_{1n}^*) for each group were averaged across triplicate LC-MS measurements and expressed as mean ± standard deviation. The X_{1n}^* were compared: (1) across groups within a formulation and (2) within a group across formulations. One-way ANOVA demonstrated that the number of peptide-peptide adducts (X_{11}^*), peptide-water adducts (X_{12}^*) and peptide-excipient adducts (X_{13}^*) are significantly different across groups within a given formulation ($p < 0.05$). Comparing X_{1n}^* values for a group across formulations ($p < 0.05$) also showed significantly different means for all groups except Group (2) for X_{11}^*, based on Tukey's post-hoc analysis.

The crosslinking numbers can be used to compare interactions within and between formulations. For the lyophilized formulation without excipients (control formulation), the sum of group X_{11}^* values (denoted as ΣX_{11}^*) for Mb-SDA was 6.8 (± 1.7) (Table 3.4), a weighted measure of the total number of distinct peptide-peptide adducts formed. Similarly, ΣX_{12}^*, the sum of X_{12}^* values for this formulation was 6.8 (± 0.9) (Table 3.4), a weighted measure of the distinct peptide-water adducts

formed. In this formulation, the greatest X_{11}^* values were observed for Groups (4), (6) and (8), consistent with greater involvement in protein-protein interactions in these regions (Fig. 3.6A, white bars). X_{11}^* values for these groups were significantly greater than values for the other groups. Group (5) showed the greatest number of peptide-water adducts (X_{12}^*), while the remaining groups did not show significantly different X_{12}^* values (Fig. 3.6B, white bars).

Table 3.4. Crosslinking numbers for peptide-peptide adducts (X_{11}^*) values (± SD, n=3) for each lyophilized formulation.

Group	X_{11}^* (± SD)		
	Control	Raffinose	Gdn HCl
1	0.7 ± 0.6	1.7 ± 0.6	4.7 ± 0.6
2	0.4 ± 0.4	1.2 ± 0.4	1.6 ± 0.7
3	0.4 ± 0.2	0.9 ± 0.0	1.9 ± 0.3
4	1.0 ± 0.1	0.9 ± 0.1	2.6 ± 0.4
5	0.3 ± 0.6	0.8 ± 0.1	1.4 ± 0.1
6	1.8 ± 0.3	2.1 ± 0.3	3.4 ± 0.2
7	0.9 ± 0.2	0.4 ± 0.2	2.2 ± 0.3
8	1.2 ± 0.1	1.6 ± 0.2	2.6 ± 0.3
Total (ΣX_{11}^*)	6.8 ± 1.7	9.7 ± 1.5	20.5 ± 1.5

Table 3.5. Crosslinking numbers for peptide-water adducts (X_{12}^{*}) values (± SD, n=3) for each lyophilized formulation.

Group	X_{12}^{*} (± SD)		
	Control	Raffinose	Gdn HCl
1	0.3 ± 0.6	1.0 ± 0.0	2.0 ± 0.0
2	0.9 ± 0.2	0.4 ± 0.0	1.5 ± 0.2
3	0.4 ± 0.0	0.3 ± 0.1	1.4 ± 0.3
4	0.3 ± 0.1	0.0 ± 0.0	0.4 ± 0.1
5	3.3 ± 0.1	2.2 ± 0.5	2.9 ± 0.4
6	0.9 ± 0.1	0.4 ± 0.0	1.0 ± 0.4
7	0.3 ± 0.1	1.0 ± 0.0	1.1 ± 0.0
8	0.3 ± 0.3	0.2 ± 0.1	0.7 ± 0.1
Total (ΣX_{12}^{*})	6.8 ± 0.9	5.6 ± 0.5	11.0 ± 1.0

Note: The moisture contents of the control, raffinose and Gdn HCl formulations were 1.03 %, 1.92 % and 0.04 % (w/w) respectively.

Table 3.6. Crosslinking numbers for peptide-raffinose adducts (X_{13}^{*}) values (± SD, n=3) for Mb-SDA lyophilized and crosslinked in the presence of raffinose.

Group	X_{13}^{*} (± SD)
1	0.0 ± 0.0
2	0.0 ± 0.0
3	0.1 ± 0.1
4	0.2 ± 0.0
5	0.1 ± 0.1
6 [a]	0.2 ± 0.1
7	0.2 ± 0.0
8	0.2 ± 0.0
Total (ΣX_{13}^{*})	1.9 ± 0.2

[a] Note that Group (6) (spanning Lys^{78}-Lys^{98}) was expanded slightly to Lys^{79}-Lys^{102} to accommodate peptide Lys^{79}-Lys^{102} that was found to form a raffinose adduct.

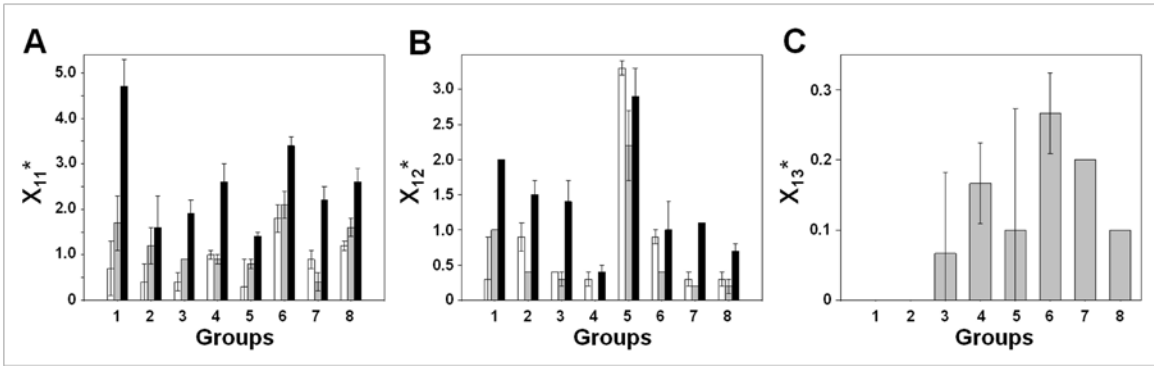


Figure 3.6. (A) Peptide-peptide adducts, (B) Peptide-water adducts and (C) Peptide-raffinose adducts detected by LC-MS. White bars represent Mb-SDA lyophilized in the absence of excipients (blank), grey bars represent Mb-SDA lyophilized with raffinose and black bars represent Mb-SDA lyophilized with Gdn HCl. X_{1n} values were counted for peptides assigned to 8 groups. Bars represent mean normalized X_{1n} values (X_{1n}^{*}) ± SD (n=3). Note that in the abscissa for panel (C),

Group (6) spanning residues Lys^{78}-Lys^{98} was expanded to Lys^{78}-Lys^{102} to accommodate peptide Lys^{79}-Lys^{102} which was found to form raffinose adducts.

In the lyophilized raffinose formulation, ΣX_{11}^* was 43 % greater than the excipient-free control (Table 3.4, Fig. 3.7A, grey bars), consistent with an increase in the number of distinct peptide-peptide adducts, although this was not a significant increase. ΣX_{12}^* for this formulation was 31% less than control, consistent with fewer distinct peptide-water adducts (Table 3.5, Fig. 3.7B, grey bars). The X_{12}^* values differed among the peptide fragment groups in the raffinose formulation (Table 3.5, Fig. 3.6B, grey bars). Group (5) again showed the greatest X_{12}^* value, while Groups (2), (3), (4), (6) and (8) showed X_{12}^* values < 1.0. Various peptide-raffinose adducts were also detected in the solid-state, with the maximum X_{13}^* for Group (6) (Table 3.6, Fig. 3.6C).

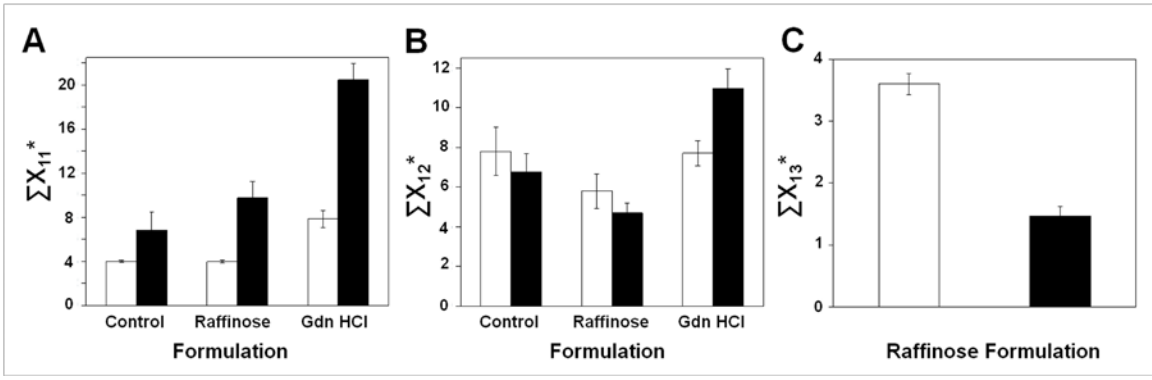


Figure 3.7. Total number of peptide-peptide (A), peptide-water (B) and peptide-raffinose (C) adducts observed for the protein in solution and solid-state (Mean ± SD (n=3)). White bars represent solution-state adducts and black bars represent adducts observed in the lyophilized formulation.

In the lyophilized Gdn HCl formulation, the ΣX_{11}^* value was 3 times greater than the control, indicating more distinct peptide-peptide adducts (Table 3.3, Fig. 3.7A, black bars). ΣX_{12}^* for this formulation was 1.6 times greater than the control, indicating more distinct peptide-water adducts (Table 3.4, Fig. 3.7B, black bars). X_{11}^* values were greater than control for all groups except Group

(2). Within the formulation, the maximum X_{11}^* values were observed for Groups (1) and (6), followed by Groups (4) and (8) (Fig. 3.6A, black bars). X_{12}^* values were significantly greater than in the control formulation for all groups except Groups (4), (5) and (6) (Fig. 3.6B, black bars). The overall increase in X_{1n}^* values is consistent with protein unfolding (as confirmed by CD and FTIR spectroscopy) and increased interactions with the matrix.

Comparing ΣX_{11}^* and ΣX_{12}^* values across lyophilized formulations, the numbers of peptide-peptide and peptide-water interactions were significantly greater in the Gdn HCl formulation (Fig. 3.7A, B). Comparing ΣX_{11}^* values across solution state formulations, peptide-peptide interactions were significantly greater in the Gdn HCl formulation, while ΣX_{12}^* values were similar across all three solution formulations (Fig. 3.7A, B). Comparing solution- and solid-state formulations, ΣX_{11}^* values were greater in the lyophilized raffinose and Gdn HCl formulations than in the corresponding solution formulations, while ΣX_{12}^* for lyophilized Gdn HCl formulation was significantly greater than in the solution state. Peptide-raffinose adducts (ΣX_{13}^*) for the solution-state raffinose formulation were significantly greater than in the solid state (Fig. 3.7C).

To determine the physical form of the excipient in the solid state, the lyophilized formulations were examined using X-ray diffraction. The control and raffinose formulations remained amorphous while the Gdn HCl formulation showed crystalline features, suggesting that the excipient had crystallized (data not shown). To relate the formation of peptide-water adducts to overall moisture content, the moisture content was determined using gravimetric analysis. The moisture contents of the control, raffinose and Gdn HCl formulations were 1.03%, 1.92% and 0.04% (w/w), respectively (Appendix, Fig. A8). The raffinose formulation showed the fewest peptide-water adducts (Table 3.4, Fig. 3.7B), although it had the highest gravimetric water content. Conversely, the Gdn HCl formulation had the lowest water content and the most peptide-water adducts.

3.5 DISCUSSION

The results presented here demonstrate that ssPC-MS can be used to map the protein microenvironment in lyophilized formulations with peptide-level resolution, providing information on the interactions of protein side chains with water, excipients and other protein molecules. Methods such as FTIR and DSC are routinely used to characterize lyophilized proteins, but provide only bulk information for the protein or matrix as a whole. ssPC-MS probes the protein side-chain environment with high resolution at the local level, based on qualitative determination of the types of adducts formed and quantitative crosslinking numbers (X_{1n}). To our knowledge, this is the first time that protein-protein and protein-matrix interactions have been mapped directly in the solid state.

The interaction maps show specific protein-matrix interactions at the peptide level and reflect the heterogeneous nature of the lyophilized matrix (Fig. 3.4 and 3.5, Appendix Fig. A5 and A6). Not all theoretically possible adducts were observed, as shown by the white boxes in the interaction maps (Fig. 3.4 and 3.5, Appendix Fig. A5 and A6). The distribution of the peptide-peptide adducts (colored or shaded boxes) across the maps suggests that Mb molecules are oriented in the solid matrix in several different ways, allowing different adducts to be formed with the same peptide. Despite the lack of long-range order, there appear to be constraints in the control and raffinose formulations that prevent the formation of many of the theoretically possible adducts (white boxes). That these constraints are related to protein structure is supported by the presence of a greater number of unique adducts in the Gdn HCl formulation.

The interactions detected by ssPC-MS provide additional information about protein structure and environment in the solid matrix. For example, several peptide-peptide adducts were observed in the control and raffinose formulations for peptides spanning the CD, DE and EF loop regions (Fig. 3.4A, B). Motions of loop regions are linked to conformational transitions involving the helices of Mb[25]. It has been shown experimentally and computationally that the CD and EF loops of holoMb

are especially flexible, allowing for efficient ligand binding[26-28]. This loop flexibility may result in better protein-protein contacts in the solid state. Any disruption of the salt bridge between residues Lys[45] and Asp[60] that normally stabilizes the DE link with the CD loop [29, 30] could also contribute to increased loop mobility in lyophilized solids and make the loop regions more prone to interactions.

In contrast to the loop interactions, peptide-peptide adducts were rarely observed for helices A and G in the control and raffinose lyophilized formulations. In the folding pathway of holoMb, helices A, G and H fold first and form a stable molten globule core[31, 32]. This is followed by folding of helices B, C, D, E and F and heme coordination in a hydrophobic pocket between helices E and F. The structure of holoMb is further stabilized by interhelix contacts between helices B-G, B-E, G-H, F-H, A-E and A-H[25, 33]. Here, limited crosslinking for helices A and G may be explained by persistence of the molten globule in the solid state. However, helix H formed several peptide-peptide adducts despite being part of the molten globule. Previous ssHDX-MS have shown loss of backbone protection in helix H upon lyophilization[15], which may result in increased crosslinking for helix H. No peptide-peptide crosslinking was observed between helices B-G, B-E, G-H and A-E in the control and raffinose lyophilized formulations (Fig. 3.4A, B), perhaps as a result of interhelix interactions preserved in the solid state and the inability of the side-chains to participate in crosslinking.

The results show that crosslinking provides high-resolution information about protein-matrix interactions in both solution and solid state. While data matrices (Fig. 3.4 and 3.5) can be used to qualitatively describe the type of adducts formed, the number of adducts (Table 3.3) can be used as a simple metric to quantify the fraction of interactions with each matrix component. The number of peptide-matrix adducts can be affected by events such as unfolding, phase separation and aggregation. Similar numbers of peptide-peptide and peptide-water adducts in the control formulation (Table 3.3) suggest that there is equal likelihood of protein-protein and protein-water contacts in the absence of excipients, assuming similar carbene reactivity with protein and water.

The presence of interacting excipients and the nature of the interaction is expected to alter the number of adducts, as observed with raffinose and Gdn HCl (Table 3.3).

Low X_{11}^* and high X_{12}^* values for E helix in all the three formulations suggest that the side-chains in this region interact primarily with water. HoloMb contains a distal His^{64} residue (helix E) in the heme-binding pocket; this residue is involved in modulating heme-ligand affinity by binding to water [34, 35]. This suggests that there is a hydration layer around helix E, in which may be responsible for the high frequency of water adducts with helix E peptides. Interestingly, the sites (peptides) of raffinose crosslinking were not coincident with the sites for water crosslinking, even in solution (Fig. 3.5; only 2 peptides Lys^{46}-Lys^{56} and Lys^{63}-Lys^{77}/Lys^{64}-Lys^{78} out of 22 showed crosslinking with both water and raffinose). Such observations have implications regarding the water replacement hypothesis, as discussed below. The effects of Gdn HCl on local protein structure could be established, as observed by the increased peptide-peptide crosslinking in the solid state (Fig. 3.4C). That X_{11}^* and X_{12}^* values for most groups were greater in the Gdn HCl formulation than in the other two is also consistent with greater matrix exposure.

The water replacement hypothesis states that lyophilized proteins are stabilized by hydrogen bonds to sugars and other excipients in the dried state, which replace the hydrogen bonds to water that stabilize the structure in solution [36]. Previous studies have tested this hypothesis by measuring the extent of hydrogen bonding using the FTIR band area at 1583 cm^{-1}, which corresponds to carboxylate- hydrogen bonding[8, 37]. The band area was found to be smaller in proteins lyophilized in the absence of carbohydrate excipients, but increased with increasing carbohydrate concentration [8]. Though FTIR results provide some support for the water replacement hypothesis, ssPC-MS allows these interactions to be interrogated directly. The presence of peptide-water adducts in all three formulations studied here confirms that residual water is present at the protein surface after lyophilization (Fig. 3.5A-F). Overall peptide-peptide interactions increased in the solid state for the raffinose and Gdn HCl formulations, compared to solution (Table 3.3, Fig. 3.7A). This

is expected as a result of freeze-concentration and increased protein-protein contacts. The magnitude of this increase in protein-protein contacts is greatest in the lyophilized Gdn HCl formulation (Fig. 3.7A). This is in part due to protein unfolding and also a possible result of Gdn HCl crystallization. In solution, Gdn HCl binds to proteins and promotes unfolding. This binding may explain the absence of more peptide-water adducts in the solution Gdn HCl formulation (Table 3.3, Fig. 3.7B), even though the protein is partially unfolded at 1.5 M Gdn HCl [38]. We hypothesize that when Gdn HCl crystallizes, the SDA-labeled residues are free to crosslink with water molecules, resulting in increased ΣX_{11}^* and ΣX_{12}^* in the solid state compared to solution.

Preferential exclusion of carbohydrates is known to occur in solution at concentrations ≥ 0.2 M [39, 40]. In this study, raffinose was present at a concentration of ~ 2 mM; in such a dilute solution, it is unlikely that there is appreciable raffinose exclusion. Hence, increased molecular mobility and diffusion in solution are more likely to contribute to the observed protein-raffinose crosslinking. While peptide-peptide crosslinking was greater in the lyophilized raffinose formulation than in solution (Table 3.3, Fig. 3.7B), peptide-raffinose adducts were fewer in the solid state than in solution (Table 3.3, Fig. 3.7C). Although the reduced mobility in the solid state is expected to produce a greater number of intermolecular contacts and crosslinked adducts, the observed results may be due to raffinose micro-phase separation in the solid state or water replacement by raffinose in the solid state. If hydrogen bonds between Mb and raffinose in the solid state indeed replaced hydrogen bonds to water in solution, one would expect to observe new peptide-raffinose adducts in the solid state that were not observed in solution. In addition, these new raffinose adducts should be detected in peptides for which peptide-water adducts were observed in solution. Neither of these was observed with SDA crosslinking in solution- and solid-state raffinose formulations. A 3:1 w/w ratio of raffinose to protein translates to about 100 molecules of raffinose per protein molecule, so that it is unlikely that the solid is too dilute in raffinose, at the bulk level, for reaction with SDA to occur. Thus, water replacement is the less likely explanation for the peptide-water and peptide raffinose crosslinking observed here (Fig. 3.5C, D, G, H, 3.7B, C). Chatterjee et al. have reported

crystallization and phase separation of raffinose during annealing, although the final lyophilized product was amorphous [41]. In this work, the lyophilized raffinose formulation was amorphous as observed by X-ray diffraction (data not shown), but raffinose crystallization during freezing or micro-phase separation in the lyophilized product may have occurred and would not be detected. The extent to which the hygroscopic nature of raffinose and raffinose-water hydrogen bonding contributes to decreased peptide-water interactions in the lyophilized raffinose formulation is also unknown. Moreover, the relative reactivity of the carbene in the solid and solution states and as well as its rates of reaction with raffinose and water may also contribute, and to date have not been explored.

While ssPC-MS offers higher resolution structural information than conventional methods such as FTIR, experimental and computational limitations remain and should be noted. A current experimental limitation is the inability to resolve the sites of crosslinking at the amino-acid level with ESI-CID-MS/MS. Higher resolution mass spectrometry instruments (e.g., FTICR-MS) may be useful for this purpose. Analysis could be simplified by better control of the sites and extent of protein derivatization. This could be accomplished through optimizing pH, SDA concentration and reaction time to limit labeling at side-chains that do not contain a primary amine, or by the use of site-specific derivatization chemistries (e.g., click chemistry). Computationally, though theoretical mass lists for derivatized and crosslinked peptides can be prepared using software such as GPMAW, the complete list can be quite long, particularly for larger proteins such as antibodies. In addition, matching the theoretical list with observed masses using software such as MassHunter can be time-consuming due to potential false positives that need to be verified manually. Recent improvements in bioinformatics such as xProphet could allow improved identification of crosslinked peptides with low false positive rates[42]. However this technique requires MS/MS information, preferably from high-resolution LTQ-Orbitrap instruments. More broadly, the effects of water activity (RH) and excipient type on protein-protein and protein-matrix interactions require further investigation, as does the relationship of the interactions detected by ssPC-MS to storage stability.

Nevertheless, the results presented here demonstrate the potential of ssPC-MS for probing protein-protein and protein-matrix interactions in lyophilized solids with high resolution.

3.6 CONCLUSIONS

ssPC-MS provided qualitative and quantitative measures of protein side-chain interactions in lyophilized formulations. The environment of lyophilized Mb could be visualized with high resolution at the peptide-level and excipient differences quantified using X_{1n}^{*} values.

3.7 REFERENCES

1. Aggarwal, R. S. What's fueling the biotech engine-2012 to 2013. *Nat Biotechnol* **2014,** *32*, (1), 32-39.
2. Kasper, J. C.; Friess, W. The freezing step in lyophilization: physico-chemical fundamentals, freezing methods and consequences on process performance and quality attributes of biopharmaceuticals. *Eur J Pharm Biopharm* **2011,** *78*, (2), 248-263.
3. Luthra, S.; Obert, J. P.; Kalonia, D. S.; Pikal, M. J. Investigation of drying stresses on proteins during lyophilization: differentiation between primary and secondary-drying stresses on lactate dehydrogenase using a humidity controlled mini freeze-dryer. *J Pharm Sci* **2007,** *96*, (1), 61-70.
4. Lai, M. C.; Topp, E. M. Solid-state chemical stability of proteins and peptides. *J Pharm Sci* **1999,** *88*, (5), 489-500.
5. Costantino, H. R.; Carrasquillo, K. G.; Cordero, R. A.; Mumenthaler, M.; Hsu, C. C.; Griebenow, K. Effect of excipients on the stability and structure of lyophilized recombinant human growth hormone. *J Pharm Sci* **1998,** *87*, (11), 1412-1420.
6. Cleland, J. L.; Lam, X.; Kendrick, B.; Yang, J.; Yang, T. H.; Overcashier, D.; Brooks, D.; Hsu, C.; Carpenter, J. F. A specific molar ratio of stabilizer to protein is required for storage stability of a lyophilized monoclonal antibody. *J Pharm Sci* **2001,** *90*, (3), 310-321.
7. Sane, S. U.; Wong, R.; Hsu, C. C. Raman spectroscopic characterization of drying-induced structural changes in a therapeutic antibody: correlating structural changes with long-term stability. *J Pharm Sci* **2004,** *93*, (4), 1005-1018.
8. Allison, S. D.; Chang, B.; Randolph, T. W.; Carpenter, J. F. Hydrogen bonding between sugar and protein is responsible for inhibition of dehydration-induced protein unfolding. *Arch Biochem Biophys* **1999,** *365*, (2), 289-298.
9. Prestrelski, S. J.; Tedeschi, N.; Arakawa, T.; Carpenter, J. F. Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers. *Biophys J* **1993,** *65*, (2), 661-671.
10. Crowe, J. H.; Carpenter, J. F.; Crowe, L. M. The role of vitrification in anhydrobiosis. *Annu Rev Physiol* **1998,** *60*, 73-103.
11. Hancock, B. C.; Shamblin, S. L.; Zografi, G. Molecular mobility of amorphous pharmaceutical solids below their glass transition temperatures. *Pharm Res* **1995,** *12*, (6), 799-806.
12. Chang, L. L.; Shepherd, D.; Sun, J.; Tang, X. C.; Pikal, M. J. Effect of sorbitol and residual moisture on the stability of lyophilized antibodies: Implications for the mechanism of protein stabilization in the solid state. *J Pharm Sci* **2005,** *94*, (7), 1445-1455.
13. Sophocleous, A. M.; Zhang, J.; Topp, E. M. Localized hydration in lyophilized myoglobin by hydrogen-deuterium exchange mass spectrometry. 1. Exchange mapping. *Mol Pharm* **2012,** *9*, (4), 718-726.

14. Li, Y.; Williams, T. D.; Topp, E. M. Effects of excipients on protein conformation in lyophilized solids by hydrogen/deuterium exchange mass spectrometry. *Pharm Res* **2008,** *25*, (2), 259-267.
15. Moorthy, B. S.; Schultz, S. G.; Kim, S. G.; Topp, E. M. Predicting protein aggregation during storage in lyophilized solids using solid state amide hydrogen/deuterium exchange with mass spectrometric analysis (ssHDX-MS). *Mol Pharm* **2014,** *11*, (6), 1869-1879.
16. Iyer, L. K.; Moorthy, B. S.; Topp, E. M. Photolytic labeling to probe molecular interactions in lyophilized powders. *Mol Pharm* **2013,** *10*, (12), 4629-4639.
17. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2005,** *2*, (3), 201-206.
18. Sato, S.; Mimasu, S.; Sato, A.; Hino, N.; Sakamoto, K.; Umehara, T.; Yokoyama, S. Crystallographic study of a site-specifically cross-linked protein complex with a genetically incorporated photoreactive amino acid. *Biochemistry* **2011,** *50*, (2), 250-257.
19. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc Natl Acad Sci U S A* **2002,** *99*, (17), 11020-11024.
20. Jumper, C. C.; Schriemer, D. C. Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Anal Chem* **2011,** *83*, (8), 2913-2920.
21. Mendoza, V. L.; Vachet, R. W. Protein surface mapping using diethylpyrocarbonate with mass spectrometric detection. *Anal Chem* **2008,** *80*, (8), 2895-2904.
22. Swaim, C. L.; Smith, J. B.; Smith, D. L. Unexpected products from the reaction of the synthetic cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate), DTSSP with peptides. *J Am Soc Mass Spectrom* **2004,** *15*, (5), 736-749.
23. Kalkhof, S.; Sinz, A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal Bioanal Chem* **2008,** *392*, (1-2), 305-312.
24. Dubinsky, L.; Krom, B. P.; Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg Med Chem* **2012,** *20*, (2), 554-570.
25. Elber, R.; Karplus, M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science* **1987,** *235*, (4786), 318-321.
26. Schulze, B. G.; Grubmüller, H.; Evanseck, J. D. Functional Significance of Hierarchical Tiers in Carbonmonoxy Myoglobin: Conformational Substates and Transitions Studied by Conformational Flooding Simulations. *Journal of the American Chemical Society* **2000,** *122*, (36), 8700-8711.
27. Cammarata, M. B.; Brodbelt, J. S. Structural characterization of holo- and apo-myoglobin in the gas phase by ultraviolet photodissociation mass spectrometry. *Chemical Science* **2015,** *6*, (2), 1324-1333.
28. Seno, Y.; Go, N. Deoxymyoglobin studied by the conformational normal mode analysis. I. Dynamics of globin and the heme-globin interaction. *J Mol Biol* **1990,** *216*, (1), 95-109.
29. Angeloni, L.; Feis, A. Protein relaxation in the photodissociation of myoglobin-CO complexes. *Photochem Photobiol Sci* **2003,** *2*, (7), 730-740.
30. Balasubramanian, S.; Lambright, D. G.; Marden, M. C.; Boxer, S. G. CO recombination to human myoglobin mutants in glycerol-water solutions. *Biochemistry* **1993,** *32*, (9), 2202-2212.
31. Barrick, D.; Baldwin, R. L. Stein and Moore Award address. The molten globule intermediate of apomyoglobin and the process of protein folding. *Protein Sci* **1993,** *2*, (6), 869-876.
32. Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993,** *262*, (5135), 892-896.
33. Pappu, R. V.; Weaver, D. L. The early folding kinetics of apomyoglobin. *Protein Science* **1998,** *7*, (2), 480-490.
34. Rohlfs, R. J.; Mathews, A. J.; Carver, T. E.; Olson, J. S.; Springer, B. A.; Egeberg, K. D.; Sligar, S. G. The effects of amino acid substitution at position E7 (residue 64) on the kinetics of ligand binding to sperm whale myoglobin. *J Biol Chem* **1990,** *265*, (6), 3168-3176.

35. Quillin, M. L.; Arduini, R. M.; Olson, J. S.; Phillips, G. N., Jr. High-resolution crystal structures of distal histidine mutants of sperm whale myoglobin. *J Mol Biol* **1993,** *234*, (1), 140-155.
36. Carpenter, J. F.; Crowe, J. H. An infrared spectroscopic study of the interactions of carbohydrates with dried proteins. *Biochemistry* **1989,** *28*, (9), 3916-3922.
37. Remmele, R. L., Jr.; Stushnoff, C.; Carpenter, J. F. Real-time in situ monitoring of lysozyme during lyophilization using infrared spectroscopy: dehydration stress in the presence of sucrose. *Pharm Res* **1997,** *14*, (11), 1548-1555.
38. Bismuto, E.; Colonna, G.; Irace, G. Unfolding pathway of myoglobin. Evidence for a multistate process. *Biochemistry* **1983,** *22*, (18), 4165-4170.
39. Lee, J. C.; Timasheff, S. N. The stabilization of proteins by sucrose. *J Biol Chem* **1981,** *256*, (14), 7193-7201.
40. Xie, G.; Timasheff, S. N. The thermodynamic mechanism of protein stabilization by trehalose. *Biophys Chem* **1997,** *64*, (1-3), 25-43.
41. Chatterjee, K.; Shalaev, E. Y.; Suryanarayanan, R. Raffinose crystallization during freeze-drying and its impact on recovery of protein activity. *Pharm Res* **2005,** *22*, (2), 303-309.
42. Leitner, A.; Walzthoeni, T.; Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat Protoc* **2014,** *9*, (1), 120-137.

CHAPTER 4. STUDYING PROCESS AND FORMULATION EFFECTS ON PROTEIN STRUCTURE IN LYOPHILIZED SOLIDS USING MASS SPECTROMETRIC METHODS

4.1 ABSTRACT

Myoglobin (Mb) was lyophilized in the absence (Mb-A) and presence (Mb-B) of sucrose in a pilot-scale lyophilizer with or without controlled ice nucleation. Cake morphology was characterized using scanning electron microscopy (SEM) and changes in protein structure were monitored using solid-state Fourier-transform infrared spectroscopy (ssFTIR), solid-state hydrogen-deuterium exchange-mass spectrometry (ssHDX-MS) and solid-state photolytic labeling-mass spectrometry (ssPL-MS). The results showed greater variability in nucleation temperature and irregular cake structure for formulations lyophilized without controlled nucleation. Controlled nucleation resulted in nucleation at ~ -5 °C and uniform cake structure. Formulations containing sucrose showed better retention of protein structure by all measures than formulations without sucrose. Samples lyophilized with and without controlled nucleation were similar by most measures of protein structure. However, ssPL-MS showed the greatest pLeu incorporation and more labeled regions for Mb-B lyophilized with controlled nucleation. The data support the use of ssHDX-MS and ssPL-MS to study formulation and process-induced conformational changes in lyophilized proteins.

4.2 INTRODUCTION

Proteins are often marketed in lyophilized form or stored as lyophilized powders after purification and prior to formulation. Lyophilization begins with freezing, followed by primary drying to remove bulk ice by sublimation and secondary drying to desorb unfrozen water. Proteins are subjected to various stresses during lyophilization, including freeze-concentration and denaturation at the ice-surface, pH shifts and dehydration-induced aggregation [1-4]. Proteins can be protected from some of these stresses by modifying the formulation and using stabilizing excipients [5-8]. However, the

process itself can play a role in determining critical quality attributes (CQAs) of the product. For example, the cooling rate and type of thermal treatment used may result in entrapment of metastable intermediates that can crystallize at higher temperatures during manufacturing or storage [9-11]. Inadequate drying temperature or time can also result in product failure due to increased moisture content, and processing or storage above the glass transition temperature (T_g) can result in degradation [12, 13]. Cake elegance can be adversely affected by aggressive processing above the collapse temperature, producing various degrees of macro- and micro-collapse [14, 15]. Higher temperatures during lyophilization can degrade reducing carbohydrate excipients via the Maillard reaction [16], which may reduce their stabilizing effects.

The freezing step is critical, since parameters such as the degree of supercooling and rate of freezing can affect the morphology of ice crystals, which in turn affects the rate of primary drying [17, 18]. When a solution is supercooled to a large degree, ice nucleation occurs at lower temperatures with little time for ice crystal growth, resulting in smaller pores in the dried solid. These small ice crystals result in small pores and offer greater resistance to flow of water vapor through the porous bed of partially dried solids. This necessitates the use of a longer primary drying step to remove crystalline water. In contrast, a lower degree of supercooling is associated with a slower rate of freezing from a relatively small number of large ice crystals. Since ice nucleation is stochastic, cycles without controlled freezing are expected to nucleate over a range of temperatures, resulting in longer freezing times and heterogeneous ice crystal morphology. Variability is also introduced by vial position within the lyophilizer chamber [19], since vials near the door and walls of the lyophilizer chamber receive more heat via radiation than those near the center of the chamber. Together, these factors result in inter-vial- and inter-batch heterogeneity. If left uncontrolled, this variability may be magnified when a process is scaled from a laboratory bench-top lyophilizer to a production freeze-dryer. Heat- and mass-transfer differences between pilot and production freeze-dryers may also play a role, so that the same lyophilization cycle may produce variable product CQAs at different scales. Controlling the freezing step is critical to producing uniform ice crystal morphology,

resulting in less variability between samples and faster drying. The freezing rate can also affect product stability, since smaller ice crystals formed by fast freezing present a greater surface area for potential protein adsorption and unfolding. Aggregation at the protein solution-ice interface was implicated in the lyophilization-induced instability of human growth hormone [20], recombinant human factor XIII [21], lactate dehydrogenase and immunoglobulin G [3].

Strategies to control nucleation include the use of an ice fog as a seeding technique and rapid depressurization to induce spontaneous nucleation. While the effect of controlled ice nucleation on primary drying time has been well documented [22-25], its effect on protein structure is not well understood. Controlled nucleation at a lower degree of supercooling results in larger ice crystals with lower surface area for protein adsorption. In addition, since all vials nucleate at the same time and primary drying time is decreased, the residence time of proteins at the ice surface is reduced. These two factors are expected to produce a product that is more stable than one lyophilized without controlled nucleation. The effect of depressurization-induced controlled nucleation on product characteristics was reported for a monoclonal antibody [26]. Although the drying time was reduced by ~ 10 h and cake appearance improved to some extent with controlled nucleation, there was no significant impact on aggregation as detected by UV spectroscopy and size exclusion chromatography (SEC). Secondary structure was not altered significantly, as quantified by circular dichroism (CD) spectroscopy. The process did not affect binding to protein-A, suggesting that the tertiary structure was also intact, at least at the binding site. Other studies investigated the effect of different lyophilization cycles on protein conformation and cake structure [27, 28]. While cycle variations typically led to altered cake morphology as detected using scanning electron microscopy (SEM), conformational changes could not be detected using conventional ssFTIR and solution-state CD and fluorescence spectroscopy.

In this work, the effects of controlled nucleation and lyophilizer scale on protein structure were examined. The ControLyo® depressurization technology was used for controlled nucleation.

Myoglobin (Mb) lyophilized in a LyoStar freeze-dryer with or without controlled nucleation showed no significant changes in structure at the backbone and side-chain levels, as determined by ssFTIR, solid-state hydrogen-deuterium exchange (ssHDX-MS) and solid-state photolytic labeling (ssPL-MS), respectively. However, formulation effects were dominant and protein structure was better protected at the backbone in the presence of sucrose. The results indicate that local structure remains unaltered by controlled nucleation and that ssHDX-MS and ssPL-MS can be used to detect process- and formulation-induced changes in protein structure.

4.3 MATERIALS AND METHODS

4.3.1 Materials

Equine skeletal muscle holomyoglobin (Mb), sucrose, potassium phosphate dibasic and ammonium bicarbonate were purchased from Sigma Aldrich (St. Louis, MO). Potassium phosphate monobasic (anhydrous) was purchased from Amresco (Solon, OH). D_2O was obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA) and photo-leucine (L-2-amino-4, 4′-azipentanoic acid) from Thermo Scientific (Rockford, IL). Mass spectrometry-grade water, acetonitrile and formic acid were purchased from Fisher Scientific (Fair Lawn, NJ). Spectra/Por dialysis tubing (MWCO 8000-10000 Da) was used to dialyze the protein prior to formulation (Spectrum Laboratories, Inc., Rancho Dominguez, CA). Syringes (Beckton, Dickinson and Company, Franklin Lakes, NJ) and 0.2 µm Acrodisc® syringe filters (Pall Corporation, Ann Arbor, MI) were used to filter the dialysate.

4.3.2 Sample Preparation

Mb was dissolved in potassium phosphate buffer (2.5 mM, pH 7.4) and dialyzed overnight against the same buffer using dialysis tubing. The dialyzed protein was filtered using a syringe filter and its concentration measured using UV spectroscopy (8453 UV-Vis, Agilent Technologies, Santa Clara, CA) and the molar extinction coefficient ε_{555} nm = 12.92 $mM^{-1}cm^{-1}$ (obtained from Sigma Aldrich product information sheet for equine skeletal muscle myoglobin, product M0630). This stock

solution (345 µM) was used for all formulations. A 20 mg/mL stock solution of sucrose was prepared by dissolving sucrose in potassium phosphate buffer (2.5 mM, pH 7.4) and stored at 4 °C until use. Similarly, a 30.9 mM stock solution of photo-leucine (pLeu) was prepared using the same buffer and stored at 4 °C until use.

Two formulations were prepared for lyophilization: a control formulation containing Mb and buffer ('Mb-A') and a formulation containing Mb, sucrose and buffer ('Mb-B'). Stock solutions of Mb, sucrose and buffer were mixed such that the final Mb concentration was 70 µM and the ratio of Mb to sucrose was 1:1 w/w. For photolytic labeling studies, pLeu was added to Mb-A and Mb-B such that the molar ratio of pLeu to Mb was 100:1. The weight fractions of each component are listed in Table 1. The formulations were filled in glass tubing vials (USP Type I glass; 2 ml capacity) with 13 mm necks. The fill volume was 500 µL for ssHDX-MS, ssPL-MS, thermogravimetric analysis, X-ray diffraction and solid-state Fourier-transform infrared spectroscopy samples. A 3 mL fill in 10 mL-capacity glass beakers was used for scanning electron microscopy.

Table 4.1. Weight fractions of components of lyophilized formulations

Lyophilized Formulation	% w/w			
	Mb [a]	Sucrose [a]	Buffer	pLeu [a]
Mb-A	91.7	N/A	8.3	N/A
Mb-B	42.9	42.9	14.1	N/A
Mb-A + pLeu	46.0	N/A	15.1	38.8
Mb-B + pLeu	31.5	31.5	10.4	26.6

[a] Mb, myoglobin; pLeu, photo-leucine (L-2-amino-4,4′-azipentanoic acid)

4.3.3 Lyophilization

Both Mb-A and Mb-B formulations were lyophilized with and without controlled nucleation during the freezing step. For freezing with controlled nucleation (LyoStar 3 with ControLyo®, SP Industries,

Inc., Gardiner, NY), the vials were equilibrated at 5 °C for 30 min, followed by pressurization with argon gas to 28 psig and a decrease in temperature to -5 °C. A ramp rate of 1 °C/min was used and the vials were equilibrated at -5 °C for 60 min. At the end of this step, the chamber was rapidly depressurized to 1 psig to induce controlled nucleation and the temperature was further reduced to -45 °C at a ramp rate of 1 °C/min and held overnight to complete the freezing step. For freezing without controlled nucleation (LyoStar II, SP Industries), the vials were equilibrated at 5 °C for 60 min (ramp rate 1 °C/min) and then frozen at -45 °C overnight.

To minimize process variability, vials frozen with and without controlled nucleation were dried simultaneously in the LyoStar 3 lyophilizer. Vials frozen without controlled nucleation in LyoStar II were quickly transferred to LyoStar 3 and held at -45 °C for an additional 30 min. A vacuum of 70 mTorr was used and the shelf temperature was increased stepwise with a ramp rate of 1 °C/min (-35 °C for 600 min, -20 °C for 600 min, -5 °C for 360 min, 10 °C for 360 min). Product temperature was monitored during freezing and drying using 30 gauge Type T thermocouple sensors. Thermocouples were placed either inside the solution or taped to the outside of the vial. Chamber pressure was monitored using a capacitance manometer (CM) and Pirani gauge. The end point of each drying step was determined using a pre-set CM/Pirani gauge differential, wherein the cycle advanced to the next step if the differential was reached at the end of the previous step. Since the presence of thermocouples inside the solution may affect sample integrity, thermocouple-containing vials were not used for characterization studies.

4.3.4 X-ray Powder Diffraction

Samples were analyzed by X-ray powder diffraction to detect any crystallinity after lyophilization. Diffractograms were collected on a 2θ θ scan from 7-35° 2θ with 0.02° increments using a SmartLab diffractometer (Rigaku Americas, TX).

4.3.5 Scanning Electron Microscopy (SEM)

Intact lyophilized cakes were removed from beakers and mounted on a sample holder with double-sided tape. The cakes were carefully sliced using a blade and sputter-coated with carbon graphite. Images were obtained using a JCM-6000 NeoScope benchtop instrument (JEOL USA, Peabody, MA) in the high vacuum, 15 kV mode.

4.3.6 Solid-State Fourier Transform Infrared (ssFTIR) Spectroscopy

FTIR spectra were acquired for all lyophilized samples using a Tensor 37 spectrometer (Bruker Optics, Billerica, MA), as described previously [29]. 128 scans were obtained at 4 cm^{-1} resolution and spectra were processed using OPUS software (v. 6.5, Bruker Optics), by cutting around 1600-1700 cm^{-1}, smoothing and baseline correcting before obtaining second derivative spectra.

4.3.7 Thermogravimetric Analysis (TGA)

Moisture sorption kinetics was measured using TGA (QA5000SA, TA Instruments, New Castle, DE) to study its effect on hydrogen-deuterium exchange kinetics. ~ 1 mg of lyophilized protein (with or without sucrose) was loaded onto a metallized quartz sample pan. The loosely bound water was removed by heating the sample to 40 °C, 0 % RH inside the sample chamber until the weight change was < 0.01 % and equilibrated for 1 h. The sample was then equilibrated at 5 °C, 0 % RH for 1 h, followed by moisture sorption at 5 °C, 43 % RH for 3 h inside the sample chamber. Initial moisture content was calculated from the weight change before and after equilibration at 40 °C, 0 % RH for 1 h.

4.3.8 Solid-state Hydrogen Deuterium Exchange- Mass Spectrometry (ssHDX-MS)

Lyophilized vials were uncapped and placed in a sealed desiccator equilibrated at 43 % RH over D_2O (obtained with a saturated solution of potassium carbonate in D_2O). HDX was allowed to proceed at 5 °C for 0.5, 2, 6, 12, 24, 48, 84 and 120 h. At each time point, samples were withdrawn,

stoppered and flash-frozen in liquid N_2 to quench the exchange reaction. The samples were stored at -80 °C until analysis.

HPLC-MS (1200 series HPLC, ESI-qTOF 6520, Agilent Technologies, Santa Clara, CA) was used to measure deuterium uptake at the intact level, as described previously [29, 30]. Deuterated samples were reconstituted with 2 mL of ice-cold quench buffer (5 % methanol, 0.2 % formic acid in LC-MS-grade water, pH 2.5) and injected into a refrigerated box housing the HPLC valves, tubing and protein microtrap at ~ 0 °C to reduce back-exchange. The protein was eluted with a gradient mobile phase that increased from 30 % to 80 % acetonitrile over 3 min. Mass spectra for deuterated samples were deconvoluted using MassHunter software (Agilent Technologies), and the number of deuterons incorporated was calculated by subtracting the mass of the undeuterated protein from the mass of the deuterated protein. Peak widths of the deconvoluted intact protein spectra were measured at 20 % peak height.

4.3.9 Solid-state Photolytic Labeling- Mass Spectrometry (ssPL-MS)

Lyophilized Mb-A and Mb-B vials containing pLeu were uncapped and irradiated at 365 nm for 40 min using Stratalinker 2400 (Stratagene Corp., La Jolla, CA) as described previously [31]. The cakes were then reconstituted with 500 µL ammonium bicarbonate (100 mM, pH 8.0), diluted to 20 pmol and analyzed at the intact protein level by LC-MS. The same elution parameters were used as described above for ssHDX-MS, but with analysis performed at room temperature since the pLeu label does not undergo back-exchange. The fractions of protein populations with 0-2 labels ($F_{L=0,1,2}$) were calculated from peak heights observed on the extracted ion chromatogram (EIC), as follows:

$$F_{L=0,1,2} = \frac{(\text{Peak height})_{L=0,1,2}}{(\text{Peak height})_{unlabeled\ protein\ L=0} + (\text{Peak height})_{labeled\ protein\ L=1,2}}$$

To identify the sites of labeling at the peptide level, the labeled Mb formulations were digested with trypsin (1:10 ratio of trypsin to protein) at 60 °C for 16 h. Labeled peptides were analyzed by LC-MS using the method described previously [31] and identified with MassHunter, using a theoretical mass list of tryptic peptides with up to 2 labels.

4.3.10 Statistical Analysis

Process- and excipient effects on protein structure were compared statistically using GraphPad Prism software (GraphPad software, version 6; La Jolla, CA). One-way ANOVA followed by Tukey's test was used for multiple comparisons ($p = 0.05$).

4.4 RESULTS

In the absence of solution and vial impurities, a solution can remain in a supercooled liquid state without undergoing phase transition into a solid as the temperature is lowered below its freezing point. When water crystallizes into ice, an increase in product temperature is detected associated with the latent heat of fusion. Product temperature bias due to the presence of the thermocouple in solution must be considered, since it can increase the nucleation temperature [32]. To avoid thermocouple-related artifacts in temperature, some thermocouples were taped to the outside of the vial. Vials with thermocouples also tend to dry faster, since ice nucleation at higher temperatures produces larger ice crystals with lower resistance during drying [33]. Although vials with thermocouples may not measure true product temperature, a comparison of product temperatures between processes can still be made using thermocouple-containing vials.

4.4.1 Effect of Freezing Step on Nucleation Temperature

Product temperatures for Mb-A (excipient-free formulation) and Mb-B (sucrose-containing formulation) frozen with and without controlled nucleation were recorded (Fig. 4.1 A, B). In the absence of controlled nucleation, variable nucleation temperatures were recorded ranging from -5.7 °C to -9.4 °C (mean ± SD, Fig. 4.1A) and all the thermocouple-containing vials nucleated within

10 min. With controlled nucleation, all the vials containing thermocouples inside the solution nucleated simultaneously at -5.3 ± 0.2 °C (Fig. 4.1B), irrespective of formulation. Vials with thermocouples taped to the outside reported nucleation temperatures of -4.0 °C (Mb-A) and -4.4 °C (Mb-B) with controlled nucleation.

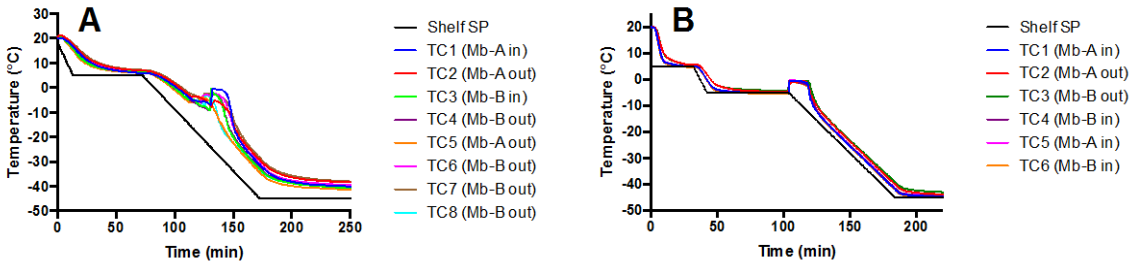


Figure 4.1. Product temperature profiles for excipient-free (Mb-A) and sucrose-containing (Mb-B) myoglobin lyophilized without controlled nucleation (panel A) and with controlled nucleation (panel B). Thermocouple probes (TC) were placed inside the vial (labeled Mb-A in, Mb-B in) or taped to the outside of the vial (labeled Mb-A out, Mb-B out). The shelf set point (shelf SP) temperature is represented by the black line in panels A and B.

4.4.2 Measurement of Supercooling during the Freezing Step

To estimate the extent of supercooling in freezing with uncontrolled nucleation, thermocouples were taped to the outside of the vial to avoid thermocouple-related bias. However, this may introduce uncertainty as a thermal lag is expected for thermocouples placed on the outside of the vial. Hence thermocouples were placed both outside and inside the vial for 2 vials each for Mb-A and Mb-B. These thermocouples reported similar temperatures, suggesting that thermal lag was minimum and the nucleation temperature recorded was reproducible (Fig. 4.2 A, B). Mb-A showed supercooling to ~ -7.5 ± 0.4 °C and Mb-B showed supercooling to ~ -9.1 ± 0.4 °C, as measured by thermocouples placed both inside and outside the vial. Moreover, vials with thermocouples only in contact with the solution showed greater supercooling than vials with thermocouples only on the outside (Fig. 4.1A).

The reason for this is unclear and the data are contrary to the expected result of increased nucleation temperature due to the presence of the thermocouple.

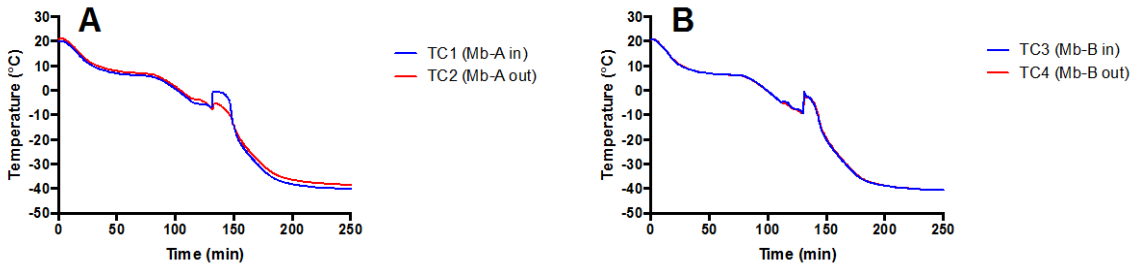


Figure 4.2. Extent of supercooling for excipient-free (Mb-A, panel A) and sucrose-containing (Mb-B, panel B) myoglobin formulations lyophilized without controlled nucleation. Thermocouples were placed inside the vial in contact with the solution and also taped to the outside of the same vial. Mb-A showed supercooling to ~ -7.5 ± 0.4 °C and Mb-B showed supercooling to ~ -9.1 ± 0.4 °C as measured by thermocouples placed both inside and outside the vial.

4.4.3 Solid-State Characterization by X-ray Diffraction and FTIR

All the formulations were amorphous after freeze-drying and produced X-ray diffraction patterns consistent with amorphous materials (Appendix Fig. A9). Both Mb-A and Mb-B retained some degree of alpha helicity, observed as a peak at about 1652 cm^{-1} in the second-derivative FTIR spectra (Fig. 4.3). Mb-B with and without controlled nucleation showed a more intense and narrower alpha helix peak than excipient-free Mb-A formulations, although there was no appearance of beta sheet peaks in any of the spectra (Fig. 4.3). Overall, the two processes appear to have affected peak intensity and position, but without the formation of non-native secondary structure.

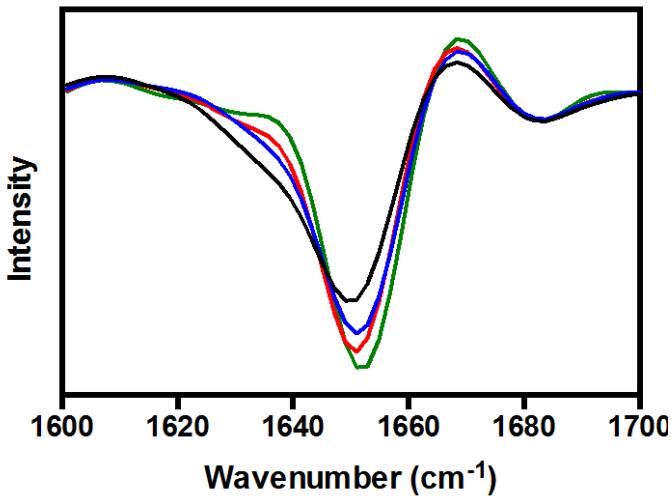


Figure 4.3. Second-derivative solid-state FTIR spectra for excipient-free (Mb-A) and sucrose-containing (Mb-B) lyophilized myoglobin formulations. Spectra were obtained for Mb-A and Mb-B lyophilized with controlled nucleation (black and red curves respectively) and without controlled nucleation (blue and green curves respectively).

4.4.4 Cake Morphology

SEM images showed porous cakes for all formulations. Mb-A and Mb-B lyophilized with controlled nucleation showed more uniform pore structure than the formulations lyophilized without controlled nucleation (Fig. 4.4, Appendix Fig. A10). Mb-A lyophilized with and without controlled nucleation showed large plate-like morphology (Appendix Fig. A10, panels A, B) and Mb-B lyophilized with and without controlled nucleation showed thinner plates (Appendix Fig. A10, panels C, D). The data are consistent with the degree of supercooling (Fig. 4.1), since the formulations lyophilized with controlled nucleation showed a lower degree of supercooling and larger pores than those lyophilized without controlled nucleation.

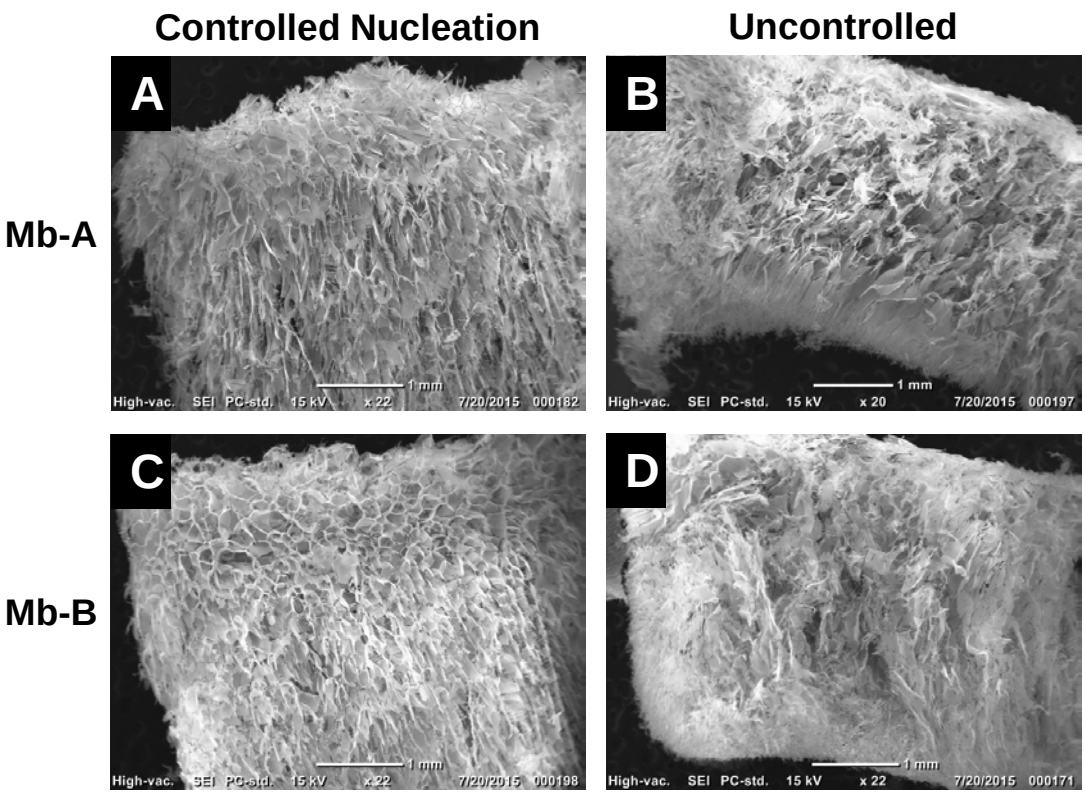


Figure 4.4. SEM images of excipient-free (Mb-A; panels A, B) and sucrose-containing (Mb-B; panels C, D) lyophilized myoglobin formulations. Panels A, C: Formulations lyophilized with controlled nucleation; Panels B, D: Formulations lyophilized without controlled nucleation. The top and bottom of each image represent the top and bottom of the cake respectively. Scale bars are set at 1 mm.

4.4.5 Effects of Formulation and Process on Protein Backbone by ssHDX-MS

Deuterated intact protein mass spectra showed an increase in mass compared to the undeuterated protein (Fig. 4.5). Since sorption and diffusion of D_2O from the vapor phase into the solid must precede the hydrogen-deuterium exchange reaction in the solid state, the observed ssHDX rate can be affected by the rate and extent of sorption [30]. To determine the effect of D_2O sorption on ssHDX kinetics, moisture uptake was measured using TGA to simulate D_2O uptake at 43 % humidity. Moisture sorption at 43 % RH was complete within 1 h for both Mb-A and Mb-B formulations, irrespective of the type of freezing (Appendix, Fig. A11). Since ssHDX continues over

several hours, it is unlikely that sorption kinetics have a significant effect on ssHDX kinetics beyond 1 h.

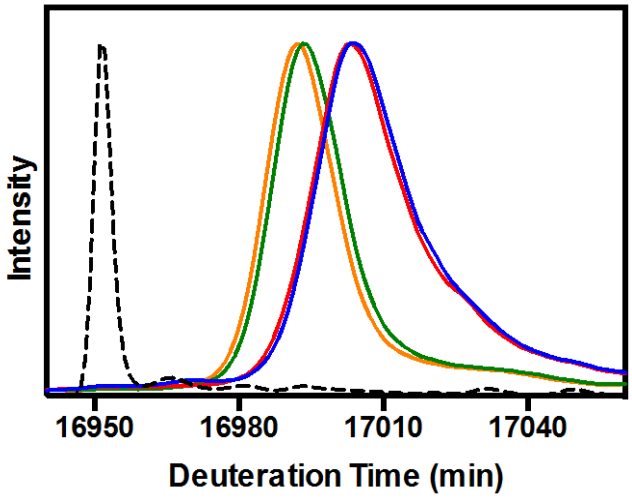


Figure 4.5. Deconvoluted mass spectra for excipient-free (Mb-A) and sucrose-containing (Mb-B) myoglobin formulations after 48 h of solid-state hydrogen-deuterium exchange at 5 °C, 43 % RH. Mass spectra were obtained for Mb-A and Mb-B lyophilized with controlled nucleation (blue and green curves respectively) and without controlled nucleation (red and orange curves respectively). The dashed curve represents the deconvoluted spectrum for undeuterated protein.

ssHDX was relatively rapid in both Mb-A and Mb-B formulations lyophilized with or without controlled nucleation up to ~ 24 h and slowed at ~ 84 h for all samples (Fig. 4.6 A, B). At each time point, sucrose-containing formulations showed significantly lower deuterium uptake than the excipient-free formulations, suggesting greater backbone protection in the solid state. Deuterium uptake in the Mb-A formulation was not significantly different ($p > 0.05$) when the protein was frozen with or without controlled nucleation (Fig. 4.6A); similar results were observed for Mb-B with or without controlled nucleation (Fig. 4.6B). The initial moisture content after lyophilization (without incubation over water or D_2O) of the formulations as measured by TGA were as follows: 2.9 % (Mb-A, controlled nucleation), 4.0 % (Mb-A, uncontrolled nucleation), 3.4 % (Mb-B, controlled nucleation) and 3.1 % (Mb-B, uncontrolled nucleation). Using TGA to simulate D_2O sorption and measure

moisture sorption at 43 % RH, differences in moisture sorption were observed between Mb-A and Mb-B formulations, but not between controlled and uncontrolled nucleation. Mb-A lyophilized with and without controlled nucleation sorbed ~ 0.12 and 0.13 g water/g dry solid respectively (~ 12.3 % and 12.7 % moisture respectively), and Mb-B lyophilized with and without controlled nucleation sorbed ~ 0.09 g water/g dry solid respectively (~ 9.4 % moisture for both; Appendix Fig. A11) within 1 h of incubation. Hence the increased deuterium uptake for Mb-A may be related to the extent of moisture sorption, at least until 1 h.

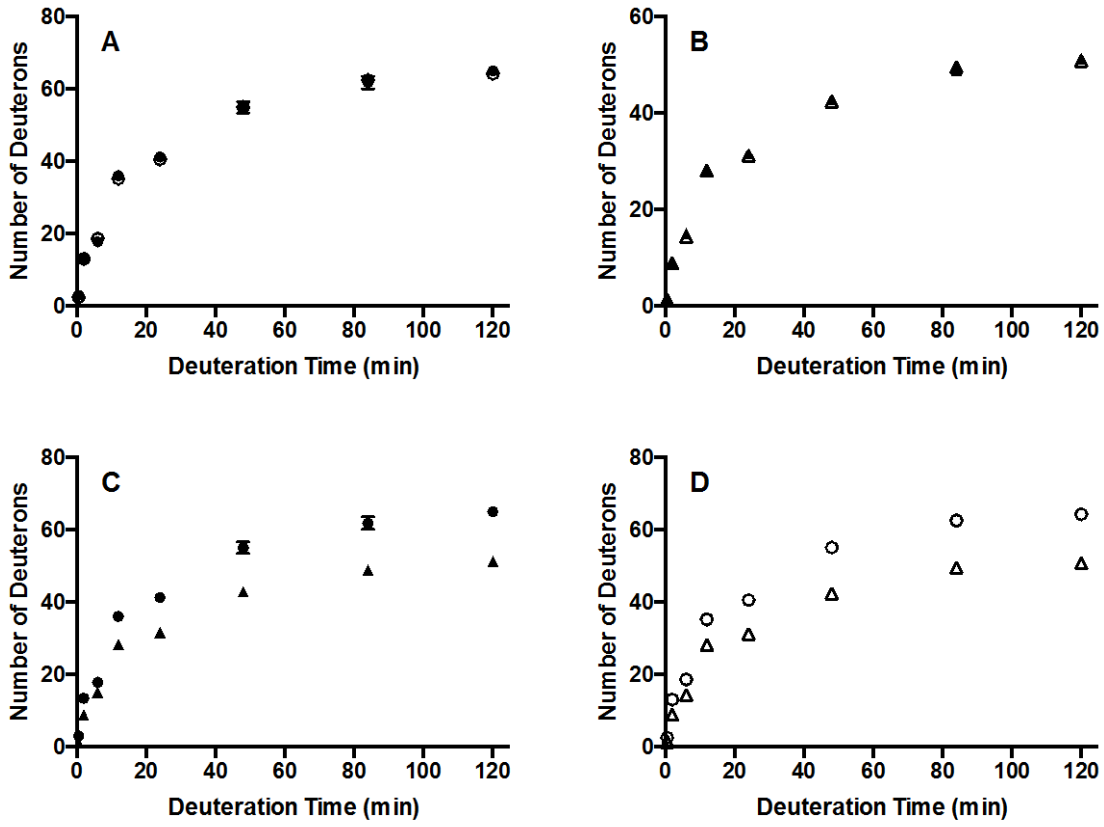


Figure 4.6. (A) Deuterium uptake kinetics for excipient-free myoglobin (Mb-A) lyophilized with controlled nucleation (closed circles) and without controlled nucleation (open circles). (B) Deuterium uptake kinetics for myoglobin-sucrose formulation (Mb-B) lyophilized with controlled nucleation (closed triangles) and without controlled nucleation (open triangles). (C) Deuterium

uptake kinetics for Mb-A (circles) and Mb-B (triangles) lyophilized with controlled nucleation. (D) Deuterium uptake kinetics for Mb-A (circles) and Mb-B (triangles) lyophilized without controlled nucleation. Each point represents the average number of deuterons (± SD) incorporated after incubation at 43 % RH, 5 °C for 0.5, 2, 6, 12, 24, 48, 84 and 120 h.

Comparison of peak widths at similar deuterium uptake levels can provide information on the conformational and/or spatial heterogeneity of protein in different formulations. The sucrose formulation lyophilized with or without controlled nucleation showed significantly narrower peak widths ($p < 0.05$) than the excipient-free formulation at ~ 28 % deuteration (Fig. 4.5, Table 4.2). Peak widths for Mb-B were about 26 % smaller than those for Mb-A formulations, irrespective of the type of freezing. No significant differences in peak widths were observed between controlled and uncontrolled nucleation within each formulation.

Table 4.2. ssHDX-MS peak widths for myoglobin formulations in the absence (Mb-A) and presence (Mb-B) of sucrose lyophilized with or without controlled nucleation. Peak widths at ~28 % deuteration were calculated from the deconvoluted mass spectra for each formulation at 20 % peak height.

Formulation	Peak Width (Da)	
	Controlled Nucleation	Uncontrolled Nucleation
Mb-A [a]	38.0 ± 1.0	38.3 ± 0.6
Mb-B [a]	28.0 ± 0.0	27.7 ± 0.6

[a] Mb-A, myoglobin lyophilized without excipients; Mb-B, myoglobin lyophilized with sucrose

4.4.6 Effects of Formulation and Process on Protein Side-Chain by ssPL-MS

Up to 2 pLeu labels were detected for all formulations lyophilized with pLeu and irradiated with UV light (Fig. 4.7A). Mb-A lyophilized with and without controlled nucleation showed similar pLeu incorporation (Fig. 4.7B). About 7 (± 1) % and 6 (± 1) % of the protein population was labeled for

Mb-A with and without controlled nucleation respectively, as detected by LC-MS. Similar pLeu incorporation (7 ± 1 %) was observed for Mb-B without controlled nucleation. The greatest pLeu incorporation was detected for Mb-B with controlled nucleation (11 ± 1 %). Comparing Mb-A and Mb-B formulations lyophilized with controlled nucleation, the fraction of intact protein labeled increased significantly for Mb-B. There was no significant difference in the fraction of labeled protein in Mb-A and Mb-B formulations lyophilized without controlled nucleation.

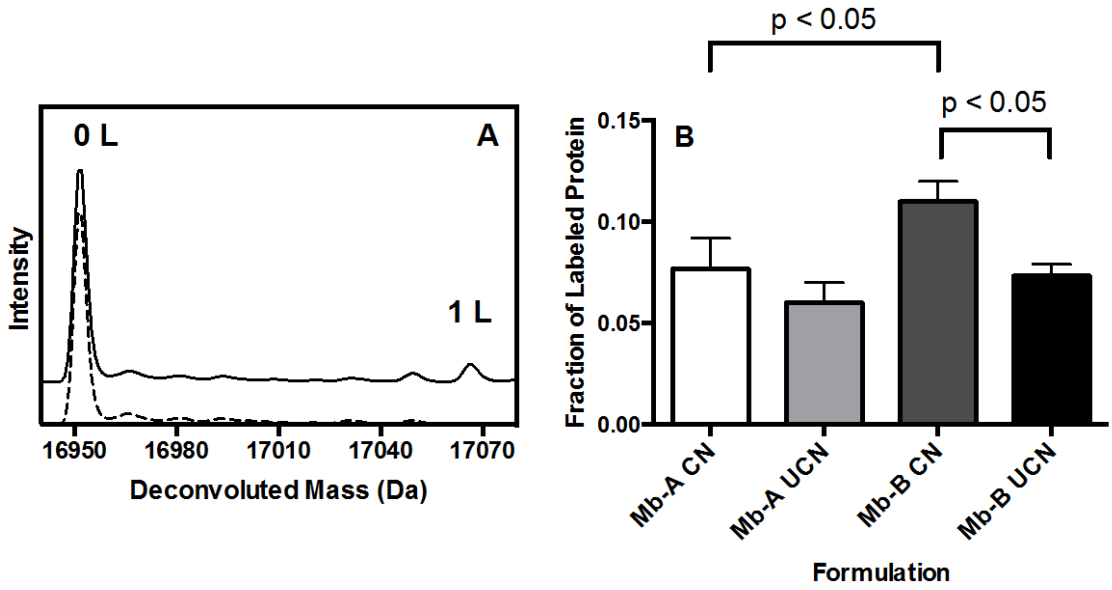


Figure 4.7. (A) Deconvoluted mass spectra for unlabeled myoglobin (dotted line) and myoglobin labeled with photo-leucine (solid line). The spectrum for labeled myoglobin has been offset vertically to show differences. Peaks corresponding to the mass of unlabeled (0 L, ~ 16951 Da) and protein with 1 label (1 L, ~ 17066 Da) were detected. A peak for myoglobin labeled with 2 labels (~ 17182 Da) was also detected, but not shown here since the intensity was low. Similar spectra with up to 2 labels were obtained for excipient-free (Mb-A) and sucrose-containing (Mb-B) myoglobin formulations lyophilized with and without controlled nucleation. (B) Fraction of protein labeled with photo-leucine for Mb-A and Mb-B formulations lyophilized with controlled nucleation (CN) or without controlled nucleation (UCN). The fraction was calculated using peak heights of

labeled protein on the extracted ion chromatogram (EIC) obtained by LC-MS. The bars show the mean (± SD) of three LC-MS injections. The following pairs showed significantly different means ($p < 0.05$, GraphPad Prism) using one-way ANOVA: Mb-A CN vs. Mb-B CN and Mb-B CN vs. Mb-B UCN.

At the peptide level, complete sequence coverage was obtained after tryptic digestion and LC-MS analysis. Since photolytic labeling is associated with some degree of variability due to promiscuity of the carbene reaction, matrix heterogeneity and ionization efficiencies of labeled and unlabeled peptides, only labeled peptides that were detected in at least 2 of triplicate LC-MS injections were considered. Mb-A lyophilized with controlled nucleation and Mb-A and Mb-B lyophilized without controlled nucleation were labeled at peptide His^{103}-Lys^{118} (Fig. 4.8A). Mb-B lyophilized with controlled nucleation was labeled at peptides Leu^{32}-Lys^{42} and His^{103}-Lys^{118} (Fig. 4.8B).

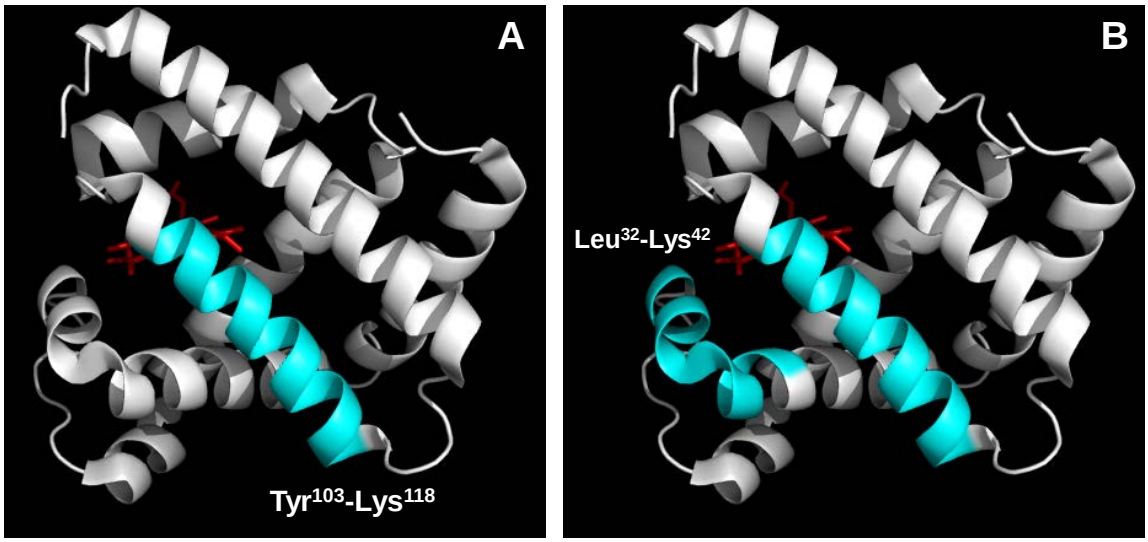


Figure 4.8. Peptide-level labeling with photo-leucine for (A) excipient-free (Mb-A) myoglobin lyophilized with controlled nucleation. (B) Sucrose-containing (Mb-B) myoglobin lyophilized with controlled nucleation. Labeled peptides Leu^{32}-Lys^{42} and Tyr^{103}-Lys^{118} are represented in blue and the heme group is shown in red. Mb-A and Mb-B lyophilized without controlled nucleation showed the same labeling pattern on Tyr^{103}-Lys^{118} as represented in panel (A). The crystal structure for

holomyoglobin was obtained from the Protein Data Bank (www.rcsb.org; PDB ID 1WLA) and PyMOL was used to generate labeled ribbon diagrams (PyMOL Molecular Graphics System, Version 1.4.1, Schrödinger, LLC).

4.5 DISCUSSION

During scale-up in an aseptic environment, vials are expected to supercool to a greater degree than in a laboratory environment where there are more particulates. This, combined with vial position effects and the stochastic nature of ice nucleation, produces heterogeneous nucleation temperatures and ice crystal sizes in the sterile environment, resulting in prolonged drying cycles. Controlled nucleation can help improve inter-vial and inter-batch homogeneity and reduce drying times. Lyophilization-induced structural changes may affect protein stability when a process is transferred from laboratory-scale to production, however, and these changes may not be detected using conventional, global methods such as ssFTIR and differential scanning calorimetry. Here, we report the use of high-resolution ssHDX-MS and ssPL-MS to detect conformational changes in lyophilized formulations with process and formulation differences.

ssHDX-MS results indicate that deuterium incorporation was affected significantly by formulation, but not the freezing step (Figs. 4.5, 4.6). Deuterium incorporation in the excipient-free formulation (Mb-A) was greater than in the formulation containing sucrose (Mb-B), consistent with greater structural perturbation and/or a decrease in matrix interactions in Mb-A (Fig. 4.5). Deuterium incorporation in both the excipient-free formulation and the sucrose-containing formulation was relatively unaffected by the freezing step (Fig. 4.5, orange vs. green curves, blue vs. red curves). That the type of freezing (controlled vs. uncontrolled nucleation) does not affect deuterium incorporation for both Mb-A and Mb-B suggests that the process does not significantly alter protein structure in this study. These findings are supported by ssHDX-MS kinetics (Fig. 4.6) and are consistent with trends in FTIR band position and band intensity (Fig. 4.3).

In analyzing deuterium uptake kinetics, the peak width provides a measure of protein structural heterogeneity resulting from the distribution of deuterated populations [34, 35]. Small increases in mass due to deuteration may not be detected if the peaks are not sufficiently resolved. This can cause peak broadening, wherein multiple deuterated populations comprise a wider peak. Peak widths for the Mb-A formulation were significantly greater than for Mb-B for both processes, consistent with greater structural and/or spatial heterogeneity in Mb-A than in Mb-B (Table 4.2). Within Mb-A or Mb-B formulations, peak widths for samples from the two processes were not significantly different (Table 4.2), suggesting similar structural and/or spatial heterogeneity with controlled and uncontrolled nucleation.

Like ssHDX-MS, ssPL-MS provides qualitative and quantitative information on proteins in the solid state. Since the mass of a pLeu label (~ 115 Da) is much greater than that of a deuterium label, mass spectrometric peaks for unlabeled and singly labeled protein can be resolved easily and peak broadening is not observed. Assuming uniform pLeu distribution in the matrix, the fraction of protein labeled by pLeu quantifies the fraction of protein with side-chain exposure to pLeu. The fraction of labeled protein and map of the labeled regions (Figs. 4.7B, 4.8) provide information on pLeu exposure at the side-chain level. In the presence of sucrose in the lyophilized matrix, it is expected that pLeu labeling efficiency will be somewhat diluted, resulting in decreased labeling compared to an excipient-free formulation. However, pLeu labeling was similar (uncontrolled nucleation) or greater (controlled nucleation) in the Mb-B formulation compared to Mb-A (Fig. 4.7). Within Mb-A formulations, the fraction of labeled protein was similar (Fig. 4.7) and within Mb-B formulations, the fraction of labeled protein was greater for controlled nucleation than for uncontrolled nucleation, consistent with greater interactions with the matrix. This suggests that while uncontrolled nucleation results in equivalent side-chain matrix exposure in the presence or absence of sucrose, controlled nucleation affects side-chain exposure in the presence of sucrose.

Intact- and peptide-level pLeu labeling data indicate that side-chain exposure to pLeu is different for Mb-B lyophilized with controlled nucleation than for the other formulations, suggesting more favorable contacts between the protein and pLeu (Figs. 4.7, 4.8). The reasons for the increased fraction of labeled protein in Mb-B with controlled nucleation are not clear, but may be related to the distribution of pLeu and protein in the formulation and/or changes in protein conformation. Inhomogeneity in the freeze-concentrated liquid after ice crystallization has been reported previously [36-38]. Efficiency of UV light penetration may also affect protein labeling. Smaller fill volumes result in lower cake height and better labeling efficiency (Appendix Fig. A12). Hence, ssPL-MS may not be representative of protein structure across the entire cake. Moreover, the uniform and large pore structure for Mb-B lyophilized with controlled nucleation may permit more efficient irradiation and pLeu labeling. Labeling of sucrose by pLeu was not detected by LC-MS, but may also affect protein labeling. Previous ssPL-MS results in our lab showed greater protein labeling in the presence of sucrose compared to guanidine hydrochloride [31]. It is difficult to distinguish the effect of excipients from the effect of protein conformation on the nature of solid-state labeling observed, and stability studies are needed to correlate the fraction of labeled protein with structure retention and interpret the side-chain labeling results.

Together, the results show that controlled nucleation did not significantly affect protein conformation in this study as determined by ssFTIR and ssHDX-MS (Figs. 4.5, 4.6, 4.7 and 4.8) and may offer the advantage of reduced drying time. A similar absence of structural changes has been reported for IgG lyophilized with and without controlled nucleation [26], although solution-state analytical methods were used. In this study, formulation effects were more dominant than process effects. Additional studies on the effects of controlled nucleation on protein structure are needed in order to extend these results to other proteins, using a variety of protein-excipient systems.

There were several unexpected observations in this study that merit further investigation. In preliminary controlled nucleation experiments using a fill volume of 200 μL and nitrogen as the gas

for pressurization, nucleation did not take place at -5 °C. Instead, nucleation proceeded in an uncontrolled manner. Controlled nucleation at -5 °C did take place when a fill volume of 500 μL was used in combination with argon as the pressurization gas. Further investigation is needed in order to understand the role of vial size, relative fill volume, and pressurization gas on the robustness of the nucleation process using rapid depressurization. Controlled nucleation by rapid depressurization is highly directional, always proceeding from the top of the fill volume downward. For very small fill volumes, the dynamics of this top-down process is quickly interrupted. It would be useful to study the effect of relative fill volume on protein structure perturbation using controlled nucleation by rapid depressurization. Generally speaking, the vials containing thermocouples nucleate before the unmonitored vials. That did not happen when monitoring uncontrolled nucleation in this study. In fact, the vials containing thermocouples nucleated last. This could have been a random occurrence, but it may be useful to further examine the influence of thermocouples on nucleation when using very small fill volumes.

Previous ssHDX-MS studies in our lab have shown a correlation between deuterium incorporation in freshly lyophilized samples and aggregation during storage over a year, with greater stability for formulations showing lower deuterium incorporation [29]. It is reasonable to expect a similar correlation for process-induced differences in ssHDX, though extended storage stability studies were not conducted here. Based on this previous report and the ssHDX-MS results for intact Mb presented here (Fig. 4.5), stability would be expected to decrease in the order: (Mb-B without controlled nucleation) = (Mb-B with controlled nucleation) > (Mb-A without controlled nucleation) = (Mb-A with controlled nucleation). A similar trend in structure retention was observed by ssFTIR, although the data are qualitative. To our knowledge, the relationship between ssPL-MS results and storage stability has not yet been explored.

4.6 CONCLUSIONS

Two formulations of Mb (with or without sucrose) were lyophilized according to the same lyophilization cycle with or without controlled nucleation and the effects on Mb conformation in the lyophilized solids were assessed using ssFTIR, ssHDX-MS and ssPL-MS. Formulation effects were dominant, with formulations containing sucrose showing better retention of structure by all measures than formulations without sucrose. Samples lyophilized with controlled nucleation did not differ from those lyophilized without controlled nucleation by most measures of structure. ssPL-MS showed greater pLeu incorporation and the involvement of more regions of the Mb molecule in Mb lyophilized with controlled nucleation in the presence of sucrose than for other conditions. The data support the use of ssHDX-MS and ssPL-MS to study formulation and process-induced conformational changes in lyophilized proteins.

4.7 REFERENCES

1. Bhatnagar, B. S.; Bogner, R. H.; Pikal, M. J. Protein stability during freezing: separation of stresses and mechanisms of protein stabilization. *Pharm Dev Technol* **2007,** *12*, (5), 505-523.
2. Anchordoquy, T. J.; Carpenter, J. F. Polymers protect lactate dehydrogenase during freeze-drying by inhibiting dissociation in the frozen state. *Arch Biochem Biophys* **1996,** *332*, (2), 231-238.
3. Schwegman, J. J.; Carpenter, J. F.; Nail, S. L. Evidence of partial unfolding of proteins at the ice/freeze-concentrate interface by infrared microscopy. *J Pharm Sci* **2009,** *98*, (9), 3239-3246.
4. Luthra, S.; Obert, J. P.; Kalonia, D. S.; Pikal, M. J. Investigation of drying stresses on proteins during lyophilization: differentiation between primary and secondary-drying stresses on lactate dehydrogenase using a humidity controlled mini freeze-dryer. *J Pharm Sci* **2007,** *96*, (1), 61-70.
5. Prestrelski, S. J.; Arakawa, T.; Carpenter, J. F. Separation of freezing- and drying-induced denaturation of lyophilized proteins using stress-specific stabilization. II. Structural studies using infrared spectroscopy. *Arch Biochem Biophys* **1993,** *303*, (2), 465-473.
6. Chang, B. S.; Kendrick, B. S.; Carpenter, J. F. Surface-induced denaturation of proteins during freezing and its inhibition by surfactants. *J Pharm Sci* **1996,** *85*, (12), 1325-1330.
7. Pikal-Cleland, K. A.; Carpenter, J. F. Lyophilization-induced protein denaturation in phosphate buffer systems: monomeric and tetrameric beta-galactosidase. *J Pharm Sci* **2001,** *90*, (9), 1255-1268.
8. Izutsu, K.; Kojima, S. Excipient crystallinity and its protein-structure-stabilizing effect during freeze-drying. *J Pharm Pharmacol* **2002,** *54*, (8), 1033-1039.
9. Cao, W.; Xie, Y.; Krishnan, S.; Lin, H.; Ricci, M. Influence of process conditions on the crystallization and transition of metastable mannitol forms in protein formulations during lyophilization. *Pharm Res* **2013,** *30*, (1), 131-139.

10. Kauppinen, A.; Toiviainen, M.; Aaltonen, J.; Korhonen, O.; Jarvinen, K.; Juuti, M.; Pellinen, R.; Ketolainen, J. Microscale freeze-drying with Raman spectroscopy as a tool for process development. *Anal Chem* **2013,** *85*, (4), 2109-2116.
11. Kim, A. I.; Akers, M. J.; Nail, S. L. The physical state of mannitol after freeze-drying: effects of mannitol concentration, freezing rate, and a noncrystallizing cosolute. *J Pharm Sci* **1998,** *87*, (8), 931-935.
12. Breen, E. D.; Curley, J. G.; Overcashier, D. E.; Hsu, C. C.; Shire, S. J. Effect of moisture on the stability of a lyophilized humanized monoclonal antibody formulation. *Pharm Res* **2001,** *18*, (9), 1345-1353.
13. Pikal, M. J.; Rigsbee, D.; Roy, M. L. Solid state stability of proteins III: calorimetric (DSC) and spectroscopic (FTIR) characterization of thermal denaturation in freeze dried human growth hormone (hGH). *J Pharm Sci* **2008,** *97*, (12), 5122-5131.
14. Jiang, S.; Nail, S. L. Effect of process conditions on recovery of protein activity after freezing and freeze-drying. *Eur J Pharm Biopharm* **1998,** *45*, (3), 249-257.
15. Schersch, K.; Betz, O.; Garidel, P.; Muehlau, S.; Bassarab, S.; Winter, G. Systematic investigation of the effect of lyophilizate collapse on pharmaceutically relevant proteins, part 2: stability during storage at elevated temperatures. *J Pharm Sci* **2012,** *101*, (7), 2288-2306.
16. Fitzpatrick, K.; Kendrick, B.; Santos, C.; Green, P.; Zhang, B.; Hunt, D.; Ronk, M.; Luo, Y. Freeze-Dry Mediated Formation of 5-(Hydroxylmethyl)furfural. *Developments in Biotechnology and Bioprocessing* **2013,** *1125*, 129-145.
17. Kasper, J. C.; Friess, W. The freezing step in lyophilization: physico-chemical fundamentals, freezing methods and consequences on process performance and quality attributes of biopharmaceuticals. *Eur J Pharm Biopharm* **2011,** *78*, (2), 248-263.
18. Hottot, A.; Vessot, S.; Andrieu, J. A Direct Characterization Method of the Ice Morphology. Relationship Between Mean Crystals Size and Primary Drying Times of Freeze-Drying Processes. *Drying Technology* **2004,** *22*, (8), 2009-2021.
19. Rambhatla, S.; Tchessalov, S.; Pikal, M. J. Heat and mass transfer scale-up issues during freeze-drying, III: control and characterization of dryer differences via operational qualification tests. *AAPS PharmSciTech* **2006,** *7*, (2), E39.
20. Eckhardt, B. M.; Oeswein, J. Q.; Bewley, T. A. Effect of freezing on aggregation of human growth hormone. *Pharm Res* **1991,** *8*, (11), 1360-1364.
21. Kreilgaard, L.; Jones, L. S.; Randolph, T. W.; Frokjaer, S.; Flink, J. M.; Manning, M. C.; Carpenter, J. F. Effect of Tween 20 on freeze-thawing- and agitation-induced aggregation of recombinant human factor XIII. *J Pharm Sci* **1998,** *87*, (12), 1597-1603.
22. Nakagawa, K.; Hottot, A.; Vessot, S.; Andrieu, J. Influence of controlled nucleation by ultrasounds on ice morphology of frozen formulations for pharmaceutical proteins freeze-drying. *Chemical Engineering and Processing: Process Intensification* **2006,** *45*, (9), 783-791.
23. Rambhatla, S.; Ramot, R.; Bhugra, C.; Pikal, M. J. Heat and mass transfer scale-up issues during freeze drying: II. Control and characterization of the degree of supercooling. *AAPS PharmSciTech* **2004,** *5*, (4), e58.
24. Passot, S.; Trelea, I. C.; Marin, M.; Galan, M.; Morris, G. J.; Fonseca, F. Effect of controlled ice nucleation on primary drying stage and protein recovery in vials cooled in a modified freeze-dryer. *J Biomech Eng* **2009,** *131*, (7), 074511.
25. Konstantinidis, A. K.; Kuu, W.; Otten, L.; Nail, S. L.; Sever, R. R. Controlled nucleation in freeze-drying: effects on pore size in the dried product layer, mass transfer resistance, and primary drying rate. *J Pharm Sci* **2011,** *100*, (8), 3453-3470.
26. Awotwe-Otoo, D.; Agarabi, C.; Read, E. K.; Lute, S.; Brorson, K. A.; Khan, M. A.; Shah, R. B. Impact of controlled ice nucleation on process performance and quality attributes of a lyophilized monoclonal antibody. *Int J Pharm* **2013,** *450*, (1-2), 70-78.
27. Parker, A.; Rigby-Singleton, S.; Perkins, M.; Bates, D.; Le Roux, D.; Roberts, C. J.; Madden-Smith, C.; Lewis, L.; Teagarden, D. L.; Johnson, R. E.; Ahmed, S. S. Determination of the influence of primary drying rates on the microscale structural attributes and

physicochemical properties of protein containing lyophilized products. *J Pharm Sci* **2010,** *99*, (11), 4616-4629.
28. Lewis, L. M.; Johnson, R. E.; Oldroyd, M. E.; Ahmed, S. S.; Joseph, L.; Saracovan, I.; Sinha, S. Characterizing the freeze-drying behavior of model protein formulations. *AAPS PharmSciTech* **2010,** *11*, (4), 1580-1590.
29. Moorthy, B. S.; Schultz, S. G.; Kim, S. G.; Topp, E. M. Predicting protein aggregation during storage in lyophilized solids using solid state amide hydrogen/deuterium exchange with mass spectrometric analysis (ssHDX-MS). *Mol Pharm* **2014,** *11*, (6), 1869-1879.
30. Sophocleous, A. M.; Zhang, J.; Topp, E. M. Localized hydration in lyophilized myoglobin by hydrogen-deuterium exchange mass spectrometry. 1. Exchange mapping. *Mol Pharm* **2012,** *9*, (4), 718-726.
31. Iyer, L. K.; Moorthy, B. S.; Topp, E. M. Photolytic labeling to probe molecular interactions in lyophilized powders. *Mol Pharm* **2013,** *10*, (12), 4629-4639.
32. Roy, M. L.; Pikal, M. J. Process control in freeze drying: determination of the end point of sublimation drying by an electronic moisture sensor. *J Parenter Sci Technol* **1989,** *43*, (2), 60-66.
33. Patel, S. M.; Doen, T.; Pikal, M. J. Determination of end point of primary drying in freeze-drying process control. *AAPS PharmSciTech* **2010,** *11*, (1), 73-84.
34. Sophocleous, A. M.; Topp, E. M. Localized hydration in lyophilized myoglobin by hydrogen-deuterium exchange mass spectrometry. 2. Exchange kinetics. *Mol Pharm* **2012,** *9*, (4), 727-733.
35. Weis, D. D.; Wales, T. E.; Engen, J. R.; Hotchko, M.; Ten Eyck, L. F. Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *J Am Soc Mass Spectrom* **2006,** *17*, (11), 1498-1509.
36. Dong, J.; Hubel, A.; Bischof, J. C.; Aksan, A. Freezing-induced phase separation and spatial microheterogeneity in protein solutions. *J Phys Chem B* **2009,** *113*, (30), 10081-10087.
37. Heller, M. C.; Carpenter, J. F.; Randolph, T. W. Manipulation of lyophilization-induced phase separation: implications for pharmaceutical proteins. *Biotechnol Prog* **1997,** *13*, (5), 590-596.
38. Izutsu, K.; Kojima, S. Freeze-concentration separates proteins and polymer excipients into different amorphous phases. *Pharm Res* **2000,** *17*, (10), 1316-1322.

CHAPTER 5. CONCLUSIONS

This dissertation has described novel, high-resolution techniques to probe protein structure and environment in the solid state. These techniques are orthogonal to conventional analytical methods such as FTIR and Raman spectroscopy and provide peptide- to amino acid-level information about changes in protein structure and microenvironment in the solid state.

The research presented in Chapter 2 shows the potential of solid-state photolytic labeling- mass spectrometry (ssPL-MS) to study protein structure with high resolution. Although labeling approaches have been described in solution, these have not been applied to the solid state previously, to the authors' knowledge. Solid-state labeling overcomes the low resolution of methods such as Fourier-transform infrared (FTIR) spectroscopy and allows direct visualization of side-chain matrix accessibility. It does not depend on pH and can label the entire protein surface.

ssPL-MS can be used to study storage stability in lyophilized formulations by measuring the change in side-chain accessibility. For example, vials containing lyophilized protein formulation with pLeu in the matrix can be stored at high temperature and/or humidity for accelerated stability studies. Samples can be withdrawn at definite intervals, irradiated and analyzed by LC-MS at the intact protein- and peptide level. Side-chain accessibility of peptides can be quantified using relative peak heights of unlabeled and labeled peptides obtained by LC-MS. This metric can be correlated with % monomeric protein observed by size exclusion chromatography. ssPL-MS can also be used to study the mechanism of solid-state aggregation by labeling protein formulations over the time course of aggregation and studying changes in side-chain matrix accessibility.

The crosslinking approach described in Chapter 3 advances the labeling technique by allowing the protein structure as well as environment to be probed. Thus far, the effect of excipients on lyophilized protein structure has only been probed indirectly, based on FTIR band areas [1, 2]. However, it is not clearly understood how the mechanism of protein stabilization changes when a protein in solution is lyophilized. To the authors' knowledge, this is the first report of direct visualization of protein-matrix interactions in the solid state. The microenvironment around the protein could be examined with 3.9 Å resolution, comparable to X-ray crystallography resolution but without the need for large amounts of protein, isotopic labeling or crystallinity requirements.

The effects of the excipients' physical form on protein local structure were also observed using photolytic crosslinking. It is generally accepted that a stabilizing excipient must be in an amorphous state with the protein to allow better mixing and physical contact. Hence, excipients that remain amorphous during lyophilization, such as sucrose and trehalose, are expected to stabilize the protein to a greater degree than crystallizing excipients such as sodium chloride and mannitol. As described in Chapter 2, the microenvironment around the protein changed significantly when it was lyophilized, as observed by changes in the crosslinked adducts observed by LC-MS. Although guanidine hydrochloride (Gdn HCl) was in intimate contact with the protein in solution and caused protein unfolding, the excipient appeared to crystallize during lyophilization and produced different crosslinking patterns compared to solution. The raffinose formulation was expected to be more stable since carbohydrates such as sucrose and trehalose tend to remain in an amorphous phase with the protein during lyophilization [3]. Although crystallinity was not detected by X-ray diffraction, the crosslinking patterns indicated possible micro-phase separation in the solid state. Such phase separating, albeit non-crystallizing, excipients may not provide adequate stabilization during lyophilization and/or storage.

This research can be applied to other amorphous protein systems spanning a range of secondary structure content, with commonly used disaccharide excipients such as sucrose and trehalose.

Concentrations of disaccharides in the formulations can also be varied to study their effect on crosslinking patterns. Crosslinking data obtained from different protein and excipient systems can help build a model to describe and predict protein stability in the solid state. Furthermore, this method has the potential to provide insight into the mechanisms of protein stabilization by excipients, both in solution and solid state. Thus protein crosslinking can aid rational design of formulations.

The Food and Drug Administration's (FDA) emphasis on Quality by Design recognizes the need to improve product quality and reduce the risk of failure. This requires better analytical and predictive tools to identify process- and product-related variables and ultimately control them, thereby creating a Design Space. The results described in Chapter 4 showed that controlled ice nucleation did not affect local protein conformation significantly and that process- and excipient-related effects on protein local structure in the solid state can be monitored using high-resolution solid-state hydrogen-deuterium exchange-mass spectrometry (ssHDX) and solid-state photolytic labeling-mass spectrometry (ssPL-MS). The results highlight the potential of these analytical methods as QbD tools to provide predictive measures of protein stability.

Gaps still exist in our knowledge of process effects on protein structure and function. While it is of interest to make the lyophilization process more efficient, the consequences of process-related stresses on protein stability must be evaluated during manufacture as well as storage. Proteins that are sensitive to lyophilization-induced structural changes such as lactate dehydrogenase and human growth hormone can be used as model proteins. ssHDX-MS, ssPL-MS and ssPC-MS can be used to study protein conformation changes as a function of lyophilization cycle parameters such as freezing and drying temperatures and times and chamber pressure. The effect of protein concentration, fill volume, type of excipient, vial shape and material must also be investigated. Storage stability studies at different temperatures and relative humidity can be performed with ssHDX and photolytic methods. The use of high-resolution metrics such as the number of

exchangeable amides, number of pLeu labels and number of protein-matrix adducts to describe product stability (using % monomer determined by size exclusion chromatography) must be evaluated. Thus, high-resolution methods for backbone and side-chain conformational change in the solid state have potential as tools for rational formulation design, storage stability and product quality evaluation.

5.1 REFERENCES

1. Allison, S. D.; Chang, B.; Randolph, T. W.; Carpenter, J. F. Hydrogen bonding between sugar and protein is responsible for inhibition of dehydration-induced protein unfolding. *Arch Biochem Biophys* **1999,** *365*, (2), 289-298.
2. Carpenter, J. F.; Crowe, J. H. An infrared spectroscopic study of the interactions of carbohydrates with dried proteins. *Biochemistry* **1989,** *28*, (9), 3916-3922.
3. Carpenter, J. F.; Chang, B. S.; Garzon-Rodriguez, W.; Randolph, T. W. Rational design of stable lyophilized protein formulations: theory and practice. *Pharm Biotechnol* **2002,** *13*, 109-133.

APPENDIX

APPENDIX

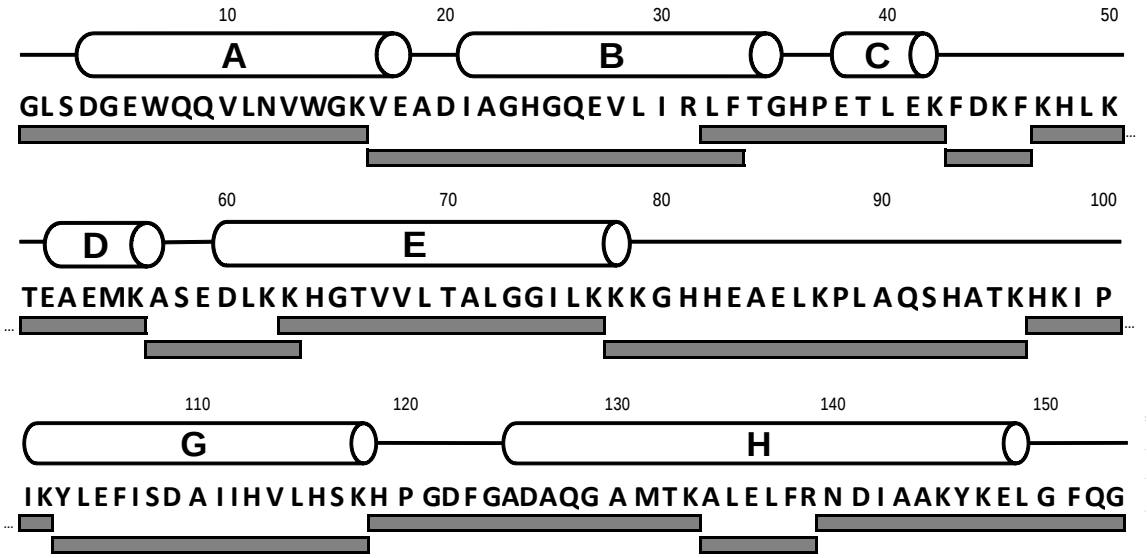


Figure A1. Digest map of native apoMb digested with a combination of trypsin and chymotrypsin. A total of 36 peptides were produced, of which the 13 shown by the shaded bars were selected to provide 100% sequence coverage.

Table A1. Theoretical and observed *b*- and *y*-ions from MS/MS analysis of native and labeled L32-K42 in apoMb labeled with pLeu in lyophilized solids.

I. Product ions with *z*=+1 produced by fragmenting native L32-K42 (*m/z* = 424.5609; *z*=+3)

	b-ions	**Theoretical m/z**[a]	**Observed m/z**[b]	**y-ions**	**Theoretical m/z**[a]	**Observed m/z**[b]
L	b_1	114.0919		y_{11}	1271.6636	
F	b_2	261.1604	261.1586	y_{10}	1158.5796	
T	b_3	362.2080	362.2050	y_9	1011.5111	1011.5070
G	b_4	419.2295	419.2223	y_8	910.4635	910.4603
H	b_5	556.2884	556.2834	y_7	853.442	853.4381
P	b_6	653.3412		y_6	716.3831	716.3800
E	b_7	782.3838	782.3758	y_5	619.3303	619.3267
T	b_8	883.4314	883.4253	y_4	490.2877	490.2852
L	b_9	996.5155		y_3	389.2401	389.2385
E	b_{10}	1125.5581		y_2	276.1560	276.1544
K	b_{11}	1253.6531		y_1	147.1134	147.1119

II. Product ions with *z*=+2 produced by fragmenting native L32-K42 (*m/z* = 424.5609; *z*=+3)

	***y*-ions[c]**	**Theoretical m/z[a]**	**Observed m/z[b]**
L	y_{11}	636.3357	636.3344
F	y_{10}	579.7937	579.7910
T	y_{9}	506.2595	506.2572
G	y_{8}	455.7357	455.7332
H	y_{7}	427.2249	427.2230
P	y_{6}	358.6955	358.6935
E	y_{5}	310.1691	
T	y_{4}	245.6478	
L	y_{3}	195.1240	
E	y_{2}	138.5819	
K	y_{1}	74.0606	

III. Product ions with *z*=+2 produced by fragmenting labeled L32-K42 (*m/z* = 462.9133; *z*=+3)

	y-ions	(A) Theoretical *m/z* (labeled)	(B) Observed *m/z* (labeled)	(C) Observed *m/z* (unlabeled)	Mass difference (u)[d] $M = M_{labeled} - M_{unlabeled}$
L	y_{11}	693.8674	693.8649		
F	y_{10}	637.3254	637.3184	579.789	115.0588
T	y_{9}	563.7912	563.7873	506.2524	115.0698
G	y_{8}	513.2673	513.2592	455.7365	115.0454
H	y_{7}	484.7566	484.7525	427.2245	115.0560
P	y_{6}	416.2271	416.2189	358.6996	115.0386
E	y_{5}	367.7008			
T	y_{4}	303.1795			
L	y_{3}	252.6556			
E	y_{2}	196.1136			
K	y_{1}	131.5923			

IV. Product ions with *z*=+1 produced by fragmenting labeled L32-K42 (*m/z* = 462.9133; *z*=+3)

	y-ions[c]	(A) Theoretical *m/z* (labeled)	(B) Observed *m/z* (labeled)	(C) Observed *m/z* (unlabeled)	Mass difference M = $M_{labeled}$ – $M_{unlabeled}$
L	y_{11}	1386.7269			
F	y_{10}	1273.6429			
T	y_9	1126.5744			
G	y_8	1025.5268			
H	y_7	968.5053			
P	y_6	831.4464	831.4462	716.3819	115.0643
E	y_5	734.3936	734.3989	619.3205	115.0784
T	y_4	605.3510	605.3535	490.289	115.0645
L	y_3	504.3034		389.2466	
E	y_2	391.2193		276.1541	
K	y_1	262.1767		147.1118	

[a] Calculated *m/z* values.

[b] *m/z* values obtained experimentally using mass spectrometry.

[c] No *b*-ions were detected by MS for *z* = +2.

[d] Mass difference M was calculated from *m/z* values in columns (B) and (C), using the formula M =(m/z)*n – nH, where n is the number of charges on the *y*-ion and H is the mass of a proton (H=1.01 u).

Table A2. Theoretical and observed *b*- and *y*-ions from MS/MS analysis of GCG (1-8)* dimer from formulation containing peptide lyophilized with L-leu. F* denotes p-benzoyl-L-phenylalanine (pBpA). Calculated m/z values are denoted as 'Theoretical m/z' while m/z values obtained experimentally using mass spectrometry are denoted as 'Observed m/z'.

I. Internal fragment (non-cross-linked) product ions with z=+1 produced by fragmenting GCG (1-8)* dimer (m/z = 646.2783; z=+3)

	***b*-ions**	**Theoretical *m/z*[b]**	**Observed *m/z*[c]**	***y*-ions**	**Theoretical *m/z*[b]**	**Observed *m/z*[c]**
H	b_1	138.0668	138.0656	y_8	968.4110	
S	b_2	225.0988	225.0973	y_7	831.3521	
Q	b_3	353.1574	353.1502	y_6	744.3201	
G	b_4	410.1789		y_5	616.2615	
T	b_5	511.2265	511.2234	y_4	559.2400	559.2367
F*[a]	b_6	762.3207		y_3	458.1923	458.1916
T	b_7	863.3684		y_2	207.0981	207.0963
S	b_8	950.4004		y_1	106.0505	106.0497

II. Cross-linked product ions with *z*=+2 produced by fragmenting GCG (1-8)* dimer (*m/z* = 646.2783; *z*=+3)

	***b*~α-ions[d]**	**Theoretical *m/z*[b]**	**Observed *m/z*[c]**	**α~*y*-ions**	**Theoretical *m/z*[b]**	**Observed *m/z*[c]**
H	b_1	553.2400		y_8	968.9173	968.9178
S	b_2	596.7560		y_7	890.8773	
Q	b_3	660.7853		y_6	847.3613	847.3526
G	b_4	689.2960	689.2939	y_5	783.3320	783.3290
T	b_5	739.8198	739.8130	y_4	754.8213	
F*	b_6	865.3669	865.3627	y_3	704.2975	
T	b_7	915.8908	915.8929	y_2	578.7504	
S	b_8	959.4068		y_1	528.2265	

[a] F* = p-benzoyl-L-phenylalanine (pBpA).

[b] Calculated *m/z* values.

[c] *m/z* values obtained experimentally using mass spectrometry.

[d] α = GCG (1-8)* monomer.

Table A3. List of SDA-labeled tryptic peptides selected for LC-MS/MS analysis.

(Gly[1]-Lys[16]) L1 [a]	(His[48]-Lys[56]) L1
(Leu[32]-Lys[45]) L1	(Lys[79]-Lys[96]) L1
(Leu[32]-Lys[47]) L1	(His[97]-Lys[118]) L1
(Phe[43]-Lys[56]) L4 [a]	(Tyr[146]-Gly[153]) L1
(Phe[46]-Lys[50]) L2 [a]	

[a] L1 denotes 1 SDA label on the peptide, L2, 2 SDA labels and L4, 4 SDA labels.

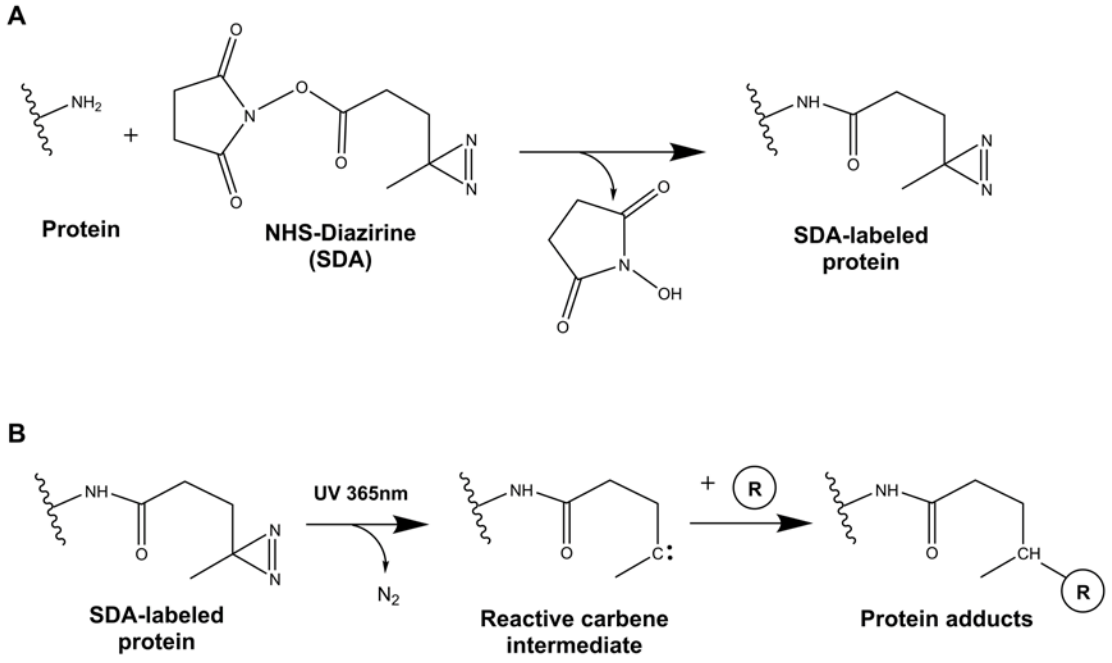


Figure A2. Mechanism of crosslinking using succinimidyl 4,4'- azipentanoate (NHS-diazirine; SDA). In the first step (A), the protein is incubated with SDA which results in derivatization of primary amine containing side-chains. In the second step (B), the SDA-labeled protein is exposed to UV light (365 nm) and forms a reactive carbene intermediate with the loss of N_2. The carbene forms covalent adducts with reactants (R) within the distance of the spacer arm (R = water, formulation additives or protein).

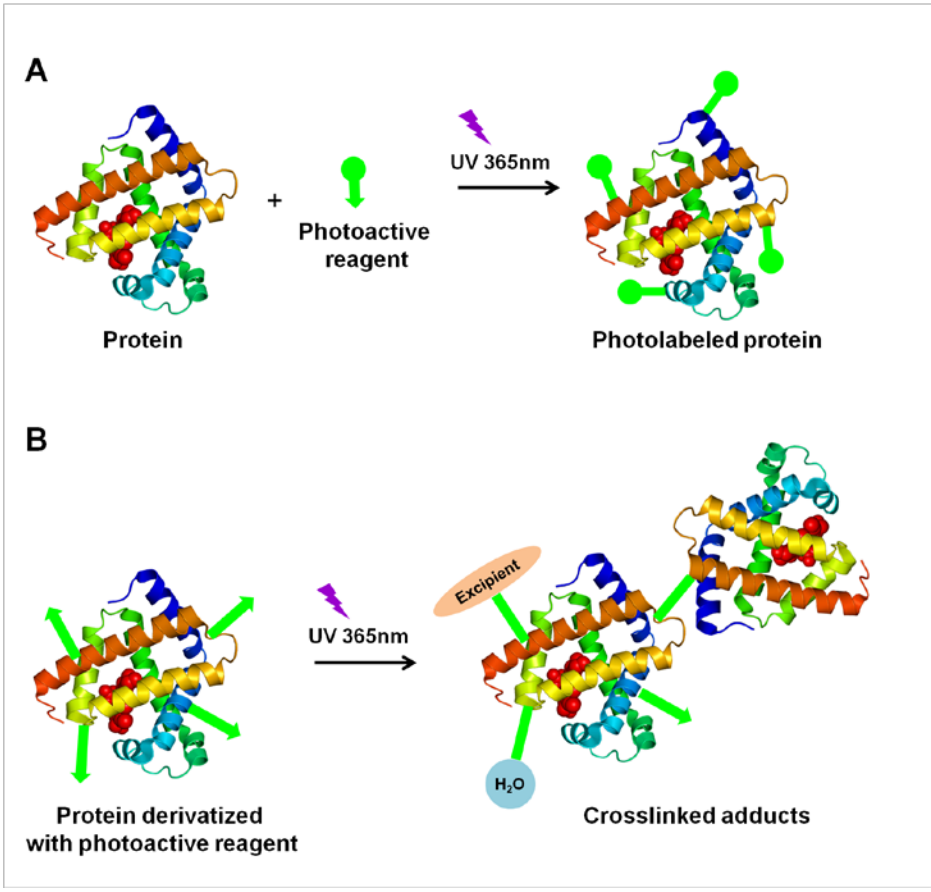


Figure A3. Comparison of (A) photolabeling and (B) photocrosslinking methods. In photolabeling, the protein is exposed to UV irradiation (365nm) in the presence of a photoactive reagent in the excipient matrix. In photocrosslinking, a protein derivatized with a bifunctional photoactive reagent is exposed to UV light (365 nm) in the presence of other matrix components (e.g. water, formulation additives or protein), which produces crosslinked molecules.

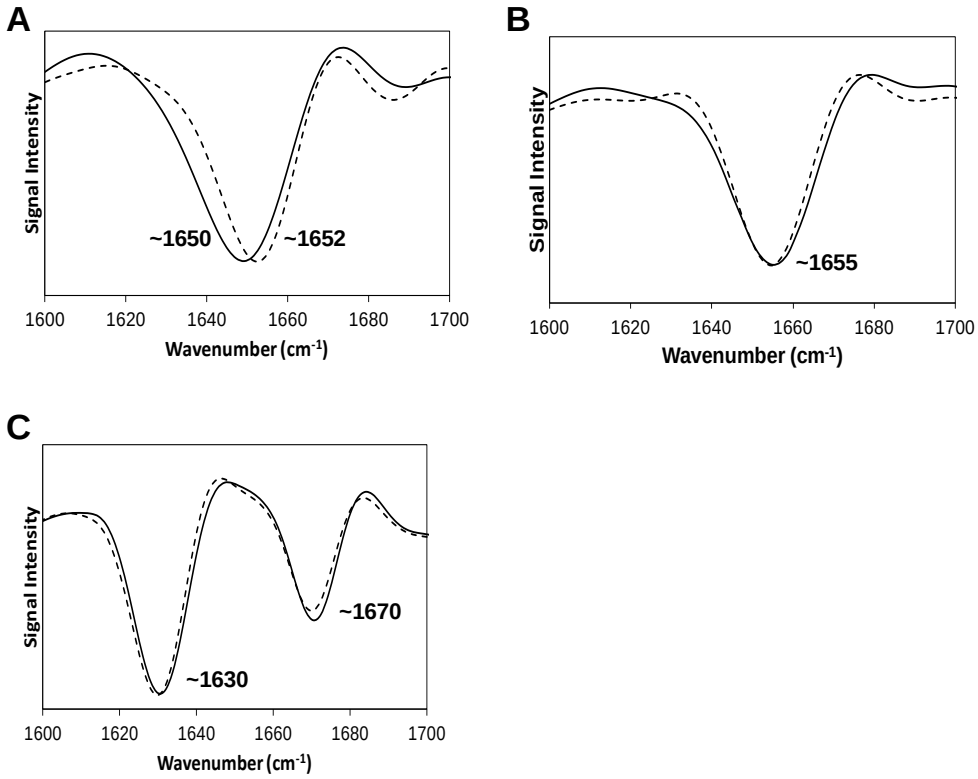
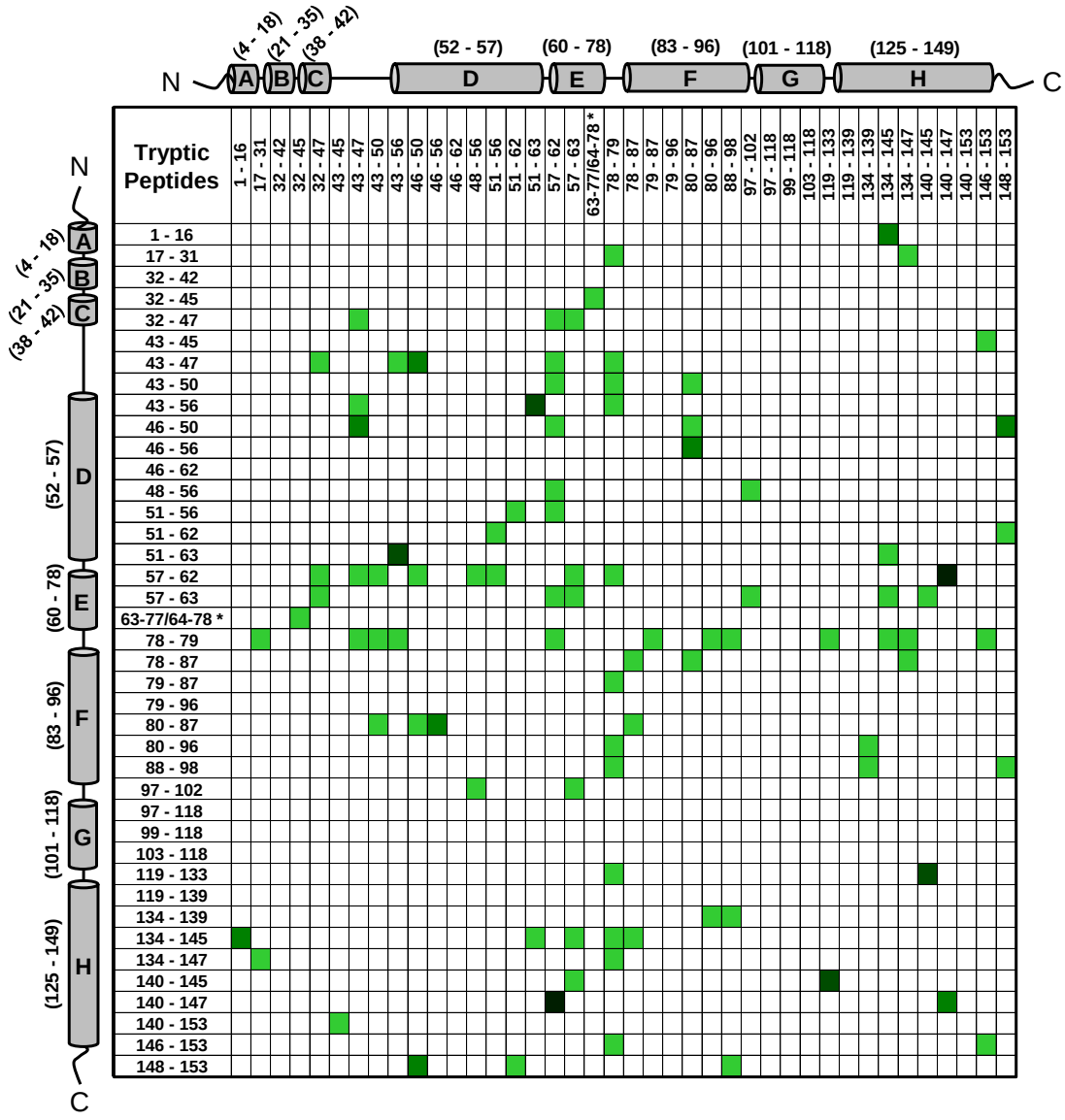
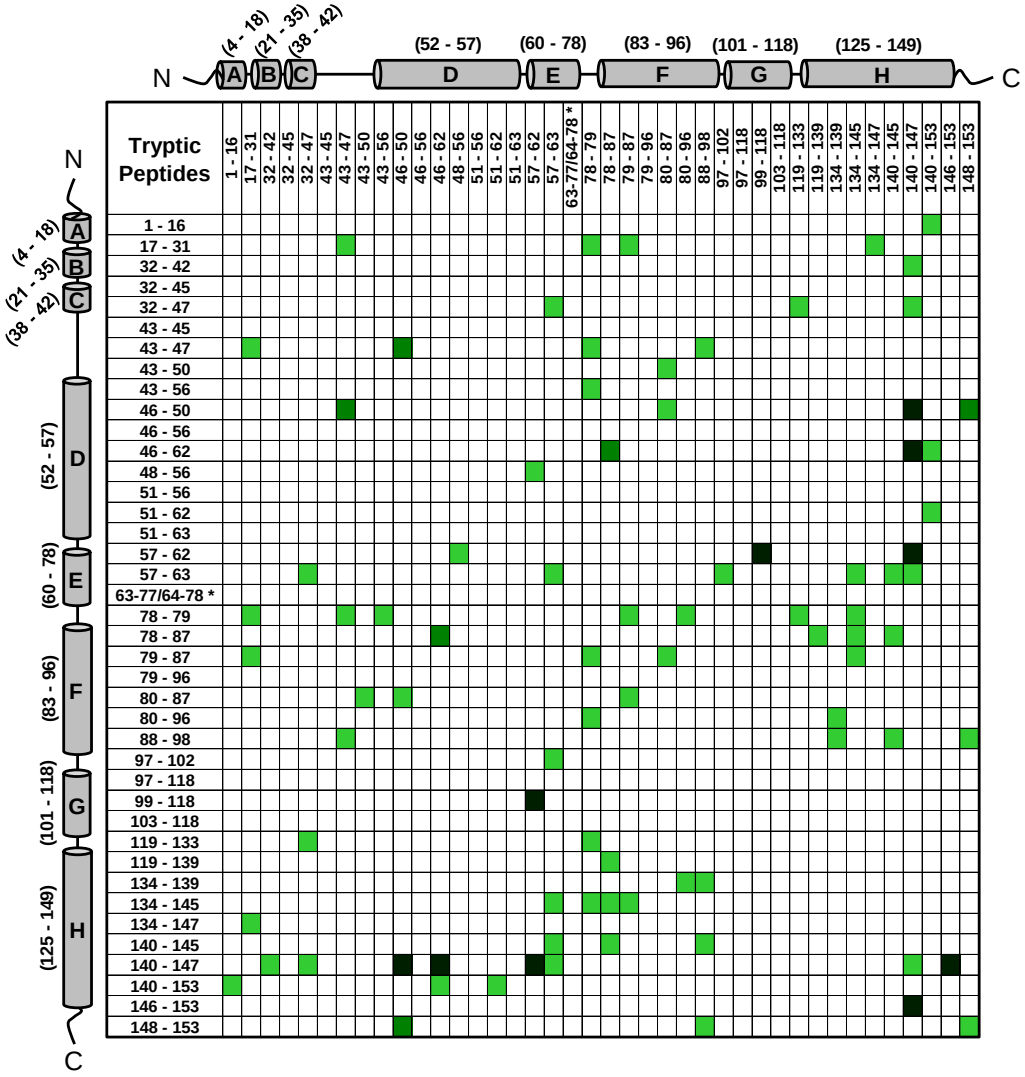


Figure A4. Second derivative amide I FTIR spectra of Mb-SDA (solid line) and unlabeled Mb (dashed line) Mb in formulations (A) Mb alone (control), (B) Mb with raffinose and (C) Mb with Gdn HCl. The band intensity (1650-1655 cm^{-1}) observed for A and B indicates the presence of α-helix in Mb, whereas the band intensity (~1630 and ~1670 cm^{-1}) for (C) is mainly from the beta sheet content.

Table A4. List of all possible peptide-water adducts that can be formed by crosslinking with SDA. A maximum of 4 SDA labels per peptide (and hence up to 4 H_2O and 4 raffinose molecules per peptide) were considered. The same list was considered for all possible peptide-raffinose adducts.

Peptide + 1SDA – $1N_2$ + $1H_2O$	Peptide + 2SDA – $1N_2$ + $1H_2O$	Peptide + 3SDA – $1N_2$ + $1H_2O$	Peptide + 4SDA – $1N_2$ + $1H_2O$
	Peptide + 2SDA – $2N_2$ + $1H_2O$	Peptide + 3SDA – $2N_2$ + $1H_2O$	Peptide + 4SDA – $2N_2$ + $1H_2O$
	Peptide + 1SDA – $2N_2$ + $2H_2O$	Peptide + 3SDA – $2N_2$ + $2H_2O$	Peptide + 4SDA – $2N_2$ + $2H_2O$
		Peptide + 3SDA – $3N_2$ + $1H_2O$	Peptide + 4SDA – $3N_2$ + $1H_2O$
		Peptide + 3SDA – $3N_2$ + $2H_2O$	Peptide + 4SDA – $3N_2$ + $2H_2O$
		Peptide + 3SDA – $3N_2$ + $3H_2O$	Peptide + 4SDA – $3N_2$ + $3H_2O$
			Peptide + 4SDA – $4N_2$ + $1H_2O$
			Peptide + 4SDA – $4N_2$ + $2H_2O$
			Peptide + 4SDA – $4N_2$ + $3H_2O$
			Peptide + 4SDA – $4N_2$ + $4H_2O$


(4 - 18)
(21 - 35)
(38 - 42)
(52 - 57)
(60 - 78)
(83 - 96)
(101 - 118)
(125 - 149)
N
A
B
C
D
E
F
G
H
C
Tryptic Peptides
1 - 16
17 - 31
32 - 42
32 - 45
32 - 47
43 - 45
43 - 47
43 - 50
43 - 56
46 - 50
46 - 56
46 - 62
48 - 56
51 - 56
51 - 62
51 - 63
57 - 62
57 - 63
63-77/64-78 *
78 - 79
78 - 87
79 - 87
79 - 96
80 - 87
80 - 96
88 - 98
97 - 102
97 - 118
99 - 118
103 - 118
119 - 133
119 - 139
134 - 139
134 - 145
134 - 147
140 - 145
140 - 147
140 - 153
146 - 153
148 - 153


N
C
(4 - 18)
(21 - 35)
(38 - 42)
(52 - 57)
(60 - 78)
(83 - 96)
(101 - 118)
(125 - 149)
A
B
C
D
E
F
G
H
Tryptic Peptides
1 - 16
17 - 31
32 - 42
32 - 45
32 - 47
43 - 45
43 - 47
43 - 50
43 - 56
46 - 50
46 - 56
46 - 62
48 - 56
51 - 56
51 - 62
51 - 63
57 - 62
57 - 63
63-77/64-78 *
78 - 79
78 - 87
79 - 87
79 - 96
80 - 87
80 - 96
88 - 98
97 - 102
97 - 118
99 - 118
103 - 118
119 - 133
119 - 139
134 - 139
134 - 145
134 - 147
140 - 145
140 - 147
140 - 153
146 - 153
148 - 153

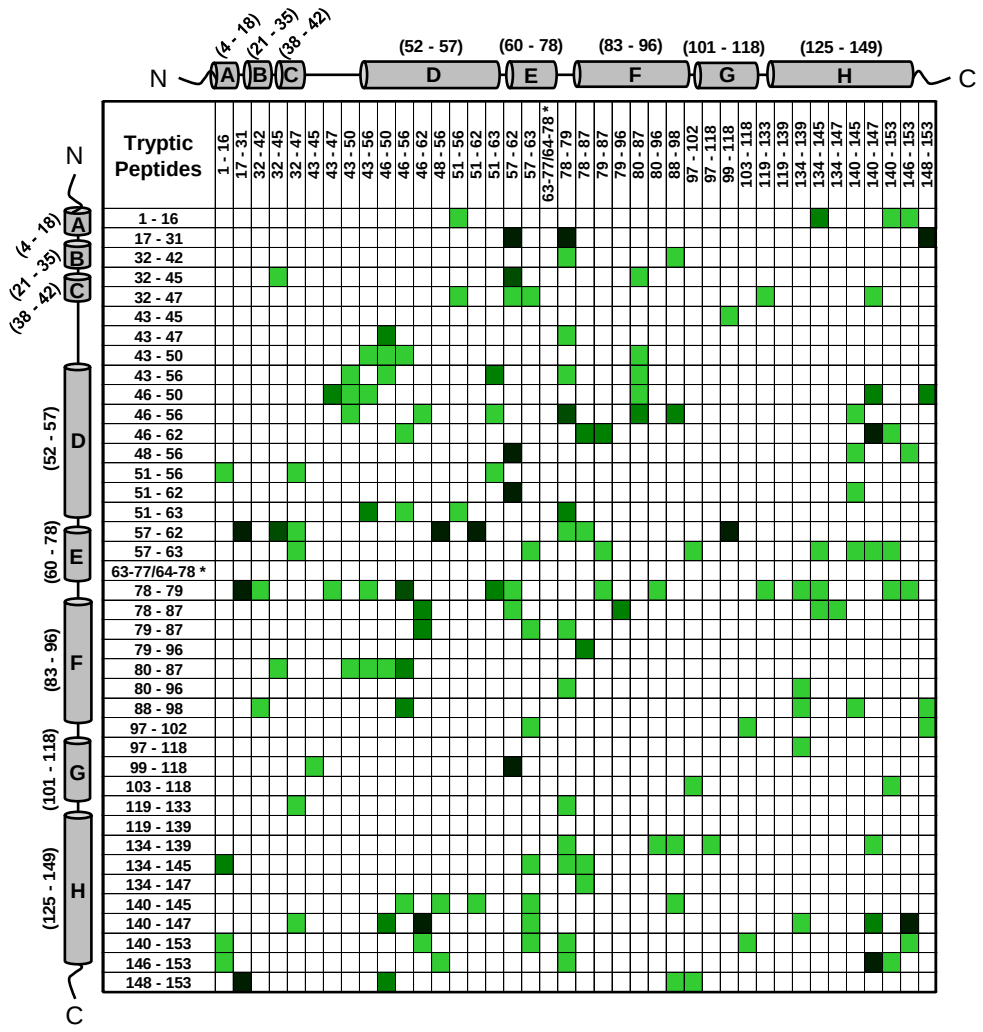


Figure A5. Peptide-peptide adducts formed in (A) Mb-SDA alone (control), (B) Mb-SDA with raffinose and (C) Mb-SDA with Gdn HCl lyophilized formulations. Peptide-peptide adducts formed by maximum one (■), two (■), three (■) and four (■) SDA molecules from a single sample injection are plotted. The α-helices from N-terminus to C-terminus in Mb are represented by cylinders labeled A to H respectively.

* The molecular mass of tryptic fragments 63-77 and 64-78 are identical and cannot be differentiated.

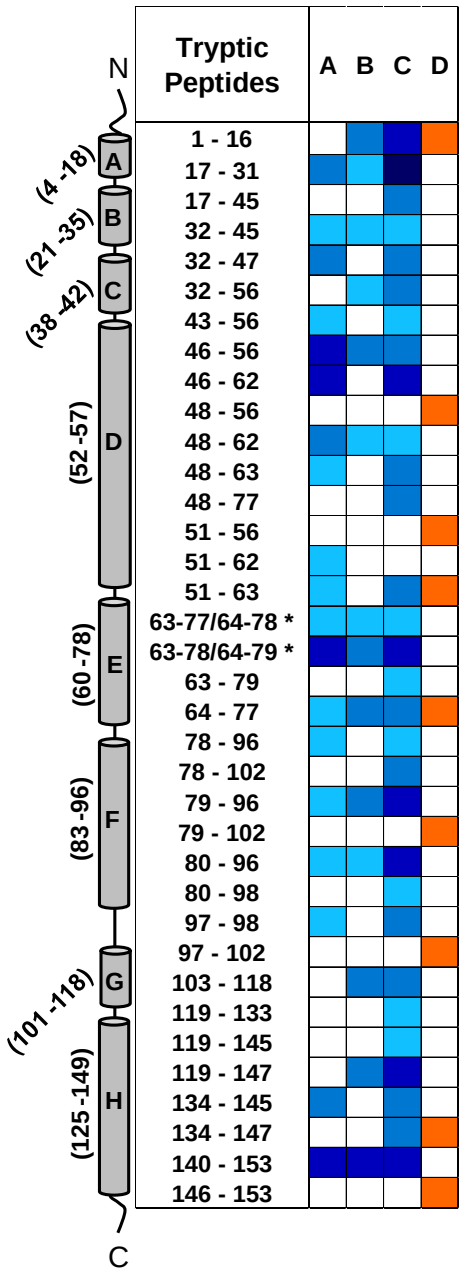
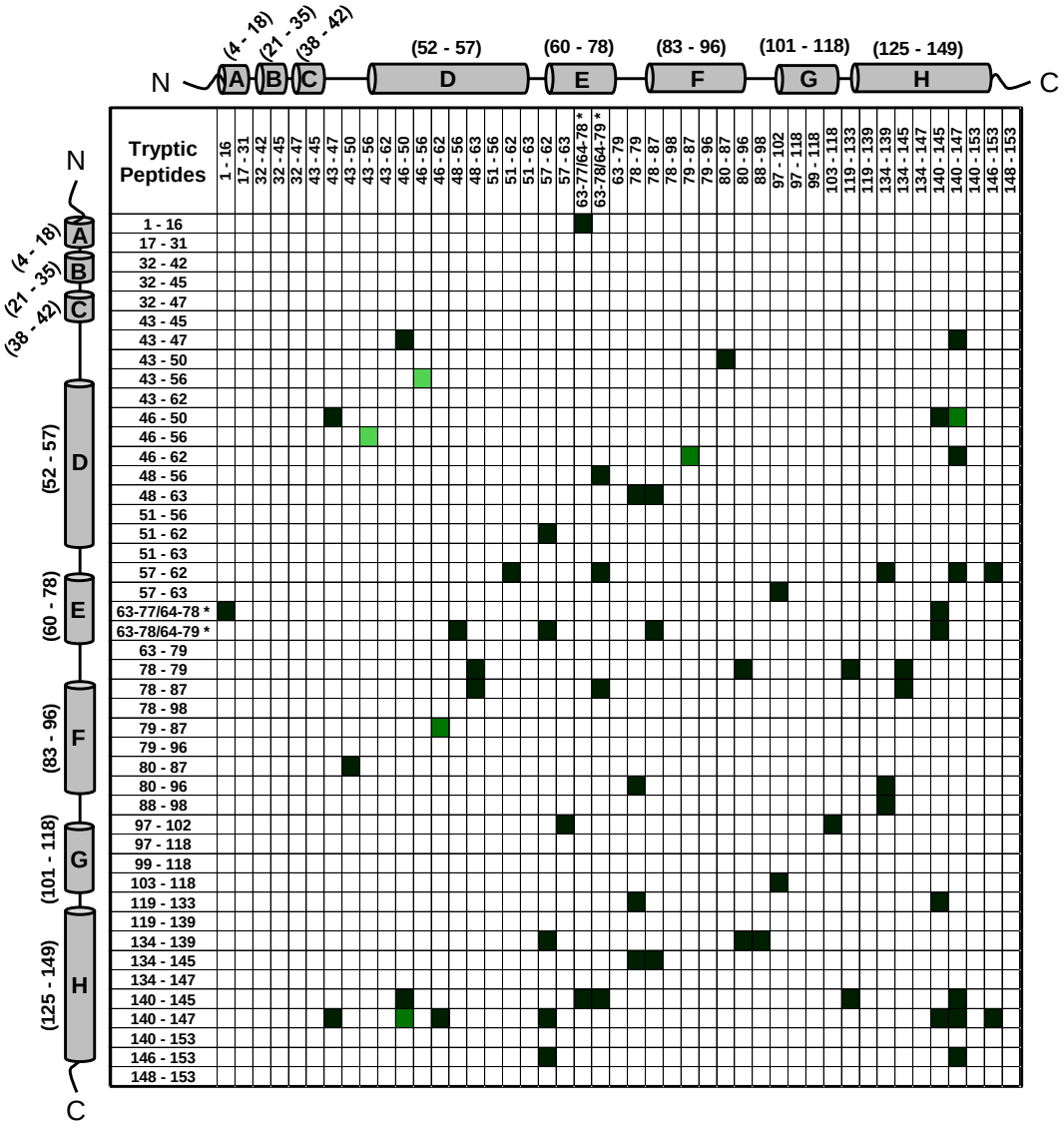
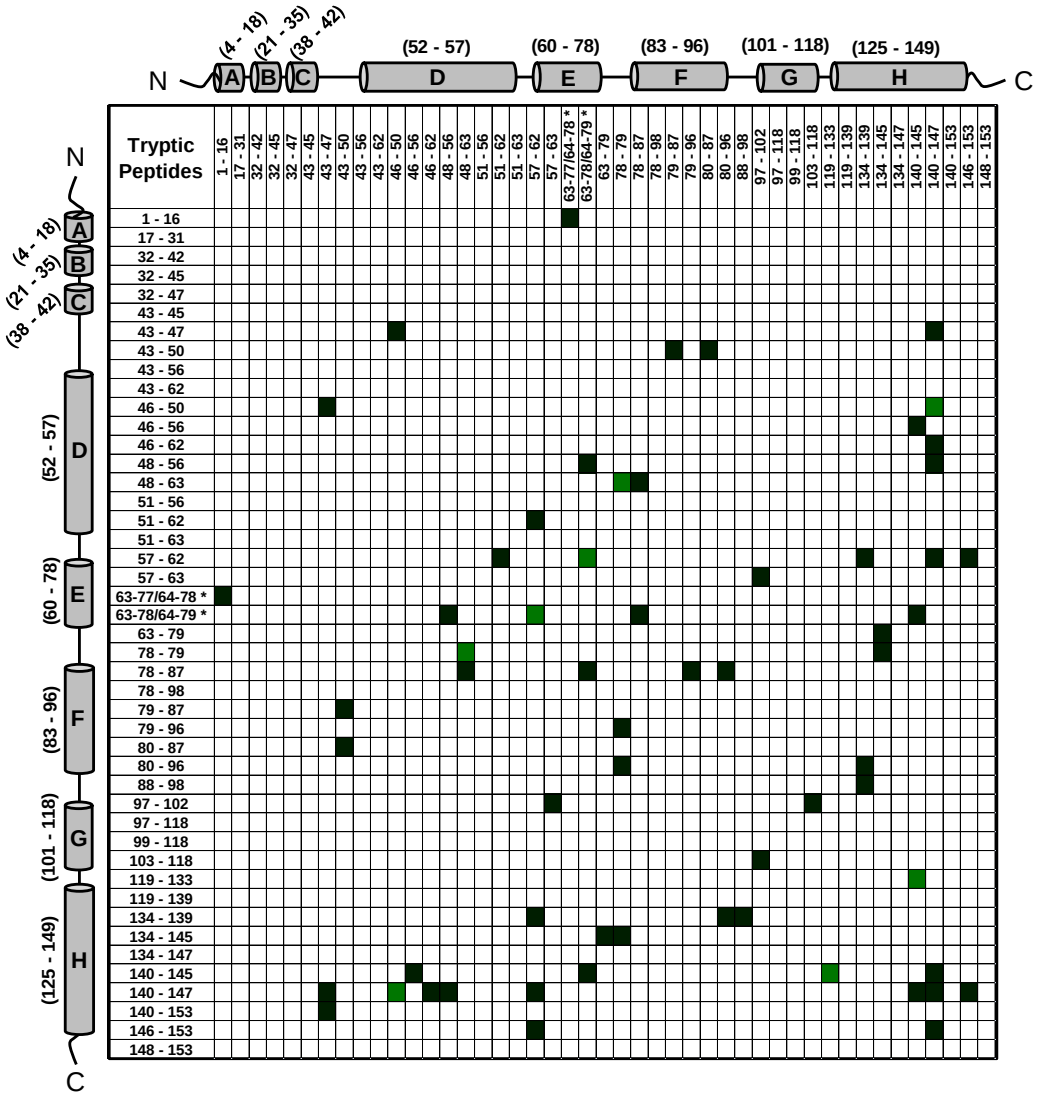


Figure A6. Tryptic peptides of Mb detected as peptide-water adducts in (A) Mb-SDA alone (control), (B) Mb-SDA with raffinose, (C) Mb-SDA with Gdn HCl lyophilized formulations and (D) Tryptic peptides of Mb detected as peptide-raffinose adducts in Mb-SDA with lyophilized raffinose formulation. Peptide-water adducts formed by maximum one (■), two (■), three (■) and four (■) molecules of water from a single sample injection are plotted. Peptide-raffinose adducts formed

with maximum one molecule of raffinose are plotted in orange. The α-helices from N-terminus to C-terminus in Mb are represented by cylinders labeled A to H respectively.

* The molecular mass of tryptic fragments 63-77 and 64-78; 63-78 and 64-79 are identical and cannot be differentiated.


N
(4 - 18)
(21 - 35)
(38 - 42)
(52 - 57)
(60 - 78)
(83 - 96)
(101 - 118)
(125 - 149)
A
B
C
D
E
F
G
H
C
Tryptic Peptides
1 - 16
17 - 31
32 - 42
32 - 45
32 - 47
43 - 45
43 - 47
43 - 50
43 - 56
43 - 62
46 - 50
46 - 56
46 - 62
48 - 56
48 - 63
51 - 56
51 - 62
51 - 63
57 - 62
57 - 63
63-77/64-78 *
63-78/64-79 *
63 - 79
78 - 79
78 - 87
78 - 98
79 - 87
79 - 96
80 - 87
80 - 96
88 - 98
97 - 102
97 - 118
99 - 118
103 - 118
119 - 133
119 - 139
134 - 139
134 - 145
134 - 147
140 - 145
140 - 147
140 - 153
146 - 153
148 - 153


N
(4 - 18)
(21 - 35)
(38 - 42)
(52 - 57)
(60 - 78)
(83 - 96)
(101 - 118)
(125 - 149)
A
B
C
D
E
F
G
H
C
Tryptic Peptides
1 - 16
17 - 31
32 - 42
32 - 45
32 - 47
43 - 45
43 - 47
43 - 50
43 - 56
43 - 62
46 - 50
46 - 56
46 - 62
48 - 56
48 - 63
51 - 56
51 - 62
51 - 63
57 - 62
57 - 63
63-77/64-78 *
63-78/64-79 *
63 - 79
78 - 79
78 - 87
78 - 98
79 - 87
79 - 96
80 - 87
80 - 96
88 - 98
97 - 102
97 - 118
99 - 118
103 - 118
119 - 133
119 - 139
134 - 139
134 - 145
134 - 147
140 - 145
140 - 147
140 - 153
146 - 153
148 - 153

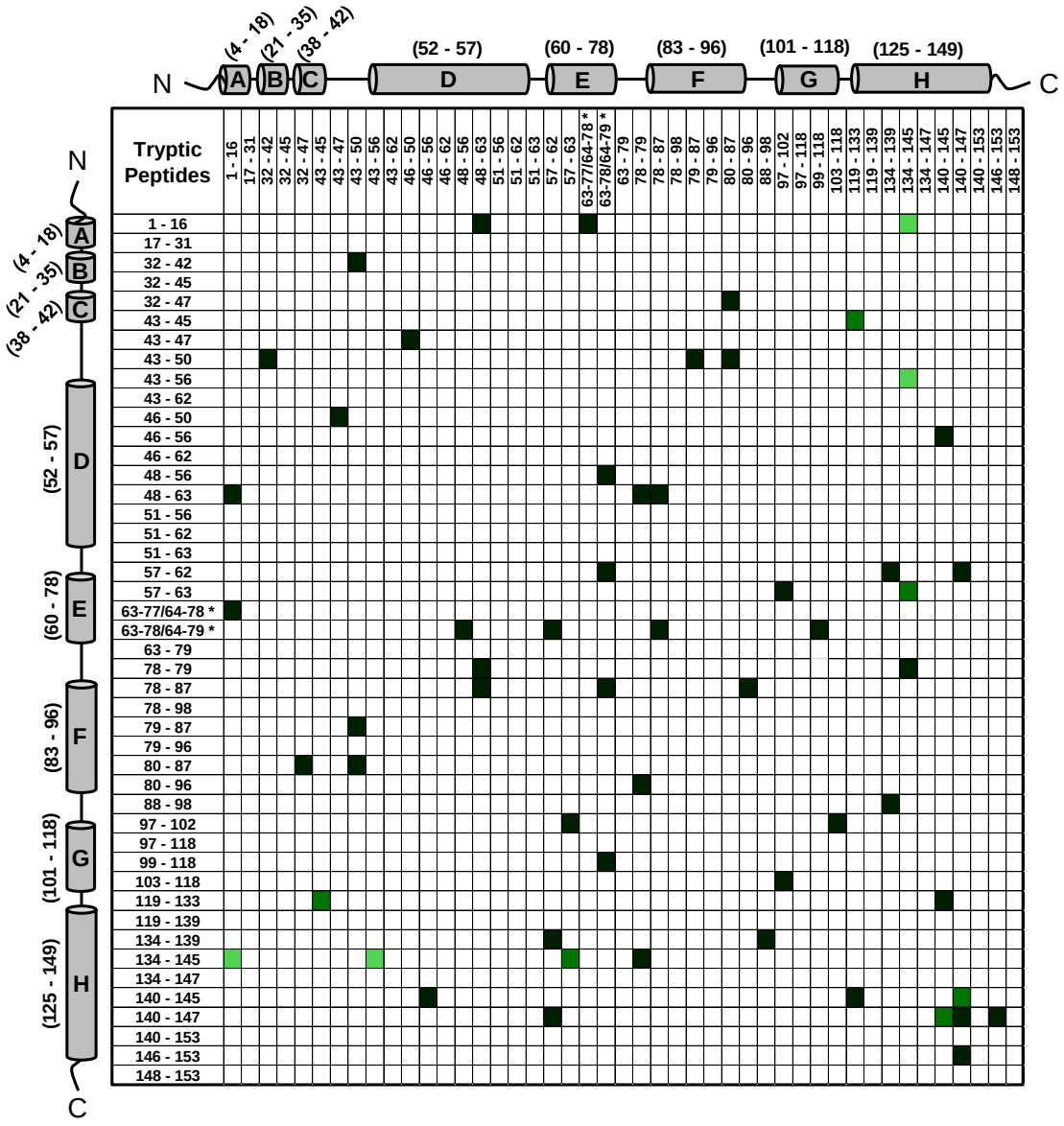


Figure A7. Peptide-peptide adducts formed in (A) Mb-SDA control, (B) Mb-SDA with raffinose and (C) Mb-SDA with Gdn HCl solution formulations. Peptide-peptide adducts detected in single (■), duplicate (■) and triplicate (■) sample injection(s) are plotted. Adducts were mapped irrespective of the number of SDA linkages (1-4 SDA). The α-helices from N-terminus to C-terminus in Mb are represented by cylinders labeled A to H respectively. ***** The molecular mass of tryptic fragments 63-77 and 64-78; 63-78 and 64-79 are identical and cannot be differentiated. The molecular mass for

peptide-peptide adducts (32-45 x 43-47) and (32-47 x 43-45); (79-87 x 51-63) and (78-87 x 51-62); (63-78 x 57-63) and (63-79 x 57-62) are identical and cannot be differentiated.

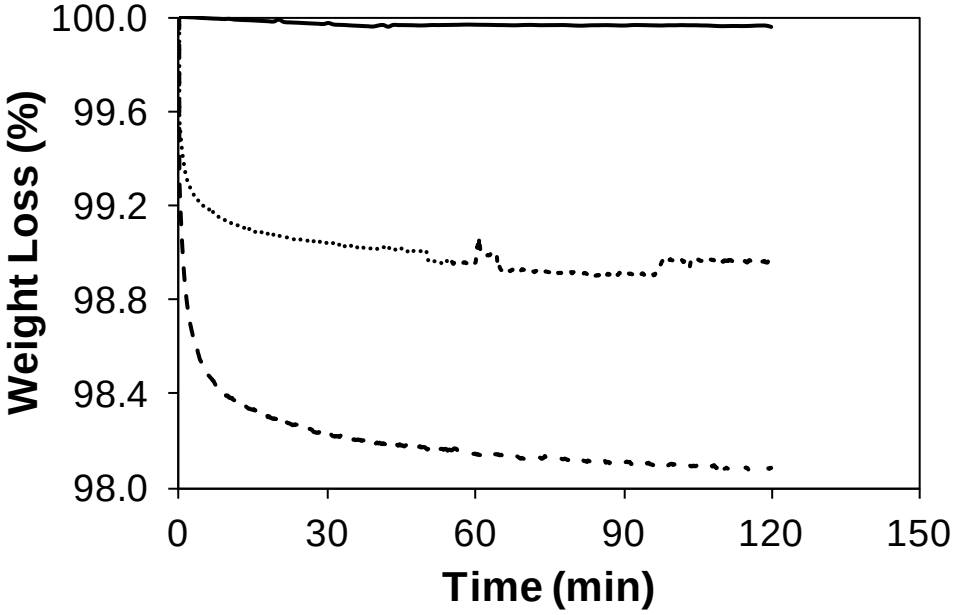


Figure A8. Percent weight loss with time at 50 ˚C, 0 % RH for Mb-SDA alone (dotted line), Mb-SDA with raffinose (dashed line) and Mb-SDA with Gdn HCl (solid line).

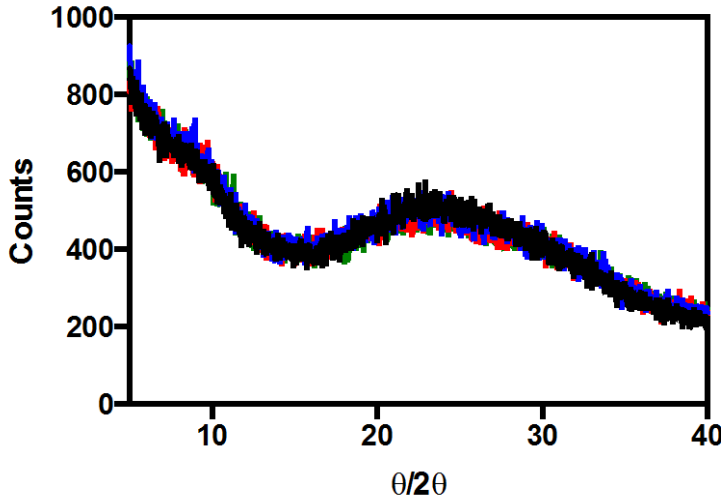


Figure A9. X-ray diffractograms of lyophilized excipient-free (Mb-A) and sucrose-containing (Mb-B) myoglobin formulations. Crystalline features were not observed for Mb-A lyophilized with (black line) and without (blue line) controlled nucleation. Similar diffractograms were obtained for Mb-B lyophilized with (red line) and without (green line) controlled nucleation.

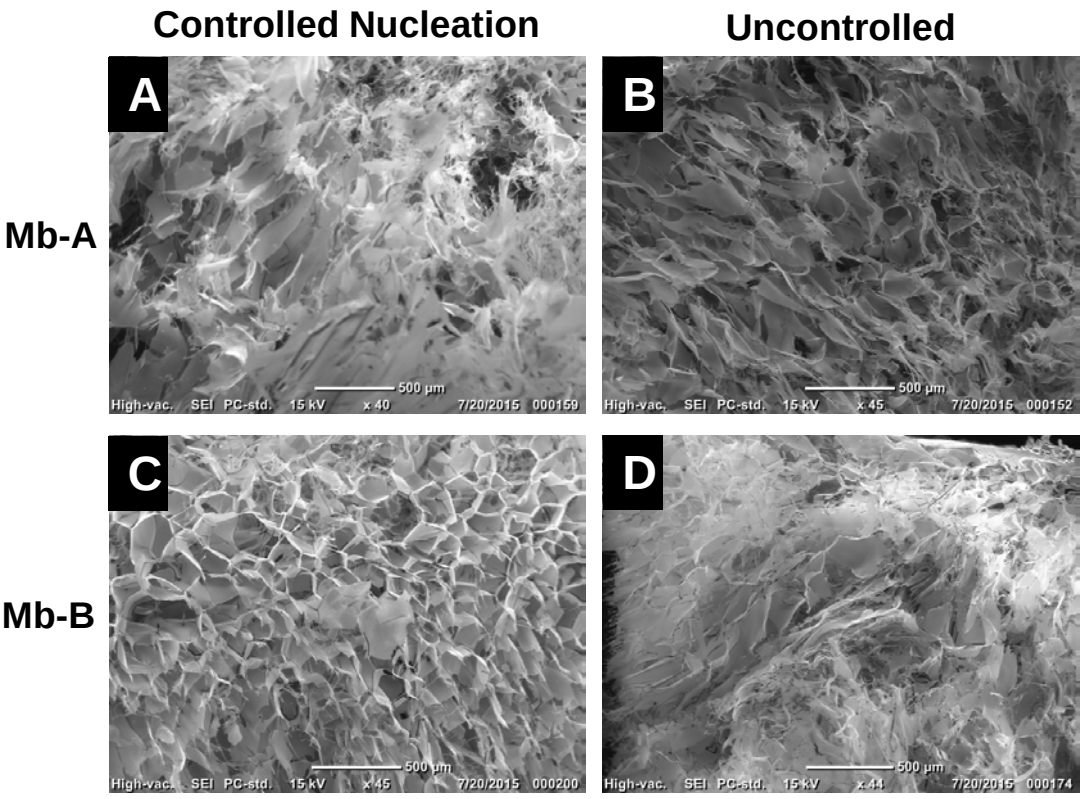


Figure A10. SEM images of excipient-free (Mb-A; panels A, B) and sucrose-containing (Mb-B; panels C, D) lyophilized myoglobin formulations. Panels A, C: Formulations lyophilized with controlled nucleation; Panels B, D: Formulations lyophilized without controlled nucleation. The top and bottom of each image represents the top and bottom of the cake respectively. Scale bars are set at 500 µm.

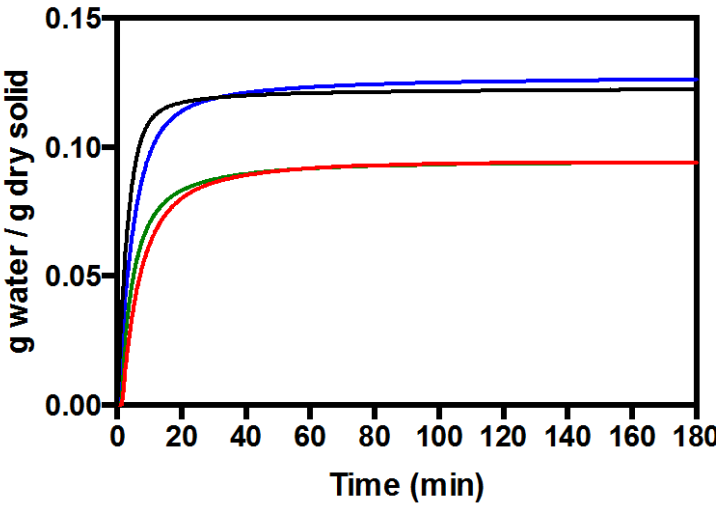


Figure A11. Moisture sorption kinetics for excipient-free (Mb-A) myoglobin formulations lyophilized with (black line) and without (blue line) controlled nucleation and sucrose-containing (Mb-B) myoglobin formulations lyophilized with (red line) and without (green line) controlled nucleation. Moisture sorption was measured at 43 % RH, 5 °C for 3 h.

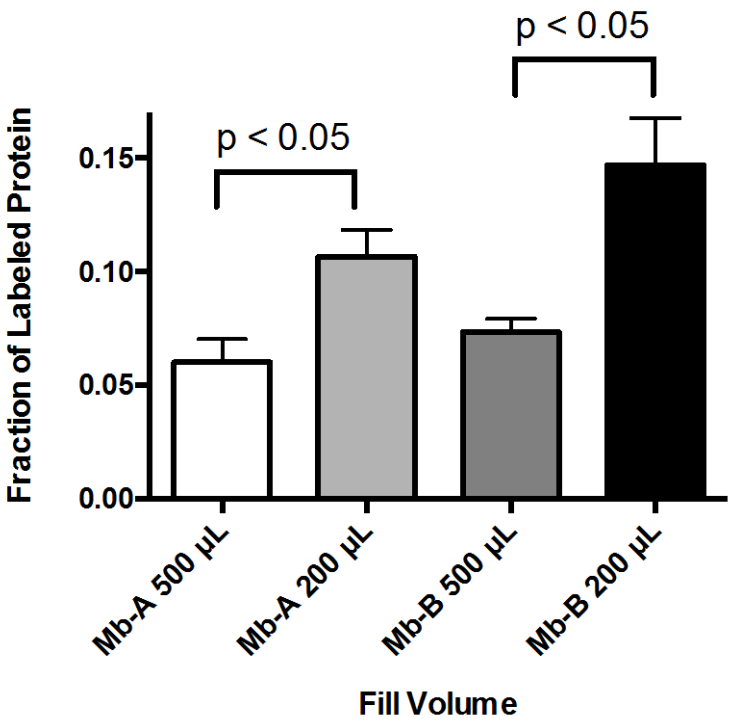


Figure A12. Fraction of protein labeled with photo-leucine for excipient-free (Mb-A) and sucrose-containing (Mb-B) myoglobin formulations lyophilized without controlled nucleation with a fill volume of 200 or 500 µL. All lyophilized formulations contained 100:1 molar ratio of pLeu to protein and were irradiated for 40 min at 365 nm. The fraction was calculated using peak heights of labeled protein on the extracted ion chromatogram (EIC) obtained by LC-MS. The bars show the mean (± SD) of three LC-MS injections. The two fill volumes showed significantly different means using a paired t-test ($p < 0.05$, GraphPad Prism).

VITA

VITA

Lavanya K. Iyer

EDUCATION

Doctor of Philosophy (Industrial and Physical Pharmacy) Aug 2010 - Aug 2015

Purdue University

Doctoral Thesis: High-resolution Mass Spectrometric Approaches to Study Protein Structure and Environment in Lyophilized Formulations

Committee: Dr. Elizabeth Topp (advisor), Dr. Gregory Knipp, Dr. Steven Nail, Dr. Lynne Taylor

Graduate Certificate in Applied Statistics Jan 2013 - Aug 2014

Purdue University

Master of Science in Pharmaceutical Sciences Sep 2008 - June 2010

State University of New York at Buffalo

Masters thesis: Phosphatidylserine/phosphatidylinositol-containing liposomes reduce immunogenicity of recombinant human factor VIII in hemophilia A mice.

Bachelor of Pharmacy June 2004 - June 2008

L.M. College of Pharmacy, Ahmedabad, India (Gujarat University)

RESEARCH SKILLS AND TECHNIQUES

- Biophysical characterization of proteins and peptides
    - Mass Spectrometry **(ESI-Q-TOF)**- LC/MS, MS/MS, Peptide Mapping, Hydrogen-Deuterium Exchange, Covalent Labeling
    - Gel Electrophoresis
    - Chromatography- Size Exclusion Chromatography, HPLC
    - Spectroscopy- UV-Visible, CD, FTIR, Fluorescence
    - Differential Scanning Calorimetry
- Protein derivatization and crosslinking

- Protein formulation
- Protein lyophilization
- Protein expression (bacterial systems) and purification
- ELISA
- Statistics: Design and Analysis of Experiments using SAS, JMP, OriginPro, Prism

WORK EXPERIENCE

Graduate Research Assistant, Purdue University **Aug 2010-May 2015**

- Applied novel covalent labeling and crosslinking approaches with mass spectrometry to study the local environment of protein side-chains in solutions and lyophilized solids.
- Enabled prediction of protein aggregation in the solid state by development of a statistical model to predict aggregation propensity in the presence of different excipients in lyophilized formulations.
- Engineering unnatural amino acids in a protein by successfully incorporating L-photo-methionine into the sequence of calmodulin.

Intern at Baxter Healthcare (Bloomington, IN) **Aug 2014- Oct 2014**

- Worked on lyophilization process design and control with focus on studying formulation characteristics.

Teaching Assistant, Parenteral Products (Purdue University) **Jan 2012- May 2012**

- Trained undergraduate students in aseptic techniques for preparation and testing of sterile dosage forms.

Research Assistant, State University of New York at Buffalo **Sep 2008-Jun 2010**

Research Advisor: Dr. Sathy Balu-Iyer

- Formulated factor VIII in liposomes and assessed its immunogenicity in hemophilic mice.
- Gained considerable experience with ELISA and understanding the effects of neutralizing antibodies on protein pharmacodynamics.

ACTIVITIES AND ORGANIZATIONS

- Member, American Association of Pharmaceutical Scientists (AAPS) 2012- present
- Secretary, AAPS- Student Chapter (Purdue University) 2013-14

PUBLICATIONS

- Characterizing Protein Structure, Dynamics and Conformation in Lyophilized Solids. Balakrishnan S. Moorthy, **Lavanya K. Iyer** and Elizabeth M. Topp, *Current Pharmaceutical Design* (in press, June 2015)
- Photolytic Crosslinking to Probe Protein-Protein and Protein-Matrix Interactions in Lyophilized Powders. **Lavanya K. Iyer**, Balakrishnan S. Moorthy and Elizabeth M. Topp, *Molecular Pharmaceutics* (in press, Aug 2015)
- Effect of hydrolytic degradation on the *in vivo* properties of monoclonal antibodies. Balakrishnan Shenbaga Moorthy, Bo Xie, Ehab Moussa, **Lavanya K. Iyer**, Saradha Chandrasekhar, Jainik Panchal, Elizabeth M. Topp, chapter in Biobetters: Protein Engineering to Approach the Curative (2015)
- Mass Spectrometric Approaches to Study Protein Structure and Interactions in Lyophilized Powders. Balakrishnan S. Moorthy, **Lavanya K. Iyer** and Elizabeth M. Topp. *Journal of Visualized Experiments* (2015) 98, e52503.
- Photolytic Labeling To Probe Molecular Interactions in Lyophilized Powders. **Lavanya K. Iyer**, Balakrishnan S. Moorthy and Elizabeth M. Topp, *Molecular Pharmaceutics* (2013) 10, 4629-39.
- Protein aggregation and lyophilization: Protein structural descriptors as predictors of aggregation propensity. Brock C. Roughton, **Lavanya K. Iyer**, Esben Bertelsen, Elizabeth M. Topp, Kyle V. Camarda, *Computers and Chemical Engineering* (2013) 58:11, 366-77.
- Microarrays and microneedle arrays for delivery of peptides, proteins, vaccines and other applications. S. Chandrasekhar, **L.K. Iyer**, J.P. Panchal, E.M. Topp, J.B. Cannon and V.V. Ranade, *Expert Opinion in Drug Delivery* (2013) 10:8, 1155-70.
- Proteins and Peptides: Chemical and Physical Stability. Andreas M. Sophocleous, Jun Zhang, **Lavanya K. Iyer**, Saradha Chandrasekhar, Esben Bertelsen and Elizabeth M. Topp (2013), Encyclopedia of Pharmaceutical Science and Technology, 4th ed.

SELECTED PRESENTATIONS

- Photolytic Crosslinking-Mass Spectrometry to Probe the Environment of Lyophilized Proteins. **Lavanya K. Iyer** and Elizabeth M. Topp, Workshop on Protein Aggregation and Immunogenicity, Breckenridge, CO, Jul 15-17, 2014
- Photolytic Labeling-Mass Spectrometry to Probe Protein-Matrix Interactions in Lyophilized Solids. **Lavanya K. Iyer** and Elizabeth M. Topp, AAPS National Biotechnology Conference, San Diego, CA, May 20-22, 2013 (oral presentation)

- Protein Aggregation in Lyophilized Solids: Protein Structural Descriptors as Predictors of Aggregation Propensity. **Lavanya K. Iyer** and Elizabeth M. Topp, AAPS Annual Meeting, Chicago, IL, Oct 10-14, 2012
- Aggregation and Stabilization of Lyophilized Proteins by Excipients. **Lavanya K. Iyer** and Elizabeth M. Topp, Pharmaceutics Graduate Students Research Meeting, Madison, WI, June 23-25, 2011

HONORS AND ACHIEVEMENTS

- Awarded the Jenkins-Knevel Award for Outstanding Graduate Research, Industrial and Physical Pharmacy, Purdue University (2014)
- Awarded the Allen Chao Fellowship in Industrial Pharmacy, Purdue University (2014-15)
- Recipient of AAPS Graduate Student Symposium Award sponsored by Eli Lilly & Company at National Biotechnology Conference (NBC), May 2013
- Awarded the Ronald Dollens Graduate Scholarship in Life Sciences, 2011-12 and 2013-14
- Awarded prize for best poster in Biotechnology at Pharmaceutics Graduate Student Research Meeting (PGSRM), June 2011