Published in final edited form as:

Mol Pharm. 2015 September 8; 12(9): 3237–3249. doi:10.1021/acs.molpharmaceut.5b00183.

Photolytic Crosslinking to Probe Protein-Protein and Protein-Matrix Interactions In Lyophilized Powders

Lavanya K. Iyer, Balakrishnan S. Moorthy, and Elizabeth M. Topp*

Department of Industrial and Physical Pharmacy, Purdue University, West Lafayette, IN

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.

Graphical Abstract

*To whom correspondence should be addressed: Department of Industrial and Physical Pharmacy, Purdue University, 575 Stadium Mall Drive, Room 124D, West Lafayette, IN 47907-2091, Phone: 765-494-1450, Fax: 765-494-6545, topp@purdue.edu.

SUPPORTING INFORMATION

Additional information can be found in Supporting Information available online. Detailed information: Data Analysis for Crosslinking Numbers and Qualitative Matrices. Additional tables: Table S1 - List of SDA-labeled tryptic peptides selected for LC-MS/MS analysis; Table S2 - List of all possible peptide-water adducts that can be formed by crosslinking with SDA. Additional figures: Figure S1- Mechanism of SDA crosslinking; Figure S2- Comparison of labeling and crosslinking; Figure S3 - Second derivative FTIR spectra of Mb-SDA and unlabeled Mb in control, raffinose and Gdn HCl formulations; Figure S4 - Peptide-peptide adducts formed in Mb-SDA alone (control), Mb-SDA with raffinose and Mb-SDA with Gdn HCl lyophilized formulations by a maximum of one, two, three and four SDA molecules; Figure S5 - Peptide-water and peptide-raffinose adducts formed in Mb-SDA alone (control), Mb-SDA with raffinose and Mb-SDA with Gdn HCl lyophilized formulations with a maximum of one, two, three and four water or raffinose molecules; Figure S6 - Peptide-peptide adducts formed in Mb-SDA alone (control), Mb-SDA with raffinose and Mb-SDA with Gdn HCl solution formulations detected in a single, duplicate or triplicate injection(s); Figure S7 - Moisture content as measured by percent weight loss with time at 50 °C, 0 % RH for Mb-SDA alone, Mb-SDA with raffinose and Mb-SDA with Gdn HCl. This material is available free of charge via the Internet at <http://pubs.acs.org>.

Keywords

Photolytic crosslinking; mass spectrometry; lyophilized formulations; protein-protein interactions; protein-excipient interactions; protein-water interactions; myoglobin; succinimidyl 4,4'-azipentanoate; raffinose

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 ^{2,3}, 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 ⁴⁻⁶. 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, conventional techniques such as Fourier-transform infrared spectroscopy (FTIR) provide only low-resolution information about lyophilized protein secondary structure, 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 ⁷⁻⁹. 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 ^{10–12}. 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 FTIR 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, the glass transition temperature (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 ^{13,14}. 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 ^{15,16}, and, in a recent study of lyophilized myoglobin formulations, provided significantly higher correlation with aggregation during storage than FTIR ¹⁷. 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 an approach that is complementary to ssHDX-MS called solid-state photolytic labeling- mass spectrometry (ssPL-MS) ¹⁸. 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 ¹⁸.

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^{19–21}, 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 (Supporting Information, Fig. S1). 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 (Supporting Information, Fig. S2). 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 (Supporting Information, Fig. S1). 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 molecules²². 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 N₂ and the formation of a reactive carbene. 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.

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).

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 $\epsilon_{555\text{nm}} = 12.92 \text{ mM}^{-1}\text{cm}^{-1}$). 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.

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.

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 10× 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 concentration²³. 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):

$$F_{L,i} = \frac{(\text{Peak height})_{L,i}}{\sum_{i=0}^{10} (\text{Peak height})_{L,i}} \quad (1)$$

where $F_{L,i}$ is the fraction of protein containing i SDA labels ($i = 0, 1, \dots, 10$), 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 ($i = 2, \dots, 10$). The concentrations of each labeled species ($P_{L,i}$) were calculated by multiplying $F_{L,i}$ by the initial protein concentration (P_0).

$$P_{L,i} = P_0 F_{L,i} \quad (2)$$

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

$$P = P_0 F_{L,i=0} \quad (3)$$

$$X = X_0 - \sum_{i=1}^{10} i P_{L,i} \quad (4)$$

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.

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 1). A second formulation containing Mb-SDA and Gdn was prepared with a final concentration of 1.5 M Gdn (Table 1). The formulations were lyophilized as described previously¹⁸. 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, 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.

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 previously¹⁵. 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 previously¹⁸. 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.

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.

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 × 50 mm, particle size 3.5 µm) using a gradient, as described previously¹⁸. MS/MS was performed on selected peptides labeled with SDA (Supporting Information, Table S1). The peptides were fragmented using CID (13 V) and the product ions analyzed using MassHunter software.

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 (C₅H₆O) 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-N₂), in which N₂ 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 in Supporting Information (refer section ‘Data Analysis for Crosslinking Numbers (X_{1n})’).

RESULTS

Intact Protein Labeling with SDA

Following initial succinimidyl derivatization, Mb-SDA carrying up to five labels was detected by LC-MS (Fig. 1). No significant secondary structural changes after SDA-labeling were detected using CD spectroscopy (Fig. 2A) and solid-state FTIR spectroscopy (Fig. S3). 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. 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.

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). LC-MS/MS analysis conclusively established that labeling occurred on the N-terminal Gly¹, Lys⁴², Lys⁵⁰, Lys⁵⁶, Lys⁸⁷ and Lys¹⁴⁷, 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 Lys¹⁶, Lys⁷⁷, Lys⁷⁸, Lys⁷⁹, Lys⁹⁶ and Lys¹¹⁸. 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 Asn¹⁴⁰-Lys¹⁴⁷ showed 4 SDA labels, although it contains only two Lys. Similarly, peptides Val¹⁷-Arg³¹ (containing no Lys), His¹¹⁹-Lys¹³³ (one Lys) and Ala⁵⁷-Lys⁶³ (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^{24,25}.

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 S2, 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 Supporting Information (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. 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. 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 light²⁶. 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. 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. 4B). In the Gdn HCl formulation, the map shows a number of interactions not detected in the control and raffinose formulations (Fig. 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.

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. 5 and Fig. S6 in Supporting Information). 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. 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.

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 2). 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.

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). Peptide crosslinking numbers (X_{1n}) were calculated as described in Supporting Information. Normalized X_{1n} values (denoted X_{1n}^*) for each group were averaged across triplicate LC-MS measurements and expressed as mean $\pm$ 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 $6.8 (\pm 1.7)$ (Table 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 $6.8 (\pm 0.9)$ (Table 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. 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. 6B, white bars).

In the lyophilized raffinose formulation, ΣX_{11}^* was 43 % greater than the excipient-free control (Table 4; Fig. 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 5; Fig. 7B, grey bars). The X_{12}^* values differed among the peptide fragment groups in the raffinose formulation (Table 5; Fig. 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 6; Fig. 6C).

In the lyophilized Gdn HCl formulation, the ΣX_{11}^* value was 3 times greater than the control, indicating more distinct peptide-peptide adducts (Table 4; Fig. 7A, black bars). ΣX_{12}^* for this formulation was 1.6 times greater than the control, indicating more distinct peptide-water adducts (Table 5; Fig. 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. 6A, black bars). X_{12}^* values were significantly greater than in the control formulation for all groups except Groups (4), (5) and (6) (Fig. 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. 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. 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. 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 (Supporting Information, Fig. S7). The raffinose formulation showed the fewest peptide-water adducts (Table 5, Fig. 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.

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. 4 and 5, Fig. S4 and S5 in Supporting Information). Not all theoretically possible adducts were observed, as shown by the white boxes in the interaction maps (Fig. 4 and 5, Fig. S4 and S5 in Supporting Information). 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. 4A, B). Motions of loop regions are linked to conformational transitions involving the helices of Mb²⁷. It has been shown experimentally and computationally that the CD and EF loops of holoMb are especially flexible, allowing for efficient ligand binding^{28–30}. This loop flexibility may result in better protein-protein contacts in the solid state. Any disruption of the salt bridge between residues Lys⁴⁵ and Asp⁶⁰ that normally stabilizes the DE link with the CD loop^{31,32} 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^{33,34}. 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^{27,35}. 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 data have shown loss of backbone protection in helix H upon lyophilization¹⁷, 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. 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. 4 and 5) can be used to qualitatively describe the type of adducts formed, the number of adducts (Table 2) 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 2) 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 2).

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 His⁶⁴ residue (helix E) in the heme-binding pocket; this residue is involved in modulating heme-ligand affinity by binding to water^{36,37}. This suggests that there is a hydration layer around helix E, 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. 5; only 2 peptides Lys⁴⁶-Lys⁵⁶ and Lys⁶³-Lys⁷⁷/Lys⁶⁴-Lys⁷⁸ 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. 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 solution³⁸. 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^{10,39}. The band area was found to be smaller in proteins lyophilized in the absence of carbohydrate excipients, but increased with increasing carbohydrate concentration¹⁰. 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. 5A-F). Overall peptide-peptide interactions increased in the solid state for the raffinose and Gdn HCl formulations, compared to solution (Table 2, Fig. 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 (Table 2, Fig. 7A). This may be due in part to protein unfolding and/or 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 2, Fig. 7B), even though the protein is partially unfolded at 1.5 M Gdn HCl⁴⁰. We hypothesize that when Gdn HCl crystallizes, the SDA-labeled residues are free to crosslink with water molecules, resulting in increased ΣX_{12}^* in the solid state compared to solution.

Preferential exclusion of carbohydrates is known to occur in solution at concentrations 0.2 M^{41, 42}. 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 2, Fig. 7B), peptide-raffinose adducts were fewer in the solid state than in solution (Table 2, Fig. 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. 5C, D, G, H, 7B, C). Chatterjee et al. have reported crystallization and phase separation of raffinose during annealing, although the final lyophilized product was amorphous⁴³. 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 rates⁴⁴. 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.

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_{In}^* values.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from NIH RO1 GM085293 and thank Dr. Niraj Trasi for assistance with X-ray diffraction experiments.

REFERENCES

1. Aggarwal RS. What's fueling the biotech engine-2012 to 2013. *Nat Biotechnol.* 2014; 32(1):32–39. [PubMed: 24406926]
2. Wang W, Nema S, Teagarden D. Protein aggregation--pathways and influencing factors. *Int J Pharm.* 2010; 390(2):89–99. [PubMed: 20188160]
3. Chi EY, Krishnan S, Randolph TW, Carpenter JF. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res.* 2003; 20(9):1325–1336. [PubMed: 14567625]
4. Kasper JC, 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. [PubMed: 21426937]
5. Luthra S, Obert JP, Kalonia DS, Pikal MJ. 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. [PubMed: 17031859]
6. Lai MC, Topp EM. Solid-state chemical stability of proteins and peptides. *J Pharm Sci.* 1999; 88(5):489–500. [PubMed: 10229638]
7. Costantino HR, Carrasquillo KG, Cordero RA, Mumenthaler M, Hsu CC, Griebenow K. Effect of excipients on the stability and structure of lyophilized recombinant human growth hormone. *J Pharm Sci.* 1998; 87(11):1412–1420. [PubMed: 9811499]
8. Cleland JL, Lam X, Kendrick B, Yang J, Yang TH, Overcashier D, Brooks D, Hsu C, Carpenter JF. 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. [PubMed: 11170024]
9. Sane SU, Wong R, Hsu CC. 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. [PubMed: 14999736]
10. Allison SD, Chang B, Randolph TW, Carpenter JF. Hydrogen bonding between sugar and protein is responsible for inhibition of dehydration-induced protein unfolding. *Arch Biochem Biophys.* 1999; 365(2):289–298. [PubMed: 10328824]
11. Prestrelski SJ, Tedeschi N, Arakawa T, Carpenter JF. Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers. *Biophys J.* 1993; 65(2):661–671. [PubMed: 7693001]
12. Crowe JH, Carpenter JF, Crowe LM. The role of vitrification in anhydrobiosis. *Annu Rev Physiol.* 1998; 60:73–103. [PubMed: 9558455]

13. Hancock BC, Shamblin SL, Zografi G. Molecular mobility of amorphous pharmaceutical solids below their glass transition temperatures. *Pharm Res.* 1995; 12(6):799–806. [PubMed: 7667182]
14. Chang LL, Shepherd D, Sun J, Tang XC, Pikal MJ. 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. [PubMed: 15920766]
15. Sophocleous AM, Zhang J, Topp EM. Localized hydration in lyophilized myoglobin by hydrogen-deuterium exchange mass spectrometry 1. Exchange mapping. *Mol Pharm.* 2012; 9(4):718–726. [PubMed: 22352965]
16. Li Y, Williams TD, Topp EM. Effects of excipients on protein conformation in lyophilized solids by hydrogen/deuterium exchange mass spectrometry. *Pharm Res.* 2008; 25(2):259–267. [PubMed: 17597380]
17. Moorthy BS, Schultz SG, Kim SG, Topp EM. 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. [PubMed: 24816133]
18. Iyer LK, Moorthy BS, Topp EM. Photolytic labeling to probe molecular interactions in lyophilized powders. *Mol Pharm.* 2013; 10(12):4629–4639. [PubMed: 24125175]
19. 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. [PubMed: 15782189]
20. 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. [PubMed: 21128684]
21. Chin JW, Martin AB, King DS, Wang L, Schultz PG. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2002; 99(17):11020–11024. [PubMed: 12154230]
22. Jumper CC, Schriemer DC. Mass spectrometry of laser-initiated carbene reactions for protein topographic analysis. *Anal Chem.* 2011; 83(8):2913–2920. [PubMed: 21425771]
23. Mendoza VL, Vachet RW. Protein surface mapping using diethylpyrocarbonate with mass spectrometric detection. *Anal Chem.* 2008; 80(8):2895–2904. [PubMed: 18338903]
24. Swaim CL, Smith JB, Smith DL. 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. [PubMed: 15121203]
25. 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. [PubMed: 18724398]
26. Dubinsky L, Krom BP, Meijler MM. Diazirine based photoaffinity labeling. *Bioorg Med Chem.* 2012; 20(2):554–570. [PubMed: 21778062]
27. Elber R, Karplus M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science.* 1987; 235(4786):318–321. [PubMed: 3798113]
28. Schulze BG, Grubmüller H, Evanseck JD. 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.
29. Cammarata MB, Brodbelt JS. Structural characterization of holo- and apomyoglobin in the gas phase by ultraviolet photodissociation mass spectrometry. *Chemical Science.* 2015; 6(2):1324–1333.
30. 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. [PubMed: 2231732]
31. Angeloni L, Feis A. Protein relaxation in the photodissociation of myoglobin-CO complexes. *Photochem Photobiol Sci.* 2003; 2(7):730–740. [PubMed: 12911220]
32. Balasubramanian S, Lambright DG, Marden MC, Boxer SG. CO recombination to human myoglobin mutants in glycerol-water solutions. *Biochemistry.* 1993; 32(9):2202–2212. [PubMed: 8443162]
33. Barrick D, Baldwin RL. Stein Moore Award address. The molten globule intermediate of apomyoglobin and the process of protein folding. *Protein Sci.* 1993; 2(6):869–876. [PubMed: 8318892]

34. Jennings PA, Wright PE. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science*. 1993; 262(5135):892–896. [PubMed: 8235610]
35. Pappu RV, Weaver DL. The early folding kinetics of apomyoglobin. *Protein Science*. 1998; 7(2): 480–490. [PubMed: 9521125]
36. Rohlfis RJ, Mathews AJ, Carver TE, Olson JS, Springer BA, Egeberg KD, Sligar SG. 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. [PubMed: 2303446]
37. Quillin ML, Arduini RM, Olson JS, Phillips GN Jr. High-resolution crystal structures of distal histidine mutants of sperm whale myoglobin. *J Mol Biol*. 1993; 234(1):140–155. [PubMed: 8230194]
38. Carpenter JF, Crowe JH. An infrared spectroscopic study of the interactions of carbohydrates with dried proteins. *Biochemistry*. 1989; 28(9):3916–3922. [PubMed: 2526652]
39. Remmele RL Jr, Stushnoff C, Carpenter JF. 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. [PubMed: 9434273]
40. Bismuto E, Colonna G, Irace G. Unfolding pathway of myoglobin Evidence for a multistate process. *Biochemistry*. 1983; 22(18):4165–4170. [PubMed: 6626499]
41. Lee JC, Timasheff SN. The stabilization of proteins by sucrose. *J Biol Chem*. 1981; 256(14):7193–7201. [PubMed: 7251592]
42. Xie G, Timasheff SN. The thermodynamic mechanism of protein stabilization by trehalose. *Biophys Chem*. 1997; 64(1–3):25–43. [PubMed: 9127936]
43. Chatterjee K, Shalaev EY, Suryanarayanan R. Raffinose crystallization during freeze-drying and its impact on recovery of protein activity. *Pharm Res*. 2005; 22(2):303–309. [PubMed: 15783079]
44. 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. [PubMed: 24356771]

Figure 1. Deconvoluted mass spectrum of Mb labeled with 10× 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 2.

(A) Far-UV CD spectra of Mb without SDA labeling (dotted line) and Mb labeled with 10 $\times$ 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; [P₀], initial protein concentration; [X], SDA remaining unused after quenching the labeling reaction; [X₀], 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.

Figure 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).

Figure 4.

Peptide-peptide adducts formed in (A) Mb-SDA alone (control), (B) Mb-SDA with raffinose and (C) Mb-SDA with Gdn HCl formulations. 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 $\times$ 43–47) and (32–47 $\times$ 43–45);

$(79-87 \times 51-63)$ and $(78-87 \times 51-62)$; $(63-78/64-79 \times 57-63)$ and $(63-79 \times 57-62)$ are identical and cannot be differentiated.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 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. 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.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 6.

Comparison of X_{11}^* (A), X_{12}^* (B) and X_{13}^* (C) values for Mb in Mb-SDA lyophilized with no excipients (white bars), Mb-SDA lyophilized with raffinose (grey bars) and Mb-SDA lyophilized with Gdn HCl (black bars). See text for details of calculations (Mean $\pm$ SD (n=3)). Note that in the abscissa for panel (C), Group (6) spanning residues Lys⁷⁸-Lys⁹⁸ was expanded to Lys⁷⁸-Lys¹⁰² to accommodate peptide Lys⁷⁹-Lys¹⁰² which was found to form raffinose adducts.

Figure 7. Comparison of ΣX_{11}^* (A), ΣX_{12}^* (B) and ΣX_{13}^* (C) values for Mb in solution-state (white bars) and lyophilized formulation (black bars) (Mean $\pm$ SD (n=3)). See text for details of calculations.

Table 1

Composition of lyophilized formulations.

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

^aMb, myoglobin; Gdn HCl, guanidine hydrochloride; SDA, succinimidyl 4,4'-azipentanoate.^bMb-SDA (10×) denotes Mb labeled with 10× molar excess of SDA in solution.

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 ($\pm$ SD) from three LC-MS injections.

Table 2

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 $\pm$ 3.5	30.3 $\pm$ 0.6	50.7 $\pm$ 3.2	31.3 $\pm$ 1.2	105.0 $\pm$ 12.3	28.3 $\pm$ 2.1
Peptide-Water	42.7 $\pm$ 3.8	30.0 $\pm$ 1.7	11.3 $\pm$ 1.2	19.7 $\pm$ 1.2	74.3 $\pm$ 10.2	34.3 $\pm$ 1.5
Peptide-Excipient	N/A	N/A	11.3 $\pm$ 0.6	41.3 $\pm$ 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.

Table 3

Classification of peptides based on trypsin digestion pattern.

Group	Amino Acids Included	Number of SIDA 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
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 ⁸⁸ 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

Group	Amino Acids Included	Number of SDA labels ^a	Number of trypsin cleavage sites	Tryptic Peptides included	Secondary Structure Elements
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.

Table 4

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

Group	X_{11}^* ($\pm$ 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 5

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

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

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

Table 6

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

Group	X_{13}^* ($\pm$ 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

^aNote that Group (6) (spanning Lys⁷⁸-Lys⁹⁸) was expanded slightly to Lys⁷⁹-Lys¹⁰² to accommodate peptide Lys⁷⁹-Lys¹⁰² that was found to form a raffinose adduct.