Louisiana State University
LSU Digital Commons

LSU Doctoral Dissertations | Graduate School

2014

Reductive Alkylation of Proteins Towards Structural and Biological Applications

Kevin Jerome Roberson
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Chemistry Commons

Recommended Citation

Roberson, Kevin Jerome, "Reductive Alkylation of Proteins Towards Structural and Biological Applications" (2014). *LSU Doctoral Dissertations*. 1267.
https://digitalcommons.lsu.edu/gradschool_dissertations/1267

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contactgradetd@lsu.edu.

REDUCTIVE ALKYLATION OF PROTEINS TOWARDS STRUCTURAL AND BIOLOGICAL APPLICATIONS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Chemistry

by
Kevin Jerome Roberson
B.S., Georgia Southern University, 2007
August 2014

To the pursuit of consciousness…surviving is not enough!

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Megan Macnaughtan, for guidance and support throughout my graduate studies. I would also like to thank my committee members, Dr. Isiah Warner, Dr. Kermit Murray, Dr. William Crowe, and Dr. William Monroe for their advice and support. Thanks to Dr. Fareed Aboul-ela for his support and advice as well. I sincerely thank Dr. Dale Treleaven and Dr. Thomas Weldeghiorghis for their invaluable help and teachings of NMR.

I thank the Macnaughtan Research Group for their support and friendship over the years. Octavia “Buddarris” Goodwin, you have been one of my biggest supporters in a place so far away from family. I thank you and Mrs. Judy for being my family when I couldn’t reach mine.

I give a special thank Dr. Karen Terry Welch and Marion Welch. I am blessed to have crossed paths with you. I thank you for being selfless enough to invest in the potential I couldn’t see in myself. Although these words do no justice to my gratitude, I offer them in a lifelong attempt to fully express how much you mean to me and my family. Just know I’m working on something big…

I thank my mom, Diane Roberson, for the many prayers and sacrifices you have made on my behalf so that I may be in this position. My siblings for your unwavering support. Fred, I still want to be like you when I grow up. Michelle, the bond we share has always been special and I’m a much better person because of you. My angels, Jessica and Vanessa, I was always jealous cause I didn’t have a twin. I wouldn’t work and push as hard as I do in life if not for you. This would not be possible without any of you. I love you!

Last but definitely not least, I thank Dr. Marsha Cole. Your heart is as pure as they come. I thank you for being there for me throughout this journey. Your support always seem to grow

when it was hardest to be there for me. Thank you for sacrificing sleep with me, challenging me, and somehow always polishing any idea I presented to you. It did not go unnoticed. I thank you for helping me achieve such a milestone in my life, and I look forward to the amazing things you will produce in the future. I love you!

-Kevin J. Roberson

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF SCHEMES ix

LIST OF FIGURES x

ABSTRACT xii

CHAPTER 1 INTRODUCTION 1
- 1.1 Protein Structure and Stability 1
  - 1.1.1 Protein Structure 1
  - 1.1.2 Classification of Protein Structure 6
  - 1.1.3 Protein Denaturation 8
- 1.2 Protein Modification 11
  - 1.2.1 Protein Synthesis 12
  - 1.2.2 Post-translational Modification 13
  - 1.2.3 Recombinant Protein Expression 15
  - 1.2.4 Chemical Modification 17
  - 1.2.5 Side Chain Selective Modifications 19
- 1.3 Reductive Methylation of Proteins 26
  - 1.3.1 Techniques to Study Reductive Methylation of Proteins 28
  - 1.3.2 Protein Structure: NMR and X-Ray Crystallography 30
  - 1.3.3 Protein Samples for NMR 34
  - 1.3.4 NMR Assignment Strategies 36
  - 1.3.5 Assignment of Reductively Methylated HEWL 42
- 1.4 Conclusion 48
- 1.5 References 48

CHAPTER 2 ATTEMPTS TOWARD UNAMBIGUOUSLY ASSIGNING ^{13}C-DIMETHYLAMINE NMR RESONANCES 59
- 2.1 Introduction 59
- 2.2 Experimental 61
  - 2.2.1 Reductive Methylation in the Presence of 18C6 61
  - 2.2.2 Fully ^{13}C-labeled Control Sample 61
  - 2.2.3 Partial ^{13}C-labeling 62
  - 2.2.4 Excess Natural Abundance Formaldehyde to Complete Methylation 62
  - 2.2.5 Reductive Methylation Using Multiple Reducing Agents 62
  - 2.2.6 Preparation for NMR 63

2.2.7 NMR 63
2.2.8 ^{13}C Percent Incorporation Calculation from NMR Data 63
2.2.9 Non-destructive Edman degradation 64
2.3 Results and Discussion 64
2.3.1 Reductive Methylation in the Presence of 18-Crown-6-Ether 64
2.3.2 Reductive Methylation Using Multiple Reducing Agents....67
2.3.3 Non-destructive Edman Degradation 69
2.4 Conclusion 71
2.5 References....71

CHAPTER 3 METHOD TO IDENTIFY THE NMR RESONANCES OF THE ^{13}C-DIMETHYL *N*-TERMINAL LYSINE ON REDUCTIVELY METHYLATED PROTEIN....75
3.1 Introduction....75
3.2 Experimental....78
3.2.1 Materials and Methods 78
3.2.2 Reductive Methylation Lysozyme (Generic Procedure) 78
3.2.3 MALDI-TOF Sample Preparation....79
3.2.4 MALDI-TOF Analysis 80
3.2.5 Determination of the Percent ^{13}C Labelling at Each Site: NMR 80
3.2.6 Mass Spectrometry 81
3.3 Results....82
3.3.1 Reductive Methylation and pH....82
3.4 Discussion....85
3.5 Conclusions....85
3.6 Acknowledgements....86
3.7 References....86

CHAPTER 4 SYNTHESIS AND APPLICATION OF PROTEIN MODIFIER AS A GLYCOSYLATION MIMETIC AND CROSS-LINKING AGENT....89
4.1 Introduction....89
4.2 Materials and Methods....91
4.2.1 Experimental....91
4.2.2 Coupling the Protein Modifier to Lysozyme....91
4.2.3 Synthesis of 1,2-O-isopropylidene-α-D-glucofuranose-5-carbaldehyde....92
4.3 Results and Discussion 94
4.3.1 Synthesis of 1,2-O-isopropylidene-3-deoxy-5-pentafuranal....94
4.3.2 Coupling the Modifier to Lysozyme....95
4.3.3 Glycosylation Mimetic 96
4.3.4 Protein Cross-linking....98
4.4 Conclusions....100

4.5 References..101
4.6 Spectra..105

CHAPTER 5 CONCLUSIONS AND FUTURE STUDIES..113
5.1 Concluding Remarks..113
5.2 Future Studies..114

APPENDIX: LETTERS OF PERMISSION..116

VITA 119

LIST OF TABLES

Table 1.1. Acid dissociation constants (pK_a), basic constituents, and isoelectric points (pI) of free amino acids at 25°C. Horton, H. R.; Moran, L. A.; Scrimgeour, K. G.; Perry, M. D.; Rawn, J. D., *Principles of Biochemistry*. 4th ed.; Pearson Prentiss-Hall: Saddle River, New Jersey, 2006. 5

Table 1.2. Assignments of dimethylamino NMR peaks to the lysyl residue number and N-terminal amine of reductively methylated hen egg white lysozyme with peak numbers corresponding to those in Figure 3. 43

LIST OF SCHEMES

Scheme 2.1. The reductive methylation reaction: In the presence of formaldehyde and a reducing agent, the primary amine is reductively methylated to produce monomethylamine. In the presence of excess formaldehyde, the monomethylamine undergoes a second reductive methylation to produce dimethylamine. 60

Scheme 2.2. Reductive methylation in the presence of 18C6.. 65

Scheme 3.1. Reductive methylation reaction.. 75

Scheme 4.1. Reagents and conditions: (a) NaH, CS_2, MeI, anhydrous THF, quantitative yield; (b) Bu_3SnH, toluene, reflux, 65%; (c) 0.8% aqueous H_2SO_4, MeOH, rt, 83%; (d) $NaIO_4$, EtOH, H_2O, 40% 95

Scheme 4.2. Schematic of the second alkylation reaction for cross-linking and intramolecular reaction. 98

LIST OF FIGURES

Figure 1.1. Diagram of an amino acid 1

Figure 1.2. Mirror image pair of L-Cysteine and D-Cysteine. The two are not identical; they cannot be superimposed. 2

Figure 1.3. The 20 common amino acids group according to side chain (green) functionality.. 3

Figure 1.4. The four classifications of protein structure. (Adapted from http://en.wikipedia.org/wiki/Protein_ structure#mediaviewer/File:Main_protein_structure_levels_en.svg).. 7

Figure 1.5. Flow chart outlining experiments for chemical modification of proteins. (Adapted from Matthews, et al. [54] with permission from Elsevier)... 19

Figure 1.6. Reductive methylation reaction. Reproduced from Roberson *et al*., *All Res. J. Chem*, 2013, 4, 10-16) 27

Figure 1.7. Comparison of the α-carbon positions for native (open circles) and reductively methylated (solid circles) hen egg white lysozyme. Reproduced from (Rypniewski, W.; Holden, H.; Rayment, I., Structural consequences of reductive methylation of lysine residues in hen egg-white lysozyme - an x-ray analysis at 1.8-angstrom resolution. *Biochemistry* **1993,** *32* (37), 9851-9858.) with permission from the publisher.. 29

Figure 2.1. (Left) Normalized NMR ^{13}C percent incorporation of each reaction site of partially-labeled lysozyme at 1:5 ratio (moles of reactive amine: moles of formaldehyde) in the presence of 18C6 and sodium ion. (Right) Normalized NMR ^{13}C percent incorporation of each reaction site of partially-labeled lysozyme at 1:5 ratio (reactive amine: formaldehyde) in the presence of excess 18C6 over sodium and lithium ions. 66

Figure 2.2. (a-e) Using the crystal structure of reductively methylated lysozyme by Rypniewski, et. al (PDB ID: 132L),[8] the dimethylamino groups (orange) and hydrophobic residues (yellow) are highlighted. (f) The structures of the borane reducing agents. 67

Figure 2.3. ^{13}C percent incorporation of partially ^{13}C- reductively methylated lysozyme with different reducing agents normalized to fully ^{13}C-methylated lysozyme with DMAB. 68

Figure 3.1. 1D HSQC NMR spectra obtained at 25°C and pH 8.5 of reductively ^{13}C-methylated lysozyme reacted at (a) pH 6.0, (b) pH 7.5, and (c) pH 10.0. 82

Figure 3.2. NMR spectra of lysozyme reductively ^{13}C-methylated with 7 moles of reactive amine to 2 moles of ^{13}C-formaldehyde. (a) 2D ^{1}H-^{13}C HSQC and (b) 1D ^{1}H-^{13}C HSQC. 83

Figure 3.3. MALDI mass spectrometry isotopic profile of peptide 634 from lysozyme reductive methylated with (a) natural abundance formaldehyde, (b) sub-stoichiometric ^{13}C-formaldehyde followed by natural abundance formaldehyde, and (c) ^{13}C-formaldehyde. 84

Figure 4.1. MALDI-TOF analysis of a) protein modifier coupled to hen egg white lysozyme and b) hen egg white lysozyme cross-linked with protein modifier........ 96

Figure 4.2. Overlay of 2D ^{1}H-^{13}C HSQC spectra of (red) native lysozyme and (green) lysozyme coupled to the protein modifier. 97

Figure 4.3. 2D ^{1}H-^{13}C HSQC of modified lysozyme after removal of the acetonide protecting group. 99

Figure 4.4. 1D gel protein profile of unmodified lysozyme (lane 4), modified lysozyme after the second reductive alkylation reaction (lane 3), and the cross-linked lysozyme sample (lane 2)........ 100

ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is a proven technique for protein structure and dynamic studies, typically requiring the incorporation of stable magnetic isotopes to improve sensitivity and assign resonances. Degenerate levels of ^{13}C-incorporation have been the biggest obstacle for mass spectrometry-assisted assignment of ^{13}C-dimethylamine resonances in nuclear magnetic resonance spectroscopy (NMR). Reductive ^{13}C-methylation is an alternative labeling method for proteins not amenable to bacterial host overexpression. Because reductive ^{13}C-methylation adds sparse, isotopic labels, traditional methods of assigning the NMR signals are not applicable. The research presented in the first part of this dissertation explores several methods used to break the degeneracy in ^{13}C-labeling of lysozyme. To overcome the degeneracy in labeling lysozyme with the reductive methylation reaction, we investigated two methods: 1) reductive methylation in the presence of 18-crown-6-ether (18C6) and 2) reductive methylation using multiple reducing agents. To assign the α- and ε-dimethylamine resonances of the *N*-terminal lysine residue of lysozyme, a non-destructive Edman degradation method was explored. The second part of this research discusses an alternative assignment method based on mass spectrometry to aid in the assignment of the NMR signals from reductively ^{13}C-methylated proteins. Because assignment is increasingly difficult when lysine is the N-terminal residue of the protein, one method is described to identify the NMR resonance of the ^{13}C-methyls associated with both the N-terminal α-amine and the side chain ε-amine. The NMR signals of the N-terminal α-dimethylamine and the side chain ε-dimethylamine of hen egg white lysozyme Lys1 are identified in ^{1}H-^{13}C heteronuclear single-quantum correlation spectra. Protein chemical modification is a well-established field that continues to impact leading research today including glycomimetics and cross-linking of proteins. Current protein chemical modifications like polyethylene glycol are

proven useful for increasing the lifetime of several therapeutic enzymes but are also toxic to the body. In the last chapter, we present the use of sugar derivatives as a possible less toxic alternative for synthetic glycoproteins. The synthesis of a protein modifier is described and preliminary data of its application as a glycomimetic and cross-linking agent is presented.

CHAPTER 1 INTRODUCTION

1.1 Protein Structure and Stability

1.1.1 Protein Structure

Proteins are macromolecules assembled from the arrangement of covalently-linked amino acids. Twenty standard amino acids compose the structures of proteins and are critical to all living elements used throughout nature and the human body. Structurally, each amino acid contains four common features: 1) α-carbon, 2) an amine, 3) a carboxyl, and 4) a side chain or R group (Figure 1.1).

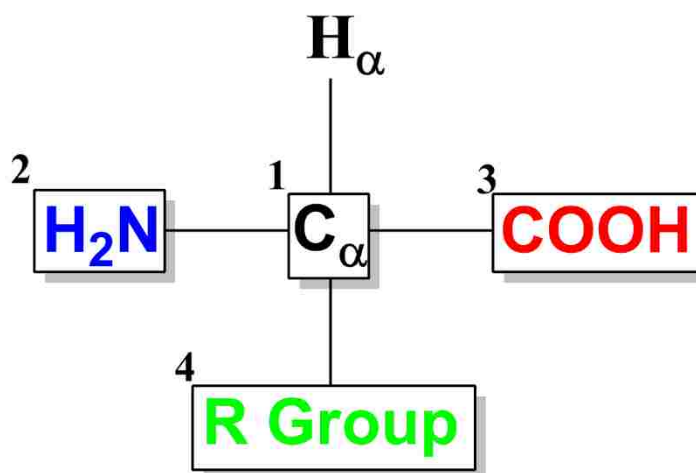


Figure 1.1. Diagram of an amino acid

Three components (i.e. amine, carboxyl, and side chains) are covalently bound to the α-carbon, which is the central atom for each amino acid, causing the amino acids to be asymmetrical, or chiral. Nineteen of the twenty common amino acids follow this pattern, giving them the ability to exist as two non-superimposable mirror forms, stereoisomers, in nature. The stereoisomers are designated D (*dextro*, right) or L (*levo*, left) for the direction they rotate the plane of a beam of polarized light (Figure 1.2). The amino acids found in protein and peptide structures are

synthesized by forming peptide bonds (or amide bonds) during condensation reactions between the α-carboxyl group and the α-amine group of adjacent amino acids. A peptide is a compound consisting of at least 2 amino acids bound by a peptide bond. Oligopeptides are peptides consisting of 10 or fewer amino acids. Polypeptides are peptides with more than 10 amino acids, and proteins are polypeptides with more than 50 amino acids. All proteins translated in the ribosome are constructed with L-amino acids, although, a few D-amino acids are found in proteins that undergo post-translational modifications.

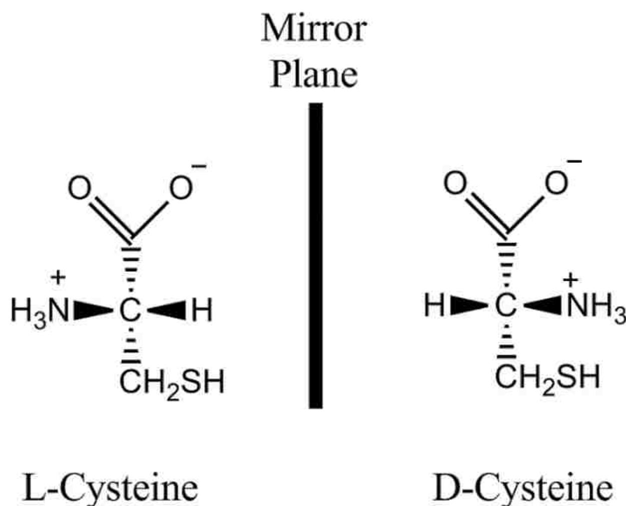
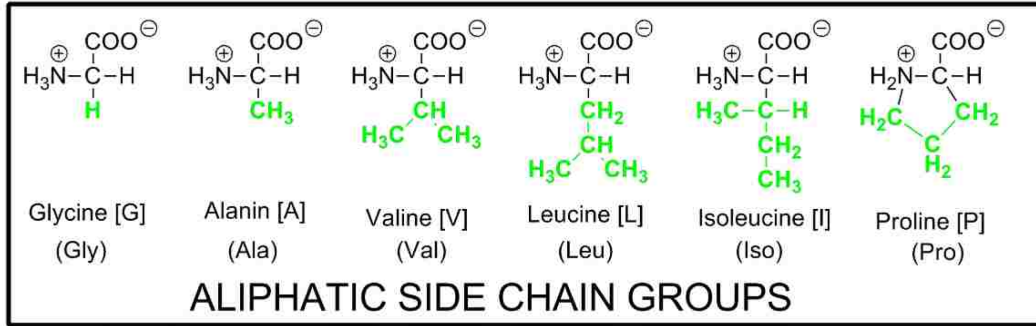
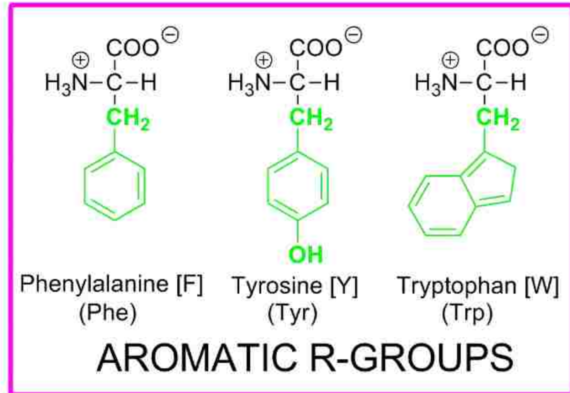
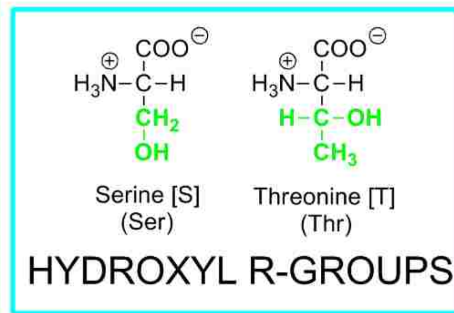


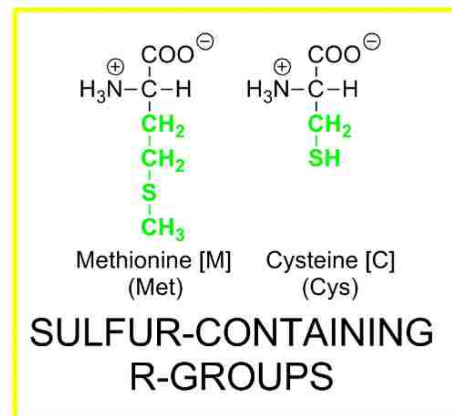
Figure 1.2. Mirror image pair of L-Cysteine and D-Cysteine. The two are not identical; they cannot be superimposed.

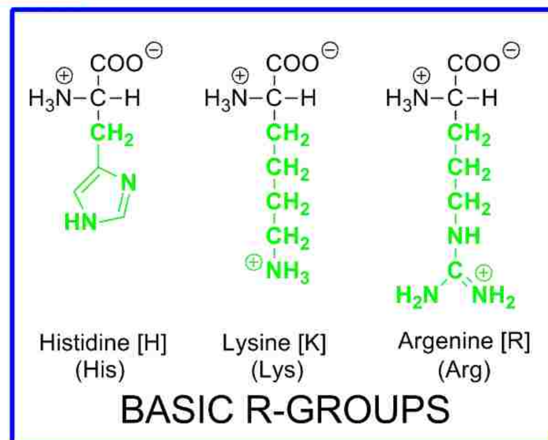
Amino acids range in charge, polarity, hydrophilicity, and hydrophobicity, which results in proteins with varying affinities to polar or nonpolar solvents (Figure 1.3). For example, when a protein is dissolved in aqueous solution at a particular pH, the charged terminal amine and carboxyl groups on a protein can react with either acids or bases. Not only does this feature make proteins ampholytes, it also classifies them as zwitterions since the side-chain moiety can exist as either neutral or as positively or negatively charged molecules. Amino acids that can be positively or negatively charge are lysine (+), arginine (+), aspartate (-), and glutamate (-).











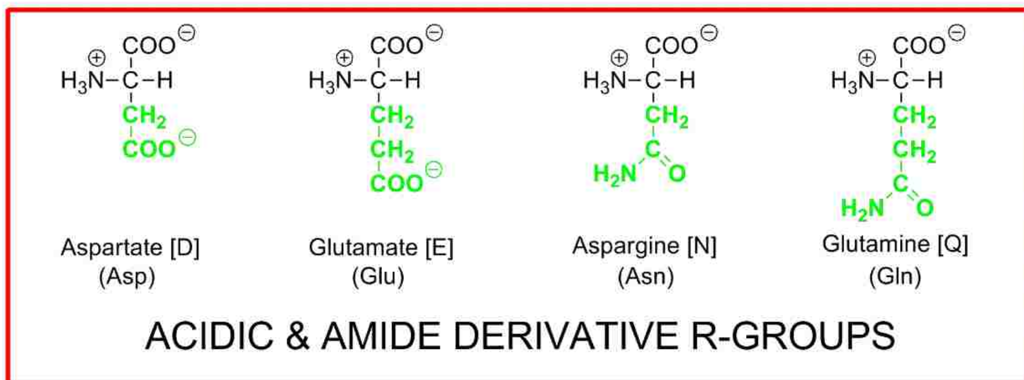


Figure 1.3. The 20 common amino acids group according to side chain (green) functionality.

Moreover, the functional groups within proteins allow it to act as an acid (a proton donor) or a base (a proton acceptor). When the sum of charges from the proton donors ($^{+}H_3N$-CHR-COOH) and proton acceptors ($^{+}H_3N$-CHR-COO^{-}) are equal, the net charge is zero. The pH of the aqueous media that results in a zero net charge is known as the isoelectric point (pI). The pH of a solution affects the protein's charge and aqueous solubility. Therefore, when proteins are suspended in media at a pH below its isoelectric point, the protein has a net positive charge; whereas, protein solutions with pH values above its isoelectric point results in a negative net charge. In either case, the protein is charged and dissolved in the aqueous media. However, at pH values within 1 deviation of the pI, proteins are typically unstable and will precipitate.

Acid-dissociation constants (pK_a) represent the pH-dependent characteristics of the functional groups (carboxyl, amino, and any ionizable side chain) within the protein (Table 1.1). Each ionizable functional group has a specific pK_a value that corresponds to a pH in which the concentration of the protonated form is equal to that of the unprotonated form. When the pH of the solution is below the pK_a of the ionizable group, the protonated species dominates and it is capable of acting as an acid by donating a proton. When the pH of the solution is above the pK_a, the unprotonated species dominates allowing the functional group to act as a proton acceptor. Each amino acid has at least two pK_a values associated with its α-amine and α-carboxyl groups, which can be determined by titration curves. Amino acids that do not have an ionizable functional group on its side chain exists as a zwitterion, a neutral molecule with both positive and negative charges, in solution at pH values between the pK_a values of the carboxyl and amine groups.

The chemical properties of proteins affect its propensity to be in contact with polar or nonpolar molecules. More particularly, the side chains govern the interactions the protein has with solvents. Polar and charged amino acids such as serine, threonine, asparagine, aspartate,

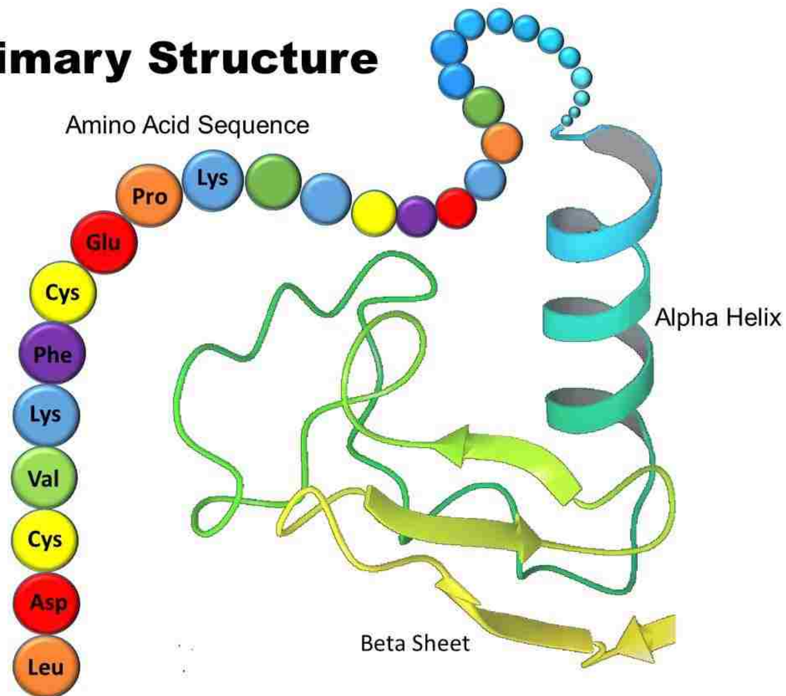
Table 1.1. Acid dissociation constants (pK_a), basic constituents, and isoelectric points (pI) of free amino acids at 25°C. Horton, H. R.; Moran, L. A.; Scrimgeour, K. G.; Perry, M. D.; Rawn, J. D., *Principles of Biochemistry*. 4th ed.; Pearson Prentiss-Hall: Saddle River, New Jersey, 2006.

Amino Acid	**Acid Dissociation Constant (pKa)**			**Isoelectric Point (pI)**
	Carboxyl Group	Amino Group	Side Chain	
Glycine	2.34	9.60	-	5.97
Alanine	2.34	9.69	-	6.00
Valine	2.32	9.62	-	5.96
Leucine	2.36	9.60	-	5.98
Isoleucine	2.36	9.60	-	6.02
Methionine	2.28	9.21	-	5.74
Proline	1.99	10.60	-	6.30
Phenylalanine	1.83	9.13	-	5.48
Tryptophan	2.83	9.39	-	5.89
Serine	2.21	9.15	-	5.58
Threonine	2.09	9.10	-	5.60
Cysteine	1.96	10.28	8.18	5.07
Tyrosine	2.20	9.11	10.07	5.66
Asparagine	2.02	8.80	-	5.41
Glutamine	2.17	9.13	-	5.65
Aspartic Acid	1.88	9.60	3.65	2.77
Glutamic Acid	2.19	9.67	4.25	3.22
Lysine	2.18	8.95	10.5	5.98
Arginine	2.17	9.04	12.48	10.76
Histidine	1.82	9.17	6.00	7.59

glutamine, glutamate, histidine, lysine, arginine, cysteine, and tyrosine result in proteins that have energetically favorable interactions with polar solvents like water. These interactions are due to the hydrogen bonding capabilities of polar amino acids. The hydrophobic amino acids (e.g. alanine, valine, leucine, isoleucine, proline, phenylalanine, and tryptophan) are typically inaccessible to polar solvents and have a lower propensity to interact with water. Although protein structure and classification will be discussed in the next section, these amino acids control protein structure in aqueous Solvents and facilitate protein solubility in nonpolar solvents.

1.1.2 Classification of Protein Structure

The structural components of each protein have a critical role in its size, conformation, stability, and function. In particular, the assembly of amino acids, amino acid sequence, and their impact in higher-order protein structure and function were first investigated by William Atsbury (1933),[1] Walter Kauzmann (1946),[2] Kaj Linderstrom-Lang (1949),[3] Linus Pauling *et al.* (1951),[4] and Christian Anfinsen (1954).[5] Protein structures are generally classified into four groups: primary (1°), secondary (2°), tertiary (3°), and quaternary (4°). The linear sequence of amino acids results in the primary structure of a protein. Conventionally, the order of amino acids is interpreted from the N-terminus to the C-terminus, or the terminal amine to the terminal carboxyl group of the protein, respectively. Intra-molecular forces (i.e. hydrogen bonding, Van der Waals, and dipole-dipole) among amino acid residues result in the formation of secondary structure. Examples of secondary structures are alpha helix, beta sheet, random coil, turns, and loops. Studies premised from the original work of Pauling (1951) have demonstrated that protein structure (e.g. α-helix and β-sheet) and subsequent folding are preferentially formed from the properties of amino acids in aqueous solution.[4] As a result, α-helices and β-sheets (Figure 1.4) are the two most common, stable secondary structures that exist. Tertiary structure is the three-dimensional shape formed mainly by hydrophobic interactions. In aqueous solutions, the hydrophobic amino acids orient themselves away from the water resulting in higher order folding. As the protein chain folds, this situates some nonpolar amino acids in the core of the protein in closer proximity to others allowing disulfide and salt bridges to be formed and stabilize the structure.[3] The accumulation of intramolecular forces results in the protein's native conformation, which is the state that a protein is most active and functional. Quaternary structure results from the intermolecular assembly of lower-ordered protein structures into larger complexes, called



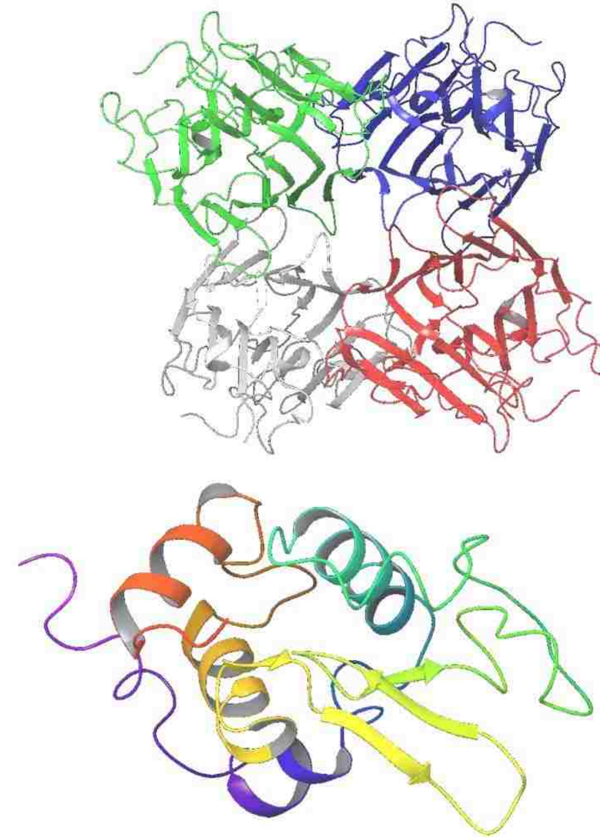


Figure 1.4. The four classifications of protein structure. (Adapted from http://en.wikipedia.org/wiki/Protein_structure#mediaviewer /File:Main_protein_structure_levels_en.svg)

multimers (i.e. dimer, trimer, tetramer, etc.). Similar to tertiary structures, quaternary structures are stabilized by non-covalent interactions. Many known functional biological proteins exist as quaternary proteins (i.e. hemoglobin, viral capsids, ion channels, DNA polymerase, etc.). Changes to protein structure or conformation will be discussed in the next section.

1.1.3 Protein Denaturation

Protein denaturation or unfolding is the process in which the secondary and tertiary structures of a protein, in its native (folded) state, become unfolded. Since the native conformation is usually the most water soluble, denatured proteins typically exhibit changes in solubility. Several approaches have been used to denature proteins.[6-7] For most proteins in solution, the equilibrium between the native conformation and the unfolded state is modulated by pH, ionic strength, and temperature of the aqueous surroundings or the addition of co-solvent or chaotropic agents.[7] The unfolded state is a result of losses in secondary and tertiary structure by disrupting intramolecular interactions. In particular, amino acids bound by disulfide and salt bridges, polar side chains and solvent, and hydrogen-bonded amino acids are affected.[8] It must be noted that this process is not strong enough to cleave peptide bonds; as such, the primary structure remains unchanged.

Heat is used extensively to reversibly (temporarily) and irreversibly (permanently) denature proteins. Exposing proteins to increasing temperatures weakens the protein structure by first affecting the long-range interactions in the tertiary structure. Depending on the proteins' thermal stability, heat increases the kinetic energy of the macromolecule. When the kinetic energy surpasses the hydrogen bond and van der Waals forces, increased vibrations and long-range flexibility permit water to interact with the nonpolar regions of the protein, typically through the backbone carbonyl and amine groups. The effect of water on the hydrophobic regions of the

protein results in unfolding, at the cost of the protein minimizing its free energy and exposing as many polar amino acids to offset the high entropy caused by the exposed hydrophobic residues. Since both short and long-range interactions are disturbed, the protein may or may not return to its native, functional conformation. In the case of the former, cooling allows the protein to return to its native state. However, in irreversibly heat-denatured proteins, the kinetic barriers required to return to the native conformation must be overcome, typically in a sequential fashion to reach the proper conformation. Calorimetric methods are the premier approaches to determine the thermodynamic transitions that proteins undergo when heated.[9] Lyubarev and Kurganov extensively covered the use of calorimetry and other approaches to understand thermal denaturation of proteins.[10] In particular, if the original thermal character is observed after reheating the sample then the protein is considered reversible; whereas, irreproducible thermal transitions observed in calorimetric curves indicate an irreversible denaturation induced by heat.

Acids and bases are also used to disrupt bonds between amino acids in the tertiary structure. As alluded to earlier in Section 1.1, pH is critical to the stability and structure of proteins since it affects its charge distribution and intramolecular forces. This type of denaturation is mostly attributed to the distortion of charge balance on the protein.[7,11] In water, proteins fold with the aggregation of the hydrophobic residues internally, exposing the polar residues to interact with the solvent. During this process, salt bridges form between acidic and amine side chain ionic pairs in close proximity. These ionic parings are stronger and have a bigger role in stabilizing the protein structure than the polar side chain interactions with the solvent. In extreme acidic conditions, the acidic side chain of the salt bridge is neutralized and the protonated amine associates to the anion of the acid. This process creates an overabundance of positive charges, leading to repulsion of the

charged sites and the structure of the protein is destroyed. The effects of base are analogous to the use of acid but causes an excess of negative charges.

Alcohols and other organic solvents, with dielectric constants lower than water, are also known to denature proteins.[12-14] Not only do these solvents increase the strength of the electrostatic interactions between the water molecules and the charged residues in the protein (increased organic solvent results in a smaller hydration sphere around the protein), but the hydrophobic bonds governing the tertiary protein structure weaken allowing for greater interactions between the organic solvent and protein. In this case, the degree of folding depends largely on the properties of the primary structure and the concentration of organic solvent introduced into the system. For example, large amounts of organic solvent would typically result in further denaturation as compared to larger amounts of aqueous solvent. This effect was evident in the investigation reported by Voets *et al.* (2010), in which smaller fractions of DMSO in water were required to maintain lysozyme stability.[15] At concentrations exceeding 70% DMSO, lysozyme became unfolded and collapsed and was attributed to the impact of the protein microenvironment at varying concentrations of DMSO, since it strongly disrupts hydrogen bonds with the water and polar amino acids. Ultimately, the apolar components of the protein became unfolded and better dissolved in the solvent mixture. In conditions in which solvent exchange occurs slowly, the protein would be able to maintain stability as it adjusts to the conditions in which it is exposed. However, this phenomenon is not accurate for every protein and solvent mixture studied. Griebenow and Klibanov (1996) reported that an organic and aqueous solvent mixture was actually detrimental to the secondary structures of lysozyme and subtilisin.[16] In their study, the presence of polar, pure organic (anhydrous) solvent did not change the protein secondary structure; whereas organic-aqueous solvent mixtures resulted in losses of α-helical content.[16] However, due to the

limited solubility of proteins in organic solvents, they are typically not used as denaturants. As such, the incorporation of chaotropic agents is a preferred method to denature proteins when using chemical additives.

Chaotropic agents, such as urea and guanidine hydrochloride, are used to denature proteins by disrupting the hydrogen-bonding network between water molecules and weakening the hydrophobic interactions within the protein structure.[11,17] These agents specifically interfere with the stability of the hydrophobic regions of the protein by increasing the entropy of the system after exposing the hydrophobic amino acids to the surrounding aqueous environment. As a result, the hydrophobic amino acids become solubilized and the native conformation becomes destroyed. In addition to urea and guanidinium chloride, univalent salts are effective protein denaturants. Zhang and Cremer (2010) extensively reviewed the interactions of proteins with many osmolytes and anions that follow the Hofmeister series,[18-19] which has been known to be a useful guide to determine effective protein-dissociating chaotropes.[17,19-22] For example, Adebowale and Adebowale (2007) found that the solubility of *Mucuna pruriens* protein isolates increased when placed in chaotropic solutions that followed the Hoffmeister series.[23-24] In depth details regarding the molecular-level interactions of chaotropes and macromolecules using sodium dodecyl sulfate, guanidinium chloride, and urea, as example agents known to disrupt protein structure, can be found in reviews published by Alonso and Dill (1991) and Zhang and Cremer (2010).[25-26]

1.2 Protein Modification

Protein modification plays an essential role in the production of target proteins and peptides and the understanding of their biological functions and applications. Rooted in the last decade of genomics research, the journey from gene to protein revealed the complexity, diversity, and abundance of proteins that encompass the human proteome. While it is estimated that

approximately 25,000 genes comprises the human genome,[27] it is believed that the human proteome consists of over 1 million proteins.[28] Evidenced by studies showing that a single gene can encode multiple proteins, the complexity of understanding protein production, function, and application is further facilitated by protein post-translational modifications (PTMs).[29] To understand the occurrence of PTMs, the processes (i.e. transcription and translation) in which a gene is expressed to produce proteins will be briefly discussed.

1.2.1 Protein Synthesis

The first step of gene expression, which occurs in the cell nucleus, is known as transcription. Transcription is the genetic process in which RNA polymerase copies DNA into a complementary antiparallel RNA strand to form a transcription unit that encodes one gene. Encoded genes for protein synthesis use messenger RNA (mRNA) to convey genetic information out of the nucleus to the ribosomal complex in the cytoplasm where the second step of gene expression called translation occurs. Translation is a four-step process (i.e. initiation, elongation, translocation, and termination) in which clusters of three mRNA bases along its sequence are translated as codons to link individual amino acids. Transfer RNA (tRNA) carries the amino acids as directed by the codons to the ribosome, which assembles the protein until it encounters three bases that code for protein assembly to stop. Post-translational modifications are events in which the amino acids produced within the ribosome are enzymatically converted into a non-standard amino acid. Typically, PTMs include, but are not limited to, proteolysis, acetylation, lipidation, glycosylation, ubiquitination, phosphorylation, and methylation. Because of this, the twenty amino acids commonly observed in proteins have now been extended to more than 400.

1.2.2 Post-translational Modification

Post-translational modification of proteins can be summarized into three major categories: 1) modifications that involve peptide bond cleavage and formation, 2) modifications that involve amino (N-) and carboxyl (C-) terminal amino acids, and 3) modifications that involve specific amino acid side chain moieties.[30-31] Peptide bond cleavage, or proteolysis, is the most common protein modification. Typically, large precursor proteins or polyprotein chains are assembled in the ribosomal complex and later cleaved by various enzymes to form their final active structures. Since many proteins have been found to begin with an N-terminal amino acid other than L-methionine, which is usually the first amino acid because its codon is the start codon for protein synthesis, it is believed that peptide cleavage occurs soon after protein synthesis.[32] Peptide cleavage directly affects protein folding by exposing specific structural domains of proteins that dominate their tertiary structures and, if enzymes, their ability to interact with a substrate.[33] Therefore, the consequence of cleaving specific polypeptide bonds is the production of a new, diverse set of proteins.

Of the several PTMs, acetylation is the most common irreversible and reversible modification to occur at the N-terminal amino acid. N-terminal acetylation is a chemical reaction that occurs in more than 80% of human and bacterial proteins in which an acetyl functional group is used to replace L-methionine after it has been cleaved by methionine-aminopeptidase (MAP). Although this type of reaction is classified as a PTM, it can occur while the protein is still undergoing synthesis in the ribosomal complex, or co-translationally.[34] In many instances, N-acetylated proteins have L-alanine, glycine, or L-serine as N-terminal residues. However, proteins are typically acetylated on L-lysine residues to regulate gene transcription, although many cytoplasmic proteins have been reported to undergo acetylation.[34] The exact biological

significance of N-acetylation is unclear, but is postulated to protect the protein from cleavage by N-terminal peptidases.[31]

Other N-terminal addition reactions involve lipidation and glycosylation. Lipidation is a method in which proteins are modified by the addition of fatty acids so they can be taken in by organelles in the plasma membrane.[35] There are four types of lipidation PTMs, but only N-myristoylation is specific to the N-terminus. This type of lipidation functionalizes an amino acid with a myristoyl group to change the localized hydrophobicity of the protein and drive its affinity and ability to dock the protein to the endoplasmic membranes.[34,36] N-myristoylation is initiated by N-myristoyltransferase (NMT) using myristoyl-CoA as the substrate to attach the myristoyl group to the N-terminal glycine. This PTM also requires that L-methionine be cleaved using MAP before the myristoyl group can be added.

There are several types of C-terminal modifications, but ubiquitination is the most common PTM. Ubiquitination is a three-step, cascading enzymatic modification in which one or more ubiquitin polypeptides are covalently attached to side chain amino groups of lysine through the C-terminal glycine of ubiquitin.[37-38] Ubiquitin, a 76 amino acid protein consisting of seven lysine residues, plays a key role in cellular processes.[39] The roles of each lysine residue in this PTM have yet to be elucidated, but many studies have concluded that K48 and K63 primarily regulate protein degradation and coordinate protein processes, respectively.[40-41]

All other PTMs, like glycosylation, phosphorylation, and methylation, also result in the addition of small molecules to the protein. Glycosylation, which can occur during co- or post-translational modification, are a series of reactions in which carbohydrate groups are attached to protein moieties.[11] Specifically, N-linked glycosylation involves the covalent attachment of a complex, branched oligosaccharide consisting of N-acetylglucosamine (GlcNAc), mannose,

different sugars, and a terminal sialic acid to the side chain of an L-asparagine residue.[11] The synthesis of N-linked oligosaccharides for plants, animals, and single-celled eukaryotes begins in the rough endoplasmic reticulum with the basic scaffold: three glucose, nine mannose, and two GlcNAc molecules. These oligosaccharides help to direct the newly formed glycoprotein to specific locations in the cell or help with stability and folding of proteins.[42] When on the surface of eukaryotic cells, glycoproteins facilitate cell-cell attachment. Overall, glycosylation reactions contribute greatly (>50% of human proteins are glycosylated[43]) to the variety of proteins in the human proteome because all four components unique to glycosylation (i.e., glycosidic bond, glycan composition, branching structure of the glycan, and the length of the oligosaccharides attached to the protein) can be altered.[42-43] Phosphorylation is an enzymatic modification in which a phosphoryl group is added to the hydroxyl side of L-serine, L-tyrosine, or L-threonine residues transferring the regional hydrophobicity of a protein to one of greater polarity and hydrophilicity.[44-45] Also catalyzed by enzymes, methylation usually incorporates one or more methyl groups to an L-lysine or L-arginine residue.[46] Like all other PTMs, incorporation of either phosphoryl or methyl groups onto a protein helps to govern cellular regulatory signaling.[39]

1.2.3 Recombinant Protein Expression

Proteins are the foundation for virtually every cellular activity and knowledge of their roles in cell processes is essential to understanding biological systems. The vast diversity of the human proteome necessitates the ability to acquire proteins in their native conformation, at usable quantities and lower costs, and ideally from the original source. Substantial progress in genetic engineering has allowed proteins to be expressed from recombinant DNA in both prokaryotic and eukaryotic systems for application-driven structural and activity studies.

Recombinant protein expression is the ability to exchange or transfer unmodified DNA from one host to another to produce proteins. There are several attractive reasons for which recombinant technology is used: 1) proteins can be produced in organisms that have fast growth rates and are easy to cultivate;[47] 2) promoters can be used to potentiate the yields of recombinant protein by allowing the target protein to be overexpressed relative to its other cellular components;[48] and 3) translational fusion can be incorporated to facilitate protein purification.[49-50] Ultimately, recombinant protein expression provides a medium to overcome low yields of newly discovered proteins so that they can undergo sequencing, biochemical, and structural studies.

The host used to express recombinant proteins for structural analysis is usually *Escherichia coli* because altering its genetic information is relatively easy and inexpensive.[47] Extensive use of *E. coli* in bacterial genetics has resulted in a multitude of systems available to facilitate its utility in recombinant protein expression.[51] Recombinant expression to obtain high yields of purified proteins usually occurs when the DNA sequence of a target protein is cloned into a vector and subsequently transformed into the host cell. The bacterium uses glucose and ammonium chloride as carbon and nitrogen sources, respectively, to produce the necessary amino acids to express the proteins encoded in the vector.[11] Although the impact of this technology has been extended to the development of therapeutics and diagnostic kits, among many others, this aspect of recombinant protein expression usually occurs after the protein has been fully characterized. Recombinant protein expression can also be used to incorporate labels in the protein that allow for complete characterization of its structure and interactions with ligands and other proteins to better understand its function.[29,47,50] Labeling can be achieved by feeding the bacterial host labeled carbon and nitrogen so that amino acids with the necessary isotopes can be expressed. For

example, if *E. coli* is fed ^{13}C-glucose and ^{15}N-ammonium chloride, then the target protein is fully expressed with ^{13}C and ^{15}N labels and suitable for structurally characterization by NMR.

Despite its many advantages, recombinant expression in bacteria is not amenable to eukaryotic proteins that require PTM.[51] Since bacteria lack the ability to produce eukaryotic PTMs (i.e. glycosylation) and do not contain chaperone proteins (disulfide bonds), there is the likelihood that improperly folded proteins would be expressed.[47,51] Additionally, recombinant expression suffers from incorrectly degraded, rearranged, or poorly expressed proteins when codon usage differs between source and host ribosomal complexes.[47] These features critically limit the feasible use of protein-based applications, especially when these aberrations affect its industrial production, safety, and efficacy in therapeutic applications.[51-53] As a result, other bacteria and eukaryotic cells have also been sourced as hosts for this technology, since it is a rate-limiting step in protein structural studies and therapeutic applications.[52] To date, no single host exists that allows all proteins to be expressed in an active form.[47]

With over 300 different types of PTMs, the applicability of incorporating isotopic labels via recombinant expression for structural studies of proteins, specifically of unknown structure or sequence, is limited.[34] It is possible to express these proteins in eukaryotic hosts. However, since eukaryotic cells cannot make all amino acids for protein synthesis like prokaryotic cells, uniform isotopic labeling is prohibitively expensive. Thus, chemical techniques that enable modification of proteins are highly attractive tools that allow the roles of amino acids to be evaluated in protein structural and activity studies.

1.2.4 Chemical Modification

Chemical modifications have been essential to protein structure and activity studies. Like PTMs, chemical derivatization takes advantage of the chemical reactivity of the amino acid side chains to

achieve specific functional outcomes after proteins have been isolated and purified. In fact, an understanding of organic chemistry and the reactivity of amino acids has allowed the foundation for non-enzymatic and simulated post-translational modifications to be conducted using chemical means. For example, chemistry of amino acids have sought to preserve or alter electrostatic charge, increase or decrease hydrophobicity, denature or renature proteins, or incorporate or remove specific chemical moieties with the intention to study their role in protein tertiary structure and activity.

Figure 1.5 outlines the experimental approach generally used in the chemical modification of proteins and subsequent validation studies. In summary, targeted amino acids in the protein backbone are matched with chemical reagents under optimal reaction conditions. After the protein is reacted, excess reagent is removed and the protein is prepared for biochemical studies. Then, the success of the reaction, site and number of residues modified, and its effects on protein activity are studied using various assays and elucidative techniques. Depending upon findings, modifications to the existing experiment are made and repeated to achieve desirable outcomes. Earlier attempts of protein modification were once limited by the lack of sensitive methods, accuracy in controlling the number and type of amino acids undergoing reaction, and denaturation of the protein. Yet, the introduction of specialized reagents and advanced methods has permitted site-selective chemical reactions of amino acids with better sensitivity and accuracy than before.

Site-selective chemical modification is a process in which a single amino acid residue is covalently derivatized without modifying a second, separate amino acid or disrupting the protein's conformation. Examples of amino acid functional groups, mainly sulfhydryl residues (cysteine), phenols (tyrosine), aromatic heterocycles (tryptophan and histidine), and amines (arginine and lysine) are detailed below.

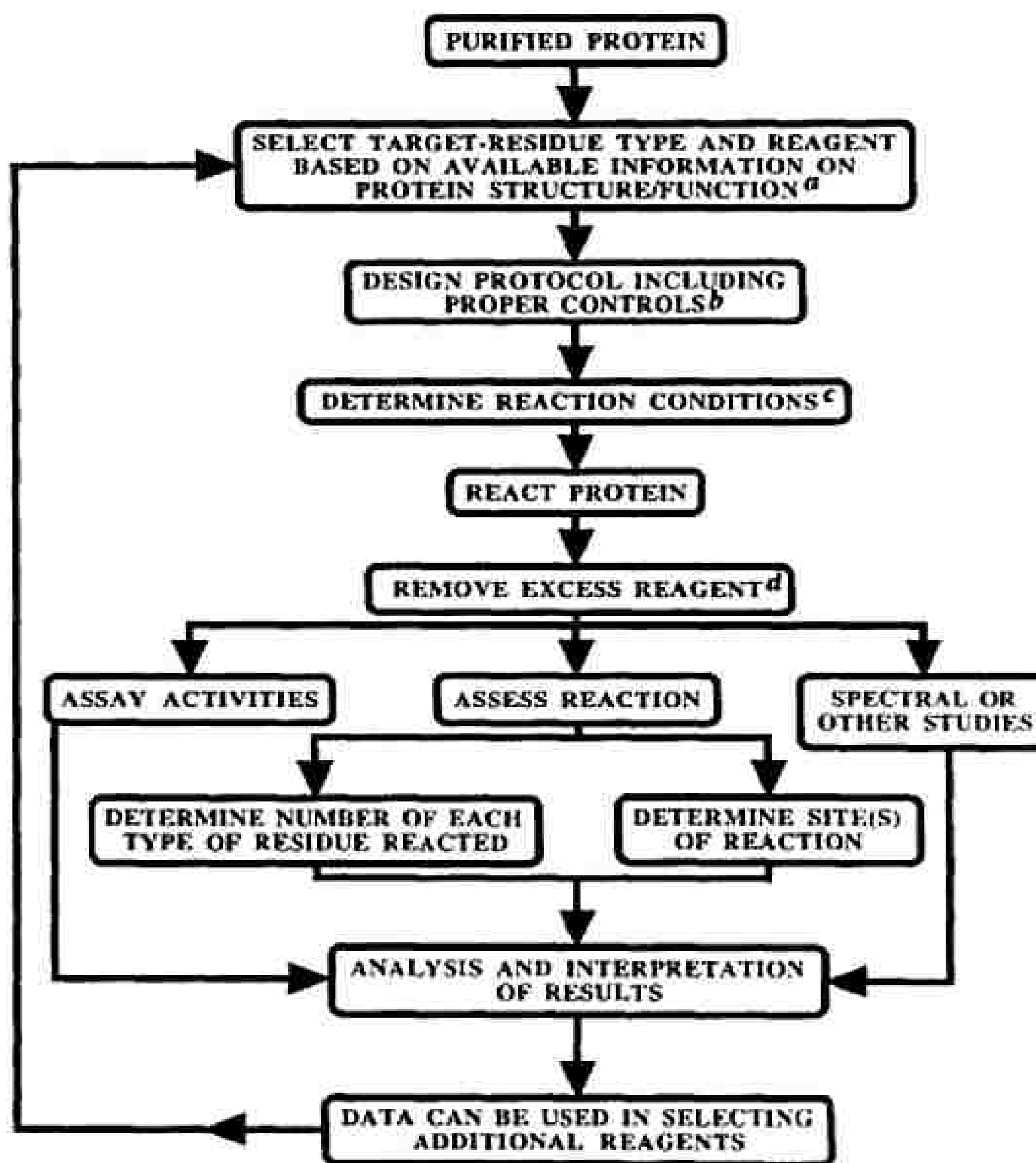


Figure 1.5. Flow chart outlining experiments for chemical modification of proteins. (Adapted from Matthews, et al. [54] with permission from Elsevier).

1.2.5 Side Chain Selective Modifications

1.2.5.1 Sulfhydryl Residues

Sulfhydryl residues, also known as thiols or mercaptans, are carbon-bonded molecules consisting of a sulfur bonded to a hydrogen (–S-H group).[55] Because oxygen and sulfur are in group VIA in the periodic table, there are similarities in their reactivities, however, there exist differences in the chemical properties of the two because they are on different rows. For example, oxygen is

relatively more electronegative than sulfur, but thiols (-S-H) are more acidic than alcohols (-O-H). The general reaction to prepare thiols results from an S_N2 reaction between sodium hydrosulfide and an alkyl halide.[55] Since the thiol product is very nucleophilic, if excess hydrosulfide is used, then it will be alkylated again to give a sulfide product (R-S-R).

As major component of the amino acid cysteine, these alcohol analogues are one of the most reactive amino acid functionalities observed in proteins.[11] In fact, thiols are easier to oxidize into a dimer than alcohols. For example, when two cysteine amino acids are in close proximity, they readily form a covalent disulfide bond in which the –S-H group is changed into an –S-S- group and called cystine.[11] The oxidation of cysteine residues to form disulfide bonds is the most common naturally occurring post-translational/ chemical modification of proteins, which is critical for protein structure and stability. Also, disulfide linkages can be cleaved to return cystine back to two separate cysteine amino acids.

Because the sulfhydryl side chain is very reactive, a wide variety of reagents are available for modification of cysteine residues. Most of these reagents have reactivities with other side chains. However, careful consideration of reaction conditions can afford the necessary selectivity of most reactions. Early studies of cysteine residues used direct alkylation with α-iodocarbonyls at varying pH values.[56-57] Derivatization of these alkylating reagents increased the applicability of this method. For example, in the lactose repressor protein, none of the three cysteine residues is reactive iodoacetic acid or iodoacetamide, but two of them are modified with 2-bromoacetamido-4-nitrophenol (BNP).[58] Understanding disulfide linkages in PTMs was initiated from the observation of spontaneous oxidation of cysteine residues by molecular oxygen. Although too slow to account for rapid rates of disulfide bond formation in the cell, this observation led to the discovery of protein disulfide isomerase, a eukaryotic cell catalyst responsible for this activity. Oxidation of cysteine

side chains has been used to form disulfide bridges in proteins and to functionalize cysteine residues with small molecules.[59] However, these simple reactions are challenged by long reaction times and an inability to control the reaction product, where either mixed disulfides or dimers would be formed. One alternative reaction that allows for control of the reaction product is to use Ellman's Reagent (5,5-dithobis(2-nitrobenzoate); DTNB) which has historically been used to quantify sulfhydryl content in proteins through disulfide exchange reactions.[60] Sulfenyl halides, particularly methanethiosulfonates, have improved the reaction time, reduced the reagent quantities, and increased the selectivity compared to conventional disulfide formation reactions.[54,61]

There are several elements required for the execution of these reactions. In order for disulfide bridges to be formed between thiol groups, the proximity between cysteine residues must be favorable and not hindered by neighboring, large side chains. Also, reaction conditions must be favorable in that the cysteine residue can be independently modified without inducing protein denaturation.[62] Avoidance of denaturation can be a challenge since fundamental studies have revealed that most sulfhydryl groups within the protein are submerged in hydrophobic locations, in which functionalization most often produces an altered protein structure.[62] Disulfides may also present some instability in reducing environments thereby limiting their *in vivo* application with high intracellular glutathione concentrations.[56] Because of this, the feasibility of known thiol-selective chemical modifiers and subsequent analyses of modified proteins were limited. Recent methods have approached these challenges by 1) desulfurization of the cysteine residue, 2) oxidative elimination of cysteine, or 3) metal-mediated modification of cysteine and its derivatives.[56] Understanding sulfhydryl reactions and its applicability to cysteine and proteins, have promoted critical insight into chemistries needed to further explore protein structural studies.

1.2.5.2 Phenols

Some of the most widely observed compounds in nature and industry are alcohols, or organic compounds that contain hydroxyl (-O-H) groups in their structure. Structurally, alcohols are similar to water in that they both have sp^3 hybridized oxygen atoms, but the bond angle in alcohols are larger than the bond angle in water.[55] The bond lengths of the hydroxyl moieties are approximately the same. Alcohols differ by the type of carbinol carbon (i.e., carbon conjugated to the hydroxyl group) and, depending on its number of functional groups attached, are classified as primary, secondary, or tertiary alcohols.[55] Compounds that have a hydroxyl group bonded directly to an aromatic ring are called phenols. Phenols have properties similar to the aromatic and alcohol component of its structure. As a result, alcohols have variable solubility in polar and nonpolar solvents. Generally, the hydroxyl proton of alcohols is weakly acidic and can be removed easily in the presence of a strong base.[55] However, the acid dissociation constants of alcohols vary depending on the order of bonding (i.e., primary, secondary, or tertiary) and their subsequent reactions.[55]

In proteins, only three amino acids contain hydroxyl groups on their side chains, namely serine, threonine, and tyrosine.[11] Serine, a primary alcohol, is very important in biosynthesis and enzyme catalysis reactions. Threonine is an essential amino acid with a secondary alcohol that participates in numerous PTM reactions. Tyrosine is the only standard amino acid that has a phenolic functional group in its side chain, and it participates in signal transduction processes and photosynthesis reactions. Because the hydroxyl group on tyrosine has a lower pK_a (pK_a ~10) than serine or threonine (pK_a ~13), it is much more reactive.[63] None of the hydroxyl groups on serine, threonine, or tyrosine residues are ionizable.

Chemical modification of serine or threonine groups is not a trivial task, because neither

amino acid is a strong nucleophile. However, their nucleophilicity is greatly improved when spatially located near a histidine residue or the hydroxyl group is transformed into an alkoxide ion.[63] Several literature reports have discussed the use of N-hydroxysuccinimide (NHS) derivatization of serine, but since it is dependent on the nature of adjacent amino acids, it has lost popularity.[63] Attempts to oxidize serine and threonine with carbodiimide in the presence of dimethylsufoxide and phosphoric acid resulted in severe protein aggregation.[64]

The molecular structure of tyrosine allows for functionalization through aromatic substitution or hydroxyl group reactions. Irreversible modification of tyrosine residues occurs when the aromatic ring is functionalized. Functionalization of the aromatic ring can be easily achieved using iodination reactions, but is not ideal for structural studies since it results in changes to the protein that affect folding and biophysical properties.[54] Functionalization of the hydroxyl group is typically the easiest and reversible approach to modify tyrosine, leaving the native structure of the protein unperturbed.[54] Contrary to serine and threonine, tyrosine residues can be modified under mild conditions with high selectivity.[62] As a result, there are numerous derivatizing agents and chemical conditions available to facilitate tyrosine-based reactions. For example, tyrosine has been modified through esterification, photooxidation, acetylation, nitration, and iodination reactions.[62,65] In particular, iodination reactions of the hydroxyl group have been an essential approach to incorporate iodine radioisotopes (^{125}I and ^{131}I) for protein labeling and structural studies.[65] Typically, tyrosyl residues that can be iodinated can also be O-acetylated with acetic anhydride. These reactions are not as specific as N-acetylimidazole and tetranitromethane reagents because it is very difficult to selectively modify tyrosine (or serine and threonine) when carboxylic acid anhydrides preferentially react with primary amines.[62-63] Because the reagents available to functionalize tyrosine also derivatize other amino acids, it is best to use

other approaches to determine the presence, abundance, and function of tyrosine in the protein structure.[54]

1.2.5.3 Aromatic Heterocycles

Aromatic heterocycles are cyclic molecules containing a heteroatom such as nitrogen, oxygen, or sulfur in its structure.[55] Essentially, the number of aromatic heterocycles is substantial since they vary in ring size, conjugation, and the number and type of heteroatom included. There are only two standard amino acids that are classified as aromatic heterocycles, histidine and tryptophan, which contain imidazole and indole functional groups, respectively.

Histidine is an essential amino acid with an imidazole side chain, an aromatic five-membered heterocycle containing two nitrogen atoms. Since one nitrogen contains a lone pair and the other a hydrogen, the basicity of the nitrogens differ, where the unprotonated nitrogen has higher basicity.[55] Histidine is important to the catalytic mechanism of many enzymes. Functionalization of histidine has been explored using the same reactions executed for phenols, namely photooxidation, alkylation, and iodination reactions.[54] Initially, photooxidation was not specific towards histidine residues since a number of dyes can also oxidize tryptophan, tyrosine, cysteine, and methionine.[62] The dye, Rose Bengal, seems to be the most specific towards histidine residues and has been essential to study the roles of histidines in the active site of enzymes.[62] Alkylation reactions are not very selective towards histidine, but can be improved if a suitable pH is used for the reaction. The histidyl side chain undergoes alkylation in the unprotonated form, which requires a pH above 5. Selectivity for histidine modification is optimum in pH range of 5 – 7, where the protonated form of other side chain amines dominate to minimize competition.[66] Most importantly, alkylating histidine does not typically alter binding or catalytic efficiency.

Iodination reactions preferentially modify tyrosine residues, but can be histidine-selective at increased pH values.[62]

Tryptophan, another essential amino acid, has a fused, bicyclic indole moiety containing one nitrogen. In general, fused-ring heterocycles have similar properties to simple heterocycles. Tryptophan is important in protein synthesis and serves as a biochemical precursor to many compounds. Chemical modification of tryptophan can be achieved through photooxidation, iodination, ozonization, and treatment with N-bromosuccinimide. Overall, these methods show poor selectivity to tryptophan since they react with other amino acid residues.[54]

Aromatic sulfenyl halides have been shown to specifically and quantitatively modify tryptophan residues in strongly acid conditions.[67] The only competing reaction is the formation of mixed disulfides with cysteine residues. Additionally, tryptophan residues can be selectively functionalized with a formyl group in anhydrous formic acid saturated with gaseous HCl without cleavage of peptide bonds.[67] Although structure has been shown to be conserved, protein activity is most often adversely effected. The harsh acidic conditions have severely limited the wide spread use of sulfenyl halides and formylation in protein structural studies.

1.2.5.4 Amines

Amines are compounds containing one or more alkyl or aryl groups bonded to a nitrogen. Similar to carbon atoms, amines are classified as primary, secondary, or tertiary depending if they have one, two, or three substituents covalently attached to the nitrogen center.[55] Overall, amines are strongly polar because they have contributing dipole moments from the lone pair, C-N, and N-R bonds. As a result, they form hydrogen bonds very easily; however, tertiary amines can only act as hydrogen bond acceptors. In comparison to O-H bonds, amines form weaker hydrogen

bonds because the N-H bond is less polar due to nitrogen's lower electronegativity compared to oxygen.[55]

Asparagine, glutamine, lysine, and arginine are four standard amino acids with amine-based functionalities. Both glutamine and asparagine are amides, and lysine and arginine are a primary amine and guanidine, respectively.[55] Owing to the strong hydrogen bonding capability of these amino acids, one function is to maintain the secondary and tertiary structure of proteins.[11] Each amino acid can also have a critical role in cellular processes and enzyme function.

Similar to other pursuits to chemically modify amino acids, a variety of reagents have been used to functionalize amines. In particular, lysine can undergone acylation, esterification, guanidination, acetoacetylation, carbamylation and arylation reactions. Most of those reactions are limited because of selectivity as threonine, serine, histidine, cysteine, and tyrosine residues are chemically derivatized. Improved selectivity has been achieved with an alkaline pH (i.e. above 8) with a number of reagents used for these reactions. However, elevated pH values can sometimes prove to be harsh to the structural integrity of the protein. Reductive alkylation is a popular method to specifically modify lysine residues requiring mild reaction conditions. Particularly, reductive methylation reaction has become the primary method used to study the role of lysine residues in protein structure and activity, which is discussed in more detail in the next section.

1.3 Reductive Methylation of Proteins

Reductive methylation is a good tool for studying protein structure and function and has been used since 1968 to study the roles of specific lysines in biological function.[68] This sparse labelling method allows for lysines to be chemically modified under slightly basic conditions in the presence of a reducing agent and formaldehyde. The protein *N*-terminal α-amine and the lysyl side-chain ε-amines undergo methylation to produce monomethylamines, which have a slightly

higher acid dissociation constant (pK_a) compared to the unmodified primary amines.[69] In the presence of excess formaldehyde, the monomethylamine readily reacts to produce a dimethylamine (Figure 1.6). Proteins can be isotopically labeled in this manner with the use of labeled formaldehyde (^{13}C and ^{14}C) and/or reducing agent (^{2}H and ^{3}H).[69-72] Since lysines are polar and the side-chains are typically solvent-exposed, it is easy to homogeneously modify most lysines by reductive methylation. Some exceptions are when the lysine is involved in a salt bridge or buried in the tertiary or quaternary structures. The reductive methylation reaction conditions are mild, so proteins can be modified without denaturation. Reductive methylation of proteins is usually performed at pH 7.5-10.0; but, in an unpublished study, we found that reductive methylation is achievable as low as pH 4.0, extending the applicability of the method. Another attractive feature is the retained charge of the amines with reductive methylation, closely maintaining the isoelectric point (pI) of the protein. Due to the small size of the methyl groups, the modification rarely introduces global changes to the protein structure or interferes with its activity, the most important feature of reductive methylation.

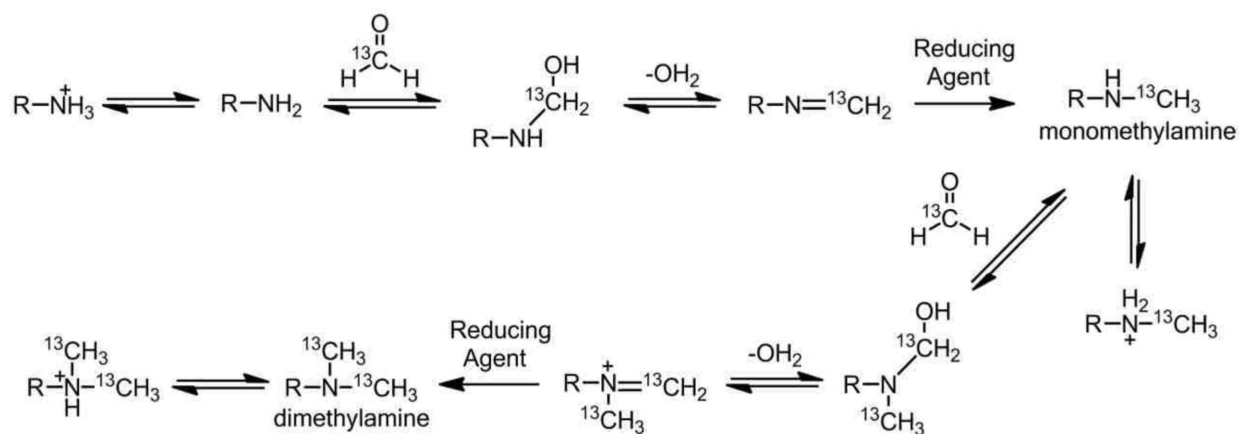


Figure 1.6. Reductive methylation reaction. Reproduced from Roberson *et al.*, *All Res. J. Chem*, 2013, 4, 10-16)

1.3.1 Techniques to Study Reductive Methylation of Proteins

Reductive methylation of proteins has been used to advance many techniques in the study of protein structure and function. Reductive methylation can facility protein crystallization by altering the dynamics of the lysyl side-chains through steric and hydrophobic interactions. Aiding protein crystallization is one of the early applications of the reductive methylation reaction and is still a major application today.[73-74] Rypniewski *et. al* demonstrated the first high-resolution three-dimensional (3D) structure of a reductively methylated protein.[75] In this work, the structures of lysozyme and reductively methylated lysozyme were compared, revealing highly similar structures with a root mean squared deviation of 0.40 Å in the backbone atoms (Figure 1.7). As a result, reductive methylation has had a major impact on studying the biophysical properties of proteins, especially structure.

Other techniques, such as bioactivity assays, electrophoresis, and circular dichroism (CD), have also benefited from reductive methylation.[71,76-79] Electrophoresis is a useful technique for assessing biophysical properties, such as the molecular weight and the pI, of medium to large proteins. Reductive methylation has been used in electrophoresis to radiolabel protein samples with ^{14}C and ^{3}H, improving the detection sensitivity over the traditional Coomassie Blue staining method.[80] CD is a technique that can probe the secondary structure of a protein by monitoring the differential absorption of circularly polarized light at ultraviolet wavelengths. Each unique combination of amino acids and secondary structure give a distinct absorption fingerprint. Reductively methylated proteins have been shown to conserve the structural fingerprint of their native counters, which allows CD to be used to analyze the effects of methylation on protein binding.[81-82] Mass spectrometry (MS) is a powerful technique for analyzing elemental and/or

isotopic composition, mass, and chemical structures of proteins and peptides. Reductive methylation has provided many advantages in MS proteomics studies for

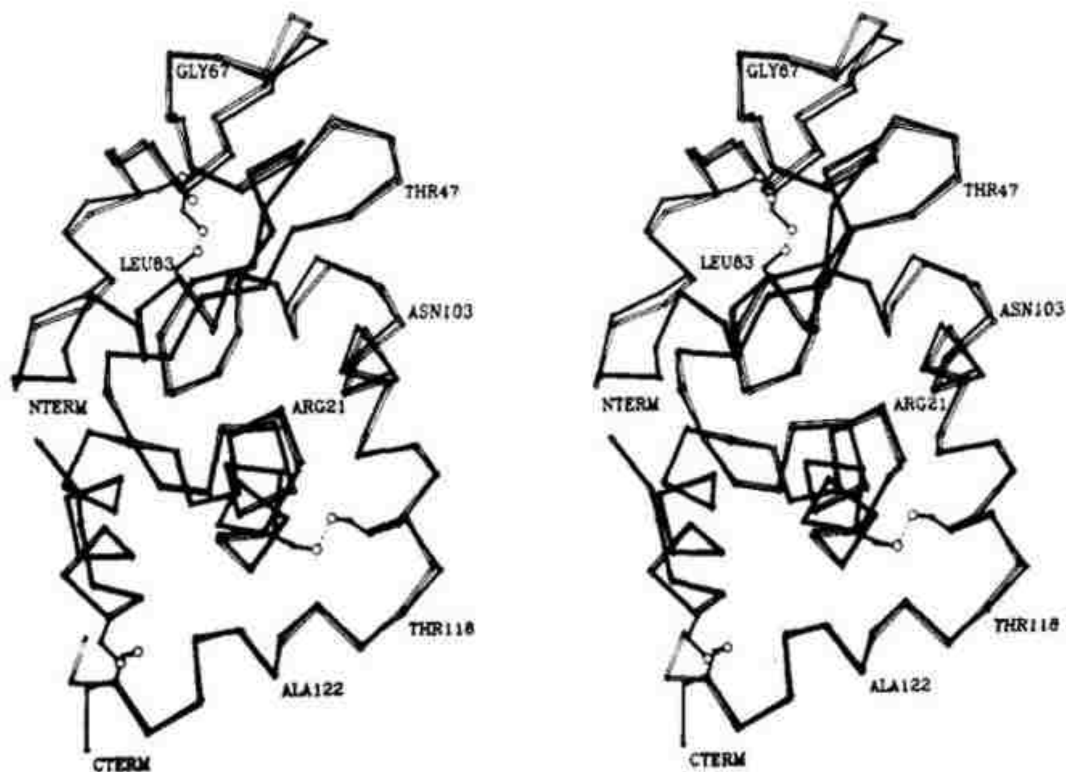


Figure 1.7. Comparison of the α-carbon positions for native (open circles) and reductively methylated (solid circles) hen egg white lysozyme. Reproduced from (Rypniewski, W.; Holden, H.; Rayment, I., Structural consequences of reductive methylation of lysine residues in hen egg-white lysozyme - an x-ray analysis at 1.8-angstrom resolution. *Biochemistry* **1993,** *32* (37), 9851-9858.) with permission from the publisher.

peptide enrichment [83], separations [84], quantitation [85], detection [86], and signal enhancement.[87] In addition, reductive methylation has been used to identify and associate protein biomarkers with breast cancer using a proteomics MS approach.[88] MS is also being used to determine tertiary and quaternary protein structure,[89-90] and the reductive methylation reaction could be used in this way to identify structurally protected lysines or the *N*-terminus.

It is well known that protein function is directly related to its structure; hence, the importance of probing protein structure on all levels. Crystallization of proteins was one of the early applications of the reductive methylation reaction to protein studies and is still a major application today.[73-74,91-93] X-ray crystallography was the first to take advantage of reductively methylated crystallized proteins to produce a 3D structure.[75] Reductive methylation created a boom in protein structure determination by X-ray crystallography that is still increasing today with the efforts of the Protein Structure Initiative. While X-ray crystallography is the most successful technique in protein structural determination, accounting for approximately 90% of all protein structures in the protein data bank, it is not without limitations. Obtaining high quality crystals is difficult, leaving a large number of structures that have yet to be solved. NMR is another technique used to determine the 3-dimensional (3D) structure of reductively methylated proteins. Reductive methylation impacted the field of NMR by the discovery of the associated roles of lysine residues with the activity of proteins.[94-97] While activity and function of these extensively researched proteins were known, these studies identified active site residues that improved understanding of protein function and mechanisms of activity. NMR has the added benefit of analyzing each modified lysine in various conditions to study the microenvironment of the lysine side chain, intra/intermolecular interactions involving the lysine, and distance restraints associated with metal binding sites for 3D structure determination.

1.3.2 Protein Structure: NMR and X-Ray Crystallography

NMR has been instrumental in the determination of solution-state protein structure through its usefulness in probing the atomic detail of both small and large macromolecules of varying complexity up to 80 kDa.[98] The technique is based on the production of measureable perturbations in a sample’s nuclear magnetization when placed inside a large magnet and exposed to powerful

radio frequency pulses of short durations. NMR resonances can be analyzed from any magnetic nuclei that can be resolved by multi-dimensional experiments. Enrichment of a protein sample with these nuclei (i.e. ^{13}C, ^{15}N, and sometimes ^{2}H) is typically required for NMR protein structural studies. To date, NMR is the only method that provides an in-solution snapshot of a protein, allowing for a unique perspective in investigating protein structure. In-solution atomic probing not only supplies information about the structure and function of macromolecules, but also provides beneficial information about dynamics, conformational equilibria, folding, and intra-/intermolecular interactions. A 3D image of a protein sample can be determined through indirect interpretation of NMR data. Typically, amide chemical shifts are used to probe the structure of a protein in its native or non-native conformation or as a result of conformational changes that arise from partial unfolding or binding. More sophisticated NMR tools, in combination with chemical shifts, reveal information about proton-proton distances through the nuclear Overhauser effect (NOE) and the orientation of nuclei within the protein structure with residual dipolar couplings (RDCs). The unique fingerprint of proteins, as observed through various NMR methods, can provide exclusive information about the relative spatial locations of different residues in a protein sequence and details critical to its function in the presence or absence of ligand or cofactor binding. Ligand dependent function is particularly important in applied biological research. The determination of a 3D structure is made possible through NMR measurements, assignments of the signals to the respective protein residues, and identification of conformational restraints, the determination of a 3D structure can be realized. In turn, the large amount of data represented by common nuclei (^{1}H, ^{13}C, and ^{15}N) in known 3D NMR structures of proteins form the basis for a multitude of empirical and semi-empirical correlations with structural parameters.

X-ray crystallography has been a vital tool in the advancement of 3D protein structure determination. Not only have the majority of the protein information known to date arisen from X-ray crystallography, it has facilitated an in-depth understanding about protein function at atomic resolution and structural information at the molecular level of proteins. This technique is built around measured angles and intensities of diffracted X-rays by crystalline atoms. These measurements are then used to create a 3D map of electron densities in the crystal. The successes of X-ray crystallography rely on the ability to crystallize pure proteins or small complexes that are easily acquired through recombinant expression. When considering larger molecules or complexes (> 100 kDa), X-ray studies are faced with the daunting requirements of high purity and large quantities of protein. If the protein sample is not stoichiometrically homogenous, it can have a detrimental impact on the crystallization of the protein and its crystal diffraction properties. Moreover, strong beam intensities and long exposure times are required to obtain high-resolution diffraction data on large macromolecules; yet, it is to the expense of severe radiation damage. Authors have used cryogenics to reduce radiation damage and increase the lifetime of the crystals that encase large macromolecules.[99] However, these crystals typically have lattice disorders from the rapid freezing process, which is known to reduce the diffraction limit. This effect is attributed mostly to the spatially unequal contraction of the solvent and crystal lattice. Larger crystals contain greater volumes of solvent and larger protein complexes, which increase the obscurity in crystal diffraction properties. To overcome this disadvantage, an X-ray free-electron laser (FEL) is used to reduce radiation damage to cryogenized macromolecules and vastly increases the range of proteins whose structures can be solved.

NMR and X-ray crystallography are complementary methods for protein structure determination. Structures solved by both methods typically agree within error. On the other hand,

since solution NMR does not require crystals, it eliminates the potential crystal packing effects known to influence the surface structure of a protein. Despite the incredible amount of structural data garnered by X-ray crystallography over the years, the acquisition of crystals and meaningful data for macromolecules of differing size and complexity is not a trivial task. In fact, obtaining highly ordered crystals is the limiting step to X-ray structure determination, especially since some proteins crystallize with ease while others are quite challenging. Today, crystallization remains more of an art than a science, which is evident from structural genomics and proteomics data that state, for approximately 34,000 cloned proteins, less than 4% have been successfully crystallized and 2% structurally defined.[100] The Protein Structure Initiative reports similar results with a 12% success rate from purified to structurally defined protein. Recent advances in high throughput protein NMR methods to quickly assess the foldedness of proteins, which has been correlated to the ability to crystallize, show promise in the production of protein crystals that had previously eluded crystallizatioin.[101-102]

Although limited to relatively smaller proteins and complexes, NMR has the advantage over X-ray crystallography in protein dynamics studies. Proteins are known to be flexible molecules with vibrational motions. Proteins often contain flexible linker regions and/or unstructured domains in which NMR can be used to identify them.[103] Monitoring protein folding is another advantage NMR has over X-ray crystallography. Hydrogen exchange experiments have been used to determine the degree of protection of hydrogen donors due to protein structure.[104-106] Real-time NMR experiments are also used to monitor time-dependent structure formation.[107-109] NMR experiments can give chemically specific microenvironment information such as inter-atomic distances from nuclear Overhauser effects (NOEs), dihedral angle restraints from J couplings, and orientation information from RDCs.[110-112] The aforementioned capabilities allow

for dynamic studies of proteins using NMR that are not accessible with X-ray crystallography through static crystals.

1.3.3 Protein Samples for NMR

Protein samples for NMR structural studies can either be isolated from natural sources or expressed and purified in either prokaryotic or eukaryotic cells. Both methods must reproducibly yield homogenous protein in milligram quantities.[113] Early experiments of NMR structural studies relied on proteins that were abundant in natural resources such as myoglobin (blood), lysozyme (egg white), and hexokinase (every domain of life from bacteria, yeast, humans and other vertebrates). However, proteins existing in abundant quantities from natural sources are rather limited. Structural studies of proteins isolated from natural sources utilized the proton, which is the most essential nucleus in protein NMR. The ^{1}H isotope is the most sensitive nuclei available for protein NMR, existing in 99.985% of all hydrogen in nature.[114] Two other useful isotopes for protein NMR are ^{13}C and ^{15}N, which exist in nature at 1.1 and 0.37%, respectively, thus requiring artificial enrichment to make most modern NMR experiments possible.[114] This realization led to the emergence of recombinant DNA expression of target proteins in prokaryotic cells, which is still the most common source of protein sample for NMR and X-ray crystallography studies performed today. In this process, the DNA sequence of a target protein is cloned into a vector and subsequently transformed into the host cell, most commonly *Escherichia coli*. The bacteria uses glucose and ammonium chloride as carbon and nitrogen sources, respectively, to produce the necessary amino acids to express the proteins encoded in the vector. If the bacteria is fed ^{13}C-glucose and ^{15}N-ammonium chloride, then the target protein is fully expressed with ^{13}C and ^{15}N increasing the sensitivity of the sample and allowing magnetization transfer experiments for NMR studies.

Recombinant expression in bacteria is not amenable to eukaryotic proteins that require post-translational modifications. This limitation is attributed to the inability of the bacteria to produce the eukaryotic post-translational modification and the absence of chaperone proteins necessary for proper folding of the target protein. With over 300 different types of post-translational modifications, the applicability of bacterial metabolic isotopic labeling is limited. It is possible to express these proteins in eukaryotic hosts, but uniform isotopic labeling is prohibitively expensive. Chemical modification is an alternative approach to incorporate isotopes into a target protein from source. In contrast to protein expression, chemical modification of proteins introduces sparse isotopic labels via selective reactions at specific amino acid side chains. The chemical modification we have focused on in our research is reductive methylation of lysine side chains.

The most daunting obstacle of the reductive methylation labeling method for NMR applications is assigning the resonances to the respective amines. This labeling method introduces sparse ^{13}C-labels that can be used as sensitive probes for protein-protein interactions and distinct isolated sites to collect distant restraints from paramagnetic resonance perturbations. However, assignment of the NMR resonances is required before the applications can be realized. Traditional NMR assignment strategies rely on through bond (1 – 4 bond distance) or through space (< 4 Å) coupling. Because the reductive methylation reaction is a sparse labeling technique, through bond and through space distances of the ^{13}C-methyls lie outside the capabilities of traditional assignment strategies, thus emphasizing the need for an alternative assignment strategy. Here, we review the strategies designed to address the difficult task of assignment and compare results for the assignment of reductively ^{13}C-methylated hen egg white lysozyme.

1.3.4 NMR Assignment Strategies

Assignment of the dimethyllysyl NMR peaks was pioneered in 1973 by Bradbury and Brown.[115] This work established the utility of the dimethylamino group to probe the microenvironment of lysines by studying the dimethylamino ^{1}H-NMR peaks of lysozyme as a function of gadolinium (III) (Gd^{3+}) concentration. Through PRE, Gd^{3+} has a broadening effect on NMR peaks, which is inversely related to the distance to the sixth power between Gd^{3+} and the nucleus of interest. Broader lines were observed as the concentration of Gd^{3+} bound to lysozyme increased, with the group closest to the paramagnetic ion broadening the most. Assignments made with this method were based on a single Gd^{3+} binding site observed in the x-ray crystal structure determined by Blake and Rabstein (not published). Corroboration of the results with experimental pK_a values and the 3D structure allowed for assignment of the six ^{1}H-NMR peaks of the ε-dimethyllysines. It should be noted that lysozyme has six lysines, so seven dimethylamino peaks are expected; however, inadequate modification of the *N*-terminal α-amino group, confirmed by experimental pK_a values, resulted in six peaks. While absolute assignment of *ε*-dimethyllysyl peaks was accomplished, this assignment method is limited to proteins that bind paramagnetic ions and have known structures with ion-to-methyl distances in the range of 15 – 24 Å. In 1975, Brown and Bradbury published an addendum to this procedure using reductively methylated ribonuclease A (RNase-A).[116] In this work, the authors compared the pK_a values obtained from pH titrations of the dimethylamines to those of model compounds to make assignments. The pK_a of the *N*-terminal α-dimethylamine of a protein is distinct and typically around 7.0. The lysyl ε-dimethylamines fall within the range of 9.5 – 11.0, depending on their microenvironment. When the dimethylamines are involved in a salt bridge, the pK_a is typically lower. Because of their distinct pK_a values, the authors were able to assign the α-dimethylamino peak ($pK_a = 6.6$) as well as the ε-dimethylamino

peak of Lys41 ($pK_a = 8.6$), with the aid of the crystal structure, which shows Lys41 in a salt bridge. The remaining dimethylamines had pK_a values in the range of (10.6 – 11.2) and could not be assigned by titration alone.

In 1979, Jentoft and Dearborn described new methods for reductive methylation and studying the environments of dimethylamines on proteins.[117] Reductive methylation of proteins with sodium borohydride ($NaBH_4$), the method introduced by Means and Feeney, suffers from low efficiency due to the strength of this reducing agent. In addition to reducing the Schiff base, $NaBH_4$ also reduces aldehydes and ketones, thus the reduction of formaldehyde to methanol competes with the desired, forward reaction. Other problems associated with $NaBH_4$ are its strong pH dependence (pH 9.0) and the reduction of disulfide bonds. The method presented by Jentoft and Dearborn uses sodium cyanoborohydride ($NaCNBH_3$), a milder reducing agent that does not reduce aldehydes or ketones at neutral pH, to reductively methylate proteins. Additional benefits to using $NaCNBH_3$ are the increased pH range (pH 7.0 – 9.0) and a higher degree of lysine modification. Another difference between the methods is that Jentoft and Dearborn used carbon-13 NMR (^{13}C-NMR) to study the dimethylamino peaks instead of ^{1}H-NMR. Reductive methylation of the lysyl side-chains can yield monomethylamines and dimethylamines depending on the efficiency of the reaction and limiting reagents. The ^{1}H chemical shifts of both mono- and dimethyl peaks are similar and often overlap each other as well as other ^{1}H signals from the protein. Titration (pH) experiments are often used to differentiate the α- and ε-dimethylamino peaks, however significant ^{1}H signal overlap due to an inefficient reaction can render this method ineffective as signals shift in the same direction with pH. On the other hand, the ^{13}C chemical shifts of methylamines are sensitive to the “amine order.” Monomethylamines (δ ~32 – 34) and dimethylamines (δ ~41 – 44) are isolated from each other as well as other signals from the protein.

Jentoft and Dearborn demonstrated that α-dimethylamino peaks shift upfield while the ε-dimethylamino peaks shift downfield with increasing pH. Using this new method of reductive methylation and ^{13}C NMR, Jentoft *et al*. studied RNase-A. Peak assignments of RNase-A were made by comparing the ^{13}C-spectra of dimethylated RNase-A to the spectra of model compounds and verified both the α-dimethylamino and Lys41 ε-dimethylamino peaks by pH titrations. This work demonstrated the benefits of using the ^{13}C NMR peaks and became the basis for ensuing experiments.

Variations of Jentoft and Dearborn's method were later used by Gerken *et al*. to study lysozyme and Sparks *et al*. to study apolipoprotein A-1 (apoA-1).[118-120] Instead of using model compounds to verify assignments made from experimental pK_a values, Gerken *et al.* used the x-ray structure of lysozyme, while Sparks *et al.* used the predicted secondary structure of apoA-1 to make and verify assignments. Like their predecessors, each of the methods is limited by the number of distinct pK_a values. Additionally, altered methods used by Gerken *et al*. and Sparks *et al*. are limited by their reliance on determined and predicted structural data, respectively. Using pH titrations for assigning dimethylamino peaks is proven to be a useful tool; however, supplementing with a secondary technique is necessary to achieve absolute assignment. In order to use reductive methylation as a strategy to study protein structure by NMR, it is important that the assignment techniques not rely on prior knowledge of the protein other than the amino acid sequence.

In 1988, Dick *et al*. used a ^{13}C-NMR, pH titration method on fd gene 5 protein (G5P) with peptide sequencing of partially radiolabeled (^{3}H and ^{14}C) G5P to make assignments.[94] This method advanced the development of an ideal assignment strategy and was the first to utilize partial methylation for assignment. In this method, G5P was reductively ^{13}C-methylated both free

(control) and bound to a ligand followed by subsequent ^{14}C-methylation to completion (unbound) to incorporate radiolabels. The protein was enzymatically digested and sequenced by Edman degradation. Radioactivity was monitored during degradation to assess the level of protection by bound ligand of each lysyl residue. Those lysyl residues with higher radioactivity exhibited lower peak intensity in the ^{13}C-NMR spectrum. One of the lysines of G5P is the *C*-terminal residue, which Dick *et al*. cleaved using carboxypeptidase B. Subsequent difference spectra of methylated G5P and the carboxypeptidase-modified G5P were used to assign the *C*-terminal lysyl residue. Degenerate levels of radioactive isotopes led to group-wise assignments; however, absolute assignments were possible by comparing the data to the x-ray crystal structure.

The assignment of dimethylamino NMR peaks soon progressed with the introduction of two dimensional (2D) NMR. The 2D experiments improved the resolution of the dimethylamino peaks by combining the resolution and chemical shift dispersion of dimethylamino peaks in the ^{13}C-dimension with the separation of the peaks in the ^{1}H-dimension and sensitivity of ^{1}H detection. Zhang and Vogel used site-directed mutagenesis to assign dimethylamino peaks in a ^{1}H-^{13}C heteronuclear multiple quantum coherence (HMQC) spectrum of reductively methylated mammalian calmodulin (CaM).[121] Individual mutants (Lys → Gln) of CaM were generated for each of the 7 lysl residues followed by reductive methylation with ^{13}C-formaldehyde. Each peak was assigned by its absence in the spectrum of the corresponding mutant. This assignment method is the first to achieve absolute assignment without prior knowledge of the target protein's structure and has significant implications towards protein structural studies. Although technically sound and successful, site-directed mutagenesis can require extensive work for those proteins containing many lysines.

In 2000, Ashfield *et al.* presented an assignment method that utilized matrix assisted laser desorption ionization – time of flight (MALDI-TOF) MS to identify partially ^{13}C-methylated lysines on tryptic peptides.[122] The use of MS to probe the extent of methylation at each lysine is an alternative to peptide sequencing (Edman degradation) used by Dick *et al.* and does not require radio-labeling. Dimethylamino NMR peaks of human MIP-1α D26A (hMIP-1α), which contains 3 lysines, were assigned by correlation of the disappearance of unmodified peptide masses observed in MS with the appearance of mono- and dimethylamino peaks seen in ^{13}C 1-D and 2D ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) NMR spectra. Reductive methylation of hMIP-1α under conditions in which the protein forms a dimer revealed that one of the lysyl residues was completely protected from modification. The authors were able to make absolute assignments of the 3 reductively methylated lysines of hMIP-1α with only the amino acid sequence as prior knowledge.

The MS-assisted assignment method, developed by Macnaughtan *et al.,* expands on the Ashfield method with the introduction of using MS to measure the amount of ^{13}C incorporated at each dimethylamine.[123] This method utilized partial labeling for assignment based on the varying rates of the reductive methylation reaction at each site. Partial labeling with ^{13}C-formaldehyde followed by labeling with excess natural abundance formaldehyde resulted in varying levels of ^{13}C at each dimethylamine. The amount of ^{13}C-incorporation at each site was measured by NMR using the peak volumes in a 2D ^{1}H-^{13}C HSQC spectrum and MS using the shift in the isotopic profile of peptides produced by trypsin digestion. MS also provides the identity of the dimethylamine through the unique mass-to-charge (*m/z)* of the peptide, which can also be verified by tandem MS experiments. By correlating the NMR and MS data for ^{13}C-incorporation, the dimethylamino peaks can be assigned knowing only the amino acid sequence of the target protein. Assignments

of the dimethylamino NMR peaks of hen egg white lysozyme were made through the correlation of ^{13}C-incorporation percent values calculated for both tryptic peptides with MS and NMR peaks of 2D ^{1}H-^{13}C HSQC NMR experiments. While the range of partial labeling was good (30.2 – 71.4 %), the accuracy of the NMR and/or MS measurements of ^{13}C-incorporation did not allow for complete assignment. In addition, the MS data was incomplete because individual isotopic profile measurements could not be made for the two *N*-terminal dimethylamines or for the tandem lysines, Lys96 and Lys97. The inherent difficulty in assigning the *N*-terminal lysine led to the supplemental method presented by Roberson *et al.*[124] This method presented two techniques to identify and assign the *N*-terminal lysine α- and ε-dimethylamino NMR peaks. The first method used aminopeptidase to identify the two peaks belonging to the *N*-terminal lysine, and the second method used pH-induced chemoselectivity to modify either the α-amine or the lysine side-chain ε-amines. Assignments were made by the loss of peaks in 2D ^{1}H-^{13}C HSQC spectra and by using the MS-assisted assignment method on selectively labeled samples. The pH technique provides advantages over the titration approach in assigning the α-dimethylamino peak because it does not require a supplemental technique or structural knowledge to confirm the assignments. It also has the added benefit of assigning the *N*-terminal ε-dimethyllysyl peak, which has a pK_a similar to other ε-dimethyllysines. Establishment of these methods to supplement the original work by Macnaughtan *et al.* aided in further assignments of lysozyme; however, limited accuracy and degenerate labeling did not allow for absolute assignments, despite intense efforts to vary the extent of labeling.[125]

Recently, Larda *et al.* published a new approach for the assignment of ε-methyllysyl peaks for protein structural studies.[69] Instead of focusing on dimethylamines, the authors presented a protocol that favors monomethylation of the amino groups to take advantage of the improved

resolution and overcome exchange broadening. In addition, deuterium tags allowed the use of pulse sequences that filter out natural abundance ^{13}C signals. Assignments of methylamino peaks were made by comparing aromatic nuclear Overhauser effects (NOEs) and line broadening effects from dissolved nitroxide spin labels to that of predicted effects from the x-ray crystal structure. Dependence on the assignments made by Macnaughtan *et al*. and the crystal structure limits this method, although absolute assignments were achieved.

1.3.5 Assignment of Reductively Methylated HEWL

Lysozyme has been the model protein most often used in developing an assignment strategy for dimethyllysyl NMR peaks. With six lysines and the *N*-terminal amine, seven sites of methylation are possible and will be referred to as αLys1, εLys1, Lys13, Lys33, Lys96, Lys97, and Lys116. Depending on the pH, ionic strength, and temperature of the reductively methylated lysozyme sample, the methyl groups of a single dimethylamine can be equivalent and exhibit one NMR peak or they can be non-equivalent and exhibit two NMR peaks. The number of peaks and the peak widths depend on the dynamics associated with the methyl groups, such as bond rotations, amine-inversion, and pK_a. Assigning the NMR peaks of the dimethylamino groups to the lysines and *N*-terminal amine is usually performed within the pH range of 8 – 9 at room temperature, when each site has equivalent methyl groups and seven peaks are observed as shown in Figure 1.8. Assignments of the dimethylamino peaks have been made in part or in absolution by five of the methods discussed, which we refer to as Bradbury, Gerken, Macnaughtan, Roberson, and Larda (Table 1.2). Each assignment method will be discussed in detail and their results compared.

In the first study, conducted by Bradbury and Brown, six resolved peaks corresponding to the six ε-dimethyllysines were titrated to obtain pK_a values and then assigned using selective broadening effects with increasing concentrations of Gd^{3+}.[115] This method is based on the

Table 1.2. Assignments of dimethylamino NMR peaks to the lysyl residue number and N-terminal amine of reductively methylated hen egg white lysozyme with peak numbers corresponding to those in Figure 3.

Peak number	Bradbury[a]	Gerken[b]	Macnaughtan[c]	Roberson[d]	Larda[e]
1	97	~97	ε1 or 13	13	13
2	96		33 or 116	33 or 116	116
3	116	13	ε1 or 13	97	96
4	13	~97	α1	96	97
5	33		33 or 116	33 or 116	33
6	ε1	ε1	96	ε1	ε1
7	α1	α1	97	α1	α1

[a] Data from reference [115].
[b] Data from reference [119].
[c] Data from reference [123].
[d] Based on data from reference [124] and [123].
[e] Data from reference [69].

assumption that the paramagnetic ion binds in a single location between Glu35 and Asp52. Each residue was assigned based on the observed pK_a values compared to those of model compounds and the side-chain interactions interpreted from the x-ray crystal structure. The authors noted that the only ambiguity in the assignments was between peaks 1 and 2, corresponding to Lys97 and Lys96, respectively, citing identical chemical shifts at pH 4.5. Paramagnetic perturbation studies have been useful in protein structure determination; however, when used to assign NMR peaks, this technique requires structural knowledge of the protein. For that reason this method is not applicable to proteins whose structures have not been solved. In addition, the assumption made in this assignment strategy was questioned by Gerken *et al.*, citing multiple binding sites for Gd^{3+} at high concentrations.[119]

Gerken *et al.* used observed pK_a values and chemical shifts to make assignments of lysozyme dimethylamino NMR peaks.[119] The αLys1 peak was assigned by the distinct pK_a and chemical shift. A suggested "salt bridge" between Lys13 and the *C*-terminal Leu129 in the x-ray crystal structure aided in the assignment of Lys13's dimethylamino peak. This assignment was

confirmed with the altered pK_a value of Lys13 when Leu129 was cleaved. Likewise, the εLys1 was assigned with assistance from the suggested ion pairing between Lys1 and Glu7. Lys97 was deduced to be either peak 1 or 4 (peak numbering based on Figure 1.8 and not on Gerken *et al.*), however, the remaining peaks could not be unambiguously assigned. Table 1.2 summarizes the peak assignments for each method for direct comparison. Gerken's assignments agree with Bradbury's for αLys1, εLys1, and Lys97 (peak 1), but other assignments were different or missing. Like Bradbury and Brown's method, this method is limited to those proteins whose 3D structures have been solved. Even though both methods used suggested salt bridges and ion pairings from the crystal structure to justify their assignments, the assignments of all, except for αLys1 and εLys1, conflict. Viable arguments were presented to rationalize the assignments from both

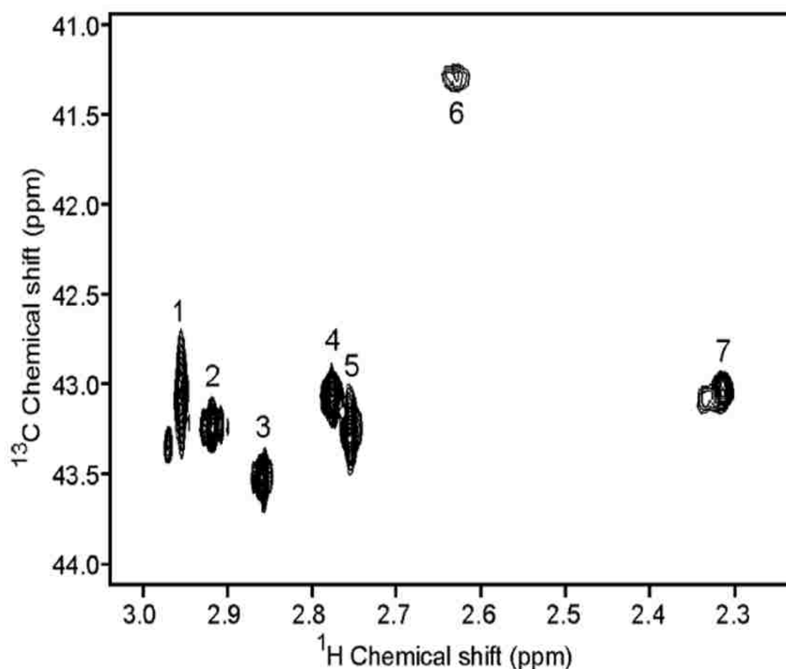


Figure 1.8. ^{1}H-^{13}C HSQC of ^{13}C-dimethylated HEWL (120 μM) at pH 8.1 on a 600 MHz spectrometer in approximately 5 min. with the same numbering scheme of Table 1.2. Reproduced from (Macnaughtan, M. A.; Kane, A. M.; Prestegard, J. H., Mass spectrometry assisted assignment of nmr resonances in reductively c-13-methylated proteins. *J Am Chem Soc* **2005,** *127*, 17626-17627.) with permission of the publisher.

methods, which suggests that reliance on the crystal structure introduces bias when assigning the NMR peaks.

Assignments of lysozyme dimethylamino NMR peaks by Macnaughtan *et al*. relied on the correlation of isotope incorporation measurements using NMR of the intact protein and MS of dimethylamino, tryptic peptides.[123] Definitive assignments were made for three of the seven peaks (αLys1, Lys96, and Lys97), but the remaining four peaks were assigned pairwise (εLys1 and Lys13; Lys33 and Lys116) as shown in Table 1.2. Absolute assignment of all dimethylamino peaks were prohibited by two problems associated with this method. First, isotope incorporation was degenerate for two pairs of sites. The reaction rates at these sites are comparable, likely due to similar microenvironments. Various methods have been tested to break the degeneracy, including varying the reducing agent and using 18-crown-6-ether as a selective protecting agent.[124] Improving the accuracy of the NMR and/or MS measurements may reveal a difference in the level of labeling. The second problem is that individual isotope incorporation values could not be measured for all dimethylamines with MS. For peptides that contain two dimethylamines, only the average isotope incorporation can be measured with MS. There are a couple of instances in which this can happen: (1) if a peptide contains more than one lysine and (2) if the *N*-terminal residue of the protein is a lysine. Trypsin is commonly used to digest proteins for MS analysis, but it does not cleave at dimethylated lysines. Consecutive lysines, such as Lys96 and Lys97 in lysozyme, pose a significant challenge for digestion. One possible solution is to use high-resolution tandem MS to measure the isotopic profile of peptide fragments. On the other hand, the selectivity of trypsin can be an advantage. The assignments of Lys96 and Lys97 are based on the observation of peptide Lys97-Arg112 in the MS, which suggests that some Lys96 was not methylated, even under conditions of excess formaldehyde. Consequently, the NMR peak with

the lowest peak area (peak 6) was assigned to Lys96 (Table 1.2). The α- and ε-dimethylamines of the *N*-terminal lysine cannot be separated using a protease and other methods must be used.

Roberson *et al.* published two methods to distinguish, identify, and assign both the αLys1 and εLys1 NMR peaks as a supplement to the MS-assisted assignment method.[124] The NMR spectrum of lysozyme treated with aminopeptidase displayed the disappearance of peaks 6 and 7, identifying these peaks as the dimethylamino groups of the *N*-terminal lysine. Further investigation of the peaks with pH-induced selectivity of the reductive methylation reaction confirmed those results. At low pH, the α-amine methylated at a higher rate than any of the side-chain ε-amines, but at high pH, no methylation was observed for the α-amine. The high pH sample made it possible to distinguish the αLys1 peak from the εLys1 peak. With little to no methylation at the α-amine, the *N*-terminal peptide could be analyzed with MS to give a ^{13}C-incorporation value for the ε-dimethylamine, which made the assignment of εLys1 possible (independent of the aminopeptidase method). Although peaks 6 and 7 were previously assigned to Lys96 and Lys97, respectively, data from the *N*-terminal methods were conclusive and these peaks are εLys1 and αLys1, respectively. Combining these new assignments with Macnaughtan's data[123] provides new assignments for peaks 1 (Lys13), 3 (Lys97), and 4 (Lys96), but peaks 2 and 5 are still ambiguous (Lys33 or Lys116) due to degenerate labeling (Table 1.2). The assignment of αLys1 and εLys1 may provide insight to improve the MS-assisted assignment method. The average ^{13}C-incorporation measurements of αLys1 and εLys1 are 54.2% by NMR (Peak 6 = 30.2%, Peak 7 = 78.3%) and 44.8% by MS (peptide K1-R5). The difference between the values is very large and clearly indicates that the accuracy of the NMR and/or MS measurement must be improved for the method to be viable.

Larda *et al.* introduced the use of monomethylamines and deuterium-filtered NMR to improve the quality of the NMR peaks for assignment.[69] Using data from Macnaughtan *et al.* and predicted, spin-label-induced line broadening from the x-ray crystal structure, all monomethyl- and dimethylamino signals were assigned (Table 1.2). The assignment method is limited by its reliance on the crystal structure, the same as the Bradbury and Gerken methods. The adjusted assignments made by combining Roberson and Macnaughtan's data agree with Larda's assignments, with the exception of Lys96 and Lys97. It is not surprising that the assignments of Lys96 and Lys97 are unresolved, since a pair of consecutive lysines is difficult to distinguish using spin-labels or MS.

To date, the only consensus and conclusive assignments of dimethylamino peaks of hen egg white lysozyme are the αLys1 and the εLys1. Although there is some overlap in the assignment of the remaining peaks, they are difficult to accept as definite. The use of crystal structures to aid in the assignments produces conflicting results and limits the applicability of the method. While proteomics MS is useful, it is not without its limitations, including the need to separate, identify, and measure every dimethylamine. Advanced tandem MS methods and more accurate modeling of dimethylated lysines from crystal structures may alleviate the current challenges to dimethylamino NMR peak assignment.

To date, the only consensus and conclusive assignments of *N*-dimethylamino resonances of hen egg white lysozyme are the α-amine and the K1 ε-amine. Although there is some overlap in the assignment of the remaining resonances, not enough evidence has been presented to make definite assignments. Each of the methods presented here has its limitations and limited reliability. However, as a whole, they represent how much the assignment of dimethyllysine resonances has come.

1.4 Conclusion

In 40 years, labeling proteins with reductive methylation has proved useful to many analytical techniques and permitted the study of proteins with limited prior knowledge of structural properties. Using isotopically labeled reagents, radio- or magnetic-isotopes can be incorporated in high yield and with little impact on the protein's overall fold or activity. Labeling a protein with ^{13}C-dimethylamino groups provides a means of studying proteins by NMR and is particularly useful for labeling proteins from source or purified from eukaryotic cell cultures. The ^{13}C-dimethylamino NMR peaks can be studied in a number of ways to characterize the protein's activity and probe its tertiary and quaternary structures. The applicability of the reductive methylation strategy to NMR is limited by the lack of a universal assignment strategy that does not require a high resolution structure or recombinant expression. Our review of five assignment strategies for reductively ^{13}C-methylated hen egg white lysozyme exemplifies the difficulty of the task and creativity devoted to overcoming this hurdle. Using partial labeling of the protein (radio- or magnetic-isotopes) and a secondary method to identify and correlate the data (Edman degradation/scintillation or MS) has the advantage of not requiring a protein structure for assignment. Such an assignment strategy has the potential to be universal, but requires higher accuracy in measuring the amount of partial labeling for absolute correlation. When the structure of a protein is available, soluble (nitroxide) or bound (Gd^{3+}) spin-labels and measured pK_a values can be used to make assignments. This strategy produced conflicting results for hen egg white lysozyme, but improved computational modeling of reductively methylated proteins may provide more accurate predictions of paramagnetic effects and the state of salt bridges in solution.

1.5 References

1. Atsbury, W. T.; Woods, H. J., The molecular weights of proteins. *Nature* **1931,** *127*, 663-665.

2. Kauzmann, W., Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959,** *14*, 1-63.

3. Jacoesen, C. F.; Linderstrom-Lang, K., Salt linkages in proteins. *Nature* **1949,** *164*, 411-412.

4. Pauling, L.; Corey, R. B.; Branson, H. R., The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **1951,** *37* (4), 205-211.

5. Afinsen, C. B.; Redfield, R. R.; Choate, W. I.; Page, J.; Carroll, W. R., Studies on the gross structure, cross-linkages, and terminal sequences in ribonuclease. *J. Biol. Chem.* **1954,** *207*, 201-210.

6. Bischof, J. C.; He, X., Thermal stability of proteins. *Ann. N. Y. Acad. Sci.* **2006,** *1066* (1), 12-33.

7. Dill, K. A.; Shortle, D., Denatured states of proteins. *Annu. Rev. Biochem.* **1991,** *60* (1), 795-825.

8. Neurath, H.; Greenstein, J. P.; Putnam, F. W.; Erickson, J. A., The chemistry of protein denaturation. *Chem. Rev. (Washington, DC, U. S.)* **1944,** *34* (2), 157-265.

9. Uversky, V. N.; Permiakov, E. A., *Methods in protein structure and stability analysis: Conformational stability, size, shape, and surfaceof protein molecules*. Nova Publishers: 2007; p 404.

10. Lyubarev, A. E.; Kurganov, B. I., Methods in protein structure and stability analysis -- conformational stability, size, shape and surface of protein molecules. *Biochemistry (Moscow)* **1998,** *63*.

11. Horton, H. R.; Moran, L. A.; Scrimgeour, K. G.; Perry, M. D.; Rawn, J. D., *Principles of Biochemistry*. 4th ed.; Pearson Prentiss-Hall: Saddle River, New Jersey, 2006.

12. Kumar, A.; Venkatesu, P., Overview of the stability of α-chymotrypsin in different solvent media. *Chem. Rev. (Washington, DC, U. S.)* **2012,** *112* (7), 4283-4307.

13. Sashi, P.; Yasin, U. M.; Bhuyan, A. K., Unfolding action of alcohols on a highly negatively charged state of cytochrome C. *Biochemistry* **2012,** *51* (15), 3273-3283.

14. Giugliarelli, A.; Sassi, P.; Paolantoni, M.; Morresi, A.; Dukor, R.; Nafie, L., Vibrational circular dichroism spectra of lysozyme solutions: Solvent effects on thermal denaturation processes. *J. Phys. Chem. B* **2013,** *117* (9), 2645-2652.

15. Voets, I. K.; Cruz, W. A.; Moitzi, C.; Lindner, P.; Arêas, E. P. G.; Schurtenberger, P., DMSO-induced denaturation of hen egg white lysozyme. *J. Phys. Chem. B* **2010,** *114* (36), 11875-11883.

16. Griebenow, K.; Klibanov, A. M., On protein denaturation in aqueous−organic mixtures but not in pure organic solvents. *J. Am. Chem. Soc.* **1996,** *118* (47), 11695-11700.

17. Sawyer, W. H.; Puckridge, J., The dissociation of proteins by chaotropic salts. *J. Biol. Chem.* **1973,** *248* (24), 8429-8433.

18. Kunz, W.; Henle, J.; Ninham, B. W., 'Zur lehre von der wirkung der salze' (about the science of the effect of salts): Franz hofmeister's historical papers. *Curr. Opin. Colloid Interface Sci.* **2004,** *9* (1–2), 19-37.

19. Zhang, Y.; Cremer, P. S., Interactions between macromolecules and ions: The hofmeister series. *Curr. Opin. Chem. Biol.* **2006,** *10* (6), 658-663.

20. Romsted, L. S., Do amphiphile aggregate morphologies and interfacial compositions depend primarily on interfacial hydration and ion-specific interactions? The evidence from chemical trapping. *Langmuir* **2006,** *23* (2), 414-424.

21. Zavitsas, A. A., Properties of water solutions of electrolytes and nonelectrolytes. *J. Phys. Chem. B* **2001,** *105* (32), 7805-7817.

22. Salvi, G.; De Los Rios, P.; Vendruscolo, M., Effective interactions between chaotropic agents and proteins. *Proteins: Struct., Funct., Bioinf.* **2005,** *61* (3), 492-499.

23. Adebowale, Y. A.; Adebowale, K. O., The influence of kosmotropic and chaotropic salts on the functional properties of mucuna pruriens protein isolate. *Int. J. Biol. Macromol.* **2007,** *40* (2), 119-125.

24. Adebowale, Y. A.; Adeyemi, I. A.; Oshodi, A. A.; Niranjan, K., Isolation, fractionation, and characterisation of proteins from mucuna bean. *Food Chem.* **2007,** *104* (1), 287-299.

25. Alonso, D. O. V.; Dill, K. A., Solvent denaturation and stabilization of globular proteins. *Biochemistry* **1991,** *30* (24), 5974-5985.

26. Zhang, Y.; Cremer, P. S., Chemistry of hofmeister anions and osmolytes. *Annu. Rev. Phys. Chem.* **2010,** *61* (1), 63-83.

27. International, H. G. S. C., Finishing the euchromatic sequence of the human genome. *Nature* **2004,** *431* (7011), 931.

28. Nørregaard Jensen, O., Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **2004,** *8* (1), 33-41.

29. Ayoubi, T. A.; Van De Ven, W. J., Regulation of gene expression by alternative promoters. *FASEB J.* **1996,** *10* (4), 453-60.

30. Wold, F., In vivo chemical modification of proteins (post-translational modification). *Annu. Rev. Biochem.* **1981,** *50* (1), 783-814.

31. Rucker, R. B.; McGee, C., Chemical modifications of proteins in vivo: Selected examples important to cellular regulation. *J. Nutr.* **1993,** *123* (6), 977-990.

32. Chattapadhyay, R.; Pelka, H.; Schulman, L. H., Initiation of in vivo protein synthesis with non-methionine amino acids. *Biochemistry* **1990,** *29* (18), 4263-4268.

33. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133-7155.

34. Beck-Sickinger, A. G.; Mörl, K., Posttranslational modification of proteins. Expanding nature's inventory. By Christopher T. Walsh. *Angew. Chem., Int. Ed.* **2006,** *45* (7), 1020-1020.

35. Orlean, P.; Menon, A. K., Thematic review series: Lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: How we learned to stop worrying and love glycophospholipids. *J. Lipid Res.* **2007,** *48* (5), 993-1011.

36. Fujita, M.; Jigami, Y., Lipid remodeling of GPI-anchored proteins and its function. *Biochim. Biophys Acta, Gen. Subj.* **2008,** *1780* (3), 410-420.

37. Hershko, A., Ubiquitin-mediated protein degradation. *J. Biol. Chem.* **1988,** *263* (30), 15237-40.

38. Zeng, L.-R.; Vega-Sanchez, M. E.; Zhu, T.; Wang, G.-L., Ubiquitination-mediated protein degradation and modification: An emerging theme in plant-microbe interactions. *Cell Res.* **2006,** *16* (5), 413.

39. Leach, M. D.; Brown, A. J. P., Posttranslational modifications of proteins in the pathobiology of medically relevant fungi. *Eukaryotic Cell* **2012,** *11* (2), 98-108.

40. Jacobson, A. D.; Zhang, N.-Y.; Xu, P.; Han, K.-J.; Noone, S.; Peng, J.; Liu, C.-W., The lysine 48 and lysine 63 ubiquitin conjugates are processed differently by the 26 s proteasome. *J. Biol. Chem.* **2009,** *284* (51), 35485-35494.

41. Sadowski, M.; Sarcevic, B., Mechanisms of mono- and poly-ubiquitination: Ubiquitination specificity depends on compatibility between the E2 catalytic core and amino acid residues proximal to the lysine. *Cell Div.* **2010,** *5* (1), 19.

42. Gavel, Y.; Heijne, G. V., Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: Implications for protein engineering. *Protein Eng.* **1990,** *3* (5), 433-442.

43. Wong, C.-H., Protein glycosylation: New challenges and opportunities. *J. Org. Chem.* **2005,** *70* (11), 4219-4225.

44. Ubersax, J. A.; Ferrell Jr, J. E., Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **2007,** *8* (7), 530.

45. Cohen, P., The regulation of protein function by multisite phosphorylation: a 25 year update. *Trends Biochem. Sci.* **2000,** *25* (12), 596-601.

46. Polevoda, B.; Sherman, F., Methylation of proteins involved in translation. *Mol. Microbiol.* **2007,** *65* (3), 590-606.

47. Balbas, P.; Lorence, A., *Recombinant gene expression: Reviews and protocols*. 2nd ed.; Humana Press: 2004; Vol. 267, p 508.

48. Studier, F. W.; Roseberg, A. H.; Dunn, J. J.; Dubendorff, J. W., Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **1990,** *185*, 60-89.

49. Palva, E. T.; Silhavy, T., Lacz fusions to genes that specify exported proteins: A general technique. *Mol. Gen. Genet.* **1984,** *194* (3), 388-394.

50. Maina, C. V.; Riggs, P. D.; Grandea Iii, A. G.; Slatko, B. E.; Moran, L. S.; Tagliamonte, J. A.; McReynolds, L. A.; Chu, d. G., An *escherichia coli* vector to express and purify foreign proteins by fusion to and separation from maltose-binding protein. *Gene* **1988,** *74* (2), 365-373.

51. Zerbs, S.; Giuliani, S.; Collart, F., Chapter eleven - small-scale expression of proteins in *e. Coli*. In *Methods enzymol.*, Jon, L., Ed. Academic Press: 2014; Vol. Volume 536, pp 117-131.

52. Andersen, D. C.; Krummen, L., Recombinant protein expression for therapeutic applications. *Curr. Opin. Biotechnol.* **2002,** *13* (2), 117-123.

53. Chu, L.; Robinson, D. K., Industrial choices for protein production by large-scale cell culture. *Curr. Opin. Biotechnol.* **2001,** *12* (2), 180-187.

54. Matthews, K.; Chakerian, A.; Gardner, J., [23] protein chemical modification as probe of structure-function relationships. In *Protein 3- DNA Interactions*, Sauer, R., Ed. Academic Press: 1991; Vol. Volume 208, pp 468-496.

55. Wade Jr., L. G., *Organic Chemistry*. 5th ed.; Pearson Education, Inc.: Upper Saddle River, NJ, 2003.

56. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G., Chemical modification of proteins at cysteine: Opportunities in chemistry and biology. *Chem. – Asian J.* **2009,** *4* (5), 630-640.

57. Lindley, H., A new synthesis substrate for trypsin and its application to the determination of the amino-acid sequence of proteins. *Nature* **1956,** *178* (4534), 647-648.

58. Sams, C. F.; Friedman, B. E.; Burgum, A. A.; Yang, D. S.; Matthews, K. S., Spectral studies of lactose repressor protein modified with nitrophenol reporter groups. *J. Biol. Chem.* **1977,** *252* (10), 3153-3159.

59. Crankshaw, M. W.; Grant, G. A., Modification of cysteine. In *Curr. Protoc. Protein Sci.*, John Wiley & Sons, Inc.: 2001.

60. Ellman, G. L., Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **1959,** *82* (1), 70-77.

61. Fontana, A.; Scoffone, E.; Benassi, C. A., Sulfenyl halides as modifying reagents for polypeptides and proteins. II. Modification of cysteinyl residues. *Biochemistry* **1968,** *7* (3), 980-986.

62. Glazer, A. N., Specific chemical modification of proteins. *Annu. Rev. Biochem.* **1970,** *39* (1), 101-130.

63. Lundbald, R. L., *Modification of hydroxyl and carboxyl groups in proteins*. Taylor & Francis Group, LLC: Boca Raton, FL, 2012; p 426.

64. Marzotto, A.; Giormani, V., Attempts of chemical modification of threonine and serine residues in RNase A. *Experientia* **1970,** *26* (8), 833-834.

65. Wood, E. J., Chemical modification of proteins: By G.E. Means and R.E. Feeney. pp. 254. Holden-Day Inc., San Francisco, 1971. (from mcgraw-hill in the U.K.). *Biochem. Educ.* **1974,** *2* (2), 34-34.

66. Gurd, F. R. N., Carboxymethylation. *Methods Enzymol.* **1967,** *Volume 11*, 532-541.

67. Previero, A.; Colettip.Ma; Cavadore, J. C., A reversible chemical modification of tryptophan residue. *Biochim. Biophys. Acta* **1967,** *147* (3), 453-&.

68. Means, G.; Feeney, R., Reductive alkylation of amino groups in proteins. *Biochemistry* **1968,** *7* (6), 2192-2201.

69. Larda, S. T.; Bokoch, M. P.; Evanics, F.; Prosser, R. S., Lysine methylation strategies for characterizing protein conformations by NMR. *J. Biomol. NMR* **2012,** *54* (2), 199-209.

70. Rice, R. H.; Means, G. E., Radioactive labeling of proteins in-vitro. *J. Biol. Chem.* **1971,** *246* (3), 831-832.

71. Ottesen, M.; Svensson, B., Use of reductive methylation for radioactive labeling of proteins. *C. R. Trav. Lab. Carlsberg* **1971,** *38* (23), 445-456.

72. She, Y. M.; Rosu-Myles, M.; Walrond, L.; Cyr, T. D., Quantification of protein isoforms in mesenchymal stem cells by reductive dimethylation of lysines in intact proteins. *Proteomics* **2012,** *12* (3), 369-379.

73. Rayment, I., Reductive alkylation of lysine residues to alter crystallization properties of proteins. In *Macromolecular Crystallography, Pt A*, 1997; Vol. 276, pp 171-179.

74. Walter, T. S.; Meier, C.; Assenberg, R.; Au, K. F.; Ren, J. S.; Verma, A.; Nettleship, J. E.; Owens, R. J.; Stuart, D. I.; Grimes, J. M., Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **2006,** *14*, 1617-1622.

75. Rypniewski, W.; Holden, H.; Rayment, I., Structural consequences of reductive methylation of lysine residues in hen egg-white lysozyme - an X-ray analysis at 1.8-angstrom resolution. *Biochemistry* **1993,** *32* (37), 9851-9858.

76. Vanderwel, H.; Wiersma, A.; Brouwer, J. N., Tritium labeling of thaumatin I, a sweet-tasting protein from thaumatococcus-daniellii benth, by reductive methylation. *J. Labelled Compd. Radiopharm.* **1978,** *14* (5), 735-740.

77. Khanh, N. Q.; Linde, R.; Lipecky, R.; Gassen, H. G., Inactivation of ribosomal-protein S1 by means of reductive methylation. *Hoppe-Seyler's Z.Physiol. Chem.* **1978,** *359* (3), 284-284.

78. Guha, S.; Szulmajster, J., Analysis of ribosomal-protein conformation in bacillus-subtilis by reductive methylation - identification of proteins with different conformation in monosomes and polysomes. *FEBS Lett.* **1980,** *118* (1), 103-108.

79. Heacock, C. S.; Bernstein, B. W.; Duhaiman, A. S.; Amorese, D. A.; Bamburg, J. R., In vitro labeling of proteins by reductive methylation - application to proteins involved in supramolecular structures. *J. Cell Biochem.* **1982,** *19* (1), 77-91.

80. Finger, J. M.; Choo, K. H., Double-label reductive methylation of tissue proteins for precision two-dimensional polyacrylamide-gel electrophoretic analysis *Biochem. J.* **1981,** *193* (1), 371-374.

81. Gray, D. M.; Sherry, A. D.; Teherani, J.; Kansy, J. W., Reductive methylation of the lysyl residues in the fd gene 5 DNA-binding protein - CD and C-13 NMR results on the modified protein. *J. Biomol. Struct. Dyn.* **1984,** *2* (1), 77-91.

82. Fretheim, K.; Egelandsdal, B.; Harbitz, O. L. E., Effect of alkylation with different sized substituents on the conformation of ovomucoid, lysozyme and ovotransferrin. *Int. J. Pept. Protein Res.* **1985,** *- 25* (- 6), - 607.

83. Prokai-Tatrai, K.; Guo, J.; Prokai, L., Selective chemoprecipitation and subsequent release of tagged species for the analysis of nitropeptides by liquid chromatography-tandem mass spectrometry. *Mol. Cell Proteomics.* **2011,** *10* (8).

84. Hsu, J. L.; Huang, S. Y.; Shiea, J. T.; Huang, W. Y.; Chen, S. H., Beyond quantitative proteomics: Signal enhancement of the a(1) ion as a mass tag for peptide sequencing using dimethyl labeling. *J. Proteome Res.* **2005,** *4* (1), 101-108.

85. Guo, J.; Prokai-Tatrai, K.; Prokai, L., Relative quantitation of protein nitration by liquid chromatography-mass spectrometry using isotope-coded dimethyl labeling and chemoprecipitation. *J. Chromatogr. A* **2012,** *1232*, 266-275.

86. Chicooree, N.; Connolly, Y.; Tan, C.-T.; Malliri, A.; Li, Y.; Smith, D. L.; Griffiths, J. R., Enhanced detection of ubiquitin isopeptides using reductive methylation. *J. Am. Soc. Mass Spectrom.* **2013,** *24* (3), 421-430.

87. Hsu, J. L.; Huang, S. Y.; Chen, S. H., Dimethyl multiplexed labeling combined with microcolumn separation and ms analysis for time course study in proteomics. *Electrophoresis* **2006,** *27* (18), 3652-3660.

88. Cohen, A.; Wang, E.; Chisholm, K. A.; Kostyleva, R.; O'Connor-McCourt, M.; Pinto, D. M., A mass spectrometry-based plasma protein panel targeting the tumor microenvironment in patients with breast cancer. *J. Proteomics* **2013,** *81*, 135-147.

89. Kriwacki, R.; Reisdorph, N.; Siuzdak, G., Protein structure characterization with mass spectrometry. *Spectroscopy (N. Y., NY, U. S.)* **2004,** *18* (1), 37-47.

90. Kussmann, M.; Roepstorff, P., Characterisation of the covalent structure of proteins from biological material by maldi mass spectrometry – possibilities and limitations. *Spectroscopy (N. Y., NY, U. S.)* **1998,** *14* (1), 1-27.

91. Claesson, M.; Siitonen, V.; Dobritzsch, D.; Metsa-Ketela, M.; Schneider, G., Crystal structure of the glycosyltransferase snogd from the biosynthetic pathway of nogalamycin in streptomyces nogalater. *FEBS J.* **2012,** *279* (17), 3251-3263.

92. Buttner, C. R.; Sorg, I.; Cornelis, G. R.; Heinz, D. W.; Niemann, H. H., Structure of the yersinia enterocolitica type III secretion translocator chaperone SYCD. *J. Mol. Biol.* **2008,** *375* (4), 997-1012.

93. Au, K.; Ren, J. S.; Walter, T. S.; Harlos, K.; Nettleship, J. E.; Owens, R. J.; Stuart, D. I.; Esnouf, R. M., Structures of an alanine racemase from bacillus anthracis (BA0252) in the presence and absence of (R)-1-aminoethylphosphonic acid (l-ALA-P). *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2008,** *64*, 327-333.

94. Dick, L.; Sherry, A.; Newkirk, M.; Gray, D., Reductive methylation and C-13 NMR studies of the lysyl residues of fd gene 5 protein - lysines 24, 46, and 69 may be involved in nucleic acid binding. *J. Biol. Chem.* **1988,** *263* (35), 18864-18872.

95. Dick, L. R.; Geraldes, C.; Sherry, A. D.; Gray, C. W.; Gray, D. M., C-13 NMR of methylated lysines of fd gene-5 protein - evidence for a conformational change involving lysine-24 upon binding of a negatively charged lanthanide chelate. *Biochemistry* **1989,** *28*, 7896-7904.

96. Jentoft, J.; Gerken, T.; Jentoft, N.; Dearborn, D., [C-13] methylated ribonuclease-A - C-13 NMR studies of the interaction of lysine 41 with active site ligands. *J. Biol. Chem.* **1981,** *256* (1), 231-236.

97. Huque, M.; Vogel, H., C-13 NMR studies of the lysine side chains of calmodulin and its proteolytic fragments. *J. Protein Chem.* **1993,** *12* (6), 695-707.

98. Mainz, A.; Religa, T. L.; Sprangers, R.; Linser, R.; Kay, L. E.; Reif, B., NMR spectroscopy of soluble protein complexes at one mega-dalton and beyond. *Angew. Chem., Int. Ed.* **2013,** *52* (33), 8746-8751.

99. Higashiura, A.; Ohta, K.; Masaki, M.; Sato, M.; Inaka, K.; Tanaka, H.; Nakagawa, A., High-resolution X-ray crystal structure of bovine H-protein using the high-pressure cryocooling method. *J. Synchrotron Radiat.* **2013,** *20*, 989-993.

100. Chayen, N. E., Turning protein crystallisation from an art into a science. *Curr. Opin. Struct. Biol.* **2004,** *14* (5), 577-583.

101. Bailey-Kellogg, C.; Widge, A.; Kelley, J. J.; Berardi, M. J.; Bushweller, J. H.; Donald, B. R., The NOESY jigsaw: Automated protein secondary structure and main-chain assignment from sparse, unassigned NMR data. *J. Comput. Biol.* **2000,** *7* (3-4), 537-558.

102. Mielke, S. P.; Krishnan, V. V., Characterization of protein secondary structure from NMR chemical shifts. *Prog. Nucl. Magn. Reson. Spectrosc.* **2009,** *54* (3-4), 141-165.

103. Muchmore, S. W.; Sattler, M.; Liang, H.; Meadows, R. P.; Harlan, J. E.; Yoon, H. S.; Nettesheim, D.; Chang, B. S.; Thompson, C. B.; Wong, S. L.; Ng, S. C.; Fesik, S. W., X-ray and NMR structure of human BCL-X(l), an inhibitor of programmed cell death. *Nature* **1996,** *381* (6580), 335-341.

104. Okazaki, H.; Ohori, Y.; Komoto, M.; Lee, Y. H.; Goto, Y.; Tochio, N.; Nishimura, C., Remaining structures at the *N*- and *C*-terminal regions of alpha-synuclein accurately elucidated by amide-proton exchange NMR with fitting. *FEBS Lett.* **2013,** *587* (22), 3709-3714.

105. Englander, S. W., Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **2000,** *29*, 213-238.

106. Engel, M. F. M.; Visser, A.; van Mierlo, C. P. M., Conformation and orientation of a protein folding intermediate trapped by adsorption. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (31), 11316-11321.

107. Rennella, E.; Brutscher, B., Fast real-time NMR methods for characterizing short-lived molecular states. *Chem. Phys. Chem.* **2013,** *14* (13), 3059-3070.

108. Roche, J.; Dellarole, M.; Caro, J. A.; Norberto, D. R.; Garcia, A. E.; Garcia-Moreno, B.; Roumestand, C.; Royer, C. A., Effect of internal cavities on folding rates and routes revealed by real-time pressure-jump NMR spectroscopy. *J. Am. Chem. Soc.* **2013,** *135* (39), 14610-14618.

109. Yamasaki, K.; Ohara, Y.; Hasegawa, M.; Tanaka, H.; Yamasaki, T.; Wakuda, T.; Okada, M.; Kohzuma, T., Real-time NMR monitoring of protein-folding kinetics by a recycle flow system for temperature jump. *Anal. Chem.* **2013,** *85* (20), 9439-9443.

110. Wuthrich, K., *NMR of Proteins and Nucleic Acids*. John Wiley and Sons: New York, 1986.

111. Wuthrich, K., NMR - this other method for protein and nucleic-acid structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995,** *51*, 249-270.

112. Gronenborn, A. M.; Clore, G. M., Structures of protein complexes by multidimensional heteronuclear magnetic-resonance spectroscopy. *Crit. Rev. Biochem. Mol. Biol.* **1995,** *30* (5), 351-385.

113. Kim, Y.; Babnigg, G.; Jedrzejczak, R.; Eschenfeldt, W. H.; Li, H.; Maltseva, N.; Hatzos-Skintges, C.; Gu, M. Y.; Makowska-Grzyska, M.; Wu, R. Y.; An, H.; Chhor, G.; Joachimiak, A., High-throughput protein purification and quality assessment for crystallization. *Methods* **2011,** *55* (1), 12-28.

114. Jacobsen, N. E., *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. Wiley: 2007.

115. Bradbury, J. H.; Brown, L. R., Determination of dissociation constants of lysine residues of lysozyme by proton-magnetic-resonance spectroscopy. *Eur. J. Biochem.* **1973,** *40*, 565-576.

116. Brown, L.; Bradbury, J., Proton magnetic resonance studies of lysine residues of ribonuclease-A. *Eur. J. Biochem.* **1975,** *54* (1), 219-227.

117. Jentoft, N.; Dearborn, D. G., Labeling of proteins by reductive methylation using sodium cyanoborohydride. *J. Biol. Chem.* **1979,** *254*, 4359-4365.

118. Jentoft, J. E.; Jentoft, N.; Gerken, T. A.; Dearborn, D. G., C-13 NMR studies of ribonuclease-A methylated with [formaldehyde-C-13]. *J. Biol. Chem.* **1979,** *254*, 4366-4370.

119. Gerken, T.; Jentoft, J.; Jentoft, N.; Dearborn, D., Intramolecular interactions of amino groups in C-13 reductively methylated hen egg-white lysozyme. *J. Biol. Chem.* **1982,** *257* (6), 2894-2900.

120. Sparks, D.; Phillips, M.; Lundkatz, S., The conformation of apolipoprotein A-I in discoidal and spherical recombinant high density lipoprotein particles - C-13 NMR studies of lysine ionization behavior. *J. Biol. Chem.* **1992,** *267* (36), 25830-25838.

121. Zhang, M.; Vogel, H., Determination of the side chain pka values of the lysine residues in calmodulin. *J. Biol. Chem.* **1993,** *268* (30), 22420-22428.

122. Ashfield, J.; Meyers, T.; Lowne, D.; Varley, P.; Arnold, J.; Tan, P.; Yang, J.; Czaplewski, L.; Dudgeon, T.; Fisher, J., Chemical modification of a variant of human MIP-1 alpha; implications for dimer structure. *Protein Sci.* **2000,** *9* (10), 2047-2053.

123. Macnaughtan, M. A.; Kane, A. M.; Prestegard, J. H., Mass spectrometry assisted assignment of NMR resonances in reductively C-13-methylated proteins. *J. Am. Chem. Soc.* **2005,** *127*, 17626-17627.

124. Roberson, K. J.; Brady, P. N.; Sweeney, M. M.; Macnaughtan, M. A., Methods to identify the NMR resonances of the 13C-dimethyl *N*-terminal amine on reductively methylated proteins. *J. Vis. Exp.* **2013,** (82).

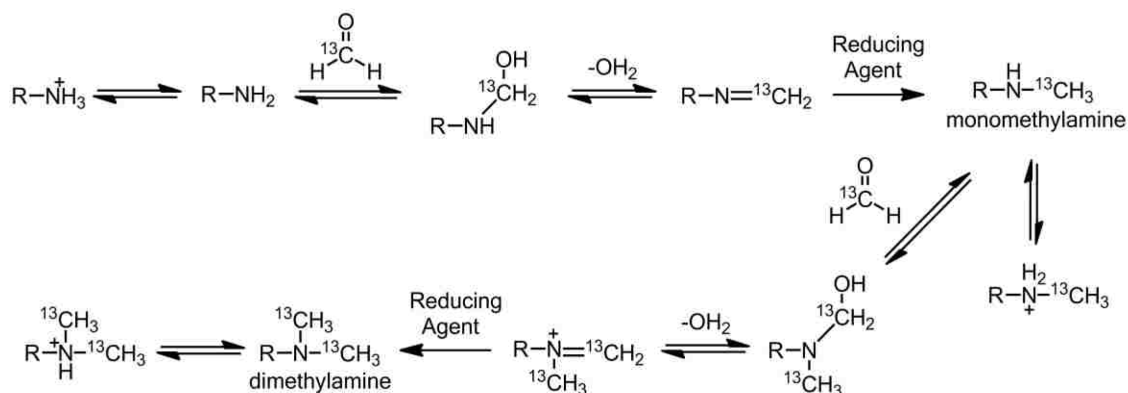
125. Roberson, K. J.; Macnaughtan, M. A., Attempts towards unambiguously assigning 13C-dimethylamine NMR resonances. *All Results J.: Chem.* **2013,** *4*, 10-16.

CHAPTER 2 ATTEMPTS TOWARD UNAMBIGUOUSLY ASSIGNING ^{13}C-DIMETHYLAMINE NMR RESONANCES[1]

2.1 Introduction

As a tool for protein structure and dynamics studies, nuclear magnetic resonance spectroscopy typically requires stable magnetic isotope enrichment to improve signal intensity and allow assignment.[1] Complete isotopic labeling is most commonly achieved through overexpression in *Escherichia coli* with ^{15}N-ammonium chloride and ^{13}C-glucose.[2-3] Because many human proteins require eukaryotic hosts for proper folding and post-translational modifications, metabolic labeling is not always viable. An alternative approach is sparse isotopic labeling using protein-chemical modification with highly selective reactions for specific amino acids.[4] ^{13}C-methyl tagging via reductive methylation has been successful in both NMR and X-ray crystallographic protein studies. Under non-denaturing conditions, the *N*-terminal α-NH_2 and the lysine side chain ε-NH_2 groups are selectively methylated in the presence of formaldehyde and a reducing agent (Scheme 2.1). The reaction produces dimethylamino groups because the monomethylamine readily reacts due to a higher pK_a than the non-methylated primary amine.[5-6] By using ^{13}C-formaldehyde, ^{13}C-labels are incorporated into the protein in the form of ^{13}C-methyls.[7] Once incorporated, protein structure and dynamics are usually not perturbed.[8] The ^{13}C-methyls can be used as probes in NMR experiments to study protein-protein interactions and as sites to collect distance constraints using paramagnetic perturbations.[9-10]

[1] This chapter previously appeared as Roberson, K. J.; Macnaughtan, M. A., Attempts towards unambiguously assigning ^{13}C-dimethylamine NMR resonances. *All Results J.: Chem.* **2013,** *4*, 10-16. (Open source) All rights reserved.



Scheme 2.1. The reductive methylation reaction: In the presence of formaldehyde and a reducing agent, the primary amine is reductively methylated to produce monomethylamine. In the presence of excess formaldehyde, the monomethylamine undergoes a second reductive methylation to produce dimethylamine.

Since its introduction as a means for protein modification by Means and Feeney in 1968,[5] reductive methylation has been used to incorporate probes of structural and dynamic properties.[9-16] When the ^{13}C-methyl groups are used to study proteins with NMR, the utility of the labeling method is limited by difficulties in assigning the ^{13}C-methyl resonances with their corresponding lysine residues. Past assignment approaches have relied on a small number of methylation sites,[14,17] known structural properties,[7,11,15,18-20] or genetic modifications.[21] One strategies used matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) to identify partially-methylated lysines of tryptic peptides of human MIP-1α by correlating the disappearance of unmodified peptide masses in MS with the appearance of mono- and dimethylamine signals in 2D ^{1}H-^{13}C HSQC-NOESY NMR spectra.[17] In 2005, we presented an improved method for assigning dimethylamine resonances in 2D ^{1}H-^{13}C HSQC NMR spectra that requires no prior knowledge of the protein except the amino acid sequence.[22] The mass spectrometry-assisted assignment method uses isotopic ratio measurements. Using limiting ^{13}C-formaldehye, this technique takes advantage of slightly differing reaction rates at each site, based on pK_a values and steric accessibility, to

distinguish sites in NMR and MS data. Data presented on this procedure using lysozyme showed two inherent problems with the strategy: 1) degenerate levels of ^{13}C-incorporation and 2) no MS method to independently measure the ^{13}C-incorporation of the α- and ε-dimethylamines on Lys1.[22]

To overcome the degeneracy in labeling lysozyme with the reductive methylation reaction, we investigated two methods: 1) reductive methylation in the presence of 18-crown-6-ether (18C6) and 2) reductive methylation using multiple reducing agents. To assign the α- and ε-dimethylamine resonances of the *N*-terminal lysine residue of lysozyme, we explored a non-destructive Edman degradation method.

2.2 Experimental

2.2.1 Reductive methylation in the presence of 18C6

Lysozyme was reductively methylated to produce two samples: (1) a fully ^{13}C-labeled sample and (2) a partially ^{13}C-labeled sample. The fully labeled sample is a control sample and percent ^{13}C-incorporation calculated for the partially labeled sample is normalized to the fully labeled sample. After the partially labeled sample is modified with ^{13}C-formaldehyde, it was set to undergo another reductive methylation with natural abundance ^{12}C-formaldehyde to produce a fully dimethylated protein sample, which make interpreting the NMR spectrum easier.

2.2.2 Fully ^{13}C-labeled control sample

To an aqueous solution of lysozyme (2.5 mg, 5 mg/mL), DMAB (6.13 μL, 1 M) was added, followed by ^{13}C-formaldehyde (12.25 μL, 1 M). The reaction mixture was shaken at 4 °C for 2 hours. A second aliquot of DMAB and ^{13}C-formaldehyde was added, and the mixture was shaken at 4 °C for an additional 2 hours. DMAB (3.06 μL, 1 M) was added, and the mixture was shaken at 4 °C overnight for a total reaction time of 24 hours.

2.2.3 Partial ^{13}C-labeling

To an aqueous solution of lysozyme (5 mg/mL), an aliquot of 18C6 (0.66 mg, 5 μmol or 1.32 mg, 10 μmol) was added for final concentrations of 5 and 10 mM respectively, followed by DMAB (1.53 μL, 1 M), then ^{13}C-formaldehyde (3.06 μL, 1 M). The mixture was shaken at 4 °C for 2 hours. A second aliquot of DMAB and ^{13}C-formaldehyde were added, and the mixture was shaken at 4 °C for an additional 2 hours. DMAB (0.77 μL, 1 M) was then added, and the mixture was shaken at 4 °C overnight for a total of 24 hours.

For reactions at ratios 2:3 and 2:5, an aliquot of 18C6 (4.0 mg, 15 μmol or 6.6 mg, 25 μmol) was added for final concentrations of 30 and 50 mM, respectively.

2.2.4 Excess natural abundance formaldehyde to complete methylation

To an aqueous solution of lysozyme (5 mg/mL), DMAB (6.13 μL of 1 M) was added, followed by formaldehyde (12.25 μL of 1 M). The mixture was shaken at 4 °C for 2 hours. A second aliquot of DMAB and formaldehyde were added, and the mixture was shaken at 4 °C for an additional 2 hours. DMAB (3.06 μL, 1 M) was then added, and mixture was shaken at 4 °C overnight for a total of 24 hours.

2.2.5 Reductive methylation using multiple reducing agents

The procedures for fully and partially ^{13}C-labeling and excess natural abundance formaldehyde described above were used without 18C6 and with various reducing agents. One fully ^{13}C-labeled control sample with DMAB and partially ^{13}C-labeled samples with each reducing agent were prepared. All reducing agent aliquots were from 1 M stock solutions except sodium borohydride, which was added as a solid in small portions over time due to its high reactivity.

2.2.6 Preparation for NMR

The protein samples were exchanged into a D_2O, 50 mM sodium borate buffer at pH 8.5 using a 4mL centrifugal filter with a 3kDa molecular weight cutoff.

2.2.7 NMR

All 1D ^{1}H - ^{13}C heteronuclear single-quantum coherence (HSQC) NMR spectra were collected using a 700 MHz Varian spectrometer equipped with a 5 mm-HCN-5922 probe. 1,2-dichloroethane-$^{13}C_2$ (26 mM in D_2O) was used as an external reference via coaxial insert. The protein concentrations were 50 μM. Experiments were acquired at 25 °C. All experiments were acquired using a relaxation delay of 5s, 256 scans, and a 4529Hz spectral width. The total acquisition time for each experiment was approximately 25 minutes. 1,2-dichloroethane-$^{13}C_2$ was used as a chemical shift reference instead of the traditional DSS. The 1D ^{1}H-NMR of the PITC-lysozyme was acquired using a relaxation delay of 1.5 s, 4 scans, and a 7022.5 spectral width. The spectra were referenced to DSS.

2.2.8 ^{13}C percent incorporation calculation from NMR data

1D ^{1}H-^{13}C HSQC was performed on a fully ^{13}C-dimethylated (control) protein sample. The area under each dimethylamine peak in the spectrum was integrated and set to 100% ^{13}C-incorporation for each site. The areas of the corresponding peaks in the spectra for the partially ^{13}C-dimethylated samples were integrated. ^{13}C percent incorporation of the dimethylamine peaks of the partially ^{13}C-methylated samples was calculated as a fraction of the respective peak for the control protein sample.

2.2.9 Non-destructive Edman degradation

2.2.9.1 Coupling

Lysozyme (10 mg, 0.680 μmol) was dissolved in pyridine, triethylamine, and water (1.2:15:10) co-solvent. PITC (1.3 mg, 9.55 μmol) was added, and the reaction was stirred at 50 °C for 30 minutes.

2.2.9.2 Cleavage

(1): PTH-lysozyme was dissolved in 100% TFA or concentrated HCl and stirred for 30 minutes at room temperature.

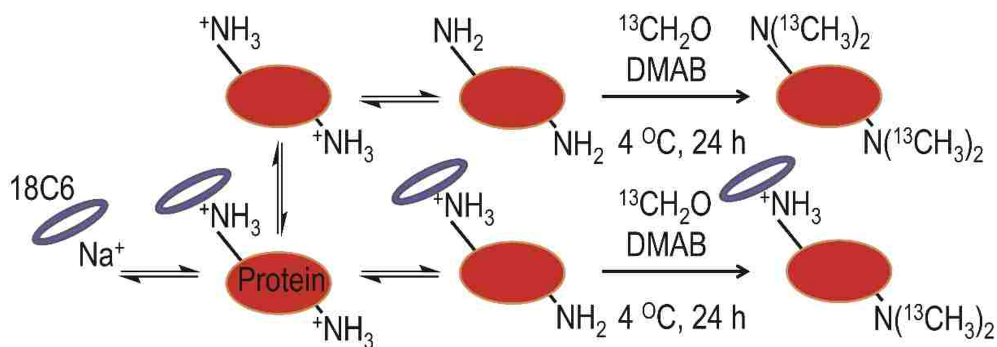
(2): PTH-lysozyme was dissolved in an organic solvent. TFA was added to make a 75 % solution and stirred for 30 minutes at room temperature.

(3): PTH-lysozyme was dissolved in triethylamine, acetic acid, and acetonitrile (7.5:3:5) and stirred at 50 °C for 30 minutes.

2.3 Results and Discussion

2.3.1 Reductive Methylation in the Presence of 18-Crown-6-Ether

18C6 is known for its ability to form non-covalent complexes to metal cations and protonated primary amines in both solution and gas phase,[23] hence its extensive use in peptide synthesis[23-26] and purification.[27-30] The 18C6 and protonated primary amine complex forms through a combination of three hydrogen bonds and ion-dipole interactions. It is through this chemistry we postulated the concept shown in Scheme 2.2 to break the degenerate levels of ^{13}C-incorporation for reductively methylated lysozyme. In solution, the lysine side chain amine exists in equilibrium between the protonated primary amine and the neutral primary amine. 18C6



Scheme 2.2. Reductive methylation in the presence of 18C6.

complexes with the protonated form, which is not the reactive species, and should hinder methylation at the site. Binding affinities of 18C6 should be slightly different at each site based on steric accessibility and influences from surrounding residues, altering the relative reductive methylation rates.

Initial methylation attempts in the presence of 18C6 were performed by varying the ratio of reactive amine concentration to 18C6. Results showed that the competitive binding of the buffer-counter ion inhibited binding of 18C6 to the reactive amine. Potassium ion binds better to 18C6 than the sodium ion, which binds slightly better than the ammonium ion, while the lithium ion does not bind as well.[31-32] For this reason, the potassium buffer was avoided and the sodium and lithium buffers were tested. Lysozyme was reductively methylated in the presence of a sodium phosphate buffer (20 mM Na^+) at ratios of 2:1 and 1:1 (moles of Na^+: moles of 18C6). In each experiment, lysozyme was reacted in the presence of 18C6 at pH 7.5.

As shown in Figure 2.1, 18C6 had no effect on the methylation of lysozyme at these molar ratios, presumably because of the poor fit of the protein ammonium ion in the 18C6 cavity or the stronger binding of 18C6 to the sodium ions. Both possibilities were tested by methylating lysozyme in the presence of excess 18C6 at ratios of 2:3 and 2:5 (counter ion: 18C6), where the

counter ion is either sodium or lithium (Figure 2.1). In each of these reactions, lysozyme was reacted in the presence of 18C6 at pH 7.5. When in excess of the counter ion, 18C6 hinders reductive methylation of lysozyme; however, the effect is equivalent across each reactive site and did not solve the degenerate labeling problem, indicating that the affinity of the protein ammonium ions for 18C6 are not significantly different. Switching the counter ion from sodium to lithium showed no significant differences in methylation, indicating that excess 18C6 is more important than counter-ion selection.

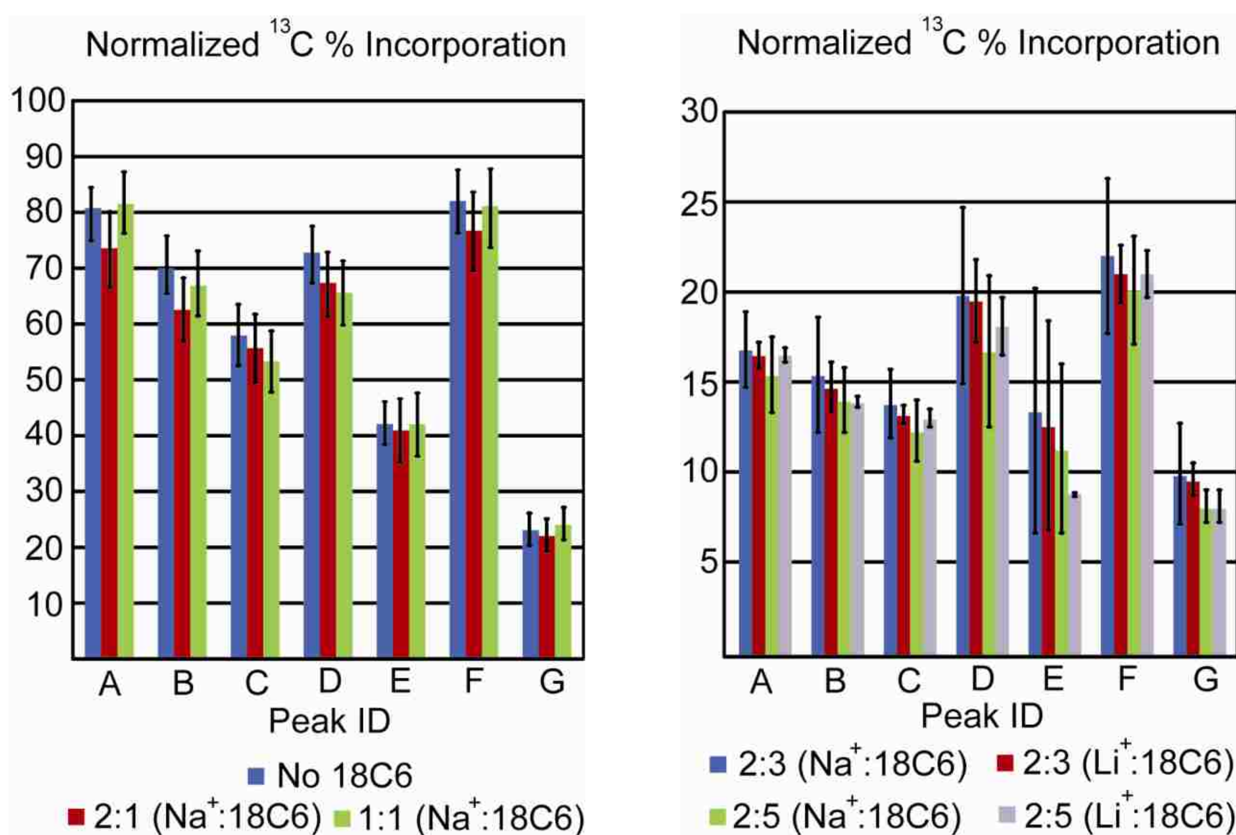


Figure 2.1. (Left) Normalized NMR 13C percent incorporation of each reaction site of partially-labeled lysozyme at 1:5 ratio (moles of reactive amine: moles of formaldehyde) in the presence of 18C6 and sodium ion. (Right) Normalized NMR 13C percent incorporation of each reaction site of partially-labeled lysozyme at 1:5 ratio (reactive amine: formaldehyde) in the presence of excess 18C6 over sodium and lithium ions.

2.3.2 Reductive Methylation Using Multiple Reducing Agents

In 1995, Means and Feeney reviewed the reductive methylation reaction and its efficiency as it related to different reducing agents.[33] The strength of the reducing agent was found to be inversely related to the efficiency of methylation. Here, we attempted to take advantage of the inefficiency of the stronger reducing agents, hypothesizing that the kinetic mode of action will be determined by the physical properties of the reducing agent and reactive site. For instance, a reducing agent with a hydrophobic structure, like pyridine borane, would likely reduce a Schiff base in a hydrophobic pocket faster than a hydrophilic reducing agent and, in turn, break the degeneracy in the ^{13}C-incorporation across the reactive sites. Figure 2.2 a-e are structural representations of the hydrophobic regions of lysozyme (yellow) with each lysine residue (orange) labeled. Reducing agents used in these experiments, in order of increasing strength, were dimethylamine borane complex (DMAB), pyridine borane complex, and sodium borohydride (Figure 2.2f).

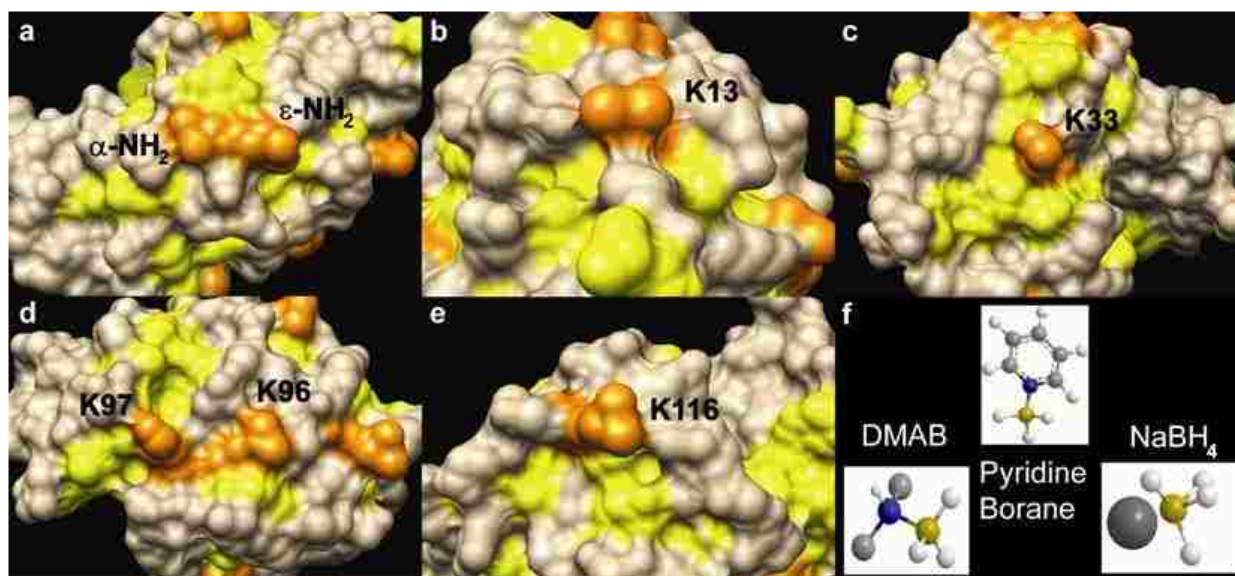


Figure 2.2. (a-e) Using the crystal structure of reductively methylated lysozyme by Rypniewski, et. al (PDB ID: 132L),[8] the dimethylamino groups (orange) and hydrophobic residues (yellow) are highlighted. (f) The structures of the borane reducing agents.

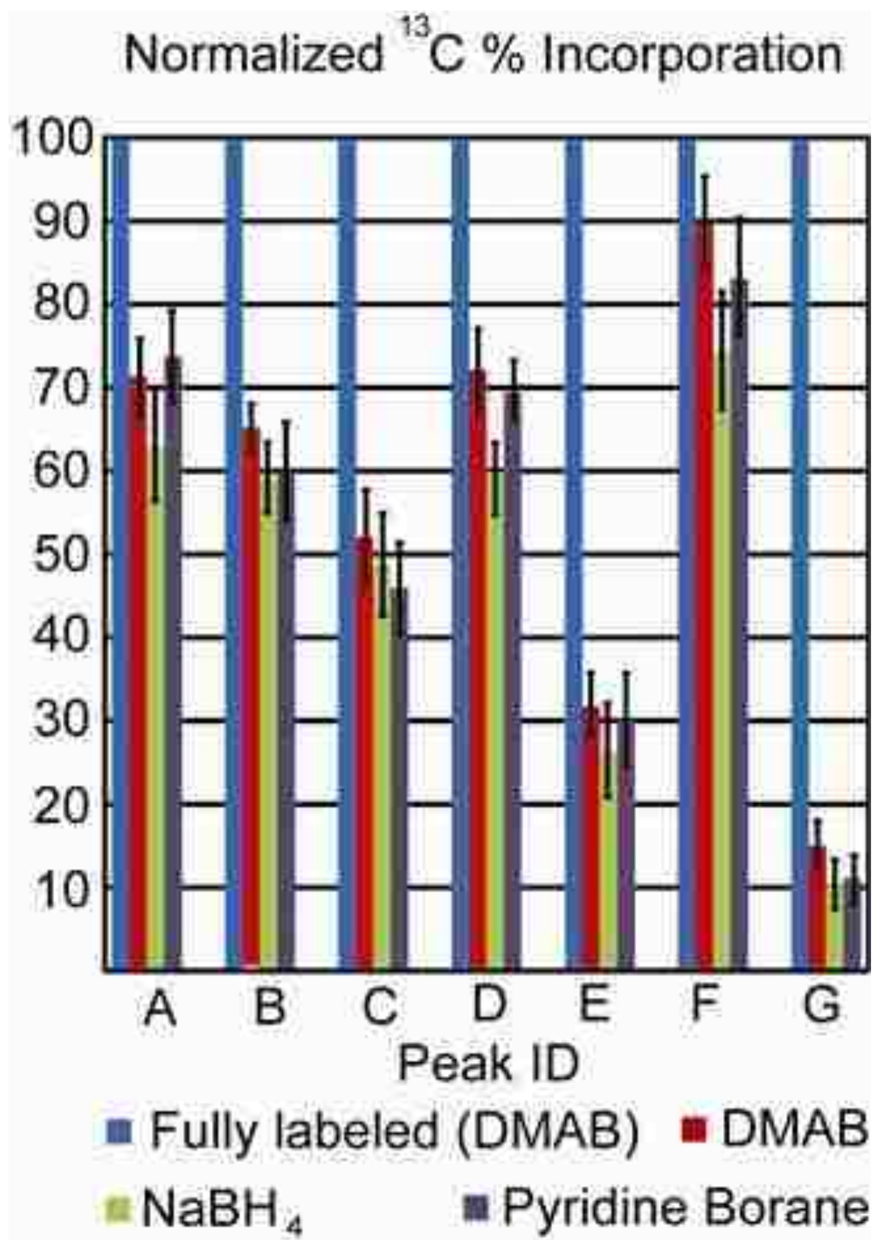


Figure 2.3. ^{13}C percent incorporation of partially 13C- reductively methylated lysozyme with different reducing agents normalized to fully 13C-methylated lysozyme with DMAB.

Lysozyme was reductively methylated with a sub-stoichiometric amount of ^{13}C-formaldehyde in the presence of each reducing agent for partial labeling, and then methylated with excess natural abundance formaldehyde in the presence of dimethylamine borane complex to complete methylation. Reactions were performed at both 4 and 25°C and yielded the same results

(data shown for 4°C). Changing the reducing agent had no significant effect on the degenerate labeling between peaks A and D, as shown in Figure 2.3. While the different reducing agents produced slightly different levels of ^{13}C, most of the rates were degenerate within error.

Interestingly, peaks D and F were reductively methylated at a faster rate with the small, hydrophobic DMAB than the hydrophilic $NaBH_4$. Previous studies have assigned Peak D as K97[19,34] and peak F as the *N*-terminal α-NH_2,[19,34-36] which were found to be sandwiched between hydrophobic regions (Figure 2.2). These findings confirm our hypothesis that the hydrophobic DMAB would reduce a Schiff base in a hydrophobic region faster than a more hydrophilic reducing agent.

2.3.3 Non-destructive Edman Degradation

Since its introduction in 1950, Edman degradation has been used to determine the amino acid sequence of proteins until recent advances in mass spectrometry.[37-38] Phenylisothiocyanate (PITC) is reacted with the *N*-terminal amine under mildly alkaline conditions to form a cyclical phenylthiocarbamoyl derivative. When treated with acid, the derivatized terminal amino acid is cleaved. After extraction, it is further treated with acid to form the more stable phenylthiohydantoin (PTH)-amino acid derivative. During this process, the protein is sequentially cleaved to determine the amino acid sequence. For our purpose, we were not necessarily interested in the cleaved PTH-amino acid derivative, but rather the remaining protein for structural studies; therefore it was important that the truncated protein remain folded. Our plan was to use Edman degradation to remove the *N*-terminal lysine residue of lysozyme (Lys1) to assign the NMR peaks belonging to the α-, and ε- dimethylamine of Lys1. Upon NMR analysis, the Lys1 ε-dimethylamine peak will be absent and a new *N*-terminal α-dimethylamine resonance will be present allowing the assignment of both the α- and ε-dimethylamine resonances of Lys1.

PITC was coupled to the *N*-terminus of lysozyme, the sample was freeze-dried, reconstituted, and analyzed with ^{1}H NMR to verify the tertiary structure. The broad range of resonances in the ^{1}H NMR spectrum (not shown) confirmed that the PTH-protein was folded. Difficulty followed in attempts to cleave the derivatized amino acid from the protein. Initially, lysozyme was dissolved in pure trifluoroacetic acid (TFA) or HCl for cleavage. The acid was then exchanged for acetonitrile before the buffer was exchanged to phosphate to prevent further degradation of the protein by acid hydrolysis, since cleavage is sensitive to aqueous conditions. NMR analysis showed that the protein was unfolded and prompted another approach to preserve the protein's tertiary structure.

Since lysozyme unfolded after cleaving with pure TFA and HCl, we investigated the use of other organic solvents in which lysozyme was soluble and not destructive to the protein structure. In a study conducted by Chin *et al.*, lysozyme was reported to be soluble in many organic solvents at concentrations greater than 10mg/mL.[39] Here, we used methanol, dimethyl sulfoxide (DMSO), glycerol, and formamide as solvents for lysozyme before adding TFA (final concentration of 75%) for cleavage. After cleavage, NMR studies showed that lysozyme unfolded in all solvents when treated with TFA. Attempts to refold the protein were also unsuccessful. In summary, all experiments performed under acidic conditions resulted in unfolded lysozyme. We also attempted cleavage studies in basic conditions using a similar co-solvent system to that of the coupling reaction previously described by Barrett and Penglis.[40] During the cleavage step at 75°C, the protein precipitated. Precipitation was avoided when cleavage was performed at 50°C, but NMR showed the protein unfolded. Moreover, cleavage studies were also unsuccessful using the same conditions as the coupling reaction, with a mixture of pyridine, triethylamine, and water

(1.2:15:10). To date, protein stability in association with the cleavage of the phenylthiocarbamoyl amino acid derivative has yet to be resolved.

2.4 Conclusion

Here we presented two theoretically sound methods to eradicate the degeneracy of ^{13}C-labeling via reductive methylation of lysozyme and one method to assign the α- and ε-dimethylamine of Lys1. The reductive methylation of lysozyme in the presence of 18C6 was found to hinder the extent of methylation but was not selective. Although a series of reducing agents were successful in reductively methylating lysozyme, they also did not produce the desired selectivity. Non-destructive Edman degradation was promising, but unsuccessful in preserving the protein's tertiary structure.

2.5 References

1. Wuthrich, K., *NMR of Proteins and Nucleic Acids*. John Wiley and Sons: New York, 1986.

2. Shimba, N.; Yamada, N.; Yokoyama, K.; Suzuki, E., Enzymatic labeling of arbitrary proteins. *Anal. Biochem.* **2002,** *301*, 123-127.

3. Palomares, L.; Estrada-Moncada, S.; Ramírez, O., Production of recombinant proteins. In *Recombinant gene expression*, Second ed.; Balbás, P.; Lorence, A., Eds. Humana Press: New York City, New York, USA, 2004; Vol. 267, pp 15-51.

4. Lundblad, R., *Chemical Reagents for Protein Modification*. CRC Press: Boca Raton, FL, 1991.

5. Means, G.; Feeney, R., Reductive alkylation of amino groups in proteins. *Biochemistry* **1968,** *7* (6), 2192-2201.

6. Abraham, S.; Hoheisel, S.; Gaponenko, V., Detection of protein–ligand interactions by nmr using reductive methylation of lysine residues. *J. Biomol. NMR* **2008,** *42* (2), 143-148.

7. Jentoft, J. E.; Jentoft, N.; Gerken, T. A.; Dearborn, D. G., C-13 nmr-studies of ribonuclease-a methylated with [formaldehyde-c-13]. *J. Biol. Chem* **1979,** *254*, 4366-4370.

8. Rypniewski, W.; Holden, H.; Rayment, I., Structural consequences of reductive methylation of lysine residues in hen egg-white lysozyme - an x-ray analysis at 1.8-angstrom resolution. *Biochemistry* **1993,** *32* (37), 9851-9858.

9. Moore, G. R.; Cox, M. C.; Crowe, D.; Osborne, M. J.; Rosell, F. I.; Bujons, J.; Barker, P. D.; Mauk, M. R.; Mauk, A. G., N-epsilon,n-epsilon-dimethyl-lysine cytochrome c as an nmr probe for lysine involvement in protein-protein complex formation. *Biochem. J.* **1998,** *332*, 439-449.

10. Dick, L. R.; Geraldes, C.; Sherry, A. D.; Gray, C. W.; Gray, D. M., C-13 nmr of methylated lysines of fd gene-5 protein - evidence for a conformational change involving lysine-24 upon binding of a negatively charged lanthanide chelate. *Biochemistry* **1989,** *28*, 7896-7904.

11. Brown, L.; Bradbury, J., Proton magnetic resonance studies of lysine residues of ribonuclease-a. *Eur. J. Biochem.* **1975,** *54* (1), 219-227.

12. Jentoft, J.; Gerken, T.; Jentoft, N.; Dearborn, D., [c-13] methylated ribonuclease-a - c-13 nmr studies of the interaction of lysine 41 with active site ligands. *J. Biol. Chem.* **1981,** *256* (1), 231-236.

13. Goux, W.; Teherani, J.; Sherry, A., Amine inversion in proteins - a c-13-nmr study of proton exchange and nitrogen inversion rates in n-epsilon,n-epsilon,n-alpha,n-alpha-[c-13]tetramethyllysine, n-epsilon,n-epsilon,n-alpha,n-alpha[c-13]tetramethyllysine methyl-ester, and reductively methylated concanavalin-a. *Biophys. Chem.* **1984,** *19* (4), 363-373.

14. Dick, L.; Sherry, A.; Newkirk, M.; Gray, D., Reductive methylation and c-13 nmr studies of the lysyl residues of fd gene 5 protein - lysines 24, 46, and 69 may be involved in nucleic acid binding. *J. Biol. Chem.* **1988,** *263* (35), 18864-18872.

15. Jentoft, J. E., Reductive methylation and c-13 nuclear-magnetic-resonance in structure-function studies of fc fragment and its subfragments. *Methods Enzymol.* **1991,** *203*, 261-295.

16. Sparks, D.; Phillips, M.; Lundkatz, S., The conformation of apolipoprotein a-i in discoidal and spherical recombinant high density lipoprotein particles - c-13 nmr studies of lysine ionization behavior. *J. Biol. Chem.* **1992,** *267* (36), 25830-25838.

17. Ashfield, J.; Meyers, T.; Lowne, D.; Varley, P.; Arnold, J.; Tan, P.; Yang, J.; Czaplewski, L.; Dudgeon, T.; Fisher, J., Chemical modification of a variant of human mip-1 alpha; implications for dimer structure. *Protein Sci.* **2000,** *9* (10), 2047-2053.

18. Sherry, A.; Teherani, J., Physical studies of c-13-methylated concanavalin a - ph- and co^{2+}-induced nuclear magnetic resonance shifts. *J. Biol. Chem.* **1983,** *258* (14), 8663-8669.

19. Gerken, T.; Jentoft, J.; Jentoft, N.; Dearborn, D., Intramolecular interactions of amino groups in c-13 reductively methylated hen egg-white lysozyme. *J. Biol. Chem.* **1982,** *257* (6), 2894-2900.

20. Huque, M.; Vogel, H., C-13 nmr studies of the lysine side chains of calmodulin and its proteolytic fragments. *J. Protein Chem.* **1993,** *12* (6), 695-707.

21. Zhang, M.; Vogel, H., Determination of the side chain pka values of the lysine residues in calmodulin. *J. Biol. Chem.* **1993,** *268* (30), 22420-22428.

22. Macnaughtan, M.; Kane, A.; Prestegard, J., Mass spectrometry assisted assignment of nmr resonances in reductively c-13-methylated proteins. *J. Am. Chem. Soc.* **2005,** *127* (50), 17626-17627.

23. Julian, R. R.; May, J. A.; Stoltz, B. M.; Beauchamp, J. L., Biomimetic approaches to gas phase peptide chemistry: Combining selective binding motifs with reactive carbene precursors to form molecular mousetraps. *Int. J. Mass Spectrom.* **2003,** *228* (2-3), 851-864.

24. Botti, P.; Ball, H. L.; Lucietto, P.; Pinori, M.; Rizzi, E.; Mascagni, P., The use of crown ethers in peptide chemistry – v. Solid-phase synthesis of peptides by the fragment condensation approach using crown ethers as non-covalent protecting groups. *J. Pept. Sci.* **1996,** *2* (6), 371-380.

25. Botti, P.; Ball, H. L.; Rizzi, E.; Lucietto, P.; Pinori, M.; Mascagni, P., The use of crown-ethers in peptide chemistry .4. Solid-phase synthesis of peptides using peptide-fragments n-alpha protected with 18-crown-6. *Tetrahedron* **1995,** *51* (18), 5447-5458.

26. Hyde, C. B.; Mascagni, P., The use of crown ethers in peptide chemistry .3. Synthesis of an enkephalin derivative using 18-crown-6 as a non-covalent amino protecting group. *Tetrahedron Lett.* **1990,** *31* (3), 399-402.

27. Poll, D. J.; Harding, D. R. K., Column-friendly reversed-phase high-performance liquid-chromatography of peptides and proteins using formic-acid with sodium-chloride and dynamic column coating with crown ethers. *J. Chromatogr.* **1991,** *539* (1), 37-45.

28. Josic, D.; Reutter, W.; Reusch, J., Crown ethers as ligands for high-performance liquid-chromatography of proteins and nucleic-acids. *J. Chromatogr.* **1989,** *476*, 309-318.

29. Julian, R. R.; Beauchamp, J. L., Site specific sequestering and stabilization of charge in peptides by supramolecular adduct formation with 18-crown-6 ether by way of electrospray ionization. *Int. J. Mass Spectrom.* **2001,** *210*, 613-623.

30. Julian, R. R.; Beauchamp, J. L., The unusually high proton affinity of aza-18-crown-6 ether: Implications for the molecular recognition of lysine in peptides by lariat crown ethers. *J. Am. Soc. Mass Spectrom.* **2002,** *13* (5), 493-498.

31. Liou, C. C.; Brodbelt, J. S., Comparison of gas-phase proton and ammonium ion affinities of crown ethers and related acyclic analogs. *J. Am. Chem. Soc.* **1992,** *114* (17), 6761-6764.

32. Maleknia, S.; Brodbelt, J., Gas-phase selectivities of crown ethers for alkali-metal ion complexation. *J. Am. Chem. Soc.* **1992,** *114* (11), 4295-4298.

33. Means, G. E.; Feeney, R. E., Reductive alkylation of proteins. *Anal. Biochem.* **1995,** *224*, 1-16.

34. Bradbury, J. H.; Brown, L. R., Determination of dissociation constants of lysine residues of lysozyme by proton-magnetic-resonance spectroscopy. *Eur. J. Biochem.* **1973,** *40*, 565-576.

35. Larda, S. T.; Bokoch, M. P.; Evanics, F.; Prosser, R. S., Lysine methylation strategies for characterizing protein conformations by nmr. *J. Biomol. NMR* **2012,** *54* (2), 199-209.

36. Roberson, K. J.; Brady, P. N.; Sweeney, M. M.; Macnaughtan, M. A., Methods to identify the nmr resonances of the 13c-dimethyl n-terminal amine on reductively methylated proteins *J. Visualized Exp.* **2013,** *"In Press"*.

37. Edman, P., Preparation of phenyl thiohydantoins from some natural amino acids. *Acta Chem. Scand.* **1950,** *4* (2), 277-282.

38. Edman, P., Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.* **1950,** *4* (2), 283-293.

39. Chin, J. T.; Wheeler, S. L.; Klibanov, A. M., Communication to the editor on protein solubility in organic-solvents. *Biotechnol. Bioeng.* **1994,** *44* (1), 140-145.

40. Barrett, G. C.; Penglis, A. a. E.; Penrose, A. J.; Wright, D. E., Edman stepwise degradation of polypeptides - a new strategy employing mild basic cleavage conditions. *Tetrahedron Lett.* **1985,** *26* (36), 4375-4378.

CHAPTER 3 METHOD TO IDENTIFY THE NMR RESONANCES OF THE ^{13}C-DIMETHYL *N*-TERMINAL LYSINE ON REDUCTIVELY METHYLATED PROTEINS[2]

3.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a valuable structural elucidation tool for proteins.[1] NMR spectroscopy provides an in-solution, average snap shot of the protein in its native state. To overcome the low natural abundance of stable magnetic isotopes, it is necessary to incorporate ^{13}C and ^{15}N into the protein of interest. The most common method employed is recombinant expression in a bacterial host.[2] One downside to bacterial host over-expression is it cannot produce post-translational modifications and does not work for all proteins.[3] When bacterial expression is not a viable route for protein production, such as for glycosylated proteins and membrane proteins, over-expression in non-bacterial hosts can be used, but is difficult and expensive. Alternative methods for incorporating ^{13}C and ^{15}N isotopes into proteins for NMR analysis include sparse labeling techniques using metabolic precursors for methyl labeling[4] and single ^{13}C, ^{15}N amino acids[5-7] in non-bacterial expression hosts. A chemical approach to sparse labeling used herein is the well-established reductive ^{13}C-methylation reaction (Scheme 3.1), where the primary amino groups on a protein—the *N*-terminal α-amine and the lysine, side chain ε-amines—react to form dimethylamines.

$$RNH_3 + {}^{13}CH_3O \longrightarrow RNH({}^{13}CH_3)_2 + H_2O$$

Scheme 3.1. Reductive methylation reaction

[2] This Chapter previously appeared as (Roberson, K. J.; Brady, P. N.; Sweeney, M. M.; Macnaughtan, M. A., Methods to identify the NMR resonances of the 13C-dimethyl *N*-terminal amine on reductively methylated proteins. *J. Vis. Exp.* **2013,** (82) doi:10.3791/50875.). Reprinted with permission from MyJoVE Corporation.

Reductive methylation was first introduced as a method to chemically modify proteins by Means and Feeney.[8] The advantages of this reaction are it is applicable to any protein and works under mild conditions at buffered, physiological pH and low temperatures. In the presence of formaldehyde and a reducing agent, such as dimethylamine borane complex (DMAB), the lysine ε-amino groups and the *N*-terminal α-amino group are selectively methylated to produce dimethylated amines.

Reductive methylation has been successfully used with both NMR and X-ray crystallography. Reductive methylation is used to facilitate the crystallization of otherwise intractable proteins.[9] Hen egg white lysozyme was the first protein crystallized in the dimethylated form. The root-mean square difference of the heavy atoms in the methylated and unmethylated lysozyme structures is 0.40 Å.[10] This comparison demonstrates that the protein structure can be maintained after reductive methylation, making the reaction a viable, labeling tool for structure elucidation.

By using ^{13}C-labeled formaldehyde in the reductive methylation reaction, ^{13}C-dimethylated amines are produced. The ^{13}C-dimethylamines are NMR-detectable probes that have been used to study protein dynamics and structure.[11-16] Jentoft *et al.* used ^{13}C-formaldehyde to reductively methylate ribonuclease A to study the dynamics, kinetics, and substrate binding of the dimethyl-lysines using ^{13}C NMR.[12] Gerken *et al.* performed similar studies on lysozyme.[17]

The reductive ^{13}C-methylation labeling technique has always been limited by the lack of NMR signal assignment.[15] Most assignment strategies have relied on small numbers of sites, known structural properties, or extensive genetic modifications.[18-20] One assignment strategy requires prior knowledge of the protein structure,[15] which limits its applicability. Another strategy

employs site-directed mutagenesis, which is time consuming.[18] The use of mass spectrometry (MS) to aid in the NMR assignment of reductively ^{13}C-methylated amines was first reported by Fisher *et al.* Matrix assisted laser desorption ionization time-of-flight (MALDI-TOF) MS was used to identify the lysines at the interface of a dimer. The partially methylated lysines of human MIP-1α tryptic peptides were identified with MS and correlated with the appearance of mono- and dimethylamine signals of the intact protein observed in 2D ^{1}H-^{13}C NMR spectra.[21] Our group expanded on the use of MS and presented an assignment method that requires no prior knowledge of the protein's structure or properties other than the amino acid sequence.[22] This method is applicable to most proteins except for proteins with an *N*-terminal lysine because the MS isotopic profile of the *N*-terminal lysine α- and ε-^{13}C-dimethylamines cannot be independently measured.

To distinguish the α- and ε-amino groups on proteins with capillary electrophoresis, Córdova *et al.* presented a method to selectively acetylate the *N*-terminal α-amine or side chain ε-amine by controlling the reaction pH.[23] The reaction selectively favored the *N*-terminal α-amine at low pH and the side chain ε-amines at high pH. Similarly, Gerken *et al.* used low pH and a sub-stoichiometric amount of ^{13}C-formaldehyde to selectively methylate the α-amine of lysozyme and identify its resonance after denaturation.[17]

Here we present one method to identify the *N*-terminal α-dimethylamine and the side chain ε-dimethylamine sites of an *N*-terminal lysine residue using pH. Because this method maintains the structural integrity of the protein, the NMR assignments can be used for protein structure and dynamic studies.

3.2 Experimental

3.2.1 Materials and Methods

Lysozyme from hen egg white, borane-dimethylamine complex, formaldehyde (36.5% wt/wt), acetonitrile (HPLC grade), trifluoroacetic acid (99%), ammonium sulfate, ammonium bicarbonate, sodium borate, potassium phosphate, and sodium phosphate (monobasic and dibasic) were supplied from Sigma. 1,2-dichloroethane-$^{13}C_2$, ^{13}C-formaldehyde (99%) and deuterium oxide (99.9%, D_2O) was supplied from Cambridge Isotope Laboratories. Centrifugal filters (4mL with 3K MWCO) were supplied from Fisher Scientific. Sinipinic acid was supplied by Fluka.

3.2.2 Reductive methylation Lysozyme (Generic Procedure)

To an aqueous solution of buffered lysozyme (0.5 mL, 5 mg/mL) was added an aliquot of DMAB (1M) followed by an aliquot of ^{13}C-formaldehyde (1M) twice the volume of the DMAB aliquot. The amount of ^{13}C-formaldehyde used for each sample will be given as the ratio of moles of reactive amine to moles of ^{13}C-formaldehyde, referred to as ratio throughout. The sample was shaken and incubated at 25 °C for an hour. Aliquots of DMAB and ^{13}C-formaldehyde were added again followed by an hour of shaking and incubation at 25 °C. The sample was buffer-exchanged into sodium phosphate buffer at pH 7.5 to remove excess reagents and side products. 50 µg of the sample was set aside for MALDI analysis to calculate the percentage of ^{13}C incorporation. For all partially labeled samples, unless otherwise noted, the reaction was repeated with excess natural abundance formaldehyde at a ratio of 1:10 (1 mole of reactive amine: 10 moles of formaldehyde) to produce homogeneous, dimethylated protein. The sample was buffer-exchanged into D_2O buffer at pH 8.5 for NMR analysis.

3.2.2.1 Fully ^{13}C-labeled Lysozyme

The lysozyme sample was reacted in 50 mM sodium phosphate buffer at pH 7.5 at a ratio of 1:10.

3.2.2.2 Partially ^{13}C-labeled Lysozyme

The lysozyme sample was reacted in 50 mM sodium phosphate buffer at pH 7.5 at a ratio of 1:5.

For reactions at ratios 2:3 and 2:5, an aliquot of 18C6 (4.0 mg, 15µmol or 6.6mg, 25µmol) was added for final concentrations of 30 and 50mM, respectively.

3.2.2.3 Partial ^{13}C-methlation at Low pH

Two lysozyme samples were reacted in 50 mM sodium phosphate buffer at pH 6.0 at a ratio of 1:5. One was not methylated to completion with natural abundance formaldehyde.

3.2.2.4 Partial ^{13}C-methlation at High pH

The lysozyme sample was reacted in 50 mM sodium borate buffer at pH 10.0 at a 1:5 ratio.

3.2.3 MALDI-TOF Sample Preparation

The high pH protein sample (50 µg), 15 µL of 5X ammonium bicarbonate buffer (500 mM), and 2 µL of ultrapure water were added to a 0.5 mL microcentrifuge tube and incubated at 60 °C for 60 minutes. After cooling to room temperature, 2 µL of iodoacetamide was added and the sample was covered with foil and incubated at room temperature for 25 minutes. Dithiothreitol (2 µL, 500 mM) was used to quench the reaction. Trypsin (2 µg) was added and the sample was incubated at 37 °C for 18 – 24 hours. The reaction was quenched with trifluoroacetic acid (1.5 µL) and analyzed with MALDI TOF-TOF MS.

3.2.4 MALDI-TOF Analysis

Data was acquired using a Bruker UltrafleXtreme MALDI-TOF-TOF instrument equipped with a Nd:YAG in linear mode, with 89-92% laser energy, and an average of 4 x 500 shots. The spectra were smoothed using the Savitzky-Golay algorithm, the baseline was subtracted, and the apex peak was manually selected using the Bruker FlexAnalysis software.

3.2.4.1 Determination of the Total Percent ^{13}C percent Labeling by MALDI

After the samples were reductively methylated with ^{13}C-formaldehyde, the extent of modification was determined using MALDI-MS. The protein sample was co-spotted in equal parts on the MALDI target plate with unmodified lysozyme and the matrix, sinipinic acid.

The percentage of ^{13}C incorporation for the modified lysozyme sample ($\%^{13}C_{protein}$) was determined using Equation 1:

Equation 3.1: $\%^{13}C_{protein} = \frac{M - M_0}{210.4} * 100,$

where M is the average *m/z* of the reductively ^{13}C-methylated lysozyme sample, M_0 is the average *m/z* of the unmodified lysozyme sample, and 210.4 is the theoretical mass shift corresponding to 100% reductive ^{13}C-methylation, i.e. all 7 primary amino groups labeled with 2 ^{13}C-methyl groups.

3.2.5 Determination of the percent ^{13}C labelling at each site: NMR

Each peak in the 1D ^{1}H-^{13}C HSQC of each sample was fit using Equation 2 and Mathcad 15.0 software (Parametric Technology Corporation). Equation 2 is the sum of a Lorentzian and Gaussian line shape with six variables: baseline offset (b), peak amplitude (A), Lorentzian lineshape fraction (f), peak frequency (ν_0), Lorentzian linewidth (w_L), and the Gaussian linewidth (w_G). The variables were optimized iteratively using 10 iterations until minimum error reached.

Equation 3.2: $b + A\left[\frac{|f|}{\left(1+\frac{2(\nu-\nu_0)}{w_L}\right)^2} + |1-f|e^{-\frac{(\nu-\nu_0)^2}{2w_G^2}}\right]$

Using the best fit parameter for each peak, the peak areas were calculated and normalized to the total percent ^{13}C incorporation using equation 3.

Equation 3.3: $A_i' = \frac{A_i}{A_{total}} * \%^{13}C_{\boldsymbol{protein}}$,

where A_i is the individual peak area and A_{total} is the sum of all peak areas.

The percentage of reductive ^{13}C-methylation at each primary amino group (i) for the partially labeled lysozyme samples ($\%^{13}C_i$) was calculated by comparing the normalized NMR peak areas of a fully, labeled control sample ($A_{i,control}'$) to the normalized NMR peaks areas of the partially labeled sample (A_i') using Equation 4.

Equation 3.4: $\%^{13}C_i = \frac{A_i'}{A_{i,control}'}$

3.2.6 Mass Spectrometry

From the mass spectrum, a weighted average using the peak intensity and isotope number (where monoisotopic = 0) was calculated for each peptide (m_{avg}). Additionally, theoretical isotopic distributions were generated using MS-Isotope (Protein Prospector, UCSF)[24] for the same peptide sequence with natural abundance labeling and 99.9% ^{13}C labeling; these were called the ^{12}C and ^{13}C standards, respectively. Using the calculated m_{avg} values for the standards, the % ^{13}C labeling was calculated using Equation 5.

Equation 3.5: $\%^{13}C = 100 - \left[\left(\frac{m_{avg}-m_{13C}}{m_{12C}-m_{13C}}\right) * 100\right]$

3.3 Results

3.3.1 Reductive Methylation and pH

The NMR spectra of the reductively ^{13}C-methylated lysozyme samples at pH 6.0 (low), 10.0 (high), and 7.5 (the optimum pH for the reducing agent, DMAB) are shown in Figure 3.1. For the pH 10.0 sample, the intensity of peaks 1 – 6 are nearly the same as the pH 7.5 sample, but peak 7 is missing (Figure 3.1c), indicating that the amino group corresponding to peak 7 was not modified with ^{13}C-methyl groups. The spectrum of the pH 6.0 sample (Figure 3.1a) shows no discernible preference for an amino group when compared to the pH 7.5 sample (Figure 3.1b).

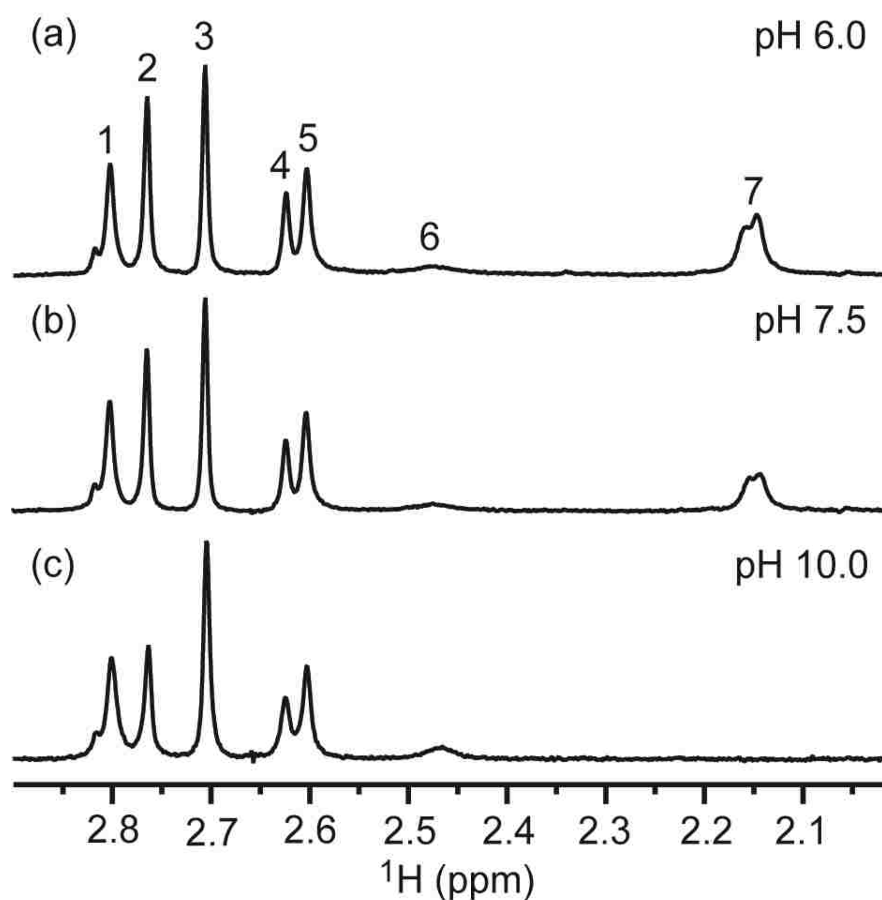


Figure 3.1. 1D HSQC NMR spectra obtained at 25°C and pH 8.5 of reductively ^{13}C-methylated lysozyme reacted at (a) pH 6.0, (b) pH 7.5, and (c) pH 10.0.

To better visualize the results of reductive ^{13}C-methylation at pH 6.0, a second experiment with 7 moles of reactive amine to 2 moles of ^{13}C-formaldehyde was performed without completing the methylation with natural abundance formaldehyde. As shown in Figure 2, the NMR spectra of this sample show both mono-methylamine (~33 ppm) and dimethylamine groups (~43 ppm). Only three of the seven reactive sites were dimethylated as indicated by the numbers on the spectra in Figure 3.2. Of those three, the amine corresponding to peak 7 was the most highly modified (Figure 3.2b).

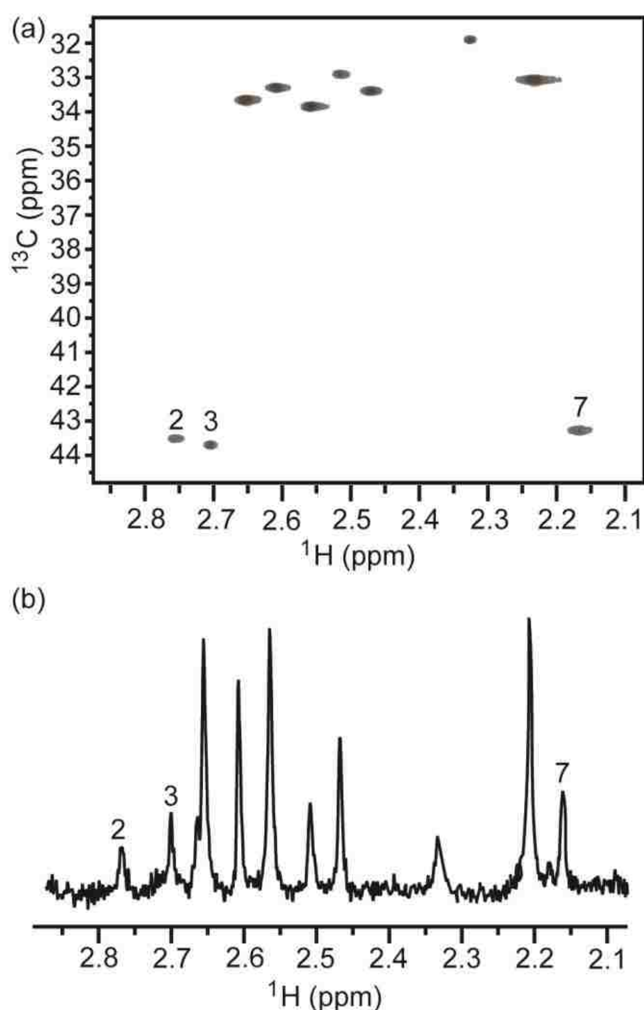


Figure 3.2. NMR spectra of lysozyme reductively ^{13}C-methylated with 7 moles of reactive amine to 2 moles of ^{13}C-formaldehyde. (a) 2D ^{1}H-^{13}C HSQC and (b) 1D ^{1}H-^{13}C HSQC.

The MALDI TOF-TOF MS analysis of the partially, reductively ^{13}C-methylated pH 6.0 sample shows two peptides corresponding to the *N*-terminus of the protein. The peptide with a mass-to-charge ratio (*m/z*) of 634 corresponds to the *N*-terminal peptide ($K(CH_3)_2VFGR$) with two of the four methyl groups, and the peptide with 662 *m/z* corresponds to the peptide with all four methyl groups ($K(CH_3)_4VFGR$). The percent ^{13}C incorporation of the protein was calculated by both NMR and MS as described in the methods section. The percentage of ^{13}C incorporation for peak 6 (NMR) and peptide 634 (Figure 3.3b) were calculated with values of 32.9% and 33.3%, respectively.

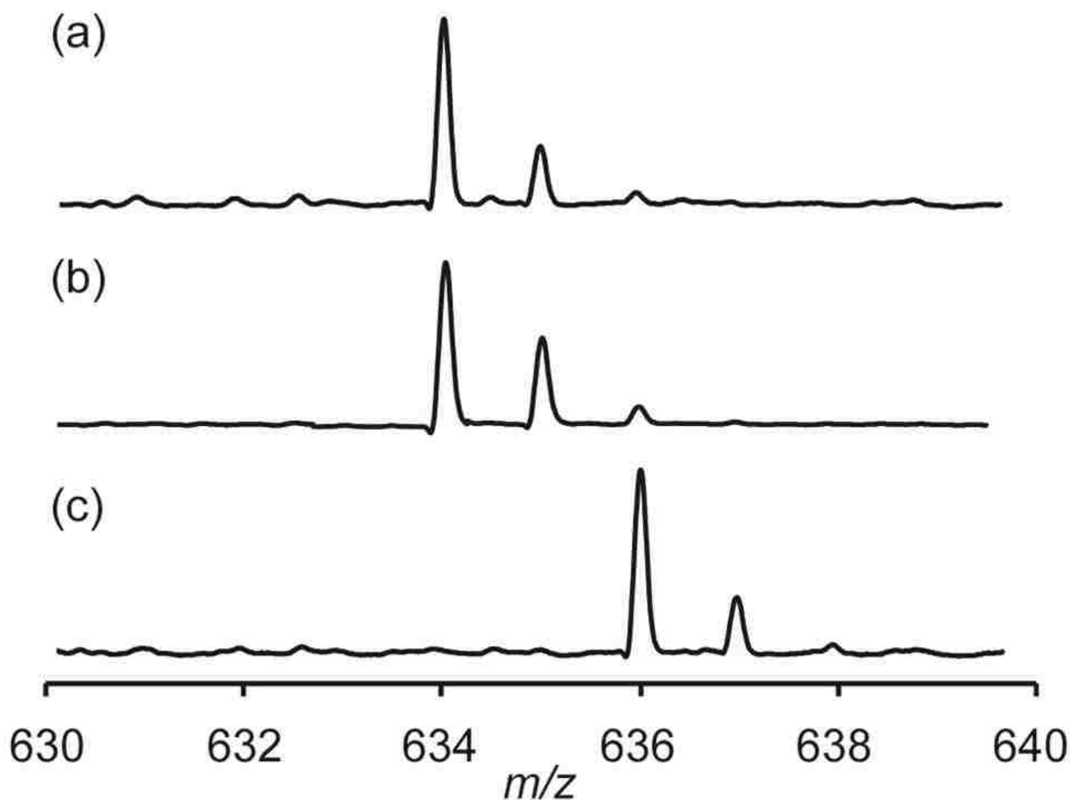


Figure 3.3. MALDI mass spectrometry isotopic profile of peptide 634 from lysozyme reductive methylated with (a) natural abundance formaldehyde, (b) sub-stoichiometric ^{13}C-formaldehyde followed by natural abundance formaldehyde, and (c) ^{13}C-formaldehyde.

3.4 Discussion

In solution, the amino groups exist in equilibrium between the free amine and the conjugate acid, which is tunable with pH. The reductive methylation reaction preferentially modifies the *N*-terminal α-amine at pH 6.0 due to the lower acid dissociation constant (pK_a) of the α-amine (~7)[17] versus the ε-amines (~10).[25] At pH 6.0 the equilibrium is drastically shifted to favor the conjugate acid; the α-amine has the highest concentration of free amine, the reactive form in the reductive methylation reaction. At pH 10.0, the equilibrium favors the free amine. Even though the α-amine with its lower pK_a exists nearly entirely in the free amine form, the ε-amines are stronger bases and more reactive. In the presence of limiting ^{13}C-formaldehyde, we are able to take advantage of the differing rates of reaction between the ε-amines and the α-amine. The weaker nucleophilic nature of the α-amine resulted in lower α-^{13}C-dimethylamine as is observed in Figure 3.1c for peak 7. By using a correlation experiment (^{1}H-^{13}C HSQC), we were able to resolve the resonances without denaturing the protein such as the protocol used by Gerken *et al.*

The high pH sample was also used to identify the NMR peak corresponding to the side-chain ε-dimethylamine of Lys1. Since the *N*-terminal side-chain ε-amino group is selectively labeled over the α-amino group, the mass spectrum of the *N*-terminal tryptic peptide 634 was used to determine the percentage of ^{13}C-incorporation for the ε-dimethylamine (33.3%). This percentage corresponds to the percentage of ^{13}C-incorporation calculated from the peak area of peak 6 (32.9%) in the NMR spectrum (Figure 3.1c), identifying peak 6 as the ε-dimethylamine of Lys1.

3.5 Conclusions

Two methods are introduced to assign the NMR signals of the α-dimethylamine and the side chain ε-dimethylamine of an *N*-terminal lysine residue. The methods are applicable to

assigning signals from any *N*-terminal α-dimethylamine and a lysine ε-dimethylamine near the *N*-terminus. The low/high pH method can be applied without further manipulation of the sample when 2D ^{1}H-^{13}C NMR spectra provide resolved signals. Another advantage of the low/high pH method is the protein remains intact, minimizing the possibility of structure degradation. With the *N*-terminal α-amine and Lys1 side-chain ε-amine identified in the NMR spectrum, further studies are currently being performed with partially ^{13}C-labeled protein to unambiguously assign the other dimethylamine resonances with their respective lysine residues.

3.6 Acknowledgements

This research was supported by Award Number R00RR024105 from the National Center For Research Resources, National Institute of Health.

3.7 References

1. Bax, A.; Grzesiek, S., Methodological advances in protein NMR. *Acc. Chem. Res.* **1993,** *26* (4), 131-138.

2. Shimba, N.; Yamada, N.; Yokoyama, K.; Suzuki, E., Enzymatic labeling of arbitrary proteins. *Anal. Biochem.* **2002,** *301*, 123-127.

3. Graslund S., Nordlund P., Weigelt J., Hallberg B.M., Bray J., Gileadi O., Knapp S., Oppermann U., Arrowsmith C., Hui R., Ming J., dhe-Paganon S., Park H.W., Savchenko A., Yee A., Edwards A., Vincentelli R., Cambillau C., Kim R., Kim S.H., Rao Z., Shi Y., Terwilliger T.C., Kim C.Y., Hung L.W., Waldo G.S., Peleg Y., Albeck S., Unger T., Dym O., Prilusky J., Sussman J.L., Stevens R.C., Lesley S.A., Wilson I.A., Joachimiak A., Collart F., Dementieva I., Donnelly M.I., Eschenfeldt W.H., Kim Y., Stols L., Wu R., Zhou M., Burley S.K., Emtage J.S., Sauder J.M., Thompson D., Bain K., Luz J., Gheyi T., Zhang F., Atwell S., Almo S.C., Bonanno J.B., Fiser A., Swaminathan S., Studier F.W., Chance M.R., Sali A., Acton T.B., Xiao R., Zhao L., Ma L.C., Hunt J.F., Tong L., Cunningham K., Inouye M., Anderson S., Janjua H., Shastry R., Ho C.K., Wang D., Wang H., Jiang M., Montelione G.T., Stuart D.I., Owens R.J., Daenke S., Schutz A., Heinemann U., Yokoyama S., Bussow K., Gunsalus K.C., Protein production and purification. *Nat Meth* **2008,** *5* (2), 135-146.

4. Goto, N. K.; Kay, L. E., New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr. Opin. Struct. Biol.* **2000,** *10*, 585-592.

5. Lian, L.; Middleton, D., Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001,** *39* (3), 171-190.

6. Mason, A.; Siarheyeva, A.; Haase, W.; Lorch, M.; van Veen, H.; Glaubitz, C., Amino acid type selective isotope labelling of the multidrug ABC transporter lMRA for solid-state NMR studies. *FEBS Lett.* **2004,** *568* (1-3), 117-121.

7. Chen, C.; Cheng, C.; Chen, Y.; Lee, J.; Chou, S.; Huang, W.; Chuang, W., Preparation of amino-acid-type selective isotope labeling of protein expressed in pichia pastoris. *Proteins:Struct., Funct., Bioinf.* **2006,** *62* (1), 279-287.

8. Means, G.; Feeney, R., Reductive alkylation of amino groups in proteins. *Biochemistry* **1968,** *7* (6), 2192-2201.

9. Walter, T. S.; Meier, C.; Assenberg, R.; Au, K. F.; Ren, J. S.; Verma, A.; Nettleship, J. E.; Owens, R. J.; Stuart, D. I.; Grimes, J. M., Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **2006,** *14*, 1617-1622.

10. Rypniewski, W.; Holden, H.; Rayment, I., Structural consequences of reductive methylation of lysine residues in hen egg-white lysozyme - an X-ray analysis at 1.8-angstrom resolution. *Biochemistry* **1993,** *32* (37), 9851-9858.

11. Brown, L.; Bradbury, J., Proton magnetic resonance studies of lysine residues of ribonuclease-A. *Eur. J. Biochem.* **1975,** *54* (1), 219-227.

12. Jentoft, J.; Gerken, T.; Jentoft, N.; Dearborn, D., [c-13] methylated ribonuclease-a - C-13 NMR studies of the interaction of lysine 41 with active site ligands. *J. Biol. Chem.* **1981,** *256* (1), 231-236.

13. Goux, W.; Teherani, J.; Sherry, A., Amine inversion in proteins - a C-13-NMR study of proton exchange and nitrogen inversion rates in n-epsilon,n-epsilon,n-alpha,n-alpha-[c-13]tetramethyllysine, n-epsilon,n-epsilon,n-alpha,n-alpha[c-13]tetramethyllysine methyl-ester, and reductively methylated concanavalin-a. *Biophys. Chem.* **1984,** *19* (4), 363-373.

14. Dick, L.; Sherry, A.; Newkirk, M.; Gray, D., Reductive methylation and C-13 NMR studies of the lysyl residues of fd gene 5 protein - lysines 24, 46, and 69 may be involved in nucleic acid binding. *J. Biol. Chem.* **1988,** *263* (35), 18864-18872.

15. Jentoft, J. E., Reductive methylation and C-13 nuclear-magnetic-resonance in structure-function studies of fc fragment and its subfragments. *Methods Enzymol.* **1991,** *203*, 261-295.

16. Sparks, D.; Phillips, M.; Lundkatz, S., The conformation of apolipoprotein A-1 in discoidal and spherical recombinant high density lipoprotein particles - C-13 NMR studies of lysine ionization behavior. *J. Biol. Chem.* **1992,** *267* (36), 25830-25838.

17. Gerken, T.; Jentoft, J.; Jentoft, N.; Dearborn, D., Intramolecular interactions of amino groups in C-13 reductively methylated hen egg-white lysozyme. *J. Biol. Chem.* **1982,** *257* (6), 2894-2900.

18. Zhang, M.; Vogel, H., Determination of the side chain pka values of the lysine residues in calmodulin. *J. Biol. Chem.* **1993,** *268* (30), 22420-22428.

19. Huque, M.; Vogel, H., C-13 NMR studies of the lysine side chains of calmodulin and its proteolytic fragments. *J. Protein Chem.* **1993,** *12* (6), 695-707.

20. Sherry, A.; Teherani, J., Physical studies of C-13-methylated concanavalin A - pH- and CO^{2+}-induced nuclear magnetic resonance shifts. *J Biol Chem* **1983,** *258* (14), 8663-8669.

21. Ashfield, J.; Meyers, T.; Lowne, D.; Varley, P.; Arnold, J.; Tan, P.; Yang, J.; Czaplewski, L.; Dudgeon, T.; Fisher, J., Chemical modification of a variant of human MIP-1 alpha; implications for dimer structure. *Protein Sci.* **2000,** *9* (10), 2047-2053.

22. Macnaughtan, M.; Kane, A.; Prestegard, J., Mass spectrometry assisted assignment of NMR resonances in reductively C-13-methylated proteins. *J. Am. Chem. Soc.* **2005,** *127* (50), 17626-17627.

23. Cordova, E.; Gao, J.; Whitesides, G. M., Noncovalent polycationic coatings for capillaries in capillary electrophoresis of proteins. *Anal. Chem.* **1997,** *69* (7), 1370-9.

24. Clauser, K. R.; Baker, P.; Burlingame, A. L., Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing ms or ms ms and database searching. *Anal. Chem.* **1999,** *71* (14), 2871-2882.

25. Bradbury, J. H.; Brown, L. R., Determination of dissociation constants of lysine residues of lysozyme by proton-magnetic-resonance spectroscopy. *Eur. J. Biochem.* **1973,** *40*, 565-576.

CHAPTER 4 SYNTHESIS AND APPLICATION OF PROTEIN MODIFIER AS A GLYCOSYLATION MIMETIC AND CROSS-LINKING AGENT

4.1 Introduction

Chemical modification is a formidable approach, based on the orthogonal reactivities of amino acids, used to manipulate and mimic protein structure and function. Understanding the relationship between protein structure and function has been the target for many biological studies today,[1-5] particularly in regards to their use in application-driven technologies. The advent of specialized reagents, advanced methods, and novel chemical reactions specific to certain amino acid side chains has driven this field from a restricted set of reactions with irrepressible results to a well-rounded arsenal of selective chemistries. Chemical modification approaches have not only contributed to protein structure and function, but to enzyme catalysis,[6-9] post-translational modification (PTM) mimetics,[10] protein cross-linking for protein-protein interaction studies,[11-13] analytical methods such as mass spectrometry,[14-17] and the development of new therapeutics.[18]

In nature, proteins are modified through a host of enzymes to induce either co- or post-translational modifications that are critical to protein activity. Fundamental studies have imparted the role of these enzymes in the *in vivo* chemical modification of target proteins and cooperative recombinant expression systems that allow for their large-scale production. However, the necessary post translational modification of glycosylated proteins (50 % of eukaryotic proteins) is not amenable in prokaryotic systems and prohibitively expensive in eukaryotic systems; requiring other means for production. To mimic the glycosylation PTM of proteins expressed in bacterium or other prokaryotic systems, glycan groups are incorporated on the target protein using reactive chemistries and site-selective strategies.[10,19] Typically, cysteine (-SH) and lysine ($-NH_2$) side chains are the targeted sites used to incorporate glycans on natural amino acids, but there are

research efforts exploiting the unconventional side chain functionalities of unnatural amino acids.[20] Site-selective glycosylation has been pursued using the following approaches: 1) synthesizing complex glycans before coupling to the protein;[21-22] 2) coupling a single glycan and using glycosyltransferases to produce complex, synthetic glycoproteins;[23] and 3) attaching covalent polymers, like poly(ethylene glycol), instead of a glycan.[24] Each of these approaches has facilitated studies to understand the effects of glycosylation in proteins, improved resistance to enzymatic degradation, increased protein solubility and/or bioavailability, and prolonged protein drug half-life, making therapeutics based upon glycosylated proteins more robust.

Chemical cross-linking of proteins has been essential in understanding protein-protein interactions and complex biological processes. Similar chemistries used to mimic PTMs at cysteine or lysine residues can be used to functionalize proteins at more than one site. This method allows multiple peptides and proteins to become "linked" using site-selective reagents and offers a versatile approach that directly identifies subsequent protein-protein interactions.[25] In fact, elucidating the locations of disulfide bonds at cystine residues, the major naturally occurring cross-linkage in proteins, has been an important step in solving tertiary structures in insulin, lysozyme, ribonuclease, and chymotrypsin.[25] Preservation of structure, stability, and functionality of proteins and enzymes has largely benefited from advances in cross-linking chemical modification.[26-29] However, chemical reagents to date are unilateral and only permit molecular incorporation or protein-protein binding at select amino acid sites, but are not able to do both. In this chapter, we present the synthesis of a protein modifier and preliminary data on its application as a glycosylation mimetic and a cross-linking agent

4.2 Materials and Methods

Diacetone-D-glucose, sodium hydride, carbon disulfide, iodomethane, tributyltin hydride, sodium periodate, imidazole, tetrahydrofuran, toluene, methanol, acetic acid, ethyl acetate, hexanes, petroleum ether, acetonitrile, sodium bicarbonate, magnesium sulfate, and chloroform were obtained from Sigma-aldrich. All solvents were used as received without further distillation.

The NMR ^{1}H and ^{13}C spectra were collected on a Bruker DPX-250 at 250 and 62.5 MHz respectively. ^{1}H spectra were collected with a relaxation delay of 5 seconds, 4 scans, and a 6000 Hz spectral width. ^{13}C spectra were collected with a relaxation delay of 5 seconds, 1000 scans, and a 15723 Hz spectral width. 2D NMR spectra were collected on a Varian 700 MHz equipped with a 5 mm-HCN-5922 probe. 2D NMR spectra were collected with a relaxation delay of 5s, 256 scans, and a 4529 spectral width. TMS was used as a standard for the ^{1}H and ^{13}C spectra and DSS was used for 2D experiments.

Mass spectrometry data was acquired using a Bruker UltrafleXtreme MALDI-TOF-TOF instrument equipped with a Nd:YAG in linear mode, with 89-92% laser energy, and an average of 4 x 500 shots. The spectra were smoothed using the Savitzky-Golay algorithm, the baseline was subtracted, and the apex peak was manually selected using the Bruker FlexAnalysis software.

4.2.1 Experimental

4.2.2 Coupling the protein modifier to Lysozyme

To an aqueous solution of lysozyme (5 mg/mL), DMAB (6.13 μL of 1 M) was added, followed by protein modifier (3.5 mg, 20.2 μmol). The mixture was shaken at 4°C for 4 hours. A second aliquot of DMAB was added, and the mixture was shaken at 4°C for an additional 24 hours. DMAB (3.06 μL, 1 M) was then added, and mixture was shaken at 4°C overnight for a total of 48 hours.

Note: Reactions with 1:10, 1:5 and 1:1 molar ratios produced little to no modification, in decreasing order.

4.2.3 Synthesis of 1,2-O-isopropylidene-α-D-glucofuranose-5-carbaldehyde

1,2:5,6-Di-O-isopropylidene-3-O-(S-methylxanthate)- α -D-glucofuranose 2: Diacetone-D-glucofuranose (4.08 g, 7.68 mmol), imidazole (4.0 mg, 0.0282 mmol), tetrahydrofuran (80 mL), and a magnetic stir bar were added to a 500 – mL round-bottom flask. The reaction vessel was flushed with nitrogen and a nitrogen atmosphere was maintained during the following steps. 60 % Sodium hydride dispersion (1.23 g, 51.2 mmol) was added over time at 0 °C. Vigorous bubbling was observed and reaction mixture turned cloudy. After the reaction was stirred for 30 min, carbon disulfide (2.80 mL/3.55 g, 46.6 mmol) was added. The reaction mixture turned yellow. The reaction was stirred for 45 min, and then iodomethane (1.92 mL/4.37 g, 30.8 mmol) was added. The reaction mixture turned orange. The reaction mixture was stirred for an additional 20 min. Glacial acetic acid was added drop wise until vigorous bubbling ceased. The solution was filtered and concentrated on a rotary evaporator. The residue was dissolved in 100 mL of ethyl acetate and washed with two, 50 mL portions of saturated sodium bicarbonate. The ether solution was dried over magnesium sulfate, and the solvent was removed by rotary evaporation. The resulting orange syrup was purified using a flash column (hexane-ethyl acetate 10:1) to give **2** as a yellow syrup; ^{1}H NMR (250 MHz, $CDCl_3$): δ 5.89 - 5.90 (m, 2H), 4.67 (d, *J* = 3.82, 1H), 4.43 – 4.19 (m, 2H), 4.17 – 3.93 (m, 2H), 2.58 (s, 3H), 1.53 (s, 3H), 1.41 (s, 3H), 1.31 (d, *J* = 2.6 Hz, 6H); ^{13}C NMR (62.5 MHz, $CDCl_3$): δ 214.64, 112.28, 109.22, 104.90, 84.08, 82.65, 79.61, 72.23, 66.81, 26.68, 26.52, 26.11, 25.12, 19.18; ESI-MS: *m/z* 351 [M + H^+]

3-Deoxy-1,2:5,6-di-O-isopropylidene-α-D-glucofuranose 3: A 100 mL round-bottom flask was equipped with a magnetic stir bar and a condenser to which a bubbler was attached to maintain

a nitrogen atmosphere. Tributyltin hydride (1.49 g, 5.13 mmol) in 10 mL of anhydrous toluene was added. Xanthate (1.0 g, 2.85 mmol) in 10 mL of toluene was added over time. The reaction mixture was heated at reflux until TLC analysis indicated the disappearance of starting materials (approximately 7 h). The reaction solution changed from deep yellow to nearly colorless. Toluene was removed by rotary evaporation and the oily residue was partitioned between 100 mL portions of petroleum ether and acetonitrile. The acetonitrile layer was separated and washed with three, 50 mL portions of petroleum ether and then concentrated on a rotary evaporator. The yellow oil was taken up in hexane-ethyl acetate (10:1) and filtered. The filtrate was concentrated and the oil was purified using a flash column (hexane-ethyl acetate 3:1) to give (450 mg, 65 %) **3** as a colorless syrup; ^{1}H NMR (250 MHz, $CDCl_3$): δ 5.80 (d, *J* = 3.7 Hz, 1H), 4.74 (dd, *J* = 4.8, 3.7 Hz, 1H), 4.37 – 3.91 (m, 4H), 3.80 (td, *J* = 4.3, 2.5 Hz, 1H), 2.17 (dd, *J* = 13.4, 4.0 Hz, 1H), 1.94 – 1.67 (m, 1H), 1.49 (s, 3H), 1.40 (s, 3H), 1.34 (s, 3H), 1.30 (s, 3H); ^{13}C NMR (62.5 MHz, $CDCl_3$): δ 111.15, 109.49, 105.44, 80.27, 78.47, 76.64, 67.02, 35.06, 26.61, 26.31, 25.98, 25.01; ESI-MS: *m/z* 245 $[M + H^+]$.

3-Deoxy-1,2-O-isopropylidene-α-D-glucofuranose 4: To a stirred solution of 3-deoxyglucofuranose (200 mg, 0.819 mmol) in methanol (2.8 mL) was added 0.8 % H_2SO_4 solution (700 µL). After completion of the reaction, a saturated $NaHCO_3$ (3 mL) solution was added after which methanol was removed under reduced pressure. It was extracted into chloroform (5 x 10 mL) and dried over anhydrous $MgSO_4$. Solvent was removed under vacuum, and the residue was purified by flash column chromatography (hexane/ethyl acetate = 1:9) to afford (139 mg, 83 % yield) **4** as a white solid; ^{1}H NMR (250 MHz, $CDCl_3$): δ 5.78 (d, J = 3.7 Hz, 1H), 4.73 (t, *J* = 4.2 Hz, 1H), 4.20 (dt, *J* = 10.8, 4.4 Hz, 1H), 3.98 – 3.79 (m, 1H), 3.77 – 3.46 (m, 2H), 2.79 (s, 1H), 2.51 (s, 1H), 2.04 (dd, *J* = 13.4, 4.5 Hz, 1H), 1.82 (ddd, *J* = 13.5, 10.7, 4.7 Hz, 1H), 1.49 (s, 3H),

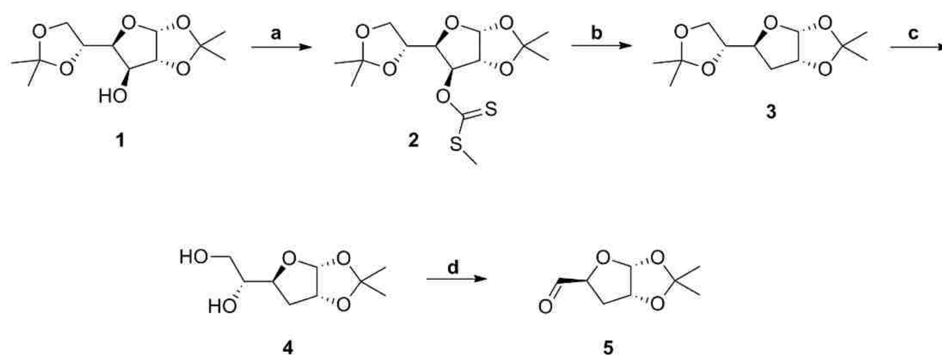
1.30 (s, 3H); ^{13}C NMR (62.5 MHz, $CDCl_3$): δ 111.20, 105.00, 80.41, 78.44, 71.91, 63.36, 33.49, 26.57, 25.95; ESI-MS: *m/z* 227 [M + H^+].

1,2-O-isopropylidene-α-D-glucofuranose-5-carbaldehyde 5: $NaIO_4$ (131 mg, 0.613 mmol) was dissolved in hot water (2.25 mL). To a stirred solution of the diol (50 mg, 0.245 mmol) in ethanol (520 µL) was added the $NaIO_4$ solution. After completion of the reaction, ethanol was added dropwise until no more precipitate formed. The solution was filtered and the ethanol was evaporated. The mixture was extracted into chloroform (5 X 50 mL). The organic layer was dried over magnesium sulfate and the solvent was removed by rotary evaporation. The residue was purified by flash chromatography to give (17 mg, 40 % yield) **5**; ^{1}H NMR (250 MHz, $CDCl_3$): δ 9.78 (d, J = 2.0 Hz, 1H), 6.03 (d, J = 3.6 Hz, 1H), 5.97 – 5.81 (m, 1H), 4.84 (dt, J = 15.7, 4.2 Hz, 1H), 2.04 (dd, *J* = 13.4, 4.5 Hz, 1H), 1.82 (ddd, *J* = 13.5, 10.7, 4.7 Hz, 1H), 1.49 (s, 3H), 1.30 (s, 3H); ^{13}C NMR (62.5 MHz, $CDCl_3$): δ 203.76, 112.71, 104.11, 83.32, 79.47, 35.77, 27.34, 25.38; ESI-MS: *m/z* 173 [M + H^+].

4.3 Results and Discussion

4.3.1 Synthesis of 1,2-O-isopropylidene-3-deoxy-5-pentafuranal

Synthesis of 1,2-*O*-isopropylidene-3-deoxy-5-pentafuranal was achieved with an initial Barton-McCombie deoxygenation of diacetone-D-glucofuranose **1** (Scheme 4.1). The secondary alcohol of **1** was converted to an S-methyl dithiocarbonate (xanthate) of **2**. Deoxygenation was completed by reacting **2** with tributylstannane via a radical process to give the 3-deoxy-glucofuranose **3**. When stirred at room temperature in 0.8 % sulfuric acid, the 3-deoxy-glucofuranose **3** was converted to the diol **4** without hydrolysis of the sterically hindered acetonide. A subsequent oxidation of the diol **4** yielded the aldehyde **5**.



Scheme 4.1. Reagents and conditions: (a) NaH, CS_2, MeI, anhydrous THF, quantitative yield; (b) Bu_3SnH, toluene, reflux, 65%; (c) 0.8% aqueous H_2SO_4, MeOH, rt, 83%; (d) $NaIO_4$, EtOH, H_2O, 40%

4.3.2 Coupling the Modifier to Lysozyme

Hen egg-white lysozyme was modified with the aldehyde **5** via reductive alkylation. Reaction times of the reductive alkylation reaction are directly correlated to the size of the carbonyl compound, in this case the aldehyde. Reductive alkylation of lysozyme with formaldehyde, the smallest aldehyde, is reacted between 18 to 24 hours at 4 °C with dimethylamine borane to completely modify the protein, so reaction times exceeding 24 hours are expected for this reaction. The completeness of the reaction is heavily dependent on the reducing agent and concentration of the aldehyde. In addition to reducing the Schiff base, reducing agents can reduce the aldehyde to an alcohol in a competing reaction. Stronger reducing agents have a higher rate of the competing reaction. Reacting the protein with 20 molar excess of the aldehyde (compared to reactive amines in the protein) for five days at 4 °C with dimethylamine borane yielded the highest degree of modification as shown by MALDI-TOF analysis (Figure 4.1a). The mass-to-charge (*m/*z) ratio of 16162.9 corresponds to 10 aldehydes being coupled to lysozyme, which has 7 reactive amines; therefore, some sites have multiple aldehydes coupled to the lysine side chain amine. Reactions with smaller aldehyde to protein

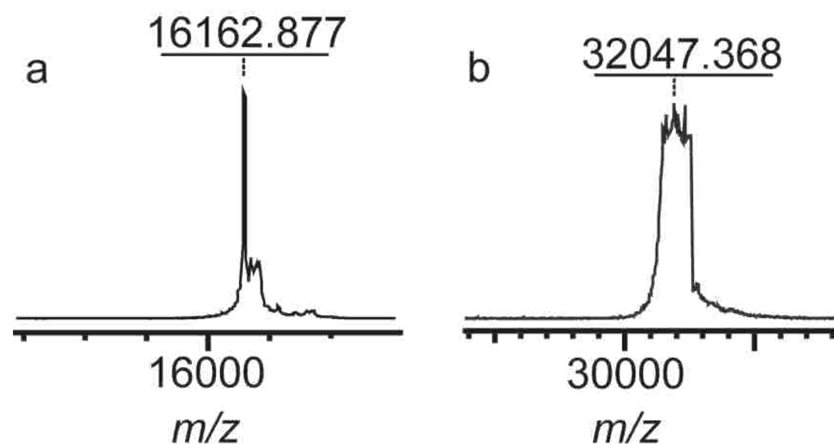


Figure 4.1. MALDI-TOF analysis of a) protein modifier coupled to hen egg white lysozyme and b) hen egg white lysozyme cross-linked with protein modifier.

reactive amine ratios produce minimum to no modification. The protein sample was then analyzed by NMR spectroscopy. 2D ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) NMR spectrum of the modified protein was compared to the spectrum of unmodified lysozyme (Figure 4.2). Overlay of the two spectra showed the appearance of two new peaks in the methyl region of the spectra corresponding to the acetonide methyls of the protein modifier, confirming its addition to lysozyme. The new peaks integrated to a 1:1 ratio with the other isolated methyl peaks in the spectrum, which conflicts with the MALDI data. Integration of the peaks and similar chemical shifts suggest they are from the same modification, in which case there should be 9 more pairs according to the MALDI spectrum. Although conflicting, NMR and MALDI analysis revealed modification of the protein. Additionally, overlap of the NMR signals of modified and unmodified lysozyme suggested minimum structural perturbations due to the modification.

4.3.3 Glycosylation Mimetic

Application of the protein modifier as a glycosylation mimetic requires the removal of the acetonide protecting group of the 1,2 diol of the aldehyde. Once the protecting group is removed from the small molecule, a ring opening step occurs that establishes an equilibrium between the

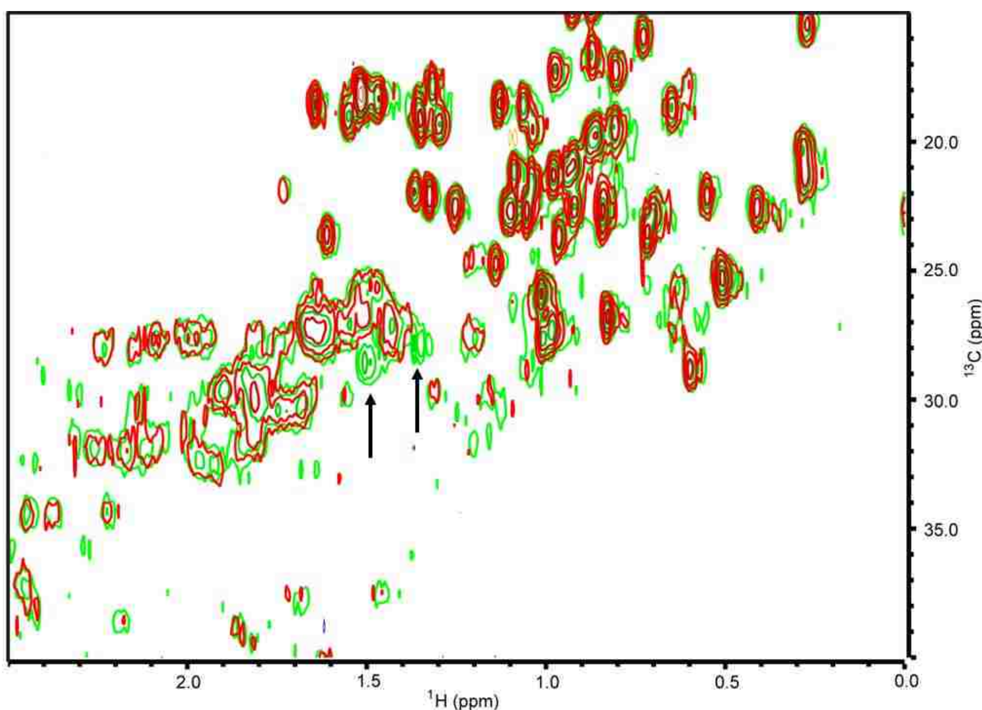
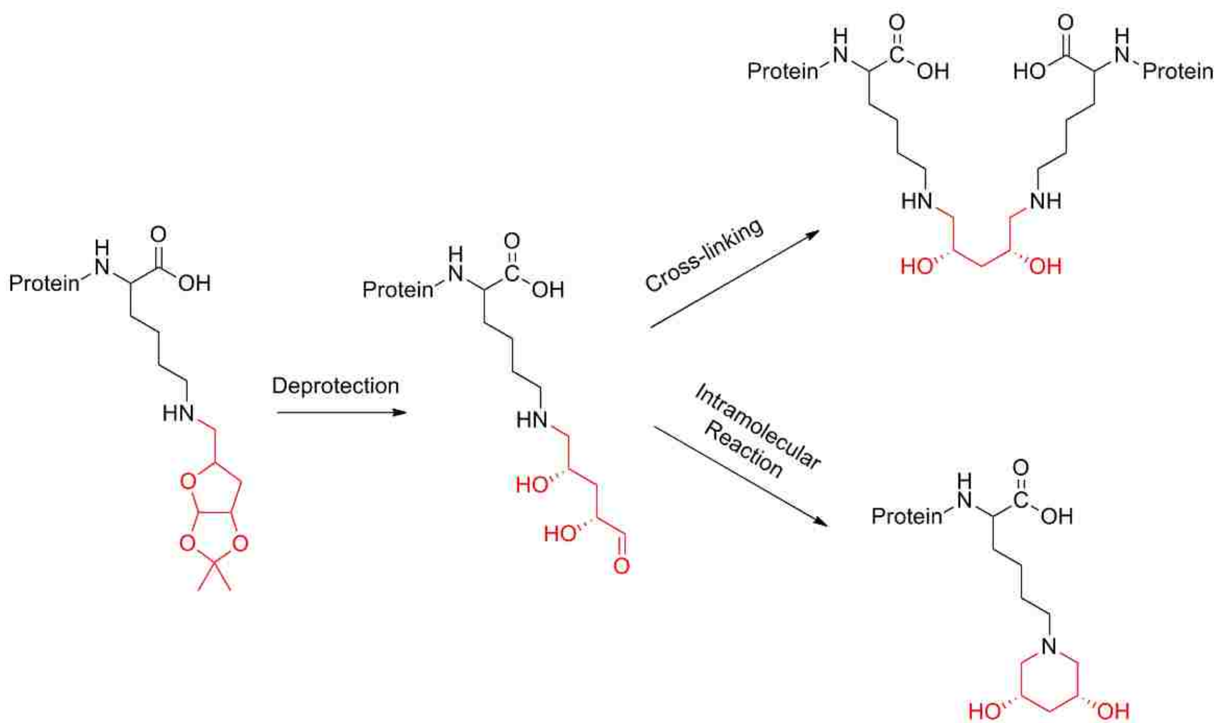


Figure 4.2. Overlay of 2D ^{1}H-^{13}C HSQC spectra of (red) native lysozyme and (green) lysozyme coupled to the protein modifier.

furanose ring and the straight chain glucose derivative, which has a terminal aldehyde. Subsequent intramolecular reductive alkylation will give the desired cyclic product for glycosylation mimetics. However, there is the possibility of an intermolecular reductive alkylation with another lysozyme molecule (Scheme 4.2). The rate of hydrolysis of the 1,2-*O*-alkylidene-α-glucofuranose (the parent compound of the aldehyde derivative) has a half-life of 20 hours in dilute sulfuric acid.[30] The pH of those conditions is less than 1, which is extremely harsh for the protein and would cause denaturation. However, lysozyme can survive at pH 4.0 for 1 week at 4 °C before denaturation. The protein sample was diluted to 2.5 mg/mL in phosphate buffer at pH 4.0 and stirred at 55 °C for five days for deprotection. The methyl region of the ^{1}H-^{13}C HSQC NMR spectrum was analyzed to confirm the removal of the acetonide and showed a decreased intensity of the two methyl resonances previously associated to the



Scheme 4.2. Schematic of the second alkylation reaction for cross-linking and intramolecular reaction.

protecting group methyls (Figure 4.3). The sample was also analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) to access how much of the protein had cross-linked (Figure 4.4). Analysis of the protein gel reveals a faint band around 30 kDa for cross-linked protein and a significantly more intense band around 16 kDa for the modified protein.

4.3.4 Protein Cross-linking

To date, there is not a single compound used as both a glycomimetic and a protein cross-linking agent. Experiments were conducted to assess the possibility of this glucose derivative performing as both. Whether the intramolecular or intermolecular reaction occurs depends on the conditions immediately following the removal of the acetonide protecting group. The protein sample was concentrated to 10 mg/mL before the deprotection step, increasing the proximity of

protein molecules to induce cross-linking. Upon completion of the deprotection, the pH was adjusted to neutral and the sample was stirred at 4°C for two days. Initial attempts to cross-link

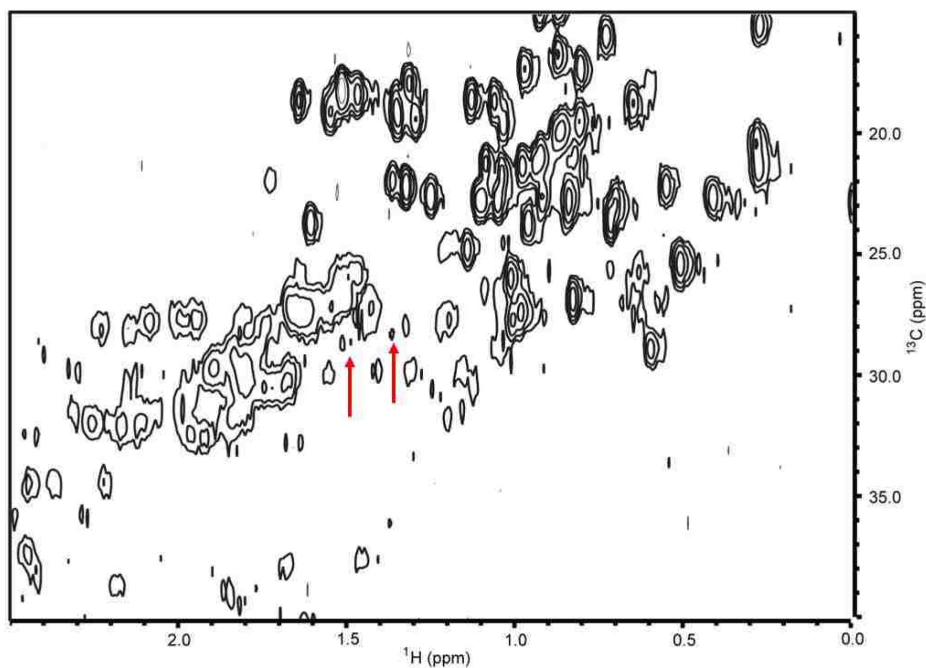


Figure 4.3. 2D ^{1}H-^{13}C HSQC of modified lysozyme after removal of the acetonide protecting group.

lysozyme were not successful. However, one experiment, produced a higher order species as analyzed by MALDI-TOF (Figure 4.1b). The shape of the cross-linked protein peak in the mass spectrum was odd with spikes and a width of 1500 mass units. The *m/z* values of the spikes in the peak appear random and do not clearly correspond to different correspond to units of modification. The sample was also analyzed by SDS-PAGE revealing a band around 30 kDa, corresponding to cross-linked dimer (Figure 4.4) with 20 total protein modifiers coupled. It must be noted that the solvent dried up during the deprotection of the sample that produced the higher order species. These results taken together suggest that the intermolecular cross-linking reaction is not very efficient.

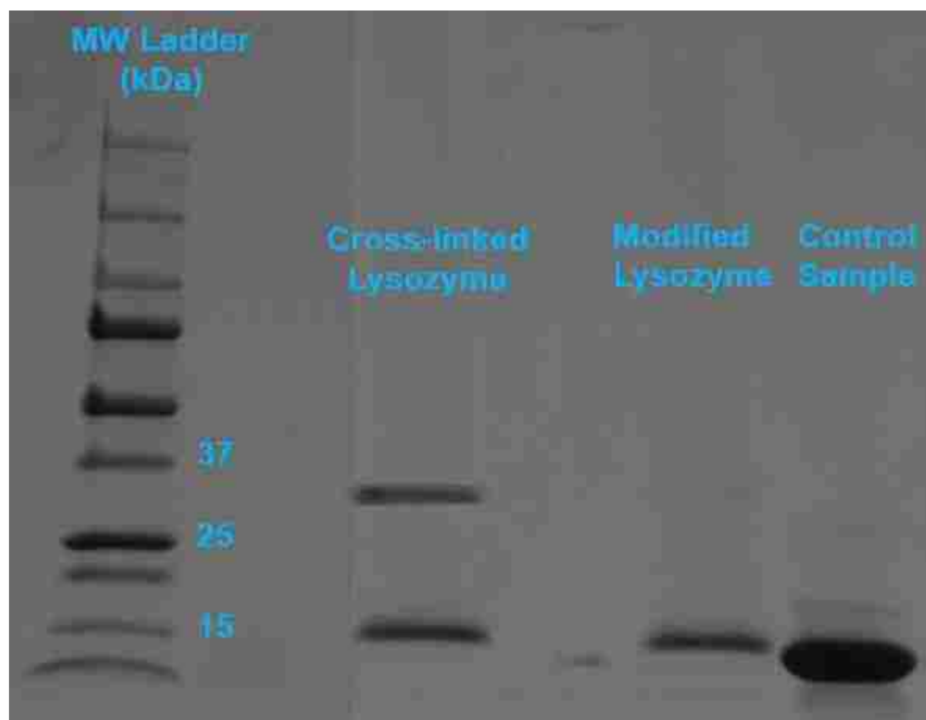


Figure 4.4. 1D gel protein profile of unmodified lysozyme (lane 4), modified lysozyme after the second reductive alkylation reaction (lane 3), and the cross-linked lysozyme sample (lane 2).

4.4 Conclusions

We have presented the synthesis of 1,2-*O*-isopropylidene-3-deoxy-5-pentafuranal and preliminary results of its use to modify proteins. The protein modifier has demonstrated potential as a glycosylation mimetic. The intramolecular reaction of the same amine creates a cyclic modification on the lysine side chain and/or *N*-terminal amine. The cyclic modification of proteins with this reagent mimics the structure of a glycan and is a potential site for elongation of complex glycans through glycosyltransferases. Mimicking the structure of sugar molecules may have an added benefit over current glycomimetics (i.e. polyethylene glycol) of lower toxicity. This modification also has potential to be used as a modification of protein drugs to prevent premature enzymatic digestion and novel incorporation of iminosugars in proteins.

Successful coupling of the modifier to lysozyme was demonstrated. However, the reaction with 20 molar excess of the modifier produced a less than optimal lysozyme sample (7 reactive amines) with 10 modifications. Ideally, a single alkylation at any reactive site of the protein is desired to accommodate the subsequent intramolecular reaction of the same amine for application as a glycosylation mimetic.

The intramolecular reaction was shown to occur without significant competition from intermolecular cross-linking. While intermolecular cross-linking was not an issue, it is possible that intramolecular cross-linking occurred between two lysine residues rather than the cyclization reaction. Further analysis is necessary to demonstrate that the targeted intramolecular, cyclization reaction occurred and the aldehyde acts as a glycosylation mimetics.

Intermolecular cross-linking with this reagent was not efficient. Although a cross-linked dimer was created, it was most likely facilitated by the evaporation of the solvent during the reaction; previous attempts, in which the solvent had not evaporated, failed to produce cross-linked protein. Production of the dimer with evaporated solvent was also inefficient as analyzed by protein gel electrophoresis, with an equally intense band around 16 kDa for modified protein and no higher order oligomers observed. Other conditions are necessary if the presented protein modifier is to be used as a cross-linking agent.

4.5 References

1. Bacot-Davis, V. R.; Palmenberg, A. C., Encephalomyocarditis virus leader protein hinge domain is responsible for interactions with ran gtpase. *Virology* **2013,** *443* (1), 177-185.

2. Chan, K. L.; Gligoris, T.; Upcher, W.; Kato, Y.; Shirahige, K.; Nasmyth, K.; Beckouet, F., Pds5 promotes and protects cohesin acetylation. *Proc. Natl. Acad. Sci. U S A* **2013,** *110* (32), 13020-13025.

3. Sasaki, S.; Kobayashi, M.; Futagi, Y.; Ogura, J.; Yamaguchi, H.; Takahashi, N.; Iseki, K., Crucial residue involved in l-lactate recognition by human monocarboxylate transporter 4 (HMCT4). *PLoS One* **2013,** *8* (7).

4. Wasmuth, A.; Ludwig, C.; Mootz, H. D., Structure-activity studies on the upstream splice junction of a semisynthetic intein. *Bioorg. Med. Chem.* **2013,** *21* (12), 3495-3503.

5. Zhou, Y.; Kierzek, E.; Loo, Z. P.; Antonio, M.; Yau, Y. H.; Chuah, Y. W.; Geifman-Shochat, S.; Kierzek, R.; Chen, G., Recognition of rna duplexes by chemically modified triplex-forming oligonucleotides. *Nucleic Acids Res.* **2013,** *41* (13), 6664-6673.

6. Gomez, A.; Cardoso, C.; Genta, F. A.; Terra, W. R.; Ferreira, C., Active site characterization and molecular cloning of tenebrio molitor midgut trehalase and comments on their insect homologs. *Insect Biochem. Mol. Biol.* **2013,** *43* (8), 768-780.

7. Real-Guerra, R.; Carlini, C. R.; Staniscuaski, F., Role of lysine and acidic amino acid residues on the insecticidal activity of jackbean urease. *Toxicon.* **2013,** *71* (1), 76-83.

8. Rodrigues, R. C.; Berenguer-Murcia, A.; Fernandez-Lafuente, R., Coupling chemical modification and immobilization to improve the catalytic performance of enzymes. *Adv. Synth. Catal.* **2011,** *353* (13), 2216-2238.

9. Cowan, D. A.; Fernandez-Lafuente, R., Enhancing the functional properties of thermophilic enzymes by chemical modification and immobilization. *Enzyme Microb. Technol.* **2011,** *49* (4), 326-346.

10. van Kasteren, S. I.; Garnier, P.; Davis, B. G., Chemical methods for mimicking post-translational modifications. In *Protein Engineering*, Köhrer, C.; RajBhandary, U., Eds. Springer Berlin Heidelberg: 2009; Vol. 22, pp 97-125.

11. Arenas, A.; Lopez-Alarcon, C.; Kogan, M.; Lissi, E.; Davies, M. J.; Silva, E., Chemical modification of lysozyme, glucose 6-phosphate dehydrogenase, and bovine eye lens proteins induced by peroxyl radicals: Role of oxidizable amino acid residues. *Chem. Res. Toxicol.* **2013,** *26* (1), 67-77.

12. Bhardwaj, A.; Myers, M. P.; Buratti, E.; Baralle, F. E., Characterizing tdp-43 interaction with its rna targets. *Nucleic Acids Res.* **2013,** *41* (9), 5062-5074.

13. Ismaya, W. T.; Hasan, K.; Kardi, I.; Zainuri, A.; Rahmawaty, R. I.; Permanahadi, S.; El Viera, B. V.; Harinanto, G.; Gaffar, S.; Natalia, D.; Subroto, T.; Soemitro, S., Chemical modification of saccharomycopsis fibuligera r64 alpha-amylase to improve its stability against thermal, chelator, and proteolytic inactivation. *Appl. Biochem. Biotechnol.* **2013,** *170* (1), 44-57.

14. Abraham, P.; Adams, R. M.; Tuskan, G. A.; Hettich, R. L., Moving away from the reference genome: Evaluating a peptide sequencing tagging approach for single amino acid polymorphism identifications in the genus populus. *J. Proteome Res.* **2013,** *12* (8), 3642-3651.

15. Junkova, P.; Vermachova, M.; Prchal, J.; Kuckova, S.; Hrabal, R.; Hynek, R., Improved approach for the labeling of arginine, glutamic, and aspartic acid side chains in proteins using chromatographic techniques. *J. Liq. Chromatogr. Relat. Technol.* **2013,** *36* (9), 1221-1230.

16. Pecova, M.; Sebela, M.; Markova, Z.; Polakova, K.; Cuda, J.; Safarova, K.; Zboril, R., Thermostable trypsin conjugates immobilized to biogenic magnetite show a high operational stability and remarkable reusability for protein digestion. *Nanotechnology* **2013,** *24* (12).

17. Samgina, T. Y.; Vorontsov, E. A.; Gorshkov, V. A.; Artemenko, K. A.; Zubarev, R. A.; Ytterberg, J. A.; Lebedev, A. T., Collision-induced dissociation fragmentation inside disulfide c-terminal loops of natural non-tryptic peptides. *J. Am. Soc. Mass Spectrom.* **2013,** *24* (7), 1037-1044.

18. Cohen, A.; Wang, E.; Chisholm, K. A.; Kostyleva, R.; O'Connor-McCourt, M.; Pinto, D. M., A mass spectrometry-based plasma protein panel targeting the tumor microenvironment in patients with breast cancer. *J. Proteomics* **2013,** *81*, 135-147.

19. Davis, B. G., Mimicking posttranslational modifications of proteins. *Science* **2004,** *303* (5657), 480-482.

20. Liu, H. T.; Wang, L.; Brock, A.; Wong, C. H.; Schultz, P. G., A method for the generation of glycoprotein mimetics. *J. Am. Chem. Soc.* **2003,** *125* (7), 1702-1703.

21. Andre, S.; Unverzagt, C.; Kojima, S.; Dong, X.; Fink, C.; Kayser, K.; Gabius, H. J., Neoglycoproteins with the synthetic complex biantennary nonasaccharide or its alpha 2,3/alpha 2,6-sialylated derivatives: Their preparation, assessment of their ligand properties for purified lectins, for tumor cells in vitro, and in tissue sections, and their biodistribution in tumor-bearing mice. *Bioconjugate Chem.* **1997,** *8* (6), 845-855.

22. Andre, S.; Unverzagt, C.; Kojima, S.; Frank, M.; Seifert, J.; Fink, C.; Kayser, K.; von der Lieth, C. W.; Gabius, H. J., Determination of modulation of ligand properties of synthetic complex-type biantennary *N*-glycans by introduction of bisecting glcnac in silico, in vitro and in vivo. *Eur. J. Biochem.* **2004,** *271* (1), 118-134.

23. Cornish, V. W.; Hahn, K. M.; Schultz, P. G., Site-specific protein modification using a ketone handle. *J. Am. Chem. Soc.* **1996,** *118* (34), 8150-8151.

24. Harris, J. M.; Chess, R. B., Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discovery* **2003,** *2* (3), 214-221.

25. Wold, F., In vivo chemical modification of proteins (post-translational modification). *Annu. Rev. Biochem.* **1981,** *50* (1), 783-814.

26. Wong, S. S.; Wong, L. J. C., Chemical cross-linking and the stabilization of proteins and enzymes. *Enzyme Microb. Technol.* **1992,** *14* (11), 866-874.

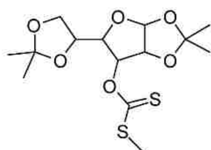
27. DeSantis, G.; Jones, J. B., Chemical modification of enzymes for enhanced functionality. *Curr. Opin. Biotechnol.* **1999,** *10* (4), 324-330.

28. Govardhan, C. P., Crosslinking of enzymes for improved stability and performance. *Curr. Opin. Biotechnol.* **1999,** *10* (4), 331-335.

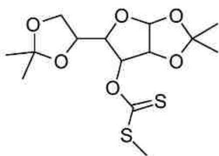
29. Tyagi, R.; Gupta, M. N., Chemical modification and chemical cross-linking for protein/enzyme stabilization. *Biochemistry-Moscow* **1998,** *63* (3), 334-344.

30. Vanheeswijk, W. A. R.; Goedhart, J. B.; Vliegenthart, J. F. G., Acid-catalyzed hydrolysis of 1,2-*O*-alkylidene-alpha-D-glucofuranoses. *Carbohydr. Res.* **1977,** *58* (2), 337-344.

4.6 Spectra



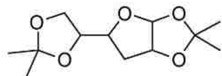
1,2:5,6-Di-O-isopropylidene-3-O-(S-methylxanthate)- α -D-glucofuranose - ^{1}H NMR in $CDCl_3$ at 250 MHz



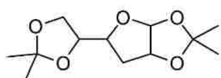
220 200 180 160 140 120 100 80 60 40 20 0

(^{13}C ppm)

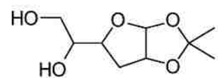
1,2:5,6-Di-O-isopropylidene-3-O-(S-methylxanthate)- α -D-glucofuranose - ^{13}C NMR in $CDCl_3$ at 62.5 MHz



3-Deoxy-1,2:5,6-di-O-isopropylidene-α-D-glucofuranose - ^{1}H NMR in $CDCl_3$ at 250 MHz



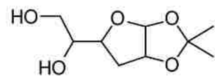
3-Deoxy-1,2:5,6-di-O-isopropylidene-α-D-glucofuranose - ^{13}C NMR in $CDCl_3$ at 62.5 MHz



7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0

(^{1}H ppm)

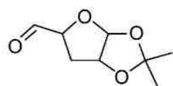
3-Deoxy-1,2-O-isopropylidene-α-D-glucofuranose - ^{1}H NMR in $CDCl_3$ at 250 MHz



110 100 90 80 70 60 50 40 30 20 10 0

(^{13}C ppm)

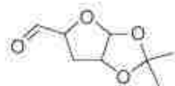
3-Deoxy-1,2-O-isopropylidene- α -D-glucofuranose - ^{13}C NMR in $CDCl_3$ at 62.5 MHz



10.0 9.0 8.0 7.0 6.0 5.0 4.0 3.0 2.0 1.0 0.0

(^{1}H ppm)

1,2-O-isopropylidene- α -D-glucofuranose-5-carbaldehyde - ^{1}H NMR in $CDCl_3$ at 250 MHz



220 200 180 160 140 120 100 80 60 40 20 0

(^{13}C ppm)

1,2-O-isopropylidene-α-D-glucofuranose-5-carbaldehyde - ^{13}C NMR in $CDCl_3$ at 62.5 MHz

CHAPTER 5 CONCLUSIONS AND FUTURE STUDIES

5.1 Concluding Remarks

Orthogonal chemical modifications are the key to completely understanding the roles of protein function in biological activity. Typically, this is accomplished through X-ray crystallography and activity assays. However, those methods do not provide an in-solution perspective of protein structure or direct correlation to change in activity, respectively. NMR spectroscopy offers a means to provide an in-solution snapshot of proteins, in which specific amino acid residues can be correlated to structural and functional perturbations. In this dissertation, we used hen egg white lysozyme as a model protein to study in depth reductive alkylation of lysine residues to improve NMR methodologies for protein structural studies.

In Chapter 2, the development of chemical modification methodologies for ultimate protein structural studies were presented. Crown ether (18C6) was used as a reagent to selectively alter the reaction rates of lysine side chain amines, however, the crown ether only showed effects in concentrations of excess and had a universal effect on all sites. A second method utilized the varying strengths and hydrophobicities of different reducing agents to alter the rates of the side chain amines. Although proven effective, the degeneracy in ^{13}C incorporation was not overcome to assign lysine dimethylamino NMR resonances. The third method presented in this chapter aimed to develop a non-destructive Edman degradation to cleave the *N*-terminal amino acid for identification and assignment of dimethylamino resonances of lysine 1. Cleavage of the *N*-terminal residue without denaturation of the protein proved elusive, thereby, rendering this method inoperable.

In Chapter 3, we presented a method to identify and assign both the α- and ε-dimethylamino NMR resonances of the *N*-terminal lysine residue of hen egg white lysozyme. This method was

based on a pH induced chemoselectivity of the reductive methylation reaction, in which the α-amine reacted at a faster rate than the side chain ε-amines at pH 6.0 and vice versa at pH 10.0. This method made it possible to obtain an individual isotope incorporation measurement for the ε-amine of an *N*-terminal lysine residue, which had previously never been accomplished.

In Chapter 4, the synthesis and application of 1,2-O-isopropylidene-3-deoxy-5-pentafuranal as a protein modifier was discussed. Preliminary results show the potential of this individual reagent as a glycomimetic; however, its use as a cross-linking agent does not seem viable.

5.2 Future Studies

To date, many approaches to chemically modify proteins for structural and functional gains have been used. We have successfully developed a method to identify and assign *N*-terminal dimethylamino NMR resonances as well as synthesized a protein modifier with dual application. However, this work only scratches the surface of the larger problems they target.

The need to make absolute assignments of NMR resonances without extensive prior knowledge of the target protein makes the *N*-terminal assignment strategy valuable. However, more work is needed to overcome the degenerate isotope incorporation problem with the mass spectrometry assisted assignment strategy. Degenerate modification necessitates advancements in altering the rates of reaction at selective sites. Therefore, reductive methylation in the presence of ligands may be suitable for this effort. In this manner, lysine residues shielded by the bound ligand or structural perturbations due to binding are not modified as extensively as those not affected by the binding of the ligand.

The presentation of the protein modifier clearly demonstrated application as a protein modifier. However, future work on the optimization of the coupling of the modifier to the protein is necessary. The reaction of the lysozyme and aldehyde at a 1:20 molar ratio, respectively,

produced a sample with 10 modifications. Lysozyme has 7 reactive sites, 6 lysines and the α-amine. This extent of modification is not optimal for use as a glycosylation mimetic. Ideally, the protein should have no more than 7 modifications to facilitate the cyclization reaction for the glycosylation mimicking modification. Also, future studies on the characterization of the modified product is needed for confidence that the desired modification occurred. MS/MS may be the best option for mapping the location of modification of the protein. The utility of this reagent as a protein modifier can be realized with the accomplishment of these future studies.

APPENDIX: LETTERS OF PERMISSION

RightsLink®

Home | Account Info | Help

ACS Publications
High quality. High impact.

Title: Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: An x-ray analysis at 1.8-.ANG. resolution

Author: Wojciech R. Rypniewski, Hazel M. Holden, and Ivan Rayment

Publication: Biochemistry

Publisher: American Chemical Society

Date: Sep 1, 1993

Copyright © 1993, American Chemical Society

Logged in as:
Kevin Roberson
Account #:
3000758358

LOGOUT

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK | CLOSE WINDOW

Copyright © 2014 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

ELSEVIER LICENSE
TERMS AND CONDITIONS

Mar 12, 2014

This is a License Agreement between Kevin J Roberson ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Kevin J Roberson
Customer address	418 Choppin Hall
	Baton Rouge, LA 70803
License number	3346311229340
License date	Mar 12, 2014
Licensed content publisher	Elsevier
Licensed content publication	Elsevier Books
Licensed content title	Methods in Enzymology, Volume 208
Licensed content author	Kathleen S. Matthews,Artemis E. Chakerian,Joseph A. Gardner
Licensed content date	1991
Number of pages	29
Start Page	468
End Page	496
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier chapter?	No
Will you be translating?	No
Title of your thesis/dissertation	Reductive Alkylation of Proteins towards Biological Applications
Expected completion date	May 2014
Estimated size (number of pages)	130
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD

VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. Byclicking "accept" in connection with completing this licensingtransaction, you agree that the following terms and conditions apply to thistransaction (along with the Billing and Payment terms and conditionsestablished by Copyright Clearance Center, Inc. ("CCC"), at the timethat you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementionedmaterial subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example,figures) has appeared in our publication with credit or acknowledgement toanother source, permission must also be sought from that source. If suchpermission is not obtained then that material may not be included in yourpublication/copies. Suitable acknowledgement to the source must be made, eitheras a footnote or in a reference list at the end of your publication, asfollows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title ofarticle / title of chapter, Pages No., Copyright (Year), with permission fromElsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit -"Reprinted from The Lancet, Vol. number, Author(s), Title of article, PagesNo., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media forwhich permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures andillustrations may be altered/adapted minimally to serve your work. Any otherabbreviations, additions, deletions and/or any other alterations shall be madeonly with prior written authorization of Elsevier Ltd. (Please contact Elsevierat permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived inthis instance, please be advised that your future requests for Elseviermaterials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specificallygranted in the combination of (i) the license details provided by you andaccepted in the course of this licensing transaction, (ii) these terms andconditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rightslicensed immediately upon issuance of the license at the end of the licensingprocess for the transaction, provided that you have disclosed complete andaccurate details of your proposed use, no license is finally effective unlessand until full payment is received from you (either by publisher or by CCC) asprovided in CCC's Billing and Payment terms and conditions. If

VITA

Kevin J. Roberson earned his B.S. in chemistry from Georgia Southern University in May 2007. Kevin began his career as a researcher under Dr. Karen Welch investigating the ligand binding sites of C*oncanavalin A* through molecular modeling. From this work, Kevin wrote and secured research funding from the College of Science and Technology at Georgia Southern University. In August 2007, Kevin began his graduate studies as an organic chemist under Dr. Robert Hammer studying the synthesis and incorporation of unnatural amino acids in proteins by recombinant expression. In 2009, he joined Dr. Megan Macnaughtan's research group investigating reductive alkylation of proteins towards biological applications.

Kevin Roberson is a member of National Organization for the Professional Advancement of Black Chemists and Chemical Engineers and the American Chemical Society. In his spare time, he served as Director of the K-12 tutoring program and Co-Director of the Introduction to Chemistry Program at Rose Hill Baptist Church and Donaldsonville Baptist Church, respectively. Additionally, he has served as a mentor through the STEM NOLA Program by mentoring the scientific development of K-12 minority students in the Greater New Orleans area. Kevin is now pursuing opportunities for the next phase of his young career and has recently began starting his own chemistry research firm. Kevin will graduate with the degree of Doctor of Philosophy in chemistry from Louisiana State University in May 2014.