Summer 2014

MASS SPECTROMETRY BASED EXPERIMENTAL STRATEGIES TO CHARACTERIZE NATIVE AND NON-NATIVE DISULFIDE BONDS IN CYSTEINE-RICH PROTEIN THERAPEUTICS

Adriana Z. Kita
University of Massachusetts - Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2
Part of the Analytical Chemistry Commons

Recommended Citation
Kita, Adriana Z., "MASS SPECTROMETRY BASED EXPERIMENTAL STRATEGIES TO CHARACTERIZE NATIVE AND NON-NATIVE DISULFIDE BONDS IN CYSTEINE-RICH PROTEIN THERAPEUTICS" (2014). Doctoral Dissertations. 221.
https://scholarworks.umass.edu/dissertations_2/221

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
MASS SPECTROMETRY BASED EXPERIMENTAL STRATEGIES TO
CHARACTERIZE NATIVE AND NON-NATIVE DISULFIDE BONDS IN
CYSTEINE-RICH PROTEIN THERAPEUTICS

A Dissertation Presented

by

ADRIANA ZELEDON KITA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2014

Molecular and Cellular Biology
MASS SPECTROMETRY BASED EXPERIMENTAL STRATEGIES TO CHARACTERIZE NATIVE AND NON-NATIVE DISULFIDE BONDS IN CYSTEINE-RICH PROTEIN THERAPEUTICS

A Dissertation Presented

by

ADRIANA ZELEDON KITA

Approved as to style and content by:

______________________________
Igor A. Kaltashov, Chair

______________________________
Richard W. Vachet, Member

______________________________
Stephen J. Eyles, Member

______________________________
Scott C. Garman, Member

______________________________
Barbara A. Osborne, Director
Molecular and Cellular Biology
DEDICATION

To my husband Daniel who has been there for me with love and encouragement

To my daughter Sofia

To Mom and Dad
ACKNOWLEDGMENTS

I would like to thank my advisor, Igor Katashov, for his support and guidance throughout my graduate studies. I have learned a tremendous amount of knowledge about mass spectrometry and protein therapeutics with his guidance. I would also like to thank my committee members for their scientific wisdom and the work they put in to helping me with my graduate degree.

I truly enjoyed working for Igor and with all the people in the Kaltashov lab. Their presence made the lab environment very enjoyable. I would like to thank Rinat Abzalimov and Cedric Bobst who helped me learn about the multiple mass spectrometers in the lab. I would also like to thank Virginie Sjoelund who taught me some fundamentals of cellular biology at the beginning of my graduate studies at UMass. Guanbo, Shuhai, and Burcu have been there through a significant portion of my time at UMass and I would like to thank them for making the lab a truly enjoyable place to work.

I would also like to thank my family. Dan Kita who has been with me through the entire graduate degree process, and I would truly like to thank him for all his love, support, and encouragement. I would also like to thank Sofia, Mom, and Dad for always being there for me when I needed a little extra encouragement. Thank you all!
ABSTRACT

MASS SPECTROMETRY BASED EXPERIMENTAL STRATEGIES TO CHARACTERIZE NATIVE AND NON-NATIVE DISULFIDE BONDS IN CYSTEINE-RICH PROTEIN THERAPEUTICS

SEPTEMBER 2014

ADRIANA ZELEDON KITA, B.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Igor A. Kaltashov

The impact post-translational modifications (PTMs) can have on the structure, function, and immunogenicity of protein therapeutics makes it especially important for these protein-based treatments to be well characterized. Mass spectrometry has become instrumental in the examination of enzymatic and non-enzymatic PTMs. Disulfide bonds fall into both of these categories in which native disulfide bonds are formed in the ER by disulfide mediated enzymes and non-native disulfide bonds are often formed by non-enzymatic reducing/oxidizing reactions. Disulfide bonds are particularly important for protein folding and reinforcing higher order structure, and are typically characterized by LC-MS of non-reduced peptides. However, characterizing the disulfide connectivity can be challenging when cysteine residues lie in close proximity within the primary sequence. Here we developed a dual proteolytic method with several gas-phase fragmentation techniques to map the cysteine-rich N-terminus of the protein therapeutic β-glucocerebrosidase (GCase). We used this approach to map the native disulfide
connectivity of GCase and also identified non-native disulfide bonds in a long-term stability sample.

Investigating non-native disulfide bonds can be challenging because they often exist at low levels and fully oxidized isoforms do not exhibit a change in molecular weight. Here we used lysozyme (LYZ) as a model to develop a rapid characterization strategy to monitor non-native disulfide conformers by electrospray ionization (ESI) MS. We demonstrate that this technique can be used to monitor large-scale conformational changes that often accompany disulfide scrambling. Initially, LYZ was subjected to disulfide scrambling and their disulfides were mapped. Then the charge state distribution of each scrambled species was monitored by ESI-MS. We show that we can distinguish non-native conformers from natively oxidized LYZ by comparing their extent and distribution of protonation during ESI-MS.

Here we demonstrate the application of mass spectrometry based experimental strategies that can be used to monitor large-scale conformational changes and characterize challenging native and non-native disulfide bonds in proteins. These methods are intended to improve the strategies that are currently used to characterize disulfide bonds in protein therapeutics and further detect non-native conformers.
# TABLE OF CONTENTS

| CHAPTER |
|-----------------|----------------------------------|
| 1. INTRODUCTION | 1.1 Recombinant DNA technology... |
|                 | 1.2 Insulin: The First Recombinant Biologic |
|                 | 1.3 Human Growth Hormone          |
|                 | 1.4 Gaucher Disease               |
|                 | 1.5 Enzyme Replacement Therapy for Gaucher Disease |
|                 | 1.6 Future Treatments for Neuronopathic Forms of Gaucher Disease: |
|                 | 1.7 Regulation of Protein Therapeutics and Biosimilars |
|                 | 1.8 Introduction to Mass Spectrometry |
|                 | 1.9 Characterization of Protein Therapeutics by Mass Spectrometry: |
|                 | 1.10 Cysteine Modifications in Protein Therapeutics |
|                 | 2. MAPPING DISULFIDES WITHIN THE CYSTEINE-RICH REGION OF THE |
|                 | PROTEIN THERAPEUTIC BETA GLUCOCEREBROSIDASE |
|                 | 2.1 Abstract                      |
|                 | 2.2 Introduction                  |
|                 | 2.3 Materials and Methods         |

| TABLE OF CONTENTS |
|-------------------|-----------------|
| ACKNOWLEDGMENTS   | vii              |
| ABSTRACT          | vi               |
| LIST OF FIGURES   | xi               |

| CHAPTER |
|-----------------|----------------------------------|
| 1. INTRODUCTION | 1.1 Recombinant DNA technology... |
|                 | 1.2 Insulin: The First Recombinant Biologic |
|                 | 1.3 Human Growth Hormone          |
|                 | 1.4 Gaucher Disease               |
|                 | 1.5 Enzyme Replacement Therapy for Gaucher Disease |
|                 | 1.6 Future Treatments for Neuronopathic Forms of Gaucher Disease: |
|                 | 1.7 Regulation of Protein Therapeutics and Biosimilars |
|                 | 1.8 Introduction to Mass Spectrometry |
|                 | 1.9 Characterization of Protein Therapeutics by Mass Spectrometry: |
|                 | 1.10 Cysteine Modifications in Protein Therapeutics |
|                 | 2. MAPPING DISULFIDES WITHIN THE CYSTEINE-RICH REGION OF THE |
|                 | PROTEIN THERAPEUTIC BETA GLUCOCEREBROSIDASE |
|                 | 2.1 Abstract                      |
|                 | 2.2 Introduction                  |
|                 | 2.3 Materials and Methods         |

| TABLE OF CONTENTS |
|-------------------|-----------------|
| ACKNOWLEDGMENTS   | vii              |
| ABSTRACT          | vi               |
| LIST OF FIGURES   | xi               |

## CHAPTER 1. INTRODUCTION

### Subsection 1.1 Recombinant DNA technology

### Subsection 1.2 Insulin: The First Recombinant Biologic

### Subsection 1.3 Human Growth Hormone

### Subsection 1.4 Gaucher Disease

### Subsection 1.5 Enzyme Replacement Therapy for Gaucher Disease

### Subsection 1.6 Future Treatments for Neuronopathic Forms of Gaucher Disease: Substrate Reduction Therapy

### Subsection 1.7 Regulation of Protein Therapeutics and Biosimilars

### Subsection 1.8 Introduction to Mass Spectrometry

### Subsection 1.9 Characterization of Protein Therapeutics by Mass Spectrometry:

#### Subsubsection 1.9.1 Intact Mass and Top-Down MS

#### Subsubsection 1.9.2 Bottom-up MS and Peptide Mapping by LC-MS

#### Subsubsection 1.9.3 Subunit and Middle-Down Analysis of mAbs by MS

#### Subsubsection 1.9.4 Glycan Analysis by MS

#### Subsubsection 1.9.5 Structural Analysis of Protein Therapeutics by MS

### Subsection 1.10 Cysteine Modifications in Protein Therapeutics

#### Subsubsection 1.10.1 Free Cysteines

#### Subsubsection 1.10.2 Trisulfides and Cysteinylation

#### Subsubsection 1.10.3 Alternate Disulfide Isoforms

#### Subsubsection 1.10.4 Disulfide Scrambling

## CHAPTER 2. MAPPING DISULFIDES WITHIN THE CYSTEINE-RICH REGION OF THE PROTEIN THERAPEUTIC BETA GLUCOCEREBROSIDASE

### Subsection 2.1 Abstract

### Subsection 2.2 Introduction

### Subsection 2.3 Materials and Methods

#### Subsubsection 2.3.1 Endoproteinase Lys-C digest of GCase

#### Subsubsection 2.3.2 LC-MS identification of monoglycosylated and deglycosylated L1L2 GCase peptide

#### Subsubsection 2.3.3 RP-HPLC purification of the monoglycosylated and deglycosylated L1L2 peptide
2.3.4 Proteolytic digest of purified L1L2 with pepsin and Asp-N.............27
2.3.5 Asp-N digest of intact GCase .........................................................29
2.3.6 Thermolysin digest of GCase A1 N-terminal peptide .....................30
2.4 Results and Discussion ........................................................................30
2.4.1 Glycosylation at N19 prevents disulfide bond characterization of L1L2..........................................................30
2.4.2 CAD fragmentation of deglycosylated L1L2 occurs mainly on the C-terminus .................................................................34
2.4.3 Sequence coverage of the L1L2 C-terminus by pepsin ....................36
2.4.4 CAD of Lys-C/Asp-N digest reveals the native disulfide connectivity of GCase .................................................................37
2.4.5 Asp-N digestion and LC-MS reveal partially reduced species of A1 .................................................................42
2.4.6 All disulfide configurations within the cysteine-rich region of GCase can be probed by digesting A1 with thermolysin ...............44
2.4.7 Characterizing native and non-native disulfide bonds in a GCase stability sample .................................................................45
2.5 Conclusions ..........................................................................................48

3. MASS SPECTROMETRY BASED METHODS FOR DETECTION AND CHARACTERIZATION OF PROTEINS WITH NON-NATIVE DISULFIDE BONDS.................................................................52
3.1 Abstract ..............................................................................................52
3.2 Introduction ..........................................................................................53
3.3 Materials and Methods .........................................................................56
3.3.1 Materials .......................................................................................56
3.3.2 Circular Dichroism .......................................................................57
3.3.3 Disulfide Scrambling .....................................................................57
3.3.4 Disulfide Mapping of LYZ ............................................................58
3.3.5 ESI-MS .........................................................................................59
3.4 Results and Discussion ......................................................................59
3.4.1 The Disulfide Bonds in LYZ Impose Conformational Constraints under Denaturing Conditions .................................................59
3.4.2 The Production, Purification, and Characterization of Non-Native LYZ Isoforms .................................................................63
3.4.3 Changes in the Charge State Distribution of Scrambled LYZ Variants Observed by ESI-MS .................................................................68
3.4.4 Disulfide Scrambling and ESI-MS of the Protein Therapeutic IFNβ-1a .........................................................................................71
3.4.5 ESI-MS of GCase Stability Sample .............................................73
3.4.6 Disulfide Scrambling of GCase ....................................................76
3.5 Conclusions ..........................................................................................77

4. CONCLUSIONS AND FUTURE DIRECTIONS..............................................82
4.1 Conclusions ..........................................................................................82
4.2 Structural Analysis of Disulfide Scrambled Lysozyme .........................84
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2. 1</td>
<td>Beta-glucocerebrosidase (GCase) primary and tertiary structure.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 2. 2</td>
<td>Nomenclature for N-terminal peptide of GCase digested with Lys-C.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2. 3</td>
<td>Nomenclature for natively oxidized N-terminal peptides generated by a Lys-C/Asp-N digest.</td>
<td>28</td>
</tr>
<tr>
<td>Figure 2. 4</td>
<td>Natively oxidized N-terminal peptide generated by an Asp-N digest of GCase.</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2. 5</td>
<td>Nomenclature for native and scrambled species of A1 thermolysin digest.</td>
<td>30</td>
</tr>
<tr>
<td>Figure 2. 6</td>
<td>Identification of monoglycosylated N-terminal peptide by LC-MS.</td>
<td>31</td>
</tr>
<tr>
<td>Figure 2. 7</td>
<td>Purification of monoglycosylated N-terminal peptide.</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2. 8</td>
<td>Purification of GCase L1L2 N-terminal peptide.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2. 9</td>
<td>CAD fragmentation of L1L2 peptide.</td>
<td>36</td>
</tr>
<tr>
<td>Figure 2. 10</td>
<td>Mass spectrum of the identified peptic peptides from the L1L2 pepsin digest.</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2. 11</td>
<td>ECD of GCase peptide trimer L1A1A2.</td>
<td>40</td>
</tr>
<tr>
<td>Figure 2. 12</td>
<td>CAD of GCase peptide dimer L1A2.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2. 13</td>
<td>All possible disulfide patterns in GCase N-terminal segment.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2. 14</td>
<td>Asp-N digestion efficiency is low in disulfide containing region.</td>
<td>42</td>
</tr>
<tr>
<td>Figure 2. 15</td>
<td>GCase peptide map of an Asp-N digest.</td>
<td>43</td>
</tr>
<tr>
<td>Figure 2. 16</td>
<td>GCase A1 thermolysin digest.</td>
<td>45</td>
</tr>
<tr>
<td>Figure 2. 17</td>
<td>Partial reduction observed in GCase stability and control.</td>
<td>47</td>
</tr>
<tr>
<td>Figure 2. 18</td>
<td>nanoLC-MS of GCase control and stability A1 thermolysin digest.</td>
<td>48</td>
</tr>
</tbody>
</table>
Figure 3.1 Charge state distribution of natively oxidized and reduced LYZ ...........62

Figure 3.2 Promoting disulfide scrambling with thermal and chemical
denaturation of LYZ. ........................................................................................................64

Figure 3.3 Characterizing LYZ disulfide scrambled isoforms .........................67

Figure 3.4 Charge state distribution of LYZ disulfide scrambled isomers ..........70

Figure 3.5 Crystal structure of IFNβ-1a (1AU1) with cysteine residues shown
in blue................................................................................................................................71

Figure 3.6 RP-HPLC purification of IFNβ-1a disulfide scrambling ....................72

Figure 3.7 Charge state distribution of RP-HPLC fractions of IFNβ-1a ..........73

Figure 3.8 GCase charge state distribution of stressed and stability samples ..........75

Figure 3.9 All possible disulfide scrambled species in GCase N-terminal
region ..................................................................................................................................76

Figure 3.10 GCase incubated in conditions favorable for disulfide scrambling.....77

Figure 3.11 A representation of LYZ structurally distorted by disulfide
scrambling..............................................................................................................................79

Figure A.1 Mass spectrometry analysis of GALNS with cysteine and
formylglycine.........................................................................................................................89
CHAPTER 1
INTRODUCTION

1.1 Recombinant DNA technology

The use of medicinal compounds from natural sources, such as microorganisms, plants, and animals has and continues to be exploited to develop treatments for diverse diseases. However, modern drug manufacturing processes have limited the need to extract and purify these compounds from their natural sources due to the improved production and safety from recombinant DNA technology. Paul Berg and colleagues revolutionized protein and peptide drug development with the introduction of recombinant DNA technology and allowed protein-based drugs, or biologics, to be generated at industrial scales capable of supplying worldwide need (Jackson, Symons et al. 1972). This technology allows for the precise generation of genetically modified organisms that can express virtually any coding sequence of interest. In this way, proteins can be engineered for specific properties that may enhance their pharmacokinetic and pharmacodynamic properties to ultimately produce a safer and more effective drug.

1.2 Insulin: The First Recombinant Biologic

In 1982, the Food and Drug Administration (FDA) approved Eli Lilly’s recombinant human insulin produced in E. coli (Bliss 1993). As a result of it being the first biologic, it was met with increased scrutiny by regulatory agencies and the public. Despite the initial fears, recombinant insulin was found to be a safe and effective treatment for diabetes mellitus, which improved the lives of millions by allowing pure protein to be produced at high yield to treat patients with the widespread and debilitating
disease. Prior to the advent of this technology, insulin was purified from the pancreas of various animals and administered as daily injections to patients (Roth, Qureshi et al. 2012). This production strategy could not keep up with the medical need, in part, due to the demand for meat products that influenced the amount of pancreas tissue available to extract insulin. Furthermore, insulin derived from animal sources, particularly bovine, can result in immunogenicity (Wilson, Douglas et al. 1985; Schernthaner 1993). The approval of insulin as a recombinant protein therapeutic opened the door to use this technology for the production of biologics targeting diverse diseases.

1.3 Human Growth Hormone

An advantage of using host organisms to produce recombinant protein therapeutics is the increased amounts that these expression systems are capable of producing beyond that of natural sources. However, this strategy can also significantly increase the safety of these products because it avoids potential infectious agents that may be present in natural sources. This was particularly evident in the case of human growth hormone (hGH), which was originally purified from pooled pituitary glands of cadavers. As a result of contaminated hGH, some patients developed iatrogenic Creutzfeldt-Jakob’s disease (iCJD) and resulted in the worldwide withdrawal of this treatment option in 1985 (Dukes 1996; Ayyar 2011; Blizzard 2012). This was a devastating situation for patients with severe growth hormone deficiency, but was resolved soon thereafter with the accelerated approval of Genentech’s recombinant hGH by the FDA (Ayyar 2011).
1.4 Gaucher Disease

Similar to diabetes or hypopituitarism, all lysosomal storage diseases (LSDs) are a consequence of dysfunctional proteins. LSDs are rare metabolic disorders that result from inherited mutations, which disrupt the activity of enzymes in the lysosome. The lumen of this membrane-enclosed organelle exists at a pH of 5.0 and contains an array of acid hydrolase enzymes that are necessary to degrade biological components including protein, sugars, and lipids from intra- and extracellular origins (Brady 2006; Fujiwara, Kikuchi et al. 2013). These catabolic activities are necessary to ensure cellular homeostasis, and if disrupted can lead to the buildup of unprocessed substrates, which can impede cellular function and cause disease with varying physical manifestations and severity depending on the specific mutations involved.

Gaucher disease is the most prevalent LSD and results from inherited mutations that severely compromise or completely render inactive the β-glucocerebrosidase (GCase) enzyme (Brady, Kanfer et al. 1966; Brady 2006). GCase functions to breakdown glucoceramide or glucocerebroside (GC) into glucose and ceramide and when aberrant, results in the accumulation of GC particularly in the lysosomes of macrophages.

Gaucher disease is categorized based on its severity and clinical presentation. There are three major subclasses of the disease, which include Types I, II, and III. Type I is the most common and is defined by its lack of involvement of the central nervous system (non-neuronopathic). It affects major organs such as the liver and spleen and is also recognized by severe anemia and bone malformities (Weinreb, Charrow et al. 2002). Type II and III are neuronopathic and affect the central nervous system with varying degrees of severity. Patients with type II Gaucher disease exhibit acute symptoms early in
development and die in infancy. Type III is characterized by a later onset and slower
disease progression within the central nervous system (Jmoudiak and Futerman 2005).

The most direct use of protein therapeutics is to replace deficient or dysfunctional
proteins that cause the disease. In the 1960’s de Duve and Brady postulated that LSDs
might be treated with what is now termed enzyme replacement therapy (ERT), which is a
therapeutic strategy that provides a functional enzyme that compensates for the defunct
endogenous protein (Brady 2006; Desnick and Schuchman 2012). Indeed the first attempt
to provide ERT was to an infant with the LSD Sandhoff disease which results from
inherited mutations in the Hexosaminidase B (HEXB) gene (Johnson, Desnick et al. 1973;
Mahuran 1999). The trial consisted of purifying β-hexosaminidase from human urine and
injecting it intravenously into the patient. Unfortunately, the enzyme was unsuccessful in
translocating past the blood-brain barrier and could not revert the accumulation of GM2
ganglioside in neuronal cells (Johnson, Desnick et al. 1973; Mahuran 1999). It is
important to note that hydrolysis of the glycosphingolipid, globoside had occurred in
peripheral tissues with the addition of the exogenous protein. This demonstrated that the
injected enzyme could function in cells that were accessible. Soon thereafter, efforts were
focused on the treatment of the most common LSD, Gaucher disease.

1.5 Enzyme Replacement Therapy for Gaucher Disease

With efforts focused on GCase as a viable option to treat patients with Gaucher
disease, an enzyme source and purification method were pursued by Roscoe Brady and
colleagues at the NIH. Initial attempts that isolated GCase from human spleen, rat
intestine, and cow spleen yielded only partially purified enzyme and were insufficient for
ERT (Pentchev, Brady et al. 1973). Therefore, alternative options were explored that
focused on human-derived sources in order to reduce the likelihood of immunogenicity (Pentchev, Brady et al. 1973). From here, human placental tissue was chosen due to its availability and methods of purification were established (Pentchev, Brady et al. 1973; Furbish, Blair et al. 1977). Initial treatments with placental-derived GCase had promising results, with two patients showing diminished accumulation of GC in their livers and blood. Improved targeting of the enzyme to the lysosomes of macrophages was achieved by modifying its glycosylation pattern with exoglycosidases to expose mannose sugar moieties, which are recognized by mannose receptors on the macrophage cell surface (Brady, Pentchev et al. 1974; Stahl, Rodman et al. 1978; Furbish, Steer et al. 1981).

Henry Blair, a co-founder of Genzyme, had been producing GCase for Roscoe Brady at the NIH (Furbish, Blair et al. 1977; Deegan and Cox 2012). With the ability to purify GCase from human placental tissue at a large scale, Genzyme received approval by the FDA in 1991 to treat patients with type I Gaucher disease. Thus, GCase became the first protein therapeutic used for ERT. As was the case for insulin and hGH, this method could not provide enough protein to meet demand. However, recombinant GCase proved to be more challenging to produce because of the necessity for N-linked glycosylation to be present on the protein. This was addressed by expressing GCase in mammalian CHO cells that, unlike *E. coli*, provided the appropriate cellular machinery to produce this post-translational modification (PTM) (Kacher, Brumshtein et al. 2008; Deegan and Cox 2012). However, similar to the protein purified from placental tissues, treatment with exoglycosidases to expose core mannose residues is required to efficiently target the recombinant protein to macrophages (Deegan and Cox 2012). In 1994, GCase
expressed in CHO cells (Cerezyme®) was approved to treat type I Gaucher disease and replaced the need to purify GCase from placental tissue (Deegan and Cox 2012).

The development of recombinant GCase as a successful ERT led others to pursue similar approaches. Lysosomal enzymes were of particular interest due to their ability to internalize into macrophages and provide a targeted approach to treat LSDs. After the approval of Cerezyme®, two more GCase enzymes, VPRIV® from SHIRE Human Genetic Therapies and Elelyso® from Pfizer, were being developed as alternative options to treat type I Gaucher disease. These efforts were exceptionally important in 2009 when a viral contamination led to the shutdown of one of Genzyme’s manufacturing facilities that led to a shortage of Cerezyme® (Deroma, Sechi et al. 2013). VPRIV® was already in phase III trials at the time and was granted accelerated approval by the FDA to compensate for the shortage (Zimran 2011).

Unlike Cerezyme®, VPRIV® contains the full GCase amino acid sequence and is produced in human fibroblasts using gene-activated technology. An advantage of producing VPRIV® in the fibroblast cell line is that it bypasses the need for glycosidase treatment and adds the appropriate glycans required for macrophage internalization (Zimran, Brill-Almon et al. 2011; Deegan and Cox 2012). Similar to VPRIV®, Elelyso® does not require the additions of glycosidases and is more cost effective due to its production in genetically modified carrot cells (Deegan and Cox 2012). Currently, Cerezyme®, VPRIV®, and Elelyso® are all used as ERTs to treat type I Gaucher disease.
1.6 Future Treatments for Neuronopathic Forms of Gaucher Disease: Substrate Reduction Therapy

ERT has been successfully shown to treat patients with the nonneuronal form of Gaucher disease (type I); however, the ability for the enzyme to alleviate symptoms associated with the nervous system has been unsuccessful and is mainly due to the protein’s inability to pass the blood-brain barrier (Weinreb, Charrow et al. 2002). Therefore, other non-protein based therapeutic strategies must be explored. Currently, Zavesca® is the only small-molecule that is approved for the treatment of type I Gaucher disease which blocks the activity of glucosylceramide synthase and therefore inhibits the production of GC. This treatment is known as substrate reduction therapy (Ficicioglu 2008). Zavesca® is an orally available drug has the potential to cross the blood-brain barrier and some initial studies have shown that in combination with ERT, Zavesca® may be beneficial to help treat patients with neuronopathic forms of Gaucher disease (Capablo, Franco et al. 2007).

1.7 Regulation of Protein Therapeutics and Biosimilars

Unlike Zavesca® and other small molecules, protein therapeutics are exponentially more complex to characterize due to their size and heterogeneity. Furthermore, small molecules can be synthesized, whereas protein therapeutics are produced in living cells making it more likely to have impurities in the product. These factors make it a long and challenging road for both the pharmaceutical companies and regulatory agencies to monitor these processes, which are major contributors to the lengthy process of going from development to approval.
Monoclonal antibodies are the most common class of protein therapeutics and present many challenges to produce and characterize. A typical antibody has a molecular weight (MW) of about 150 kDa and contains a variety of post-translational modifications (PTMs), which are sensitive to changes in cell culture conditions or the expression system. Therefore, it is critical to maintain strict control over the manufacturing process such that the final protein product remains constant.

It is the responsibility of regulatory agencies to assure that the final product has been shown to be consistent, safe, and effective. However, with the patent lifetimes of many protein therapeutics nearing an end, regulatory agencies are faced with a new challenge in monitoring the safety and efficacy of “generic” versions of protein therapeutics. For small molecules, generics can typically bypass preclinical and clinical trials as long as they meet the criteria for bioequivalence. The FDA defines a generic as “one that is comparable to an innovator drug product in dosage form, strength, route of administration, quality, performance characteristics and intended use (2009). The issue of “generic” protein therapeutics, known as follow-on biologics and biosimilars, stem from the impact of production variables on protein heterogeneity. Therefore, biosimilars are assessed on a case-by-case prior to approval and require extensive protein characterization efforts to show similarity (Carter 2011). Currently, mass spectrometry (MS)-based methods are at the forefront of these efforts due to their sensitivity, specificity, and analysis speed together with their use in diverse applications.

1.8 Introduction to Mass Spectrometry

Mass spectrometry (MS) has become an indispensable tool for characterizing proteins within the pharmaceutical industry. It provides sensitive and accurate
measurements of both small and large molecules by separating and detecting ions based on their mass-to-charge ratio (m/z). A mass spectrometer is comprised of three main parts that include an ionization source, mass analyzer, and detector. Depending on the instrument, the sensitivity, mass accuracy, and resolution vary. An instrument’s sensitivity permits analysis of trace analytes; however this factor is largely dependent on the sample composition and method for introducing the sample into the mass spectrometer. Mass accuracy is important for obtaining reliable data, and is defined as an instrument’s ability to obtain measurements that are as close as possible to the theoretical molecular weight of analytes of interest. Lastly, resolution is largely dependent on the mass analyzer, in which higher resolution mass spectrometers can identify multiple species within close molecular weight. All of these factors contribute to the quality of the data that can be obtained when using MS for characterizing protein therapeutics.

Advancements in MS hardware, software, and applications have grown substantially in the past thirty years with electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) being two milestones for characterization efforts of protein therapeutics. Compared to classical ionization methods such as plasma desorption (PD), chemical ionization (CI), and fast atom bombardment (FAB), these two techniques are considered “softer” ionization techniques because proteins are analyzed by MS without fragmentation (Tanaka, Waki et al. 1988; Fenn, Mann et al. 1989). The impact and significance of these methods are exemplified by the Nobel Prize in Chemistry being awarded to John Fenn and Koichi Tanaka in 2002 for their contributions of using ESI and MALDI to analyze biomolecules, respectively.
ESI involves a high voltage that is applied to a metal-tipped needle as solvent is pushed through. The resulting charged droplets are desolvated with the help of a nebulizing gas, typically N\textsubscript{2}, and transferred to the capillary through a voltage potential (Fenn, Mann et al. 1989). ESI allows analytes in solution to be transferred seamlessly into multiply charged ion species in the gas phase to be detected by MS. The extent of protonation that occurs during this process has been shown to retain information pertaining to its higher order structure; therefore, the charge state profile is often used to monitor solvent accessibility of a protein during ESI MS. Furthermore, this capability allows for the coupling of high performance liquid chromatography (HPLC) onto the front end of a mass spectrometer, which provides an orthogonal separation prior to MS. LC-MS separates and identifies macromolecules based on UV and/or fluorescence detection before subjecting them to MS for accurate mass measurements. However, there are limitations to LC-MS, as a result of the sensitivity of MS to non-volatile buffers that are commonly used in chromatography methods such as ion exchange and size exclusion. Therefore, the most common technique in LC-MS analysis is reverse phase (RP) chromatography in which organic solvents and acids are used as compatible mobile phases.

Several gas-phase fragmentation techniques can be used for sequence identity, each with their own advantages, and include collision activated dissociation (CAD), electron capture dissociation (ECD), and electron transfer dissociation (ETD). CAD preferentially fragments at the peptide bond generating b and y fragment ions and is the most widely used fragmentation technique. CAD is typically used for peptide sequencing and localizing PTMs. ECD and ETD are electron based fragmentation techniques that
provides an alternative approach to peptide identification. Electron based fragmentation typically fragments between the N and Cα which generates c and z fragment ions. This technique also provides added benefits in fragmenting disulfide bonds which may be beneficial for characterization. Furthermore, electron based fragmentation leaves PTMs intact, such as phosphorylation, which would typically not remain in CAD.

1.9 Characterization of Protein Therapeutics by Mass Spectrometry

MS is capable of in-depth structural analyses of proteins, which include verification of amino acid sequence, identification of PTMs, examination of N- and C-terminal sequences, determination of glycan composition and site occupancy, mapping disulfide connectivity, and monitoring global and local structural changes (Nemeth-Cawley, Tangarone et al. 2003; McLafferty, Breuker et al. 2007). There are two complementary approaches to gain structural information with MS including analysis of intact and enzymatically digested proteins.

1.9.1 Intact Mass and Top-Down MS

Intact analysis of proteins by MS is typically performed by MALDI MS or ESI MS. This can provide accurate molecular weight measurements on the protein of interest and determine the presence of PTMs including glycan heterogeneity (Nguyen, Becker et al. 1995). Intact mass measurements require the least amount of sample handling, which reduces the chances for assay-induced modifications and therefore is the most reliable method for assessing a protein’s molecular weight. A limitation to this method includes the need to remove complex sugars moieties for heavily glycosylated proteins due to overlapping charge states for each glycoforms. Another limitation to intact molecular
weight analysis includes the inability to resolve modifications that result in subtle changes in molecular weight.

Top-down measurements can also be employed on intact proteins, where fragmentation is most often used to gain information pertaining to the sequence on the N- and C-terminal portions of the molecule. This strategy eliminates the need for lengthy proteolytic digestion procedures. MALDI MS is often used in top-down MS and is an orthogonal method to the traditional N-terminal sequencing by Edman degradation.

### 1.9.2 Bottom-up MS and Peptide Mapping by LC-MS

Enzymatic digestion followed by LC-MS (“bottom-up” approach or peptide mapping) is the most commonly utilized strategy for obtaining detailed information pertaining to protein therapeutics. Enzymatic digestion is typically performed with trypsin or Lys-C followed by reducing disulfide bonds and alkylating free thiols. This is followed by reverse-phase chromatography directly infused into a mass spectrometer for MS and MS/MS analysis. With this technique amino acid composition can be verified, PTMs can be identified, and their location deduced. This includes glycan site occupancy and composition, the presence of N- and C-terminal variants, and other PTMs such as the presence of phosphorylation (Larsen, Trelle et al. 2006). Also if performed under non-reducing conditions disulfide bonds can be mapped (Nguyen, Becker et al. 1995). Peptide mapping is also commonly used to monitor non-enzymatic post-translational modifications, which include oxidation, deamidation, and disulfide scrambling in which non-native disulfides are formed (Kaltashov, Bobst et al. 2012).
1.9.3 Subunit and Middle-Down Analysis of mAbs by MS

A more recent technique used to characterize mAbs is through the use of Immunoglobulin G-degrading enzyme of S.pyogenes (IdeS), which cleaves below the hinge region of immunoglobulin G (IgG) antibodies. This enzyme is specific toward IgG mAbs and does not digest immunoglobulins M, A, D and E (von Pawel-Rammingen, Johansson et al. 2002).

Subunit or middle-up analysis uses the IdeS enzyme in conjunction with reducing agent to characterize the Fc’, Fd’, and light chain of a mAb. Subunit analysis provides a middle ground between intact and peptide mapping, where more information can be obtained about the molecule while minimizing sample preparation and data analysis (Fornelli, Ayoub et al. 2014). Sequence information can also be obtained on each subunit by employing fragmentation methods such as CAD or ETD, which is known as middle-down analysis (Fornelli, Ayoub et al. 2014).

1.9.4 Glycan Analysis by MS

MALDI and ESI MS are commonly used in the field of glycan analysis. The sensitivity of MS allows for the identification and characterization of minor glycan species. Unlike the classical method of fluorescence HPLC for detecting oligosaccharides, MALDI MS can obtain information pertaining to the identity of low-level glycans. However, methods such as 2-AA and 2-AB labeling with fluorescence HPLC are quantitative whereas MALDI MS can only provide the identity of the molecule. More recently, fluorescence HPLC coupled directly with ESI MS has yielded a powerful strategy by providing quantitative analysis of glycans with their simultaneous identification by MS (Chen and Flynn 2007).
1.9.5 Structural Analysis of Protein Therapeutics by MS

Methods such as hydrogen/deuterium exchange (HDX) MS and ESI-MS can provide detailed information pertaining to molecular weight, heterogeneity, protein dynamics, protein folding, and non-covalent protein interactions. MALDI is particularly useful when monitoring complex protein mixtures, because of relative ease in sample analysis. ESI of intact proteins generates multi-charged gas phase ions, which collectively form its charge state distribution. The first studies examining the charge state distribution of native vs. denatured proteins demonstrated that ESI-MS of unfolded or reduced proteins produced a wide distribution of ions with increased charge states, whereas ESI-MS of native proteins produced a narrow distribution of ions with lower overall charge state envelope (Chowdhury, Katta et al. 1990; Loo, Edmonds et al. 1990; Loo, Loo et al. 1991). This effect is a result of the increased accessibility of denatured proteins, which can accept more protons (Krusemark, Frey et al. 2009; Bobst and Kaltashov 2011). In this way, ESI-MS of intact proteins can provide a rapid and simple way to assess a protein’s higher order structure.

HDX MS is currently being implemented into the pharmaceutical industry to monitor the dynamics of protein therapeutics in solution (Berkowitz, Engen et al. 2012; Kaltashov, Bobst et al. 2012). The concept of HDX is straightforward in which liable amide hydrogens on the peptide backbone are exchanged for deuterium in solution under the specified experimental conditions. This is followed by quenching the reaction by bringing the pH down to 2.5 and the temperature to 0°C in order to slow down the exchange rate of hydrogen to deuterium and visa versa. From here, the protein is digested with a protease capable of digesting at such a low pH, typically pepsin, and separated by
LC-MS. The average incorporation of deuterium is then measured for each peptic peptide and mapped back to the crystal structure to observe areas of high or low flexibility.

Previously, these experiments were lengthy and the data could take months to interpret. However, with advancements in automating HDX experiments as well as data analysis, these experiments can be accomplished in less than a day and therefore feasible for implementing this instrumentation into the fast pace nature of the pharmaceutical industry (Berkowitz, Engen et al. 2012). HDX has shown to be useful for monitoring structural changes in protein therapeutics due to oxidation, Ca$^{2+}$ binding, formulation, and aggregation due to thermal stress (Bobst, Thomas et al. 2010; Houde and Berkowitz 2012; Zhang, Singh et al. 2012; Houde and Engen 2013). A limitation to HDX MS is that it is poor in understanding structural dynamics in complex mixtures. An example would be a mixture of disulfide scrambled species, and therefore each species would need to be purified prior to these experiments.

1.10 Cysteine Modifications in Protein Therapeutics

Disulfide bonds are covalently linked cysteine residues that form through oxidation of their sulfhydryl groups. A main function of disulfide bonds is to stabilize a protein’s structure. It is thought that this function is achieved by restricting the motion of specific protein segments and by reducing the enthalpy and entropy of the unfolded state. These modifications can also facilitate protein-protein interactions such as dimerization or oligomerization, as is the case for antibodies. However, besides the classical view of cysteine residues as participants in disulfide bond formation, many other cysteine variants have been reported, which include the presence of free thiols, trisulfides, cysteinylation, alternative disulfide isoforms, and disulfide scrambling (Kleinova, Belgacem et al. 2005;
Gu, Wen et al. 2010; Wang, Kumar et al. 2011; Liu and May 2012). Monoclonal antibodies are cysteine-rich proteins that represent the largest class of protein therapeutics. Cysteine residues play a critical role in regulating their oligomeric state and structure and in many cases have been shown to contain the aforementioned modifications.

1.10.1 Free Cysteines

Theoretically, all of the disulfide bonds in IgGs should be fully oxidized. However, free cysteines are commonly observed from recombinantly produced IgGs as well as those isolated directly from serum (Liu, Chumsae et al. 2010; Liu, Manuilov et al. 2011). The presence of free cysteines can lead to aggregation due to the formation of aberrant disulfide linkages and thermal instability. Intramolecular disulfide bonds are known to reinforce a protein’s higher order structure (Trivedi, Laurence et al. 2009), (Wang, Meng et al. 2004). When engineered with an extra disulfide bond, the half-life of substilisin E in serum increased three-fold and it was significantly more active at higher temperatures (Takagi, Takahashi et al. 1990). However, thermal instability was observed in the protein therapeutic omalizumab when free cysteines were present (Harris 2005).

1.10.2 Trisulfides and Cysteinylation

Other, less commonly observed modifications include the presence of an additional sulfur within a typical disulfide (trisulfide bond) or the addition of a free cysteine amino acid forming a disulfide bond with a cysteine in the molecule (cysteinylation). Trisulfides have been reported in every IgG subtype and are most often located within the interchain linkages (Gu, Wen et al. 2010). In one study, the trisulfide
in an IgG1 was shown to be maintained after prolonged storage and when incubated with rat serum. However, when injected into rats, the trisulfide rapidly converted to the native disulfide bond (Gu, Wen et al. 2010). Cysteinylation has previously been shown to occur on a cysteine residue in the complementarity determining region 3 (CDR3) region of an IgG1 monoclonal antibody (mAb) (Gadgil, Bondarenko et al. 2006; Banks, Gadgil et al. 2008). This modification negatively impacted its structural integrity and the potency (Banks, Gadgil et al. 2008).

1.10.3 Alternate Disulfide Isoforms

The classical view of the IgG subclasses (IgG1, IgG2, IgG3, and IgG4) was that they all had distinct interchain disulfide bonds, but each existed as single conformer (Liu and May 2012). However, it was later discovered that the two disulfide bonds within the hinge region of an IgG4 existed in equilibrium as two intra- or two interchain bonds (Angal, King et al. 1993; Bloom, Madanat et al. 1997; Schuurman, Perdok et al. 2001). Interestingly, this results in an unstable heavy chain-heavy chain interaction and is thought to be the root cause for the generation of bispecific IgG4 antibodies found in serum (Schuurman, Van Ree et al. 1999). In other words, as a result of the weakened interchain interaction, IgG4s are able to associate with other IgG4 half molecules and form antibodies with two distinct antigen-binding sites. When one of the cysteine residues involved in the intramolecular disulfide bond formation was mutated to a Ser, a more stable interaction between the two heavy chains was observed. Additionally, mutating the hinge sequence of the IgG4 (Cys-Pro-Ser-Cys) to the IgG1 sequence (Cys-Pro-Pro-Cys) also increased the stability of the heavy chain-heavy chain interaction (Schuurman, Perdok et al. 2001).
MS has played an important role in disulfide bond characterization. For example, MS was critical in the recent demonstration that the human IgG2 subclass exists in multiple disulfide isoforms (Dillon, Ricci et al. 2008; Wypych, Li et al. 2008). This explained the structural changes observed by reverse phase and ion exchange chromatography, and capillary electrophoresis. That this discovery took until 2008 illustrates the challenges in identifying these distinct, but subtle differences prior to the use of high-resolution analytical techniques. Moreover, the structural variations did not alter the intact molecular weight of IgG2 antibodies because they are fully oxidized species. However, when a non-reduced Lys-C peptide map was used to characterize the disulfide connectivity, three distinct disulfide-linked peptides were observed which led to the discovery that the IgG2 molecule existed in three isoforms (Wypych, Li et al. 2008). These three peptides included a 5 kDa peptide dimer with four intermolecular disulfide bonds (IgG2A), a ~15 kDa peptide pentamer of one light chain and four heavy chain peptides (IgG2A/B), and a 25 kDa peptide octamer consisting of two light chain and six heavy chain peptides (IgG2B) (Wypych, Li et al. 2008). The sensitivity, wide dynamic range, and mass accuracy of MS were fundamental in elucidating the disulfide isoforms present in IgG2 antibodies.

1.10.4 Disulfide Scrambling

The rearrangement of a protein’s disulfide bonds can cause significant alterations in its higher order structure, which can impact the stability, efficacy, and safety of a protein therapeutic. Changes in disulfide connectivity can occur normally as is the case for IgG2 conformers, or they can form in response to thermal stress, alkaline pH, and the presence of oxidizing and reducing agents. Conditions that can lead to disulfide
scrambling in protein therapeutics include unfavorable storage conditions, sample handling, and formulation. Unlike other non-enzymatic PTMs, disulfide scrambling occurs without changing the molecular weight of the molecule. Therefore, specific methods are required to be able to monitor protein samples for these changes.

LC-MS of non-reduced enzymatic digests has been the most widely used method to characterize disulfide bonds and map their connectivity. However, identifying non-native disulfide bonds can be challenging if proteins contain a significant number of cysteine residues because of the large number of possible disulfide linkages. With an increase in disulfide bonds, analyzing the disulfide-containing peptide fragments becomes progressively more challenging. Moreover, disulfide scrambling can be difficult to detect especially for low abundance species, but technological advancements to improve the sensitivity and resolution in both chromatography and MS have significantly improved the capability of examining this modification.

Gas-phase fragmentation methods that preferentially cleave disulfide bonds have also played a large role in confirming native and non-native disulfide-linked peptides. These methods include negative ion CAD, ECD, and ETD (Zhang and Kaltashov 2006; Wu, Jiang et al. 2009; Wang, Lu et al. 2011). ETD has gained immense popularity for characterizing protein therapeutics and was utilized to successfully map disulfide bonds in hGH, multiple IgG1 mAbs, and a fusion protein bearing an IgG4 Fc region (Wu, Jiang et al. 2009; Wang, Lu et al. 2011). Furthermore, ETD was used to examine disulfide scrambled isoforms of an IgG1 mAb following exposure to thermal stress. Specifically, these species were separated by SDS-PAGE, the bands excised for a non-reduced in-gel digestion, and characterized by LC-MS peptide mapping with disulfide bond
fragmentation by ETD (Wang, Lu et al. 2011). Despite the benefits of the current advancements in methodology and instrumentation, probing for disulfide scrambling remains a lengthy process. Therefore, we sought to investigate ESI-MS as a method to probe charge-state distribution as a means to quickly monitor for non-native disulfides.

Disulfide scrambling is a major concern that can affect the structural integrity of protein therapeutics and can lead to undesirable consequences. This dissertation focuses on MS-based strategies that can provide a rapid assessment of proteins with native and non-native disulfide bonds using ESI-MS. Furthermore, distinguishing native from non-native connections is challenging within cysteine-rich protein segments. We demonstrate that with nanoLC-MS in combination with multiple proteolytic digests and several gas phase fragmentation techniques, we can map the disulfide bonds within the cysteine-rich N-terminal region of the protein therapeutic GCase.
CHAPTER 2

MAPPING DISULFIDES WITHIN THE CYSTEINE-RICH REGION OF THE PROTEIN THERAPEUTIC BETA GLUCOCEREBROSIDASE

This work was accomplished in collaboration with SHIRE Human Genetic Therapies and Dr. Cedric Bobst.

2.1 ABSTRACT

Proper formation of disulfide bonds is critical for ensuring the safety and efficacy of many protein therapeutics. Therefore it is essential to have methods that can quickly and completely characterize all of the disulfide bonds within each protein. This is a priority in order to produce a safe, effective, and stable product. A common strategy typically involves analyzing a trypsin digest with liquid chromatography and mass spectrometry (LC-MS). In many cases, this method is suitable to completely map disulfide connectivity. However, when cysteine residues are located within close proximity in the primary sequence, it can be difficult to determine which cysteines are participating in bond formation. Here we use the protein therapeutic β-glucocerebrosidase (GCase) as an example to develop a strategy to map disulfides in cysteine-rich protein segments. We demonstrate that by using multiple proteases, in combination with nanospray LC-MS and several gas-phase fragmentation techniques, we can map the disulfides within the cysteine-rich N-terminal region of GCase. Moreover, we established a method to detect and identify minor disulfide scrambled species within a long-term GCase stability sample.
2.2 INTRODUCTION

Beta-glucocerebrosidase (GCase), also known as acid β-glucosidase, is an enzyme present in the lysosome that is responsible for the degradation of glucocerebroside (GC) into ceramide and glucose (Grabowski, Gatt et al. 1990). Ceramides are building blocks for higher order glycosphingolipids, which are important components of cell membranes that regulate numerous cellular pathways. In general, ceramide suppresses the cell cycle, whereas GC can promote cell-specific proliferation in keratinocytes (Messner and Cabot 2010). Ceramide also promotes apoptosis by mediating the cleavage of BID to tBID in the lysosome, which in turn activates caspase-3 and caspase-9 (Ogretmen and Hannun 2004). With a decrease or loss in GCase enzymatic activity, GC accumulates in the lysosomes of phagocytic cells causing them to swell. This causes an enlargement of the liver and spleen, anemia, and fatigue, which are all symptoms of type I Gaucher disease (Peters, Lee et al. 1977; Barton, Furbish et al. 1990).

Type I Gaucher disease is a rare autosomal recessive disorder affecting 1:40,000 to 1:60,000 in the general population (Meikle, Hopwood et al. 1999; Morris 2012). In most cases, treatment for this disease is enzyme replacement therapy (ERT) in which GCase is injected intravenously to compensate for the dysfunctional endogenous enzyme.

There are three marketed variants of GCase, which are Imiglucerase (Cerezyme®), Velaglucerase alfa (VPRIV®), and Taliglucerase alfa (Elelyso®). Cerezyme®, approved in 1994, was the first recombinant protein therapy used for the treatment of type I Gaucher disease. This product is produced in Chinese hamster ovary
(CHO) cells and contains a single amino acid mutation compared to that of the endogenous enzyme (Grabowski, Barton et al. 1995).

VPRIV® is an alternative biologic produced using human gene activated technology (Moran 2010). After the 2009 viral contamination of one of Genzyme’s manufacturing facilities that led to a shortage of Cerezyme®, VPRIV® was granted accelerated approval to ensure all type I Gaucher disease patients would receive treatment (Moran 2010). VPRIV® consists of the native GCase protein sequence with high mannose N-linked glycans produced using a gene-activated human cell line (Brumshtein, Salinas et al. 2010; Moran 2010). These changes have been shown to improve GCase internalization by human macrophages (Brumshtein, Salinas et al. 2010).

Lastly, Elelyso® was approved by the FDA in 2012 to be the first plant-based biologic approved in the United States. Plant-derived products are appealing because they do not require the investment of a large-scale bioreactor and are cheaper to produce (Kaiser 2008; Zimran, Brill-Almon et al. 2011). Similar to VPRIV®, Elelyso® does not require treatment in vitro with exoglycosidases to expose the core mannose residues necessary for macrophage internalization (Shaaltiel, Bartfeld et al. 2007; Zimran, Brill-Almon et al. 2011). This is achieved through the addition of a C-terminal sequence that results in the formation of high mannose sugars in vivo (Zimran, Brill-Almon et al. 2011). All three GCase biologics can replace the defective enzyme in patients with type I Gaucher disease.

GCase is a 63 kDa protein consisting of three domains observed by x-ray crystallography (Figure 2. 1)(Dvir, Harel et al. 2003) (Brumshtein, Salinas et al. 2010). Domain III contains the active site, which is positioned within a TIM barrel and functions
to hydrolyze the glycosidic bond in GC leading to free glucose and ceramide (Dvir, Harel et al. 2003) (Meikle, Hopwood et al. 1999). Domain I includes the N-terminal segment and contains the only two disulfide bonds within the protein separated by a single amino acid (Dvir, Harel et al. 2003). This presents a challenge to corroborate what is observed by x-ray crystallography using traditional disulfide mapping techniques.

Figure 2. 1 Beta-glucocerebrosidase (GCase) primary and tertiary structure. Highlighted within the primary sequence are the disulfide bonds, free cysteines (orange), and N-linked glycosylation sites (red). The structure of GCase is also shown observed by x-ray crystallography (2WKL). Shown are the three domains with a focus on the three free cysteines (orange) and the zoomed in disulfide bond containing region (Brumshtein, Salinas et al. 2010).
Post-translational modifications (PTMs) are present in most protein therapeutics and can influence their bioavailability, structure, and function. Disulfide bonds are particularly important to reinforce the structure of cysteine-rich protein therapeutics, such as monoclonal antibodies (Zhang, Marzilli et al. 2002). The complexity and heterogeneity of post-translational modifications (PTMs) makes the assessment and regulation of protein-based therapeutics exponentially more complex than organically synthesized small molecules (Berkowitz, Engen et al. 2012; Sethu, Govindappa et al. 2012). Current methods used to characterize protein therapeutics involve the analysis of their primary structure and PTMs by a combination of proteolysis, typically trypsin or Lys-C, followed by liquid chromatography and mass spectrometry (LC-MS) (Srebalus Barnes and Lim 2007) (Zhang, Pan et al. 2009). However, enzymes such as trypsin and Lys-C are not ideal for deciphering the disulfide connectivity when multiple cysteine residues reside within the same proteolytic fragment. Furthermore, alternative disulfide linkages, which exist in substantially lower quantities, are difficult to detect, especially within cysteine-rich protein segments. Here we utilize multiple proteolytic enzymes in combination with nanospray MS and high resolution MS to confirm the native disulfide connectivity within the cysteine-rich region of GCase. Moreover, we developed a method to monitor the presence of disulfide scrambled species.

2.3 Materials and Methods

2.3.1 Endoproteinase Lys-C digest of GCase

GCase was generously donated to by SHIRE human genetic therapies (Lexington, MA). GCase was alkylated with 10 mM iodoacetic acid (Sigma-Aldrich St.Louis, MO) in
a denaturing buffer of 6 M guanidine hydrochloride (GuHCl) (Thermo Fisher Waltham, MA), 1 mM ethylenediaminetetraacetic acid (EDTA) (Sigma-Aldrich), and 100 mM tris (hydroxymethyl) aminomethane hydrochloride (Tris-HCl) at pH 8.5. After incubating GCase with alkylating agent for 30 minutes in the dark, the protein was buffer exchanged into 100 mM Tris HCl pH 8.5 using NAP5 desalting columns (GE Healthcare Life Sciences Pittsburg, PA) and subsequently digested with endoproteinase Lys-C (Promega Madison, WI) at a ratio of 1:400 (w:w). After five hours of digestion, the protein was deglycosylated with PNGase F (New England Biolabs Ipswich, MA) for an additional two hours at 25 °C or 37 °C. Incubation at 25°C led to the generation of the monoglycosylated L1L2 peptide, while incubation at 37°C removed all N-linked glycans. Nomenclature for the native N-terminal peptide generated by a Lys-C digest is shown in Figure 2. 2.

**Monoglycosylated L1L2:**

\[
1\text{ARPCIPK}^{7} 8\text{SGYSSVVVCVCNATYCD}^{24} \\
25\text{SFDPPTFPALGTSRYESTSGRM}^{49} \\
50\text{ELSMGPIQADHTGTGLLLTLQPEQK}^{74}
\]

**L1L2:**

\[
1\text{ARPCIPK}^{7} 8\text{SGYSSVVVCVDATYCD}^{24} \\
25\text{SFDPPTFPALGTSRYESTSGRM}^{49} \\
50\text{ELSMGPIQADHTGTGLLLTLQPEQK}^{74}
\]

- **Mannose**
- **N-Acetyl Glucosamine**
- **D** = Deamidated Asn due to PNGaseF treatment

**Figure 2. 2 Nomenclature for N-terminal peptide of GCase digested with Lys-C**
2.3.2 LC-MS identification of monoglycosylated and deglycosylated L1L2 GCase peptide

GCase peptide mixture, subsequent to deglycosylation, was injected into an Agilent 1100 HPLC for reverse phase separation using a Gemini C-18, 5 µm, 3x150 mm column (Phenomenex Torrance, CA) directly infused into a hybrid quadrupole time-of-flight (Q-TOF) mass spectrometer (QSTAR-XL AB SCIEX Framingham, MA). LC mobile phases included water with 0.1 % formic acid (FA) (Thermo Fisher) and acetonitrile (ACN) (Thermo Fisher) with 0.1 % FA. Identification of the monoglycosylated and deglycosylated L1L2 peptide was accomplished by RP-HPLC separation online with MS.

2.3.3 RP-HPLC purification of the monoglycosylated and deglycosylated L1L2 peptide

GCase peptides, generated from a Lys-C digest, were separated by RP-HPLC using a Gemini C-18 reverse phase column and mobile phases of water with 0.1 % trifluoroacetic acid (TFA) and ACN with 0.1 % TFA. Peptides were fractionated manually to purify monoglycosylated L1L2 peptide followed by lyophilization. Similarly, the deglycosylated L1L2 peptide was isolated by RP-HPLC and subsequently lyophilized. These fractions were resuspended in 50 % ACN with 0.1 % FA to confirm the presence of the N-terminal peptide. During identification of purified L1L2 GCase peptides, collision activated dissociation (CAD) was performed for sequence specific information.

2.3.4 Proteolytic digest of purified L1L2 with pepsin and Asp-N

GCase N-terminal peptide, L1L2, was digested individually with pepsin (Thermo Fisher) and Asp-N (Promega). The pepsin digest included a resuspension of lyophilized
L1L2 at a ratio of 1:3 (v:v) into an acidic solution of 0.1 % FA. Digestion was performed at 25 °C with continuous agitation for five minutes. Subsequently, the immobilized pepsin was removed by centrifugation and the peptic peptides of L1L2 were analyzed by nanoLC-MS.

A digest with endoproteinase Asp-N was performed in parallel to the pepsin digest in which Asp-N was added to the resuspended L1L2 peptide (50 mM Tris pH 8.0) at a ratio of 1:50 (w:w) and incubated for twenty minutes at 37°C. The peptides generated by this digest are L1A1A2 and L1A2 (Figure 2.3). Subsequently, the Asp-N digest was quenched with a 2-fold dilution of 0.1 % FA and separated by reverse phase on an Agilent 1100 HPLC. The collected fractions were lyophilized and resuspended in 50 % ACN and 0.1 % FA. The fractions were then analyzed by direct ESI-MS on a Solarix 7T FT-ICR (Bruker Daltonics Billerica, MA). Two MS gas phase fragmentation techniques were performed on the disulfide containing peptides generated by the Asp-N digest. CAD and electron capture dissociation (ECD) were used, separately, on each peptide for peptide backbone and disulfide fragmentation.

**Figure 2.3 Nomenclature for natively oxidized N-terminal peptides generated by a Lys-C/Asp-N digest**

L1A1A2:

\[ ^{1}\text{ARP}\text{CIPK}^{7} \text{SFGYSSVVCCV}^{18} \text{DATYC}^{23} \]

L1A2:

\[ ^{1}\text{ARP}\text{CIPK}^{7} \text{SFGYSSVVCCVDATYC}^{23} \]

\( D = \) Deamidated Asn due to PNGaseF treatment
2.3.5 Asp-N digest of intact GCase

GCase stability sample was kept at 4°C for about 5 years and donated by SHIRE human genetic therapies. GCase and GCase stability samples were alkylated using iodoacetic acid (IAA) in a denaturing solvent of 6M GuHCl for 30 minutes. With all free cysteines alkylated, GCase was digested using endoproteinase Asp-N to generate a short peptide (A1), which consists of the first twenty-three amino acids, including the two disulfide bonds of GCase (Figure 2.4). Each sample was digested for five hours and subsequently deglycosylated with PNGase F (NEB) for an additional two hours at 37°C. The peptide mixture was analyzed by nanoLC-MS/MS using LC Packings Ultimate nano LC (Dionex/Thermo Fisher Scientific) and a Q-TOF mass spectrometer (QSTAR-XL AB SCIEX) for the identification of the A1 peptide and any corresponding N-terminal peptides. Following identification of A1 and partially reduced A1 by nanoLC-MS, the A1 peptide was isolated by RP-HPLC with a Gemini C-18 reverse phase column. To monitor the purity of the isolated A1 peptide ESI-MS was performed with a soliriX 7T FTMS (Bruker Daltonics). Further investigation of the A1 peptide was accomplished by CAD, following the isolation of the [M+3]^{3+} charge state at 830.3 m/z.

\[
\text{A1:} \quad \begin{array}{c}
\text{ARP} \text{CIPKSFGYSVVCCVCDATYC}^{23} \\
\text{D = Deamidated Asn due to PNGaseF treatment after Asp-N treatment}
\end{array}
\]

Figure 2.4 Natively oxidized N-terminal peptide generated by an Asp-N digest of GCase
2.3.6 Thermolysin digest of GCase A1 N-terminal peptide

The A1 peptide of GCase collected by RP-HPLC fractionation was resuspended in 50 mM Tris-HCl and digested with thermolysin from *Bacillus thermoproteolyticus rokko* (Sigma-Aldrich). The theoretical peptides generated by an Asp-N/thermolysin digest for native and disulfide scrambled species (N, S1, and S2) are shown in Figure 2.5 with spaces indicating the cleavage sites. The enzyme was added at a 1:500 (w:w) E:S ratio and the reaction was performed for 20 minutes at 37°C. The peptide mixture was separated and detected using nanoLC-MS and all potential disulfide-linked peptide dimers corresponding to native and disulfide scrambled species (N, S1, and S2) within the N-terminal segment of GCase were explored by manually tabulating their molecular weights. Characterization of the disulfide linked peptides included analysis by Analyst QS (AB SCIEX), protein prospector (University of California San Francisco, CA), and manual interpretation.

![Native (N): ARPCIPKS FGYSS V VC VCD ATYC\textsuperscript{23}](image)

**Figure 2.5 Nomenclature for native and scrambled species of A1 thermolysin digest**

2.4 Results and Discussion

2.4.1 Glycosylation at N19 prevents disulfide bond characterization of L1L2

To characterize the disulfide bonds associated with GCase, we aimed to avoid undesired modifications (i.e. oxidation, deamidation, and disulfide scrambling) that might occur during sample preparation and analysis. It is known that alkaline pH and elevated
temperatures are conditions that favor these modifications (Browning, Mattaliano et al. 1986; Liu, Gaza-Bulseco et al. 2007). Unfortunately, these conditions are also optimal for the activity of most proteolytic enzymes. To prevent non-native disulfide formation due to thermal stress, GCase was digested and deglycosylated at 25°C with Lys-C and PNGase F respectively. The peptides were analyzed by LC-MS and a molecular weight of ~10 kDa was identified as the L1L2 peptide with one N-linked Man9 glycan (Figure 2. 6).

![Figure 2. 6 Identification of monoglycosylated N-terminal peptide by LC-MS.](image)

LC-MS was performed on a Lys-C digest of GCase. All free cysteine residues were alkylated with IAA and glycans subsequently cleaved with PNGase F at 25°C. A total ion chromatogram (TIC) is shown with the identification of GCase peptides. The mass spectrum (red) reveals L1L2 N-terminal peptide with the most abundant glycoform containing (Man)$_9$(GlcNAc)$_2$ and a lower abundance of (Man)$_8$(GlcNAc)$_2$. 
There are two N-linked glycans (N19 and N59) located within the L1L2 protein segment. To determine the glycan site occupancy, the monoglycosylated L1L2 peptide was isolated by RP-HPLC fractionation (Figure 2. 7A) and was subjected to CAD. We observed the removal of the N59 glycan and deamidation due to PNGase F treatment, and predominantly (Man)_9(GlcNAc)_2 at position N19. Glycan composition was previously determined at SHIRE (Brumshtein, Salinas et al. 2010). Further investigation of the GCase N-terminus by CAD resulted in fragmentation of the C-terminus of L1L2 and within the N19 glycan (Figure 2. 7B). However, peptide backbone fragmentation did not occur within the disulfide-containing region, which prevented our ability to map the disulfide connectivity. These results demonstrate that the N19 high mannose glycan prevents disulfide characterization by blocking fragmentation within the N-terminal segment of L1L2 and therefore its removal is critical.

Treatment of GCase with PNGase F removed the other two N-linked glycans at N146 and N260 as determined by the loss of the glycan plus a peptide mass increases of 1 Da (deamidation) occurring on each of the known glycopeptides. Taken together, these results suggest that steric hindrance imposed by disulfide bonds prevents deglycosylation of N19, which in turn impedes disulfide mapping by CAD fragmentation.
Figure 2. Purification of monoglycosylated N-terminal peptide.
(A) RP-HPLC separation of GCase Lys-C digest. Each fraction was collected manually from a reverse phase C-18 column and subsequently lyophilized. Fraction 15 contained the monoglycosylated L1L2 peptide of interest. (B) Collision activated dissociation (CAD) of L1L2 monoglycosylated peptide. With the glycan present on the L1L2 peptide, fragmentation is occurring at the C-terminal region of the peptide (top) as well as fragmentation of the sugar moieties (bottom).
2.4.2 CAD fragmentation of deglycosylated L1L2 occurs mainly on the C-terminus

GCase treatment with PNGase F was adjusted to include an incubation temperature at 37°C, which resulted in the removal of the remaining N-linked glycan on the L1L2 peptide. The purity was assessed by RP-HPLC fractionation and online LC-MS (Figure 2.8). Upon removal of the glycans, the fraction containing the L1L2 deglycosylated peptide was subjected to CAD for characterization (Figure 2.9). By analyzing the b and y fragment ions corresponding to the L1L2 peptide, the C-terminal region was predominantly observed and very few ions pertaining to the N-terminal region were detected. These results are similar to CAD fragmentation of the mono-glycosylated L1L2, with a decrease in the complexity due to the absence of the glycan. However, fragment ions necessary for determining disulfide bond orientation were not observed. For both mono- and deglycosylated L1L2, fragmentation occurred mainly at the flexible C-terminal region of the peptide. Therefore, additional enzymes were used to isolate smaller disulfide-containing peptides of GCase.
Figure 2. Purification of GCase L1L2 N-terminal peptide.
(A) Reverse phase chromatogram of GCase Lys-C digest. Highlighted in blue contains the deglycosylated collected fraction that contains L1L2 peptide. (B) NanoLC-MS of purified L1L2. L1L2, collected from RP-HPLC was analyzed using nano-LCMS. The blue trace is a TIC of purified L1L2. The inner box shows the mass spectrum of the purified fraction of L1L2 with co-eluting L15 peptide.
Figure 2. 9 CAD Fragmentation of L1L2 Peptide
The b and y ions identified by CAD pertain mainly to the C-terminal region of the L1L2 peptide and do not provide the necessary fragmentation data to decipher the disulfide connectivity of GCase.

2.4.3 Sequence coverage of the L1L2 C-terminus by pepsin

As mentioned previously, conditions favorable for disulfide scrambling include alkaline pH. Therefore, an enzyme whose activity is optimal at an acidic pH is ideal. Pepsin is one of the few enzymes active under these conditions and was chosen in an attempt to produce smaller peptides that contain the N-terminal disulfide bonds. Due to its broad specificity, pepsin can complicate data analysis by generating a large number of fragments. We attempted to circumvent this problem by purifying the L1L2 peptide prior to digestion with pepsin. ESI-MS measurements revealed numerous peptides relating to the C-terminal portion of L1L2, whereas the N-terminal region was not detected (Figure 2. 10). These results are similar to trypsin digests of full length GCase, in which recovery of the N-terminal region is minimal and may be due to the steric hindrances imposed by the presence of disulfide bonds.
2.4.4 CAD of Lys-C/Asp-N digest reveals the native disulfide connectivity of GCase

Thus far, both trypsin and Lys-C/pepsin digests failed to facilitate disulfide mapping of GCase. As a result, we searched for less commonly used enzymes that would produce smaller, soluble peptides for MS analysis. The presence of aspartic acid at residue D24, led us to examine Asp-N as another candidate for further digestion of L1L2 because it specifically hydrolyzes N-terminal to aspartic acids. If successful, Asp-N would cleave between C23 and D24 in the N-terminal region of the protein, generating a short [1-23] peptide (L1A2) containing both disulfide bonds (Figure 2. 3). Furthermore, an additional Asp-N cleavage site is introduced due to the Asn to Asp conversion as a result of PNGase F treatment and if L1L2 is cleaved at both sites, then the peptide trimer L1A1A2 is produced (Figure 2. 3). To assess the L1L2 Asp-N digest, we collected fractions from RP-HPLC and analyzed the peptides by FTMS.
As expected, analysis of the Asp-N L1L2 digest revealed the presence of the peptide trimer L1A1A2 held together by two disulfide bonds. Therefore, Asp-N does indeed cleave between C18 and N19 following its conversion to aspartic acid. Gas phase fragmentation of either disulfide in L1A1A2 would allow us to determine peptide species that are in the native (N) disulfide configuration (C4/C16 and C18/C23) or S2 disulfide configuration (C4/C18 and C16/C23). We proceeded by using ECD as a fragmentation technique because of its ability to dissociate disulfide bonds (Roman A. Zubarev 1999). Specifically we show one disulfide bond (C4/C16 or C4/C18) is fragmented by ECD and separates L1 from A1A2. Additionally, the second disulfide bond was also cleaved (C18/C23 or C16/C23) producing ions corresponding to the molecular weights of L1A1 and A2 peptides (Figure 2.11).

L1A2 was also produced and co-purified with the L1A1A2 peptide. When subjected to CAD a variety of b and y ions were produced, including b10, b11, y6, and y7 (Figure 2.12). These ions are present only if the disulfide arrangement is in the native (N) configuration and in agreement with the crystal structure of this molecule (2WKL).

We have successfully established a method to examine the native disulfide connectivity of GCase using a combination of Lys-C and AspN digestion coupled with CAD. However, the detection of disulfide scrambled species is an important factor when analyzing cysteine-rich protein segments, where many non-native disulfides may exist in low abundance. S1 disulfide configuration would separate out L1A2 and A1 peptides and be detected by accurate mass measurements. It is also possible to determine the presence of the native configuration (Figure 2.13), with CAD on the L1A2 peptide or the L1A1A2 peptide. Furthermore, it is possible to determine the S2 disulfide scrambled species by
CAD on the L1A1A2 peptide. However, neither S1 nor S2 were determined by LC/MS in combination with CAD or ECD fragmentation. This was largely due to the inefficient production of L1A1A2 by Asp-N (Figure 2.14) and led us to examine other proteolytic strategies.
Figure 2. 11 ECD of GCcase peptide trimer L1A1A2.
ECD of the peptide trimer L1A1A2 was performed on a solariX™ 7T FTICR-MS (Bruker Daltonics, Billerica, MA). L1A1A2 (m/z 626.3) was subjected to ECD, which is known to preferentially fragment the disulfide bond. The red boxes depict the corresponding sequence of each zoomed in fragment ion. Based on the results from ECD of GCcase L1A1A2 peptide, we can conclude that the disulfide pattern is either in the N or S2 configuration. We can also rule out the presence of S1 in this sample due to the lack of its corresponding peptide.
Figure 2. 12 CAD of GCase peptide dimer L1A2.
Various fragment ions were observed including four ions unique to the disulfide pattern shown in the crystal structure (2WKL). These ions are b_{10}^{2+}, b_{11}^{2+}, y_{7}^{2+}, and y_{6}^{2+}. With these ions we can unequivocally conclude GCase contains the disulfide pattern consistent with its crystal structure.

Figure 2. 13 All possible disulfide patterns in GCase N-terminal segment.
Figure 2. **Asp-N digestion efficiency is low in disulfide containing region.**

nanoLC-MS was performed on an Asp-N digest of L1L2. The black trace represents the TIC of this digest while the red and blue traces are extracted ion chromatograms (XICs) corresponding to L1A2 and L1A1A2 peptide. Asp-N efficiency to hydrolyze N-terminal to D19 is low and therefore, may not be suitable for purification and identification of any minor scrambled species.

### 2.4.5 Asp-N digestion and LC-MS reveal partially reduced species of A1

Enzymatic digestion of GCase with Lys-C leads to the large 8 kDa N-terminal peptide L1L2, which we show that in combination with Asp-N is useful for gaining disulfide-linked peptide information. However, it is not ideal to determine the presence of partial reduction or disulfide scrambled species. This is largely due to a change in retention time of partially reduced species, which results in their loss during L1L2 purification and the inefficiency of Asp-N to generate the L1A1A2 peptide. Therefore,
Asp-N was used as our primary protease to generate a small twenty-three amino acid peptide (A1) (Figure 2. 4) followed by deglycosylation.

The peptide mixture generated from the Asp-N digestion of GCase was analyzed by nanoLC-MS, where we were able to identify the A1 peptide and detect low abundant partially reduced species of A1 in which two cysteine residues were carboxymethylated (Figure 2. 15). Similar to the L1A2 peptide, the native disulfide configuration can be determined by CAD fragmentation of the N- or C-terminus of V17 on the A1 peptide. However, characterization of alternative disulfides cannot be determined with the sole use of Asp-N, and therefore a second protease was introduced.

Figure 2. 15 GCase peptide map of an Asp-N digest
Nano-LCMS was performed on an Asp-N digest of GCase and identified peptides are shown on the TIC. The N-terminal peptide A1 (A1) and A1 with one disulfide bond and two carboxymethyl cysteines (A1∗) are shown to elute at ~38 minutes.
2.4.6 All disulfide configurations within the cysteine-rich region of GCase can be probed by digesting A1 with thermolysin

The ability of Asp-N to generate a short N-terminal peptide of GCase allows us to detect the presence of any partially reduced species and characterize the predominant disulfide species present in GCase. We show the majority of the disulfide connectivity in GCase is consistent with its crystal structure as determined by x-ray crystallography (2WKL) (Figure 2.1). However, a single proteolytic digest of Asp-N is not sufficient to detect the presence of other minor disulfide scrambled species.

Pepsin and thermolysin were chosen as a second protease to digest the A1 peptide. A pepsin digest of A1 showed poor digestion efficiency and after extended periods of incubation only the intact A1 peptide was present. Similar to the L1L2 peptide, we believe the digestion inefficiency is due to steric hindrance imposed by the presence of the disulfide bonds, and the inability for pepsin to effectively cleave within this region.

Next, thermolysin was chosen as a second protease whose enzymatic activity favors valine and other hydrophobic residues. Thermolysin significantly improved digestion efficiency with no intact A1 remaining following a twenty-minute incubation. If N, S1, and S2 disulfide configurations are present, a thermolysin digest can generate disulfide linked peptide dimers distinct to all three species. Analysis by nanoLC-MS demonstrated that thermolysin completely digested A1 and resulted in the identification of peptide dimers distinct to N and S2, with S1 not observed (Figure 2.16). An extracted ion chromatogram shows the relative abundance of both N and S2 species with low levels of S2 disulfide scrambled species observed. Together with the analysis of the GCase Asp-N digest (Figure 2.15) and the dual proteolytic digest of Asp-N and thermolysin (Figure
2.16), we are able to monitor for all possible disulfide linkages and partial reduction within the cysteine-rich protein segment of GCase.

Figure 2.16 GCase A1 thermolysin digest.
Thermolysin digest of A2 generated peptide dimers unique to N and S2 disulfide scrambled species, while S1 scrambled species was not observed. The black trace represents the TIC of A1A2 thermolysin digest. Below the TIC are extracted ion chromatograms of the m/z value unique to N, S1, and S2 disulfide linked peptides. A mass spectrum with m/z 594.8 and m/z 602.8 that represent a peptide dimer unique to N (orange) and S2 (green) disulfide configuration of GCase.

2.4.7 Characterizing native and non-native disulfide bonds in a GCase stability sample

To test whether the Asp-N and Asp-N/thermolysin digest can identify partial cysteine reduction and disulfide scrambling of GCase, we performed a side-by-side comparison of a control and stability sample. The stability sample was kept at 4 °C for
five years, while the control sample was stored lyophilized at -80 °C prior to digestion and analysis by nanoLC-MS. An extracted ion chromatogram normalized to the total ion counts from each spectrum shows the relative abundance of fully oxidized A1 and also A1 with two cysteine residues carboxymethylated (Figure 2.17). We observed similar levels of partially reduced A2 species following Asp-N digestion and nanoLC-MS. This indicates that long-term storage does not affect the amount of partially reduced cysteine residues in GCase.

Next, we sought to use our disulfide-mapping strategy to assess whether changes in GCase disulfide connectivity occur in the stability sample as compared to control. A1 digestion with thermolysin allowed us to probe for all possible disulfide configurations (native, S1, and S2) within the N-terminus of GCase and an increase in S2 disulfide scrambled species was observed in the 5 year stability sample probed by this method (Figure 2.18). The theoretical and experimental molecular weights were calculated for their mass accuracy. For the control sample, the N species resulted in ±5 ppm error and the S2 species resulted in ±8 ppm error. The mass accuracy for the stability sample resulted in ±6 ppm for the N species and ±8 ppm error for the S2 scrambled species. This data demonstrates that when stored over long periods of time, a detectable percentage of GCase undergoes disulfide scrambling, which favors the S2 configuration above the S1 state. This is likely due to the close proximity of the C16 and C18 residues, which would not require a dramatic change in three-dimensional structure when shifting from the native (C4/C16) to the S2 (C4/C18) scrambled configuration. On the other hand, making the S1 (C4/C23) linkage would require a significant rearrangement of the folded protein and is not energetically favorable.
Figure 2. 17 Partial reduction observed in GCase stability and control
GCase Asp-N digest of GCase control (purple) and GCase stability sample (grey) were digested with Asp-N and analyzed by nanoLC-MS. The top trace for both control and stressed show a TIC of the digest, and below two extracted ion chromatograms of A1 oxidized and partially reduced. (GCase N-terminal peptide [1-23]) and the A1 peptide partially reduced, with one disulfide bond and two cysteine residues carboxymethylated.
Figure 2. 18 nanoLC-MS of GCase control and stability A1 thermolysin digest
Top four traces represent GCase control and bottom four traces represent the stability sample. A TIC of of A1 thermolysin digest for GCase control and stability samples is shown by the black traces. Below the TIC are XICs of m/z representative of N, S1, and S2 disulfide species. An increase in the presence of S2 disulfide species (green) is seen with the sample sitting at 4 °C for about 5 years.

2.5 Conclusions

Advances in MS have propelled this technique into an instrumental component of pharmaceutical drug development. The diversity of applications for MS has made it a powerful and robust strategy for characterizing diverse protein PTMs as well as subtle structural differences (Bobst, Thomas et al. 2010). Here we addressed the challenges that arise when characterizing complex cysteine-rich protein segments, as seen with the
GCase N-terminus. It is difficult to probe disulfide configurations in which cysteine residues are in close proximity to one another in the primary sequence. Also, complex bond configurations are challenging to characterize. An example is the cystine knot structural motif, which has three intertwined disulfide bridges present in certain growth factors and toxins (Sun and Davies 1995; Craik, Daly et al. 2001). In most cases, disulfide mapping will need to be tailored to each protein of interest and may require a multi-tiered approach.

Developing a method to determine the native disulfide linkages is an important step in characterizing the production of a protein therapeutic. However, of equal importance is the ability to examine the presence of disulfide scrambled species within protein samples. We accomplished both of these goals by using a multiple proteolytic approach in combination with the sensitivity of nanoLC-MS. It is likely that mapping the disulfide bonds in native and non-native configuration will require empirical testing of multiple proteolytic enzymes in a strategy similarly described for GCase.

The properties and distribution of residues within the region of interest are important to consider when designing an approach to analyze cysteine-rich protein segments. In the case of mapping GCase, a sequential digest using two enzymes with different proteolytic activities proved to be the best strategy to discern all possible disulfide linkages present within our samples. First, Asp-N was used to generate a short peptide containing the cysteine residues to be analyzed and minimize the complexity of the MS analysis for the subsequent digestion reaction. Then we used thermolysin, which cleaved hydrophobic residues within the A1 peptide and allowed us to probe for disulfide-linked peptides unique to the native and scrambled disulfide configurations.
Using this approach we confirmed the native disulfide connectivity of GCase and detected the presence of a minor S2 scrambled species in a long-term stability sample (Figure 2.18).

Traditionally, trypsin and Lys-C are used for peptide mapping and often generate peptides that are too large to gain significant sequence data or have poor peptide recovery. Indeed, both of these situations occurred with GCase in which Lys-C generated a peptide too large for significant fragmentation data and trypsin led to poor recovery of the disulfide containing N-terminal peptide.

There are numerous proteolytic enzymes commercially available that are specific for protein characterization experiments. These enzymes are critical to achieve consistent reactions necessary for similarity and comparability studies. Recently, commercially available on-column immobilized enzymes have been shown to be fast, efficient, and reproducible alternative methods for automating digestions and LC-MS analysis of protein therapeutics. Despite this growing collection of enzymes, steric hindrances imposed by disulfide bonds may impede the ability of glycosidases or proteases to interact with their substrates. PNGase F activity (at 25 °C), and proteolytic activity of pepsin and Asp-N was diminished at regions juxtaposed with N-terminal cysteine residues. For example, the ability of Asp-N to hydrolyze N-terminal to D19 was significantly diminished.

Protein therapeutics require extensive monitoring throughout discovery, development, and formulation. This is critical to bring a safe and effective drug to market. The more we know about complex protein therapeutics in vitro and in vivo will in turn improve the quality of the product we provide to the patients. An important factor in
making this possible is the continuous advancement of the technology to address the challenges we may face as protein therapeutics become more complex.
CHAPTER 3

MASS SPECTROMETRY BASED METHODS FOR DETECTION AND CHARACTERIZATION OF PROTEINS WITH NON-NATIVE DISULFIDE BONDS

3.1 Abstract

Disulfide bonds have a major influence on a protein’s higher order structure and therefore it is necessary to maintain a solid network of techniques to identify non-native configurations. Unlike other non-enzymatic post-translational modifications (PTMs), the presence of fully oxidized disulfide-scrambled isomers will not alter a protein’s molecular weight and is indistinguishable from the native state. Therefore, this modification is more challenging to detect and is typically identified through the characterization of disulfide-containing peptides by liquid chromatography and mass spectrometry (LC-MS). However, sample preparation can be laborious and LC-MS data can be time-consuming to analyze. Furthermore, this method fails to provide information pertaining to the effect non-native disulfides impart on protein structure.

Here we used lysozyme (LYZ) as a model protein to develop a strategy to detect the presence of disulfide scrambling. We generated a mixture of disulfide-scrambled isomers, from which three were isolated by reverse phase chromatography. The purified isomers were then characterized and disulfides mapped by proteolysis and LC-MS. These species were subjected to electrospray ionization (ESI) MS to assess their solvent-exposed surface area by examining their charge state distribution. In this way, we generated a spectral fingerprint of each specific isomer, which was clearly distinct from
the native profile. We then applied charge state distribution analysis to two protein therapeutics, interferon β-1a (IFNβ-1a) and GCase. The appeal of using ESI-MS for monitoring overall solvent accessibility is that it is a quick experiment and requires little sample. Under carefully controlled conditions, a protein’s charge state distribution can reveal its structural integrity. This strategy is shown to be a rapid, effective, and sensitive assessment tool to monitor for non-native disulfide bonds.

3.2 Introduction

With the introduction of recombinant DNA technology, biologics have become widely used in the pharmaceutical industry with the intent of generating protein-based treatments for disease. As of 2010, protein therapeutics consisted of more than 30% of pharmaceutical products that include treatments for cancer, genetic disorders, and autoimmune diseases (DiMasi, Feldman et al. 2010; Sethu, Govindappa et al. 2012). However, unlike organically synthesized small molecules, protein therapeutics are immensely more complex in their production, composition, and structure, making them more challenging to monitor (Berkowitz, Engen et al. 2012). Protein therapeutics are produced in living cells which can add to their complexity through the presence of host cell proteins and the addition of post-translational modifications (PTMs) that can affect the proper formation of tertiary and quaternary structures. Beyond their production, it is necessary for protein therapeutics to be formulated and stored appropriately such that they retain their higher order structure (Manning, Patel et al. 1989; Sofer 1995; Chi, Krishnan et al. 2003).
Characterization of protein therapeutics can present challenges largely due to the heterogeneity of PTMs. For example, glycosylation is present in many protein therapeutics and can affect their stability, efficacy, solubility, and immunogenicity (Walsh and Jefferis 2006; Li and d'Anjou 2009). Often glycoproteins exhibit heterogeneity in their glycan-site occupancy and composition. This can result in difficulties when characterizing glycoproteins and require specific methods to assess these samples for their glycosylation pattern.

Disulfide bonds are another type of PTM that are critical for protein stability and higher order structure (Berkowitz, Engen et al. 2012). This covalent modification occurs when two free thiols oxidize to form a bridge between two cysteine residues, which can occur enzymatically or chemically depending on the surrounding environment (Manning, Chou et al. 2010). Proteins that contain disulfide bonds are largely dependent on their proper formation to ensure the correct higher order structure is maintained. Within the cell, proper disulfide connectivity is typically formed in the oxidative environment of the ER with the aid of protein disulfide isomerases (Depuydt, Messens et al. 2011). Native disulfide bond formation is also largely dependent on the formation of non-covalent structure thereby enabling proper disulfide linkages to form (Weissman and Kim 1995). Once secreted, the chemical environment is important for maintaining the presence of native disulfides.

For a purified protein, altering the chemical environment to any unfavorable conditions can result in the formation of non-enzymatic PTMs. These include deamidation, oxidation, and disulfide scrambling, which can affect its higher order structure (Bobst, Abzalimov et al. 2008; Kaltashov, Bobst et al. 2012). These chemical
modifications can also affect the safety and efficacy of a protein therapeutic (Chi, Krishnan et al. 2003; Zhang, Chou et al. 2011). Disulfide scrambling possesses a unique quality in that it may alter a protein’s conformation, however its molecular weight remains unchanged and therefore will not be detected by a screening method such as intact MW analysis (Kaltashov, Bobst et al. 2012). Therefore, given the influence non-enzymatic PTMs may have on the structure of a protein therapeutic, it is important to develop methods to monitor for these modifications (Berkowitz, Engen et al. 2012)(1993).

Currently, the method used to characterize protein therapeutics involves detailed analysis of the primary sequence and PTMs by a combination of proteolysis followed by liquid chromatography and mass spectrometry (LC-MS) (Kaltashov, Bobst et al. 2010). This strategy lacks information pertaining to the structural impact non-enzymatic PTMs may have. While there are many techniques that can probe the tertiary and quaternary structure of a protein (i.e. size exclusion chromatography, light scattering, nuclear magnetic resonance, x-ray crystallography), they can be time consuming and require substantial protein sample. ESI-MS can be used as an alternative approach to detect conformational changes of a protein while minimizing experimental time, sample handling and quantity.

Under carefully controlled conditions, the charge state distribution can be used to gain insight into the solvent-exposed surface area of a protein (Konermann and Douglas 1998; Kaltashov and Mohimen 2005). Additionally, ESI-MS can simultaneously provide intact molecular weight analysis as well as information pertaining to a protein’s PTMs. Therefore, ESI-MS is valuable as an initial characterization technique to detect structural changes that may occur due to enzymatic and non-enzymatic PTMs.
Using lysozyme (LYZ) as our model protein, we aimed to assess whether ESI-MS can be used to detect the non-enzymatic PTM disulfide scrambling. LYZ is a 14 kDa protein with four disulfide bonds and has the capability of forming 28 fully oxidized disulfide isoforms. We generated and purified disulfide scrambled isoforms of LYZ and monitored their impact on structure by charge state distribution analysis. Furthermore, the protein therapeutic IFN-β1a was subjected to disulfide scrambling conditions and analyzed by ESI MS charge state distribution. GCase, which was shown to have a small amount of disulfide scrambling in the N-terminal region of the protein and was also analyzed by charge state distribution. Here we demonstrate that ESI-MS is a powerful tool that can distinguish structural changes that accompany non-native disulfide bonds.

3.3 Materials and Methods

3.3.1 Materials

Chicken egg white LYZ L4919 (Sigma-Aldrich), was buffer exchanged into 50 mM Tris pH 8.0 (Fisher Scientific) using a 5,000 MWCO Amicon ultra centrifugal filter (Millipore). Mobile phases for RP-HPLC included water with 0.1 % trifluoroacetic acid (TFA) (Fisher Scientific) and acetonitrile (ACN) (Fisher Scientific) with 0.1 % TFA. A substitution of TFA for formic acid (FA) (Acros) was used for LC-MS. Sequencing grade immobilized trypsin (Promega) and immobilized pepsin from *Bacillus thermoproteolyticus rokko* (Sigma-Aldrich) were used for characterizing LYZ disulfide isomers. IFNβ-1a was given to us from Biogen Idec. and both GCase control and stability samples were given to us from SHIRE Human Genetic Therapies.
3.3.2 Circular Dichroism

LYZ was dissolved in either 10 mM ammonium acetate as a near-native solvent or a 50/50 mixture of ACN/H₂O with 0.05 % FA as a denaturing solvent to a final concentration of 5 μM. Circular dichroism was performed with the following parameters: scanning speed of 20 nm/sec, accumulation of 5, data pitch at 0.2, band width of 1 nm, while scanning the UV range of 190-250 nm.

3.3.3 Disulfide Scrambling

Thermal and chemical disulfide scrambling was adapted from methods derived from the Chang laboratory at the University of Texas (Houston, TX) (Chang and Li 2002; Chang 2009; Chang 2009; Chang, Lu et al. 2009). Thermal disulfide scrambling was prepared by incubating 0.5 mg/mL LYZ, 0.1 mM beta-mercaptoethanol (BME) (Sigma-Aldrich) and 50 mM Tris-HCl pH 8.0 (Fisher Scientific) at 75°C. Each sample was incubated for 0, 5, and 10 minutes. Subsequently, the reaction was quenched through the addition of 0.1 % TFA. The 0.1 % TFA quench solution was also added to the control sample for thermal disulfide scrambling, but was not subjected to elevated temperatures and reducing agent (0.1 mM BME).

Samples for chemical-induced disulfide scrambling were prepared by incubating 0.5 mg/mL LYZ with 6 M GuHCl (Thermo Fisher), 0.1 mM BME (Sigma-Aldrich), and 50 mM Tris HCl pH 8.0 (Fisher Scientific) for 45 minutes, 1.5 hours, and 4.5 hours. A control sample was incubated in a denaturing solution of 6 M GuHCl lacking reducing agent (0.1 mM BME). The samples were immediately injected into the HPLC for separation by reverse phase chromatography.
Both thermal and chemical disulfide scrambling samples were injected into a 50 µL loop on an Agilent 1100 HPLC using a 4.6 x 250 mm C-18 reverse phase column (Varian 218TP54). The gradient used was from 0-5 minutes at 15 % M.P. B, 5-15 minutes from 15-35 % M.P. B, 15-40 minutes from 35-45 % M.P. B., and washed using 95 % M.P. B. LYZ variants elute during the shallow gradient of 34-45 % B. Each fraction was collected manually and subsequently lyophilized. Chemical disulfide scrambling for 1.5 hours was chosen for further characterization by mass spectrometry.

IFNβ-1a was subjected to chemical-induced disulfide scrambling with 6 M GuHCl, 0.05 mM BME, and incubated at room temperature for 3 hours. Following treatment, the sample was injected onto a C-18 column (Varian 218TP54) for reverse phase (RP) HPLC separation. GCase was also exposed to disulfide scrambling conditions that included 6 M GuHCl and 0.05 mM BME. This sample was incubated for 0.5, 1, 2, 4, and 6 hours.

3.3.4 Disulfide Mapping of LYZ

Chemical-induced disulfide scrambled species of LYZ were purified by RP-HPLC and characterized by liquid chromatography-mass spectrometry (LC-MS). Each fraction was re-suspended in 50 mM Tris-HCl pH 8 with 30 % ACN and digested with immobilized trypsin for three hours. Immobilized trypsin was removed by centrifugation and the tryptic peptides were injected onto a Gemini C-18 reverse phase column (Phenominex) using an Agilent 1100 HPLC directly connected to a QSTAR-XL quadrupole time-of-flight mass spectrometer (AB SCIEX Framingham, MA). The data
was analyzed manually using the instrument software, Analyst QS (AB SCIEX Framingham, MA).

To characterize the disulfide bonds within the T9-T11 peptide dimer, a second protease, pepsin, was introduced. The T9-T11 peptide dimer, which contains four of the eight cysteines, was isolated from reverse phase chromatography and lyophilized. The peptide was then resuspended in a solution of 0.1 % FA and 10 % ACN with a 1:4 (v:v) ratio of immobilized pepsin to T9-T11 peptide and digested for five minutes. The sample was then analyzed by direct infusion into a QSTAR-XL mass spectrometer.

3.3.5 ESI-MS

Native and disulfide scrambled isomers of LYZ and fractions of IFN-β1a were fractionated manually by RP-HPLC, lyophilized, and analyzed by mass spectrometry. Each sample was re-suspended in a denaturing solution of (50/50) ACN and water (v:v) with 0.05 % FA and immediately analyzed by ESI-MS using a Q-STAR XL mass spectrometer at 5 µL/min. Spray conditions and instrument controls were kept mild at an ion spray voltage of 4500 V and declustering potential of 60 V. Data analysis for the molecular weight of each scrambled species was calculated as an average mass (Mavg). The average mass for each charge distribution was calculated separately.

3.4 Results and Discussion

3.4.1 The Disulfide Bonds in LYZ Impose Conformational Constraints under Denaturing Conditions

The solvent-exposed surface area of a protein provides a useful and rapid assessment of its three-dimensional structure. The extent of protonation during ESI-MS is
directly correlated with its solvent-exposed surface area (Konermann and Douglas 1998; Kaltashov and Mohimen 2005). ESI-MS can use native-like solvents compatible with MS to maintain conformation and therefore retain information pertaining to its structure in solution (Dobo and Kaltashov 2001; Bobst, Abzalimov et al. 2008). In most cases ammonium acetate is used due to its volatility while maintaining necessary salt bridges at a physiologically relevant pH. However, it is critical to perform these experiments under carefully controlled conditions, which include minimizing the capillary voltage, and lowering the declustering potential to prevent charge stripping.

Following these guidelines we used ESI-MS to examine the charge-state distribution of natively oxidized LYZ. Under these conditions, we observed a narrow charge state distribution with the predominant charge state of [M+8H]^{8+} (Figure 3. 1A). This finding demonstrates the rigidity of LYZ, which is consistent with higher order structural data demonstrated by NMR and hydrogen-deuterium exchange MS (Redfield and Dobson 1988; Chung, Nettleton et al. 1997).

Proteins exposed to a high concentration of organic solvent typically result in denaturation of its native conformation. Similarly, an acidic pH (typically accomplished with the addition of acetic, formic, or trifluoroacetic acid) can lead to protein unfolding. As a result, denatured proteins typically exhibit a wide charge state distribution with increased protonation. To examine the effect of denaturing conditions on charge state distribution, we reconstituted LYZ in 50% ACN with 0.05% FA. In this denaturing environment the solvent accessibility remained relatively unchanged, with a subtle shift towards a higher average charge of [M+9H]^{9+} (Figure 3. 1B). The limited impact
denaturing conditions had on LYZ’s charge state distribution is likely a result of the significant structural constraints provided by its four disulfide bonds.

To investigate the influence disulfide bonds have on LYZ’s charge state distribution, we examined reduced LYZ in both near-native and denaturing conditions. When performed in a native buffer (10 mM ammonium acetate pH 6.7), LYZ was unstable and precipitated out of solution. However, reduced LYZ under denaturing conditions was soluble and was examined by ESI-MS. Upon thiol reduction, we observed a dramatic widening in its charge envelope with a significant shift towards higher charge states (Figure 3. 1C). These results demonstrate the polypeptide is no longer constrained, which caused an increase in protonation during ESI-MS. Furthermore, the reduced species has a molecular weight of 14313 Da, a mass increase of 8 Da compared to fully oxidized LYZ with a molecular weight of 14305 Da (Figure 3. 1D).

Next we used CD to probe the secondary structure of LYZ in native and denaturing solvents and no changes were observed (Figure 3. 1E). However, although minimal, the tertiary structure observed by ESI-MS showed a slight change in its overall solvent accessibility prior to disulfide reduction (Figure 3. 1A,B). Taken together, we conclude that the change in solvent accessibility when exposed to acidic and organic conditions is not due to the alteration of secondary structure present in native LYZ.

Our results provide a useful framework for developing a method to rapidly assess the integrity of a protein’s higher order structure as a result of the formation of non-native disulfides. We propose that the native charge envelope of a protein can be used as a control for measuring protein misfolding. In the case of LYZ, disulfide bonds are a major factor in maintaining its three dimensional state.
Figure 3. Charge state distribution of natively oxidized and reduced LYZ. (A) ESI-MS of LYZ in a near native solvent shows a narrow charge distribution with the $[\text{M}+7\text{H}]^{7+}$ and $[\text{M}+8\text{H}]^{8+}$ charge states as the most abundant. (B) When LYZ is subjected to an organic solvent of 50% acetonitrile with 0.05% formic acid (pH ~3.5) its charge distribution shifts slightly to a more abundant $[\text{M}+8\text{H}]^{8+}$ charge state. (C) Upon disulfide bond reduction, there is a dramatic change in the protein’s solvent accessibility shown by the wide charge state distribution and high charge states. (D) A deconvoluted spectrum of oxidized and reduced LYZ can be distinguished by a molecular weight shift of 8 Da (MW of oxidized LYZ is 14,305 Da and reduced LYZ is 14,313). (E) Near UV CD traces show the similarity in secondary structure between natively oxidized LYZ in near-native and organic solvent.
3.4.2 The Production, Purification, and Characterization of Non-Native LYZ Isoforms

In order to investigate whether ESI-MS is capable of detecting changes in protein structure as a result of disulfide scrambling, we attempted to generate LYZ with non-native disulfide bonds. We chose to use LYZ as a model protein because it has the capability of forming up to 27 fully oxidized disulfide scrambled isoforms that are structurally distinct from native LYZ.

We aimed to generate and purify disulfide scrambled species using both thermal and chemical approaches. Incubating LYZ with a low amount of reductant at a temperature of 75°C produced multiple LYZ variants (Figure 3.2A). However, this mixture was largely heterogeneous which resulted in poor separation by RP-HPLC and prevented charge state distribution analysis of pure LYZ disulfide scrambled species. Furthermore, thermal-induced disulfide scrambling may result in complications when analyzing mass spectra, as a result of other non-enzymatic modifications. Thermal stress is known to induce many protein modifications such as oxidation, deamidation, and degradation that can complicate downstream MS charge state distribution analysis.

Alternatively, we used a chemical approach to induce disulfide scrambling in LYZ while minimizing other modifications. LYZ was incubated for 45 minutes, 1.5 hrs, and 4.5 hrs in 6 M GuHCl with a low concentration of reducing agent followed by separation by RP-HPLC (Figure 3.2B). This chemical disulfide scrambling procedure led to an abundant LYZ variant with a retention time of 26.5 minutes after incubation for 4.5 hrs. At 1.5 hrs there were four abundant species identified by RP-HPLC and were purified for further evaluation of disulfide scrambling (Figure 3.2C).
Figure 3. 2 Promoting disulfide scrambling with thermal and chemical denaturation of LYZ.

(A) Thermal disulfide scrambling. LYZ (0.5 mg/mL) was subjected to an elevated temperature of 75°C with a thiol reactive agent of 0.05 mM BME. After incubation at elevated temperatures for a series of timepoints, the sample was quenched 2-fold with 1% TFA and immediately injected into the HPLC for reverse phase separation. Fractions were collected manually as they eluted from a C-18 reverse phase column and detected at a wavelength of 280 nm. (B) Chemical disulfide scrambling. LYZ (0.5 mg/mL) was subjected to a denaturant of 6 M guanidine hydrochloride, with a thiol reactive agent of 0.05 mM BME. After incubation in a denaturing solution over a period of various times, the sample was subsequently injected into the HPLC for reverse phase separation. Similar to thermal disulfide scrambling of LYZ, fractions were collected manually and subsequently lyophilized. (C) Fraction collection of 1.5 hrs of chemical disulfide scrambling. Fractions labeled 1-4 were chosen for further characterization.
Characterization of LYZ disulfide connectivity in native and chemically scrambled LYZ was accomplished by a combination of proteolytic enzymes and LC-MS and MS-MS. Native LYZ (control) was subjected to RP-HPLC and fractionated into two species, with the major species eluting at 31 minutes (Figure 3.2B). The collected fractions were digested with immobilized trypsin and characterized by LC-MS. We were able to confirm two of the four disulfide bridges to be C6-C127 and C30-C115. Furthermore, tryptic peptide T11 contained three cysteine residues linked to the T9 peptide with one cysteine. To identify the connectivity we isolated the T9/T11 peptide and digested it with immobilized pepsin. Through MS analysis of the digest, we were able to map the remaining disulfide linkages to be C64-C80 and C76-C94 (Figure 3.3), which is consistent with the established disulfide connectivity of LYZ (Canfield and Liu 1965). We attempted to characterize the minor species detected by RP-HPLC, which eluted at 28 minutes (Figure 3.2A, B). However, we were unable to determine its disulfide connectivity due to its low abundance.

The 1.5 hr chemical treatment condition provided four abundant LYZ variants as observed by RP-HPLC (Figure 3.2C). We executed a similar strategy outlined above and performed LC-MS of proteolytic digests of each of the purified species. As expected, fraction 3 had the same retention time as the untreated control sample and contained the native disulfide connectivity (Figure 3.2C, Figure 3.3). Next, fractions with distinct retention times (1, 2, and 4) were characterized for their disulfide connectivity. Fraction 1 contained two native disulfide bonds (C6-C127 and C30-C115), but also contained two atypical connections (C64-C76 and C80-C94). Similarly, fraction 4 also contained two native (C64-C80 and C76-C94) and two scrambled disulfide bonds (C6-C30 and C115-
Finally, fraction 2 contained all non-native disulfide bonds (C6-C30, C64-C76, C80-C94, and C115-C127), in which all disulfide connections are formed following a sequential order (Figure 3. 3A,B). As shown in Figure 3. 2B, the abundance of fraction 2 increased with longer incubation times. This demonstrates that prolonged chemical treatment results in an increasingly disordered LYZ. To investigate charge state distribution as an approach to monitor disulfide scrambling, we subjected each of the LYZ variants to ESI-MS analysis.
Figure 3. 3 Characterizing LYZ disulfide scrambled isoforms.
LYZ disulfide scrambled species were characterized by a trypsin digestion and subsequent pepsin digestion of the T9-T11 disulfide linked peptide. A. Shows a mass spectrum of one native disulfide bond present in fraction 3 and one non-native disulfide bond present in fraction 4. By observing the presence and absence of disulfide linked peptides, we were able to characterize all disulfide bonds for all fractions (F1-F4). B. Disulfide linkages present in each fraction collected from RP-HPLC (F1-F4).
3.4.3 Changes in the Charge State Distribution of Scrambled LYZ Variants Observed by ESI-MS

To validate charge state distribution as a method to detect proteins containing non-native disulfide bonds, we examined each individual disulfide scrambled isoform of LYZ identified in Figure 3.3 by ESI-MS. LC-MS analysis of fraction 3 revealed it to be natively oxidized LYZ (Figure 3.3). However, a bimodal charge distribution was observed by ESI-MS (Figure 3.4). Both charge state distributions consist of fully oxidized LYZ identified by a molecular weight of 14305 Da. This pattern is likely a result from one of two possibilities. Either a minor disulfide scrambled species co-elutes in fraction 3 and was not detected by LC-MS peptide mapping, or while incubating with the reductant, LYZ’s disulfide bonds re-formed into their correct orientation, but other structural rearrangements occurred and were trapped in a disordered state by disulfide bonds.

Next, we examined the charge state distribution of the purified disulfide scrambled species. Analysis of fraction 1 revealed widely disseminated charge states that showed overlap between native and reduced LYZ (Figure 3.4). RP-HPLC of chemically induced disulfide scrambling showed that fraction 1 increased in abundance until 1.5 hrs and subsequently decreased over time (Figure 3.2B). This finding suggests that fraction 1 is an intermediate species that ultimately transitions to the formation of disulfide bonds in a sequential order.

Fractions 2 and 4 have a molecular weight of fully oxidized LYZ and exhibit a charge state distribution almost identical to the reduced form (Figure 3.4). These results illustrate the comparable extent of solvent accessibility between the denatured state of reduced LYZ and these two non-native species. Collectively, these findings show that
ESI-MS is a rapid and sensitive technique to probe for structural changes that occur due to disulfide scrambling.
Figure 3. Charge state distribution of LYZ disulfide scrambled isomers. ESI mass spectra are shown for native, reduced, and disulfide scrambled isomers of LYZ. All fractions were re-suspended in a 50/50 ACN/H₂O solution with 0.05% formic acid and analyzed by ESI-MS. The charge distributions of all non-native LYZ isomers more closely represent the distribution portrayed by the reduced LYZ in comparison to that of the charge distribution of natively oxidized LYZ. The [M+9H]^{9+} and [M+14H]^{14+} are zoomed in for fractions 1-4 to show the calculated MW of fully oxidized LYZ with a MW of 14305.
3.4.4 Disulfide Scrambling and ESI-MS of the Protein Therapeutic IFNβ-1a

IFNβ-1a is a 22.5 kDa protein therapeutic currently approved to alleviate symptoms of multiple sclerosis. This protein contains three cysteine residues, two of which form a disulfide bond (Karpusas, Nolte et al. 1997). These attributes make IFNβ-1a an ideal candidate to pursue because only two non-native disulfide bonds are possible, which simplifies the analysis. Structurally, its free sulfhydryl is not localized within close proximity to the native disulfide bond and would most likely not form scrambled disulfide bonds within the monomer (Figure 3. 5). However, the presence of a free sulfhydryl increases the possibility for multimers forming through intermolecular disulfide bonds, which was hypothesized to be the root cause for the formation of aggregates for deglycosylated human recombinant IFNβ-1a (Karpusas, Whitty et al. 1998). This is particularly relevant, as aggregation of IFNβ-1a elicits an immune response and promotes the production of neutralizing antibodies that reduce its efficacy (van Beers, Jiskoot et al. 2010) (Bertolotto, Deisenhammer et al. 2004).

Figure 3. 5 Crystal structure of IFNβ-1a (1AU1) with cysteine residues shown in blue.
To investigate ESI-MS charge state distribution as a method to assess disulfide scrambling in IFNβ-1a, we used the procedure that we developed to scramble LYZ. Similarly, a non-scrambled sample of IFNβ-1a was denatured in 6 M GuHCl and used as a reference for its retention time profile by RP-HPLC and charge state distribution by ESI-MS (Figure 3. 6 and Figure 3. 7). ESI MS CSD revealed a bimodal charge distribution for fraction 1. The poor separation from the main peak is likely the result of these two distributions, one resulting from the native conformer and the other with the wider distribution in the lower m/z region may be due to disulfide scrambling which caused an increase in solvent accessibility (Figure 3. 7). Fraction 2, which is consistent with the retention time of the control, had a similar charge state distribution as unscrambled IFNβ-1a. Fraction 3 displayed a similar charge state distribution as the control and fraction 2, had the most abundant charge state of [M+14H]^{14+}. Additionally, we observed a fragment of IFNβ-1a that had a loss of 893 Da.

![Figure 3. 6 RP-HPLC purification of IFNβ-1a disulfide scrambling](image)
**Figure 3.7 Charge state distribution of RP-HPLC fractions of INFβ-1a**

INFβ-1a was purified from RP-HPLC and analyzed by ESI-MS. The charge state distribution (CSD) for four samples is shown above. The untreated control sample contains a distinct CSD, which is similar to fraction 2 and fraction 3. Moreover, fraction 3 shows a fragment with a loss of 893 Da shown as orange circles. Fraction 1 shows a wider CSD at the lower m/z region of the spectrum that is indicative of an increase in solvent accessibility.

### 3.4.5 ESI-MS of GCase Stability Sample

A 5 year stability sample with a known presence of disulfide scrambling, as shown in chapter 2, was analyzed by ESI MS charge state distribution (Figure 3. 8). Although slight, the subtle shift in charge state distribution may be due to the presence of a minor population of GCase with non-native disulfide bonds. The small shift, is likely due to a combination of factors, which include its low abundance and that the non-native
disulfide bonds are shifted by a single amino acid in the primary sequence, which would likely result in a modest change in its tertiary structure (Figure 3.9). It is possible that the sensitivity of ESI-MS is able to detect minor structural changes in a small subpopulation of a protein sample. However, this hypothesis would need to be supported by further structural analysis of S2 disulfide scrambled species. To accomplish this goal, we would need to develop a method to generate and isolate S2 prior to analysis by ESI-MS. This could be achieved through RP-HPLC. Furthermore, other structural techniques such as hydrogen deuterium exchange (HDX) MS, can be used to demonstrate the structural changes that accompany disulfide scrambling into the S2 isoform.
Figure 3. 8 GCase charge state distribution of stressed and stability samples
GCase analyzed by ESI-MS in 10mM ammonium acetate with their spectra shown in black and red, with a y-axis offset of 10% applied to GCase control. GCase was also diluted to 10 μM in denaturing organic solvent and shown in the grey trace. The largest difference observed is the increasing [M+17H]^{17+} in the 5 year at 4°C stability sample with the calculated average charge for both GCase control and 5 year stability sample in near-native solvent also shown above.
3.4.6 Disulfide Scrambling of GCase

To enhance disulfide scrambling, in particular the S2 species, we incubated GCase in conditions favorable for disulfide scrambling. In Chapter II, we showed that a small portion of GCase was disulfide scrambled in a stability sample that was incubated at 4 °C for over five years. However, due to the 5 cysteine residues present in GCase, many other possible disulfide scrambled isomers are possible. To be able to use ESI-MS to examine those distinct species, we attempted to generate non-native disulfide bonds by incubating GCase in denaturing solution of 6M GuHCl and 0.05 mM BME. Under these conditions, a new peak at about 19 minutes was observed and increased over time (Figure 3. 10). In the future we would like to isolate this peak and map its disulfide bonds. Ideally we would like to isolate the S2 scrambled species and show the charge state distribution of the pure S2 species. This would provide further evidence of whether or not the difference in charge state distribution observed in the 5 year stability sample is related to the presence of non-native disulfide isoforms.
Figure 3. 10 GCase incubated in conditions favorable for disulfide scrambling
GCase (0.5 mg/mL) was injected onto an HPLC after incubating in 6M GuHCl with 0.05 mM BME for increasing amounts of time ranging from 0 to 6 hours.

3.5 Conclusions

Here we developed a method to monitor non-native disulfide-induced changes in the solvent-exposed surface area and hence structure of LYZ. It is our aim that this strategy be extended to analyze any protein of interest. Moreover, this method can be tailored to provide a basic assessment of a protein’s structure, or it can be designed to yield a more detailed analysis. To determine whether a protein deviates from its native conformation, a comparison of the native and sample’s charge state distribution is
required (Konermann and Douglas 1998; Kaltashov and Mohimen 2005). However, a more in depth analysis may help to elucidate whether specific modifications, such as disulfide scrambling, have occurred. We first generated disulfide scrambled species and then profiled their charge state distribution. In this way, a fingerprint of each isomer’s charge distribution can be used to monitor whether structural changes are present due to specific non-native conformers.

We characterized three disulfide scrambled isomers of LYZ. Fraction 2 represented the most extreme case in which all disulfide bonds are non-native (Figure 3. 3). As a consequence, the structure is significantly impacted and has a charge state distribution similar to that of reduced and denatured LYZ (Figure 3. 4). Figure 3. 11 shows a model of the species present in fraction 2 and illustrates the likely unstructured nature of this isoform.
Figure 3.11 A representation of LYZ structurally distorted by disulfide scrambling. (A) Crystal structure of lysozyme (1GXV) in which cysteine residues and its side chains are in blue. (B) A PyMol distortion of lysozyme, 1GXV, to represent a scenario where disulfide scrambling has occurred. In this scenario disulfide scrambling caused LYZ to generate a disulfide pattern with sequential disulfides representing the “beads on a string” formation.

The four disulfide bonds in LYZ exert significant conformational restraints on the protein and are a major factor in its higher order structure. Therefore it is not surprising that the chemical-induced rearrangement of these disulfide bonds significantly altered its solvent-exposed surface area. For proteins whose disulfide-bonds do not play as significant a role in modulating protein structure, more sensitive techniques such as hydrogen-deuterium exchange (HDX) MS, NMR, or x-ray crystallography may be required. However, these techniques are often time consuming and require significant amounts of protein.

Applying this method to a protein therapeutic shows promise. Another species of IFN-β1a is shown to be eluting prior to the control through RP-HPLC and this fraction
was analyzed by ESI MS charge state distribution. We saw that fraction 1 contained a species that had a much wider charge state distribution when subjected to disulfide scrambling conditions. However, further analysis, such as accurate molecular weight measurements with a high resolution mass spectrometer can confirm that the species observed is indeed fully oxidized. Furthermore, an increase in separation of the main species and scrambled species will be beneficial. Unfortunately, the remaining IFNβ-1a contained a significant portion of aggregate and therefore improved chromatography elution patterns could not be repeated.

The second protein therapeutic, GCase, was analyzed by ESI MS without prior separation. Although we are aware of the presence of GCase S2 scrambled species within the 5 year stability sample, further confirmation will be needed to show that the change in charge state distribution is due to disulfide scrambling. This includes isolating the pure S2 scrambled species and analyzing its charge state distribution similar to what is shown in Figure 3. 8.

A benefit of ESI-MS is its ability to rapidly provide an assessment of a protein’s structure while using minimal sample. This can be particularly useful in initial stages of development where sample quantity might be low and a large number of proteins are screened. Additionally, ESI-MS requires reduced sample handling which may decrease the chance for assay-induced modifications.

The ability to quickly and easily monitor changes in protein structure due to alterations in disulfide connectivity has significant practical applications. For a disulfide-containing protein therapeutic, charge state distribution analysis can be a tool that can rapidly assess whether a protein has been properly folded. This can be used throughout
development including examining the impact of the expression system, cell culture conditions, purification strategies, formulation, and storage conditions on a protein therapeutic. This technique can be appealing in the development of biosimilars in which initial screens could be performed with ESI-MS to hone in on the most representative drug candidate as compared to the innovator drug.

Non-enzymatic post-translational modifications such as oxidation, deamidation, and disulfide scrambling can impact the structure and therefore efficacy and safety of a protein therapeutic. However, disulfide scrambling is the only modification that does not alter the molecular weight of a protein, but can substantially influence its higher order structure. Therefore, alternative methods are required to monitor changes in disulfide connectivity. Here we developed a rapid and effective strategy using ESI-MS to monitor global structural changes in LYZ that occur due to disulfide scrambling.
CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Conclusions

MS has become an indispensable tool for understanding protein therapeutics at the primary biomolecular level and has become widely accepted as a method to probe their higher order structure. Advancements in MS instrumentation are rapidly evolving, which provide the opportunity to use this technology to develop novel MS-based methods that can be used to better understand protein therapeutics. In this work, we focused on MS-based experimental strategies that detected and characterized native and non-native disulfide bonds at both the primary and tertiary biomolecular level.

Here we examined CSD by ESI MS as a tool to rapidly assess disulfide scrambling in protein therapeutics. CSD profiles are often overlooked and underutilized when analyzing proteins by ESI MS, and under carefully controlled conditions is a powerful tool to quickly assess the structural integrity of protein therapeutics. We demonstrate how ESI MS charge state distribution is used as a rapid tool to assess the structural integrity of a protein due to the formation of non-native disulfide bonds. This was accomplished by isolating and characterizing non-native isoforms of LYZ and assessing the CSD for each species compared to natively oxidized LYZ. Here, we showed that the isolated species of LYZ with non-native disulfide bonds was clearly distinguishable from the native form.
Furthermore, the ability to characterize native and non-native disulfide bonds within cysteine-rich protein segments can be challenging with traditional peptide mapping methods. Therefore, we developed a unique method to identify native and non-native disulfide bonds within the cysteine-rich protein segment of the protein therapeutic GCase. As the portfolio of proteins used for therapeutic purposes broadens, there is an increasing need to expand the characterization-based methods that can accommodate these changes. Here, we demonstrate a tailored method to address this issue in GCase, which contains a cysteine-rich protein segment; however, this approach can also be used as a platform to address this same challenge within similar protein therapeutics. Furthermore, the techniques demonstrated were able to characterize the native disulfide-connectivity by multiple approaches and also detect non-native disulfides in a stressed sample of GCase.

This work provides insight in how MS can be used to address current issues with detecting and characterizing native and non-native disulfide bonds in proteins with a focus on applications for protein therapeutics. There are many questions that can be followed up with the work presented here. This includes the application of inducing disulfide scrambling on a protein therapeutic. How does CSD analysis of disulfide scrambled isoforms compare to the model presented in chapter 3? Initial studies which include inducing disulfide scrambling and separation via RP-HPLC were successful, however the analysis of their CSD and disulfide-bond characterization is yet to be determined. Furthermore, can CSD detect disulfide scrambled conformers without prior isolation? To what extent is the higher-order structure of proteins impacted when
disulfide scrambling occurs? Each section below shows initial work and future directions to address this issue of structural characterization and quantifying disulfide scrambling.

4.2 Structural Analysis of Disulfide Scrambled Lysozyme

In chapter III, we demonstrated that ESI-MS could be used to monitor the higher order structure of LYZ disulfide scrambled isomers. However, it is important to compare ESI-MS observations to other higher order structure techniques such as HDX MS, NMR, and X-ray crystallography. HDX can be used with MS as a method to probe conformational dynamics by incubating proteins with deuterium and measuring its incorporation on the intact or digested proteins. Non-native isoforms of LYZ that were purified and analyzed by ESI-MS showed differences in their charge state distribution, however the specific structural changes that occurred due to non-native disulfide bonds are not known. Here, HDX can be used for detailed structural analysis of non-native disulfide scrambled species.

4.3 Quantitative Analysis of Disulfide Scrambling in Protein Therapeutic

The relative or absolute amounts of proteins, peptides, or post-translational modifications in complex or simple solutions can be quantified using mass spectrometric-driven approaches (Gerber, Rush et al. 2003) (Tao and Aebersold 2003; Beynon, Doherty et al. 2005). Strategies such as the absolute quantification (AQUA) method are particularly relevant in the development of protein therapeutics. The need to monitor the presence and amount of post-translational modifications is important to develop methods that ensure uniformity in the final product. Quantitative analysis of protein deamidation, methionine oxidation, and glycosylation has been performed by mass spectrometry (Liu,
Manuilov et al. 2011). However, there are few methods that can quantify disulfide scrambling. Reverse phase chromatography can theoretically be used to quantify disulfide scrambling if the resolution of peaks between disulfide isoforms is sufficient and if the species are within the detectable range (Chang, Lu et al. 2005). Therefore, a more robust method is needed to effectively quantify non-native disulfide bonds. One way to achieve this goal would be to use synthetic stable isotope labeled internal standard of disulfide-linked peptides that are specific to the disulfide scrambled species of interest. It would be very interesting to perform this experiment with the five year GCase stability sample described in chapter II. There I described the presence of a low abundant disulfide scrambled isoform, which was directly compared to a control sample. Although the S2 disulfide scramble species was identified, we were unable to quantify it due to differences in ionization efficiency between the disulfide-linked peptides.
THE STRUCTURE OF HUMAN GALNS REVEALS THE MOLECULAR BASIS FOR MUCOPOLYSACCHARIDOSIS IV A

Published in final edited form as:

Yadilette Rivera-Colóna, Emily K. Schutskya, Adriana Z. Kitab, and Scott C. Garmana,b,*
aDepartment of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst MA 01003 USA
bDepartment of Chemistry, University of Massachusetts Amherst, Amherst MA 01003 USA

Monitoring Cysteine Conversion to Formylglycine by Mass Spectrometry

The lysozome contains a large number of sulfatases that aid in the degradation of a variety of substrates including carbohydrates, glycosaminoglycans, and proteins. This is accomplished though hydrolysis of the sulfate ester bond. Similar to other enzymes in the lysosome, when dysfunctional, sulfatase deficiency can lead to a variety of lysosomal storage diseases. The activity of sulfatases depends on the recognition of its “CXPXR” motif by formylglycine-generating enzyme to convert a conserved cysteine residue to the rare amino acid formylglycine.

Galactosamine-6-sulfatase (GALNS) is an enzyme present in the lysosome that hydrolyses N-acetylgalactosamine-6-sulfate into sulfate and N-acetylglactosamine. A deficiency in this enzyme leads to the accumulation of mucopolysaccharides in the lysosome, which is known as mucopolysaccharidosis IV A (MPS IV A). This disease is rare and affects approximately 1 in 20,000 live births. Here, we used mass spectrometry to monitor the efficiency of an insect cell expression system to enzymatically convert the cysteine within the “CXPXR” motif to the rare amino acid formylglycine. MS is a powerful tool to determine modifications such as cysteine modifications. Previous
chapters demonstrate methods to detect the formation of non-native disulfides that can render an enzyme inactive. Here, we display MS as a method to detect and confirm the presence of a cysteine modification that is necessary for enzymatic activity of GALNS. Furthermore, with an extracted ion chromatogram (XIC) of both the T4 peptide with and without formylglycine, we can determine their relative abundance and therefore the efficiency of insect cells to convert the Cys in GALNS to the rare amino acid formylglycine.

In supplemental figure S1a we showed a TIC of a reduced and alkylated trypsin digest of GALNS analyzed by nanoLC-MS/MS. An XIC of the tryptic peptide (T4) showed that the majority had Cys79 converted to a formylglycine residue (Figure S1b).

**Materials and Methods to Characterize Cysteine Conversion by Mass Spectrometry**

GALNS was denatured by addition of guanidinium HCl (GuHCl) and ethylenediaminetetraacetic acid (EDTA) to a final concentration of 50 μM GALNS, 6 M GuHCl, and 1 mM EDTA. The protein was then reduced in 20 mM dithiothreitol for 30 minutes and cysteines carboxymethylated in 20 mM iodoacetic acid for 30 minutes in the dark. The protein was exchanged into 100 mM Tris HCl, pH 8.0 on a NAP-5 desalting column prior to trypsin incubation for five hours with immobilized trypsin (Promega). After trypsin removal, GALNS peptides were injected into a nanoLC system (Dionex) for separation on a C18 reverse phase column and analysis by a QStar XL hybrid quadrupole time-of-flight mass spectrometer (AB Sciex). In parallel with LC/MS, the more abundant peptide ions were isolated and subjected to tandem mass spectrometry (MS/MS) for fragmentation by collision-activated dissociation (CAD). This method allowed peptide identification and measurement of the mass of the amino acid at position 79.
Figure A. 1 Mass spectrometry analysis of GALNS with cysteine and formylglycine. GALNS was reduced, alkylated, and subsequently digested with trypsin. The peptide mixture was separated by nanoLC and analyzed using online mass spectrometry for peptide identification and simultaneous CAD for residue specific information. A) nanoLC separation of peptides is shown by the TIC (black) and XIC of the tryptic peptide containing Cα-formylglycine (blue) and carboxymethylated Cys (red) at residue 79. B) Tryptic peptide containing formylglycine was subjected to CAD and various fragment ions were observed including y4 and y5 that represent fragmentation at either end of the formylglycine residue.
BIBLIOGRAPHY


90


