Characterization of Protein Therapeutics by Advanced Liquid Chromatography Coupled with Mass Spectrometry (LC-MS)

By Yu Wang

B.S. in Biological Science, Jilin University, Jilin, China

A dissertation submitted to

The Faculty of
The College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

August 1st, 2017

Dissertation directed by
William S. Hancock
Professor and Fellow, Barnett Institute of Chemical and Biological Analysis
Department of Chemistry and Chemical Biology
Raymond Bradford Bradstreet Chair in Bioanalytical Chemistry
Acknowledgements

I would like to take this chance to thank everyone who helped with my PhD studies.

I am deeply grateful and thankful for my advisor, Prof. William Hancock, who accepts me as a PhD student and shows me the value and charm of analytical science. He is always cheery and supportive. Even as an expert in protein chemistry and biosimilars, he is always patient and sheds light on not only the fundamentals but also the big picture of the projects. His words “as an analytical scientist, you will always be under pressure because your work is so important” will always be inspiring to me along my professional career.

I would like to thank my supervisor, Shiaw-lin (Billy) Wu for his guidance and help with my projects. What he has taught me about LC-MS and biosimilar characterization is of great significance to me.

I would like to thank Dr. Jared Auclair for his help with my projects and lab management.

I would like to thank my committee members, Prof. Michael Pollastri, Prof. Olga Vitek, Prof. George O’Doherty and Dr. Jared Auclair for their help and suggestions about my projects and thesis work.

I would like to thank my collaborators, the Love Lab from MIT, Koch Institute of Integrative Cancer Research for this great opportunity on the Biomedicine on Demand project. I would like to thank Defense Advanced Research Projects Agency (DARPA) for funding us.

I would like to thank the staff in the Department of Chemistry and Chemical Biology, and the Barnett Institute, Cara Shockley, Richard Pumphrey and Emanuelle Hestermann for their help during my study in Northeastern University.

I would like to thank my colleagues in Hancock Lab, Dr. KyOnese Taylor, Dr. Fan Zhang, Di Wu, Yanjun Liu, John De La Parra for their help during my study.
I would like to thank BioAnalytix Inc. for giving me the opportunity of internship and getting exposure to industry projects.

I would like to thank my friends in Northeastern University, Dr. Wanlu Qu, Dr. Yuanwei (Abby) Gao, Shanshan Liu and Dr. Jingkun Li for their support and help in my study and life.

Finally, I would like to thank my parents for their unconditional love and support on my life and study.
Abstract of Dissertation

Statement of challenges of thesis research projects and conclusion based on the result of the studies:

Characterization of proteins produced from novel production platforms are always challenging especially at the process development stages. First, the concentration might be very low, which requires extra steps to concentrate the protein before analysis and high sensitivity analytical workflows. Second, there might be unknown proteolysis or modifications caused by new production platform. Third, high level of contaminants and various products might be in the product in the process development stage.

It is very important to choose the suitable analytical method for the characterization of process development stage materials. Stability study of the reference material is always a good start. Stability study with stressed condition can be used to accelerate the degradation of the proteins. The degradation products characterization can provide information to identify the weak points of the protein. In-solution digestion is suitable for purified material characterization, crude sample such as strain supernatant needs a SDS-PAGE gel separation at first, followed by in-gel digestion. Bottom-up method with multi enzyme digestion strategy can be used to generate optimal size peptides containing the sites of degradation. Complementarily, top-down approach is a quick way to monitor all the product variants in the sample at protein level instead of peptide level.
Table of Contents

Acknowledgements............................................................................................................ ii
Abstract of Dissertation....................................................................................................... iv
Table of Contents ................................................................................................................ v
List of Figures....................................................................................................................... xiv
List of Tables ...................................................................................................................... xvi
List of abbreviations .......................................................................................................... xvii

Chapter 1: Introduction to characterization of biopharmaceuticals using advanced liquid chromatography and mass spectrometry (LC-MS) ............ 1

1.1. Abstract............................................................................................................................. 2
1.2. Protein Therapeutics and Biosimilars ............................................................................ 4
1.3. Post-translational modification of proteins ................................................................... 4
    1.3.1. Oxidation.................................................................................................................. 5
    1.3.2. Deamidation.......................................................................................................... 6
    1.3.3. Glycosylation........................................................................................................ 7
1.4. LC-MS for Characterization of Protein Therapeutics .................................................. 8
    1.4.1. High-performance liquid chromatography (HPLC).............................................. 8

v
1.4.2. Mass Spectrometry

1.4.2.1. Ionization method

1.4.2.1.1. Electrospray (ESI)

1.4.2.1.2. Matrix-Assisted Laser Desorption Ionization (MALDI)

1.4.2.2. Fragmentation method

1.4.2.2.1. Collision Induced Ionization (CID) and High Energy Collision dissociation (HCD)

1.4.2.2.2. Electron Transfer Dissociation (ETD)

1.4.2.2.3. EThcD

1.4.2.2.4. Surface Induced Dissociation (SID)

1.4.2.2.5. Photodissociation

1.4.2.3. Mass Analyzers

1.4.2.3.1. Magnetic Sector

1.4.2.3.2. Quadrupole Mass Spectrometers

1.4.2.3.3. Time of Flight (TOF)

1.4.2.3.4. Quadrupole Ion Trap

1.4.2.3.5. Fourier Transform (Fourier Transform Ion Cyclotron Resonance)

1.4.2.3.6. Orbitrap

1.4.2.3.7. Tandem mass

1.4.2.4. Data Acquisition Method
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4.2.4.1</td>
<td>Data Dependent Mode (DDA)</td>
<td>19</td>
</tr>
<tr>
<td>1.4.2.4.2</td>
<td>Selected Ion Monitoring (SIM) and Selected Reaction Monitoring (SRM)</td>
<td>19</td>
</tr>
<tr>
<td>1.4.2.4.3</td>
<td>Data Independent Mode (DIA)</td>
<td>19</td>
</tr>
<tr>
<td>1.4.2.5</td>
<td>Data Interpretation</td>
<td>20</td>
</tr>
<tr>
<td>1.4.2.5.1</td>
<td>Database search Software</td>
<td>20</td>
</tr>
<tr>
<td>1.4.2.5.2</td>
<td>Manual Inspection</td>
<td>20</td>
</tr>
<tr>
<td>1.4.3</td>
<td>Sample Preparation</td>
<td>21</td>
</tr>
<tr>
<td>1.4.3.1</td>
<td>Bottom-Up</td>
<td>21</td>
</tr>
<tr>
<td>1.4.3.1.1</td>
<td>In-gel digestion</td>
<td>21</td>
</tr>
<tr>
<td>1.4.3.1.2</td>
<td>In solution digestion</td>
<td>21</td>
</tr>
<tr>
<td>1.4.3.1.3</td>
<td>Novel sample preparation methods</td>
<td>22</td>
</tr>
<tr>
<td>1.4.3.1.4</td>
<td>Enzymes</td>
<td>23</td>
</tr>
<tr>
<td>1.4.3.2</td>
<td>Top-Down</td>
<td>24</td>
</tr>
<tr>
<td>1.4.4</td>
<td>Critical Quality Attributes (CQAs)</td>
<td>24</td>
</tr>
<tr>
<td>1.4.4.1</td>
<td>Peptide mapping</td>
<td>25</td>
</tr>
<tr>
<td>1.4.4.2</td>
<td>Correct amino acid sequence</td>
<td>25</td>
</tr>
<tr>
<td>1.4.4.3</td>
<td>Characterization of critical quality attributes (CQAs)</td>
<td>29</td>
</tr>
<tr>
<td>1.5</td>
<td>Isoelectric Focusing analysis of protein therapeutics</td>
<td>29</td>
</tr>
<tr>
<td>1.5.1</td>
<td>Capillary isoelectric focusing</td>
<td>30</td>
</tr>
<tr>
<td>1.5.2</td>
<td>Small scale multi-compartment electrolyzers (MCE)</td>
<td>30</td>
</tr>
</tbody>
</table>
1.6. Continuous Manufacturing ........................................................................................................ 31

1.6.1. Bio-MOD platform ................................................................................................................ 32

1.6.2. Process Related Impurity and Product Related Impurity .................................................. 33

1.6.2.1. Incomplete Removal of leader sequence ........................................................................ 33

1.6.2.2. Cleavage of target protein by proteases in the host cell .................................................. 34

1.7. References ................................................................................................................................ 35

Chapter 2: Integrated bottom-up and top-down liquid chromatography-mass spectrometry (LC-MS) approach for characterization of recombinant human growth hormone degradation products. ........................................................................ 51

2.1. Abstract ..................................................................................................................................... 52

2.2. Introduction ............................................................................................................................... 53

2.3. Materials and Methods ............................................................................................................. 54

2.3.1. Reagents ............................................................................................................................... 54

2.3.2. Stability of human growth hormone ..................................................................................... 55

2.3.3. Oxidation of human growth hormone ................................................................................... 55

2.3.4. Preparation of 2-chain human growth hormone ................................................................. 55

2.3.5. SDSPAGE of hGH degradation products ............................................................................ 55

2.3.6. Analysis of tryptic digest of human growth hormone (bottom-up method) ................. 56

2.3.7. LC-MS: Dionex Ultimate 3000 LC-LTQXL ......................................................................... 57
2.3.8. Analysis of intact human growth hormone (top-down method), LC-MS: Eksigent LC-Bruker SolariX (FTMS) ................................................................................................................. 58

2.3.9. Data Analysis ......................................................................................................................... 59

2.3.9.1. Bottom up .......................................................................................................................... 59

2.3.9.2. Top-down .......................................................................................................................... 59

2.4. Results and Discussion ........................................................................................................... 60

2.5. Conclusion .................................................................................................................................. 70

2.6. References .................................................................................................................................. 71

Chapter 3: The application of HPLC/MS analysis with a multi enzyme digest strategy to characterize different interferon products produced from Pichia pastoris 74

3.1. Abstract ....................................................................................................................................... 75

3.2. Introduction ................................................................................................................................... 76

3.2.1. Recombinant interferon α2b ....................................................................................................... 76

3.2.2. Recombinant interferon gamma ................................................................................................. 76

3.2.3. Leader sequence: alpha-factor signal from Pichia pastoris ................................................... 76

3.2.4. High mannose structure of glycosylation of protein expressed in Pichia pastoris . 77

3.3. Materials and Methods ............................................................................................................. 77

3.3.1. Chemical and reagents ............................................................................................................. 78

3.3.2. Sample preparation .................................................................................................................. 78
3.3.2.1. SDS-PAGE of interferon α2b................................................................. 78
3.3.2.2. Isoelectric focusing (IEF) gel of interferon α2b ............................................ 79
3.3.2.3. Trypsin digestion of interferon α2b .......................................................... 79
3.3.2.4. GluC digestion of interferon α2b .............................................................. 80
3.3.2.5. In-gel digestion of IFN gamma gel .......................................................... 81
3.3.2.6. LC-MS/MS ...................................................................................... 81

3.4. Result and Discussion ............................................................................. 82

3.4.1. Part A: Analysis of Interferon (IFN) α2b reference material and Pichia pastoris

IFN α2b 82

3.4.1.1. SDS-PAGE of interferon α2b ................................................................... 82
3.4.1.2. IEF gel of interferon α2b ...................................................................... 83
3.4.1.3. Trypsin digestion and LC/MS analysis (IFN α2b reference material and Pichia pastoris sample) ......................................................................................... 84
  3.4.1.3.1. Determination of amino acid sequence .............................................. 84
  3.4.1.3.2. Analysis of residual leader sequence ............................................... 88
  3.4.1.3.3. Oxidation analysis ........................................................................ 95

3.4.2. Part B: Analysis of Interferon (IFN) γ reference material (Abchem) and Pichia pastoris IFN γ ................................................................. 95

3.4.2.1. In-gel digestion of IFN γ SDS-PAGE gel bands with trypsin: N-glycosylation of Pichia pastoris IFN γ ......................................................................................... 95
3.4.2.2. Oxidation and deamidation .................................................................. 100
3.4.2.3. In-gel digestion of IFN γ SDS-PAGE gel bands with Glu-C: C-terminal proteolysis 101

3.5. Discussion ................................................................................................................................. 102

3.6. Conclusion ................................................................................................................................. 107

3.7. Reference ................................................................................................................................ 108

Chapter 4: Complementary top-down and bottom up approach for the characterization of G-CSF variants produced from Pichia pastoris ..........111

4.1. Abstract .................................................................................................................................... 112

4.2. Introduction ............................................................................................................................... 113

4.2.1. Recombinant Granulocyte-colony stimulating factor ......................................................... 113

4.2.2. Free cysteine in G-CSF induces aggregation at high pH .................................................... 113

4.2.1. O-Glycosylation of G-CSF at Thr 134 .................................................................................. 113

4.2.1. G-CSF Expressed in Pichia pastoris .................................................................................... 114

4.2.2. N-terminal cleavage of G-CSF ............................................................................................ 114

4.3. Experiment Section .................................................................................................................. 115

4.3.1. Reagent and chemicals ......................................................................................................... 115

4.3.2. Preparation of a G-CSF stability sample (high temperature and pH stress) ............... 116

4.3.3. SDS-PAGE gel analysis of G-CSF stability ........................................................................ 116

4.3.4. Pepsin digestion of G-CSF stability sample ......................................................................... 116

4.3.5. Glu-C digestion of G-CSF .................................................................................................... 116
4.3.6. LC-MS analysis of pepsin and Glu-C digested sample ........................................ 117
4.3.7. Intact mass measurement of G-CSF standard and G-CSF produced in Pichia pastoris 118

4.4. Results ....................................................................................................................... 119

4.4.1. Aggregation of reference G-CSF stressed under high pH and Temperature .... 119
4.4.2. Determination of primary structure of G-CSF ....................................................... 122
4.4.3. Oxidation and deamidation analysis .................................................................. 123
4.4.4. Characterization of N-terminal cleavage by Glu-C digestion ....................... 124
4.4.5. Characterization of O-mannosylation at T134 by Glu-C digestion ............... 125
4.4.6. Intact mass measurement to identify the O-mannosylation at T134 and N-terminal proteolysis at between leucine 4 and glycine 5 ......................................................... 126

4.5. Discussion .................................................................................................................. 129

4.5.1. Aggregation of G-CSF stressed under high pH and Temperature ............... 130
4.5.2. Determination of primary structure of G-CSF ....................................................... 130
4.5.3. Oxidation and deamidation analysis ................................................................. 131
4.5.4. Characterization of N-terminal cleavage by Glu-C digestion ....................... 132
4.5.5. Characterization of O-mannosylation at T134 by Glu-C digestion ............... 132
4.5.6. Intact mass measurement to identify the O-mannosylation at T134 and N-terminal proteolysis at between leucine 4 and glycine 5 ......................................................... 133

4.6. Conclusion .................................................................................................................. 133
4.7. References ......................................................................................................................... 134

Chapter 5: Conclusions ............................................................................................................ 138
List of Figures

Figure 1-1: Formation of isoaspartic acid via deamidation of asparagine or isomerization of aspartic acid. ................................................................. 7

Figure 1-2: Schematic depiction of an ESI source operated in positive ion mode ................. 10

Figure 1-3: Cross-section of the C-trap and orbitrap analyzer. ........................................ 18

Figure 1-4: Integration of four-column PCC system with a perfusion cell culture bioreactor to continuously capture the target protein ................................................. 32

Figure 2-1: Sequence map of hGH modifications sites identified by either the bottom-up and top-down method ............................................................... 61

Figure 2-2: SDS-PAGE of hGH variants under non-reduced and reduced conditions .......... 63

Figure 2-3: Deconvoluted spectra of hGH standard and hGH degradation products .......... 66

Figure 2-4: Analysis of methionine oxidation in hGH oxidized sample using top-down method under reduced condition .................................................. 67

Figure 2-5: Cleavage of hGH 2-chain sample determined by the top-down method ......... 68

Figure 3-1: SDS-PAGE of IFN α2b reference material and Pichia pastoris IFN α2b .......... 83

Figure 3-2: IEF gel of IFN α2b reference material and Pichia pastoris IFN α2b .............. 84

Figure 3-3: Confirmation of K112 of standard IFN α2b by GluC digestion ..................... 87

Figure 3-4: Confirmation of amino acid substitution K112 to N 112 of Pichia pastoris IFN α2b by GluC digestion ................................................................. 88

Figure 3-5: Extracted Ion Chromatogram (XIC) of “EEGVSLEK” .............................. 89

Figure 3-6: Monoisotopic mass and CIDMS2 of EEGVSLEK in Pichia pastoris sample ... 91

Figure 3-7: Extracted Ion Chromatogram (XIC) of RCDLPQTHSLGSR ................... 92
Figure 3-8: Monoisotopic mass of leader sequence “RCDLPQTHSLGSR” and “CDLPQTHSLGSR” .................................................................................................................................................. 93

Figure 3-9: CID MS2 of leader sequence peptide “RCDLPQTHSLGSR” in Pichia pastoris sample. .......................................................................................................................................................................................................................................................... 94

Figure 3-10: SDS-PAGE of IFN γ strain supernatant................................................................. 96

Figure 3-11: Glycosylation of Asn98 at 17kDa band. ............................................................... 98

Figure 3-12: Glycosylation of Asn98 at 19kDa band. ............................................................... 99

Figure 3-13: Glycosylation of Asn26 at 19kDa band. ............................................................. 100

Figure 3-14: Identification of C-terminal cleavage at K129 in Pichia pastoris IFN γ sample... 102

Figure 4-1: SDS-PAGE of high temp & pH stressed G-CSF. .................................................. 120

Figure 4-2: Disulfide connectivity confirmation in the G-CSF stressed aggregates. .............. 121

Figure 4-3: Identification of L4 cleaved N-terminal peptide.................................................. 125

Figure 4-4: Identification of O-mannosylation structures observed at T134 LGMAPALQPTQGAMPAPA. .................................................................................................................................................................................. 126

Figure 4-5: Deconvoluted spectra of G-CSF standard............................................................ 127

Figure 4-6: LC chromatogram of Pichia pastoris derived G-CSF ........................................... 128

Figure 4-7: Deconvoluted spectra of Pichia pastoris G-CSF.................................................. 129
List of Tables

Table 1-1: List of molecular variants CQAs for a monoclonal antibody. ................................. 27
Table 1-2: Quality attributes generally observed in biopharmaceutical proteins. ...................... 28
Table 2-1: Comparison of Top Down and Bottom Up characterization of human growth hormone .................................................................................................................................................. 62
Table 2-2: Bottom-up result of oxidation and deamidation analysis of the stability samples..... 63
Table 2-3: Glutamine 141 and Threonine 142 cleavage are determined from a plasmin cleaved reference material using bottom up method........................................................................................................................................ 64
Table 2-4: Peptide sequences detected in the analysis of 2-chain variants from a *Pichia pastoris* derived bioreactor sample using the bottom up method ................................................................. 70
Table 3-1: Peptide mapping result of IFN α2b standard using trypsin in-gel digestion............ 86
Table 3-2: Oxidation analysis of IFN α2b reference material and *Pichia pastoris* material ...... 95
Table 3-3: Peptide mapping of IFN γ reference material with trypsin digestion....................... 97
Table 3-4: Oxidation analysis of IFN γ reference material and *Pichia pastoris* material. ........ 101
Table 3-5: Variants identified for *Pichia pastoris* IFN α2b....................................................... 104
Table 3-6: Variants identified for *Pichia pastoris* IFN γ ............................................................. 105
Table 4-1: Peptide mapping of G-CSF with Glu-C digestion. .................................................... 123
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2AA</td>
<td>2-Aminobenzoic Acid</td>
</tr>
<tr>
<td>ABC</td>
<td>Ammonium Bicarbonate</td>
</tr>
<tr>
<td>ACN</td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>APCI</td>
<td>Atmospheric Pressure Chemical Ionization</td>
</tr>
<tr>
<td>Bio-MOD</td>
<td>Biomedicine on Demand</td>
</tr>
<tr>
<td>CE</td>
<td>Capillary Electrophoresis</td>
</tr>
<tr>
<td>CID</td>
<td>Collision Induced Dissociation</td>
</tr>
<tr>
<td>cIEF</td>
<td>Capillary Isoelectric Focusing</td>
</tr>
<tr>
<td>CQAs</td>
<td>Critical Quality Attributes</td>
</tr>
<tr>
<td>DC</td>
<td>Direct Current</td>
</tr>
<tr>
<td>DDA</td>
<td>Data Dependent Mode</td>
</tr>
<tr>
<td>DIA</td>
<td>Data Independent Mode</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EI</td>
<td>Electron Impact Ionization</td>
</tr>
<tr>
<td>ESI</td>
<td>Electrospray</td>
</tr>
<tr>
<td>ETD</td>
<td>Electron Transfer Dissociation</td>
</tr>
<tr>
<td>EThcD</td>
<td>Electron-Transfer and Higher-Energy Collision Dissociation</td>
</tr>
<tr>
<td>FAB</td>
<td>Fast Atom Bombardment</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug administration</td>
</tr>
<tr>
<td>FTICR</td>
<td>Fourier Transform Ion Cyclotron Resonance</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>G-CSF</td>
<td>Granulocyte Colony-Stimulating Factor</td>
</tr>
<tr>
<td>GHD</td>
<td>Growth Hormone Deficiency</td>
</tr>
<tr>
<td>HCD</td>
<td>High Energy Collision Dissociation</td>
</tr>
<tr>
<td>HCP</td>
<td>Host Cell Proteins</td>
</tr>
<tr>
<td>hGH</td>
<td>Human Growth Hormone</td>
</tr>
<tr>
<td>HILIC</td>
<td>Hydrophilic Interaction Chromatography</td>
</tr>
<tr>
<td>IAM</td>
<td>Iodoacetamide</td>
</tr>
<tr>
<td>IEF</td>
<td>Isoelectric Focusing</td>
</tr>
<tr>
<td>IEX</td>
<td>Ion Exchange Chromatography</td>
</tr>
<tr>
<td>IFN γ</td>
<td>Interferon Gamma</td>
</tr>
<tr>
<td>IFNa2b</td>
<td>Interferon Alpha 2b</td>
</tr>
<tr>
<td>InSCyT</td>
<td>Integrated and Scalable Cyto Technology</td>
</tr>
<tr>
<td>IPG</td>
<td>Immobilized pH Gradient</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid Chromatography-Mass Spectrometry</td>
</tr>
<tr>
<td>MALDI</td>
<td>Matrix Assisted Laser Desorption Ionization</td>
</tr>
<tr>
<td>MAM</td>
<td>Multi-Attribute Method</td>
</tr>
<tr>
<td>MCE</td>
<td>Multi-Compartment Electrolyzers</td>
</tr>
<tr>
<td>NPC</td>
<td>Normal Phase Chromatography</td>
</tr>
<tr>
<td>PCC</td>
<td>Periodic Counter-Current Chromatography</td>
</tr>
<tr>
<td>PD</td>
<td>Photodissociation</td>
</tr>
<tr>
<td>PQA</td>
<td>Product Quality Attributes</td>
</tr>
<tr>
<td>PTMs</td>
<td>Post-Translational Modifications</td>
</tr>
<tr>
<td>QbD</td>
<td>Quality by Design</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive Oxygen Species</td>
</tr>
<tr>
<td>RPC</td>
<td>Reverse Phase Chromatography</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium Dodecyl Sulfate Polyacrylamide gel</td>
</tr>
<tr>
<td>SEC</td>
<td>Size Exclusion Chromatography</td>
</tr>
<tr>
<td>SID</td>
<td>Surface Induced Dissociation</td>
</tr>
<tr>
<td>SIM</td>
<td>Selected Ion Monitoring</td>
</tr>
<tr>
<td>SRM</td>
<td>Selected Reaction Monitoring</td>
</tr>
<tr>
<td>TOF</td>
<td>Time of Flight</td>
</tr>
<tr>
<td>UPLC</td>
<td>Ultra Performance Liquid Chromatography</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction to characterization of biopharmaceuticals using advanced liquid chromatography and mass spectrometry (LC-MS)
1.1. Abstract

Protein Therapeutics, also referred as recombinant biologics, are an important class of today’s medicines. According to FDA guidance, biosimilars are highly similar to pharmaceutical reference material that has already been authorized for marketing as a biological drug. Biosimilars are developed with the purpose of reducing cost and thus increasing patient access. As the patent coverage of some of the best-selling biological medicines will expire over the next few years, there is considerable interest for companies to produce biosimilars. Thus, development of analytical tools to ensure the similarity of biosimilars to reference material is in high demand recently.

Comprehensive and extensive characterization of protein drugs is of great significance of ensuring the product quality, safety and efficacy. LC-MS Liquid chromatography coupled with mass spectrometry (LC-MS) for is an important tool for the characterization of protein therapeutics and biosimilars. Impressive advancements in LC-Ms resolution and throughput have been made it possible to provide more efficient and thorough characterization of protein therapeutics and biosimilars.

Top-down and bottom-up are two common approaches used complementarily for the extensive characterization of protein therapeutics. The bottom-up approach involves proteolytic digestion of proteins to peptides and the following analysis of the peptides. It is more robust and reproducible, however, the sample preparation is time consuming and may cause artifacts interfering the results. The top-down approach, on the other hand, avoids tedious sample preparation and enables the analysis of all the proteoforms together. Nevertheless, high resolution and sensitivity of a powerful LC-MS system is required for accurate intact mass measurement, which is a limitation sometimes. Thus, bottom-up and top-down are often used together for the in-depth analysis of protein therapeutics.
Quality by Design (QbD) proposes to build the quality of the product into the process by design rather than testing and has brought revolutionary changes in the field of biomanufacturing. The first and most important step for QbD is to determine the Critical Quality Attributes (CQAs) for the protein therapeutics. Ensuring the biosimilars to have same level of CQAs with the reference drug is critical for biosimilar development.

The focus of this thesis is to develop advanced analytical methods for the characterization of biosimilars produced by a Biomedicine on Demand (Bio-MOD) platform. Bio-MOD aims to design and assemble a portable and integrated platform to enable the delivery of drugs to patient whenever and wherever needed especially battlefield threat and medicine needs. *Pichia Pastoris* was chosen as the host strain and alpha pre-pro leader signal peptide was used for protein secretion. This platform aims to enable small scale and high throughput of biosimilar production with high purity, safety and efficacy. High-mannose glycosylation structure and residual leader sequence were unique features for biosimilars produced from *Pichia Pastoris*. 
1.2. **Protein Therapeutics and Biosimilars**

Protein therapeutics first emerged in the early 1980s. The first protein therapeutic approved by FDA Food and Drug Administration (FDA) was the recombinant version of human insulin in 1982. Although a few approved protein drugs were extracted in their natural forms, most of them were produced by recombinant DNA technology. Seven of the top ten best selling drugs of 2015 were protein drugs. Protein therapeutics are becoming the major class of drugs in the current drug market.

Protein therapeutics are advantageous in their specific and biological functions when compared to small molecules. First, as the functional agents in an organism, proteins exhibit higher specificity and complexity than small molecules. Additionally, biopharmaceutical companies could obtain patent protection due the uniqueness of proteins, this could lead to considerable financial benefits. Herein, three catalogs of protein therapeutics are classified by functions: protein therapeutics with enzymatic or regulatory activities with specific targeting activities and protein vaccines.

A new class of recombinant biologics referred to as biosimilars, is just starting to enter the US Market. Biosimilars, according to FDA guidance, are highly similar to a biological reference material that has already been authorized for marketing as a biological drug. Biosimilars are developed with the purpose of reducing cost and increasing access of the treatment. As the patent coverage of some of the best-selling biological medicines will have expired in a few years, there is considerable interest for companies to produce biosimilars. The first biosimilar in US, Zarxio (Sandoz, subsidiary of Novartis), which is the generic form of Amegen’s Neupogen, was approved in 2015.

1.3. **Post-translational modification of proteins**
Post-translational modifications (PTMs) are the main reason for proteome complexity and diversity. Protein properties are changed by these covalent processing events by addition of functional groups, degradation of entire proteins, and proteolytic cleavage of regulatory subunits. The modification of the primary structure by PTMs can have a positive or negative effect on the structure and function of a protein therapeutic.\(^4\) The various PTMs almost have influence on all aspects of protein function and cell biology.\(^4\)-\(^7\) Consequently, characterization of PTMs is critical to understanding the disease mechanism, prevention and treatment.\(^8\) However, the diversity of PTMs enlarges the difficulty of protein characterization, and application of more accurate and efficient detection methods is of growing interest and demand. Several common and important PTMs were discussed below.

1.3.1. **Oxidation**

Proteins could be oxidized easily by reactive oxygen species (ROS) in biological systems, because of the high abundance and rate constants. Oxidations usually happen on both side chains of amino acids, as well as the backbones of proteins, resulting in hydrophilicity increase, side-chain backbone fragmentations, protein aggregations, and even conformation changes. Biological consequences of oxidation include aging and disease.\(^9\)-\(^10\) When it comes to protein therapeutics, oxidation is hard to avoid, as it happens at almost every step of manufacturing, including production, purification, formulation, transportation and storage. Comprehensive characterization of oxidation is of great significance, as it may change protein properties and increase clinical risks, such as aggregations, loss of biological activities, and immune response.\(^11\) Methionine is often the target for oxidation, and other amino acids such as tryptophan, histidine and cysteine could also be oxidized. The oxidation characterization of several protein therapeutics are studied extensively,
examples include but not limited to monoclonal antibodies\textsuperscript{12-13} human growth hormone,\textsuperscript{14-15} insulin,\textsuperscript{16} interferon alpha,\textsuperscript{17-18} interferon beta and granulocyte colony stimulating factor.\textsuperscript{19-20}

1.3.2. **Deamidation**

Deamidation is another common PTM of protein therapeutics, occurring as asparagine (Asn) or glutamine (Gln). Deamidation of glutamine is hundred times slower than asparagine deamidation, therefore, asparagine is usually the site for the characterization of deamidation. Deamidation of asparagine is a spontaneous, nonenzymatic reaction resulting in the formation of a five-ring succinimide intermediates, and then aspartic acid or isoaspartic acid are formed via the hydrolysis of succinimide as shown in Figure 1-1.\textsuperscript{21-23} Deamidation could happen on almost every Asp and Gln except “NP” and “DP” sequence, while “NS”, “NG” and “DG” were reported to be the “hot spot” for deamidation.\textsuperscript{21} Deamidation could be accelerated by many factors such as alkaline pH and high temperature. Deamidation contributes to the charge heterogeneity and lead to the formation of acidic variants in therapeutic proteins. A number of protein therapeutics are prone to deamidation.\textsuperscript{24-28} One of deamidation products, isoaspartic acid, may cause the loss of activity or protein aggregation in protein therapeutics, more over isoaspartic acid plays a critical role in many biological malfunction such as neurodegeneration, autoimmune and cancer.\textsuperscript{29-30}
Figure 1-1: Formation of isoaspartic acid via deamidation of asparagine or isomerization of aspartic acid.


1.3.3. Glycosylation

Glycosylation can be classified into different categories by the types of sugar-amino acid linkages as N-link glycosylation, O-link glycosylation, C-mannosylation, phosphoglycation and glypiation. Glycosyltransferases and glycosidases are the key enzymes that are responsible for biosynthesis of glycans. Cellular glycans mostly attach to proteins and lipids and play essential roles in the biological functions, such as, cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction and endocytosis. As reported, two-third of the protein drugs in the market are glycoproteins with N-linked the forms most dominant. Glycosylation has influence on the activity, safety and efficacy of the protein therapeutics, thus profiling various
forms of glycosylation and accurate quantification is crucial for the characterization of protein therapeutics.  

1.4. **LC-MS for Characterization of Protein Therapeutics**

1.4.1. **High-performance liquid chromatography (HPLC)**

HPLC or High-performance liquid chromatography is used to separate, identify and quantify the compounds in an analyte by column chromatography. It was revolutionized by the introduction of ultra-performance liquid chromatography (UPLC) with dramatic increases in resolution, selectivity and sensitivity by the advanced in instrumentation and column technology. HPLC can be classified by the phase system as Normal Phase Chromatography (NPC), Reversed Phase chromatography (RPLC), Size Exclusion Chromatography (SEC), Ion Exchange Chromatography (IEX) and Bio-Affinity chromatography. All of which have broad application in the biological and pharmaceutical field. For example, RPLC is routinely used for peptide mapping, C4 material is used for intact protein analysis and C18 is usually used for peptide mapping analysis. NPC is applied to glycan analysis, SEC is widely used for molecular weight determination and aggregation identification, IEX is useful for charge heterogeneity and PTMs analysis. Bio-affinity is an important purification and enrichment tool, for example the first step of antibody purification is done by protein-A affinity chromatography.

1.4.2. **Mass Spectrometry**

Mass Spectrometry (MS) is one of the most highly utilized analytical techniques in pharmaceutical development and research. Because of the sensitivity and selectivity, it has been extensively applied for the characterization of protein therapeutics. While well-developed platforms are well
used by pharmaceutical companies benefiting from its robustness and reproducibility, improvement in both ionization methods and mass analyzer have also been pushing its limits. All mass spectrometers contain three main functions ion formation, mass analysis and ion detection. With the improvement in both ionization method and mass analyzer, LC-MS becomes a powerful technique to characterize protein therapeutics with high sensitivity, high selectivity and high throughput. The introduction and application for the ionization methods, fragmentation methods and mass analyzers will be discussed in the following sections.

1.4.2.1. Ionization method

There are two classes of ionization method based on the change of molecules during the ionization process, which are, hard and soft ionization. Hard ionization such as Electron Impact Ionization (EI), involves the fragmentation of the molecule with m/z lower than the intact molecule and it was replaced by APCI (Atmospheric Pressure Chemical Ionization).

Soft ionization, such as Fast Atom Bombardment (FAB), Electrospray (ESI), APCI and Matrix-Assisted Laser Desorption Ionization (MALDI) induces very few fragmentations of the molecular ion to avoid ambiguous identification caused by the ionization process. ESI and MALDI are the most commonly used ionization methods currently.

1.4.2.1.1. Electrospray (ESI)

ESI produces ions by spraying charged droplets into a strong electrical field. It involves three steps: formation of charged droplets, progressive solvent evaporation to shrink the droplet and finally transfer the ions to gas phase as shown in Figure 1-2. ESI generally produces intact charged ions in either positive-ion mode or negative-ion mode. One important advantage is that it enables
analysis of large molecules by generating multiple charged ions. Nano ESI (nESI) is the most popular derivative technique of ESI, as it greatly increased the assay sensitivity due to a very low flow rate.\textsuperscript{39-40} DESI is another variation that analyze the sample on a surface while acquiring a similar spectra to ESI.\textsuperscript{41}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{esi_source.png}
\caption{Schematic depiction of an ESI source operated in positive ion mode.\textsuperscript{40}}
\end{figure}


1.4.2.1.2. Matrix-Assisted Laser Desorption Ionization (MALDI)

Matrix-assisted laser desorption ionization is another commonly used ionization technique. MALDI requires mixture of the sample with a specific matrix that absorbs the laser radiation and transfers a proton to the sample, then the sample is bombarded with a laser and ionized. MALDI
mostly generates singularly (sometimes doubly) charged ions and it is often always coupled with time-of-flight (TOF) mass analyzer because of its large dynamic range. MALDI imaging mass spectrometry is one of the major applications of MALDI these days, which enables the mass measurement of targeted biomolecules in biological tissues.42-43

1.4.2.2. **Fragmentation method**

After ionization, the charged ions are subjected to certain fragmentation methods to generate different types of fragment ions for identification.

**1.4.2.2.1. Collision Induced Ionization (CID) and High Energy Collision dissociation (HCD)**

In CID, a precursor ion collides with a neutral gas (nitrogen, helium or argon), and part of the kinetic energy can be converted to internal energy, leading to the subsequent decomposition of the precursor ions. Even though CID has limitations with large peptide and proteins due to the low efficiency of energy redistribution, it is widely applied to peptide sequencing. Typical CID fragmentation generates b-type and y-type ions. High-energy collision dissociation (HCD) is a specific type of CID technique in an Orbitrap MS with higher collision energy and shorter activation time. Similarly, HCD generates b- and y- type ions, with y-ions most abundant. Additionally, the high energy induces further fragmentation from b-ions to a-ions. Moreover, HCD can also induce dissociation of the amino acid side chains forming d-, w- and v-type ions which are useful for the identification and quantification of peptides and proteins with isobaric labeling.44 While CID has been broadly implemented in protein characterization, it has limitation in the analysis of PTMs. Labile PTMs, such as glycosylation are usually disassociated under CID as it
linked by a weaker bond than the peptide bond and lead to no peptide backbone sequence information.\textsuperscript{45-49}

1.4.2.2.2. Electron Transfer Dissociation (ETD)

ETD fragmentation involves the transfer of an electron from a radical anion to a protonated peptide, resulting c- and z- type ions. Peptides with higher charge state tend to be better fragmented under ETD than CID. Charge reduced species are often the high abundance ions of ETD, from which CID can be further applied to get more fragmentation data. Different from CID, typical ETD cleaves the alpha carbon bond of amino acid, therefore, preserves the side chain information. Thus one important applications of ETD on protein therapeutics analysis is the preservation of labile PTMs, for example backbone information can be obtained for a glycopeptide without losing glycosylation during fragmentation process.\textsuperscript{50} ETD is also a useful tool for the analysis of disulfide connectivity, which breaks the disulfide into two peptides ions.\textsuperscript{51} Additionally, ETD is widely applied to the analysis of isoaspartic acid, where ETD can generate diagnostic ions for verification of isoaspartic acid.

1.4.2.2.3. EThcD

EThcD, as its name indicates, is a fragmentation method combing ETD and HCD. It was commercially implemented in a Thermo Orbitrap Fusion platform. EThcD involves sequential events, in which ETD is firstly applied to fragment the precursor ions, followed by HCD for the sequential fragmentation of all the product ions formed in the ETD event. Thus, all the product ions, and unreacted charge reduced species will be further fragmented resulting a mixture of b/y
and c/z fragmentations in the EThcD MS/MS spectrum. This overcomes the limitation of limited sequence coverage in CID as CID only breaks the weakest bonds and the requirement of a high charge state for ETD.\textsuperscript{52} The idea of combined ETD and CID firstly came from an experiment performing CID of a charge reduced species in ETD in a MS3 event.\textsuperscript{53} As described, EThcD combined not only the two fragment cells, ETD and HCD, but also the advantages of the two fragmentation methods to fulfill full sequence coverage and characterization of PTMs with site information. With the accessibility of EThcD, complex MS method can be simplified while reducing the data acquisition time and obtaining more sequence information. EThcD has been widely applied in the characterization of primary protein sequences and labile PTMs using top-down approach.\textsuperscript{54}

1.4.2.2.4. Surface Induced Dissociation (SID)

Surface-induced dissociation (SID) is another dissociation method, where the collision is induced on a surface of voltage, instead of neutral gas. SID was demonstrated to be a promising alternative dissociation method in getting more details for protein complexes, especially for non-covalent protein complexes, where CID produced undesired unfolding of protein subunits. In recent years, Wysocki group has been pioneering in analyzing structures of protein complexes using SID.\textsuperscript{55-56}

1.4.2.2.5. Photodissociation

Photodissociation (PD) generally involves the irradiation of ions with light and consecutive internal decomposition. PD has been an outstanding alternative dissociation method in several perspectives. Firstly, it enables high-throughput analysis with ultrafast energy deposition. Secondly, it only needs a simple trap mass spectrometer to trap the ions for light radiation. Thirdly,
ion-scattering effects can be mitigated using photons, instead of collisions. Brodbelt group has propelled the usage of PD in several fields including protein-protein interaction and intact protein sequencing using both a top down and bottom up strategy. UVPD was a newly commercialized Photodissociation method in the Orbitrap Fusion Lumos Tribrivd Mass Spectrometer, showing higher selectivity and specificity of dissociation patterns via 213 nm UV laser, which provides 1000000 FWHM ultra-high resolution for improved structural elucidation.

1.4.2.3. Mass Analyzers

Today mass analyzer uses are divided into two groups: beam and trapping analyzers.

1.4.2.3.1. Magnetic Sector

From 1950s to 1980s, mass spectrometers with a sector analyzer were the predominant types, combining both magnetic sector and an electric sector. In the default mode, the electric sector was kept at constant value allowing only ions with a specific kinetic energy pass through, while the magnetic field was scanning ions exponentially or linearly to obtain corresponding mass spectra. Sector mass spectrometers are not suitable for MALDI and had limited MS/MS application.

1.4.2.3.2. Quadrupole Mass Spectrometers

The quadrupole mass analyzer is commonly used in mass spectrometers, behaving as a “mass filter”. Combined direct current (DC) and radio frequency (RF) potentials are applied to quadrupole rods to allow a selected m/z to ion pass for detection. All other ions without the defined m/z will not have a stable trajectory through and collide with the rods, never reaching the detector.

1.4.2.3.3. Time of Flight (TOF)
A TOF mass spectrometer is conceptually the simplest type of mass spectrometer which can be considered as a race from the ionization (starting point) to the detector. Theoretically, the ions are accelerated with a fixed potential into a TOF drift tube with a fixed distance. The ions with different m/z values will abstain different velocities and then arrive to the detector at different time. In this case, by measuring the velocity of different ions, their corresponding m/z can be determined and reflected in the mass spectra data. A reflectron was then developed to compensate for the velocity difference of ions with same m/z but different sizes and thus increasing the resolution of TOF. TOF has the advantages of high resolution, high speed, and high mass range and is often coupled with pulsed ionization method such as MALDI. However, the poor sensitivity and selectivity of precursor ion for MS/MS fragmentations is often the limitation of TOF. In order to minimize the limitation, the combination of quadrupole and TOF (QTOF) has been developed for high resolution accurate mass spectrometry, for example Water Xevo QTOF and Sciex X500B QTOF.

1.4.2.3.4. Quadrupole Ion Trap

Quadrupole ion trap is a type of a dynamic trap, along with ion cyclotron resonance (static traps) are both operated by storing ions in the trap and manipulating the ions by DC and RF electric fields. Some unique capabilities for ion traps include extended MS/MS analysis, high resolution and sensitivity. On the other hand, the tradeoff is that a long period of analysis time may cause unwanted ion-ion interactions which spontaneously fall apart.

The ion trap has electric field in all three dimensions, where x and y are in quadrupole and the ions move perpendicular to the field in the z direction. In this way, ions can be trapped in the field. Ion traps have the m/z range and accuracy similar to quadrupole, however, resolution can be
improved. The ease of use, throughput, low cost are the main reasons that ion traps are mostly used in the lab. Space charge effects, also referred as ion-ion repulsion, has been the concern for ion traps. What’s more, the ions in the trap can react with the neutral species resulting in changes in the mass spectrum. Nowadays, the ion trap is always incorporated in tandem MS to increase the performance, such as Thermo LTQ series.\textsuperscript{61-63}

\subsection{Fourier Transform (Fourier Transform Ion Cyclotron Resonance)}

In a FTICR mass analyzer, ions have the kinetic energy of a few tens of electron volts (eV). With such low kinetic energies, ions are trapped in the magnetic field instead of a pass through. By calculating the cyclotron frequency, the m/z can be determined. To increase the resolution, the ions are excited by an RF pulse to a larger ICR radius. The packet of ions passes the receiver plates and induces image current and the fourier transform converts the image currents to the mass spectrum. The measurement of different masses is achieved by using a frequency sweep that contains all the cyclotron frequencies. FTICR is best known for its high resolution and accuracy. However, the instrument costs more than US $ 1 million and requires much more lab space and superconducting magnetic.\textsuperscript{64-65}

\subsection{Orbitrap}

The shortcomings of current mass analyzers, for example, high cost and size of FTICR, low accuracy of ion trap have driven the development of the orbitrap analyzer. The term “orbitrap analyzer” was first brought up by Alexander Makarov at the 1990 ASMS conference where proof-of-principle results were presented. The first commercial release of orbitrap was in 2005 as a
hybrid mass analyzer LTQ Orbitrap. It inherited some features from other analyzers, the principle of current detection from FTICR and pulsed injection from TOF.\textsuperscript{66-67}

The orbitrap mass analyzer consists of an outer barrel like electrode and a central spindle like electrode, both are connected to independent voltage supplies. The working principle of the orbitrap is based on the orbital trapping of ions. The injected ions cycle around the central electrode and oscillate along the horizontal axis at the same time. Mass analysis in an orbitrap can be carried in two modes: Fourier Transform (FT) mode and Mass Selective Instability (MSI) mode. The principle analysis method of the orbitrap is the FT mode as it allows the highest mass resolution. Fourier Transform mode measures coherent oscillations in the axial direction. As the ions axially oscillate back and forth, the outer electrode measures their image current. The image current is then converted to a mass spectrum by Fourier Transform. There are some instances in which the Mass Selective Instability (MSI) mode should be used, for example, high intensity signals from unwanted compounds can be ejected to improve dynamic range.\textsuperscript{68}

1.4.2.3.7. Tandem mass

Tandem Mass Spectrometry usually referred to as MS/MS, involves the use of two or more mass analyzers for example, the Orbitrap XL consists of a linear ion trap and an orbitrap. Orbitrap XL was most used for the experiment in this thesis.
Figure 1-3: Cross-section of the C-trap and orbitrap analyzer.


Orbitrap XL is a hybrid mass spectrometer that enables accurate peptide identification and quantitation. It incorporates the ruggedness and sensitivity of a linear ion trap and high mass accuracy and high resolution of an orbitrap. It can utilize multiple ionization methods including ESI APCI and APPI. After the analyte is ionized, the linear ion trap can store, filter and fragment the ions to be transferred to orbitrap for analysis. The orbitrap has a spindle shaped center electrode and a pair of bell shaped outer electrodes. Ions go through C trap and an octupole collision cell before being transferred to the orbitrap as shown in Figure 1-3. The C trap initially served with the function of both transferring ions and for collision. Then an additional octupole collision cell was
added next to C trap to avoid making the compromise of losing high mass ions or to change low mass cut-off.\textsuperscript{69}

Thermo Fisher scientific also launched a series of high resolution mass spectrometers after Orbitrap XL including Q Exactive and Orbitrap Fusion. Both of which have similar but improved configurations to push the limit of resolution and mass accuracy. For example, parallel measurement of MS and MS/MS has been achieved and top speed instead of top N mode has also been implemented.\textsuperscript{70-71}

1.4.2.4. Data Acquisition Method

1.4.2.4.1. Data Dependent Mode (DDA)

Data Dependent Mode (DDA), DDA is a robust and traditional data acquisition method in mass spectrometry. Following a full scan MS, fragmentation will be applied on the precursors ions with top N highest intensity. The product ions are searched against a database for identification. Although powerful in high abundance peptide identification, DDA has constrains in low abundant species.

1.4.2.4.2. Selected Ion Monitoring (SIM) and Selected Reaction Monitoring (SRM)

Selected Ion Monitoring (SIM) is used as a complementary method with DDA for the low abundance peptide identification, in which a precursor ion with specific m/z values is selected for fragmentation regardless of its intensity in the full scan. Selective Reaction Monitoring (SRM) is analogous to SIM, except that it monitors both a selected precursor ion and a product ion.

1.4.2.4.3. Data Independent Mode (DIA)
Data Independent Mode (DIA) conceptually combines the advantages of DDA and SIM, in which all the precursor ions within a defined m/z window in one full scan are subjected to fragmentation until the full mass range is reached. Although complex, an accurate set of data for complete profiling of the peptides are collected without losing the low abundance peptides information in as in DDA. The data complexity and requirement for a high-resolution mass spectrometer are the major drawback of DIA, notwithstanding, DIA is expected to live up to its potential with an increasing number of practical implementations of DIA.²²-⁷⁴

1.4.2.5. **Data Interpretation**

1.4.2.5.1. **Database search Software**

PepFinder is a powerful software designed specifically for protein therapeutic characterization. Peptide mapping can be processed with high confidence and efficiency. Novel features like a MS2 prediction algorithm, PTM identification and quantification, disulfide bridge mapping, sequence variant identification are included. Later 2016, PepFinder is incorporated in the software Biopharmfinder which also has the function of characterization of intact molecular weights of biotherapeutics.

1.4.2.5.2. **Manual Inspection**

Although database search is indispensable and helpful for peptide identification and modification quantification, the limitation of false positive identification cannot always be avoided. Manual inspection is necessary as a complementary method. Firstly, theoretical mass of each fragment ion can be predicted by Fragment Ion Calculator, secondly ions in a MS/MS spectrum can be matched with the theoretical fragment masses for the identification.
1.4.3. Sample Preparation

1.4.3.1. Bottom-Up

Bottom-up is a robust way to characterize a protein’s primary sequence and involves the proteolytic digestion of a given protein to a mixture of peptides. Multiple enzymes can be used to improve the robustness of the analysis, providing 100% sequence coverage and allowing an in-depth analysis of PTMs and degradation products. However, drawbacks include tedious sample preparation, artifacts induced during the sample manipulation, and the difficulty in obtaining 100% for some protein sequences.

1.4.3.1.1. In-gel digestion

Polyacrylamide gel based digestion offers several advances over solution based digestion. Firstly, SDS efficiently removes mass spec interfering impurities from a protein sample which might be detrimental to mass spectrometric sequencing. Secondly, gel fractions reduce complexity of protein sample. For example, heavy and light chains, Fab and Fc regions of monoclonal antibodies can be separated on gel before digestion for the purpose of analyzing a specific part of the antibodies. On the other hand, the drawbacks of gel based digestion include: extensive peptide loss decreases efficiency of the subsequent mass spec analysis and it can be laborious and tedious procedures.

1.4.3.1.2. In solution digestion

Other than gel based digestion, in-solution digestion requires less time for sample preparation and less artifacts such as oxidation and deamidation is induced with in-solution digestion as it does not require a high temperature incubation and avoids the high voltage of gel electrophoresis step.
Another key advantage is the use of this approach for complete sequence mapping of therapeutic proteins. Different from proteomics sample preparation which uses only trypsin digestion in most cases, therapeutic proteins digestion always requires sequential digestion or cocktail enzyme digestion to achieve 100% sequence coverage. With in-solution digestion, digested peptides can be collected after each step of the sequential enzyme digestion for analysis which is not feasible for in-gel digestion.51

There are several disadvantages for in-solution digestion as well. Firstly, it always involves detergents such as urea or SDS to solubilize and unfold the protein to increase digestion efficiency. However, these detergents may not be compatible with mass spectrometry and can depress the ionization of the peptides. Thus, a dialysis or buffer exchange step is a requirement but may cause sample loss.79 Moreover, for process development samples such as a strain or fermentation samples, in solution digestion is not suitable as the proteins are mixed together with HCPs in the media which will suppress the signal of the drug protein.

1.4.3.1.3. Novel sample preparation methods

Recent years, there have been monumental advances in mass spectrometry instrumentations, sample preparation has enjoyed the same rate of development as well. Some technologies currently employed for biopharmaceutical sample preparation are suffering from high levels of irreproducibility, poor sensitivity and high levels of time-consuming manual work. These issues associated with limitations of manual sample preparation stimulates the development of automated sample preparation platforms including automated SPE (solid phase extraction), automated LLE (liquid-liquid extraction) for protein purification and automated protein digestion platform.80-82 Advancements in nanoscale sample handling and on-probe, chip, membrane digestion is being
investigated as well. For example, Gyrolab MALDI SP1 workstation by GyroAb (Uppsala, Sweden) has been developed to enable simultaneously digestion of 96 protein samples.\(^8^3\)

1.4.3.1.4. Enzymes

Trypsin is the predominant enzyme for bottom up analysis owing to its high specificity, wide availability and easy usage. Trypsin cleaves the carboxylic side of lysine and arginine residues except when these two amino acids are next to proline. The C-terminal end of the tryptic peptide always has a positive charge which makes it easy to be ionized. Biotechnology companies have made various versions of commercial trypsin serving different purposes. For example, immobilized trypsin significantly reduces the digestion time and expands the protein concentrations,\(^8^4\) sequence modified trypsin can avoid autolysis of trypsin, trypsin/LysC (cleaves at lysines residues) mixtures can increase the digestion efficiency and reproducibility. In a typical trypsin digest, 10-30% cleavages sites are missed due to the trypsin proteolytic inefficiency at lysine sites which can be compensated by the action of Lys-C in the trypsin/Lys-C mix.\(^8^5\)\(^-^8^6\) Recently, some trypsin enhancers are also available for better performance.\(^8^7\) However, trypsin alone sometimes is not enough for complete sequence mapping and comprehensive PTM identification and quantitation. It may generate peptides too small or too large for mass spectrometry analysis, and ones not ideal for PTMs analysis. Alternative enzymes have been explored as a complementary solution, either alone or combined with trypsin. These enzymes do not always have the same optimal digestion condition as trypsin. To achieve the best digestion efficiency, experimental condition for each of the enzyme was optimized and published.\(^8^8\)
There are also a group of enzymes for a specific PTM such as glycosylation. Various enzymes are available now for the cleaving or adding different glycan structures, which are quite helpful for glycan analysis.

In regard to monoclonal antibodies, FabRICATOR (IdeS) has been developed to cleaves the hinge region of IgG and generate F(ab’)2 and two Fc regions of antibodies. It has higher specificity compared to the alternative enzyme papain.

1.4.3.2. Top-Down

The top-down is used to analyze the intact protein giving a global perspective of all the proteoforms for a given sample, including labile PTMs. One key advantage of top-down is the ability to perform such an analysis without the requirement of prior knowledge. Major disadvantages of the top-down method include separation and analysis of unstable proteins, limitation of the MW range of the technique, complexity of PTMs such as glycosylation and the challenge of analysis of a range of closely related proteoforms.

1.4.4. Critical Quality Attributes (CQAs)

The concept of Quality by Design (QbD) was first brought up as a revolutionary step for the biomanufacturing industry. QbD proposes to build the quality of the product into the process by design rather than test for it. Biological drugs has a number of features called attributes, those are known, or potentially, to be critical to the safety, efficacy and pharmacokinetics of the drug and are referred as critical quality attributes (CQAs) as in Table 1-1 and Table 1-2. CQAs are determined by the DNA sequence, cell line and manufacturing process. The first step of QbD is to define the CQAs of a product. One important example is an understanding of the CQAs facilities

24
the development of high-quality biosimilars, although not identical to original drug, a biosimilar has the same level of CQAs and thus same biological functions. CQAs are determined by a risk assessment as per the ICH guidance Q9, and evaluates the attributes’ impact on biological activity, pharmacokinetics/pharmacodynamics (PK/PD), immunogenicity and overall safety/toxicity. Prior product knowledge is key in making the risk assessments which provides a rationale for the relating CQAs. The result of risk assessment would be a list of CQAs ranking in order of importance. Generally observed CQAs in protein therapeutics are categorized as product-related impurities, process related impurities and contaminants. The focus of this thesis is product-related impurities.

1.4.4.1. Peptide mapping

Peptide mapping is a robust and reproducible workflow for elucidating the primary amino acid structure of therapeutic proteins. Peptide mapping first involves proteolytic digestion of proteins into peptides. The following separation and identification steps provide insights into the proteins’ amino acid sequence and modifications such as site-specific glycosylation, amino acid substitution and truncations.

For the characterization of biosimilars, peptide mapping is usually performed in a comparative manner due to its inherent variability. In a side-by-side experiment, biosimilars are compared with a reference product (innovator).

1.4.4.2. Correct amino acid sequence

The correct amino acid sequence ensures the primary structure of the protein. Sequence variant generally means any unintentional amino acid substitutions, omissions or insertions during protein
biosynthesis. Three main causes are: mutation at the DNA level, misincorporation at the protein level due to mistranslation or improper tRNA acylation and miscleavage during the post-translational modification. Sequence variants contributes to the heterogeneity of protein therapeutics and may potentially affect the immunogenicity and potency of the drug.²⁹⁴

Peptide mapping with liquid chromatography-mass spectrometry is currently employed as an advanced and sensitive tool to detect and characterize sequence variant at low level. Pioneer work has been done by Yate’s lab employing the concept of single-nucleotide polymorphisms and an updated version of the SEQUEST program. Later, this concept was adopted in to the error-tolerant search (ETS) program in MASCOT. In this approach, a substitution matrix is generated by translating the amino acid to the corresponding codon, after making all possible single base substitutions and re-translating to amino acid sequence. Each unassigned tandem MS spectrum is matched with the “new” amino acid sequence to identify the variant. In addition to Mascot-ETD, several softwares have been developed based on this workflow such like SIEVE and MassAnalyzer to identify low-level sequence variants.⁹⁵-⁹⁹ One significant advantages of the mass spectrometric based approaches is their ability to detect non-genetic mutation derived sequence variants caused by tRNA mischarging or codon mischarging or codon misreading. Notwithstanding, if ionization efficiency of the peptide is poor, this approach might miss a certain amino acid variation.⁹⁸

Another approach complementary to MS based method is transcriptome sequencing (RNAseq) technology. RNAseq enables detection of the mutation at each stage of cell line development (CLD) as early assurance of the genetic homogeneity to avoid the tedious and even infeasible separation of the variant during downstream purification of the proteins. The LC-MS approach can be used to confirm the mutation detected in RNAseq and further assist in the cell line screening step. Any
genetic mutation detected by RNAseq provides the information of potential mass shift for mass spectrometric analysis.\textsuperscript{98}

<table>
<thead>
<tr>
<th>Category</th>
<th>Quality attribute\textsuperscript{a}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size-related Variants</td>
<td>High Molecular Weight Species (HMWS)</td>
</tr>
<tr>
<td></td>
<td>Low Molecular Weight Species (LMWWS)</td>
</tr>
<tr>
<td>Charge-related Variants (Acidic)</td>
<td>Deamidation in CDR</td>
</tr>
<tr>
<td></td>
<td>Deamidation in Non-CDR</td>
</tr>
<tr>
<td></td>
<td>Glycation in CDR</td>
</tr>
<tr>
<td></td>
<td>Glycation in Non-CDR</td>
</tr>
<tr>
<td>Charge-related Variants (Basic)</td>
<td>Aspartic Acid Isomerization in CDR</td>
</tr>
<tr>
<td></td>
<td>Aspartic Acid Isomerization in Non-CDR</td>
</tr>
<tr>
<td></td>
<td>N-Terminal Leader Sequence (may be molecule specific)</td>
</tr>
<tr>
<td></td>
<td>N-Terminal Pyroglutamic Acid</td>
</tr>
<tr>
<td></td>
<td>C-Terminal Lysine</td>
</tr>
<tr>
<td></td>
<td>C-Terminal Proline (IgG1) or Leu (IgG4) Amidation</td>
</tr>
<tr>
<td>Oxidation-related Variants</td>
<td>Oxidation in CDR (Met, Trp)</td>
</tr>
<tr>
<td></td>
<td>Oxidation in Non-CDR (Met, homo-variant)</td>
</tr>
<tr>
<td></td>
<td>Oxidation in Non-CDR (Met, hetero-variant)</td>
</tr>
<tr>
<td>Fc Glycosylation</td>
<td>Afucosylation</td>
</tr>
<tr>
<td></td>
<td>Galactosylation</td>
</tr>
<tr>
<td></td>
<td>High-Mannose</td>
</tr>
<tr>
<td></td>
<td>Sialylation (NANA, NGNA)</td>
</tr>
<tr>
<td></td>
<td>Non-Glycosylated Heavy Chain</td>
</tr>
<tr>
<td>Structural Variants</td>
<td>Cysteine Forms</td>
</tr>
<tr>
<td></td>
<td>Sequence Variants</td>
</tr>
<tr>
<td></td>
<td>Protein Structure</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Certain low abundance variants may need to be added to the list of general known variants such as advanced glycation end-products, hydroxlysine, or oxidative carbonylation.

Table 1-1: List of molecular variants CQAs for a monoclonal antibody. \textsuperscript{100}

<table>
<thead>
<tr>
<th>Product-Related Impurities and Substances</th>
<th>Process-Related Impurities</th>
<th>Contaminants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregation</td>
<td>Residual DNA</td>
<td>Adventitious agents</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(bacteria, mycoplasma, fungi, and viruses)</td>
</tr>
<tr>
<td>Fragmentation</td>
<td>Residual host cell proteins</td>
<td>Endotoxins</td>
</tr>
<tr>
<td>C- and N-terminal modifications</td>
<td>Raw material-derived impurities</td>
<td></td>
</tr>
<tr>
<td>Oxidation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deamidation/isomerization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycosylation (N-linked)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site occupancy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galactosylation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sialylation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fucosylation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oligomannose forms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bisecting GlcNAc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycosylation (O-linked)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disulfide bond modifications/free thiols</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GlcNAc, N-acetylglucosamine.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1-2: Quality attributes generally observed in biopharmaceutical proteins.\(^{93}\)

Reprinted with permission from (Eon-Duval, A.; Broly, H.; Gleixner, R., Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach. Biotechnology progress 2012, 28 (3), 608-622). Copyright 2012 American Institute of Chemical Engineers (AIChE).
1.4.4.3. **Characterization of critical quality attributes (CQAs)**

The traditional strategy for the characterization of proteins includes multiple assays. Capillary electrophoresis (CE) and ion exchange chromatography are generally used to monitor clipping or deamidation of proteins. Nevertheless, the modification cannot be assigned to a specific amino acid. Although the variants peaks can be collected and digested for further site-specific characterization, it is time-consuming and not cost effective. In regard to glycan analysis, hydrophilic interaction chromatography (HILIC) assay is the commonly used approach. The glycans can be released from proteins by specific enzymes, followed by labeling with 2-aminobenzoic acid (2AA) for the fluorescence based HILIC analysis. The striking advancement of high resolution mass spectrometers and statistical analysis software (Thermo LCQUAN and Pinpoint) enable monitoring multiple quality attributes with one analysis which is termed the Multi-attribute method (MAM). The workflow involves peptide mapping of the therapeutic protein followed by identification of product quality attributes (PQAs) peptides (deamidation, oxidation, glycosylation and clipping) with a data base search in PepFinder. Then a Pinpoint workbook is generated for each of the modification that will be monitored. Afterwards, only MS data needs to be collected in MS1 mode for automatic quantitation of each peptide. Potentially this MAM approach could replace the conventional electrophoretic and chromatographic methods currently employed in quality control. It is a universal platform that can be used in almost every step of process development including clone screening.

1.5. **Isoelectric Focusing analysis of protein therapeutics**

Isoelectric focusing (IEF) provides separation of proteins according to the exposed charged residues that behave as weak acids or bases. Proteins will migrate in the presence of a pH gradient...
and electric field until the net charge of the protein is zero, in which migration will cease. The pH at which the protein has zero net charge is termed as the pI of the protein.

1.5.1. Capillary isoelectric focusing

Solution-based IEF is achieved with the development of carrier ampholytes. cIEF is typically performed in a coated capillary where a pH gradient is generated under an electric field using a mixture of mobile carrier ampholytes. The analyte protein will migrate through the ampholyte solution toward the oppositely charge electrode until the pH environment equals its pI.\textsuperscript{105}

Capillary isoelectric focusing (cIEF) allows the separation of protein/peptide mixtures, protein glycoforms and other charge variants, based on their isoelectric point (pI). When compared to conventional isoelectric focusing (IEF) in gels, cIEF allows for higher resolution, faster sample analysis and has an approximately 10-fold lower limit of detection. In addition, with cIEF the conventional one-point detection is overcome by continuous detection over the whole capillary, allowing real-time observation and preventing sample degradation before detection.\textsuperscript{106}

1.5.2. Small scale multi-compartment electrolyzers (MCE)

To overcome the limitation of gel based IEF method, MCE has been developed for the application of proteomics analysis and the key advantages lie in the increased dynamic range and sample loading capacity. On the other hand, an important limitation of solution based IEF is the requirement to remove carrier ampholytes before injection to mass spectrometer.\textsuperscript{107}

One of the commercial available MCE devices is Zoom IEF (ThermoFisher Scientific). It enables solution-based IEF fractionation with an acrylamide buffering membranes. The prototype Zoom
IEF device was first described by Zuo et al and was used to separate metabolically radiolabeled Escherichia coli extracts and serial proof-of-principle experiments. It can reduce highly complex protein samples into discrete fractions before analysis by two dimension electrophoresis.\textsuperscript{108}

The OFFGEL Fractionator (Agilent Technologies) is another widely used MCE that is used to separate proteins based on the IEF process. It provides high resolution as immobilized pH gradient (IPG) strips but recovers separated proteins in the liquid phase. A number of reports are on the utilization of the OFFGEL device for peptide-based fractionation before LC-MS analysis.\textsuperscript{109}

1.6. Continuous Manufacturing

Unlike oil, gas and food industries that employ an integrated continuous manufacturing platform, pharmaceutical manufacturing has been performed using batch processing for more than a century. The conventional batch to batch mode is costly and inefficient, and what’s more, it does not provide the flexibility and on-line quality control required for a modern manufacturing process.\textsuperscript{110} Over the last few decades, there has been a growing interest in implementing the continuous operation in biotherapeutics manufacturing. Continuous manufacturing means material is simultaneously charged and discharged from the process. The major benefits include reduced costs, small scale production and real time quality control.\textsuperscript{111} Sanofi first reported an integrated and automated system of perfusion bioreactor and a four-column periodic counter-current chromatography (PCC) system for the continuous capture of protein therapeutics,\textsuperscript{112} addressing the need for consistent product quality control, high process output and low cost as shown in Figure 1-4.
Figure 1-4: Integration of four-column PCC system with a perfusion cell culture bioreactor to continuously capture the target protein.\textsuperscript{112}


1.6.1. Bio-MOD platform

Biologically-Derived Medicine on Demand (Bio-MOD) project is led by Love Lab in MIT. The initiative is to develop a flexible, portable device platform for manufacturing multiple biologics with high purity, efficacy and potency, at the point-of-care, in short timeframes when specific needs arise. Integrated and Scalable Cyto Technology (InSCyT) was used for a small scale integrated platform for fast production of high-quality drug substance. Three model biologics hGH, IFNα2b and G-CSF was successfully produced in the Stage I. In 7 days, InSCyT has produced hundreds of purified adult doses of hGH, 10,000 of purified doses of IFNα2b, hundred doses of purified doses of G-CSF. The correct drug sequence and PTMs were confirmed by LC-MS and to be comparable to a reference material. The host cell proteins (HCPs) and host cell DNAs were
below FDA enforced limits. The bioactivity was equivalent to a World Health Organization (WHO) standard or product specifications for the marketed drug version. Further implementation of the Bio-MOD platform is aimed to expand to the production of complex protein therapeutics such as monoclonal antibodies and vaccines. Bio-MOD platform is a type of continuous manufacturing and all the materials analyzed were produced from this platform except reference materials.

1.6.2. **Process Related Impurity and Product Related Impurity**

In the process of biomanufacturing, a number of chemicals are typically added upstream to induce expression or improve recovery. These chemicals are supposed to be entirely removed across the downstream purification process. However, trace level of these chemicals is sometimes left. To ensure that the level of these agents is within safe amount, highly sensitive analytical methods need to be used for detection and quantification of such residuals.

According to test procedures and acceptance criteria for biotechnological/biological product (CPMP/ICH/365/96), product related impurities are molecular variants arising during manufacture and/or storage, which do not have properties comparable to those of the desired product with respect to activity, efficacy and safety. These include: N-terminal truncations, PTMs, disulfide bridge scrambling, C-terminal integrity/multiple lysines and aggregation. Other process related impurities include cell substrates host cell proteins and host cell DNA.

1.6.2.1. **Incomplete Removal of leader sequence**

A signal peptide, also called leader sequence, is a short peptide presents at the N-terminus of majority of newly synthesized proteins and directs the secretory pathway. The proteins with this
are mostly secreted from the cell or reside on the cell plasma membrane. A suitable signal sequence is critical to achieve high yield of a secreted protein.\textsuperscript{113}

Recombinant proteins discussed in this thesis are expressed from \textit{Pichia Pastoris} strain using mating factor (MF alpha 1) signal sequence derived from \textit{Saccharomyces cerevisiae}. The MF signal sequence has 85 amino acids which form a pre- and pro- region. Two EA repeats are sometimes included as they can affect protein expression yield, on the other hand, this may lead to N-terminal non-homogeneity due to the inefficiency of the STE13 gene product which is responsible for the enzyme that cleaves of EA repeats. The pre- region is responsible for transferring nascent protein in to the endoplasmic reticulum (ER), followed by subsequent transfer to Golgi organelle. The KEX2 protease will cleave the dibasic KR site of the pro- region to remove the leader sequence.\textsuperscript{114}

Actually, the removal of leader sequence is not always complete due to the complex environment of protein expression and secretion. For example, inefficient processing of both enzymes STE13 and KEX 2 has been reported.\textsuperscript{115-116} Intact mass measurement and peptide finger printing with enzyme digestion are always chosen as orthogonal approaches to characterize the non-homogeneity of such proteins.

1.6.2.2. \textbf{Cleavage of target protein by proteases in the host cell}

Proteolytic enzymes also referred to as protease or peptidase, are enzymes that digest break proteins into peptides and eventually amino acids. In the biomanufacturing industry that use recombinant DNA technology to produce protein drug, proteolytic cleavage is one of the big
concerns with resulting decreases in product activity due to degradation, unexpected immunogenicity and low yields of production.\textsuperscript{117-118}

Protease knockout strains has been developed to address this proteolytic problem,\textsuperscript{119} what more, optimization of fermentation condition\textsuperscript{120} and media composition evaluation can also minimize the proteolytic cleavage.

**1.7. References**


Modification by Selective Derivatization Coupled with RPLC-EThcD-MS/MS. *J Proteome Res* 2017, 16 (2), 780-787.


46


Chapter 2: Integrated bottom-up and top-down liquid chromatography-mass spectrometry (LC-MS) approach for characterization of recombinant human growth hormone degradation products.

This chapter is based on a manuscript with the title of

“Integrated bottom-up and top-down liquid chromatography-mass spectrometry (LC-MS) approach for characterization of recombinant human growth hormone degradation products”


Submitted to Analytical Chemistry

Author Contribution:
Yu (Annie) Wang: perform most experiments and data analysis, write manuscript
Di Wu, Jared Auclair, Joseph Salisbury: assist with top down experiment
Nicholas Mozdziers, Kartik Shah: produce Pichia pastoris material
Anna Zhang: N-terminal leader sequence analysis
Shiaw-lin Wu, Jeffery Agar, Kerry Love, J. Christopher Love, William Hancock: supervisors

Based on chapter 2, the bottom-up approach with multi-enzyme digestion was applied for the study of interferon products in chapter 3.
2.1. Abstract

With the advent of biosimilars to the US market, it is important to have better analytical tools to assure product quality from batch-to-batch and with the innovator drug. In addition, the recent popularity of using a continuous process for production of biopharmaceuticals, traditional bottom-up method, alone, for product characterization and quality analysis is no-longer sufficient. Bottom-up method requires large amounts of material for analysis, is labor-intensive and time consuming. Additionally, in this analysis, digestion of the protein with enzymes such as trypsin could induce artifacts and modifications which would increase the complexity of the analysis. On the other hand, top-down method requires a minimum amount of sample, allows for analysis of the intact protein mass and sequence generated from fragmentation within the instrument. However, fragmentation usually occurs at the N-terminal and C-terminal ends of the protein with less internal fragmentation. Herein, we combine the use of the complementary techniques, top-down and bottom-up method, for the characterization of human growth hormone degradation products. Notably, our approach required small amounts of sample, which is a requirement due to the sample constraints of small scale manufacturing. Using this approach, we were able to characterize various protein variants, including post-translational modifications such as oxidation and deamidation, residual leader sequence and proteolytic cleavage. Thus, we were able to high-light the complementarity of top down and bottom up approach which achieved the characterization of a wide range of product variants in samples of human growth hormone secreted from *Pichia pastoris*. 
2.2. Introduction

Protein drugs, which we refer to as recombinant biologics, are an important class of today’s medicines and seven of the top ten best selling drugs of 2015 were protein drugs. A new class of recombinant biologics, referred to as biosimilars, is just starting to enter the US market. Biosimilars, according to FDA guidance, are highly similar to a pharmaceutical reference material that has already been authorized for marketing as a biological drug. Biosimilars are developed with the purpose of reducing cost and thus increasing patient access. As the patent coverage of some of the best-selling biological medicines will expire over the next few years, there is considerable interest for companies to produce biosimilars. The first biosimilar, Zarxio (Sandoz, subsidiary of Novartis), which is the generic form of Amgen’s Neupogen, was approved in 2015. Another important biologic, recombinant human growth hormone (hGH), which is used for the treatment of growth hormone deficiency (GHD) will be described here as an example of the application of new analytical methods to facilitate the characterization of biosimilar drugs.

Comprehensive and extensive characterization of a protein drug is a key aspect of ensuring the product quality, safety and efficacy. Two key components of a protein therapeutic characterization are complete amino acid sequence mapping and posttranslational modification (PTMs) analysis. Sequence mapping ensures the integrity of the primary structure of the protein, which is a significant attribute for biological function. In addition, the modification of the primary structure by PTMs can have a positive or negative effect on the structure and function of a protein therapeutic. Liquid chromatography coupled with mass spectrometry (LC-MS) is a state of art technique for the characterization of the protein primary structure. Two popular methods are termed as bottom-up and top-down analysis. Bottom-up analysis is a robust way to characterize a protein’s primary sequence and involves a proteolytic digestion with a carefully selected
endoprotease to generate a set of peptides suitable for LC/MS analysis. Multiple enzymes can be used to improve the completeness of the analysis, thus providing 100% sequence coverage and allowing for an in-depth analysis of PTMs and degradation products. However, drawbacks include tedious sample preparation, artifacts induced during the sample manipulation, and the difficulty in obtaining 100% coverage for some protein sequences. By contrast, top-down analysis is used to analyze the intact protein giving a global perspective of all the proteoforms in a given sample, including labile PTMs. A key advantage of top down protein analysis is the lack of requirement for prior knowledge of protein modifications. Major disadvantages of the top-down method include: separation and analysis of unstable proteins, limitation in the MW range of the technique (typically <50kD), complexity of PTMs such as glycosylation and a range of closely related proteoforms related to the complexity of the human genome. In this study, we report on an approach which is complementary to traditional approaches that uses a combination of both top-down and bottom-up methods to analyze a protein therapeutic.

Herein, we present our approach using integrated bottom-up and top-down analysis for the characterization of hGH as a model for facile and complete protein characterization which will be useful for biosimilar development. Firstly, we used stability studies to produce relevant degradation products and thus create a profile of potential modifications and thus define the potential “weak” points of the molecule. We then used both bottom-up (peptide) and top-down (intact protein) LC-MS methods, to analyze these degradation products. As a proof of the applicability of these methods this integrated method was applied to a crude sample of hGH produced by an experimental bioreactor with *Pichia pastoris* as the host organism.

### 2.3. Materials and Methods

#### 2.3.1. Reagents
Human growth hormone reference material was supplied from Sandoz. *Pichia pastoris* derived hGH was kindly provided by the Love Lab from Koch Institute. Fermentation conditions was published earlier. Sodium borate, hydrogen peroxide, Plasmin, ammonium bicarbonate, tris base, dithiothreitol (DTT), and iodoacetamide (IAM) were purchased from Sigma Aldrich. Trypsin and AspN were purchased from Promega. Formic acid, HPLC water and acetonitrile were purchased from Fisher Scientific.

2.3.2. Stability of human growth hormone

To determine the stability of human growth hormone, 100 µl of a 10.7 µg/µl human growth hormone was buffer exchanged into sodium borate, pH 9.0 using an Amicon Ultra centrifugal unit (Millipore) with a 10 KDa molecular weight cut off. Then, the sample was incubated at 37 ºC for 4 weeks and 8 weeks, respectively. Samples were stored at -80 ºC prior to enzyme digestion.

2.3.3. Oxidation of human growth hormone

To prepare the oxidized human growth hormone, 0.05 % hydrogen peroxide was added to human growth hormone in the formulation buffer (10mM sodium phosphate) and incubated at 37 ºC overnight. Samples were stored at -80 ºC prior to enzyme digestion.

2.3.4. Preparation of 2-chain human growth hormone

Two-chain human growth hormone was prepared as previously described. Briefly, human growth hormone was buffer exchanged into Tris-HCl, pH 7.5. Plasmin was added to the sample (1:50, wt/wt) and incubated at 37 ºC for 6 hours. Samples were stored at -80 ºC prior to enzyme digestion.

2.3.5. SDSPAGE of hGH degradation products
500 mL 1X NuPAGE SDS running buffer was prepared by mixing 25 mL NuPAGE MES SDS running buffer (4x) with 475 mL deionized water. 20 µl hGH standard or degradation products were mixed with 6.6 µl LDS sample buffer (4x). Half of the mixture for each sample was treated with 0.5 µl 1M DTT at 70 °C for 10 min for reduction. 5 µl protein ladder and the other samples were loaded on a NuPAGE 4-12% Bis-Tris protein gel and ran under 160V for 50 min. The gel was stained with SimpleBlue SafeStain for 30 min and destained in deionized water overnight.

2.3.6. Analysis of tryptic digest of human growth hormone (bottom-up method)

SDS-PAGE under reduced condition: 20 µl of human growth hormone, 2 µl 1 M DTT, 7 µl 4X sample loading buffer were mixed together and heated at 90 °C for 30 min, then loaded onto an SDS-PAGE gel. In-gel digestion: Gel bands were cut into 5 mm x 5 mm cubes and destained to remove the Coomassie blue. Prior to reduction, the gel slices were shrunk by incubating with neat acetonitrile and rehydrated with 200 µl 10 mM DTT in ammonium bicarbonate, pH8 for 30 min at 56 °C. After removing the DTT solution, gel slices were shrunk again with acetonitrile and rehydrated with 55 mM iodoacetamide in ammonium bicarbonate, pH 8 and incubated for 1 hour at room temperature in the dark. Then a trypsin digestion buffer (12.5 ng/µl trypsin, 5 mM CaCl2 in 50 mM ammonium bicarbonate, pH 8.0) was added to the gel slices for 35 min at 37 °C followed by incubation at 37 °C overnight. The digested peptides were extracted with 25 mM ammonium bicarbonate followed by acetonitrile for two cycles and the reaction was stopped with 5 % formic acid. All supernatants were collected together and concentrated using a speed vacuum to 10 µl for LC-MS analysis. In-solution digestion: 20 µl of 1µg/µl human growth hormone was denatured by 6 M guanidine hydrochloride, followed by reduction with 5 mM DTT for 30 min at 37 °C, and alkylation with 10 mM iodoacetamide for 1 hour at room temperature in the dark. Then human
growth hormone was buffer exchanged into ammonium bicarbonate, pH 8.0 as previously described. Trypsin digestion was performed by incubating trypsin (1/50, w/w) with human growth hormone at room temperature. After 8 h, additional trypsin was added (1/50, w/w) and the digestion continued for more 12 h. For the AspN digestion, AspN (1/100, w/w) was added to human growth hormone and incubated at 37 °C overnight. Finally, trypsin and AspN were incubated together; trypsin (1/50, w/w) was added to the protein solution and incubated at 37 °C for 4 hours, and then AspN (1/100, w/w) was added to the protein solution and incubated at 37 °C overnight.

2.3.7. **LC-MS: Dionex Ultimate 3000 LC-LTQXL**

An Ultimate 3000 nano LC pump (Dionex, Mountain View, CA) and self-packed C18 column (Magic C18, 200Å pore and 5 µm particle size, 75 µm internal diameter (ID) by 100 mm) connected to a coated 10 µm ID emitter (New Objective, Woburn, MA) were coupled online to an LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) through a nanospray ion source (New Objective, Woburn, MA). Mobile phase A was 0.1 % formic acid in HPLC grade water and mobile phase B was 0.1 % formic acid in acetonitrile. The flow rate was 250 nL/min with 2 % B for 25 min. The flow rate was maintained at 200 nL/min during the separation. The gradient is as follows: 0-60 min 2-40 % B, 60-70 min 40-90 % B, 70-75 min isocratic at 90 %B and 75-78 min 2% B. The mass spectrometer was operated in a data dependent mode (DDA) to switch between MS and CIDMS2. Briefly, after a full-scan MS spectrum from m/z 400-2000 in the linear ion-trap, 8 CIDMS2, with 28 % normalized collision energy and activation Q at 25, activation steps were performed on the 8 most intense precursor ions from the full scan.
2.3.8. **Analysis of intact human growth hormone (top-down method), LC-MS: Eksigent LC-Bruker SolariX (FTMS)**

For direct infusion experiments, protein solutions were diluted to 5 pmol/ul in 50 % acetonitrile, 0.1 % formic acid and infused into a FTICR MS (Bruker Daltonics, SolariX) at a flow rate of 120 µL/hr. Prior to LC-MS analysis human growth hormone was diluted with 0.1 % formic acid in HPLC water to approximately 5 pmol/ul and spun down at 13000 rpm for 10 min. For LC-MS and LC-MS/MS experiments, an Eksigent nano LC (AB SCIEX, Framingham, MA) system fitted with a self-packed polystyrene-based PLRP-S (Agilent) reversed-phase column (1000 Å, particle size 5 µm, 150 µm id, 3 cm length trapping, 6 cm length) was coupled to a 9.4 T SolariX XR FT-ICR mass spectrometer (Bruker Corporation, Billerica, MA). Mobile phase A was 0.1 % formic acid in water and mobile phase B was 0.1 % formic acid in acetonitrile. The gradient is as follows: 0-6 min with 10 % B, 6-16 min 10 %-60 % B, 16-19 min 60 %-100 %B, 19-24 min 100 % B, 25-30 min 10 % B. The flow rate was 3 µl/min and the injection volume was 1 µl.

The mass spectrometer was externally calibrated in the m/z domain with ESI-L low concentration tuning mix (Agilent, Santa Clara, CA, USA) using peaks at m/z 322, 622, 922, 1222, 1522, and 1822 to an RMSE of 230 ppb. Source parameters were 4.0 L/min dry gas flow (180 ºC) with capillary voltage set to 3.5 KV. The capillary exit and deflector plate were at 220 V. For MS1 acquisition, ion accumulation time was 0.2 seconds. For intact mass measurements, funnel 1 and skimmer 1 were held at 140 and 40V, respectively. Skimmer 1 was increased to 120-135 V for in-source fragmentation experiments. For CASI-CAD fragmentation, the ions between m/z 900-1500 were selected in the quadrupole with an accumulation time of 0.385 seconds and a collision energy of 14.0 V was applied to induce fragmentation. The potential of funnel 2 was held at 6.0 V in all
experiments. The data was acquired in the broadband detection mode, using 2-megaword data points, over the 150-3000 m/z range.

2.3.9. **Data Analysis**

2.3.9.1. **Bottom up**

Proteome Discover, manual inspection. For peptide identification, raw data were searched against the human growth hormone sequence using SEQUEST incorporated in Proteomic Discover 1.4 (Thermo Fisher Scientific). Peptide precursor ion mass tolerance was set to 1.0 Da, and the fragment ion mass tolerance 1.0 Da. Oxidation of Methionine (Met) and Deamidation of Asparagine (Asn) were set as dynamic modifications and carbamidomethylation was set as a fixed modification. The identified peptides were then filtered using Xcorr score (1+ precursor ion >1.5, 2+ > 2.0, and 3+ and above > 2.5). Mass accuracy was set to < 50 ppm. Final confirmation of the peptide identification was done by manual inspection, extracting the base peak from the chromatogram and matching the MS/MS fragmentation data with theoretical prediction.

2.3.9.2. **Top-down**

Data Analysis, MaxEntX deconvolution, ProSight Lite. Spectra were processed using DataAnalysis (Bruker Corporation, Billerica, MA). To identify intact masses, spectra were averaged in half minute intervals during the LC-run, and the deconvoluted spectra were obtained using Maximum Entropy Deconvolution. The low mass was set to 20000 m/z and high mass 25000 m/z. Data point spacing was set as auto. FTMS instrument MaxEntX option was chosen to get high resolution information. For MS/MS data analysis, mass lists were generated by using the SNAP algorithm. The quality factor threshold was set to 0.5, signal-to-noise threshold was set at
2, relative intensity threshold at 0.01 %, absolute intensity threshold 0, and maximum charge state 20. The mass list of fragmentation ions was inputted into the experimental data section in ProSight Lite (Northwestern University). Precursor mass type was set as monoisotopic, mass mode MH+, fragmentation method was CID and fragmentation tolerance 30 ppm. The human growth hormone sequence was entered into the candidate sequence section. For non-reduced samples, cysteine was set with “hydrogen loss” and no modification was set for the reduced samples. For oxidized samples, Met was set to have a modification of +15.9949 or +32.9989 to confirm the modification site. For 2-chain samples, truncated human growth hormone sequence from the new N-terminal residue Gln141 or Thr142 to the C-terminus was entered into the modified candidate sequence for identification of new N-terminal fragmentation.

2.4. Results and Discussion

Characterization of commercial hGH has been previously published using trypsin digestion with 100% sequence mapping, as well as characterization of degradation reactions oxidation and deamidation and determining disulfide bond connectivity. In addition, these sequence mapping studies can be used to confirm the identity of the biosimilar product with reference material and reveal the presence of any amino acid mutations. Characterization of the nature of post-translational modifications (PTMs) as well as level of structural variants is also essential as presence of the PTMs might have both structural and biological consequences. Oxidation of Met 14 and Met 125 of hGH does not induce conformational changes, however, oxidation of Met 170 will disrupt the structure of hGH as it is located in the hydrophobic core of the protein. Deamidation of asparagine is one of the most frequent modifications in proteins, which is a spontaneous and non-enzymatic reaction. Through a succinimide intermediate, asparagine is
converted to aspartic acid (Asp) or isoaspartic acid (isoAsp) which is a sign of protein aging and detrimental to biological activity.\textsuperscript{15-16}

Figure 2-1 presents the sites of modification detected by the integrated bottom-up and top-down method for the characterization of a hGH commercial standard and corresponding degradation products. The bottom-up method, which utilized both AspN and trypsin digestion, respectively, was used to generate the peptides containing the sites of modification that were suitable for LC-MS analysis. As well, complementary data was obtained by the top-down analysis of the intact protein sample. This integrated method can not only give 100\% sequence coverage of hGH, but also information of degradation sites, including the 2-chain variant, oxidation, deamidation and N-terminal leader sequence. On the other hand, neither the bottom-up or top-down method could provide the complete data set.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{sequence_map.png}
\caption{Sequence map of hGH modifications sites identified by either the bottom-up and top-down method.}
\end{figure}

Bottom up analysis: Red boxes: peptides containing modifications that were generated by AspN digestion. Blue boxes: trypsin digestion. Top down analysis: (A.): b-ion series from N-terminus denoted with blue lines. (B.): In the 2-chain sample, the b-ion series was observed from the new N-terminus which is not observed in the reference 1-chain material.
Table 2-1 summarizes the suitability of the top-down and bottom-up approach for the characterization of a hGH commercial standard and degradation products. Molecular weight differences are listed for the degradation samples including the 4-week, 8-week stability samples, an oxidized sample and 2-chain sample were determined by SDS-PAGE as shown in Figure 2-2. Degradation processes associated with chain cleavage and aggregation were observed in both the 4 week and 8 week stability samples with the appearance of both lower and higher MW bands. In the two sample sets reduction of the cysteine linkages resulted in the appearance of lower MW bands which can be related to degraded, cleaved species were held together by disulfide bonds. The oxidized samples did not contain similar bands due to either cleavage or aggregation processes. In all reduced samples, a slight increase in apparent MW of the main band was attributed to an increase in the hydrodynamic volume of the molecule due to reduction of the disulfide bonds. On SDS-PAGE analysis under reduced conditions, 2 bands for the cleaved form of 2-chain growth hormone were observed at the expected MW values. Characterization of the site of cleavage of 2-chain cleavage hGH was first reported back in 1990 and the cleaved form was reported to result in increased biological activity.17

<table>
<thead>
<tr>
<th>HGH Variants</th>
<th>Modifications</th>
<th>Amino acid position</th>
<th>Monoisotopic Mass Difference[Da]</th>
<th>Bottom-up</th>
<th>Top-down</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>Met Oxidation</td>
<td>14,125</td>
<td>+15.9949</td>
<td>√</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Deamidation</td>
<td>149</td>
<td>-0.9840</td>
<td>√</td>
<td>ND</td>
</tr>
<tr>
<td>Stability Sample</td>
<td>&quot;PhePro&quot; cleavage</td>
<td>1-2</td>
<td>-244.1211</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Met Oxidation</td>
<td>14,125</td>
<td>+15.9949</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Deamidation</td>
<td>149</td>
<td>-0.9840</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>Oxidized Sample</td>
<td>Met Oxidation</td>
<td>14,125</td>
<td>+15.9949</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Met Sulfone</td>
<td>14,125</td>
<td>+31.9898</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>2-chain Sample</td>
<td>Cleavage</td>
<td>141,142</td>
<td>+18.0106</td>
<td>√</td>
<td>√</td>
</tr>
</tbody>
</table>

Table 2-1: Comparison of Top Down and Bottom Up characterization of human growth hormone
Figure 2-2: SDS-PAGE of hGH variants under non-reduced and reduced conditions.

For each sample, the left lane is the non-reduced form and right lane is the reduced form.

Using the bottom-up method, the ratio of modifications (oxidation and deamidation) in the reference material was below 3%, with minimal artifacts detected as a result of our analytical protocols. The correct N- and C-terminal sequences and no internal cleavages were also confirmed.

In contrast, high levels of oxidation, deamidation and peptide backbone cleavages were observed in the stability samples (see Table 2-2)

<table>
<thead>
<tr>
<th></th>
<th>Standard %</th>
<th>4week %</th>
<th>8week %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxidation M14</td>
<td>1.5</td>
<td>6.3</td>
<td>10.2</td>
</tr>
<tr>
<td>Oxidation M125</td>
<td>3.9</td>
<td>8.6</td>
<td>9.3</td>
</tr>
<tr>
<td>Oxidation M170</td>
<td>4.4</td>
<td>3.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Deamidation N149</td>
<td>3.0</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

Table 2-2: Bottom-up result of oxidation and deamidation analysis of the stability samples.
For the oxidized samples, methionine oxidation to the sulfoxide (+15.99Da) was 80% for methionine 14, 30% for methionine 125 and 9% for methionine 170. The cleavage sites for the 2-chain variant produced in the degradation sample were determined using LC/MS analysis of the AspN digestion. AspN was also applicable for characterization of known two-chain variants generated by intracellular or secreted proteases with two different cleavage sites between either Gln 141 and Thr 142 or Thr142 and Tyr 143. This example shows the power of the multi-enzyme approach where the optimal protease is selected for analysis of a given variant. Conversely, for the two-chain product, trypsin will generate “TYSK” which is too short and hydrophilic to retain on the C18 column for LC-MS analysis. By contrast AspN will give “TYSKF” which can be retained on C18 column for LC-MS analysis. Detailed peptide identification for cleavage analysis using bottom-up method are summarized in Table 2-3: Glutamine 141 and Threonine 142 cleavage are determined from a plasmin cleaved reference material using bottom up method. However, a disadvantage of the bottom-up method is the amount of sample manipulation that takes place, such that, oxidation artifacts can occur as a result of sample preparation methods (e.g. high voltage during SDS-PAGE, incubation for several hours for the enzyme digestion).18

<table>
<thead>
<tr>
<th>Peptide</th>
<th>m/z</th>
<th>RT (min)</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DGSPRTGQIFKQTYSF</td>
<td>980.82 (2+)</td>
<td>30.05</td>
</tr>
<tr>
<td>2</td>
<td>DGSPRTGQIFKQ</td>
<td>567.75 (2+)</td>
<td>21.18</td>
</tr>
<tr>
<td>3</td>
<td>TYSKF</td>
<td>545.67 (1+)</td>
<td>15.19</td>
</tr>
<tr>
<td>4</td>
<td>DGSPRTGQIFKQT</td>
<td>718.25 (2+)</td>
<td>21.96</td>
</tr>
</tbody>
</table>

Table 2-3: Glutamine 141 and Threonine 142 cleavage are determined from a plasmin cleaved reference material using bottom up method.

Peptide 1 “DGSPRTGQIFKQTYSF” is the intact peptide without cleavage. Peptide 2 “DGSPRTGQIFKQ” and Peptide 4 “DGSPRTGQIFKQT” are the cleaved peptide with a cleavage site of between 141 Q and 142 T and between 142 T and 143 Y. Peptide 3 “TYSKF” is the
additional peptide generated after cleavage, which is additional evidence of cleavage between 141 Q and 142 T.

Our top-down method yielded results that were complementary to the observations from the bottom-up approach. Moreover, due to the avoidance of extensive sample handling steps, we did not have the same concern with analytical artifacts. Figure 2-3 shows the result of measurement of the reference material and stability samples by the top-down method. The reference material had the expected monoisotopic mass of 22125.58 Da and any degradation products (e.g. oxidation and deamidation) and product variants were below the detection limit of our procedure. We also used the top-down method to examine the N-terminal sequence as in source fragmentation in the FTICR MS generates fragment ions mostly from the N-terminus and C-terminus. With the reference material, we did not observe any N-terminal variants or any internal cleavages. In addition to the expected oxidation and deamidation reactions (+16Da and +1 Da peaks), both the 4-week and 8-week stability samples had an approximately -31 Da and -4 Da mass shift which remains uncharacterized (peak 2 and 3 in Fig 2-3b and peak 6 and 7 in Fig 2-3c).
Figure 2-3: Deconvoluted spectra of hGH standard and hGH degradation products.

Intact mass and any mass shift was labeled in each panel. 3a was the reference material; 3b was 4 week stability sample, peak 2 had mass shift of -31Da, peak 3 had mass shift of -4Da, peak 4 had mass shift of +1Da due to deamidation, peak 5 had the mass shift of +17Da in addition to the +1Da due to deamidation, which might be due to oxidation; 3c was 8 week stability sample which had the same peaks profile as 3b; 3d was the 2-chain sample, peak 9 was the un-cleaved form and peak 10 was the cleaved form and the +18Da mass shift was due the addition of H2O caused by the cleavage reaction; 3e was the oxidized sample, peak 11 has mass shift of +32Da meaning the addition of 2 oxygen atoms.

Another advantage of top down method was our ability to determine how many methionines were oxidized in the stability study (0.05 % H2O2 at 37 °C overnight) where the major component was a +31.99Da species which could be attributed to the addition of two oxygen atoms per protein.
molecule (see Fig 2-3e). This result could be explained by either oxidation of 2 methionine residues (addition one oxygen atom per methionine) or formation of a sulfone (double oxidation of a single methionine residue). To isolate fragment ions suitable for this determination we shifted to the LC/MS analysis to under reduced conditions where more fragmentation ions were observed in the inner regions of the protein than for the non-reduced material (see Figure 2-4). In the fragmentation data of the reduced sample for a specific precursor ion (e.g. +31.99) the b ion series of b26, b30 and b34-b36 had a mass shift of +15.99 Da, indicating Met 14 had the addition of one oxygen atom, no mass shift of y33 and y37 suggested no oxidation at Met 170, and a mass shift of +15.99 Da y86, y88 and y90 proved the oxidation of Met 125. The oxidation pattern of two methionine residues (Met 14 and 125) was thus specified for the oxidized sample. While the bottom-up approach would be able to characterize the two sites of oxidation in the corresponding enzymatically generated peptide sequences it would be difficult to determine the molecular heterogeneity from just the peptide analysis.

![Peptide Sequence](image)

**Figure 2-4:** Analysis of methionine oxidation in hGH oxidized sample using top-down method under reduced condition.

Blue arrows indicate the site of fragmentation, up arrows stand for b ions and down arrows stand for y ions. Amino acids with yellow boxes are ones could be oxidized.
Top-down intact mass measurement of the two-chain sample showed a mass shift of +18.01 (one H2O molecule) due to the polypeptide backbone cleavage as shown in the third panel of Fig 2-3d. A new N-termini was generated because of the endoproteolytic cleavage, which can be monitored by fragment ions generated using in-source fragmentation via top-down analysis. As shown in Figure 2-5a, b ion series b2, b6, b7 and b 10 as well as y ion series y2, y3 and y7 were identified to confirm the cleavage between Q141 and T142. In Figure 2-5b, b ion series b6, b10 and b11 as well as y ion series y2, y2 and y7 were identified to confirm the cleavage site between T142 and Y142. With the lack of availability of Edman N-terminal sequencing technology this method has significant potential for the monitoring of proteolysis and other cleavage reactions in recombinant protein products.

Figure 2-5: Cleavage of hGH 2-chain sample determined by the top-down method.

Black boxed amino acids are the sites of the two new N-termini. Blue arrows indicate the site of fragmentation, up arrows stand for b ions and down arrows stand for y ions. Fragmentation ions were observed in the new N-termini which were not observed in the reference material and were used to identify the cleavage site. 5a: fragment ions were identified from the new N-termini Thr 142 to C-termini Phe 191. 5b: fragment ions identified from the new N-termini Tyr 143 to C-termini Phe 191.
We then evaluated this developed method for a production sample, namely the analysis of hGH production using *Pichia pastoris* as a host.\textsuperscript{19} To promote secretion of the product, alpha factor was incorporated in the host DNA sequence and after protein expression the factor was removed by a signal peptidase.\textsuperscript{20-21} A particular concern for protein manufacture is that the N-terminal EAEA repeats in the C-terminal region of the secretion sequence were sometimes inefficiently removed by the signal peptidase.\textsuperscript{22} Thus it was necessary to check for the presence of alpha factor sequence still be attached to the N-terminus of protein and the full sequence is given here (MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEEPIGNGYSDLEGDFDVAVLPFSNSTN NGLLFINTTIASIAAKEEGVSLEKREA). Using trypsin digestion, we detected the peptide EAFPTIPLSR (33.3\%) and EAEAFPTIPLSR (54.9\%) where the residues FPTIPLSR corresponded to the N-terminus of hGH and indicated that part of the leader sequence is still attached N-terminus of hGH. Given the high level of residual leader sequence, the host strain was redesigned and EAEA repeats removed from the leader sequence. In this case, trypsin was not suitable for the detection of any residual leader sequence as it would cleave off any “K” and “R” residues still attached to the N-terminus of growth hormone and thus from the N-terminal tryptic peptide. On the other hand, AspN is ideal as it wouldn’t cleave any N-terminal “K” and “R” residues and also generates N-terminal peptides of an appropriate size for LC-MS analysis. No leader sequence was detected by AspN digestion after the removal of the EAEA repeats from the leader sequence. In this manner, the analytical results gave important feedback and allowed the production team to reengineer the host to ensure complete removal of the leader sequence. AspN is also suitable for the cleavage characterization of process development product as shown in table 3. The cleavage site (between Gln 141- and Thr 142) of the *Pichia pastoris* derived hGH (see Table 2-4:) is different from the cleaved sequence that was reported for Genentech hGH (between Thr
142- and Tyr 143)17 and for plasmin cleaved reference hGH (between Gln 141- and Thr 142 and between Thr 142- and Tyr 143) The variable cleavages in the protein is presumably due to the location of these residues in an exposed surface loop of hGH which is easily accessed by different endoproteases.

<table>
<thead>
<tr>
<th>peptide</th>
<th>m/z</th>
<th>RT (min)</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 DGSPRTGOQIKATYSKF</td>
<td>980.63(2+)</td>
<td>25.60</td>
<td>7.4956</td>
</tr>
<tr>
<td>2 DGSPRTGOQIKQ1</td>
<td>667.75 (2+)</td>
<td>21.78</td>
<td>1.8866</td>
</tr>
<tr>
<td>3 TYSKF</td>
<td>645.32 (1+)</td>
<td>15.86</td>
<td>2.7515</td>
</tr>
</tbody>
</table>

Table 2-4: Peptide sequences detected in the analysis of 2-chain variants from a Pichia pastoris derived bioreactor sample using the bottom up method

2.5. Conclusion

The complementarity of our top-down (intact proteins) and bottom-up methods (proteolytic digests) was demonstrated by profiling hGH reference material, degradation products and a Pichia pastoris derived product. Establishing an integrated method for product characterization to guide development of a new protein therapeutic or a follow on biological product is of great significance for successful upstream process development. By determining points of instability in the target protein molecule and critical quality attributes, we can minimize any elements affecting the product quality. In this study, we demonstrated that the complete amino acid sequence together with any heterogeneity in expected N and C termini as well as variants generated by oxidation, deamidation, residual leader sequence and internal cleavages could be determined in an integrated and comprehensive manner by our approach. Such information can be used to establish an efficient process for process development to optimize the fermentation process and thus ensure a correct and high-quality product.
We expect that this approach will have particular application to biosimilar development programs where the follow-on product needs to be highly similar to the innovators product. The implementation of this platform for the analysis of recombinant human growth hormone (hGH) is therefore a proof of concept. In the future, we plan to apply this analytical platform to other protein therapeutics development and characterization to demonstrate the generality of the approach.

2.6. References


Chapter 3: The application of HPLC/MS analysis with a multi enzyme digest strategy to characterize different interferon products produced from *Pichia pastoris*

This chapter is based on a manuscript with the title of

“The application of HPLC/MS analysis with a multi enzyme digest strategy to characterize different interferon product variants produced from *Pichia pastoris*”


Author Contribution:
Yu (Annie) Wang: perform most experiments and data analysis, write manuscript
Di Liu, Laura Crowell: produce *Pichia pastoris* material
Shiaw-lin Wu, Kerry Love, William Hancock: supervisors

*Based on chapter 3, the method for studying product variants of interferons produced in Pichia pastoris was applied for the study of G-CSF in chapter 4.*
3.1. Abstract

Interferons are signaling proteins that belong to the large class of proteins known as cytokines. Human interferons are classified based on the type of receptor interactions: type I, II and III. IFNα2b belongs to type I interferon class and can be used for the treatment of hepatitis B and C infections. A recombinant form of IFNα2b expressed in \textit{E.coli}, known as IntronA, has been approved by FDA. IFNγ, also known as type II interferon, plays a significant role in the inhibition of viral replication. Actimmune® is a FDA approved version of IFNγ for the indication of reducing infections associated with Chronic Granulomatous Disease and Severe Malignant Osteopetrosis.

In this study, we have applied advanced analytical methods for the characterization of IFNα2b and IFNγ produced form \textit{Pichia pastoris}. The multi-enzyme digestion approach has been developed to allow characterization of 100% sequence coverage, identification of the following variants: N-terminal leader sequence, an amino acid substitution and oxidation in IFN α2b, two sites of high-mannose N-glycosylation, C-terminal proteolysis of IFNγ, oxidation and deamidation. In this manner, the analytical program was able to support rapid process development as well as identify product variants and degradation products.
3.2. Introduction

3.2.1. Recombinant interferon α2b

Human interferon α2b is a subtype of the human alpha interferon family and the potential therapeutic applications have been investigated extensively.\(^1\,^{2}\) Recombinant IFN α2b is a 165-amino acid single chain polypeptide with a molecular weight 19.2 kDa. Licensed products include Intron A cloned in Escherichia coli and produced by Merck and Shanferon cloned in *Pichia pastoris* and sold by Shantha in India. Intron A was first marketed by Schering Corp (Merck) in 1986,\(^3\) and is widely used for the treatment of myelomas, lymphomas, solid tumors and chronic hepatitis B and C infections.

3.2.2. Recombinant interferon gamma

Interferon gamma (IFN\(\gamma\)) is the only member of the type II class of interferons and it is produced by lymphocytes activated by specific antigens. In addition to its antiviral activity, IFN\(\gamma\) has important immunoregulatory functions.\(^4\) The FDA approved interferon gamma-1b Actimmune® is a 143-amino acid protein and when produced in Escherichia coli it has an additional N-terminal Met compared residue. Unlike human interferon gamma-1b, Actimmune is not glycosylated as a result of using *E.coli* as host system.

3.2.3. Leader sequence: alpha-factor signal from *Pichia pastoris*

In the biopharmaceutical industry, there has been a growing interest in using *Pichia pastoris* a host system due to favorable economics which is related to high growth rates which allows reduced fermentation times, the use of defined media instead of complex media and secretion of low levels of host cell proteins.\(^5\) The Saccharomyces cerevisiae α-factor preprosequence is the most widely
used secretory signal in *Pichia pastoris* and is cleaved into three peptides (one very long) upon secretion from *Pichia pastoris*. The endogenous enzymes responsible for these cleavage events are Kex2 and Ste13 and often result in heterogeneity of N-terminal amino acid sequence of the secreted product due to the incomplete removal of leader sequence. The most common variant contains an additional EAEA tetrapeptide at the N-terminus. However, the α factor vector can be designed without the EAEA repeats to avoid the high level of N-terminal heterogeneity of the protein. The efficiency of Kex2 and Ste13 proteases are a key factor for the successful removal of leader sequence which can be related to the surrounding amino acid sequence which can result in steric hindrance of the cleavage process.

3.2.4. **High mannose structure of glycosylation of protein expressed in *Pichia pastoris***

A limitation with the use of *Pichia pastoris* is that N-linked glycosylation is restricted to the high-mannose type, which is immunogenic and results in poor pharmacokinetic parameters due to clearance via binding to high mannose receptors. Interferon expression in *Pichia pastoris* has previously been attempted but, however, with little protein characterization or product quality information. This paper presents the use of advanced analytical methods for the in-depth characterization for the support of process development of interferon products produced in *Pichia pastoris*. We used the multi-enzyme digestion strategy coupled with a liquid chromatography - mass spectrometry approach for the identification of product variants of interferons identified in the fermentation studies, including residual leader sequence, an amino acid substitution, backbone cleavages, glycosylation as well as degradation reactions.

3.3. **Materials and Methods**
3.3.1. Chemical and reagents

Equipment: XCell SureLock Mini-Cell electrophoresis apparatus was from Thermo Fisher Scientific. A benchtop centrifuge was purchased from Corning. Ultimate 3000 nano LC pump (Dionex, Mountain View, CA). 10 µm ID emitter (New Objective, Woburn, MA), LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA).

Ammonium bicarbonate (ABC), dithiothreitol (DTT) and iodoacetamide (IAA) was from Sigma Aldrich. Trypsin, and GluC was purchased from Promega. Guanidine hydrochloride, formic acid optima LC/MS grade, water LC/MS grade, acetonitrile (ACN), Methanol, Sulphosalicylic acid and trichloroacetic acid, hydrochloride acid was purchased from Fisher Scientific. NuPAGE Novex 4-12% Bis-Tris gel, 1.0 mm, 10 well, NuPAGE MES SDS running buffer (20X), NuPAGE LDS Sample Buffer (4X), PageRuler Prestained 10-180KDa Protein Ladder, Invitrogen SimpleBlue SafeStain, Novex pH 3-10 IEF Protein Gels, 1.0 mm, 12 well, Novex pH 3-10 Buffer Kit (includes Novex IEF Anode Buffer 50X, Novex IEF Cathod Buffer pH 3-10 10X and Novex IEF Sample Buffer pH 3-10 2X), Colloidal Blue Staining Kit and premium grade TCEP-HCl was purchased from Thermo Fisher Scientific. IEF standards was from Bio Rad. IFN γ standard material was purchased from Abcam (Cambridge, MA).

3.3.2. Sample preparation

3.3.2.1. SDS-PAGE of interferon α2b

500 mL 1X NuPAGE SDS running buffer was prepared by mixing 25 mL MES running buffer with 475 mL deionized water. 10 µl (1µg/µl) IFN α2b standard with 3.3 µl LDS sample (4x) were mixed and 10 µl (1µg/µl) Pichia pastoris IFN with 3.3 µl LDS sample buffer were mixed. Gel apparatus was set and 5µl protein ladder was loaded in lane 1, one blank lane was left between
each sample and IFN standard and *Pichia pastoris* IFN were loaded. The voltage was set at 160 V and run for 50 min. Gel apparatus was disconnected the gel was removed from two plates, stained with SimpleBlue SafeStain for 30min and destained in deionized water overnight.

3.3.2.2. **Isoelectric focusing (IEF) gel of interferon α2b**

200 ml cathode buffer for upper chamber was prepared by mixing 20 mL cathode buffer with 180 mL deionized water. 500 mL anode buffer for lower chamber was prepared by mixing 10 mL anode buffer with 490 mL deionized water. 10 µl IFN standard was mixed with 10 µl IEF sample buffer, same as *Pichia pastoris* IFN. Gel apparatus was set and 200 ml cathode buffer was filled in the upper chamber and 500 ml anode buffer was filled in the outer chamber. 2µl IEF standard was loaded on lane 1, one blank lane was left on purpose to avoid the contamination of the samples to each other. IFN standard and *Pichia pastoris* IFN were loaded. The whole apparatus was put in an ice bucket. The voltage program was set as follows: 100 V for 60 min, 200 V for 60 min and 500 V for 30 min. While in the last 30 min of the run, fixing solution was prepared by adding 11.46 g trichloroacetic acid and 4.6 g sulphosalicylic acid in 90 ml water. When the voltage program finished, the gel was washed with deionized water three times and then transferred to the fixing solution. The gel was fixed for 60 min on a shaker. Staining solution was prepared by mixing 58 mL water, 20 mL methanol, 20 mL stainer A and 2 mL stainer B of the colloidal staining kit. The gel was then destained in water for overnight.

3.3.2.3. **Trypsin digestion of interferon α2b**

20 µL 6M guanidine hydrochloride (MW 95.53 g/mol) in 0.1 M NH4HCO3 (ABC) was added to 20 µg sample of purified IFN. Followed by adding 2 µL of 1 M Dithiothreitol (DTT) solution to
the solution (Final concentration: 5mM DTT) and heat at 37°C for 30 minutes. Fresh 1M iodoacetamide (IAA) solution was made in a foil-wrapped tube to avoid exposure to light and was added to the denatured, reduced IFN solution to a final concentration of 10mM. The tube was covered with aluminum foil and incubate at room temperature (25 °C) for 1hr. After incubation, the solution was buffer exchanged to 100 mM ammonium bicarbonate pH 8 using 10KDa cutoff spin column with speed of 14000 rpm for 6 min. For buffer exchange, 100 mM ammonium bicarbonate solution was added with final volume 500 µl in the cut-off tube and reduce to 100µl each time. This step was repeated three times to make sure the protein in pH 8 solution for enzyme digestion. A first dose of trypsin (1 µl at 0.5µg/µl) was added to the solution and incubate at 37°C for 4 hours. Then a second dose of enzyme with same amount was added for incubation overnight (16 hr) at room temperature. The digestion was stopped by adding 4 µl 1% formic acid to make the final pH around 3. The peptide solution was aliquot to 30 µl/vial at -80 °C before LC/MS analysis.

3.3.2.4. **GluC digestion of interferon α2b**

20 µL 6M guanidine hydrochloride (MW 95.53 g/mol) 0.1 M ammonium bicarbonate (ABC) was added in to 20 µg sample of purified IFN. Followed by adding 2 µL of 1 M Dithiothreitol (DTT) solution to the solution (Final concentration: 5mM DTT) and heat at 37°C for 30 minutes. Fresh 1M iodoacetamide (IAA) solution was made in a foil-wrapped tube to avoid exposure to light and was added to the denatured, reduced IFN solution to a final concentration of 10mM. The tube was covered with aluminum foil and incubate at room temperature (25 °C) for 1hr. After incubation, the solution was buffer exchanged to 100 mM ammonium bicarbonate pH 8 using 10KDa cutoff spin column with speed of 14000 rpm for 6 min. For buffer exchange, 100 mM ammonium
bicarbonate solution was added with final volume 500 µl in the cut-off tube and reduce to 100µl each time. This step was repeated three times to make sure the protein in pH 8 solution for enzyme digestion. A first dose of Glu-C (1 µl at 0.5µg/µl) was added to the solution and incubate at 30 for 4 hours. Then a second dose of enzyme with same amount was added for incubation overnight (16 h) at room temperature. The digestion was stopped by adding 4 µl 1% formic acid to make the final pH around 3. The peptide solution was aliquot to 30 µl/vial at -80 °C before LC/MS analysis.

3.3.2.5. **In-gel digestion of IFN gamma gel**

SDS-PAGE gel with IFN gamma bands was provided by our collaborator as shown in Figure . The standard material was purchased from Abcam (Cambridge, MA). Day 1: The gel bands were cut into small pieces using scalpel. 200 µl of acetonitrile (ACN) and LC-MS water (1:1) was added to cover the gel pieces and shake at 800 rpm on Eppendorf incubator for 1 hour until the gel pieces are clear. The liquid was removed. 100 µl of acetonitrile was added to shrink the gel piece and the liquid was removed afterwards. The gel pieces were then rehydrated with trypsin solution (12.5 ng/ µl trypsin in 50 mM ammonium bicarbonate) at 4 °C for 30 min. After rehydration, the remaining supernatant was removed, 100 µl 50mM ammonium bicarbonate was added to cover the gel pieces followed by incubation at 37 °C overnight. Day 2: 50 µl 25mM ammonium bicarbonate was added to the gel pieces and incubate at 37 °C for 15 min with shaking. Liquid was spin down and 100 µl acetonitrile was added followed by incubation at 37 °C for 15 min with shaking. The supernatant was spin down and collected in another tube. The extraction steps were repeated 3 times. In the last extraction step, 50 µl 1% formic acid was added to stop the digestion. The combined extracts were concentrated to 10 µl using a speed vacuum concentrator.

3.3.2.6. **LC-MS/MS**
An Ultimate 3000 nano LC pump (Dionex, Mountain View, CA) and self-packed C18 column (Magic C18, 200Å pore and 5 µm particle size, 75 µm internal diameter (ID) by 100 mm) connected to a coated 10 µm ID emitter (New Objective, Woburn, MA) were coupled online to an LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) through a nanospray ion source (New Objective, Woburn, MA). Mobile phase A was 0.1 % formic acid in HPLC grade water and mobile phase B was 0.1 % formic acid in acetonitrile. The flow rate was 250 nl/min with 2 % B for 25 min. The flow rate was maintained at 200 nl/min during the separation. The gradient is as follows: 0-60 min 2-40 % B, 60-70 min 40-90 % B, 70-75 min isocratic at 90 %B and 75-78 min 2% B. The mass spectrometer was operated in a data dependent mode (DDA) to switch between MS and CIDMS2. Briefly, after a full-scan MS spectrum from m/z 400-2000 in the linear ion-trap, 8 CIDMS2, with 28 % normalized collision energy and activation Q at 25, activation steps were performed on the 8 most intense precursor ions from the full scan.

3.4. Result and Discussion

3.4.1. Part A: Analysis of Interferon (IFN) α2b reference material and Pichia pastoris IFN α2b

3.4.1.1. SDS-PAGE of interferon α2b

Molecular weight determination by SDS-PAGE can provide information about the presence of product variants such as residual leader sequence and proteolytic processing as well as a comparison between the process sample and a reference material. As shown in Figure 3-1 both IFN α2b reference material and Pichia pastoris α2b have the molecular weight of approximately 19 kDa and a single band on SDS-PAGE. For an in-depth comparison of standard and the Pichia pastoris material, LC/MS analysis was then performed on the gel bands (see later).
3.4.1.2. **IEF gel of interferon α2b**

Isoelectric focusing can be used to detect charge variants present in a process development sample such as deamidation and changes in the degree of sialylation. Theoretical isoelectric focusing point (pI) of IFN is 5.99. As shown in Figure 3-2, we compared the observed pI of a *Pichia pastoris* IFN with the reference material. While the IFN α2b reference material showed the expected pI and was consistent with the theoretical value of 5.99 (ExPASy ComputePI/MW), the *Pichia pastoris* sample showed a more acidic main band together with a faint extra acidic band.
Figure 3-2: IEF gel of IFN α2b reference material and *Pichia pastoris* IFN α2b.

Lane 1: IEF standards, Lane 2: IFN α2b reference material, Lane 3: *Pichia pastoris* IFN α2b material.

3.4.1.3. **Trypsin digestion and LC/MS analysis (IFN α2b reference material and *Pichia pastoris* sample)**

3.4.1.3.1. **Determination of amino acid sequence**

The goal of peptide mapping is to ensure that the primary structure of the drug substance is correct with 100% sequence coverage. Sometimes, one type of enzyme digestion will not give complete coverage due to factors such as limited cleavage sites in the protein with the generation of large
peptides not suitable for LC-MS analysis or peptides which are too short and polar to be retained on the HPLC column or too hydrophobic for elution. The peptides identified with trypsin digestion of IFN α2b standard and Pichia pastoris IFN α2b are listed in Table 3-1. In this analysis, tryptic peptide T10 which contains residues 84 to 112 was not identified in the Pichia pastoris IFN analysis but was observed in the reference material. While the protein sequence for approved interferon α2b remains confidential, our LC/MS analysis confirmed the DrugBank (https://www.drugbank.ca/) information which is at variance with the Uniprot (http://www.uniprot.org/) sequence information (N vs. K at residue 112 respectively). We then confirmed that the sequence of the Pichia pastoris material was consistent with the sequence reported in Drugbank by Glu-C digestion which provides a peptide TPLMNE (108-113) with a length suitable for LC-MS analysis and that contains residue 112 as in Figure 3-4 in Pichia pastoris sample and Figure 3-3 in standard material.
Table 3-1: Peptide mapping result of IFN α2b standard using trypsin in-gel digestion.

Several short peptides with two or three amino acids that were too polar to retain on C18 column for LC-MS analysis. These missing amino acids were identified by Glu-C digestion to achieve 100% sequence coverage (data not shown).
Figure 3-3: Confirmation of K112 of standard IFN α2b by GluC digestion.

A: Extracted Ion Chromatogram, B: Monoisotopic mass, C: CID MS2
Figure 3-4: Confirmation of amino acid substitution K112 to N 112 of *Pichia pastoris* IFN α2b by GluC digestion.

A: Extracted Ion Chromatogram, B: Monoisotopic mass, C: CID MS2

3.4.1.3.2. Analysis of residual leader sequence

To examine if there is any residual leader sequence present in the process development samples we combined the amino acid sequences for α factor and IFN α2b into a single FASTA file. The
trypsin digest LC/MS data was then searched against the combined FASTA file using Thermo Biopharma Finder 2.0. To effectively use Biopharma Finder for in-depth characterization of recombinant proteins there are several parameters critical for an effective database search. These parameters include S/N ratio (set at default value 20, unless the identification of very low level of variant is needed), mass accuracy (set as 5 ppm with minimum confidence 0.8 to ensure low level of false positive identifications). With this approach, we determined that “EEGV$\text{SLEKR}$” was still attached to the N-terminus of the *Pichia pastoris* IFN α2b via the observation of the peptide “EEGV$\text{SLEK}$” (see Figure 3-5) as well as “RCDLPQTHSLGSR” (see Figure 3-7, Figure 3-8 and Figure 3-9) but was not present in the reference material.

**Figure 3-5: Extracted Ion Chromatogram (XIC) of “EEGV$\text{SLEK}$”**.
Upper panel: XIC of leader sequence peptide in IFN standard, lower panel: XIC of leader sequence peptide in *Pichia pastoris* IFN.
Figure 3-6: Monoisotopic mass and CIDMS2 of EEGVSLEK in *Pichia pastoris* sample.
Figure 3-7: Extracted Ion Chromatogram (XIC) of RCDLPQTHSLGSR.

Both upper panel and lower panel were for *Pichia pastoris* IFN.
Figure 3-8: Monoisotopic mass of leader sequence “RCDLPQTHSLGSR” and “CDLPQTHSLGSR”
Figure 3-9: CID MS2 of leader sequence peptide “RCDLPQTHSLGSR” in *Pichia pastoris* sample.
3.4.1.3.3. Oxidation analysis

The oxidation level comparison of reference IFN α2b and *Pichia pastoris* material was shown in Table 3-2. The percentage of oxidation was calculated using the following equation: peak area of oxidized peak / (peak area of oxidized peak + peak area of unoxidized peak) x 100%. Except the Met16 in *Pichia pastoris* material, which is a little higher than the standard material, other four methionine residues have a similar oxidation level to the standard. The low level of oxidation observed with the reference material indicates that the in-solution digestion protocol induced little artificial oxidation.

<table>
<thead>
<tr>
<th>Site</th>
<th>Standard</th>
<th><em>Pichia</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Met 16</td>
<td>2.3%</td>
<td>6.9%</td>
</tr>
<tr>
<td>Met 21</td>
<td>2.9%</td>
<td>1.9%</td>
</tr>
<tr>
<td>Met 59</td>
<td>3.6%</td>
<td>5.8%</td>
</tr>
<tr>
<td>Met 111</td>
<td>1.0%</td>
<td>2.6%</td>
</tr>
<tr>
<td>Met 159</td>
<td>2.7%</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

Table 3-2: Oxidation analysis of IFN α2b reference material and *Pichia pastoris* material

3.4.2. Part B: Analysis of Interferon (IFN) γ reference material (Abchem) and *Pichia pastoris* IFN γ

3.4.2.1. In-gel digestion of IFN γ SDS-PAGE gel bands with trypsin: N-glycosylation of *Pichia pastoris* IFN γ
As shown in the gel image Figure 3-10, there were two bands observed for *Pichia pastoris* IFN fermentation supernatant (MW 17KDa and 19KDa) and one band for the reference material (Abcam) at 15 KDa. LC-MS analysis of the in-gel trypsin digests of these bands was performed and raw data was searched against the amino acid sequence of IFN γ together with α factor (BioPharma Finder 2.0). As in Table 3-3, for the 15KDa reference material band, except a short peptide that has “KR” repeats, all of the sequence was identified. While in the 17KDa *Pichia pastoris* IFN sample, all tryptic peptides were identified except the peptide containing residues 91-108 “DDFEKLTNYSVTDLNVQR” and the C-terminal peptide 127-144 “TGKRKRSGMLFRGRRASQ”. In the 19 kDa *Pichia pastoris* IFN γ band, three peptides 15-35 “YFNAGHSDVADNGTLFLGILK”, 91-108 “DDFEKLTNYS VTDLNVQR” and C-terminal peptide “TGKRKRSGMLFRGRRASQ” were observed. After inspecting the amino acid sequence of these two missing peptides, we found each peptide contains an N-glycosylation motif N-X-S/T, N98 (NYS) and N26 (NGT).

**Figure 3-10:** SDS-PAGE of IFN γ strain supernatant.
Lane 1: Reference R&D material from Abcam. Lane 2-10: Strain supernatant obtained from different cultivation conditions. Lane 14-15 were replicates and used for in-gel digestion. The bands in lane 14 was used for trypsin digestion and lane 15 was for Glu-C digestion.

<table>
<thead>
<tr>
<th>#</th>
<th>AA#</th>
<th>Sequence</th>
<th>Monoisotopic m/z</th>
<th>charge</th>
<th>RT(min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-14</td>
<td>MQDPYVKEAENLKK</td>
<td>570.292</td>
<td>3</td>
<td>24.33</td>
</tr>
<tr>
<td>2</td>
<td>15-35</td>
<td>YFNAGHSDVADNGTLFLGILK</td>
<td>1127.566</td>
<td>2</td>
<td>42.83</td>
</tr>
<tr>
<td>3</td>
<td>36-44</td>
<td>NWKEEsdRK</td>
<td>397.864</td>
<td>3</td>
<td>17.18</td>
</tr>
<tr>
<td>4</td>
<td>45-53</td>
<td>IMQSQIVSF</td>
<td>598.822</td>
<td>2</td>
<td>30.24</td>
</tr>
<tr>
<td>5</td>
<td>54-56</td>
<td>YFK</td>
<td>457.245</td>
<td>1</td>
<td>16.52</td>
</tr>
<tr>
<td>6</td>
<td>57-69</td>
<td>LFKNFKDDQSIQK</td>
<td>537.625</td>
<td>3</td>
<td>23.25</td>
</tr>
<tr>
<td>7</td>
<td>70-81</td>
<td>SVETIKEDMNVK</td>
<td>704.855</td>
<td>2</td>
<td>19.46</td>
</tr>
<tr>
<td>8</td>
<td>82-88</td>
<td>FFNSNNKK</td>
<td>443.228</td>
<td>2</td>
<td>18.20</td>
</tr>
<tr>
<td>9</td>
<td>89-95</td>
<td>KRDDFEK</td>
<td>313.163</td>
<td>3</td>
<td>16.45</td>
</tr>
<tr>
<td>10</td>
<td>96-108</td>
<td>LTNYSVTDLNVQR</td>
<td>761.898</td>
<td>2</td>
<td>29.53</td>
</tr>
<tr>
<td>11</td>
<td>109-118</td>
<td>KAIHELIQVM</td>
<td>591.340</td>
<td>2</td>
<td>31.10</td>
</tr>
<tr>
<td>12</td>
<td>119-126</td>
<td>AESPAAK</td>
<td>786.436</td>
<td>1</td>
<td>17.57</td>
</tr>
<tr>
<td>13</td>
<td>127-144</td>
<td>TGKRKRSQMLFRGRASQ</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

Table 3-3: Peptide mapping of IFN \( \gamma \) reference material with trypsin digestion.

Except 127-132 “TGKRKR” that has the KRKR repeats, all the peptides were covered by trypsin digestion.

In the 17 KDa band, N98 was fully glycosylated as shown in Figure 3-11 and N26 was not glycosylated. In the 19 KDa band, both N 98 Figure 3-12 and N26 Figure 3-13 were fully glycosylated. N26 had mannosylation from Man9 extending to Man16 with Man 11 the highest glycoform. N98 contained mannosylation from Man6 to Man12 with Man9 the highest abundant
glycoform. The structure of the glycan was determined by the observation of the fragmentation ions in the CID-MS2 spectrum. Mannosylation was also the reason for the molecular shift on SDS-PAGE: the mass of glycan GalNAc2Man9 is around 1865 Da, which counts for the ~2KDa shift for 17KDa band. The mass of glycan GalNAc2Man11 is around 2188 Da, which counts for the additional 2KDa mass shift for 19KDa band.

Figure 3-11: Glycosylation of Asn98 at 17kDa band.
Figure 3-12: Glycosylation of Asn98 at 19kDa band.
Oxidation and deamidation

The analysis of crude IFN γ process development sample requires an in-gel digestion analysis which can induce high levels of degradation products during the extensive sample preparation steps. The IFN γ amino acid sequence contains two deamidation hot-spots, NG of N26 and NS of N84. In the 17KDa band, the deamidation level of N26 is >90% and N84 is 20%. In the 19KDa
band, as N26 is fully glycosylated, no deamidation was observed in that site. The observed amounts of oxidation and deamidation level are summarized in Table 3-4.

<table>
<thead>
<tr>
<th>Sites</th>
<th>Modification</th>
<th>Reference</th>
<th>Pichia 17KDa</th>
<th>Pichia 19KDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met1</td>
<td>Oxidation</td>
<td>76.6%</td>
<td>87.8%</td>
<td>72.0%</td>
</tr>
<tr>
<td>Met46</td>
<td>Oxidation</td>
<td>81.0%</td>
<td>85.4%</td>
<td>77.2%</td>
</tr>
<tr>
<td>Met78</td>
<td>Oxidation</td>
<td>67.6%</td>
<td>72.4%</td>
<td>72.6%</td>
</tr>
<tr>
<td>Met118</td>
<td>Oxidation</td>
<td>22.2%</td>
<td>38.5%</td>
<td>21.7%</td>
</tr>
<tr>
<td>Asn26</td>
<td>Deamidation</td>
<td>&gt;90%</td>
<td>&gt;90%</td>
<td>NA</td>
</tr>
<tr>
<td>Asn84</td>
<td>Deamidation</td>
<td>5.2%</td>
<td>22.7%</td>
<td>6.5%</td>
</tr>
</tbody>
</table>

Table 3-4: Oxidation analysis of IFN γ reference material and *Pichia pastoris* material.

3.4.2.3. **In-gel digestion of IFN γ SDS-PAGE gel bands with Glu-C: C-terminal proteolysis**

Glu-C digestion was performed on the two bands of lane 15 in the SDS-PAGE gel as well as the reference material (Lane 1). As there are only a few glutamic acid residues (“E”) in the IFN γ sequence, the observed sequence coverage by Glu-C digestion was low. However, residues 121-129 “LSPAAKTGK” was identified in the *Pichia pastoris* IFN γ but not the reference material indicating the proteolysis had occurred at residue 129 (K) in the *Pichia pastoris* material as in Figure 3-14.
3.5. Discussion

Characterization of process development material at an early stage is a challenging yet important task because early analytical information can accelerate development projects. The analytical challenges include low concentrations of product in fermentation samples, high level of interferences caused by media components and host cell proteins (HCPs) as well as unknown
modifications in the product molecule. Recently, mass spectrometry based methods have enjoyed a dramatic growth in popularity, as they offer a variety approaches to efficiently gain diverse information on product quality. For biosimilar development, there is both the need for efficient process development material and the demonstration of a high degree of similarity between the innovator and the follow-on product. Suitable analytical methods need to be developed with a standard or reference material and validated with stability samples. For the approval process, it is necessary to characterize the protein amino acid sequence at a 100% level as well as post-translational modifications and degradation products. Since any structural changes can have effects on drug potency and have safety effects such as generation of immune responses in a patient, it is necessary to have analytical methods for complete product characterization.

As discussed previously, *Pichia pastoris* has gained wide application for protein production because of its unique features such as low cost, high growth rate, usage of defined media and low level of secretion of HCPs with the reporting of several successful cases of recombinant therapeutic protein production in yeast *Pichia pastoris*. Common characteristics of protein variants expressed in *Pichia pastoris* include residual leader sequence when using α factor signal peptide and high mannose N-glycosylation structures. The complex glycosylation patterns can impede the pharmaceutical development process and hyper mannosylation and terminal α-1, 3-mannose linkages can result in rapid in vivo clearance as well immunogenicity of the recombinant protein in clinical studies.

Since interferons α2b and γ have distinctly different amino acid sequences and contain a range of product variants that occur at various sites. Thus, multiple analytical strategies are required for the in-depth characterization of all the variants. As summarized in Table 3-5 and Table 3-6, the variants of these two IFN product requires different analytical methods and it is important to
optimize the enzyme selection to generate peptides containing the variant sites that are appropriate for LC-MS analysis. In the case of the IFNγ process development, the need for the analysis of crude fermentation samples adds an additional analytical challenge. In this case, we demonstrated that in-gel digestion rather than in-solution digestion is the method of choice. The SDS-PAGE approach used in this analysis has the following additional advantages: removal of any detergents present in the sample that can be detrimental to mass spectrometric analysis, the target protein can be separated from most HCPs, proteolysis and covalent aggregation can be observed with reduced and non-reduced gels ahead of MS analysis. One drawback of the in-gel digestion method is the high level of oxidation artifact that can be induced during the sample preparation process. On the other hand, in-solution enzyme digestion which is suitable for purified material as the case of IFN α2b can greatly reduce the level of artifacts.

<table>
<thead>
<tr>
<th>residue</th>
<th>variants</th>
<th>peptides</th>
<th>enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEGVSLEKR</td>
<td>Leader sequence</td>
<td>EEGVSLEK and RCDLPQTHSLGSR</td>
<td>trypsin</td>
</tr>
<tr>
<td>112 N</td>
<td>Amino acid substitution</td>
<td>118-123: TPLMNE</td>
<td>Glu-C</td>
</tr>
<tr>
<td>16 Met</td>
<td>Oxidation</td>
<td>14-22: TLMLLAQMR</td>
<td>trypsin</td>
</tr>
<tr>
<td>21 Met</td>
<td>Oxidation</td>
<td>14-22: TLMLLAQMR</td>
<td>trypsin</td>
</tr>
<tr>
<td>59 Met</td>
<td>Oxidation</td>
<td>50-59: AETIPVLHEM</td>
<td>trypsin</td>
</tr>
<tr>
<td>111 Met</td>
<td>Oxidation</td>
<td>116-123: TPLMNE</td>
<td>Glu-C</td>
</tr>
<tr>
<td>159 Met</td>
<td>Oxidation</td>
<td>145-149: AEMIR</td>
<td>trypsin</td>
</tr>
</tbody>
</table>

**Table 3-5: Variants identified for Pichia pastoris IFN α2b**
Table 3-6: Variants identified for *Pichia pastoris* IFN γ

Trypsin digestion of IFN α2b generated a set of peptides suitable for LC/MS analysis which allowed achievement of a high sequence coverage as well as the identification of residual N-terminal leader sequence despite the presence of several low MW peptides. For the analysis of the sequence variant with an amino acid substitution of 112K to 112N, trypsin digestion generates a peptide (84-120 “FYTELQYQLNDLEASECAVGIQGVGTTEPTPLMNEDSILAVR”) of a MW and hydrophobicity that was unsuitable for LC-MS analysis. On the other hand, Glu-C digestion generated the peptide “TPLMNE” in *Pichia pastoris* material and “TPLMKE” in reference material which allowed identification of the sequence variant.

The reference material of IFNγ had 96% sequence coverage upon trypsin digestion, however the C-terminal peptide (127-144 TGKRKRSQMLFRGRRASQ) was not identified in the *Pichia pastoris* material. Glu-C digestion lead to a low sequence coverage due to limited Glu-C cleavage sites in the sequence. However, this enzyme was useful for the detection of proteolytic processing in the two *Pichia pastoris* IFN γ bands in the SDS-PAGE gel (see Figure 3-1) by the observation
of the peptide 121-129 “LSPAAKTGK” in the *Pichia pastoris* sample but not in the reference material. Residue K129 is part of the sequence “KRKR”, which is the motif recognized by the *Pichia pastoris* Kex2 protease, the enzyme responsible for the removal of α factor leader peptide. Thus, one possibility for the C-terminal cleavage of the *Pichia pastoris* IFN γ material is that this protease also recognized the “KR” in the protein sequence and cleaved after K129. As expected for *Pichia pastoris* expression, mannosylation was identified in the two sites that contain the N-glycosylation motif N-X-S/T, N98 (NYS) and N26 (NGT). Tryptic peptides 15-35 “YFNAGHSDVADNGTLFLGILK” and 96-108 “LTNYSVTDLNVQR” were used for determination of the glycosylation sites and the corresponding glycoforms. In this analysis site N26 was observed to contain mannosylation from Man9 extending to Man16 (Man 11 most abundant), while N98 contained mannosylation from Man6 to Man12 (Man9 most abundant).

The above HPLC/MS analysis with our multi-enzyme digest strategy was developed for the characterization of IFNs α2b and γ in process development samples. For *Pichia pastoris* IFN α2b the major variants that were observed were an amino acid substitution and residual leader sequence. Amino acid substitution can occur by clonal variation in the host cell line \(^{20}\) or the use of a different cDNA sequence used in cell line construction. \(^{21}\) Such changes may obviously have effects on protein structure and function \(^{22}\) and may be detected via a complete determination of the amino acid sequence, but in the case of a charged to neutral amino acid substitution such a variant can be readily screened by isoelectric focusing (IEF). In this example, the variant was a substitution of lysine to asparagine at residue (112 K (pKa10.6, reference material) to N (neutral, *Pichia pastoris*) which results in the acidic shift observed in the observed pI in the IEF gel (see Figure 3-2). In this manner, an IEF gel can be used as a quick assessment of product variants with pKa differences that can be used to assess process development samples.
Characterization of process development material is a good implementation of Quality-by-design (QbD) strategy for biomanufacturing. Every aspect of the proteins including amino acid sequence, post-translational modifications (PTMs) is related to the characteristics of a process, including strain construction, cultivation conditions and downstream purification. This the quality of early stage process development material reflects the efficacy of the manufacturing platform and an early stage analytical evaluation is critical for host system selection, media optimization and cultivation conditions. For example, a high level of oxidation in the product may be due to high oxygen density in bioreactor, proteolysis characterization is beneficial for protease inhibitor selection, aggregation characterization is useful for the optimization of pH and the addition of detergents to the formulation buffer.

3.6. Conclusion

A multi-enzyme digestion approach was successfully developed for the characterization of interferon products expressed from *Pichia pastoris*. For IFN α2b, trypsin digestion was used initially for peptide mapping, identification of leader sequence and amino acid substitution. Glu-C digestion also provided complementary data for the characterization of an amino acid substitution. An IEF gel method followed by LC/MS analysis was developed as a quick way to monitor the acidic shift of pI due to an amino acid substitution.

In-gel digestion with trypsin or Glu-C was developed for the characterization of a crude, unpurified IFN γ strain supernatant. The material was proved to have mannosylation at the two N-glycosylation sites and C-terminal proteolysis. Such information could be extremely helpful especially for process development at early stage where prompt analytical feedback can help optimize strain development and fermentation conditions.
3.7. Reference


Chapter 4: Complementary top-down and bottom up approach for the characterization of G-CSF variants produced from *Pichia pastoris*

This chapter is based on a manuscript with the title of

“Complementary top-down and bottom up approach for the characterization of G-CSF variants produced from *Pichia pastoris*”


Yu (Annie) Wang: perform most experiments and data analysis, write manuscript

Di Wu: assist with sample preparation

Laura Crowell: produce *Pichia pastoris* material

Shiaw-lin Wu, Kerry Love, William Hancock: supervisors
4.1. Abstract

Granulocyte-colony stimulating factor (G-CSF) is one type of hematopoietins and cytokines, which plays an important role in inducing and regulating the maturation of neutrophils, proliferation and differentiation of neutrophil colonies. With the first Biosimilar G-CSF (Zarxio) approved in US, and the expiration of patent protection for Amgen G-CSF (Neupogen) the development of G-CSF biosimilars has gained broad popularity in biopharmaceutical companies. Thus, advanced analytical methods are required for the characterization of G-CSF to help accelerate the process development. Herein, we present a complementary bottom-up and top-down approach for the profiling of product variants of G-CSF expressed in *Pichia pastoris*. The propensity to aggregation due to the free cysteine, complete sequence mapping, oxidation, deamidation, O-glycosylation and N-terminal proteolysis were studied. The observed weak points and modification profile of the biotherapeutic could serve as important feedback and help with both upstream strain development and downstream purification process.
4.2. **Introduction**

4.2.1. **Recombinant Granulocyte-colony stimulating factor**

Granulocyte-colony stimulating factor (G-CSF) is one example of the class of growth factors called hematopoietins and plays an important role in inducing and regulating the maturation of neutrophils and proliferation and differentiation of neutrophil colonies.\(^1\) As the result of the advancement of recombinant DNA technology, DNA-derived recombinant therapeutic proteins have been broadly applied to medical treatments. Clinical studies have shown positive results with recombinant G-CSF (rG-CSF) decreasing the period of leukopenia, which occurs when neutrophil counts fall below the normal range. As the original patent protection for Amgen G-CSF (filgrastim) produced in *E. coli* has expired, there is a growing interest in exploring new methods to facilitate the production of biosimilar versions. Zarzio (Sandoz filgrastim) is the first biosimilar approved in US.\(^2\-4\)

4.2.2. **Free cysteine in G-CSF induces aggregation at high pH**

G-CSF has intrinsic instability due to a free cysteine at residue 17 which leads to a propensity to aggregation due to disulfide scrambling.\(^5\) Aggregation has been a major problem for the development of protein pharmaceuticals as the process of aggregation may cause the generation of oligomeric species which can be highly immunogenic.\(^6\)

4.2.1. **O-Glycosylation of G-CSF at Thr 134**

Human G-CSF is O-glycosylated at a threonine residue (133), and structure of the sugar chain was published in 1988 as NeuAcα2-3Galβ1-3(NeuAcα2-6)GalNAcol.\(^7\) Currently, there are two commercial forms of recombinant G-CSF and the first was non-glycosylated form as a result of
production using in *E coli*. It has the same amino acid sequence as human G-CSF except for an addition methionine residue at the N-terminus. A significant difference with the E coli product is the lack of O-glycosylation, which in the human version may improve stability and protect from proteolysis *in vivo*. Nevertheless, the potency of non-glycosylated G-CSF was proven to be clinically equivalent with natural G-CSF. The second commercial form of G-CSF was expressed in Chinese hamster ovarian (CHO) cell lines. In addition to containing the same protein sequence this form of rG-CSF contained the same O-glycosylation structure at T133 as the natural form. However, production of rG-CSF in CHO cell lines was limited by lower production efficiency as well as the laborious removal of the host cell proteins (HCP) during the downstream process.

4.2.1. **G-CSF Expressed in Pichia pastoris**

Recently, successful expression of G-CSF from *Pichia pastoris* was reported with improved yield using a codon-optimized synthetic gene. As such, *Pichia pastoris* is one of the ideal alternative host systems with advantages of less processing time, cost effectiveness and glycosylation similar to a eukaryotic system. However, no detailed glycosylation profiling was included in these reports. Therefore, characterization of GCSF from *Pichia pastoris*, especially O-glycosylation is required for the development of rG-CSF.

4.2.2. **N-terminal cleavage of G-CSF**

N-terminal cleavages of a G-CSF biosimilar have been previously studied in *E.coli* by a Sandoz development group and several N-terminal proteolysis events were observed, where the variant with N-terminal methionine cleavage was observed to be the most abundant. Additional N-
terminal proteolysis was observed probably because the first 10 amino acids of G-CSF form an unstructured region which makes them more susceptible to cleavage by aminopeptidases.\textsuperscript{12} Herein we present a method using a complementary bottom-up and top-down LC/MS approach for the characterization of G-CSF produced from \textit{Pichia pastoris}. In this study, we observed O-linked mannosylation and N-terminal proteolysis between residues leucine 4 and glycine 5.

\textbf{4.3. Experiment Section}

\textbf{4.3.1. Reagent and chemicals}

Equipment: XCell SureLock Mini-Cell electrophoresis apparatus was from Thermo Fisher Scientific. Bench centrifuge was purchased from Corning. Ultimate 3000 nano LC pump (Dionex, Mountain View, CA). 10 µm ID emitter (New Objective, Woburn, MA), LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA).

Sandoz G-CSF standard was kindly provided by our collaborator, Love Lab from MIT Koch Institute. Ammonium bicarbonate (ABC), dithiothreitol (DTT) and iodacetamide (IAA) was from Sigma Aldrich. Trypsin, pepsin and GluC was purchased from Promega. Guanidine hydrochloride, formic acid Optima LC/MS grade, water LC/MS grade, acetonitrile (ACN), methanol, sulphosalicylic acid and tricloroacetic acid, hydrochloric acid was purchased from Fisher Scientific. NuPAGE Novex 4-12\% Bis-Tris gel, 1.0 mm, 10 well, NuPAGE MES SDS running buffer (20X), NuPAGE LDS Sample Buffer (4X), PageRuler Prestained 10-180KDa Protein Ladder, Invitrogen SimpleBlue SafeStain, Novex pH 3-10 IEF Protein Gels, 1.0 mm, 12 well, Novex pH 3-10 Buffer Kit (includes Novex IEF Anode Buffer 50X, Novex IEF Cathod Buffer pH 3-10 10X and Novex IEF Sample Buffer pH 3-10 2X), Colloidal Blue Staining Kit and premium grade TCEP-HCl was purchased from Thermo Fisher Scientific. IEF standards were from Bio Rad.
4.3.2. **Preparation of a G-CSF stability sample (high temperature and pH stress)**

Sandoz G-CSF standard 1mg/ml was buffer exchanged to 100 mM sodium borate buffer, pH 9.0 by Amicon Ultracentrifugal units (Millipore) with a 10 kDa molecular weight cut off. Then, the samples were incubated at 37 °C for 1 h, 3 h and 24 h, respectively.

4.3.3. **SDS-PAGE gel analysis of G-CSF stability**

500 mL 1X NuPAGE SDS running buffer was prepared by mixing 25 mL MES running buffer with 475 mL deionized water. 10 µl of each G-CSF stressed sample was mixed with 3.3 µl LDS sample buffer. Gel apparatus was set and 5µl protein ladder was loaded as well as the G-CSF samples with one blank lane left between each sample. The voltage was set at 160 V for 50 min. Gel apparatus was disconnected the gel was removed from two plates, stained with SimpleBlue SafeStain for 30min at room temperature and destained in deionized water overnight.

4.3.4. **Pepsin digestion of G-CSF stability sample**

20 µl (1µg/ µl) G-CSF reference material and 20 µl (1µg/ µl) G-CSF stability sample, (stressed under pH 9 and 24h), were buffer exchange to 10 mM HCl (pH 2) using an Amicon 10 kDa ultra centrifugal filter. 2 µl pepsin (1µg/µl) was added to the samples and incubated at 37 C for 30 min. After the digestion, 4 µl 100 mM ammonium bicarbonate was added to half of each sample (11 µl) to quench the reaction. The other half of each sample was then reduced with 1 µl 50mM TCEP.

4.3.5. **Glu-C digestion of G-CSF**

20 µl 6M guanidine hydrochloride (MW 95.53 g/mol) was added in 0.1 M NH4HCO3 (ABC) to 20 µg sample of purified G-CSF. Followed by adding 2 µL of 1 M Dithiothreitol (DTT) solution
to the solution (final concentration: 5mM DTT) and heated at 37°C for 30 minutes. Fresh 1M iodoacetamide (IAA) solution was made in a foil-wrapped tube to avoid exposure to light and was added to the denatured, reduced G-CSF solution to a final concentration of 10mM. The tube was covered with aluminum foil and incubated at room temperature (25 °C) for 1h. After incubation, the solution was buffer exchanged to 100 mM ammonium bicarbonate pH 8 using 10kDa ultra centrifugal filter with speed of 14000 rpm for 6 min. For buffer exchange, 100 mM ammonium bicarbonate solution was added with a final volume 500 µl in the cut-off tube and reduced to 100ul each time. This step was repeated three times to ensure that pH 8 was obtained for the enzyme digestion. A first dose of Glu-C (1 µl at 0.4 µg/µl) was added to the solution and incubated at 37 °C for 4h. Then a second dose of enzyme with same amount was added for incubation overnight (16 h) at room temperature. The digestion was stopped by adding 4 µl 1% formic acid to make the final pH around 3. The peptide solution was aliquoted to 30 µl/vial at -80 °C before LC/MS analysis.

4.3.6. **LC-MS analysis of pepsin and Glu-C digested sample**

An Ultimate 3000 nano LC pump (Dionex, Mountain View, CA) and self-packed C18 column (Magic C18, 200Å pore and 5 µm particle size, 75 µm internal diameter (ID), 10cm long) connected to a coated 10 µm ID emitter (New Objective, Woburn, MA) were coupled online to an LTQ-XL mass spectrometer (Thermo Fisher Scientific, San Jose, CA) through a nanospray ion source (New Objective, Woburn, MA). Mobile phase A was 0.1 % formic acid in HPLC grade water and mobile phase B was 0.1 % formic acid in acetonitrile. The flow rate was 250 nL/min with 2 % B for 25 min. and was maintained at 200 nl/min during the separation. The gradient is as follows: 0-60 min 2-40 % B, 60-70 min 40-90 % B, 70-75 min isocratic at 90 %B and 75-78 min.
2% B. The mass spectrometer was operated in a data dependent mode (DDA) to switch between MS and CIDMS2. Briefly, after a full-scan MS spectrum from m/z 400-2000 in the linear ion-trap, 8 CIDMS2, with 28% normalized collision energy and activation Q at 25, activation steps were performed on the 8 most intense precursor ions from the full scan.

4.3.7. **Intact mass measurement of G-CSF standard and G-CSF produced in *Pichia pastoris***

H-Class Acquity UPLC system and an ACQUITY UPLC protein BEH300 C4 Column, 300Å, 1.7µm, 2.1 mm x 50mm were coupled to a Xevo G2-S Q-ToF mass spectrometer (Waters Corp, Milford, MA). Mobile phase A was 0.1 % formic acid in HPLC grade water and mobile phase B was 0.1 % formic acid in acetonitrile. The flowrate was maintained at 0.2 mL/min, column temperature was set at 60 °C. The gradient was as follows, 0-1 min 10% B, 1-18 min 10%-90% B, 18-23 min, kept at 90% B, 23-24 min 90%-10% B, 24-27 min, kept at 10% B. The Xevo G2-S QTof was set as sensitivity mode, capillary voltage 3.0 kV, sample cone 40 V, cone gas flow 0 L/h, desolvation gas flow 800 L/h, source temperature 150 °C and desolvation temperature 300 °C. The mass range is 400 m/z to 4000 m/z. No fragmentation method was included in this analysis.
4.4. Results

4.4.1. Aggregation of reference G-CSF stressed under high pH and Temperature

To study the propensity to aggregate, the degradation of G-CSF was accelerated by incubation under high pH and high temperatures. The stressed samples at each time point (1h, 3h and 24h) was run on SDS-PAGE to observe the level of aggregation.

As shown in Figure 4-1, the aggregation level of G-CSF increased with the incubation time 1h, 3h and 24h. Only a low level of dimer and trimer was observed for the 1 h and 3 h incubation samples. Upon the treatment with 100mM DTT at 90 °C for 30 min, most of the aggregates were reduced back to monomers, indicating that the aggregation was probably caused by disulfide scrambling and thus of a reducible nature. Pepsin digestion was performed to study the disulfide connectivity of the aggregates. Briefly, the reference material and the G-CSF 24h stability sample were subjected to pepsin digestion at pH 2 with an enzyme to substrate ratio 1:10 (wt/wt) for 30 min at 37 °C. Half of each digested sample was quenched with a pH adjustment (ammonium bicarbonate) and served as nonreduced pepsin digested sample. The other half was reduced with 50 mM TCEP after the digestion and served as the reduced pepsin digested sample. Both the reduced and non-reduced sample were run on an Ultimate 3000 UPLC-LTQ-XL mass spectrometer system coupled with a C18 capillary column (self-packed with 5 µm particles, 75 µm inner diameter, 360 µm outer diameter, 10 cm long). The separation was achieved with a gradient from mobile phase A (0.1 % formic acid) and mobile phase B (0.1 % formic acid/ acetonitrile) with a flow rate of 250 nl/min and the mass spectrometer was operated in a data dependent mode. Figure 4-2 shows the results of the LC/MS analysis of the disulfide connectivity analysis for the G-CSF stability sample. The free cysteine C18 in G-CSF was proven to be the main reason for disulfide scrambling in the G-CSF aggregates with the observation of new disulfide linkages Cys18-Cys18 and Cys18-Cys75.
Figure 4-1: SDS-PAGE of high temp & pH stressed G-CSF.

For each sample, the left and right lanes are nonreduced and reduced conditions (+DTT).
Figure 4-2: Disulfide connectivity confirmation in the G-CSF stressed aggregates.

a, CIDMS2 for the identification of disulfide bond between two C18 cysteine residues. b, CIDMS2 for the identification of disulfide bond between C18 and C75.
4.4.2. Determination of primary structure of G-CSF

Glu-C digestion of both G-CSF reference material and *Pichia pastoris* G-CSF material were performed side by side for comparison. The digestion was performed at pH8, at which Glu-C has optimal activity, with an enzyme to substrate 1:50. After the first dose incubation of 37 °C for 4h, second dose of Glu-C was added to compensate for the low protease efficiency of Glu-C. The digested samples were analyzed by LC/MS using the same conditions as for Part A of the Results section. The identification of each peptide was achieved by matching the fragment ions with theoretical prediction and 100% sequence coverage was achieved for both materials as shown in Table 4-1. Thus, the amino acid sequence of the *Pichia pastoris* G-CSF was proven to be identical to the reference material.
Table 4-1: Peptide mapping of G-CSF with Glu-C digestion.

4.4.3. Oxidation and deamidation analysis

To study the oxidation and deamidation of G-CSF, an in-solution Glu-C digestion was performed. The digestion conditions as well as the LC/MS analysis were the same as described in Part B of the Results section. The identification of each peptide was achieved by matching the fragment ions with theoretical prediction. Each peptide containing methionine as well as its oxidized counterpart were extracted from the LC-MS raw data to determine the oxidation level. Only the N-terminal methionine of the four methionines was observed to be oxidized at the level of 4.0% in the Pichia pastoris sample and 2.3% in the reference material.
Deamidation was determined in the same way as oxidation by extraction of the theoretical m/z values of deamidated peptide in the raw data and no deamidation of the 17 glutamines present in G-CSF was detected.

4.4.4. Characterization of N-terminal cleavage by Glu-C digestion

To study the N-terminal proteolysis of G-CSF, Glu-C digestion was performed in the same condition as in Part B with two additions of Glu-C and incubation at room temperature overnight. The digested samples were run on an Ultimate 3000 UPLC-LTQ-XL mass spectrometer system coupled with a C18 capillary column. All the peptides identified by Glu-C digestion in both reference material and *Pichia pastoris* material were compared. One peptide corresponding to residues G5-E20 “GPASSLPQSFLKLCE” was identified in the *Pichia pastoris* sample but not the reference material indicating a N-terminal cleavage of the first four amino acids, “MTPL” (see Figure 4-3). The cleavage between leucine 4 and glycine 5 was calculated to be 12% using the following equation: peak area of cleaved peptide/ (peak area of cleaved peptide + peak area uncleaved peptide) x 100%.
Figure 4-3: Identification of L4 cleaved N-terminal peptide.

A: extracted ion chromatogram of G5-E20 “GPASSLPQSFLKCLE” B: monoisotopic mass C: CIDMS2

4.4.5. Characterization of O-mannosylation at T134 by Glu-C digestion

To study the O-mannosylation of the threonine at residue 134 (T134), Glu-C digestion was performed in the same condition as in Part B. As the O-glycosylation of G134 was previously reported, the proteolytic peptide that contains residue T134 “LGMAPALQPTQGAMPAFA” was selected. The theoretical monoisotopic mass of this peptide, and the modified peptide containing mono and di-mannosylation was predicted. Figure 4 shows extracted ion chromatograms (XIC) of
the peptides corresponding to these expected mannosylation structures. In addition, the accurate monoisotopic mass of each base peak was listed. The mannosylation with two mannoses were found to be the most abundant. Other threonines residues (T2, T39, T103, T106, T116 and T117) were also studied and no O-glycosylation was detected.

![Graph showing O-mannosylation structures](image)

**Figure 4-4:** Identification of O-mannosylation structures observed at T134 LGMAPALQPTQGAMPFA.

Panels 1-3 are the XIC of intact peptide and with one and two mannosylation structures, Panels 4-6 are show the monoisotopic masses of panels 1-3.

4.4.6. **Intact mass measurement to identify the O-mannosylation at T134 and N-terminal proteolysis at between leucine 4 and glycine 5**
The reference material and *Pichia pastoris* G-CSF were analyzed on a Q-TOF mass spectrometer system coupled with an UPLC system. The separation was carried out on a C4 reversed phase HPLC column which provides a high efficiency separation of proteins (300Å pores and particle size of 1.7µm) and with a gradient from 0.1 % formic acid, 10% acetonitrile to 1 % formic acid, 90% acetonitrile. The mass range used in the Q-TOF analysis was 400 to 4000 m/z, and the intact mass of the G-CSF reference material was measured to be 18796 Da with a small amount of oxidation (+16Da) observed (Figure 4-5). Two peaks were observed in the elution profile for the *Pichia pastoris* G-CSF sample (Figure 4-6). The deconvoluted spectra of the two peaks are shown in Figure 5. The following proteoforms were observed, the full-length sequence with mono, di and tri-mannosylation and the G-CSF variant with N-terminal “MTPL” clipped and with mono, di and tri-mannosylation.

![Figure 4-5: Deconvoluted spectra of G-CSF standard.](image)
Figure 4-6: LC chromatogram of *Pichia pastoris* derived G-CSF
Figure 4-7: Deconvoluted spectra of *Pichia pastoris* G-CSF.

1: Proteoforms that elute at a retention time of 11.3 min. 2: Proteoforms that elute at a retention time of 11.5 min

4.5. Discussion

G-CSF is an important protein drug, which can be used in conjunction with chemotherapy to stimulate bone marrow to produce infection fighting white blood cells, or to stimulate stem cell production.\(^{13}\) As the patent coverage of G-CSF by Amgen has expired, there is a lot of interest for biopharmaceutical companies to develop G-CSF biosimilars, additionally, biosimilars can provide patients more affordable drugs. Under this circumstance, analytical methods need to be developed to ensure the similarity of biosimilars to the innovator drug to ensure the safety, purity and efficacy of the biosimilar. However, the characterization of G-CSF remains challenging in a few aspects: Firstly, the free cysteine in G-CSF makes it tend to form covalent aggregation. The presence of
sulfhydryl group (SH) in cysteines makes it easier to be oxidized to form a covalent disulfide bond, especially at high pH values where the -SH group is ionized to the thiolate ion which promotes disulfide scrambling. Secondly, the limited number of basic residues present in the amino acid sequence of G-CSF prevents the use of trypsin digestion for peptide mapping and thus requires the exploration of less widely used enzymes for peptide mapping studies.

4.5.1. Aggregation of G-CSF stressed under high pH and Temperature

Recombinant human G-CSF was reported to aggregate and precipitate at pH 6.9 and 37 °C. The published mechanism described an altered conformation of G-CSF which was aggregation prone. Under native-like conditions, G-CSF exists an equilibrium with this altered conformational form. As such, the formulation buffer for commercial forms of G-CSF are kept around pH 4.0 to avoid aggregation that occurs at neutral pH (Amgen, pH 4.0 and Sandoz, pH 4.4). In our study we prepared a stability sample by accelerating the degradation of G-CSF by use of both high temperature (37 °C) and a high pH 9. A high level of aggregation was observed on SDS-PAGE at the 24h time point (Figure 4-1). While SDS-PAGE will dissociate non-covalent aggregates the gel analysis showed bands corresponding to dimeric and trimeric material which represented covalent aggregates. The five cysteines present in G-CSF are coupled in two disulfide bonds, C37-C43, C65-C75 and one free cysteine C18. LC-MS analysis of the aggregates confirmed that the free –SH in cysteine residue 18 was indeed involved in the disulfide scrambling which is consistent with a reported study. We observed C18 formed a cysteine bridge linking another C18 residue as well as C18 also formed disulfide bond with C75 with cleavage of the C65-C75 bridge (see Figure 4-2).

4.5.2. Determination of primary structure of G-CSF
Trypsin is typically the enzyme of choice for peptide mapping of proteins owing to its high specificity for lysine and arginine residues, wide availability and easy to use. However, trypsin is not suitable for obtaining full sequence coverage of G-CSF where the amino acid sequence only contains a few trypsin cleavage sites. Thus, many of the peptides generated in a trypsin digest have either a high or very low MW and contain too many hydrophobic residues or are too polar for effective reversed phase HPLC/MS analysis. For example, sequence 42-147 “LCHPEELVLLGHLGIPWAPLSSCPSQALQLAGCLSQHSLFLYQGLLQALEGISPELG PTLDLQVDVADFATTIWQQMEELGMPALQTQGAMPFAFSQF” is one such trypsin product and such a long peptide is clearly not suitable for LC-MS analysis. Glu-C, on the other hand, proved to be a good alternative to trypsin, as there is more Glu-C digestion sites (aspartic and glutamic acids) in the G-CSF sequence. In this manner 100% sequence coverage of G-CSF was achieved by digestion with this enzyme and LC/MS analysis of the resulting digest (see Table 4-1) and is consistent with studies of G-CSF from other cell lines. As described earlier the determination of 100% sequence coverage is the primary initial step in a biosimilar development process to ensure the similarity between process development material and reference material.

4.5.3. **Oxidation and deamidation analysis**

The Glu-C digested peptides were also suitable for analysis of degradation products such as oxidation and deamidation. In this analysis only the N-terminal methionine was observed to be oxidized among the four methionines present in G-CSF and the oxidation level was 4.0% in the *Pichia pastoris* sample, which was comparable to that of the reference material, 2.3%. As the first ten amino acids in G-CSF form a flexible region, which could result in the N-terminal methionine being the most easily oxidized among the four methionines. Glutamine deamidation was reported
to be hundred times slower than asparagine due to the steric hindrance in forming the intermediate cyclic structure. Thus, even though there are 17 glutamines in G-CSF sequence, none was observed to be deamidated.

4.5.4. Characterization of N-terminal cleavage by Glu-C digestion

Glu-C digestion confirmed the occurrence of N-terminal proteolysis between residues leucine 4 and glycine 5 by observing the peptide corresponding to GPASSLPQSFLKCLE (G5-E20) in the *Pichia pastoris* sample but not reference material. The cleavage site is different from that of the Sandoz G-CCSF and one reason may be the different host system. Thus N-terminal proteolysis is a cause of major N-terminal heterogenicity and may lead to effects on protein purity, efficacy and safety. Although it remains unclear the clinical relevance of this cleavage, the generation of such a variant will need to be paid special attention during downstream process development to remove any cleaved product.

4.5.5. Characterization of O-mannosylation at T134 by Glu-C digestion

Glycosylation in *Pichia pastoris* was reported to commonly have high mannose structures as an inherent characteristic of proteins produced in this organism. Thus, it was necessary to determine the glycosylation status and the structure of any glycoforms attached to the threonine residue 134 (T134) of G-CSF produced in *Pichia pastoris*. Upon Glu-C digestion, the peptide with residues 125-142 “LGMAPALQPTGAMPAPA” which contained the potential glycosylation site was isolated and MS analysis identified the presence of glycosylation and that the major O-mannosylation structure contained two mannoses. The characterization of O-linked glycosylation remains challenging for several reasons, including no consensus sequence for O-glycosylation
sites in proteins, complexity and heterogeneity of the glycan structures, degree of occupancy, and the lack of an appropriate enzyme to release O-glycans from proteins.  

4.5.6. **Intact mass measurement to identify the O-mannosylation at T134 and N-terminal proteolysis at between leucine 4 and glycine 5**

The identification of N-terminal cleavage was achieved by comparing the peptides generated in the standard Glu-C digest and *Pichia pastoris* material Glu-C digest. This procedure is quite time consuming and laborious. Top-down, on the other hand, can be used to monitor all the proteoforms in one scan. In our case, UHPLC separation followed by Q-TOF analysis identified the correct full-length sequence of G-CSF as well as a N-terminally clipped variant (desMTPL) and both structures contained mono, di and tri mannosylation as the major structures. Nevertheless, without the information of bottom-up analysis, it is quite challenging to decipher identity of the six masses observed in the top-down experiment. Also, due to the detection limits and low oxidation level, oxidation was not observed by the top-down approach.

**4.6. Conclusion**

Propensity to aggregation of G-CSF has always been a big concern and aggregation is accelerated at higher pH values. To monitor the aggregation of early stage process development material is an important procedure to access product quality. Using our optimized enzyme digestion method, O-linked glycosylation was identified and quantified. Mannosylation with two mannose molecules was the highest abundance form. In addition, this method was also able to confirm a N-terminal cleavage of residues MTPL. The top-down method has the advantage of detecting all the proteoforms together. The three different mannosylation levels and N-terminal cleavage was
detected in one run. To conclude, bottom-up and top-down approach can provide a comprehensive and global perspective of the characterization of G-CSF. The full sequence coverage, similar oxidation level, low level of N-terminal cleavage demonstrates that the *Pichia pastoris* material is comparable to the reference material with the major difference being the presence of O-linked mannosylation.

4.7. References


characterization of O-glycan structures on alpha-dystroglycan isolated from rabbit skeletal muscle.

Chapter 5: Conclusions

Biopharmaceuticals such as proteins, peptides and monoclonal antibodies are becoming an important class of today’s medicine. Their complex nature makes it impossible to produce identical products from batch to batch. The heterogeneity of biopharmaceuticals mainly caused by the production in biological systems might result in both structural and functional consequences, thus comprehensive characterization of biopharmaceuticals is necessary to ensure the product quality, safety and efficacy.

Characterization of proteins produced by a novel production system such as Biomedicine on Demand (Bio-MOD) stimulates the development of new analytical methods due to the challenges such as low concentration, high level of contaminants and various product variants of process development samples. In addition, the small scale, high throughput features of such platforms require the complete characterization of the protein product with a minimal available sample amount.

Liquid-chromatography coupled with mass spectrometry (LC-MS) is a powerful tool for protein characterization. With the advancement of instrumentation, the sensitivity and resolution of mass spectrometers are both being improved. The combination of the bottom-up and top-down approaches illustrated in this thesis fit the need of comprehensive characterization of products with only small amounts of sample. This approach enables detailed protein characterization including complete sequence mapping, various product variants such as oxidation, deamidation, glycosylation, N-terminal truncation and residual leader sequence.

With the approval of biosimilars, huge savings to healthcare system as well as increase patient access can be expected. To ensure the safety of biosimilars, an identical amino acid sequence, comparable level of post-translational modifications (PTMs) to the reference material and
marketed product need to be confirmed with advanced analytical methods and constitute an integral part of the FDA approval process.
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

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

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

If credit is given to another source for the material you requested, permission must be obtained from that source.
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

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

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

If credit is given to another source for the material you requested, permission must be obtained from that source.
PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

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

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

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

Dear Miss. Yu Wang,

Thank you for placing your order through Copyright Clearance Center’s RightsLink® service.

Order Summary

Order Date:     Apr 21, 2017  
Order Number:  4093750526391 
Publication:  Biotechnology & Bioengineering  
Title:  Integrated continuous production of recombinant therapeutic proteins  
Type of Use:  Dissertation/Thesis  
Order Total:  0.00 USD

View or print complete details of your order and the publisher’s terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this survey to let us know.

Tel: +1-855-239-3415 / +1-978-646-2777  
customercare@copyright.com  
https://myaccount.copyright.com
Thank you for your order!

Dear Miss. Yu Wang,

Thank you for placing your order through Copyright Clearance Center’s RightsLink® service.

Order Summary

Order Date: Apr 21, 2017
Order Number: 4093730456953
Publication: Biologicals
Title: Determination of critical quality attributes for monoclonal antibodies using quality by design principles
Type of Use: reuse in a thesis/dissertation
Order Total: 0.00 USD

View or print complete details of your order and the publisher’s terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this survey to let us know.
Thank you for your order!

Dear Miss. Yu Wang,

Thank you for placing your order through Copyright Clearance Center’s RightsLink® service.

Order Summary

Order Date: Apr 21, 2017
Order Number: 4093720520905
Publication: Biotechnology Progress
Title: Quality attributes of recombinant therapeutic proteins: An assessment of impact on safety and efficacy as part of a quality by design development approach
Type of Use: Dissertation/Thesis
Order Total: 0.00 USD

View or print complete details of your order and the publisher’s terms and conditions.

Sincerely,

Copyright Clearance Center

How was your experience? Fill out this survey to let us know.

Tel: +1-855-239-3416 / +1-973-646-2777
customercare@copyright.com
https://myaccount.copyright.com