Development of Methods for the Analysis of Protein Post-translational Modifications: IsoAspartic Acid and Protein Crosslinking

by Min Liu

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

September 11, 2014

Dissertation directed by

Zhaohui Sunny Zhou
Professor of Chemistry and Chemical Biology
Da Ren
Amgen, Inc
Acknowledgments

My journey on the path of the fascinating world of protein analysis began with my advisors, Professor Zhaohui Sunny Zhou and Dr Da Ren. I would like to take immense pleasure in thanking them for the opportunity on two challenging and interesting projects. Their passion and high standard for science will inspire me for my life and whole career. Their abundant guidance, valuable suggestion, endless encouragement and patience are greatly appreciated. They have made me overcome my frustration and have made the past five years truly enjoyable and memorable for me.

I would like to express my deepest gratitude to Dr Janet Cheetham at Amgen for her endless support, many advice and encouragement. Without her support, it is impossible for me to transfer work experience from small molecule to bio-large molecules. Without her encouragement, I could not get the work done and reported here.

I would like to express my great appreciation to Dr Zhongqi Zhang at Amgen for his generous help, strong scientific guidance, and very helpful discussion. Sharing his expertise, he is my great mentor in protein characterization by mass spectrometry.

My special thanks go to my committee members: Professors Paul Vouros and William Hancock for reviewing my thesis and the valuable suggestions.

I particularly thank my coworkers and lab mates: Chris Spahr, Drs David Brems, Judy Ostovic, Nina Cauchon, Aleksander Swietlow, Peter Zhou, Wenqin Ni, Tianzhu Zang, Eddie Zhou, Kevin Moulton, Aldina Mesic, Wanlu Qu, Nathaniel Kenton and Tianyi Bai for their help and scientific discussion. I would like to especially thank Professor Richard Duclos and Kalli Catcott for their critical review and helpful suggestions on my manuscript preparations. I also
thank Drs Bin Ma, Dan Maloney, and Cassandra Wigmore at Bioinformatics Solutions for helpful discussion on de novo sequencing.

Finally and most importantly, I would like to dedicate this thesis to my family: my husband Chester Yuan, son Michael Yuan and daughter Emily Yuan. I am sincerely grateful for their generous understanding, endless love and selfless support. Their love, believing in me and encouragement made me completing this five-year endeavor in my career and life.
Abstract

Analysis of protein posttranslational modifications (PTMs) plays pivotal roles for the understanding of their biological importance. Isoaspartic acid (isoAsp) as the smallest PTM is observed in vivo and in vitro. No mass difference and subtle difference in physiochemical property between isoAsp and Asp pose a great challenging for sensitive detection and ambiguous location of isoAsp site in complex samples. A novel assay of isoAsp by exploiting methylation specificity of protein isoaspartate methyltransferase (PIMT) at isoAsp and subsequent $^{18}\text{O}$-incorporation during methyl ester hydrolysis is presented for sensitive detection and unambiguous site location of several isoAsp residues in IgG1 (Anal Chem 2012, 84, 1056-1062). The method can be applied to biological samples to understand the isoAsp process and identify biomarkers.

Ubiquitous protein crosslinks in biological systems and biopharmaceuticals are reported to result in loss of bioactivity and immunogenicity, but their characterization is poor, especially when the crosslink chemistry is undefined, due to their intrinsic structural complexity and a lack of a systematic analytical approach. A comprehensive methodology, XChem-Finder, has been developed to break down the analytical challenge via $^{18}\text{O}$ labeling and mass spectrometry, leading to the discovery of a total of 14 cross-linked thioether peptides in IgG2, including those that have not been previously reported (Anal Chem 2013, 85, 5900-5908). Furthermore, a novel Histidine-Histidine (His-His) crosslink in IgG1 was successfully discovered and characterized via our XChem-Finder (Anal Chem 2014, 86, 4940-4948). This again demonstrates the broad applicability and utility of our XChem-Finder. The further improvement of XChem-Finder is discussed. The discovery of more novel crosslinks in protein by XChem-Finder will be successful without any doubt.
# Table of Contents

Acknowledgements ........................................................................................................... ii

Abstract ................................................................................................................................. iv

Table of Contents ...................................................................................................................... v

List of Figures ........................................................................................................................... xi

List of Tables ............................................................................................................................ xvi

List of Schemes ......................................................................................................................... xviii

Abbreviations and Symbols ..................................................................................................... xix

Chapter 1: Overview of Protein Post-translational Modification and Their Analysis .......... 1

1.1 Protein Post-translational Modification ........................................................................ 1

1.2 Biological and Biopharmaceutical Importance of PTMs ............................................. 1

1.3 Analysis of PTMs .......................................................................................................... 2

1.4 Deamidation and Isomerization .................................................................................... 7

1.4.1 Isoaspartic Acid Formation ....................................................................................... 7

1.4.2 Biological and Biopharmaceutical Importance of Deamidation and Isomerization .... 11

1.4.3 Methods for Detection and Characterization of Deamidation and Isomerization ...... 12

1.4.3.1 Edman Degradation .............................................................................................. 12

1.4.3.2 Protein-L-isoaspartyl Methyltransferase (PIMT) ................................................... 13

1.4.3.3 Mass Spectrometry ............................................................................................... 14

1.4.3.4 $^{18}$O-incorporation ............................................................................................... 17
3.2 Introduction .................................................................................................................. 89

3.3 Experimental Section .................................................................................................. 91

3.3.1 Chemicals ............................................................................................................... 91

3.3.2 Generation of Stressed Sample ............................................................................... 94

3.3.3 Reduction, Alkylation, Tryptic Digestion and $^{18}$O-Labeling of the IgG2 .......... 94

3.3.4 HPLC ..................................................................................................................... 95

3.3.5 Mass Spectrometry ................................................................................................. 95

3.4 Results and Discussion .............................................................................................. 96

3.4.1 Stage 1: Identification of Crosslinked Peptides ....................................................... 96

3.4.2 Stage 2: Deduce Partial Sequence for Each Chain ............................................... 100

3.4.3 Stage 3: Inference of Full Sequence for Each Chain ........................................... 104

3.4.4 Stage 4: Deduction of Crosslinking Chemistry and Site ..................................... 111

3.4.5 Final Confirmation and Additional Support .......................................................... 115

3.4.6 Targeted Search Based on the Newly Established Crosslinking Chemistry ......... 121

3.5 Formation of Thioether .............................................................................................. 129

3.6 Conclusions .............................................................................................................. 131

3.7 References ............................................................................................................... 131
Chapter 4: Discovery and Characterization of a Novel His-His crosslink in IgG1 Utilizing $^{18}\text{O}$-labeling and Mass Spectrometry ................................................................. 137

4.1 Abstract .................................................................................................................. 138

4.2 Introduction ............................................................................................................ 139

4.3 Experimental Section .......................................................................................... 140

4.3.1 Chemicals ......................................................................................................... 140

4.3.2 Generation of Stressed Sample ...................................................................... 141

4.3.3 Aggregates by Size Exclusion Chromatography .......................................... 141

4.3.4 Reduction, Alkylation, Tryptic Digestion and $^{18}\text{O}$-Labeling of IgG1 .......... 142

4.3.5 HPLC .............................................................................................................. 143

4.3.6 Mass Spectrometry .......................................................................................... 144

4.4 Results and Discussion ....................................................................................... 145

4.4.1 Detection of Crosslinked Protein ................................................................... 145

4.4.2 Detection of Crosslinked Peptide .................................................................. 148

4.4.3 Elucidation of Crosslinking Chemistry ......................................................... 149

4.4.4 Structural Confirmation by Mass Spectrometry .......................................... 151

4.4.5 Mechanism of formation for His-His crosslink ............................................ 163
4.4.6 Other Crosslinks ................................................................. 172

4.5 Conclusions ........................................................................... 174

4.6 References ............................................................................ 174

Chapter 5: Conclusion and Future Directions ............................................. 180

5.1 isoAsp Project ...................................................................... 181

5.2 Crosslink Project ................................................................... 182

5.3 References ............................................................................ 188
List of Figures

Figure 1-1 MS/MS for mapping PTMs ................................................................. 4

Figure 1-2 Total isoAsp analysis via HPLC/UV at 260nm for SAH measurement or radioactive
detection of MeOH ................................................................. 14

Figure 1-3 Mechanism of fragmentation of the Asp and isoAsp peptides in ETD MS .......... 16

Figure 1-4 Acid-and based catalyzed deamidation ............................................. 19

Figure 1-5 Detection of deamidation and isomerization via (A) isoaspartate methyltransferase
(PIMT)-catalyzed methylation of isoaspartate and hydrazine trapping of methylester and
succinimide and (B) enrichment by hydrazide-aldehyde affinity................................. 21

Figure 1-6 Disulfide scrambling under a basic condition to form crosslink degradants ........ 24

Figure 1-7 The formation of thioether and other related degradants via dehydroalanine followed
by Michael addition ................................................................. 25

Figure 1-8 Non-disulfide crosslinking in insulin Asn deamidation followed by the reaction with
the N-terminal amine ................................................................. 26

Figure 1-9 (A) Dityrosine and Tyr-Cys crosslinks formed via tyrosyl radical (B) Other Tyrosine
related crosslink involving tyrosine oxidation followed by Michael addition with primary amines
................................................................. 28

Figure 1-10 Histidine-related crosslinks via photo-oxidation ................................. 29

Figure 1-11 Dimerization of human superoxide dismutase via a novel oxidative modification —
ditryptophan crosslink ................................................................. 30
Figure 1-12 Formaldehyde-mediated cross-linking .................................................. 31

Figure 1-13 RNase dimerization by a single amide bond between Lys66 and Glu9 under vacuum and 85 °C ........................................................................................................ 32

Figure 1-14 Isotopic labeling at N-termini ................................................................. 38

Figure 2-1 Isotopic distribution of a singly charged DSIP peptide with/without $^{18}$O tag (A) and a triply charged tryptic peptide LC69-108 from the IgG1 sample with/without $^{18}$O tag (B) ....... 63

Figure 2-2 Specificity of PIMT-mediated $^{18}$O-labeling shown in the Asp-DSIP and isoAsp-DSIP samples ........................................................................................................................................ 64

Figure 2-3 The mixture of isoAsp-DSIP and DSIP peptide was analyzed by PIMT/$^{18}$O-labeling method ....................................................................................................................................... 66

Figure 2-4 The effects of pH (A) and incubation time (B) on the hydrolysis of the succinimide and methyl ester ........................................................................................................................................ 69

Figure 2-5 Stability of Asp-DSIP during sample treatment ........................................ 70

Figure 2-6 Guanidine HCl (Gnd-HCl, 1.25 M) quenched PIMT activity during hydrolysis and thereby minimized the incorporation of two $^{18}$O-atoms into isoAsp peptides ..................... 72

Figure 2-7 Identification of isoAsp site in a doubly charged tryptic peptide HC 271-284 from the stressed IgG1 .................................................................................................................. 77

Figure 2-8 Identification of isoAsp site in a triply charged tryptic peptide LC69-108 from the IgG1 sample .......................................................................................................................... 78
Figure 2-9 Identification of isoAsp site in a doubly charged tryptic peptide HC389-405 from the IgG1 sample ................................................................. 79

Figure 2-10 Identification of isoAsp site in the stressed Asp-DSIP peptide by tandem mass spectrometry ................................................................. 82

Figure 3-1 Isotopic distributions of the cross-linked peptide HC:G118-R129/HC:C215-K240 ......................................................................................... 97

Figure 3-2 CID MS/MS spectrum of the triply charged precursor ions at m/z 1351.33 (16O-labeled C-termi) and 1354.00 (18O-labeled C-termi) ......................................................... 114

Figure 3-3a MS/MS data of the cross-link peptide HC:G118-R129/HC:K^Δ214-K240 ......................................................................................... 117

Figure 3-3b MS^3 for structure confirmation of the singly-charged fragment ion m/z 1196 from the cross-link peptide HC:G118-R129/HC:K^Δ214-K240 ......................................................... 118

Figure 3-3c MS^3 for structure confirmation of the double-charged fragment ion m/z 1521 from the cross-linked peptide HC:G118-R129/HC:K^Δ214-K240 ......................................................... 119

Figure 3-4 MS/MS data of the cross-linked peptide HC:C215-K240/HC:C215-K240 ........ 123

Figure 3-5 MS/MS data of the cross-linked peptide HC:K214-K240*/HC:K^Δ214-K240 ...... 124

Figure 3-6 MS/MS data of the cross-link peptide LC:T211-S218/HC:G118-R129 ................. 125

Figure 3-7 MS/MS data of the cross-linked peptide LC:T211-S218/HC:K214-K240 ........... 126
Figure 3-8 (A) Detection of cross-links in IgG2 by reducing SDS-PAGE and (B) Aggregation analysis by size exclusion chromatography .......................................................... 128

Figure 3-9 Major disulfide linkage isoforms in IgG2 ....................................................... 130

Figure 4-1 Detection of crosslinking in IgG1 by reduced SDS-PAGE and size exclusion chromatography (SEC) ........................................................................................................ 147

Figure 4-2 Isotopic distributions of the crosslinked peptide S215-K244/S215-K244 m/z 1673.54 (z=4) from tryptic digestion of IgG1 ................................................................. 148

Figure 4-3 CID MS/MS spectra of the quadruply charged precursor ions m/z 1673.54 (16O-labeled C-termini) and 1675.54 (18O-labeled C-termini) of the crosslinked tryptic peptide S215-K244/S215-K244 ........................................................................................................ 154

Figure 4-4 MS³ spectrum of the doubly charged fragment ion m/z 1488.35 obtained from MS/MS of the precursor ion m/z 1673.54 in Figure 4-3 ......................................................... 156

Figure 4-5 CID MS/MS spectrum of the triply charged precursor ion m/z 1178.77 of the crosslinked S215-E229/S215-E229 peptide generated from combined trypsin and GluC digestion ........................................................................................................ 157

Figure 4-6 CID MS/MS spectrum of the quadruply charged precursor ions m/z 1549.53 of the crosslinked peptide D217-K244/D217-K244 from Asp-N digestion ......................... 158

Figure 4-7 CID MS/MS spectrum of the triply charged precursor ions at m/z 1013.41 of the crosslinked peptide D217-E229/D217-E229 from digestion with Asp-N and GluC .......... 159
Figure 4-8 ETD MS/MS spectrum of the precursor ion m/z 1339.70 (z=5) of the crosslinked tryptic peptide S215-K244/S215-K244 ................................................................. 160

Figure 4-9 ETD MS/MS spectrum of the precursor ion m/z 1033.35 (z=6) of the crosslinked peptide D217-K244/D217-K244 from Asp-N digestion ............................................. 161

Figure 4-10 ETD MS/MS spectrum of the precursor ion m/z 884.33 (z=4) of the crosslinked peptide S215-E229/S215-E229 from digestion with trypsin and GluC ...................... 162

Figure 4-11 ETD MS/MS spectrum of the precursor ion m/z 760.31 (z=4) of the crosslinked peptide D217-E229/D217-E229 from digestion with Asp-N and GluC ......................... 163

Figure 4-12 CID MS/MS spectrum of the tryptic peptide containing the 2-oxo-His (+14 Da) intermediate ................................................................. 167

Figure 4-13 CID MS/MS spectrum of the tryptic peptide containing the His+32 intermediate ...................................................................................... 168

Figure 4-14 (A) Space filling illustration of the hinge region of IgG1 antibody (D\text{KTHTC}PPCP) (B) Three-dimensional (3D) structure of an IgG1 ................................................. 169

Figure 4-15 ETD MS/MS spectrum of the quadruply charged precursor ion m/z 821.09 of the crosslinked peptide D217-E229/S215-E229 generated by limited Asp-N digestion of fully digested IgG1 by trypsin and GluC .................................................................. 171

Figure 4-16 CID MS/MS spectrum of the triply charged precursor ions at m/z 1096.09 of the crosslinked peptide D217-E229/S215-E229 from limited digestion by Asp-N of the fully digested IgG1 by GluC and trypsin ........................................................................ 172
List of Tables

Table 2-1 Representative isoAsp containing peptides detected in IgG1 .......................... 74
Table 3-1 Fragmentation ions of a cross-link peptide ....................................................... 93
Table 3-2 The cross-linked peptide candidates identified by MassAnalyzer algorithm .......... 99
Table 3-3 Peptides in IgG2 that match with the mass of fragment ions of the triply charged ion m/z 1351.33 via FindPept (sorted in the order of primary sequence number) ....................... 105
Table 3-4 Peptides in IgG2 that match with the mass of fragment ions of the triply charged ion m/z 1351.33 via FindPept (grouped in the order of m/z value) ................................. 107
Table 3-5 Partial sequences determined from the mass of fragmentation ions for a triply charge precursor ion at the retention time of 91.17 min with m/z 1351.33 ................................. 109
Table 3-6 De novo sequencing for sequence tag using y-ions from the cross-linked fragments (group 4 in Table 3-1) in the cross-link peptide G118-R129/C215-K240 .................................... 110
Table 3-7 Elemental formula with mass of 149.9987 Da .................................................. 112
Table 3-8 Crosslinked peptides identified in IgG2 .......................................................... 120
Table 3-9 The cross-linking peptides identified in the IgG2 via a targeted search for cysteinyI thioether .................................................................................................................. 122
Table 3-10 Quantification of the cross-linked peptides in the IgG2 ................................. 127
Table 4-1 Partial sequences that match the mass of fragmentation ions for the precursor ion m/z 1673.54 (z=4) eluted at 112.48 min ....................................................... 150
Table 4-2 Deduction of elemental formula for the crosslinked S215-K244/S215-K244 peptide
................................................................................................................................. 151

Table 4-3 Crosslinked peptides obtained from digestion of IgG1 by various proteases and the
combination thereof .............................................................................................................. 153

Table 4-4 Peptides containing the 2-oxo-His (+14 Da) and His+32 (+32 Da) intermediates
observed in the stressed IgG1 ............................................................................................. 166

Table 4-5 Thioether crosslinks detected in IgG1................................................................. 173
List of Schemes

Scheme 1-1 Deamidation, isomerization, racemization and PIMT-dependent methylation ….. 10

Scheme 2-1 Formation of isoAsp from the isomerization of aspartic acid (Asp) or the deamidation of asparagine (Asn) …………………………………………………………………………………… 54

Scheme 2-2 Isotopic labeling of isoaspartic acid via protein isoaspartyl methyltransferase (PIMT)-catalyzed S-adenosyl-methionine (SAM or AdoMet)-dependent methylation and hydrolysis of the resulting methyl ester and succinimide in $^{18}$O-water ……………………. 56

Scheme 2-3 Identification of isoAsp peptides by mass spectrometry using the mass increase of 2 Da imparted by $^{18}$O-labeling ………………………………………………………………………………………………. 57

Scheme 3-1 Flow chart of XChem-Finder in four main stages ……………………………………… 92

Scheme 3-2 Establishment of cross link chemistry based on formula C$_4$H$_6$O$_4$S obtained from elemental composition analysis of 149.9987 Da ……………………………………………………………….. 130

Scheme 4-1 Proposed mechanism for the formation of His-His crosslink via photo-oxidation intermediates …………………………………………………………………………………………… 164

Scheme 5-1 The use of combining sample enrichment and Lys-N digestion for detection of crosslinks …………………………………………………………………………………………… 184

Scheme 5-2 Isotopic labeling at N-termini via 1) trypsin digestion; 2) protection of ε–amino group of lysine by reductive methylation; 3) specific derivatization of N-terminal amino group with a 1:1 mixture of DNFB (2, 4-dinitrofluorobenzene) and [2H$_3$]DNFB at pH 7.0 ………….. 185

Scheme 5-3 N-Terminal Succinylation via two-step chemical derivatizations ………………… 186
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>Asn (N)</td>
<td>asparagine</td>
</tr>
<tr>
<td>Asp (D)</td>
<td>aspartic acid</td>
</tr>
<tr>
<td>Asu</td>
<td>aspartyl succinimide</td>
</tr>
<tr>
<td>CDR</td>
<td>complementary-determining region of IgG</td>
</tr>
<tr>
<td>CID</td>
<td>collision induced dissociation</td>
</tr>
<tr>
<td>°C</td>
<td>degree Celsius</td>
</tr>
<tr>
<td>CEX</td>
<td>cation exchange chromatography</td>
</tr>
<tr>
<td>CHO</td>
<td>Chinese hamster ovary</td>
</tr>
<tr>
<td>Cys</td>
<td>cysteine</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DSIP</td>
<td>β-delta sleep-inducing peptide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>ECD</td>
<td>electron capture dissociation</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGFR</td>
<td>epidermal growth factor receptor</td>
</tr>
</tbody>
</table>
ELISA  enzyme-linked immunosorbent assay

ESI  electron spray ionization

ETD  electron transfer dissociation

FT-MS/MS  Fourier transform tandem mass spectrometry

GndHCl  guanidine hydrochloride

HC  heavy chain of IgG

HCD

His (H)  histidine

His-His  histidine-histidine crosslink

HPLC  high performance liquid chromatography

IAA  iodoacetic acid

ICH  The international conference on harmonization of technical requirements for registration of pharmaceuticals for human use

IgG  immunoglobulin gamma

isoAsp (isoD)  isoaspartic acid

kD  kilodalton

kV  kilovolt

LC-MS  liquid chromatography coupled with mass spectrometry
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography coupled with tandem mass spectrometry</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography, light chain of IgG</td>
</tr>
<tr>
<td>m</td>
<td>milli($10^{-3}$); meter(s)</td>
</tr>
<tr>
<td>M</td>
<td>molarity</td>
</tr>
<tr>
<td>mAb</td>
<td>monoclonal antibody</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix-assisted laser desorption ionization</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>mg/mL</td>
<td>milligram per milliliter</td>
</tr>
<tr>
<td>min</td>
<td>minute(s)</td>
</tr>
<tr>
<td>µ</td>
<td>micro ($10^{-6}$)</td>
</tr>
<tr>
<td>µL</td>
<td>microliter</td>
</tr>
<tr>
<td>mL</td>
<td>milliliter</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry, mass spectrum, mass spectroscopy</td>
</tr>
<tr>
<td>MS/MS (MS²)</td>
<td>tandem mass spectrometry</td>
</tr>
<tr>
<td>MS^n</td>
<td>multiple stage fragmentation</td>
</tr>
<tr>
<td>m/z</td>
<td>Mass-to-charge ratio</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>%</td>
<td>percentage</td>
</tr>
<tr>
<td>pI</td>
<td>isoelectric point</td>
</tr>
<tr>
<td>PIMT</td>
<td>protein isoaspartate methyltransferase</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>HNMR</td>
<td>proton nuclear magnetic resonance</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>reversed phase high performance liquid chromatography</td>
</tr>
<tr>
<td>RT</td>
<td>retention time</td>
</tr>
<tr>
<td>SAH (AdoHcy)</td>
<td>S-adenosyl-homocystein</td>
</tr>
<tr>
<td>SAM (AdoMet)</td>
<td>S-adenosyl-L-methionine</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>size exclusion chromatography</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>UPLC</td>
<td>ultra performance liquid chromatography</td>
</tr>
<tr>
<td>v/v</td>
<td>volume-to-volume ratio</td>
</tr>
</tbody>
</table>
XIC    extracted ion chromatography
Chapter 1: Overview of Protein Post-Translational Modification and Their Analysis

1.1 Protein Post-Translational Modifications

Post-translational modifications (PTMs) are enzymatic or chemical modifications occurring on the amino acid side chain or the amino and carboxyl termini of proteins[1]. Protein backbone cleavage, commonly referred to as proteolysis, is also considered as a PTM in many cases[1]. Over 300 PTMs have been reviewed in the literature[1, 2]. The common types of PTMs include oxidation, deamidation, isomerization, protein crosslinks, glycosylation, phosphorylation, ubiquitination (attachment of 76-amino acid residue ubiquitin protein), nitrosylation, pyroglutamic acid, methylation, acetylation, lipidation. This thesis will focus on Asp isomerization and non-reducible protein crosslinks.

1.2 Biological and Biopharmaceutical Importance of PTMs

PTMs are ubiquitous in biological systems and biopharmaceuticals. PTMs can potentially change protein’s physical or chemical properties, conformation, activity, cellular location or stability and serve as one of the most important regulatory mechanisms for fine-tuning protein function. Therefore, PTMs influence almost all aspects of normal cell biology and pathogenesis. PTMs can occur during manufacturing process and storage, leading to structural changes and loss of efficacy, causing immunogenic response, raising safety concern[3-5]. As examples, biological significance of Asp isomerization and protein crosslinks will be discussed in section 1.4.2 & 1.5.2, respectively.

It has been recognized the potential damage to proteins by solution pH, temperature and light exposure for years as reflected in the ICH Guideline Q1B and Q5C (the international
conference on harmonization of technical requirements for registration of pharmaceuticals for human use, www.ICH.org) [6-10]. As such, stress testing studies must be conducted in order to assess the stability of products and to facilitate improvements in the manufacturing process, formulations, package, and storage conditions. Typical stress studies include high and low pH, elevated temperature, photolysis, oxidation, and freeze-thaw cycles and shear. Photodegradation study covers direct long-term exposure of product to sunlight and common light source for indoor lighting and UV-sterilization in industries. In this thesis, acidic and basic condition leading to isomerization and thioether crosslinks in IgGs will be described in Chapter Two and Three, respectively. The formation of histidine-histidine (His-His) crosslinks in IgG resulting from light exposure will be presented in Chapter Four. A more detail overview on the formation and biological importance of isoAsp and protein crosslinks will be presented in section 1.4 and 1.5, respectively.

1.3 Analysis of PTMs

Identifying and understanding PTMs is critical in the study of cell biology and disease treatment and prevention. However, it is a great challenge to study a specific modified form in a largely heterogeneous protein pool because 1) it is often at a low abundance; 2) most post-translational modifications (e.g. isoAsp formation) alter protein in subtle ways that are not easily detected; 3) a protein can be modified by more than one type of PTMs; 4) furthermore, a protein can be multiply modified by the same PTM at different residues. This challenge will continue to promote methodology development for protein separation/detection as well as better instrumentation. In a constant effort to develop novel, highly sensitive and sophisticated PTM identification techniques, MS-based protein analysis holds great potential for the analysis of
protein PTMs. The currently reported analysis methods of isoAsp and protein crosslinks will be reviewed in sections 1.4.3 and 1.5.3, respectively. In this section, the focus will be on the overview of the determination of PTMs via MS-based approach.

The widely used strategy for protein identification is to cleave the protein with highly specific proteases, followed by LC/MS analysis. PTMs can either increase or decrease the molecular weight of peptides and result in modification-specific signals in MS/MS. In other words, the modification not only makes peptide’s molecular weight shifted, but all fragment ions containing the modified amino acid residue are also mass-shifted (Figure 1-1). However, precise identification of the modification type and the modification site can be very challenging due to 1) the mass shift in the peptide molecular weight. The difficulty increases for small mass shift which requires higher resolution MS. Modifications that are particularly large may shift total mass outside of the mass range suitable for MS/MS sequencing; 2) the overall abundance of the modified peptide. Most PTMs are low abundance and/or substoichiometric; 3) the stability of the modification and the gas phase dissociation behavior of the modified peptides. Some PTMs are labile during MS and MS/MS; 4) the effect of PTMs on protease digestion. The presence of PTMs may affect the cleavage efficiency of proteases; 5) the effect of PTMs on the peptide’s ionization efficiency. The detectability/sensitivity of a peptide is a function of its sequence and modification; 6) multiple-site PTMs. This may generate very complicated MS and MS/MS data that are difficult to interpret; 7) the effect of PTMs on peptide chromatographic behavior; 8) sample handling. Sample preparation may introduce artificial modifications to a protein, such as isoaspartic acid formation; 9) the complex sample matrix of formulation samples or biological samples (such as urine, plasma).
Therefore several strategies such as sample separation/enrichment to reduce sample complexity and to minimize ion suppression, multiple proteases with different cleavage specificity to generate complementary and redundant sets of overlapping peptides, isotope labeling, multiple stage fragmentation (MS\textsuperscript{n}), different fragmentation techniques (CID, collision induced dissociation vs ETD, electron transfer dissociation), etc. are developed for PTM analysis.

Figure 1-1. Mass spectrometry for mapping PTMs\textsuperscript{[11]}. The mass shift from PTM can be detected in MS and the modification site can be located in MS/MS.
The use of specific proteases cleaves proteins into appropriately sized peptides that can be identified easily in MS/MS experiments. It is important to note that larger peptides (>4 kDa) have poor recovery and are difficult to characterize by MS/MS on most commercial instruments (the optimal m/z range is 500-4000 Da for MALDI and 300-1500 Da for ESI), whereas very small peptides (2-3 residues) are often lost due to their poor retention on reversed-phase columns. The choice of a proper protease for digestion will be determined by the nature of amino acid sequence of the protein being analyzed. However, Trypsin and Endoproteinase LysC are the most commonly used in proteomics because they cleave Arg/Lys or Lys at the C-termini of digested peptides which aid in identification via tandem mass spectrometry. Based on the average occurrence of lysine (5.7%) and arginine (5.5%) in proteins in the Swiss-Prot database, digestion with LysC would be expected to produce peptides with an average length of about 17 residues while Trypsin would produce peptides with an average length of about 8 residues [12]. The size of those peptides produced by LysC and Trypsin can be optimally separated by reversed phase HPLC and are very suitable for mass spectrometric analysis. In some instances tryptic peptides may be too small or too large for LC MS/MS analysis. Furthermore, the proximity of post-translational modifications (PTMs) to proteolytic sites can interfere with protease efficiency, which can potentially hinder detection of the PTM. For example, the missed cleavage by trypsin at Lys residue was reported due to its close proximity to the crosslinking site[13, 14]. As such, there are cases where other proteases (such as GluC, AspN, chymotrypsin, etc) with different cleavage specificity are useful. A disadvantage of using GluC, and AspN is that they often yield peptides that are longer and contain one or more internal basic residues, which are poorly fragmentated by CID. But the alternative fragmentation strategies such as ETD and ECD are known to improve identification of long, highly-charged peptides containing basic residues.
The use of non-specific proteases are also reported[15], but these can decrease experimental reproducibility and complicate the separation and identification of proteolytic peptides.

It is important not to introduce any reagents/contamination during sample preparation, since this can inhibit enzyme activity, and generate artifacts, for example, carbamylation. Also, enzyme activity is greatly dependent on the digestion conditions such as pH, temperature, enzyme-to-substrate ratio, etc.

It has become a common practice to use a multiple protease strategy to generate complementary and redundant sets of overlapping peptides for the improvement of protein identification and sequence coverage [16-19]. In Swaney’s study, trypsin, LysC, ArgC, AspN, and GluC were used with two dissociation methods (CID, ETD) in a decision tree-driven fashion for complex protein samples[16]. All digestions were performed under optimized conditions for each protease, respectively. They observed a modest boost in protein identifications (~20%) over the use of a single protease, but a more than two fold improvement in proteome sequence coverage. The optimum digestion pH of each protease as well as the carry-over of proteases to next digestion is an important consideration in a tandem approach. Buffer exchange and other clean-up steps may be necessary for the simplification of the downstream data interpretation.

In addition, the mass spectrometry instruments have been greatly advanced in the recent years. The instruments (for example, Orbitrap and Fourier transform ion cyclotron resonance mass spectrometry) offering high resolution, good mass accuracy, high scan speed, and wide dynamic range are becoming available. In this thesis, the use of $^{18}$O-labeling, complimentary protease digestion, multiple stage fragmentation MS$^n$, different ion activation by CID and ETD
on Orbitrap mass spectrometry to detect and characterize isomerization and protein crosslinks will be discussed in great detail in Chapters Two to Four.

1.4 Deamidation and Isomerization

1.4.1 Isoaspartic Acid Formation

Isoaspartic acid (isoAsp, isoD), aspartic acid in a beta-peptide linkage, is a ubiquitous post-translation modification observed both in vivo and in vitro. IsoAsp can be spontaneously generated through the non-enzymatic deamidation of asparagine (Asn) or isomerization of aspartic acid (Asp) during manufacture process and storage. It is of great concern in the protein pharmaceutical product development. The loss of ammonia from Asn or the dehydration of Asp leads to the formation of a labile intermediate of succinimide (Asu) which readily hydrolyzes to isoAsp and Asp in about 3:1 ratio (Scheme 1-1)[20].

Many factors can influence the rates of isoAsp formation, such as pH, temperature, protein sequences, secondary structures, local three-dimensional structure, etc. Basic conditions (pH>7) favor deamidation while isomerization happens more readily in acidic conditions. In general, the half-times of aspartyl and asparaginyl peptide degradation under physiological conditions (pH 7.5, 37 °C) vary between about 1 and 1000 days[21]. In many case, deamidation rate is determined by the sequence of residues immediately adjacent in the peptide chain and by higher order protein structure[22]. The fastest deamidation sequence was reported as asparagine-glycine (NG) sequence followed by asparagine-serine (NS) sequence. The slowest is asparagine-proline (NP) sequence. Compared to Asn deamidation, Asp isomerization is about 10 times slower under physiological conditions[23]. In 1992, Kroon et al reported that OKT-3, the first marketed monoclonal antibody product(MAb), undergoes deamidation[24]. There are a number
of reports on deamidation and isomerization in MAbs and other proteins of pharmaceutical interest [25, 26].

Spontaneous direct hydrolysis of asparagine residues by water attack on the side chain amide group under pH 3 can also result in aspartyl residue formation[27, 28]. This direct hydrolysis of Asn in an acidic condition results in the formation of Asp as the only product. At neutral pH, the rate of this reaction appears to be much slower than that of the succinimide pathway. The deamidation rate reaches a minimum at approximately pH of 5.

The succinimide is also racemization-prone and can generate the D-succinimidyl, D-aspartyl, and D-isoaspartyl forms (Scheme 1-1)[29, 30]. Zhang et al reported simultaneous isomerization and racemization of Asp in Asp-Asp motif of a therapeutic protein[31]. Young et al demonstrated racemization of Asp-25 in mammalian Histone H2B[32]. Asp residues do not racemize uniformly and specific Asp residues have a greater tendency to racemize than others depending on the neighboring residue of the Asp residue as well as the higher-order structure around the Asp residues in the protein[29, 30]. UV radiation and oxidative stress can promote/induce the racemization of Asp residues[29, 33]. Among the reaction products (including L/D-succinimidyl, L/D-asp, L/D-isoAsp), L-isoAsp is typically the predominate form.

Similar degradation reactions were also reported for glutamine (Gln) or glutamic acid (Glu) residues, but the rate of these reactions via six membrane intermediate are much slower (2 orders of magnitude) of Gln than those of asparagine and aspartic acid residues via five member ring succinimide intermediate[5, 34].

Isoaspartic acid in the damaged protein can be partially repaired by protein L-isoaspartyl/D-aspartyl O-methyltransferase (PIMT, EC 2.1.1.77) which is a repair enzyme that initiates the conversion of L-isoAsp or D-Asp residues to L-Asp residues (Scheme 1-1)[20, 35].
PIMT recognizes and transfers the methyl group from S-adenosyl-L-methionine to L-isoAsp or D-Asp to form the methyl ester. The labile methylester is rapidly converted back to succinimide and subsequent hydrolysis can generate Asp and isoAsp.
Scheme 1-1. Deamidation, isomerization, racemization and PIMT-dependent methylation[29, 30, 32]. This spontaneous intramolecular rearrangement occurs most readily at Asn-Gly, Asn-Ser and Asp-Gly sequences in flexible regions of polypeptides. L-isoAsp form typically accounts for 70-85% of the succinimide hydrolysis product. Protein-L-isoaspartyl methyltransferase (PIMT) catalyzes the methylation of L-isoAsp and D-Asp in the presence of S-adenosyl-L-methionine (AdoMet).
1.4.2 Biological and Biopharmaceutical Importance of Deamidation and Isomerization

Deamidation of asparagine generates aspartate, which fundamentally changes the amino acid composition and charge of the polypeptide post-translationally[20]. Asn deamidation alters the charge of the protein from neutral to negatively charged. Formation of isoAsp from either deamidation of Asn or isomerization of Asp results in the insertion of a methylene group and D-configuration of Asp into the protein backbone[20]. These may dramatically change the protein structure/conformation[36], stability[37], bioactivity[37, 38], aggregation[39, 40], and function[41, 42] leading to aging[40], cancers[43], Alzheimer’s disease[44] and immunogenicity[45-47]. The enzymatic conversion by PIMT of the abnormal isoAsp residues to normal Asp residue in proteins prevents the accumulation of a potentially dysfunctional protein in vivo as cells and tissues age. Studies show that there is increased isoAsp accumulation in tissues (e.g. brain) and in fluids (e.g. urine) in PIMT-deficient mice compared to wild type mice[35, 48]. Furthermore, the average age for PIMT-deficient mice is 42 days, much shorter than that of 22-26 months for wild type mice. These studies clearly demonstrate the harmful consequences of isoAsp accumulation[35, 48]. The D-Asp residues were detected in various proteins from diverse tissues of elderly individuals and related to age-related diseases such as cataract and Alzheimer’s disease[29, 30, 33, 49]. Fujii and coworkers reported high level D-isoAsp at the site of 58 and 151 in the αA-crystalline from aged human lenses, which undergo abnormal aggregation and lead to the reduced chaperone activity[29, 30]. The same research group reported that human skin samples exposed to UV light exhibited significant accumulation of D-Asp by comparison to sun-protected skin[50]. This evidence led to the proposal that assay of D-amino acid accretion could serve as an indicator of sun-induced damage. In the next section, the detection of deamidation and isomerization will be discussed.
1.4.3 Methods for Detection and Characterization of Deamidation and Isomerization

Methods for detection of deamidation are usually based on the charge-sensitive techniques or mass spectrometry analysis[51]. Deamidation introduces negative charges to a protein that shift its isoelectric point (pI). It also results in a +0.984 Da mass increase from Asn to Asp/isoAsp, which can be detected and quantified by mass spectrometry. On the other hand, for isomerization, there is no difference in charge and molecular mass between Asp and isoAsp, which cannot be reliably distinguished by mass spectrometry, the method of choice for analyzing almost all other PTMs. There was no suitable method for Asp isomerization until recently several methods including chemical, instrumental and enzymatic approaches are developed. In this section, a brief review on those methods will be presented and the limitation of each method will also be discussed. The specific focus of this thesis is the determination of deamidation and isomerization using liquid chromatography-mass spectrometry (LC-MS) methods.

1.4.3.1 Edman Degradation

Edman degradation stops at isoAsp residues and was widely used to detect and identify isoAsp residues[52, 53]. Di Donato and his coworkers used chromatographic separation of tryptic peptides from RNase A followed by Edman degradation sequencing to successfully identify Asn67 in Ribonuclease A (RNase A) as the site of deamidation[53]. The Asn67 deamidation was found to impact catalytic property and refolding rate of RNase A[53]. The Edman sequencing method needs purified sample and is not suitable when N-terminal amino group is modified/blocke. In addition, relatively large quantity of protein sample is required for Edman sequencing method.
1.4.3.2 Protein-L-isoaspartyl Methyltransferase (PIMT)

Protein-L-isoaspartyl methyltransferase (PIMT, EC 2.1.1.77) specifically transfers a methyl group from S-adenosyl-L-methionine (SAM or AdoMet) to isoAsp, generating S-adenosyl-homocysteine (SAH or AdoHcy) and the corresponding isoaspartate methyl esters (Figure 1-2)[26, 32, 54, 55]. The methyl ester has increased retention time compared to isoAsp on RP-HPLC due to its increased hydrophobicity, but the methyl ester is labile and readily converts to succinimide. Therefore the methyl ester cannot be used to quantify isoAsp. As such, the methylation by-products, SAH and methanol (MeOH), are measured instead to achieve the global analysis of the isoAsp residue content in the protein samples in commercially available kit IsoQuant. In the radioactive format of IsoQuant, the methyl group donor SAM, is isotopically labeled with tritium and the resulting by-product of $[^3]$H methanol is used for isoAsp quantitation (Figure 1-2). In the HPLC format of IsoQuant, a by-product of SAH is separated by reversed phase HPLC and then quantified using a standard at UV 260 nm (Figure 1-2). The major limitation of this method is that only total isoAsp content is measured and the site of deamidation/isomerization cannot be located.
Figure 1-2. Total isoAsp analysis via HPLC/UV at 260 nm for SAH measurement or radioactive detection of MeOH.

1.4.3.3 Mass Spectrometry

Over the past decade, peptide identification by CID has become the method of choice in mass spectrometry-based proteomics. In neutral and basic solution, deamidation and isomerization involve succinimide intermediate (Asu) (Scheme 1-1). While the Asu intermediate often can be detected as a degradation product (mass decrease of 17 Da), it is readily hydrolyzed in aqueous solution to form the Asp and isoAsp products. The deamidation from Asn to isoAsp and Asp results in 1 Da mass increase which can also be readily detected by modern mass spectrometers, and the deamidation sites can be localized through tandem mass spectrometry fragmentation. In addition, mass spectrometry is a sensitive technique which requires only femtomole to attomole quantities of sample.

Unlike Asn deamidation, Asp isomerization analysis presents a significant challenge for mass spectrometry, as there is no mass change between Asp and isoAsp. Fragment ion intensity
in CID was studied to differentiate isoAsp from Asp. Lehmann et al have noticed that replacement of L-Asp by L-isoAsp resulted in 1) the ion intensity ratio of complementary b and y ions generated by cleavage of N- and C-terminal to the isoAsp decreased, and 2) the Asp ammonium ion abundance at m/z 88 also decreased[56]. However, the b/y ion intensity ratio and the ammonium ion intensity vary considerably depending on the peptide sequence and instrumental settings. Thus, the abundance changes are difficult to use in practice for detection of isoAsp. The development of alternative fragmentation techniques has extended the possibilities within tandem mass spectrometry for isomerization detection. More recently, electron capture dissociation (ECD) and electron-transfer dissociation (ETD) are used for the differentiation of isoaspartic acid and aspartic acid residues by the reporter ions c+57 and z-57 (Figure 1-3)[57-62]. Dai et al reported the identification of major isoAsp-containing proteins in the urine of PIMT-deficient mice via ETD analysis of Lys C digests[63]. The limitations of this approach 1) rely on only one pair of signature ions; 2) The abundance of the reporter ions is low. The intensity of these reporter ions is typically less than 20% of the corresponding c and z ions; 3) Also higher charge state of peptides is required for ETD fragmentation.

Like other PTMs, deamidation and isomerization often are at low abundance. Sample enrichment to improve detection and reduce sample complexity is desired. Since Asn deamidation to Asp/isoAsp changes the charge of peptides/proteins, cation exchange chromatography (CEX) has been reported to separate intact molecules from their charge variants[64, 65]. However, the mobile phase in CEX usually is not compatible with mass spectrometry and fraction collection of CEX for further mass spectrometry identification is needed.
Asp and isoAsp have identical mass and similar pI, thus their analysis remains challenging. Fortunately the structural changes induced by isomerization usually change the retention time of the peptide in reversed phase liquid chromatography (RPLC)[26, 66]. A typical peptide elution order is isoAsp, Asn, Asp, and succinimide[67]. However, caution should be taken for identification based on retention time alone since the separation sometimes varies with different chromatography conditions[61]. As such, RPLC coupled with ETD-MS provides powerful identification tool to differentiate Asp from isoAsp[57, 61].

Figure 1-3. Mechanism of fragmentation of the Asp and isoAsp peptides in ETD MS. (a) formation of c and z fragment ions of the Asp peptides, which is the same for isoAsp peptide; (b) formation of the c+57 and z-57 diagnostic ions of the isoAsp peptides[61].
1.4.3.4 $^{18}$O-Incorporation

As described in the previous section 1.4.1, IsoAsp can be generated from Asn deamidation or Asp isomerization via a common succinimide intermediate. The identification of the succinimide intermediate in proteins is challenging because it hydrolyzes rapidly under neutral to basic pH conditions—typical conditions in protease digestion. Xiao et al have developed a $^{18}$O-labeling method for identification and quantification of succinimide in proteins[68]. The method utilized $^{18}$O water in the hydrolysis of succinimide followed by tryptic digestion and LC/MS analysis to unambiguously identify the sites of deamidation and isomerization via mass increase of 3 and 2 Da comparing to their $^{16}$O counterparts, respectively[31, 68, 69].

Since Asn deamidation to convert to isomeric products (isoAsp and Asp residues) can readily occur even under the mild conditions used to digest protein for LC/MS analysis, it often overestimates the original level of deamidation. The inherent deamidation and those introduced by sample preparation can be differentiated by preparing sample in $^{18}$O-water[70-72]. The artificial deamidation from sample preparation show a 3 Da mass shift while intrinsic deamidation has a 1 Da mass shift compared to the non-deamidated peptide. When the sample preparation is conducted in $^{18}$O-water, protease can simultaneously catalyze the incorporation of up to two $^{18}$O atoms at the peptide C-terminal carboxyl groups, resulting in complicated mass spectra. This limitation can be overcome by a multiple-step calculation procedure[72, 73]. In this method, b ions were used for the calculation of Asn deamidation that occurred prior to or during sample preparation, which eliminated the complexity induced by protease C-terminal $^{18}$O-labeling.
Recently, Wang et al taking advantages of different deamidation mechanisms in acidic and basic conditions introduced isomer-specific mass tags to $^{18}$O-labeled aspartyl and isoaspartyl-containing peptides (Figure 1-4)[74]. Deamidation under basic conditions generates both aspartyl and isoaspartyl-containing peptides while acid-catalyzed deamidation only leads to aspartyl-containing peptides. When $^{18}$O-water is used in those deamidation conditions, different levels of $^{18}$O-incorporation in aspartyl- and isoaspartyl-contained peptides can be achieved. In the acid-catalyzed labeling, deamidation result in a mass increment of $4n+9$ Da (4 Da from each acidic residues, 4 Da from C-terminus and 5 Da from deamidation-formed Asp residue), where $n$ is the number of acidic residues and carboxylated Cys residues in a peptide[74]. In contrast, only one $^{18}$O atom is incorporated during the hydrolysis of succinimide intermediate under basic condition and results in mass shift of 3 Da. These different mass shifts from $^{18}$O-labeling can be exploited for unambiguous assignment of aspartyl- and isoaspartyl-containing peptides by mass spectrometry.
1.4.3.5 Selective Cleavage of isoAsp Peptides with the Asp-N Protease

Endoproteinase Asp-N selectively cleaving only at Asp, not at isoAsp, was used for enrichment of isoAsp-containing peptides for MS analysis[61, 75]. As a result, isoAsp peptide can be differentiated from the peptide containing Asp. Zhang et al employed this finding to identify isoAsp formation at Asp45, Asp47 and Asn47 of recombinant human interleukin-11 (rhIL-11)[76]. Rehder et al used Asp-N peptide mapping to identify the isomerization of Asp92 residues of anti-epidermal growth factor receptor (EGFR) immunoglobulin γ2 antibody, which is contributed to the decreased potency to bind to EGFR as measured by a cell proliferation assay[75]. The limitation of this approach is that Asp-N digestion might be not suitable to digest some proteins resulting in very long peptides which are difficult to identify by mass spectrometry.
1.4.3.6 Chemo-enzymatic Derivatization and Affinity-based Method

Alfaro and co-workers reported a chemo-enzymatic detection of protein isoaspartate by taking advantage of protein isoaspartate methyltransferase (PIMT) to selectively converts isoaspartates into the corresponding methyl esters followed by hydrazine trapping and then aldehyde affinity enrichment (Figure 1-5)[77]. Hydrazides bind to aldehyde resins at mildly acidic conditions (pH 3-6) and the trapped protein isoaspartate can be released with pH 10. The mass increase of 14 Da from isoAsp to hydrazide can be readily detected by standard mass spectrometry. This method can be used not only for site identification, but also for the detection of low abundance isoAsp peptides/proteins. The limitation of this method is that hydrazine trapping is sub-stoichiometric and hydrolysis is competing.
Figure 1-5 Detection of deamidation and isomerization via (A) isoaspartate methyltransferase (PIMT)-catalyzed methylation of isoaspartate and hydrazine trapping of methylester and succinimide and (B) enrichment by hydrazide-aldehyde affinity[77].
1.4.3.7 Protein Isoaspartate Methyltransferase-mediated $^{18}$O-Labeling

Recently, a novel approach for the detection and characterization of isomerization of Asp in IgG1 via protein isoaspartate methyltransferase-mediated $^{18}$O-labeling followed by mass spectrometry analysis has been developed in our lab[26]. In this approach, under mild basic condition, $^{18}$O has been incorporated into succinimide generated from PIMT-mediated methylation of isoAsp. Several isoAsp sites in IgG1 have been identified, which will be detailed in Chapter Two.

1.4.3.8 Methods for Racemization Detection

Racemization is much more difficult to detect because L- and D-amino acids have identical polarity, charge, and molecular weights. However, there are several methods for detection and quantification of racemization, which include chromatography, ELISA(enzyme-linked immunosorbent assay), and enzymatic assays.

In a typical chromatographic protocol, the protein of interest is acid hydrolyzed under very harsh conditions to release individual amino acids for either achiral separation after chiral derivatization or direct chiral separation on a chiral column[25, 78]. The limitation of this method includes that 1) the harsh acid hydrolysis can induce racemization; 2) the site of racemization is unknown. However, by combining Edman sequencing or peptide mapping with derivatization with chiral reagents, both the sequence and stereo-configuration of a peptide can be determined. Fujii et al tryptically digested αA-crystalline and then used reversed-phase HPLC-mass spectrometry to analyze the resulting peptides[79]. After the peptides were identified by mass and sequence analysis, the peptides were hydrolyzed and derivatized with o-
phthaldialdehyde (OPA) for fluorometric derivatization and \textit{N-tert}-butyloxy carbonyl-L-cysteine (Boc-L-Cys) for chiral specificity. They found Asp58 and Asp151 residues in aged human alpha A-crystalline were highly inverted to D-isomers. Inoue et al also reported racemization and isomerization of N-terminal Amyloid-\textbeta in Alzheimer’s brain tissues by covalent chiral derivatized UPLC-MS/MS analysis[80].

As an alternative to HPLC-based techniques, racemized amino acids can be identified and quantified using a stereo-selective enzyme. Protein isoaspartyl methyl transferase (PIMT) selectively recognizes L-isoAsp and D-Asp, which is described in the previous section 1.4.3.2. Again this method does not recognize racemization site in addition to an underestimation of total damage to protein or an overestimation of one form of specific modification since it only recognizes L-isoAsp and D-Asp, not D-isoAsp.

Currently, the most promising method for detection of racemization is to use the sequence specificity and stereospecificity of antibodies[81, 82]. The antibody-based method is a high sensitive and high throughput assay, but its method development is rather arduous.

1.5 Protein Crosslinking

1.5.1 Crosslink Formation

Protein crosslinks as one of PTMs can arise naturally or as degradation. A few protein crosslinks have been reported so far. Biological crosslinks (e.g. the crosslinks formed via transglutaminases, pentosidine and glucosepane crosslinks) are reported in the literature[83-92], therefore the focus here will be on the crosslinks formed as protein degradation[10, 24, 83].
1.5.1.1 Crosslinks as degradants

**Disulfide exchange** Disulfide scrambling, especially under basic condition, forms abnormal disulfide bonds. High pH deprotonates thiols and forms a thiol anion, which initiate thiol disulfide exchange (Figure 1-6)[93]. Therefore, lowering pH (e.g. typical formulation pH ~5) can minimize disulfide bond scrambling. A few scrambled disulfides in stressed therapeutic IgG1s, anti-HER2 and anti-CD11a, were characterized by LC/MS with ETD[94]. Disulfide scrambling is a common issue for proteins containing disulfide bonds, therefore care should be taken during protein manufacturing process.

![Figure 1-6 Disulfide scrambling under a basic condition to form crosslink degradants][93]

**Thioethers** Thioether crosslinks were reported to form as protein degradation products, especially under basic conditions, via dehydroalanine intermediate followed by Michael addition (Figure 1-7)[10, 13, 14, 83, 95-99]. During thioether formation, cysteine racemization on IgG heavy and light chains was observed[100]. The light chain sequence was reported to impact the rate of thioether formation — thioether formation rates were faster for IgG1 containing λ light chains than those containing κ light chains[101]. Mozziconacci et al also observed photolytic
conversion of a disulfide bond in IgG1 to thioether crosslink via a thyl radical-dependent mechanism[98]. As such, thioether crosslink may be a potential issue for the production and formulation of therapeutically disulfide-containing proteins. In this thesis, thioether crosslinks in IgG2 were formed under a basic condition and their characterization will be presented in Chapter Three.

Figure 1-7. The formation of thioether and other related degradants via dehydroalanine followed by Michael addition[10, 13, 14, 83, 95-98].

**Succinimide-mediated intermolecular transamidation** Covalent dimer formation in insulin has been observed both in aqueous and lyophilized formulations[102-104]. Evidence suggests it involves rate-limiting formation of a cyclic anhydride intermediate at the C-terminal Asn\(^{A-21}\) followed by intermediate partitioning to form crosslinking of Asn\(^{A-21}\)-Phe\(^{B-1}\) and Asn\(^{A-21}\)
Gly\textsuperscript{A-1} (Figure 1-8)[102-104]. Similar intermolecular amide-linked crosslinking was also observed in hen egg-white lysozyme[105].

Tyrosine-related crosslinks Dityrosine crosslink, a biomarker of oxidative/nitrative stress, is a fluorescent molecule detected as photo-degradation in many proteins (such as insulin, calmodulin, etc.) [106, 107]. The dityrosine crosslink formed by UV irradiation of bovine brain calmodulin was believed to be the intermolecular crosslinking of Tyr99 and Tyr138[108]. The mechanism of dityrosine formation begins with the generation of a tyrosinal radical which then is crosslinked to form dityrosine (Figure 1-9A)[106]. In addition, Mozziconacci et al studied
photodegradation of recombinant human insulin in the solid state[109]. They found dithiohemiacetal and Tyr-Cys crosslink by GluC digestion of the UV-irradiated human insulin followed by mass spectrometry analysis. UV-exposure of solid human insulin results in photodissociation of the C-terminal intrachain disulfide bond, leading to the formation of a thyl radical pair which reacts to proximal Tyr radical to form Tyr-Cys crosslink (Figure 1-9A)[109]. So in order to be crosslinked, the involved amino acid residues must be within a certain distance of each other.

Other Tyrosine related crosslinks were also detected in the metal-catalyzed oxidized recombinant human Interferon β-1a and recombinant human insulin (Figure 1-9B)[110, 111]. Tyrosine residues were first oxidized in the presence of copper(II) and ascorbate. The tyrosine oxidation products undergo Michael addition which is initiated by a primary amine group (Figure 1-9B). This results in the formation of crosslinks which are most likely responsible for aggregation[111].
Figure 1-9. (A) Dityrosine and Try-Cys crosslinks formed via tyrosyl radical under oxidation conditions[106, 109]. (B) Other Tyrosine related crosslink in metal catalyzed oxidized interferon β-1a involving tyrosine oxidation followed by Michael addition with primary amines of Lys side chain or N-termini[110, 111].

**Histidine-histidine crosslink** Histidine-histidine crosslink formed by photo-oxidation was reported in free histidine and peptides containing histidine (Figure 1-10)[112, 113]. The evidence obtained supports the role of singlet oxygen in the formation of reactive peroxide
intermediate on exposure of His-containing peptides to light, which leads to final histidine-histidine crosslink (Figure 1-10). Most recently, histidine-histidine crosslinks in the hinge region of light stressed IgG1 as protein photodegradation product has been discovered and characterized in our lab. The detail will be discussed in Chapter Four.

Figure 1-10. Histidine-related crosslinks via photo-oxidation[112].

**Ditryptophan crosslink** A non-disulfide covalent dimer of human superoxide dismutase 1 (hSod1), which was produced during its bicarbonate-dependent peroxidase activity in vitro, has been isolated and characterized by coupling $^{18}$O-labeling and mass spectrometry analysis
recently[114]. This covalent dimer was found to consist of two hSod1 subunits crosslinked by a ditryptophan, which contains a bond between C3 and N1 of the respective Trp32 residues (Figure 1-11)[114]. Carbonate radical was believed to promote ditryptophan crosslink[114].

Figure 1-11. Dimerization of human superoxide dismutase via a novel oxidative modification — ditryptophan crosslink[114].

**Formaldehyde-mediated crosslinking** Formaldehyde-mediated crosslinking was also reported to cause significant aggregation of lyophilized tetanus and diphtheria toxoids during storage[115]. Formaldehyde, used to prepare the toxoid from the native toxin, reacts with Lys residue and results in the reactive electrophiles which react with nucleophiles of a second vaccine molecule to form intermolecular crosslinks (Figure 1-12).
Figure 1-12. Formaldehyde-mediated cross-linking in vaccines where a formaldehyde-modified electrophile is attacked by nucleophiles to form intermolecular crosslinks[115].

**Amide crosslink** A cross-linked ribonuclease A (RNase A) dimer composed of monomeric units covalently linked by a single amide bond between the side-chain of Lys66 and Glu9 was generated without chemical reagent and characterized by mass spectrometry(Figure 1-13)[116, 117]. It is very interesting that this dimer shows a two-fold increase in activity over monomeric RNase A[116].
1.5.2 Biological and Biopharmaceutical Importance of Crosslinks

Since the technical challenge on analysis of crosslink (as described in the next section 1.5.3), the limited crosslinks (such as thioether) were discovered serendipitously and characterized with painstaking efforts. This results in limited biological knowledge on protein crosslinks. In this section, a brief review of biological and biopharmaceutical importance of crosslinks will be presented next.

Protein crosslinks play a significant role in protein structure and stability. For example, disulfide bonds between cysteine residues separated in primary sequence often help to stabilize tertiary structure and subsequently affect biological activity and stability[51, 118]. On the other hand, abnormal crosslinks often lead to protein stability problem and immunogenicity. Aggregation of type I soluble tumor necrosis factor receptor due to photoirradiation was reported via disulfide formation[119]. Aberrant disulfide linkages are also reported to associate with disease. The mutation of Cys470 to Arg in recombinant human arylsulfatase A (rhASA) and thus disruption of a disulfide linkage, has been reported in patients with metachromatic leukodystrophy (MLD), an autosomal recessive disease[120].

Figure 1-13. RNase dimerization by a single amide bond between Lys66 and Glu9 under vacuum and 85 °C[117]
Dityrosine crosslinking is likely responsible for the dimerization and decrease bioactivity of insulin[107]. Elevated levels of urinary dityrosine have been demonstrated in aging animals and patients with systemic inflammation[106].

Thioether crosslink in commercial recombinant human growth hormone (r-hGH) exhibited a significantly reduced in vivo biopotency and altered receptor-binding properties compared with a control[96].

Covalent insulin dimers formed through transamidation reactions of Asn$^{A21}$ and Phe$^{B1}$ accumulate in the circulation of type I diabetic patients undergoing prolonged insulin therapy, accounting for significantly reduced insulin biologic activity[121].

1.5.3 Methods for Detection and Characterization of Crosslinks

1.5.3.1 HPLC with Fluorescence Detection

Dityrosine crosslinks were detected and quantified by LC with fluorescence detection after acid hydrolysis of proteins[106, 122]. This method provides no structural data of crosslink site. In addition, the method is not suitable for the crosslinks which are not stable to conventional conditions of acid hydrolysis. Furthermore, HPLC results can be altered by the presence of other molecules that can coelute with the target molecule. This requires additional characterization (e.g. by MS) on the molecule of interest.

1.5.3.2 MS-based Method

Despite the excellence of mass spectrometry as an analytical tool for PTMs, it is very challenging to identify crosslinked peptides after proteolytic digestion of crosslinked proteins
because crosslinked peptides are often present at substoichiometric levels, which leads to failure in detecting them during data-dependent LC/MS analysis. Even when the crosslink peptides have been identified, it remains challenging to assign sequences and locate site of crosslinking. This is because (1) tandem mass spectra of crosslinked peptides are complicated by the presence of two sets of fragment ions; (2) the masses of crosslinked peptides are not in database if the crosslink chemistry is unknown. Therefore, traditional database search algorithms and de novo sequencing cannot be used to interpret their tandem mass spectrometry. In chemical crosslinking widely used to probe protein structure and interaction, the crosslink chemistry is known. A database of the intact mass and the tandem mass spectra for possible combination of crosslinked peptides can be computerized, and subsequently used to correlate with observed spectra to identify both the peptide sequence and sites of modification. However, this approach is futile if the crosslink chemistry is unknown. This ultimately leads to very limited crosslinks discovered. It is very helpful for interpretation of the tandem mass spectra if the fragment ions with and without crosslink site can be distinguished isotopically or chemically. The linear fragment ions then can be used for database searching as well as de novo sequencing[13]. This will be described in more detail in our XChem-Finder workflow in Chapter Three and Four. In this section, a brief summary of the techniques used to facilitate MS analysis of crosslinks will be presented next.

1.5.3.2.1 C-Terminal $^{18}$O-Labeling:

Various stable isotope labeling techniques via metabolic labeling, chemical tagging, or proteolytic $^{18}$O labeling have been developed and used for relative quantitation of change in protein abundances between two compared samples, and also for qualitative characterization of
differentially labeled proteomes[123-125]. In this section the focus will be on simple and easy enzyme-catalyzed $^{18}$O-labeling.

Protease catalytic $^{18}$O-labeling relies on the help of proteases (such as trypsin, Lys-C, Glu-C) to exchange two $^{16}$O atoms for two $^{18}$O atoms at the C-terminal carboxyl group of each newly formed digested peptide, resulting in mass shift of 4 Da[123, 126]. It was reported the optimum pH for the carboxyl oxygen exchange reaction catalyzed by Lys-C and trypsin are 5 and 6, respectively[126]. Other protease such as endoproteinase Lys-N was reported only incorporate one $^{18}$O atom and yield spectra insufficient to resolve isotope peak overlap[127]. Different peptides incorporate $^{18}$O atoms at different rates, which can complicate data analysis and limit its applications in quantitative proteomics[128]. As a matter of fact, trypsin-catalyzed $^{18}$O exchange at the carboxyl terminus is in many instances inhomogeneous/incomplete[129]. Also back exchange of the carboxyl oxygen of $^{18}$O-labeled peptide to oxygen-16 could occur in $^{16}$O-water. Therefore, several approaches including using high enzyme-to-protein ratio[130], low pH[131], heating[130], immobilized trypsin[132] were developed to optimize the exchange reaction.

Incorporation of $^{18}$O at a newly created C-termini during proteolytic digestion represents clever approach to specifically detect crosslinked peptides[133]. Four and two $^{18}$O will be incorporated at C-termini of a crosslinked peptide and linear peptide, respectively, thus resulting in a specific isotopic signature in the mass shift of 8 Da for crosslinked peptides and 4 Da for linear peptides. The proteolytic $^{18}$O-labeling achieving isotope labeling concurrent with the proteolytic digestion of proteins offers simplicity. In this thesis, the strategy of incorporation of $^{18}$O at C-terminus of peptide during tryptic digestion was used to facilitate the identification of
thioethers and histidine-histidine crosslinks, and the optimized \(^{18}\text{O}\)-labeled method will be described in more detail in Chapter Three and Four.

For large molecular weight crosslinks, this \(^{18}\text{O}\)-labeling may not produce sufficient mass difference between light and heavy form, therefore complicating detection of crosslinks. Since protease-catalyzed \(^{18}\text{O}\)-labeling is only applied to newly created C-termini, a crosslink containing C-terminus of protein has mass shift of only 4 Da compounding with linear peptides. To overcome this, other strategies such as using proteases with different specificity or N-terminal modification can be explored.

1.5.3.2.2 N-Terminal Modification

Since crosslinks contain two amino termini, isotopic labeling of \(\alpha\)-amino groups with N-terminal modification reagents will lead to incorporation of the two isotopically coded groups in crosslinked peptides versus one in case of linear peptides (Figure 1-14)[134]. Modification of crosslinked peptides with an equimolar mixture of light and heavy isotopic forms of an amine-reactive reagent e.g. \(^2\text{H}_3\)-2,4-dinitrofluorobenzen (\[^3\text{H}_3\]NDFB) results in a specific triplet of signals separated by mass according to the mass difference between the light and heavy isotopic forms of the reagent[134]. The resulting 1:2:1 intensity ratio of these peaks is due to the possible combinations of the different isoforms of the product (LL, LH+HL, HH, where L and H are light and heavy forms, respectively). On the other hand, modification of a single N-terminal amino group of linear peptides results in a doublet of signals in a 1:1 ratio due to the possible combination of L and H isoforms[134]. One complication in this approach is the possible modification of the \(\varepsilon\)-amino groups of lysine residues which will produce a false positive isotopic signature for lysine-containing linear peptides. Blocking lysine residues by reductive
methylation and then selectively hydrolyzing to release α-amino termini for derivatization with $[^2\text{H}_3]\text{DNFB}$ prior to enzymatic digest was proposed (Figure 1-17)[134]. The utility of this approach was demonstrated in the characterization of the unique crosslinks of polyubiquitin[134]. The lysine residue protection adds additional step in sample preparation and also leads to a higher molecular mass tryptic crosslinks due to missed digest sites at the modified lysine residues.
Figure 1-14. Isotopic labeling at N-termini via 1) protection of ε–amino group of lysine; 2) limited proteolysis to generate peptide; 3) specific derivatization of the liberated–amino group with \([^{2}H_{3}]DNFB[134]\).

1.5.3.2.3 Chromatographic Sample Enrichment

In order to improve identification of protein crosslinks using mass spectrometry, sample enrichment can be greatly helpful. The separation based on protein size, e.g. size exclusion chromatography(SEC), was successfully used to enrich crosslinked protein for the characterization of thioether crosslinks in monoclonal antibody by LCMS analysis[14, 95].
Also, crosslinked peptides having higher charge state than linear peptides can be enriched by cation exchange chromatography (CEX)[135].

**1.5.3.3 Antibody-based Method**

Kato and colleagues have developed rabbit polyclonal and mouse monoclonal antibodies to detect dityrosine immunohistochemically in lipofuscin granules in aged human brain and in atherosclerotic lesions in mice[136, 137]. However, method of immunodetection is semiquantitative, requires extensive sample preparation and antibody purification, and may be confounded by the presence of crossreacting proteins in the sample of interest.

**1.6 Conclusions**

Detection and characterization of PTMs remain challenging, especially when at low abundance. The high sensitive MS-based method is a method of choice for the analysis of PTMs. Various techniques as described in this chapter have been developed to enrich/facilitate the detection of PTMs. The demand to detect and characterize the low level of various PTMs for obtaining the knowledge of protein degradation pathway, controlling the quality of therapeutic proteins, and understanding PTMs biological importance will drive the continuing effort to develop novel, highly sensitive and sophisticated PTM identification techniques.
1.7 References


[36] Noguchi S. Structural changes induced by the deamidation and isomerization of asparagine revealed by the crystal structure of Ustilago sphaerogena ribonuclease U2B. Biopolymers 2010;93:1003-10.


Chapter 2: Protein Isoaspartate Methyltransferase-Mediated $^{18}\text{O}$-Labeling of Isoaspartic Acid for Mass Spectrometry Analysis


Co-authors’ work in this chapter: Min Liu: experimental design and execute, data analysis, manuscript writing and revision; Janet Cheetham: manuscript writing and revision, grant support; Nina Cauchon: manuscript writing and revision; Judy Ostovic: manuscript writing and revision; Wenqin Ni: PIMT purification, manuscript writing and revision; Da Ren: idea contribution, experimental design, manuscript writing and revision; Zhaohui Sunny Zhou: idea contribution, experimental design, data analysis, manuscript writing and revision and grant support.
2.1 Abstract

Arising from spontaneous aspartic acid (Asp) isomerization or asparagine (Asn) deamidation, isoaspartic acid (isoAsp, isoD or beta-Asp) is a ubiquitous non-enzymatic modification of proteins and peptides. Due to no mass difference between isoaspartyl and aspartyl species, sensitive and specific detection of isoAsp, particularly in complex samples, remains challenging. Here we report a novel assay for Asp isomerization by isotopic labeling with $^{18}$O via a two-step process: the isoAsp peptide is first specifically methylated by protein isoaspartate methyltransferase (PIMT, EC 2.1.1.77) to the corresponding methyl ester, which is subsequently hydrolyzed in $^{18}$O-water to regenerate isoAsp. The specific replacement of $^{16}$O with $^{18}$O at isoAsp leads to a mass shift of 2 Da, which can be automatically and unambiguously recognized using standard mass spectrometry, such as collision-induced dissociation (CID), and data analysis algorithms. Detection and site identification of several isoAsp peptides in a monoclonal antibody and the $\beta$-delta sleep-inducing peptide (DSIP) are demonstrated.

2.2 Introduction

The non-enzymatic post-translational formation of isoaspartic acid (isoAsp, isoD or beta-Asp) in oligopeptides (Scheme 2-1), arising from either the isomerization of aspartic acid (Asp) or the deamidation of asparagine (Asn) via a common succinimide intermediate, is one of the major chemical degradation pathways both in vivo and in vitro. The formation of isoAsp via Asp isomerization and Asn deamidation in a protein alters its structure by inserting an extra methylene group into the peptide backbone[1] and, in doing so, may change protein activity or trigger an immunologic response[2, 3]. As such, this protein post-translational modification
(PTM) plays critical roles in biological processes, human diseases and protein pharmaceutical development. For example, isoAsp level is elevated in amyloid-beta peptides in Alzheimer’s disease[4]. IsoAsp accumulates over time and thus is associated with aging, perhaps acting as a molecular clock[5-9]. In addition, significant amounts of isoAsp are also commonly observed in protein pharmaceuticals and represent a major contributor to heterogeneity, particularly after long-term storage[10-12]. Typically, pH is one of critical factors affecting the formation of Asn deamidation and Asp isomerization; both processes are also reported to depend on the primary sequences, higher-order structures and formulation[10, 11, 13-15].

Scheme 2-1. Formation of isoAsp from the isomerization of aspartic acid (Asp) or the deamidation of asparagine (Asn).
For analysis, it is challenging to differentiate isoAsp from Asp or Asn (particularly the former), as isoAsp and Asp have identical mass and bear similar charge and structure. Current approaches for isoAsp analysis include chemical (e.g., Edman degradation), immunological, enzymatic (e.g., isoQuant), and instrumental methods (e.g., chromatography and mass spectrometry)[16, 17]. High performance liquid chromatography (HPLC) coupled with mass spectrometry/collision induced dissociation (LC-MS/CID) is commonly used for the characterization of protein modifications, however MS/CID often fails to differentiate isoAsp and Asp[17]. Recently O’Connor and others have demonstrated that electron transfer dissociation (ETD)/electron capture dissociation (ECD) mass spectrometry is able to distinguish isoAsp from Asp peptides based on a pair of the characteristic reporter ions of isoAsp (c+58 and z-57)[16-20]. However, the peak intensity of this single pair of diagnostic ions of isoAsp may vary under different conditions, often requiring manual inspection of the spectral data and making assignment ambiguous when isoAsp is present at low abundance; moreover, multiply charged precursor ions are required as well, limiting the scope of this method.

Another commonly used assay is based on protein isoaspartate O-methyltransferase (PIMT or PCMT, EC 2.1.1.77). This enzyme specifically transfers a methyl group from S-adenosyl-L-methionine (SAM or AdoMet) to isoAsp, generating S-adenosyl-homocysteine (SAH or AdoHcy) and the corresponding isoaspartate methyl esters (Scheme 2-2)[5, 21-24]. As a result, the amount of isoAsp can be deduced from quantifying AdoHcy, the byproduct of methylation. The other methylation product, isoAsp methyl ester, is labile and spontaneously cyclize to aspartyl succinimide (Asu), which is also labile under most conditions used for the analysis of peptides and proteins[25, 26]. Therefore, the information on the specific location of isoAsp in peptides is often lost as the methyl esters and succinimides typically rapidly hydrolyze.
back to isoAsp and Asp. To overcome these limitations, trapping the labile isoaspartyl methyl esters and succinimides with hydrazines or hydroxylamines has been developed for isoAsp detection[27, 28]. However, the conversion of methyl esters to hydrazides or hydroxamic acids is not stoichiometric, as water present in the reagent solutions competes with the trapping reaction. On the other hand, as discussed below, the hydrolysis reaction in $^{18}$O-water represents an attractive method for isoAsp labeling.

Scheme 2-2. Isotopic labeling of isoaspartic acid via protein isoaspartyl methyltransferase (PIMT)-catalyzed S-adenosyl-methionine (SAM or AdoMet)-dependent methylation and hydrolysis of the resulting methyl ester and succinimide in $^{18}$O-water.
Scheme 2-3. Identification of isoAsp peptides by mass spectrometry using the mass increase of 2 Da imparted by $^{18}$O-labeling.

Stable isotope labeling combined with mass spectrometry analysis is a powerful tool for identification and quantification due to the fact that no detectable change in retention time, ionization efficiency and fragmentation patterns after isotope labeling is observed[29-37]. For example, Fenselau and coworkers have developed a general strategy to label C-termini of peptides using proteases and $^{18}$O-water for peptide quantification and identification[30, 35, 36]. The use of $^{18}$O to quantitate succinimide and to track deamidation during sample handling has also been reported[29, 31-34]. Herein we describe a novel isoAsp assay that couples PIMT-mediated methylation with $^{18}$O-labeling followed by LC-MS analysis (Scheme 2-2 & 2-3). In
the first step, PIMT specifically methylates isoAsp to form isoaspartate methyl ester. Subsequently, the labile methyl ester spontaneously cyclizes to the succinimide intermediate (Asu) which then hydrolyzes in $^{18}$O-water to produce $^{18}$O-labeled isoAsp[29, 31-34, 38]. The incorporation of $^{18}$O shifts the modified residual mass by 2 and 3 Da relative to Asp and Asn, respectively. This allows facile screening and site determination of isoAsp using standard mass spectrometry techniques (such as CID) and data analysis algorithms. Using our method, several isoAsp peptides in a recombinant monoclonal antibody and a synthetic peptide were detected and the sites of isoAsp were identified. In addition, aspartyl succinimide (Asu) and isoAsp can be distinguished by $^{18}$O-labeling in the presence and absence of PIMT. Because mass spectrometry with reasonably high resolution can distinguish isoAsp from Asn (a mass increase of 0.984 Da), our method focuses on the analysis of isoAsp from Asp (no change in mass)—the smallest protein post-translational modification.

2.3 Experimental Section

2.3.1 Chemicals

All chemicals were reagent grade or above. Guanidine hydrochloride (GndHCl) and S-adenosyl-methionine hydrochloride (AdoMet or SAM) were purchased from Sigma (St. Louis, MO, USA). $^{18}$O-water (97%) was obtained from Cambridge Isotope Laboratories (Andover, MA, USA). β-Delta sleep-inducing peptides (Asp-DSIP, Trp-Ala-Gly-Gly-Asp-Ala-Ser-Gly-Glu, and isoAsp-DSIP, Trp-Ala-Gly-Gly-isoAsp-Ala-Ser-Gly-Glu) trifluoroacetate salts were purchased from Bachem America (King of Prussia, PA, USA). Immobilized trypsin was purchased from Thermo Scientific (Rockford, IL, USA). Recombinant protein L-isoaspartyl-O-
methyltransferase (PIMT) was obtained as previously described[39]. Recombinant monoclonal antibody, anti-streptavidin immunoglobulin gamma 1 (IgG1), was produced in Chinese hamster ovary (CHO) cells, purified according to standard manufacturing procedures, formulated at a concentration of 20 mg/mL in 50 mM acetate buffer at pH 5.2, and stored at -70 °C at Amgen (Thousand Oaks, CA, USA).

2.3.2 Generation of isoAsp

The Asp-DSIP peptides were dissolved to a final concentration of 1 mg/mL in 0.1 M acetate buffer at pH 4.8 and stored at 50 °C for 3 days. After being exchanged into 0.1M acetate buffer at pH 4, the antibody IgG1 was incubated at 45 °C for 1 month.

2.3.3 Reduction, Alkylation, and Tryptic Digestion of IgG1

IgG1 (20 mg/mL) was diluted to 1 mg/mL in the denaturation buffer (7.5 M guanidine HCl, 2 mM EDTA and 0.25 M Tris-HCl, pH 7.5) to a final volume of 0.5 mL. Reduction was accomplished with the addition of 3 μL of 0.5 M dithiothreitol (DTT) followed by 30 min incubation at room temperature. S-Carboxymethylation was achieved with the addition of 7 μL of 0.5 M iodoacetic acid (IAA); the reaction was carried out in the dark for 15 min at room temperature. Excess iodoacetic acid was quenched with the addition of 4 μL of 0.5 M DTT. Reduced and alkylated IgG1 samples were exchanged into the digestion buffer (0.1 M Tris-HCl at pH 7.5) using a NAP-5 size-exclusion column (GE Healthcare, Piscataway, NJ, USA). Then, 150 μL of immobilized trypsin suspension was centrifuged at 900 x g for 10 seconds and the
supernatant was removed followed by three washes, each with 1 mL of 0.1 M Tris-HCl at pH 7.5. The washed immobilized trypsin was mixed with 300 μL of the reduced, alkylated, and buffer-exchanged antibody to achieve a 1:2 (v/v) enzyme/substrate ratio before incubation at 37 °C for 45 min. Subsequently, 300 μL of acetonitrile was added and the supernatant was collected after centrifuging at 900 x g for 10 seconds. Then, 200 μL of the supernatant was dried and then reconstituted into 98 μL of 18O-water for methylation as described next.

2.3.4 Methylation Catalyzed by PIMT

Methylation reactions were carried out in a final volume of 100 μL containing 0.1 M Tris-HCl at pH 7.5, 7 μM of DSIP peptides or the tryptic digest of IgG1, 120 μM of AdoMet and 15.9 μM of PIMT at 37 °C for 30 min. To quench methylation, guanidine HCl (GndHCl, 10 M) was added to a final concentration of 2.3 M. It should be noted that the above buffer or reagents were prepared in 18O-water, instead of normal water.

2.3.5 18O-Labeling

Sodium bicarbonate buffers at pH 8.5 in 18O-water was prepared by drying 100 μL of 1 M sodium bicarbonate solution at pH 8.5 followed by reconstitution into the same volume of 18O-water. To initiate the hydrolysis of succinimide, 60 μL of 1 M sodium bicarbonate-18O-water was added to the methylated peptide solution described above. The reaction was then conducted at 37 °C for 30 min.
2.3.6 HPLC

The separation of the DSIP peptides was carried out on an XBridge C18 column (150 x 2.1 mm, 3.5 µm, Waters, Milford, MA, USA) at column temperature of 50 °C with a flow rate of 200 µL/min. Mobile phase A consisted of 0.1% formic acid in water while mobile phase B contained 0.085% formic acid in 90% acetonitrile. A linear gradient was applied by increasing mobile phase B from 0 to 50% in 60 min. The injection volume was 25 µL. Chromatographic profiles were monitored by UV absorption at 215 nm.

Tryptic digests of IgG1 sample were separated on a Polaris Ether C18 column (250 x 2.1 mm, 3 µm, Varian, Palo Alto, CA, USA) at column temperature of 50 °C with a flow rate of 200 µL/min. Mobile phase A was 0.1% trifluoroacetic acid in water while mobile phase B contained 0.085% trifluoroacetic acid in 90% acetonitrile. A linear gradient was applied by increasing mobile phase B from 0 to 50% in 195 min. The injection volume was 100 µL. Elution profiles were monitored by UV absorption at 215 nm.

2.3.7 Mass Spectrometry

LXQ and LTQ Orbitrap mass spectrometers (ThermoFisher Scientific, San Jose, CA, USA) were used in-line with HPLC systems (Agilent 1100, Palo Alto, CA, USA) for the stressed Asp-DSIP and IgG1 samples, respectively. LXQ was operated with a full scan, zoom scan and data-independent MS/MS scan. The spray voltage was 5 kV, and the capillary temperature was 280 °C. For LTQ Orbitrap, a high resolution full MS scan at 60,000 resolution (at m/z 400), followed by data-dependent MS/MS scans of the top three most abundant ions, was set up to
acquire both the mass and the sequence information. The spray voltage was 5 kV, and the capillary temperature was 300 °C. Both instruments were tuned using the doubly charged ion of the synthetic peptide, Bradykinin. The MS/MS spectra were obtained using normalized collision energy of 35%. Mass Analyzer software developed in-house was used for peptide identification and sequencing[40, 41]. Extracted ion chromatograms (XIC) were used to quantify the relative amount of isoAsp peptide and its Asp isomer.

2.4 Results and Discussion

As expected from the general approach outlined in Scheme 2-2 & 2-3, isoAsp peptides were observed completely labeled by $^{18}$O via the sequential methylation and hydrolysis in $^{18}$O water, as evident from the mass shift of 2 Da in isotopic distribution (see Figure 2-1). By searching for potential 2 Da mass increases for each Asp residue using a standard data mining algorithm, e.g. MassAnalyzer, several $^{18}$O-labeled isoAsp peptides from the IgG1 sample were automatically identified. Furthermore, the precise locations of isoAsp were readily and unambiguously established by tandem mass spectrometry using CID.
Figure 2-1. Isotopic distribution of a singly charged DSIP peptide with/without $^{18}$O tag (A) and a triply charged tryptic peptide LC69-108 from the IgG1 sample with/without $^{18}$O tag (B). Their sequences with alkylated cysteines are

\[69\text{SGTASLAI}T\text{GLQAEDAYCQS}Y\text{isoDSSLSGLYVF}GT\text{GTK}^{108}\] and

\[69\text{SGTASLAI}T\text{GLQAEDAYCQS}Y\text{DSSLSGLYVF}GT\text{GTK}^{108}\] respectively.
2.4.1 Methylation of isoAsp

The specificity of the PIMT-catalyzed methylation of isoAsp residues has been extensively investigated, indicating that only isoAsp residues, but not Asp, are recognized[2, 5, 9, 10, 24]. In agreement with the literature, no methyl ester or succinimide was observed for the Asp-DSIP peptide (Figure 2-2).

Figure 2-2. Specificity of PIMT-mediated $^{18}$O-labeling shown in the Asp-DSIP and isoAsp-DSIP samples. Methylation via PIMT specifically occurs at isoAsp peptide (bottom trace), not Asp peptide (top trace). The corresponding methyl ester is labile and converts to the succinimide intermediate spontaneously. Hydrolysis of succinimide in $^{18}$O-water forms $^{18}$O-isoAsp (major product) and $^{18}$O-Asp (minor product).
As illustrated in Scheme 2-2, the degree of methylation of isoAsp peptides dictates the overall yield of $^{18}$O incorporation into isoAsp residues. Therefore, methylation efficiency was optimized before attempting $^{18}$O-labeling. As for a typical enzymatic transformation, methylation is faster under higher PIMT concentration, so a relatively high concentration of PIMT (16 μM) was employed. In addition, PIMT is known to be sensitive to feedback inhibition from the product, S-adenosyl-homocysteine (AdoHcy or SAH)[42], so excess of AdoMet (the methyl donor) was also used. If necessary, product inhibition can be further alleviated by the addition of AdoHcy nucleosidase (EC 3.2.2.9), as has been previously demonstrated[43-45]. In addition, pH, temperature, detergents (e.g. SDS) or chaotropic reagents (e.g. guanidine-HCl) may affect PIMT enzyme activity, therefore caution was taken during sample preparation. Under our conditions, methylation of isoAsp was completed as evidenced by the disappearance of isoAsp peptide peak and the concomitant appearance of peaks for its methyl ester and succinimide peptide (Figure 2-3). Moreover, complete methylation is also supported by the near stoichiometric incorporation of $^{18}$O into each isoAsp residue in isoAsp-DSIP and the tryptic peptides from the IgG1 antibody described next (Figure 2-1 and Figure 2-3).
Figure 2-3. The mixture of isoAsp-DSIP and DSIP peptide was analyzed by PIMT/^{18}O-labeling method. Complete conversion of isoAsp to its corresponding methyl ester and succinimide was evidenced by the disappearance of the isoAsp peak. Complete hydrolysis in the next step was supported by the disappearance of the methyl ester and succinimide peaks in buffer pH 9 for 30 min. No significant isoAsp change was observed when the sample was exposed to buffer at pH 9, 37 °C, after 30 min, 3 hrs and 24 hrs (A). No isotopic distortion was observed when the sample was exposed for 24 hrs (B).
To ensure all isoAsp residues are accessible to PIMT, the IgG1 was first digested by trypsin prior to methylation. Immobilized trypsin was used and removed from the peptides after digestion for several reasons: first and foremost, to prevent $^{18}$O-labeling at the C-termini of peptides[30, 35, 36]; second, to prevent proteolysis of PIMT. For comparison, the reduced and alkylated protein was methylated/labeled in $^{18}$O-water first and then digested with trypsin. Similar results were found (data not shown), suggesting in this case all isoAsp residues in this particular protein are accessible to the PIMT enzyme.

It is worth noting that isoAsp residues at C- or N-termini are not methylated by PIMT, thus precluding their detection by our approach. On the other hand, it has been shown that isoAsp is refractory to the proteolytic digestion by most endoproteases (such as Asp-N and trypsin)[16, 46, 47], so isoAsp is unlikely to be at the termini of the digested peptides. If such peptides do exist, N-terminal isoAsp can be detected by other methods, such as ETD mass spectrometry, as recently described[48].
2.4.2 Hydrolysis and $^{18}$O Incorporation

As illustrated in Scheme 2-2, the isoaspartate methyl ester spontaneously converts to succinimide which is hydrolyzed in $^{18}$O-water into isoAsp and Asp peptides, resulting in $^{18}$O incorporation. The mass spectra of $^{18}$O-labeled isoAsp-DSIP and IgG1 peptides showed greater than 95% isotope incorporation (Figure 2-1 and Figure 2-3), which is crucial to the subsequent mass spectrometric analysis. Considering that $^{18}$O-water contained ~ 3% $^{16}$O-water, our results again indicate that isoAsp was completely methylated under these conditions.

Buffer pH and hydrolysis time are critical parameters to completely convert succinimide to $^{18}$O-isoAsp/Asp with minimum artifact. As shown in Figure 2-4, the hydrolysis of methyl ester and succinimide was quicker under higher pH conditions, though we suspect that the abundance of deamidation artifacts from sample treatment might be increased at higher pH. Therefore, the hydrolysis conditions were examined, and pH 8.5-9.0, 37 °C and 30 min was found to be optimal. Under these conditions, both isoaspartate methyl ester and succinimide peaks disappeared with the concomitant appearance of $^{18}$O-labeled isoAsp and Asp peptide peaks (see Figure 2-3). Similar to other peptides, the isoAsp species was the preferential hydrolysis product compared to its Asp counterpart in a ratio of about 3:1. The presence of $^{18}$O-labeled Asp peptides does not affect the detection of $^{18}$O-labeled isoAsp peptides and, in fact, may provide secondary confirmation, as $^{18}$O in both isoAsp and Asp species serves as a telltale sign of the existence of isoAsp in the original samples.
Figure 2-4. The effects of pH (A) and incubation time (B) on the hydrolysis of the succinimide and methyl ester. Complete hydrolysis was observed at pH 8.5-9.0 at 37 °C after 30 min.

Since deamidation of Asn and isomerization of Asp may happen spontaneously during the sample handling process, the degree of background reactions should be measured[31, 32]. Prolonged incubation and harsh conditions should be avoided. Under our conditions, no isoAsp
from Asp was detected when the samples were incubated at pH 9, 37 °C for 30 min, and 24 hrs (Figure 2-5). Additionally, deamidation of Asn (even the NG “hot spots”) occurred to a small degree for 30 min, thus having little practical effects on the identification of isoAsp from Asp via tandem mass spectrometry.

Figure 2-5. Stability of Asp-DSIP during sample treatment. Neither isoAsp-DSIP formation nor $^{18}$O-incorporation was observed when Asp-DSIP was exposed to pH 9 buffer at 37 °C for 24 hrs.
The $^{18}$O-labeled isoAsp peptides may be methylated by PIMT again[38], leading to two $^{18}$O incorporation. This phenomenon is similar to protease-catalyzed $^{18}$O-labeling of C-termini of peptides that Fenselau’s group has developed[30, 35, 36]. This however does not affect the database search and site identification of labeled isoAsp. To simplify data analysis, incorporation of two $^{18}$O atoms can be minimized by quenching PIMT activity (e.g., by adding guanidine) after the methylation step, as demonstrated in Figure 2-6. Because of the stoichiometry may not be precisely controlled, such analysis should be treated not as absolute but rather semi-quantitative.
Figure 2-6. Guanidine HCl (Gnd-HCl, 1.25 M) quenched PIMT activity during hydrolysis and thereby minimized the incorporation of two $^{18}$O-atoms into isoAsp peptides. The blue arrow indicates that two $^{18}$O-atoms were incorporated into some peptides when 0.24 M guanidine HCl was used.
2.4.3 Screening of $^{18}$O-Labeled isoAsp by Mass Spectrometry

Compared to unlabeled isoAsp, $^{18}$O-labeling at isoAsp results in a 2 Da mass increase that can be easily detected with standard mass spectrometry (for work flow, see Scheme 2-3). As shown in Figure 2-1A, the $^{18}$O-labeled isoAsp-DSIP peptide was detected as a singly charged ion at m/z 851, 2 Da higher than that from the unlabeled Asp-DSIP peptide (m/z 849). Similarly, a shift of 0.6663 m/z was observed for the triply charged peptide LC 69-108 from IgG1 (Figure 2-1B). The high $^{18}$O-labeling efficiency (near completion) resulted in a clean shift of the isotopic pattern, thereby enabling automatic recognition of the $^{18}$O-labeled isoAsp peptides using a standard data analysis algorithm. For example, multiple isoAsp peptides in IgG1 were readily identified in this manner, some of which are listed in Table 2-1. Therefore, the method reported here is suitable for automatic, high throughput screening of isoAsp. $^{17}$O-water can also be used for labeling, but is less desirable due to its higher cost and smaller mass shift (1 Da) on the labeled peptides.
Table 2-1. Representative isoAsp containing peptides detected in IgG1.

<table>
<thead>
<tr>
<th>Retention Time (min)</th>
<th>Charge</th>
<th>Obs. m/z</th>
<th>Cal. m/z</th>
<th>Peptide Sequence</th>
<th>IsoD Site*</th>
<th>Peak Ratio in the stressed sample (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.41</td>
<td>2</td>
<td>840.4055</td>
<td>840.4073</td>
<td>FNWYV&lt;sup&gt;isoD&lt;/sup&gt;GVEVHNAK&lt;sup&gt;284&lt;/sup&gt;</td>
<td>HC, D276</td>
<td>9.2</td>
</tr>
<tr>
<td>100.07</td>
<td>2</td>
<td>839.4036</td>
<td>839.4052</td>
<td>FNWYV&lt;sup&gt;D&lt;/sup&gt;GVEVHNAK&lt;sup&gt;284&lt;/sup&gt;</td>
<td></td>
<td>90.8</td>
</tr>
<tr>
<td>114.75</td>
<td>2</td>
<td>938.4655</td>
<td>938.4672</td>
<td>TTPVLDS&lt;sup&gt;isoD&lt;/sup&gt;GFSFLYSK&lt;sup&gt;405&lt;/sup&gt;</td>
<td>HC, D397</td>
<td>0.6</td>
</tr>
<tr>
<td>117.12</td>
<td>2</td>
<td>937.4632</td>
<td>937.4651</td>
<td>TTPVLDS&lt;sup&gt;D&lt;/sup&gt;GFSFLYSK&lt;sup&gt;405&lt;/sup&gt;</td>
<td></td>
<td>99.4</td>
</tr>
<tr>
<td>131.24</td>
<td>3</td>
<td>1405.9721</td>
<td>1405.9707</td>
<td>SGTSALITGLQAEDEADYYCQSY&lt;sup&gt;isoD&lt;/sup&gt;SSLGFLVFGTGTK&lt;sup&gt;108&lt;/sup&gt;</td>
<td>LC, D94 (CDR)</td>
<td>1.4</td>
</tr>
<tr>
<td>135.25</td>
<td>3</td>
<td>1405.3058</td>
<td>1405.3026</td>
<td>SGTSALITGLQAEDEADYYCQSY&lt;sup&gt;D&lt;/sup&gt;SSLGFLVFGTGTK&lt;sup&gt;108&lt;/sup&gt;</td>
<td></td>
<td>98.6</td>
</tr>
</tbody>
</table>

*Note: LC, HC, and CDR are light chain, heavy chain, and complementary-determining region, respectively. Cysteine was alkylated.
2.4.4 Co-elution of isoAsp and Asp and Overlapping of Isotope Patterns

As shown in Table 2-1 and Figures 2-2 ~2-6, isoAsp peptide and its Asp counterpart are fully or partially resolved by liquid chromatography. However, occasionally they may co-elute[16, 49]. Under such a scenario, the isotopic envelope of the unlabeled (16O) Asp peptide peak overlaps with that of the labeled (18O) isoAsp species, potentially complicating the analysis of the isotope pattern. Although the mixed isotope patterns can always be deconvoluted[30, 34, 50], as has been done in the analysis of succinimide and deamidation by 18O-labeling[29, 32], the practical issue is whether the intensity from the 18O-species is sufficiently high (e.g., when isoAsp is in low abundance) so that the 2 Da mass shift can be automatically recognized by data analysis algorithms. In addition, the deconvolution of the mixed isotope patterns complicates isoAsp analysis. Remediation includes changing chromatographic conditions for peptide separation or using different proteases to generate isoAsp and Asp peptides with different sequences. A more direct and surefire approach, as we have reported, is to treat the sample with endoprotease Asp-N (EC 3.4.24.33), which cleaves peptides at the N-terminal side of Asp but not isoAsp residues[16, 46, 47]. As such, sample treatment with Asp-N has been used as an effective method to selectively remove Asp peptides from their isoAsp counterparts, enriching isoAsp species for subsequent analysis[16].

2.4.5 Identification of isoAsp Sites in 18O-Labeled Peptides

In addition to screening isoAsp in a given peptide, the precise location of modification can be facilely deduced by tandem mass spectrometry with high confidence, as the specific 18O
incorporation imparts 2 Da mass increases on the fragmentation ions (such as b and y ions in CID mode) that contain isoAsp (Figure 2-7–Figure 2-10). For example, the isoAsp modification site in a tryptic peptide from the IgG1 sample was identified to be at isoAsp276 by its MS/MS data (Figure 2-7). The mass increment of 2 Da corresponding to $^{18}$O incorporation was evident in a series of y9-y12 and b6-b13 ions with normal intensity as peaks from the unlabeled ($^{16}$O) peptide, leading to unambiguous identification of the modification site. In comparison, ETD/ECD mass spectrometry distinguishes isoAsp from Asp peptides based on only a single pair of characteristic reporter ions of isoAsp (c$^{+58}$ and z-57) that also are of various intensity under different conditions[16-20]. As another example, the third peptide in Table 2-1 is located in the light chain (LC) of IgG1 at amino acid positions 69-108 and contains 40 amino acids with three closely positioned Asp residues, posing challenge for site identification. Nevertheless, the modification site was automatically detected and unambiguously established to be isoAsp94 based on the isotopic patterns conferred by $^{18}$O incorporation, again exemplifying the utility of our method.
Figure 2-7. Identification of isoAsp site in a doubly charged tryptic peptide HC 271-284 from the stressed IgG1. The MS/MS spectra with (A) and without $^{18}$O tag (B) were obtained by collision induced dissociation (CID) of the (M+2H)$^{2+}$ precursor ions m/z 840.41 and 839.40 for the top and bottom traces, respectively. Cysteine is alkylated.
Figure 2-8. Identification of isoAsp site in a triply charged tryptic peptide LC69-108 from the IgG1 sample. The MS/MS spectra with (A) and without $^{18}$O tag (B) were obtained by collision-induced dissociation (CID) of the (M+3H)$^{3+}$ precursor ions, m/z 1406.64 and 1405.97, for the top and bottom traces, respectively. Cysteine was alkylated.
Figure 2-9. Identification of isoAsp site in a doubly charged tryptic peptide HC389-405 from the IgG1 sample. The MS/MS spectra with (top) and without $^{18}$O tag (bottom) (A) were obtained by collision-induced dissociation (CID) of the (M+2H)$^{2+}$ precursor ions, m/z 938.97 and 937.97, for top and bottom trace, respectively. (B) and (C) are the zoomed-in views.
Figure 2-9. Continued.
Figure 2-9. Continued.
Figure 2-10. Identification of isoAsp site in the stressed Asp-DSIP peptide by tandem mass spectrometry. The MS/MS data of isoAsp-DSIP with $^{18}$O tag (A) and Asp-DSIP without $^{18}$O tag (B) were obtained by collision induced dissociation (CID) of the (M+H)$^{+}$ precursor ions m/z 851 and 849 for the top and bottom traces, respectively.
2.5 Conclusions

We present herein an approach combining chemo-enzymatic transformations to specifically label isoAsp with $^{18}$O for facile analysis by standard mass spectrometry and routine data analysis algorithms. The complete incorporation of $^{18}$O into isoAsp makes it feasible for quantitative analysis; this project is currently under development in our laboratories.

2.6 References

[1] Noguchi S. Structural changes induced by the deamidation and isomerization of asparagine revealed by the crystal structure of Ustilago sphaerogena ribonuclease U2B. Biopolymers 2010;93:1003-10.


Chapter 3: Discovery of Undefined Protein Crosslinking Chemistry: A Comprehensive Methodology Utilizing $^{18}$O-labeling and Mass Spectrometry


Co-authors’ work in this chapter: Min Liu: experimental design and execute, data analysis, manuscript writing and revision; Zhonqi Zhang: new $^{18}$O-screening function in MassAnalyzer algorithm, data analysis, manuscript writing and revision; Tiazhu Zang: data analysis, manuscript writing and revision; Chris Spahr: data analysis, manuscript writing and revision; Janet Cheetham: idea contribution, manuscript writing and revision, grant support; Da Ren: idea contribution, experimental design, manuscript writing and revision; Zhaohui Sunny Zhou: idea contribution, experimental design, data analysis, manuscript writing and revision and grant support.
3.1 Abstract

Characterization of protein crosslinking, particularly without prior knowledge of the chemical nature and site of crosslinking, poses a significant challenge due to their intrinsic structural complexity and the lack of a comprehensive analytical approach. Towards this end, we have developed a generally applicable workflow—XChem-Finder that involves four stages. (1) Detection of crosslinked peptides via $^{18}$O-labeling at C-termini. (2) Determination of the putative partial sequences of each crosslinked peptide pair using a fragment ion mass database search against known protein sequences coupled with a de novo sequence tag search. (3) Extension to full sequences based on protease specificity, the unique combination of mass, and other constraints. (4) Deduction of crosslinking chemistry and site. The mass difference between the sum of two putative full-length peptides and the crosslinked peptide provides the formulas (elemental composition analysis) for the functional groups involved in each crosslinking. Combined with sequence restraint from MS/MS data, plausible crosslinking chemistry and site were inferred, and ultimately, confirmed by matching with all data. Applying our approach to a stressed IgG2 antibody, ten cross-linked peptides were discovered and found to be connected via thioether originating from disulfides at locations that had not been previously recognized. Furthermore, once the crosslink chemistry was revealed, a targeted crosslink search yielded four additional crosslinked peptides that all contain the C-terminus of the light chain.
3.2 Introduction

Protein crosslinking exists in a myriad of biological systems and protein pharmaceuticals, such as collagen, ubiquitylated proteins, and monoclonal antibodies[1-6]. Rich and diverse chemistry is involved as well, including disulfide[1], dityrosine[2], lysinoalanine[7, 8], lanthionine[7, 8], etc. Additionally, chemical crosslinking is widely used to probe protein structures and interactions[9, 10]. Due to their intrinsic structural complexity, characterization of crosslinked peptides is complex, but nonetheless tractable if the crosslink chemistry is pre-defined. For example, a database of the intact mass (precursor ion) and the tandem mass spectra (fragmentation ions) for all possible combinations of crosslinked peptides (e.g., two cysteines to form a disulfide bond) can be generated computationally, and subsequently, correlated with observed spectra to identify both the sequences and sites of modification. Such a database search strategy is the cardinal principle behind many common algorithms, including ASAP[11], X!link[12], BLink[13], Xlink-Identifier[14, 15] and MassAnalyzer[16]. Moreover, clever experimental tricks, such as judicious isotope labeling[17, 18], can markedly simplify the process and enhance the confidence level for assignment with the assistance of software tools e.g., Pro-Cross-link[19, 20], PepLynx[21], xQuest[22], iXlink/doXLink/XlinkViewer[23]. To date, the rapid advancements in mass spectrometers, data analysis algorithms, and computational capacity have made analyses of crosslinking with known chemistry much more accessible if not routine (for recent reviews, see [9, 10]).

Yet the aforementioned approaches are futile if the crosslink chemistry is unknown or not pre-defined; for one thing, no theoretical mass or spectrum can be simulated. Even if crosslinked peptides have been identified, it remains a tall order to deduce the sequences and sites of
crosslinking. Conceptually, de novo sequencing should provide at least partial sequences for crosslinked peptides (see review paper[24]). Under typical fragmentation conditions, however, a crosslinked peptide gives rise to at least five sets of b- and y-ions that are intertwined and indistinguishable. In addition, high-charge-states (≥3+) are typically featured in the crosslinked peptides, resulting in multiple charge fragment ions (e.g. 2+ or 3+) and further complicating data interpretation[12, 22, 25]. High resolution mass spectrometers (e.g. Orbitrap), capable of the determination of fragment ion charge state, have become widely available only recently. As such, the drastically increased complexity in tandem spectrum renders de novo sequencing ineffective in most cases. Unknown or undefined crosslinks are typically discovered serendipitously, requiring isolation of the crosslinked peptides and “old-fashioned” protein chemistry. Even so, full characterization remains elusive for many cases. For instance, the non-reducible crosslinks between an IgG heavy chain and a light chain in a murine monoclonal antibody, OKT3, and between two heavy chains of IgG2 could not be elucidated even after intensive efforts[26, 27].

To facilitate systematic and unbiased discovery of unknown crosslinks, we have developed a generally applicable workflow—XChem-Finder (Scheme 3-1). First, crosslinked peptides were isotopically labeled at the C-termini to facilitate their detection[19-21, 28]. Proteins were digested in 18O-(heavy) and 16O-(light) water, respectively, followed by LC/MS/MS analysis. At full scan, the distinct isotope pattern of the crosslinked peptides (a mass increase of 8 Da) compared to the non-crosslinked linear species (a mass increase of 4 Da) was readily detected by a spectral analysis algorithm[19-21]. The second and more challenging part is to determine the sequences, chemical nature and site of crosslink. The workflow breaks down the challenge into workable sub-steps. (a) The candidate ions of crosslinked peptides underwent
high resolution MS/MS analysis. Based on their isotope patterns, linear and crosslinked fragment ions are divided into different groups (Table 3-1). (b) Mass of linear fragment ions were searched against the protein sequence, yielding partial sequences (often sequence ladders) of each chain of the crosslinked peptides. In parallel, de novo sequencing of crosslinked fragment ions affords sequence tags. (c) Combining the partial sequences and sequence tags, putative full-length sequences of each chain were deduced based on protease specificity, the unique combination of mass, and other constraints. (d) The difference between the combined mass of the two putative full-length peptides and the observed mass of a crosslinked peptide provides the formula for the functional group involved in the crosslink (mass to formula). Combined with sequence restraint from MS/MS data, the crosslink chemistry and site were inferred, and ultimately, confirmed by matching with all data.

Applying our XChem-Finder approach to a stressed IgG2, ten crosslinked peptides were discovered and found to be linked via thioether that originated from disulfides at locations that had not been reported. Furthermore, once the crosslinking chemistry was revealed, a targeted search yielded additional four crosslinked peptides that all contain the C-terminus of light chain.

3.3 Experimental Section

3.3.1 Chemicals

All chemicals were reagent grade or above. Guanidine hydrochloride (GndHCl), dithiothreitol (DTT), iodoacetic acid (IAA), trifluoroacetic acid (TFA), acetonitrile (ACN), HPLC-grade water, and bradykinin were from Sigma-Aldrich (St. Louis, MO, USA).
Sequencing grade trypsin was obtained from Roche (Indianapolis, IN, USA). $^{18}$O-water (97%) was obtained from Cambridge Isotope Laboratories (Andover, MA, USA). Recombinant monoclonal antibody anti-streptavidin immunoglobulin gamma 2 (IgG2) was produced in Chinese hamster ovary (CHO) cells (Amgen, Thousand Oaks, CA, USA), purified according to standard manufacturing procedures, formulated at a concentration of 20 mg/mL in 50 mM sodium acetate pH 5.2, and stored at -70 °C.

<table>
<thead>
<tr>
<th>Stage 1: Detection of Crosslinked Peptides</th>
<th>Stage 2: Determination of Partial Sequences</th>
<th>Stage 3: Inference of Full Sequences</th>
<th>Stage 4: Deduction of Crosslink Site &amp; Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptic digestion in 16O- and 18O-water Group fragment ions via mass shift</td>
<td>0 or +4 Da</td>
<td>+8 Da</td>
<td>Protease specificity &amp; other constraints</td>
</tr>
<tr>
<td>(linear fragment ions) (crosslinked fragment ions)</td>
<td>Match mass of fragment ions with peptide sequences</td>
<td>de novo sequencing (manual)</td>
<td>Elementary composition analysis of mass difference (combined native linear chains vs crosslink peptide)</td>
</tr>
<tr>
<td>Mass shift of 8 Da Partial peptide sequences; often sequence ladders</td>
<td>Sequence tag</td>
<td>Putative full sequences of both chains</td>
<td>Confirmed structure (Thioether)</td>
</tr>
</tbody>
</table>

Scheme 3-1. Flow chart of XChem-Finder in four main stages: (1) detection of crosslink by $^{18}$O-labeling; (2) determination of partial sequences of crosslinked peptides from the mass of fragmentation ions; (3) inference of full sequence of each chain; (4) determination of crosslink site and chemistry by elemental composition analysis and chemical intuition.
Table 3-1. Fragmentation ions of a cross-link peptide. $^{18}$O-labeling at C-termini were shown in red. Each letter (e.g., abc and XYZ) represents one amino acid residue. The symbol (?) denotes unknown cross-link chemistry. Five—and intertwined—sets of b- and y-ions are in a cross-link peptide. The ions with consecutive bond dissociations or internal fragments are excluded due to lower abundance in CID.

<table>
<thead>
<tr>
<th>Group number</th>
<th>Structure</th>
<th>$^{18}$O mass shift (Da)</th>
<th>Linear or cross-linked</th>
<th>Searchable (Match partial sequence)</th>
<th>Sequence Tags (De novo sequence)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NH$_2$abc</td>
<td>0</td>
<td>Linear</td>
<td>Yes</td>
<td>Yes</td>
<td>b-ions, single chain</td>
</tr>
<tr>
<td>2</td>
<td>NH$_2$abc</td>
<td>+4</td>
<td>Linear</td>
<td>Yes</td>
<td>Yes</td>
<td>y-ions, single chain</td>
</tr>
<tr>
<td>3</td>
<td>NH$_2$abc</td>
<td>+4</td>
<td>Cross-linked</td>
<td>No</td>
<td>Yes</td>
<td>b-ions, cross-linked</td>
</tr>
<tr>
<td>4</td>
<td>NH$_2$abc</td>
<td>+8</td>
<td>Cross-linked</td>
<td>No</td>
<td>Yes</td>
<td>y-ions, cross-linked</td>
</tr>
<tr>
<td>5</td>
<td>NH$_2$abc</td>
<td>+4</td>
<td>Linear</td>
<td>No</td>
<td>No</td>
<td>Modified single chain from cleavage of the cross-linking</td>
</tr>
</tbody>
</table>

Precursor Ion

NH$_2$ABCD-N-WXYZ-CONH$_2$

NH$_2$-ABC-N-WXYZ-CONH$_2$
3.3.2 Generation of Stressed Sample

After being buffer exchanged into 100 mM Tris at pH 8.5, the IgG2 antibody was incubated at 50 °C for 7 days in the dark.

3.3.3 Reduction, Alkylation, Tryptic Digestion and $^{18}$O-Labeling of the IgG2

Tryptic digestion of the stressed IgG2 was performed similarly to the procedure described by Ren et al[29]. Briefly, IgG2 (20 mg/mL) was diluted to 1 mg/mL in a denaturing buffer (7.5 M GndHCl, 2 mM EDTA and 0.25 M Tris-HCl, pH 7.5) to a final volume of 0.5 mL. Reduction was accomplished with the addition of 3 μL of 0.5 M DTT followed by 30 min incubation at room temperature. S-Carboxymethylation was achieved with the addition of 7 μL of 0.5 M IAA; the reaction was carried out in the dark for 15 min at room temperature. Excess IAA was quenched with the addition of 4 μL of 0.5 M DTT. The reduced and alkylated IgG2 samples were subsequently exchanged into the digestion buffer (0.1 M Tris-HCl at pH 7.5) using a NAP-5 size-exclusion column (GE Healthcare, Piscataway, NJ, USA). After two aliquots (200 μL each) of the above buffer-exchanged antibody were completely dried via Speed Vac and reconstituted separately into the same volume of $^{18}$O-water or $^{16}$O-water, 6 μL of 1 mg/mL trypsin in $^{18}$O-water or $^{16}$O-water solution, respectively, was added to achieve a 1:25 (w/w) enzyme/substrate ratio. The reaction mixtures were incubated at 37 °C for 30 min.
3.3.4 HPLC

Tryptic digests of the IgG2 (25 μL) were separated on a Jupiter C5 column (250 x 2.0 mm, 5 μm, 300Å, Phenomenex, Torrance, CA, USA) at a temperature of 50 °C with a flow rate of 200 μL/min on a HPLC system (Agilent 1100, Palo Alto, CA, USA). Mobile phase A was 0.1% TFA in water (v/v) while mobile phase B contained 0.085% TFA / 90% ACN / 10%water. A gradient was applied by holding at 2% B for 2 min, increasing to 22% B in 38 min, then 42% B in 80 min, then 100% B in 25 min followed by holding at 100% B for 5 min. The column was re-equilibrated at 2% B for 30 min before next injection.

3.3.5 Mass Spectrometry

An LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) was used in-line with the HPLC system for the analyses of the IgG2 tryptic digests. A full MS scan (with 60,000 resolution at m/z 400 and an automatic gain control (AGC) target value of 2x10^5) followed by data-dependent MS/MS scans of the three most abundant precursor ions was set up to acquire both the peptide mass and sequence information. The spray voltage was 5.5 kV, and the capillary temperature was 250 °C. The instrument was tuned using the doubly-charged ion of a synthetic peptide, bradykinin. The MS/MS spectra were obtained using collision-induced dissociation (CID) with normalized collision energy of 35%. For MS/MS with ion detection in the Orbitrap, the AGC target was set to 3x10^6, resolution to 7,500, and the precursor isolation width to 4 m/z unit. Peptides were identified by MassAnalyzer by comparing experimental MS/MS to theoretically predicted MS/MS[16, 30-32]. Peak alignment between 16O- and 18O-
digest runs was automatically performed by MassAnalyser[33]. A new function was implemented in MassAnalyzer to calculate the level of $^{18}$O-labeling in each peptide. The number of incorporated $^{18}$O in a peptide is calculated from the following equation:

$$N_{^{18}O} = \frac{M_{\text{labeled}} - M_{\text{unlabeled}}}{2.004}$$

Eq. 1

where $M_{\text{labeled}}$ and $M_{\text{unlabeled}}$ are the average masses of the $^{18}$O-labeled and unlabeled peptides, respectively, as calculated by the centroids of their respective isotope envelopes. The value of 2.004 Da is the mass difference between an $^{18}$O atom and an $^{16}$O atom.

### 3.4 Results and Discussion

#### 3.4.1 Stage 1: Identification of Crosslinked Peptides.

$^{18}$O-labeling combined with mass spectrometry is commonly used to identify crosslinked peptides [19-21, 28]. As shown in Scheme 3-1, newly created C-termini of tryptic peptides from digestion in $^{18}$O-water were completely labeled by $^{18}$O. The distinct isotope pattern for the labeled crosslinked peptides (a mass increase of 8 Da) compared to linear (non-crosslinked) species (a mass increase of 4 Da) can be automatically detected by common spectral analysis algorithms, such as an in-house isotopic screening algorithm (MassAnalyzer[16, 30-32]).

$^{18}$O-Labeling. A general strategy to label the C-termini of peptides in $^{18}$O-water catalyzed by proteases for peptide identification and quantification is well documented[34-40]. Under our experimental conditions, near complete (four) $^{18}$O-incorporation for crosslinked
peptides was evident from the isotopic distributions (see Figure 3-1). The small amount of $^{16}$O-water (3%) in $^{18}$O-water had no significant impact in their isotopic patterns and the subsequent data analysis. $^{18}$O-labeling during tryptic digestion is only applied for newly created C-termini, not the C-termini of proteins. Hence, a crosslinked tryptic peptide that contains the C-terminus of the protein only has a mass shift of 4 Da, therefore cannot be differentiated from the linear peptides. This limitation can be overcome by using proteases with different substrate specificity or labeling N-termini (e.g., formaldehyde-$d_2$ and sodium cyanoborohydride or succinic anhydride-$d_4$)[41-43]. In this paper, this was satisfactorily addressed via a targeted mass search after the crosslink chemistry was elucidated. In addition, the deamidation of asparagine and isomerization of aspartic acid could potentially introduce $^{18}$O into peptides[44-46]; under our conditions, no isoaspartic acid was detected in the candidate peptides.

Figure 3-1. Isotopic distributions of the cross-linked peptide HC:G118-R129/HC:C215-K240 (RT at 91.17min). A mass shift of 8 Da was observed indicating four $^{18}$O-incorporation for the crosslinked peptide.
Screening of Crosslinked Peptides in Full Scan. An $^{18}\text{O}$ incorporation value (Eq. 1) of 4.0 ± 0.3 was set as cut-off in our screening. The initial screening results for the stressed IgG2 are shown in Table 3-2. Each peak was evaluated for false positive. For instance, gas phase dimerization, commonly observed in mass spectrometry[47, 48], was readily determined based on retention time (same as the monomers) and mass (exactly double that of monomers). In addition, weak precursor ions (typically with peak intensity of 50,000 count or lower) with poor or no MS/MS data were excluded. Based on these criteria, ten candidates shown in red in Table 3-2 were selected for subsequent high resolution MS/MS analysis.
Table 3-2. The cross-linked peptide candidates identified by MassAnalyzer algorithm. False hits (gas phase dimers) and weak ions (with poor or no MS/MS data) were excluded from the subsequent analysis. Ten cross-linked peptide candidates shown in red were selected for high resolution MS/MS analysis in next step. For the ions with multiple charges, the one with the highest intensity was examined. $N^{18}_O$ is the number of incorporated $^{18}$O.

<table>
<thead>
<tr>
<th>Charge</th>
<th>Retention Time (RT, min)</th>
<th>m/z</th>
<th>Average Mass (Da)</th>
<th>Monoisotopic Mass (Da)</th>
<th>Intensity</th>
<th>$N^{18}_O$</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>39.93</td>
<td>1119.24</td>
<td>3355.58</td>
<td>3353.699</td>
<td>575968</td>
<td>3.86</td>
<td>Q(-17)SVVTQPPSVSGAPGQR(dimer)</td>
</tr>
<tr>
<td>3</td>
<td>60.37</td>
<td>1163.58</td>
<td>3487.76</td>
<td>3485.707</td>
<td>511960</td>
<td>3.84</td>
<td>YAASSYLSLTPEQWK(dimer)</td>
</tr>
<tr>
<td>3</td>
<td>76.76</td>
<td>1655.45</td>
<td>4960.313</td>
<td>4960.313</td>
<td>216540</td>
<td>3.82</td>
<td>NQFSLETSVTAADTAVYYCAR(dimer)</td>
</tr>
<tr>
<td>4</td>
<td>87.91</td>
<td>1045.77</td>
<td>4177.057</td>
<td>4177.059</td>
<td>93846</td>
<td>3.78</td>
<td>Same as m/z1394.03</td>
</tr>
<tr>
<td>3</td>
<td>87.91</td>
<td>1394.03</td>
<td>4179.80</td>
<td>4179.80</td>
<td>386639</td>
<td>3.81</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>3</td>
<td>88.56</td>
<td>1413.36</td>
<td>4235.064</td>
<td>4235.064</td>
<td>796803</td>
<td>3.82</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>4</td>
<td>88.57</td>
<td>1060.53</td>
<td>4235.065</td>
<td>4235.065</td>
<td>175997</td>
<td>3.85</td>
<td>Same as m/z1413.36</td>
</tr>
<tr>
<td>5</td>
<td>88.59</td>
<td>848.42</td>
<td>4235.065</td>
<td>4235.065</td>
<td>85712</td>
<td>3.85</td>
<td>Same as m/z1413.36</td>
</tr>
<tr>
<td>3</td>
<td>91.17</td>
<td>1351.33</td>
<td>4048.963</td>
<td>4048.963</td>
<td>205489</td>
<td>3.85</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>4</td>
<td>96.15</td>
<td>1605.54</td>
<td>6414.117</td>
<td>6414.117</td>
<td>159092</td>
<td>3.77</td>
<td>Cross-link candidate, similar to m/z1620.04(-58)</td>
</tr>
<tr>
<td>4</td>
<td>97.39</td>
<td>1620.04</td>
<td>6472.120</td>
<td>6472.120</td>
<td>168467</td>
<td>3.86</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>4</td>
<td>97.88</td>
<td>1484.22</td>
<td>5928.854</td>
<td>5928.854</td>
<td>424306</td>
<td>3.79</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>4</td>
<td>98.49</td>
<td>1475.73</td>
<td>5898.65</td>
<td>5898.65</td>
<td>71227</td>
<td>3.92</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>4</td>
<td>98.99</td>
<td>1498.47</td>
<td>5986.858</td>
<td>5986.858</td>
<td>303560</td>
<td>3.82</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>6</td>
<td>99.00</td>
<td>999.32</td>
<td>5986.854</td>
<td>5986.854</td>
<td>53191</td>
<td>3.87</td>
<td>Same as m/z1498.47</td>
</tr>
<tr>
<td>4</td>
<td>99.76</td>
<td>1452.20</td>
<td>5800.765</td>
<td>5800.765</td>
<td>401159</td>
<td>3.78</td>
<td>Cross-link candidate</td>
</tr>
<tr>
<td>3</td>
<td>99.78</td>
<td>1935.93</td>
<td>5800.758</td>
<td>5800.758</td>
<td>176055</td>
<td>3.80</td>
<td>Same as m/z1452.20</td>
</tr>
<tr>
<td>3</td>
<td>101.56</td>
<td>1893.23</td>
<td>5673.664</td>
<td>5673.664</td>
<td>58750</td>
<td>3.83</td>
<td>Miss monoisotopic peak. same as m/z1419.92</td>
</tr>
<tr>
<td>4</td>
<td>101.59</td>
<td>1419.92</td>
<td>5672.670</td>
<td>5672.670</td>
<td>66658</td>
<td>3.89</td>
<td>Cross-link candidate</td>
</tr>
</tbody>
</table>
3.4.2 Stage 2: Deduce Partial Sequence for Each Chain.

As illustrated in Scheme 3-1, this stage involves (a) grouping fragment ions based on their isotope patterns imparted by their corresponding structural features (e.g., linear or cross-linked), (b) deducing partial peptide sequences via a database search (match mass with partial peptide sequences using FindPept) and de novo sequencing, and (c) determining most likely candidate peptides.

Deconvolution of Fragment Ions. Most precursor ions for cross-linked peptides were highly charged (e.g., 3+ or 4+), thus doubly and triply charged fragment ions abound, e.g. ion m/z 839.49 (2+) and 1300.47 (3+) in Figure 3-2. The high resolution of the tandem mass spectrum allowed us to measure the isotope envelope and hence determine the charge state. Also considering fragment ion type (b- vs y-ion), monoisotopic neutral mass of each fragment ion from a crosslinked peptide was calculated manually. For example, +17.0033 Da (the mass of OH\(^{-}\)) and -1.0073 Da (the mass of H\(^{+}\)) were added to a singly charged b- and y-ion, respectively, to obtain their neutral peptide mass. The high-resolution for the tandem spectra was crucial in determining the correct charge state and hence neutral mass; otherwise, incorrect monoisotopic mass would lead to false hits and even erroneous assignment.

Grouping Fragment Ions by \(^{18}\text{O}\) Incorporation. The fragment ions containing zero, one, and two C-termini displayed a mass shift of 0, 4, and 8 Da, respectively, in the corresponding MS/MS spectra obtained from \(^{18}\text{O}\)-water vs \(^{16}\text{O}\)-water (referred as \(^{18}\text{O}/^{16}\text{O}\) rule in this paper) and accordingly, are divided into different groups (Table 3-1). For each crosslinked peptide, two sets of linear fragment ions that contain no crosslink site do exist for each chain.
One set is the b-ions prior to the crosslink site, which show no mass shifts with $^{18}$O-labeling and thus are separated from other fragment ions (group 1 in Table 3-1). Another set is the y-ions to the C-terminal side of the crosslink site (group 2 in Table 3-1), which contain two $^{18}$O with a mass shift of 4 Da. Essentially, these linear fragments are searchable in standard database, i.e., the mass can be matched with the corresponding peptide fragments. The freely available FindPept (web.expasy.org/findpept/) was used for the search in this study. Each observed mass value of these linear fragment ions should match to a partial sequence of the crosslinked peptides, but also unrelated sequences (false hits). High mass accuracy (typical 10 ppm in our FT MS/MS experiments) greatly limits false positives. Furthermore, multiple fragmentation ions collectively—and in combination with de novo sequencing as described below—narrow the hits to a selected few, if not one, candidate peptides.

Isotope pattern (8 Da mass shift with $^{18}$O-labeling) can also be readily used to isolate a set of fragmentation ions that contain two C-termini (y-ions containing the crosslink site, see group 4 in Table 3-1). First, these ions were excluded from database search (which is for linear peptides), reducing false hits. Second and more importantly, this markedly simplified set of tandem spectra could be used for de novo sequencing to yield sequence tags, as it was indeed the case in our study.

**Partial Sequence Search via FindPept.** The neutral peptide monoisotopic mass (obtained from the fragment ion bins of the mass shift of 0 and 4 Da, linear peptides, as shown in Scheme 3-1) were searched against the known IgG2 sequence using FindPept with user-defined mass error (10 ppm for the resolution of 7500 in FT-MS/MS in our experiments). FindPept also allows users to define the residue modifications, for example, alkylation at all cysteine residues
(+58.005 Da for reaction with iodoacetic acid in our experiments). FindPept outputs a list of peptides that match the neutral peptide masses, and naturally, some are false hits. As such, several complimentary steps (constraints) were taken to confirm the actual sequences (higher probability and confidence level) and rule out false hits. It is worth noting that this is an iterative process, so the steps can be taken in a different order based on individual situation.

As an example, the process is demonstrated using a triply charged crosslinked peptide m/z 1351.33 (retention time at 91.17 min, G118-R129/C215-K240). The corresponding neutral monoisotopic mass of its fragment ions were searched against the IgG2 sequence. The full list of fragment ion peptides is shown in Table 3-3 and 3-4 and some are highlighted in Table 3-5.

A rewarding first step is to sort the peptides according to their positions in the full protein sequence. As illustrated in Table 3-3 and 3-5, typically, at least one sequence ladder could be readily identified. For example, the overwhelmingly large numbers of fragment ions (eight peptides #3-10 in Table 3-5) that share C-terminal sequences (CPPCPAPPVAGPSVFLFPPKPK) were found, essentially affirming this is part of the true sequence. An immediate implication is that the largest observed fragment ion (2361.190 Da) sets an upper limit for the mass of the other chain (1687 Da) by subtraction from the observed total crosslink peptide mass (4048.963 Da). Based on this criterion, fourteen peptides in Table 3-3 with a mass of significantly larger than 1687 Da were excluded.

Another two powerful constraints that can be applied to data analysis are based on the protease specificity (referred to as the tryptic rule) and mass shift conferred by $^{18}$O-labeling ($^{18}$O/$^{16}$O rule). For instance, the above mentioned eight peptides (peptides 3-10 in Table 3-5) are likely from a tryptic peptide as they all end with lysine, and indeed, a mass shift of 4 Da was observed for all the fragmentation ions from digestion in heavy and light water. Similarly, the
other two overlapping partial sequences are likely the N-terminal fragments of a single tryptic peptide containing GPSVFPLA; and again, as expected, no mass shift was observed from $^{18}$O-labeling. Conversely, false hits can be ruled out; for example, the doubly charged fragment ion m/z 1101.0885 (Table 3-4) matches four peptide sequences (W)GQGTLVTSSASTKGPSVFPLAP(C), (T)APKLLIYGNRSVPDSPGVPDRF(S), (Y)WGQGTLVTSSASTKGPSVFPL, (C)PPCPAPPVAGPSVFLFPPKPK/(D). Since a mass shift from $^{18}$O-labeling was observed, an internal fragment was ruled out, and therefore, this leaves only the last sequence with a C-terminal lysine as the only plausible choice.

At this point, the fragment ion mass search data indicated the peptide at m/z 1351.33 highly likely contains CPPCPAPPVAGPSVFLFPPKPK. For the second chain, although mass search of two b-ions suggests the presence of GPSVFPLA, additional data were warranted for higher confidence level in the assignment as described next.

**De Novo Sequencing.** This compliments nicely with the database search and afford sequence tags[25]. As shown in Scheme 3-1, the identification of the sequence tags was conducted using the crosslinked y-ions (8 Da mass shift in $^{18}$O-digest, group 4 in Table 3-1), which obviously would not match any single chain peptides in the database. In Table 3-6, the observed m/z value is from the most abundant isotopic peak in each isotopic envelope because the monoisotopic peak is weak for large ions at low level. The mass difference between a pair of adjacent y-ions was calculated and compared manually to the mass of single amino acids and dipeptides within a mass error of 0.05 Da. Matching single amino acid residues or dipeptides are shown in red. The sequence tag SVFPLA was confirmed in the crosslink peptide chain G118-
R129, lending a strong support to the existence of peptide chain G118-R129 as a component of the crosslinked peptide (Table 3-6).

In summary, the peptide chain C219-K240 (\textsuperscript{219}CPPCPAPPVAGPSVFLFPK\textsuperscript{240}) and G118-A125 (\textsuperscript{118}GPSVFPLA\textsuperscript{125}) was identified at the end of Stage 2 as parts of the crosslinked peptide of m/z 1351.33.

3.4.3 Stage 3: Inference of Full Sequence for Each Chain

**Extension to the Putative Full Sequences.** Because the peptides were generated by trypsin digestion, the putative partial sequences of crosslinked peptide chains were extended to their corresponding full tryptic peptides (G118-R129, C215-K240, V294-K314) with mass of 1287.6282, 2911.3305, and 2502.2941 Da, respectively (Table 3-5). The mass difference between the observed intact crosslinked peptide (4048 Da) and the first tryptic peptide C215-K240 (2911 Da) is 1137 Da. This narrowed down the second crosslink chain to G118-R129 (1287.6282 Da) while a putative tryptic peptide V294-K314 (2502.2941 Da) is too large (combined mass) to be the second chain. This leaves peptide G118-R129 as an only plausible choice to pair with C215-K240. We were mindful that mis-cleavage might happen, which would be considered if the initial inference did not yield correct assignment.
Table 3-3. Peptides in IgG2 that match with the mass of fragment ions of the triply charged ion m/z 1351.33 (RT at 91.17 min) via FindPept. All Cys are alkylated with IAA. The adjoining residues before cleavage are in parenthesis. NA means not available. The peptides were sorted in the order of primary sequence number. The amino acid position 1-439 and 440-657 was for HC and LC, respectively.

<table>
<thead>
<tr>
<th>m/z</th>
<th>Charge</th>
<th>¹⁸O-water</th>
<th>User mass (Da)</th>
<th>Theor mass (Da)</th>
<th>Δmass (ppm)</th>
<th>peptide</th>
<th>Position</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.173</td>
<td>0.9</td>
<td>(Q)ESGPGVLKPSGTLTCAVS GGIS(S)</td>
<td>6-30</td>
<td>Exclude (Tryptic rule, ¹⁸O/¹⁶O rule, too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>1188.6063</td>
<td>3</td>
<td>NA</td>
<td>3580.807</td>
<td>3580.817</td>
<td>2.9</td>
<td>(W)SWVRQPPGKLEWIGEISHS GTTNYPNLKSR/(V)</td>
<td>36-67</td>
<td>Exclude (too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>1188.6063</td>
<td>3</td>
<td>NA</td>
<td>3580.807</td>
<td>3580.776</td>
<td>-8.7</td>
<td>(I)GEISHGTTTNYPNLKSRVT I5GDKSKNFLSLE(L)</td>
<td>50-82</td>
<td>Exclude (Tryptic rule, too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>1101.0885</td>
<td>2</td>
<td>2</td>
<td>2218.171</td>
<td>2218.163</td>
<td>-4.3</td>
<td>(Y)WGQGTVDVTSSAKGPSVF PL(A)</td>
<td>103-124</td>
<td>Exclude (Tryptic rule, ¹⁸O/¹⁶O rule, too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>1101.0885</td>
<td>2</td>
<td>2</td>
<td>2200.162</td>
<td>2200.174</td>
<td>5.4</td>
<td>(W)GGQTVDVTSSAKGPSVFLAP(C)</td>
<td>104-126</td>
<td>Exclude (Tryptic rule, ¹⁸O/¹⁶O rule, too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.189</td>
<td>7.4</td>
<td>(W)GGQTVDVTSSAKGPSVFLAP(C)</td>
<td>104-127</td>
<td>Exclude (Tryptic rule, ¹⁸O/¹⁶O rule. Too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>470.2368</td>
<td>NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>5.1</td>
<td>(K)/GPSVF(P)</td>
<td>118-122</td>
<td>-</td>
</tr>
<tr>
<td>769.4172</td>
<td>1</td>
<td>0</td>
<td>786.420</td>
<td>786.428</td>
<td>9.6</td>
<td>(K)/GPSVFPLA(P)</td>
<td>118-125</td>
<td>-</td>
</tr>
<tr>
<td>1333.6070</td>
<td>2</td>
<td>2</td>
<td>2683.210</td>
<td>2683.218</td>
<td>3.1</td>
<td>(V)TVPSNFQTQYTCNVKHSTK V/(V)</td>
<td>183-206</td>
<td>Exclude (too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>1341.1364</td>
<td>2</td>
<td>NA</td>
<td>2680.258</td>
<td>2680.258</td>
<td>0.1</td>
<td>(T)YTCNVKHSTK V/(V)</td>
<td>194-215</td>
<td>Exclude (Tryptic rule, too large as 2ⁿ chain)</td>
</tr>
<tr>
<td>769.4172</td>
<td>1</td>
<td>0</td>
<td>804.431</td>
<td>804.434</td>
<td>3.8</td>
<td>(S)NTKVDK(T/V)</td>
<td>204-210</td>
<td>Exclude (Tryptic rule)</td>
</tr>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.19</td>
<td>8.0</td>
<td>(E)CPPCPAPPVAGPSVFLPPK PK/(D)</td>
<td>219-240</td>
<td>-</td>
</tr>
<tr>
<td>1101.0885</td>
<td>2</td>
<td>2</td>
<td>2200.162</td>
<td>2200.175</td>
<td>6.0</td>
<td>(C)CPPCPAPPVAGPSVFLPPKPK/(D)</td>
<td>220-240</td>
<td>-</td>
</tr>
<tr>
<td>1846.0661</td>
<td>1</td>
<td>4</td>
<td>1845.059</td>
<td>1845.055</td>
<td>-2.0</td>
<td>(C)APPVAGPSVFLPPKPK/(D)</td>
<td>223-240</td>
<td>-</td>
</tr>
<tr>
<td>923.5283</td>
<td>2</td>
<td>2</td>
<td>1845.042</td>
<td>1845.055</td>
<td>7.1</td>
<td>(C)APPVAGPSVFLPPKPK/(D)</td>
<td>223-240</td>
<td>-</td>
</tr>
<tr>
<td>839.4851</td>
<td>2</td>
<td>2</td>
<td>1676.956</td>
<td>1676.965</td>
<td>5.5</td>
<td>(A)PPVAGPSVFLPPKPK/(D)</td>
<td>225-240</td>
<td>-</td>
</tr>
<tr>
<td>Protein Mass</td>
<td>Start</td>
<td>End</td>
<td>Peptide Mass 1</td>
<td>Peptide Mass 2</td>
<td>Peptide</td>
<td>Charge</td>
<td>Retention Time</td>
<td>Comments</td>
</tr>
<tr>
<td>-------------</td>
<td>-------</td>
<td>-----</td>
<td>---------------</td>
<td>---------------</td>
<td>---------</td>
<td>--------</td>
<td>---------------</td>
<td>----------</td>
</tr>
<tr>
<td>1384.7951</td>
<td>1</td>
<td>4</td>
<td>1383.788</td>
<td>1383.791</td>
<td>(V)AGPSVF/FLFPPKPK/ (D)</td>
<td>2.4</td>
<td>228-240</td>
<td></td>
</tr>
<tr>
<td>470.2368</td>
<td>NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>(A)GPSVF/ (L)</td>
<td>5.1</td>
<td>229-233</td>
<td>Exclude (Tryptic rule)</td>
</tr>
<tr>
<td>1313.7528</td>
<td>1</td>
<td>4</td>
<td>1312.745</td>
<td>1312.754</td>
<td>(A)GPSVFLFPPKPK/ (D)</td>
<td>7.0</td>
<td>229-240</td>
<td></td>
</tr>
<tr>
<td>566.3647</td>
<td>1</td>
<td>4</td>
<td>565.357</td>
<td>565.359</td>
<td>(F)PPKPK/ (D)</td>
<td>3.1</td>
<td>236-240</td>
<td></td>
</tr>
<tr>
<td>1688.7689</td>
<td>1</td>
<td>4</td>
<td>1705.772</td>
<td>1705.788</td>
<td>(V)HQDWLNGKEYKCK/ (V)</td>
<td>9.4</td>
<td>302-314</td>
<td></td>
</tr>
<tr>
<td>470.2368</td>
<td>NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>(E)PQVY/ (T)</td>
<td>5.2</td>
<td>338-341</td>
<td>Exclude (Tryptic rule)</td>
</tr>
<tr>
<td>470.2368</td>
<td>NA</td>
<td>0</td>
<td>487.240</td>
<td>487.239</td>
<td>(P)PSRE/ (E)</td>
<td>-1.9</td>
<td>345-348</td>
<td>Exclude (Tryptic rule)</td>
</tr>
<tr>
<td>1676.7706</td>
<td>2</td>
<td>NA</td>
<td>3351.527</td>
<td>3351.503</td>
<td>(E)WESNGQPENNYKTTPMLDS GGSFLYSK/ (L)</td>
<td>-7.1</td>
<td>373-401</td>
<td>Exclude (too large as 2\textsuperscript{nd} chain)</td>
</tr>
<tr>
<td>1118.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.175</td>
<td>(S)DGSSFLYKLSLTVDKSRWQQG/ (N)</td>
<td>1.8</td>
<td>393-412</td>
<td>Exclude (Tryptic rule, $^{16}$O/$^{18}$O rule, too large as 2\textsuperscript{nd} chain)</td>
</tr>
<tr>
<td>1118.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.183</td>
<td>(M)HEALHNHYTQKSLSPGKQ/ (S)</td>
<td>4.8</td>
<td>421-441</td>
<td>Exclude (Tryptic rule, $^{16}$O/$^{18}$O rule, too large as 2\textsuperscript{nd} chain, HC-LC)</td>
</tr>
<tr>
<td>1676.7706</td>
<td>2</td>
<td>NA</td>
<td>3369.537</td>
<td>3369.532</td>
<td>(P)SV SGAPGQRVTISCTGSSSN IGAGYDVHWYQQ/ (L)</td>
<td>-1.4</td>
<td>448-479</td>
<td>Exclude (Tryptic rule, too large as 2\textsuperscript{nd} chain)</td>
</tr>
<tr>
<td>923.5283</td>
<td>2</td>
<td>2</td>
<td>1863.053</td>
<td>1863.041</td>
<td>(D)VHWYQQLPGTAPKLL/ (Y)</td>
<td>-6.6</td>
<td>474-489</td>
<td>Exclude (Tryptic rule, $^{18}$O/$^{16}$O rule)</td>
</tr>
<tr>
<td>1101.0885</td>
<td>2</td>
<td>2</td>
<td>2200.162</td>
<td>2200.175</td>
<td>(T)APKLLIYNSNRP5GVDRF/ (S)</td>
<td>6.0</td>
<td>484-503</td>
<td>Exclude (Tryptic rule, $^{18}$O/$^{16}$O rule, too large as 2\textsuperscript{nd} chain)</td>
</tr>
<tr>
<td>1188.6063</td>
<td>3</td>
<td>NA</td>
<td>3580.807</td>
<td>3580.801</td>
<td>(A)TLVCLISDFPFGAVTVAWKA DSSPVKAGVETTP/ (S)</td>
<td>-1.6</td>
<td>576-609</td>
<td>Exclude (Tryptic rule, too large as 2\textsuperscript{nd} chain)</td>
</tr>
<tr>
<td>470.2368</td>
<td>NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>(F)YPGAV/ (T)</td>
<td>5.1</td>
<td>585-589</td>
<td>Exclude (Tryptic rule)</td>
</tr>
<tr>
<td>751.4051</td>
<td>NA</td>
<td>0</td>
<td>786.419</td>
<td>786.412</td>
<td>(Y)LSLTPEQ/ (W)</td>
<td>-8.5</td>
<td>623-629</td>
<td>Exclude (Tryptic rule)</td>
</tr>
<tr>
<td>769.4172</td>
<td>1</td>
<td>0</td>
<td>786.420</td>
<td>786.412</td>
<td>(Y)LSLTPEQ/ (W)</td>
<td>-9.7</td>
<td>623-629</td>
<td>Exclude (Tryptic rule)</td>
</tr>
</tbody>
</table>
Table 3-4. Peptides in IgG2 that match with the mass of fragment ions of the triply charged ion m/z 1351.33 (RT at 91.17 min) via FindPept. All Cys are alkylated with IAA. The adjoining residues before cleavage are in parenthesis. NA means not available. The table was grouped in the order of m/z value.

<table>
<thead>
<tr>
<th>m/z</th>
<th>Charge</th>
<th>mass (Da)</th>
<th>User mass (m/z)</th>
<th>Theor mass (Da)</th>
<th>Δmass (ppm)</th>
<th>peptide</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>470.2368 NA</td>
<td>0</td>
<td>487.240</td>
<td>487.239</td>
<td>-1.9</td>
<td>(P)PSRE(E)</td>
<td>Exclude (Tryptic rule)</td>
<td></td>
</tr>
<tr>
<td>470.2368 NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>5.1</td>
<td>(K)/GPSVF(P)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>470.2368 NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>5.1</td>
<td>(A)GPSVF(L)</td>
<td>Exclude (Tryptic rule)</td>
<td></td>
</tr>
<tr>
<td>470.2368 NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>5.1</td>
<td>(F)YPGAV(T)</td>
<td>Exclude (Tryptic rule)</td>
<td></td>
</tr>
<tr>
<td>470.2368 NA</td>
<td>0</td>
<td>505.251</td>
<td>505.254</td>
<td>5.2</td>
<td>(E)PQVY(T)</td>
<td>Exclude (Tryptic rule)</td>
<td></td>
</tr>
<tr>
<td>566.3647 1</td>
<td>4</td>
<td>565.357</td>
<td>565.359</td>
<td>3.1</td>
<td>(F)PPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>751.4051 NA</td>
<td>0</td>
<td>786.419</td>
<td>786.412</td>
<td>-8.5</td>
<td>(Y)LSLTEQ(W)</td>
<td>Exclude (Tryptic rule)</td>
<td></td>
</tr>
<tr>
<td>769.4172 1</td>
<td>0</td>
<td>786.420</td>
<td>786.412</td>
<td>-9.7</td>
<td>(Y)LSLTEQ(W)</td>
<td>Exclude (Tryptic rule)</td>
<td></td>
</tr>
<tr>
<td>769.4172 1</td>
<td>0</td>
<td>786.420</td>
<td>786.428</td>
<td>9.6</td>
<td>(K)/GPSVFPLA(P)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>769.4172 1</td>
<td>0</td>
<td>804.431</td>
<td>804.434</td>
<td>3.8</td>
<td>(S)NTKVDKT(V)</td>
<td>Exclude (Tryptic rule)</td>
<td></td>
</tr>
<tr>
<td>839.4851 2</td>
<td>2</td>
<td>1676.956</td>
<td>1676.965</td>
<td>5.5</td>
<td>(A)PPVAGPSVLFPPKPK/(D)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>923.5283 2</td>
<td>2</td>
<td>1845.042</td>
<td>1845.055</td>
<td>7.1</td>
<td>(C)PAPPVAGPSVLFPPKPK/(O)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>923.5283 2</td>
<td>2</td>
<td>1863.053</td>
<td>1863.041</td>
<td>-6.6</td>
<td>(D)VHWYQQLQGTPKLLY</td>
<td>Exclude (Tryptic rule, 18O/16O rule)</td>
<td></td>
</tr>
<tr>
<td>1101.0885 2</td>
<td>2</td>
<td>2200.162</td>
<td>2200.174</td>
<td>5.4</td>
<td>(W)GQGTLYTVSSASTKGPSVFPLAP(C)</td>
<td>Exclude (Tryptic rule, 16O/16O rule)</td>
<td></td>
</tr>
<tr>
<td>1101.0885 2</td>
<td>2</td>
<td>2200.162</td>
<td>2200.175</td>
<td>6.0</td>
<td>(T)APKLLIY GinsNRPSGVVPDRF(S)</td>
<td>Exclude (Tryptic rule, 16O/16O rule)</td>
<td></td>
</tr>
<tr>
<td>1101.0885 2</td>
<td>2</td>
<td>2218.173</td>
<td>2218.163</td>
<td>-4.3</td>
<td>(Y)WGQGTLYTVSSASTKGPSVF PL /A</td>
<td>Exclude (Tryptic rule, 16O/16O rule)</td>
<td></td>
</tr>
<tr>
<td>Peak</td>
<td>Charge</td>
<td>Intensity</td>
<td>Retention Time</td>
<td>Mass Diff</td>
<td>pI</td>
<td>Peptide Sequence</td>
<td>Exclusion Reason</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-----------</td>
<td>----------------</td>
<td>-----------</td>
<td>---------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>1081.0885</td>
<td>2</td>
<td>2</td>
<td>2200.162</td>
<td>2200.175</td>
<td>6.0</td>
<td>(C)PPCPAPPVAGPSVFLFPPKPK/(D)</td>
<td></td>
</tr>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.175</td>
<td>0.9</td>
<td>(Q)ESGPGVTKPSGTLEAVS GGSIS(S)</td>
<td>Exclude (Tryptic rule, 18O/16O rule)</td>
</tr>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.175</td>
<td>1.8</td>
<td>(S)DGSSFLYKLTVDKSRWQQG(N)</td>
<td>Exclude (Tryptic rule, 18O/16O rule)</td>
</tr>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.183</td>
<td>4.8</td>
<td>(M)HEAHLNHYTQKSLSLSPGKQ S(V)</td>
<td>Exclude (Tryptic rule, 18O/16O rule)</td>
</tr>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.189</td>
<td>7.4</td>
<td>(W)GQGTLTVSSASTKGPSVFPLAPC(S)</td>
<td>Exclude (Tryptic rule, 18O/16O rule)</td>
</tr>
<tr>
<td>1181.5930</td>
<td>2</td>
<td>2</td>
<td>2361.171</td>
<td>2361.190</td>
<td>8.0</td>
<td>(E)CPPCPAPPVAGPSVFLFPPKPK/(D)</td>
<td></td>
</tr>
<tr>
<td>1188.6063</td>
<td>3</td>
<td>NA</td>
<td>3580.807</td>
<td>3580.776</td>
<td>-8.7</td>
<td>(I)GEISHSGTTNYNSLKSRT ISGDKSKNQFSLE(L)</td>
<td>Exclude (Tryptic rule, too large as 2(^{nd}) chain)</td>
</tr>
<tr>
<td>1188.6063</td>
<td>3</td>
<td>NA</td>
<td>3580.807</td>
<td>3580.801</td>
<td>-1.6</td>
<td>(A)TLVCLISDFYPGAVTVAWKA DSSPVKAGVETTP(S)</td>
<td>Exclude (Tryptic rule, too large as 2(^{nd}) chain)</td>
</tr>
<tr>
<td>1188.6063</td>
<td>3</td>
<td>NA</td>
<td>3580.807</td>
<td>3580.817</td>
<td>2.9</td>
<td>(W)SWVRQPPGGLEWIGEISHSGTTNYNSLKSRT/(V)</td>
<td>Exclude (Too large as 2(^{nd}) chain)</td>
</tr>
<tr>
<td>1313.7528</td>
<td>1</td>
<td>4</td>
<td>1312.745</td>
<td>1312.754</td>
<td>7.0</td>
<td>(A)GPSVFLFPPKPK/(D)</td>
<td></td>
</tr>
<tr>
<td>1333.6070</td>
<td>2</td>
<td>2</td>
<td>2683.210</td>
<td>2683.218</td>
<td>3.1</td>
<td>(V)TVPSNFGTQTYTCNDHKS SNTK/(V)</td>
<td>Exclude (too large as 2(^{nd}) chain)</td>
</tr>
<tr>
<td>1341.1364</td>
<td>2</td>
<td>NA</td>
<td>2680.258</td>
<td>2680.258</td>
<td>0.1</td>
<td>(T)YTCNDHKSNTKVDKVERK/C</td>
<td>Exclude (Tryptic rule)</td>
</tr>
<tr>
<td>1384.7951</td>
<td>1</td>
<td>4</td>
<td>1383.788</td>
<td>1383.791</td>
<td>2.4</td>
<td>(V)AGPSVFLFPPKPK/(D)</td>
<td></td>
</tr>
<tr>
<td>1676.7706</td>
<td>2</td>
<td>NA</td>
<td>3351.527</td>
<td>3351.503</td>
<td>-7.1</td>
<td>(E)WESNGQPENNYTTPMILDS DGSSFLYSK/(L)</td>
<td>Exclude (too large as 2(^{nd}) chain)</td>
</tr>
<tr>
<td>1676.7706</td>
<td>2</td>
<td>NA</td>
<td>3369.537</td>
<td>3369.532</td>
<td>-1.4</td>
<td>(P)SVSGAPQQRVTISCTGSSSIN AGAGYDVHHYQQ/(L)</td>
<td>Exclude (Tryptic rule, too large as 2(^{nd}) chain)</td>
</tr>
<tr>
<td>1688.7689</td>
<td>1</td>
<td>4</td>
<td>1705.772</td>
<td>1705.788</td>
<td>9.4</td>
<td>(V)HQDWNLNGEYKCK/(V)</td>
<td></td>
</tr>
<tr>
<td>1846.0661</td>
<td>1</td>
<td>4</td>
<td>1845.059</td>
<td>1845.055</td>
<td>-2.0</td>
<td>(C)PAPPVAGPSVFLFPPKPK/(D)</td>
<td></td>
</tr>
</tbody>
</table>
Table 3-5. Partial sequences determined from the mass of fragmentation ions for a triply charge precursor ion at the retention time of 91.17 min with m/z 1351.33 (molecular mass at 4048.963 Da). The longest b- and y-ions observed are underlined. All Cys are alkylated with IAA. The adjoining residues before cleavage are in parenthesis. $^{18}$O-labeling indicated that ions 1 and 2 are b-ions and ion 3 to 11 are y-ions. The crosslinked peptides were determined to be HC:G118-R129/HC:C215-K240

<table>
<thead>
<tr>
<th>#</th>
<th>m/z</th>
<th>Charge mass (Da)</th>
<th>Theor. mass (Da)</th>
<th>Δmass (ppm)</th>
<th>peptide</th>
<th>Corresponding Tryptic Peptide</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>470.2368</td>
<td>1</td>
<td>505.254</td>
<td>5.1</td>
<td>(K)GPSVF(P)</td>
<td>(K)$^{18}$GPSVFPLAPCSR$^{129}$/(S)</td>
<td>1287.6282 Chain 2</td>
</tr>
<tr>
<td>2</td>
<td>769.4172</td>
<td>1</td>
<td>786.428</td>
<td>9.6</td>
<td>(K)GPSVFPLA(P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>566.3647</td>
<td>1</td>
<td>565.357</td>
<td>3.1</td>
<td>(F)PPKPK/(D)</td>
<td>(K)$^{25}$CCVECPCPPVAGPSVFLFPPKPK$^{249}$(D)</td>
<td>2911.3305 Chain 1</td>
</tr>
<tr>
<td>4</td>
<td>1313.7528</td>
<td>1</td>
<td>1312.754</td>
<td>7.0</td>
<td>(A)GPSVFLPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1384.7951</td>
<td>1</td>
<td>1383.791</td>
<td>2.4</td>
<td>(V)AGPSVFLPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>839.4851</td>
<td>2</td>
<td>1676.956</td>
<td>5.5</td>
<td>(A)PPVAGPSVFLPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1846.0661</td>
<td>1</td>
<td>1845.055</td>
<td>7.1</td>
<td>(C)PAPPVAGPSVFLPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>923.5283</td>
<td>2</td>
<td>1845.059</td>
<td>-2.0</td>
<td>(C)PAPPVAGPSVFLPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1101.0885</td>
<td>2</td>
<td>2200.175</td>
<td>6.0</td>
<td>(C)PPCPAPPVAGPSVFLPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1181.5930</td>
<td>2</td>
<td>2361.190</td>
<td>8.0</td>
<td>(E)CPPCPAPPVAGPSVFLPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1688.7689</td>
<td>1</td>
<td>1705.772</td>
<td>9.4</td>
<td>(V)HQDWLNGKEYKCK/(V)</td>
<td>(R)$^{204}$VSVLTVVHQDWLNGKEYKCK$^{314}$(V)</td>
<td>2502.2941 ruled out for chain 2, see text</td>
</tr>
</tbody>
</table>

ruled out for chain 2, see text
Table 3-6. De novo sequencing for sequence tag using y-ions from the cross-linked fragments (group 4 in Table 3-1) in the cross-link peptide G118-R129/C215-K240. The observed m/z value is the most abundant isotopic peak in each isotopic envelope. The mass difference between a pair of adjacent y-ions was calculated and compared to the mass of single amino acids and dipeptides within mass error of 0.05 Da. The amino acid residues or dipeptides found in the putative cross-linked peptide chains are shown in red. The sequence tag SVFPLA is in the cross-linked peptide chain G118-R129.

<table>
<thead>
<tr>
<th>Obs. m/z</th>
<th>Charge State</th>
<th>Mass Shift in 18O-water</th>
<th>Obs. Mass (Da)</th>
<th>Mass Diff. (Da)</th>
<th>Amino Acid Residue or Di-peptide (Mass error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1633.2705</td>
<td>2</td>
<td>4</td>
<td>3264.5265</td>
<td>17.9980</td>
<td>H2O=18.0106(0.0126)</td>
</tr>
<tr>
<td>1642.2695</td>
<td>2</td>
<td>4</td>
<td>3282.5244</td>
<td>184.0776</td>
<td>IA, LA=184.1212(0.0436); PS=184.0848(0.0072)</td>
</tr>
<tr>
<td>1734.3083</td>
<td>2</td>
<td>4</td>
<td>3466.6020</td>
<td>79.0643</td>
<td>P-H2O=97.0528-18.0106=79.0422(-0.0221)</td>
</tr>
<tr>
<td>1773.8405</td>
<td>2</td>
<td>4</td>
<td>3545.6663</td>
<td>18.0336</td>
<td>H2O=18.0106(-0.0230)</td>
</tr>
<tr>
<td>1782.8573</td>
<td>2</td>
<td>4</td>
<td>3563.7000</td>
<td>147.0739</td>
<td>F=147.0684(-0.0055)</td>
</tr>
<tr>
<td>1237.6012</td>
<td>3</td>
<td>2.5</td>
<td>3709.7818</td>
<td>186.1085</td>
<td>W=186.0793(-0.0292); AD,GE=186.0641(-0.0444); SV=186.1004(-0.0081)</td>
</tr>
<tr>
<td>1299.6374</td>
<td>3</td>
<td>2.5</td>
<td>3895.8903</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
3.4.4 Stage 4: Deduction of Crosslinking Chemistry and Site

Toward this end, two pieces of information are particularly useful: (1) elemental composition of the functional group involved and (2) peptide fragments not observed in tandem mass spectrometry.

Deducing Crosslinking Chemistry: Mass to Formula (Elemental Composition). An example of the elemental composition calculation is illustrated in Table 3-7. Once the peptides G118-R129 and C215-K240 were established as the components of the cross-linked peptide, the difference between the observed mass of the crosslinked peptide (4048.9600 Da) and the combined mass of the two chains devoid of modifications (1287.6282+2911.3305=4198.9587 Da) was calculated to be 149.9987 Da. Four potential elemental compositions for the mass of 149.9987 Da were obtained via Thermo-Fisher Scientific Xcalibur (Table 3-7). The last two were eliminated based on their high delta ppm relative to the FT MS mass accuracy (typically ≤ 5 ppm). The high RDB (ring and double bond) value makes the second one unlikely too. This leaves the first one as the only plausible choice.

The elemental composition of C₄H₆O₄S contains sulfur, which only presents in cysteine and methionine. Each of the putative crosslinked peptide pair G118-R129 and C215-K240, and particularly the fragments that were not observed by tandem mass spectrometry (those not underlined in Table 3-5), contain cysteine but not methionine. During sample preparation, cysteine residues were reduced and alkylated by iodoacetic acid (IAA), so the mass for all peptides were calculated assuming cysteines are alkylated. Hence, removal of two alkyl groups (two C₂H₃O₂) and a sulfur atom exactly matches the determined elemental composition. The
mass of observed crosslink peptide (4048.963 Da) and the theoretical thioether peptide (4048.960 Da) are practically identical (with mass error of 0.74 ppm, see Table 3-8). All together, we surmised that the crosslinking chemistry is a thioether originating from a pair of cysteine residues (Scheme 3-2).

Table 3-7. Elemental formula with mass of 149.9987 Da (the mass difference between the sum of unmodified peptides and crosslinked peptide). RDB means ring and double bond; ppm is part-per-million.

<table>
<thead>
<tr>
<th>Elemental Composition</th>
<th>Proposed Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>Cal. Mass (Da)</td>
</tr>
<tr>
<td>C₄H₆O₄S</td>
<td>149.9987</td>
</tr>
<tr>
<td>C₁₀ON</td>
<td>149.9980</td>
</tr>
<tr>
<td>C₅H₂N₄S</td>
<td>150.0000</td>
</tr>
<tr>
<td>C₂H₄O₃N₃S</td>
<td>149.9973</td>
</tr>
</tbody>
</table>

**Locating Crosslink Site.** Typically, the crosslink site can be localized by the largest b- and y-ions observed. For the crosslinked peptide m/z 1351.33, the largest b- and y-ions are GPSVFPLA/(PCSR, not observed), (CCVE, not observed)/CPPCPAPPVAGPSVFLFPPPKPK for the chain HC:G118-R129 and HC:C215-K240, respectively. This indicates the crosslink site is in the corresponding PCSR and CCVE region (Figure 3-2). Compared to the highly stable
valine and proline, cysteine and glutamic acid are chemically reactive, so they are more likely candidates for crosslinking.

The elemental analysis described in the previous section indicates that a sulfur atom was removed, suggesting that a cysteine is involved. In addition, functional groups with a combined composition of $C_4H_6O_4$ are eliminated from the theoretical peptide pairs, in which cysteiny1 residues were assumed to be alkylated with IAA ($C_2H_3O_2$). Taken together, our data indicated that the crosslink site is highly likely at HC:Cys$_{127}$-HC:Cys$_{215}$ or HC:Cys$_{127}$-HC:Cys$_{216}$ as shown in Figure 3-2. Because the two cysteines (Cys$_{215}$ and Cys$_{216}$) in the heavy chain are adjacent to each other, the exact crosslink site was unable to be unambiguously determined here.
Figure 3-2. CID MS/MS spectrum of the triply charged precursor ions at m/z 1351.33 (\(^{16}\)O-labeled C-termini) and 1354.00 (\(^{18}\)O-labeled C-termini). Singly and doubly charge fragment ions that contain the individual chain of the crosslinked peptides (G118-R129/C215-K240, RT at 91.17 min) are highlighted in blue and red, respectively. Characteristic mass shift imparted by the heavier isotope \(^{18}\)O was observed (e.g., the mass shift of 4 Da for y5 in \(^{16}\)O- vs \(^{18}\)O-water).
3.4.5 Final Confirmation and Additional Support

**Confirmation by Data Matching.** Once the putative crosslinking chemistry and site have been proposed, theoretical fragmentation spectra were calculated and compared with the observed spectra. The assignment is shown in Figure 3-2 and is highly consistent with the deduced structure. A handful fragment peaks of a few crosslinks were not assigned initially, and hence were subjected to further analysis as described next.

**MS3 Analysis.** MS3 analysis may provide additional structural information, especially for fragment ions that are difficult to assign in the MS/MS. For example, in Figure 3-3a, two high intensity fragment ions at m/z 1196.50 (singly charged) and 1520.90 (doubly charged) were observed for the triply charged crosslinked peptide m/z 1413.37 (G118-R129/K214-K240), but could not be assigned to typical b- or y-ions. To ascertain, MS3 analysis of these two unassigned ions revealed the sequences to be $^{118}$GPSVFPLAPC*SR$^{129}$ and $^{214}$KCC*VECPCPPAPPVAGPSVFLFPPPK$^{240}$, in which a dehydroalanine replaces C127 in peptide G118-R129 and a free cysteine replaces the thioether at 216 in peptide K214-K240, respectively (Figure 3-3b & 3-3c). These data further supported the proposed sequence and crosslink sites. Alkylation at Lys and Met as an artifact from sample preparation in peptide mapping was reported[49]. The alkylation at K214 in the crosslink peptide G118-R129/K214-K240 is in agreement with the literature [49].
**Additional Peptides.** Following the same work flow, full sequences, crosslinking chemistry and sites have been established for all ten candidate crosslinked peptides shown in red in Table 3-2. The final results of all identified crosslinked peptides are summarized in Table 3-8.

To evaluate the sensitivity of our method, the peak intensity from LC-MS analysis for each crosslinked peptide and its related (not crosslinked) peptides was used to estimate the degree of crosslinking as described by Zhang[16], ranging from 0.2% to 5.0% (half less than 1%; see Table 3-10 for details). Comparable data were observed based on reducing SDS-PAGE, which indicated about 8% of total crosslinked species (see Figure 3-8A). It is also worth noting that no enrichment or separation was performed on the IgG2 samples prior to tryptic digestion (the first step of our work flow); in other words, the crosslinked peptides were analyzed in the presence of large excess of native peptides. Of course, considerably higher sensitivity can be achieved if the crosslinked proteins are separated or enriched prior to analysis.
Figure 3-3a. MS/MS data of the cross-link peptide HC:G118-R129/HC:K^K214-K240 (RT at 88.56 min, m/z 1413.36, charge of 3). Lys^{214} was found alkylated by IAA in this peptide.
Figure 3-3b. MS³ for structure confirmation of the singly-charged fragment ion m/z 1196 from the cross-link peptide HC:G118-R129/HC:K²¹⁴-K240 (RT at 88.56 min, m/z 1413.36, charge of 3). Lys²¹⁴ was found alkylated in this peptide. A) Isotopic distribution (observed m/z 1196.6369, calculated m/z 1196.6423, mass error of 4.5 ppm). B) MS³ spectrum.
Figure 3-3c. MS$^3$ for structure confirmation of the double-charged fragment ion m/z 1521 from the cross-linked peptide HC:G118-R129/HC:K$^q$214-K240 (RT at 88.56 min, m/z 1413.36, charge of 3). Lys$^{214}$ was found alkylated in this peptide. A) Isotopic distribution (observed m/z 1520.7060, calculated m/z 1520.7200, mass error of 9.2 ppm). B) MS$^3$ spectrum.
Table 3-8. Crosslinked peptides identified in IgG2. Thioether crosslink sites are labeled in red and bold-face. The exact site at either C215 or C216 is unable to be differentiated with available data. Unless noted, all cysteine side chains are alkylated with IAA. Alkylated Lys is shown in blue. The longest b- or y-ions observed are underlined. Peptides are shown with the adjoining amino acid residues before cleavage in parenthesis. The peptide #8 contains a thioether (in red) and dehydroalanine (in green and asterisk) at C215-C216, as shown in Figure S5.

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>m/z(charge)</th>
<th>RT (min)</th>
<th>Observed Mass (Da)</th>
<th>Theoretical Mass (Da)</th>
<th>Mass Error (ppm)</th>
<th>Sequence</th>
<th>Cross-linking Site in Heavy Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G118-R129/ C215-K240</td>
<td>1351.33(3+)</td>
<td>91.17</td>
<td>4048.963</td>
<td>4048.960</td>
<td>0.74</td>
<td>(K)127GPSVFLAPCSR129(S)</td>
<td>217 or 216</td>
</tr>
<tr>
<td>2</td>
<td>G118-R129/ K214-K240</td>
<td>1394.03(3+)</td>
<td>87.91</td>
<td>4177.059</td>
<td>4177.055</td>
<td>0.96</td>
<td>(K)127GPSVFLAPCSR129(S)</td>
<td>Same</td>
</tr>
<tr>
<td>3</td>
<td>G118-R129/ K214-K240</td>
<td>1413.36(3+)</td>
<td>88.56</td>
<td>4235.064</td>
<td>4235.060</td>
<td>0.94</td>
<td>(K)127GPSVFLAPCSR129(S)</td>
<td>Same</td>
</tr>
<tr>
<td>4</td>
<td>C215-K240/ C215-K240</td>
<td>1419.92(4+)</td>
<td>101.59</td>
<td>5672.670</td>
<td>5672.662</td>
<td>1.41</td>
<td>(K)214CCECPPCPAPPVAGPSVFLFPKPK240(D)</td>
<td>215 or 216</td>
</tr>
<tr>
<td>5</td>
<td>K214-K240/ K214-K240</td>
<td>1452.20(4+)</td>
<td>99.76</td>
<td>5800.765</td>
<td>5800.757</td>
<td>1.38</td>
<td>(K)214CCECPPCPAPPVAGPSVFLFPKPK240(D)</td>
<td>Same</td>
</tr>
<tr>
<td>6</td>
<td>K214-K240/ K214-K240</td>
<td>1484.22(4+)</td>
<td>97.88</td>
<td>5928.854</td>
<td>5928.852</td>
<td>0.34</td>
<td>(K)214CCECPPCPAPPVAGPSVFLFPKPK240(D)</td>
<td>Same</td>
</tr>
<tr>
<td>7</td>
<td>K214-K240/ K214-K240</td>
<td>1498.47(4+)</td>
<td>98.99</td>
<td>5986.858</td>
<td>5986.858</td>
<td>0.00</td>
<td>(K)214CCECPPCPAPPVAGPSVFLFPKPK240(D)</td>
<td>Same</td>
</tr>
<tr>
<td>8</td>
<td>K214-K240*/ K214-K240</td>
<td>1475.73(4+)</td>
<td>98.49</td>
<td>5894.860</td>
<td>5894.856</td>
<td>0.68</td>
<td>(K)214KCC*ECPPCPAPPVAGPSVFLFPKPK240(D)</td>
<td>Same</td>
</tr>
<tr>
<td>9</td>
<td>T210-K240/ K214-K240</td>
<td>1605.54(4+)</td>
<td>96.15</td>
<td>6414.117</td>
<td>6414.112</td>
<td>0.78</td>
<td>(K)214TVERKCCVECPPCPAPPVAGPSVFLFPKPK240(D)</td>
<td>Same</td>
</tr>
<tr>
<td>10</td>
<td>T210-K240/ K214-K240</td>
<td>1620.04(4+)</td>
<td>97.39</td>
<td>6472.120</td>
<td>6472.117</td>
<td>0.46</td>
<td>(K)214TVERKCCVECPPCPAPPVAGPSVFLFPKPK240(D)</td>
<td>Same</td>
</tr>
</tbody>
</table>
3.4.6 Targeted Search Based on the Newly Established Crosslinking Chemistry

After the thioether crosslink chemistry was established, a targeted search for this particular modification was performed following well-established protocols. First, a theoretical database was built for all combinations of a thioether crosslinking between any two cysteiny1 residues. Then, all observed precursor ions were searched against the database. When a hit was found in the targeted mass search, the corresponding MS/MS data from both the $^{18}$O-water and $^{16}$O-water digests were examined for further structural confirmation. By this approach, four additional thioether peptides were found (Table 3-9, Figure 3-6 & 3-7). All contain a light chain C-terminal peptide, so each has only one newly created C-terminus (two $^{18}$O-incorporation) and therefore was not discriminated from single chain peptides in the initial screening stage. Again, these results showcase the utility of our approach to identify crosslinks in macromolecules derived from previously unknown crosslinking chemistry.
Table 3-9. The cross-linking peptides identified in the IgG2 via a targeted search for cysteiny1 thioether (labeled in red and bold-face). The exact site at either C215 or C216 is unable to be differentiated with available data. Unless noted, all cysteine side chains are alkylated with IAA. Alkylated Lys is shown in blue and asterisk. Peptides are shown with the adjoining amino acid residues before cleavage in parenthesis.

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>m/z(charge)</th>
<th>RT (min)</th>
<th>Observed Mass (Da)</th>
<th>Theoretical Mass (Da)</th>
<th>Mass Error (ppm)</th>
<th>Sequence</th>
<th>Cross-linking Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>T211-S218/G118-R129</td>
<td>1002.50(2+)</td>
<td>52.25</td>
<td>2001.984</td>
<td>2001.983</td>
<td>0.40</td>
<td>(K)^{211}TVAPTECS^{218}</td>
<td>LC217/HC127</td>
</tr>
<tr>
<td>12</td>
<td>T211-S218/C215-K240</td>
<td>1209.90(3+)</td>
<td>78.29</td>
<td>3625.689</td>
<td>3625.685</td>
<td>1.10</td>
<td>(K)^{211}TVAPTECS^{218}</td>
<td>LC217/HC215 or 216</td>
</tr>
<tr>
<td>13</td>
<td>T211-S218/K214-K240</td>
<td>1252.94(3+)</td>
<td>76.12</td>
<td>3753.784</td>
<td>3753.780</td>
<td>1.00</td>
<td>(K)^{211}TVAPTECS^{218}</td>
<td>Same</td>
</tr>
<tr>
<td>14</td>
<td>T211-S218/K240</td>
<td>1272.27(3+)</td>
<td>77.69</td>
<td>3811.789</td>
<td>3811.786</td>
<td>0.70</td>
<td>(R)^{214}{K^{A}}CCVECPPCPAPPVAGPSVFLFPKPK^{240}(D)</td>
<td>Same</td>
</tr>
</tbody>
</table>

Notes:
1. Alkylated Lys is shown in blue and asterisk.
2. Peptides are shown with the adjoining amino acid residues before cleavage in parenthesis.
3. The exact site at either C215 or C216 is unable to be differentiated with available data.
4. Unless noted, all cysteine side chains are alkylated with IAA.
Figure 3-4. MS/MS data of the cross-linked peptide HC:C215-K240/HC:C215-K240 (RT at 101.59 min, m/z 1419.92, charge of 4).
Figure 3-5. MS/MS data of the cross-linked peptide HC:K214-K240*/HC:K*A214-K240 (RT at 98.49 min, m/z 1475.73, charge of 4).
Figure 3-6. MS/MS data of the cross-link peptide LC:T211-S218/HC:G118-R129 (RT at 52.25 min, m/z 1002.50, charge of 2).
Figure 3-7. MS/MS data of the cross-linked peptide LC:T211-S218/HC:K214-K240 (RT at 76.12 min, m/z 1252.94, charge of 3).
Table 3-10 Quantification of the cross-linked peptides in the IgG2

<table>
<thead>
<tr>
<th>#</th>
<th>Peptide</th>
<th>Level (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G118-R129/C215-K240</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>G118-R129/K214-K240</td>
<td>1.0</td>
</tr>
<tr>
<td>3</td>
<td>G118-R129/KΔ214-K240</td>
<td>1.9</td>
</tr>
<tr>
<td>4</td>
<td>C215-K240/C215-K240</td>
<td>0.3</td>
</tr>
<tr>
<td>5</td>
<td>C215-K240/K214-K240</td>
<td>2.6</td>
</tr>
<tr>
<td>6</td>
<td>K214-K240/K214-K240</td>
<td>2.4</td>
</tr>
<tr>
<td>7</td>
<td>K214-K240/KΔ214-K240</td>
<td>1.7</td>
</tr>
<tr>
<td>8</td>
<td>K214-K240*/KΔ214-K240</td>
<td>0.2</td>
</tr>
<tr>
<td>9</td>
<td>T210-K240/K214-K240</td>
<td>0.6</td>
</tr>
<tr>
<td>10</td>
<td>T210-K240/KΔ214-K240</td>
<td>0.5</td>
</tr>
<tr>
<td>11</td>
<td>T211-S218/G118-R129</td>
<td>5.0</td>
</tr>
<tr>
<td>12</td>
<td>T211-S218/C215-K240</td>
<td>0.2</td>
</tr>
<tr>
<td>13</td>
<td>T211-S218/K214-K240</td>
<td>0.2</td>
</tr>
<tr>
<td>14</td>
<td>T211-S218/KΔ214-K240</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Figure 3-8. (A) Detection of cross-links in IgG2 by reducing SDS-PAGE. Lane 1, 2, 3, and 4 are molecular weight marker, control, acid stressed sample (pH 3.0 at 50 °C for 1 week), and base stressed sample (pH 8.5 at 50 °C for 1 week), respectively. About 8% (combined) non-reducible high molecular bands were observed in the base stressed IgG2, as quantified by the software TotalLab Quant version 12.4. (B) Aggregation analysis by size exclusion chromatography. Seven percent (7%) aggregation was observed in the base stressed IgG2. Column: TSKgel G3000 SWXL 7.8 x 300 mm 5 μm (two in tandem); mobile phase: 150 mM NaCl in 100 mM sodium phosphate buffer pH 6.9; column temperature: ambient; flow rate: 0.5 mL/min; detection: 215 nm; injection: 10 μL of 2.5mg/mL sample.
3.5 Formation of Thioether

Thioether is a known modification for proteins[50-55]. For IgG1, a thioether crosslink was located at the disulfide bond of the light chain C-termini and the heavy chain hinge region[52, 55]. A generally accepted mechanism involves a β-elimination of disulfide to generate dehydroalanine followed by Michael addition by another cysteiny1 thiol[52-54, 56, 57]. Basic conditions and structural flexibility generally favor its formation[50-55]. In addition, radical intermediates have been postulated for desulfurization[57, 58]. The hinge region of IgG2 is highly flexible and solvent exposed, and therefore very susceptible to this transformation. Indeed, our results indicated it occurs more frequently at the light chain C-termini and in the hinge region of IgG2. It is very interesting that the disulfide bonds of heavy chain C127 –heavy chain C215 (or C216) in the IgG2 A/B form (or B form) are also reactive (Figure 3-9). These thioether crosslinks at HC:Cys\textsuperscript{127}-HC:Cys\textsuperscript{215} (or HC:Cys\textsuperscript{127}-HC:Cys\textsuperscript{216}), HC:Cys\textsuperscript{215}-HC:Cys\textsuperscript{215} (or HC:Cys\textsuperscript{216}-HC:Cys\textsuperscript{216}), LC:Cys\textsuperscript{217}-HC:Cys\textsuperscript{127}, and LC:Cys\textsuperscript{217}-HC:Cys\textsuperscript{215} (or LC:Cys\textsuperscript{217}-HC:Cys\textsuperscript{216}) originated from native disulfides as shown in red in Figure 3-9. Thioether linkage is in agreement with the previous reports on IgG2 disulfide bond pairing[59-61]. In Table 3-8, the crosslink peptide #8 (HC:K214-K240*/HC:K214-K240) contains a thioether and an dehydroalanine. The corresponding linear peptide K214-K240\* (\textsuperscript{214}KCC*VECPPCPAPPVAGPSVFLFPKP\textsuperscript{240}, the dehydroalanine at C215 or C216 was denoted as asterisk) was also found. All together, these data are consistent with thioether formation via dehydroalanine intermediates.
Scheme 3-2. Establishment of crosslink chemistry based on formula C₄H₆O₄S obtained from elemental composition analysis of 149.9987 Da.

Figure 3-9. Major disulfide linkage isoforms in IgG2. Those labeled in red were found to convert into thioethers in IgG2.
3.6 Conclusions

The utility of our XChem-Finder strategy for the characterization of protein crosslinking with undefined chemistry is exemplified by the discovery of fourteen thioether peptides in IgG2. Essential to our approach is $^{18}$O-isotope labeling; it allows the facile detection of crosslinked peptides, and most significantly, divides the complex tandem mass spectra to sub-sets that can be processed by standard database search (FindPept that matches fragment ions with partial peptide sequences) and de novo sequencing (sequence tags). High-resolution spectral data also dramatically improve the confidence of assignment, and moreover, reveal the chemical nature of the crosslinking. While the reported work was manually processed, most steps can be automated. Hence our XChem-Finder strategy should be generally applicable for the discovery of crosslinked proteins, without prior defined chemistry, in both biological systems and biopharmaceuticals.

3.7 References


Chapter 4: Discovery and Characterization of a Novel Photo-Oxidative Histidine-Histidine Crosslink in IgG1 Antibody Utilizing $^{18}$O-labeling and Mass Spectrometry


Co-authors’ work in this chapter: Min Liu: experimental design and execute, data analysis, manuscript writing and revision; Zhongqi Zhang: data analysis, manuscript writing and revision; Janet Cheetham: idea contribution, manuscript writing and revision, and grant support; Da Ren: manuscript writing and revision; Zhaohui Sunny Zhou: idea contribution, data analysis, manuscript writing and revision, and grant support.
4.1 Abstract

A novel photo-oxidative crosslinking between two histidines (His-His) has been discovered and characterized in an IgG1 antibody via the workflow of XChem-Finder –$^{18}$O labeling and mass spectrometry (Anal Chem 2013, 85, 5900-5908). Its structure was elucidated by peptide mapping with multiple proteases with various specificities (e.g., trypsin, Asp-N, and GluC combined with trypsin or Asp-N) and mass spectrometry with complementary fragmentation modes (e.g., collision-induced dissociation (CID) and electron-transfer dissociation (ETD)). Our data indicated that crosslinking occurred across two identical conserved histidine residues on two separate heavy chains in the hinge region, which is highly flexible and solvent accessible. Based on model studies with short peptides, it has been proposed that singlet oxygen reacts with the histidyl imidazole ring to form an endoperoxide and then converted to the 2-oxo-histidine (2-oxo-His) and His+32 intermediates, the latter is subject to a nucleophilic attack by the unmodified histidine; and finally, elimination of a water molecule leads to the final adduct with a net mass increase of 14 Da. Our findings are consistent with this mechanism. Successful discovery of crosslinked His-His again demonstrates the broad applicability and utility of our XChem-Finder approach in the discovery and elucidation of protein cross-linking, particularly without *a priori* knowledge of the chemical nature and site of crosslinking.
4.2  Introduction

Protein crosslinks are ubiquitous in biological systems and biopharmaceuticals. They are also involved in disease pathologies such as Alzheimer[1-3] and cataractogenesis[2, 4]. As one of the post-translational modifications and degradations that occur during biopharmaceutical protein production processing and storage, crosslinks have been reported to result in aggregation, loss of bioactivity, and immunogenicity[5-7].

Despite the rapid advancements in mass spectrometry and data analysis algorithms, characterization of protein crosslinks remains challenging due to their structural complexity[8]. Whereas a limited set of crosslinked structures (e.g. thioether[7, 9-12]) have been characterized, most remain unknown; for example, the non-disulfide covalent crosslinking in crystalline [4, 13, 14], collagen[15], ubiquitylated proteins[3], ribonuclease A[16] and monoclonal antibodies[17, 18]. It is particularly challenging to characterize protein crosslinking without prior knowledge of the chemical nature and sites of crosslinking as no theoretical mass or spectrum can be predicted. In contrast, numerous chemical crosslinks with well-established crosslinking chemistry have been used in the investigation of protein structures and protein-protein interactions[19-25]. Since pre-defined crosslinking chemistry is involved, various specialized algorithms have been developed for data analysis for each incorporated crosslink. Naturally, these approaches are less amenable to the identification of crosslinks with undefined crosslinking chemistry. Recently, we developed a workflow—XChem-Finder—that is generally applicable for protein crosslinking. It involves, first, the detection of cross-linked peptides via the unique isotope patterns imparted by $^{18}$O-labeling of their two termini (in comparison, one terminus for a linear peptide), and then integrated mass spectrometric and data analysis[8].
IgG1 and IgG2 are the most popular therapeutic monoclonal antibodies on the market[26]. Applying our XChem-Finder workflow, we have discovered and characterized a novel histidine-histidine (His-His) crosslink in IgG1 antibody. High molecular weight species in the light-irradiated IgG1 were detected by reduced SDS-PAGE and size exclusion chromatography (SEC). Our LC-MS analysis indicated that crosslinking occurred across two identical conserved histidine residues (His220) on two separate heavy chains in the hinge region, which is highly flexible and solvent accessible. The crosslinking chemistry is consistent with the proposed mechanism based on model peptides under photo-oxidative conditions (see Scheme 4-1) [16, 27-29]. Successful discovery of His-His crosslink in IgG1 has further demonstrated the general applicability and power of our XChem-Finder workflow. To the best of our knowledge, our work reported herein is the first example of such crosslinking in a protein.

4.3 Experimental Section

4.3.1 Chemicals

All chemicals were reagent grade or above. Guanidine hydrochloride (GndHCl), ethylenediaminetetraacetic acid (EDTA), dithiothreitol (DTT), iodoacetic acid (IAA), trifluoroacetic acid (TFA), acetonitrile (ACN), HPLC-grade water, and bradykinin were from Sigma-Aldrich (St. Louis, MO, USA). Sequencing grade trypsin, GluC, and Asp-N were from Roche (Indianapolis, IN, USA). ¹⁸O-water (97%) was from Cambridge Isotope Laboratories (Andover, MA, USA). Recombinant monoclonal IgG1 antibody (anti-streptavidin immunoglobulin gamma 1) was produced in Chinese hamster ovary (CHO) cells (Amgen,
Thousand Oaks, CA, USA), purified according to standard manufacturing procedures, formulated at a concentration of 30 mg/mL in 50 mM sodium acetate at pH 5.2, and stored at -70 °C.

4.3.2 Generation of Stressed Sample

After being exchanged into various buffers of biopharmaceutical interest (50 mM sodium acetate at pH 4.8, 50 mM sodium phosphate at pH 7.4, 50 mM sodium bicarbonate at pH 9.0 or water), the IgG1 antibody at a concentration of 5 mg/mL in a clear 3 mL glass vial was put into a light chamber (Atlas Suntest CPS+ with Xenon Lamp and ID65 solar filter, controlled irradiance at 300-800 nm, light intensity at 765 W/m²) and exposed to light irradiation for 7, 14, and 22 hrs. These conditions represent the light irradiance of 1 x, 2 x, and 3 x ICH (International Conference on Harmonization of technical requirements for registration of pharmaceuticals for human use), respectively.

4.3.3 Aggregates by Size Exclusion Chromatography

Size exclusion chromatography (SEC) analysis for reduced IgG was carried out as described[30]. Briefly, IgG1 was diluted to 1 mg/mL in a denaturing buffer (7.5 M Gnd-HCl, 2 mM EDTA and 0.25 M Tris-HCl, pH 7.5) at room temperature. Reduction was accomplished by 10 mM DTT at room temperature for 30 min. Then 50 µL of the above samples was injected onto a TSKgel G3000 SWXL column (7.8 x 300 mm 5 µm) with an isocratic mobile phase of
0.1% TFA/H₂O:ACN (80:20) and a flow rate of 0.2 mL/min. The column was set at room temperature and the UV detector was at 280 nm.

4.3.4 Reduction, Alkylation, Tryptic Digestion and ¹⁸O-Labeling of IgG1

IgG1 was digested by trypsin similarly to the procedure described by Ren et al[31]. Briefly, IgG1 was diluted to 1 mg/mL in a denaturing buffer (7.5 M GndHCl, 2 mM EDTA and 0.25 M Tris-HCl, pH 7.5) to a final volume of 0.5 mL. Reduction was accomplished with the addition of 3 μL of 0.5 M DTT followed by 30 min incubation at room temperature. S-Carboxymethylation was achieved with the addition of 7 μL of 0.5 M IAA, and resulting mixture was incubated at room temperature in the dark for 15 min. Excess IAA was quenched with the addition of 4 μL of 0.5 M DTT. The reduced and alkylated IgG1 samples were subsequently exchanged into the digestion buffer (0.1 M Tris-HCl at pH 7.5) using a NAP-5 size-exclusion column (GE Healthcare, Piscataway, NJ, USA). Next, two aliquots (200 µL each) were completely dried via SpeedVac and reconstituted separately into the same volume of ¹⁸O-water or ¹⁶O-water; then 6 µL of 1 mg/mL trypsin in ¹⁸O-water or ¹⁶O-water solution, respectively, was added to achieve a 1:25 (w/w) enzyme/substrate ratio. The reaction mixtures were incubated at 37 °C for 30 min.

Other proteolytic digestions of IgG1 (Asp-N, Trypsin combined with GluC, and Asp-N combined with GluC) were performed in ¹⁶O-water only. Proteases were added to 100 µL of the above buffer-exchanged antibody to achieve a 1:25 (w/w) enzyme/substrate ratio. The reaction mixtures were incubated at 37 °C overnight.
Limited Asp-N digestion was performed by adding 6 μg of Asp-N into 300 μL digest (of trypsin combined with GluC) and incubating at 37 °C for 1.5 hr for LC/CID-MS analysis. An aliquot of 200 μL of the above digest was dried via SpeedVac and reconstituted into 40 μL of water for LC/ETD-MS analysis.

4.3.5 HPLC

The proteolytic digests of IgG1 (25 μL) were separated on a Jupiter C5 column (250 x 2.0 mm, 5 μm, 300Å, Phenomenex, Torrance, CA, USA) at 50 °C with a flow rate of 200 μL/min on a HPLC system (Agilent 1100, Palo Alto, CA, USA). Mobile phase A was 0.1% TFA in water (v/v) while mobile phase B contained 0.085% TFA in 90% ACN / 10% water. A gradient was applied by holding at 2% B for 2 min, increasing to 22% B in 38 min, then 42% B in 80 min, then 100% B in 25 min followed by holding at 100% B for 5 min. The column was re-equilibrated at 2% B for 30 min prior to next injection.

For ETD analysis, digests of IgG1 (6 μL) were separated on a PROTO C4 column (150 x 1.0 mm, 5 μm, 300Å, Higgins Analytical, Mountain View, CA, USA) at 50 °C with a flow rate of 60 μL/min on a HPLC system (Agilent 1100, Palo Alto, CA, USA). Mobile phase A was 0.1% FA / 0.02% TFA in water (v/v) while mobile phase B contained 0.1% FA / 0.02% TFA in 90% ACN / 10% water. The same gradient as described above was applied.
4.3.6 Mass Spectrometry

An LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) was used in-line with a HPLC system for the analyses of the IgG1 proteolytic digests. A full MS scan (with 60,000 resolution at m/z 400 and an automatic gain control (AGC) target value of $2 \times 10^5$) followed by data-dependent MS/MS scans of the three most abundant precursor ions was set up to acquire both the peptide mass and sequence information. The spray voltage was 5.5 kV, and the capillary temperature was 250 °C. The instrument was tuned using the doubly-charged ion of a synthetic peptide, bradykinin. The MS/MS spectra were obtained using CID with normalized collision energy of 35%. For MS/MS with ion detection in the Orbitrap, the AGC target was set to $3 \times 10^6$, resolution to 7,500, and the precursor isolation width to 4 m/z unit. Under our experimental conditions, the typical mass accuracy in full MS scan and FT MS/MS is 5 and 10 ppm, respectively.

ETD spectra were acquired on a Thermo-Scientific LXQ-XL mass spectrometer in centroid mode with isolation width of 5, reaction time of 75 ms and reagent target value of $1 \times 10^5$, using singly charged fluoranthene anions as the ETD reagent. Both CID and ETD data were analyzed for peptide identification, using a custom-written algorithm MassAnalyzer and verified manually[32-35].
4.4 Results and Discussion

A novel His-His crosslink in proteins has been discovered via our XChem-Finder workflow, without pre-defined crosslinking chemistry. Peptide mapping with mass spectrometry has established that the crosslink occurred across two identical His220 on each of the two heavy chains in the hinge region.

4.4.1 Detection of Crosslinked Protein.

Photo-induced non-reducible high molecular weight species were detected by reducing SDS-PAGE; their intensities increased with longer light exposure (Figure 4-1A). Their formation was pH-dependent: less favorable under acidic conditions, such as pH ~5 for typical formulation of proteins (Figure 4-1B); and more favorable in neutral or basic buffers that are commonly used in protein production and purification (Figure 4-1B). The crosslinked species were also quantified by size exclusion chromatography (SEC) (Figure 4-1C and D). Mobile phase of 0.1% TFA/H₂O:ACN (80:20) was used to avoid hydrophobic interaction with stationary phase[30]. The results from SEC and SDS-PAGE were consistent. The total amounts of the early elution peaks observed were at the level of 0.2, 4.5, 9.5, & 16.5% by peak area in the control sample and samples exposed to 1x, 2x, 3x ICH irradiation, respectively (Figure 4-1C). The crosslinks were also observed to increase to 25.8% in 50 mM NaHCO₃ pH 9.0, 15.7% in 50mM sodium phosphate pH 7.4, & 6.3% in 50 mM sodium acetate pH 4.8 (Figure 4-1D). It is interesting to note that the control sample (without light stress) already contained small yet detectable amount of crosslinking (0.4%, Figure 4-1D), suggesting such modifications could
occur during routine protein production and process. The chemical nature and site of crosslinking was discovered by our XChem-Finder workflow as detailed next[8].
Figure 4-1. Detection of crosslinking in IgG1 by reduced SDS-PAGE (A & B) and size exclusion chromatography (SEC) (C & D). Samples with the corresponding lane numbers in parenthesis are molecular ladder (1), IgG1-control (without stress) (2), IgG1-water-1xLight (3), IgG1-water-2xLight (4) & IgG1-water-3xLight (5); molecular ladder (6), IgG1-control (7), IgG1-pH4.8-3xLight (8), IgG1-pH7.4-3xLight (9), & IgG1-pH 9.0-3xLight (10), respectively. Formation of the high molecular weight bands increased with longer time light exposure and was more favorable under basic conditions than acidic conditions. His-His crosslinking of two heavy chains is likely to contribute to the band at ~100 kDa. The thioether crosslink between heavy chain and light chain is probably at the band of ~92 kDa. Other high molecular weight bands could be due to other unknown crosslinking.
4.4.2 Detection of Crosslinked Peptides.

Tryptic digestion in $^{18}$O-water results in the incorporation of two $^{18}$O atoms in each of the newly generated C-termini\[36, 37\]; hence, two $^{18}$O atoms for a linear tryptic peptide (with one C-terminus) and four $^{18}$O atoms for a crosslinked peptide (with two C-termini)\[38-40\]. As shown in Figure 4-2, the isotopic distribution of the peptide at m/z 1673.54 (quadruply charged, monoisotopic mass 6687.149 Da) show mass shift of 8 Da (i.e., four $^{18}$O) in $^{18}$O-water compared to that from $^{16}$O-water (Figure 4-2), indicating it contains two C-termini and is a crosslinked peptide.

Figure 4-2. Isotopic distributions of the crosslinked peptide S215-K244/S215-K244 m/z 1673.54 ($z=4$) from tryptic digestion of IgG1. A mass increase of 8 Da was observed when the sample was digested in $^{18}$O-water instead of regular $^{16}$O-water.
4.4.3 Elucidation of Crosslinking Chemistry.

The crosslinked peptide m/z 1673.54 underwent FT MS/MS analysis. As described in our previous paper[8], the fragment ions obtained were searched against the amino acid sequence of the IgG1 via FindPept to match all possible peptide fragments, see Table 4-1. Based on the peptide ladders observed, a partial sequence K218-K244 (KTHTCPPCPAPELLGGPSVFLFPPKPK, see Table 4-1) was identified. Then, the partial sequence was extended to a putative full-length tryptic peptide S215-K244 (SCDKTHTCPPCPAPELLGGPSVFLFPPKPK, 3336.587 Da). Since the fragment ions only matched this single peptide, we surmised that crosslinking occurred across the two identical peptides. The combined mass of the two unmodified (native) peptides is 6673.174 Da, which also satisfies the mass limitation conferred by the observed mass of the crosslinked peptide (6687.149 Da, see Table 4-2).

In order to elucidate the crosslinking chemistry, elemental composition analysis of the crosslink was performed as illustrated in Table 4-2. The mass difference between the sum of the two native peptide chains and observed mass of the crosslinked peptide is 13.975 Da, for which three potential formula (O-2H, N, or CH2) were proposed. From a chemistry perspective, it is difficult to add just one nitrogen atom or a CH2 group. On the other hand, addition of one oxygen atom coupled with the loss of two hydrogen atoms (O-2H) indicates oxidation. The putative peptide chain K218-K244 contains His, of which oxidation and crosslinking have been reported[28, 29]. In addition, the formula O-2H gives the lowest mass error (0.004 Da). Therefore, a potential His-His crosslinking structure is proposed as illustrated in Table 4-2 and Figure 4-3 and verified as described next.
Table 4-1. Partial sequences that match the mass of fragmentation ions for the precursor ion m/z 1673.54 (z=4) (molecular mass at 6687.149 Da) eluted at 112.48 min. The longest peptide fragments for the observed b- and y-ions are underlined. All Cys are alkylated with IAA. The adjoining residues before cleavage are in parenthesis. Lys443 is the C-terminus of heavy chain. The crosslinked peptides were proposed to be HC:S215-K244/HC:S215-K244.

<table>
<thead>
<tr>
<th>#</th>
<th>m/z</th>
<th>Charge</th>
<th>User mass (Da)</th>
<th>Theor. mass (Da)</th>
<th>Δmass (ppm)</th>
<th>peptide</th>
<th>Corresponding Tryptic Peptide</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>566.364</td>
<td>1</td>
<td>565.356</td>
<td>565.359</td>
<td>4.8</td>
<td>(F)PPKPK/(D)</td>
<td>(K)-SCDKHTHTCPPCPAPELGGPSVFLLFPKPK_{244}(D)</td>
<td>3336.587 chain 1</td>
</tr>
<tr>
<td>2</td>
<td>1256.736</td>
<td>1</td>
<td>1255.729</td>
<td>1255.733</td>
<td>3.0</td>
<td>(G)PSVFLFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1370.777</td>
<td>1</td>
<td>1369.770</td>
<td>1369.776</td>
<td>4.2</td>
<td>(L)GGPSVFLFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1483.863</td>
<td>1</td>
<td>1482.856</td>
<td>1482.86</td>
<td>2.5</td>
<td>(L)LGPSVFLFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1823.042</td>
<td>1</td>
<td>1822.034</td>
<td>1822.039</td>
<td>2.8</td>
<td>(A)PELLGGPSVFLFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1991.144</td>
<td>1</td>
<td>1990.136</td>
<td>1990.129</td>
<td>-3.4</td>
<td>(C)PAELLGGPSVFLFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1173.626</td>
<td>2</td>
<td>2345.238</td>
<td>2345.249</td>
<td>4.8</td>
<td>(C)PPCPAPELGGPSVFLFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1304.665</td>
<td>2</td>
<td>2607.315</td>
<td>2607.312</td>
<td>-1.2</td>
<td>(H)TCP CPAELLGGPSVFLF PKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1487.750</td>
<td>2</td>
<td>2973.485</td>
<td>2973.513</td>
<td>9.4</td>
<td>(D)KTHCPCPAELLGGPSVF LFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1478.766</td>
<td>2</td>
<td>2973.528</td>
<td>2973.513</td>
<td>-4.9</td>
<td>(D)KTHCPCPAELLGGPSVF LFPPKPK/(D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>992.341</td>
<td>1</td>
<td>1009.344</td>
<td>1009.348</td>
<td>4.0</td>
<td>(K)SCDKHTHC(P)</td>
<td>(K)-SCDKHTHTCPPCPAPELGGPSVFLLFPKPK_{244}(D)</td>
<td>Same as above</td>
</tr>
<tr>
<td>12</td>
<td>1515.547</td>
<td>1</td>
<td>1532.55</td>
<td>1532.558</td>
<td>5.3</td>
<td>(K)SCDKHTHTCPPCA(P)</td>
<td></td>
<td>Same as above</td>
</tr>
<tr>
<td>13</td>
<td>1605.255</td>
<td>2</td>
<td>3244.517</td>
<td>3244.493</td>
<td>-7.3</td>
<td>(K)SRWQQGNVFSCSVMHEALHN HYTQKSL(S)</td>
<td>(K)<em>{401}SRWQQGNVFSCSVMHEALHNHYTQKSLSLPGK</em>{441}</td>
<td>too large as chain 2</td>
</tr>
</tbody>
</table>

150
Table 4-2. Deduction of elemental formula for the crosslinked S215-K244/S215-K244 peptide.

<table>
<thead>
<tr>
<th>Name</th>
<th>Mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculated mass of S215-K244 (single chain)</td>
<td>3336.587</td>
</tr>
<tr>
<td>Sum of the mass of two unmodified chains</td>
<td>6673.174</td>
</tr>
<tr>
<td>Observed mass of the crosslinked peptide</td>
<td>6687.149</td>
</tr>
<tr>
<td>Mass difference</td>
<td>+13.975</td>
</tr>
<tr>
<td>Proposed formula and calculated mass</td>
<td></td>
</tr>
<tr>
<td>(mass error in Da)</td>
<td>+O-2H: 13.979 (0.004)</td>
</tr>
<tr>
<td>Comments</td>
<td>Most likely</td>
</tr>
</tbody>
</table>

4.4.4 Structural Confirmation by Mass Spectrometry

**Full Scan and MSMS Analysis** First, the calculated mass of the His-His crosslinked peptide (6687.153 Da) is in good agreement with the observed mass (6687.149 Da, mass error 0.6 ppm, see Table 4-3). Second, the series of b- and y-ions are highly consistent with the proposed structure (Figure 4-3). The observed y-ions from y5 to y24 and the b5 ions correspond to fragment ions with no crosslinking site, while the y-ions from y27 and the b-ions from b13 to b24 are from fragments that contain the crosslinked histidine residues. These data support crosslinking at His 220. Moreover, y27* ion (in blue) and b8* ion (in blue) are peptide fragments resulting from cleaving the bond connecting the two crosslinked histidine residues (see Figure 4-3). The missed cleavage by trypsin at Lys218 is likely due to its close proximity to
the crosslinking site at His220, reminiscent of similarly missed cleavages in the case of thioether crosslinking[9]. The second missed cleavage at Lys242 is likely due to the presence of adjacent proline residues. The two missed tryptic cleavages in the crosslinked peptide would have been especially challenging to handle by traditional database-dependent algorithms, again highlighting the utility of isotope labeling and our XChem-Finder workflow[8, 41, 42].
Table 4-3. Crosslinked peptides obtained from digestion of IgG1 by various proteases and the combination thereof. The crosslinking sites are labeled in red and bold-face. All cysteines are alkylated with IAA. Peptides are shown with the amino acid residue position in IgG1 in superscript and the adjoining amino acid residues before cleavage in parenthesis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Proteases</th>
<th>Crosslinked Peptides</th>
<th>RT (min)</th>
<th>m/z (Charge)</th>
<th>Obs. Mass (Da)</th>
<th>Theor. Mass (Da)</th>
<th>Mass Error (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S215-K244/</td>
<td>Trypsin</td>
<td>(K)\textsuperscript{215}SCDKTH\textsuperscript{HTCPPCPAPEL\textsuperscript{LGPSVFLFPKPK}\textsuperscript{244}(D)}</td>
<td>112.48</td>
<td>1673.54</td>
<td>6687.149</td>
<td>6687.153</td>
<td>0.6</td>
</tr>
<tr>
<td>S215-K244</td>
<td></td>
<td>(K)\textsuperscript{215}SCDKTH\textsuperscript{HTCPPCPAPEL\textsuperscript{LGPSVFLFPKPK}\textsuperscript{244}(D)}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D217-K244/</td>
<td>Asp-N</td>
<td>(C)\textsuperscript{217}DK\textsuperscript{HTCPPCPAPEL\textsuperscript{LGPSVFLFPKPK}\textsuperscript{244}(D)}</td>
<td>113.07</td>
<td>1549.53</td>
<td>6191.064</td>
<td>6191.060</td>
<td>0.6</td>
</tr>
<tr>
<td>D217-K244</td>
<td></td>
<td>(C)\textsuperscript{217}DK\textsuperscript{HTCPPCPAPEL\textsuperscript{LGPSVFLFPKPK}\textsuperscript{244}(D)}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S215-E229/</td>
<td>Trypsin + GluC</td>
<td>(K)\textsuperscript{215}SCDKTH\textsuperscript{HTCPPCPAPE}\textsuperscript{229}(L)</td>
<td>41.19</td>
<td>1178.77</td>
<td>3531.289</td>
<td>3531.286</td>
<td>0.6</td>
</tr>
<tr>
<td>S215-E229</td>
<td></td>
<td>(K)\textsuperscript{215}SCDKTH\textsuperscript{HTCPPCPAPE}\textsuperscript{229}(L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D217-E229/</td>
<td>Asp-N + GluC</td>
<td>(C)\textsuperscript{217}DK\textsuperscript{HTCPPCPAPE}\textsuperscript{229}(L)</td>
<td>40.53</td>
<td>1013.41</td>
<td>3035.193</td>
<td>3035.193</td>
<td>0.0</td>
</tr>
<tr>
<td>D217-E229</td>
<td></td>
<td>(C)\textsuperscript{217}DK\textsuperscript{HTCPPCPAPE}\textsuperscript{229}(L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D217-E229/</td>
<td>Trypsin + GluC; then Asp-N</td>
<td>(C)\textsuperscript{217}DK\textsuperscript{HTCPPCPAPE}\textsuperscript{229}(L)</td>
<td>40.31</td>
<td>1096.09</td>
<td>3283.234</td>
<td>3283.240</td>
<td>1.9</td>
</tr>
<tr>
<td>S215-E229</td>
<td></td>
<td>(K)\textsuperscript{215}SCDKTH\textsuperscript{HTCPPCPAPE}\textsuperscript{229}(L)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4-3. CID MS/MS spectra of the quadruply charged precursor ions m/z 1673.54 (16O-labeled C-termini) and 1675.54 (18O-labeled C-termini) of the crosslinked tryptic peptide S215-K244/S215-K244. Characteristic mass shift imparted by the heavier isotope 18O was observed (e.g., the mass shift of 4 Da for y5 ions in 16O-water, 566.53 vs 570.53). The y27* ion results from cleavage of the His-His bond while the y27 ion contains the crosslinking site. MS3 spectrum of the y27* ion (m/z 1488.35) is shown in Figure 4-4.
Additional confirmation by $^{18}\text{O}/^{16}\text{O}$-isotope fragment ions pattern. Since the fragment ions containing no (zero), one, or two C-termini of the crosslinked peptides displayed a mass shift of 0, 4, and 8 Da, respectively, in the corresponding MS/MS spectra obtained from $^{18}\text{O}$- and $^{16}\text{O}$-water, the examination of mass shift of fragment ions can lend further support for the assignment of fragment ions. For example, as shown in Figure 4-3, b-ions prior to the crosslinking site (e.g., b5) have no mass shift between the $^{18}\text{O}$-water and $^{16}\text{O}$-water digests. On the other hand, the y-ions without the crosslinking site (e.g., y5) gave mass shift of 4 Da. All assignments were verified by their distinct mass shift in $^{18}\text{O}$, depending on the number of C-termini they contain.

Additional confirmation by MS$^3$ analysis. Several abundant fragment ions shown in Figure 4-3 were selected for MS$^3$ analysis which simplified and further confirmed data interpretation. For example the fragment ion m/z 1488.35 shown in Figure 4-3 could not be assigned initially, so it was selected for MS$^3$ analysis (Figure 4-4). The analysis established that it was the y27* ion (in blue) generated from cleaving the bond connecting the two crosslinked histidine residues.
Peptide mapping with multiple proteases. Since this is the first report of His-His crosslinking in a protein, peptide mapping with additional proteases was carried out to glean complimentary data[43, 44]. In addition to trypsin, proteases with different sequence specificity (e.g. Asp-N[45-47] or GluC [48]) and combined proteases (e.g., trypsin with GluC, and Asp-N with GluC) were employed. Additional crosslinked peptides containing His220 were detected and analyzed: D217-K224/D217-K224 from Asp-N, S215-E229/S215-E229 from trypsin with GluC, and D217-E229/D217-E229 from Asp-N and GluC, respectively.

In each case, the observed mass was in good agreement with its theoretical mass with mass errors ranging from 0.0-0.6 ppm (see Table 4-3). The y- and b-ions were also consistent with the corresponding structure (Figure 4-5, 4-6, & 4-7). Similar to the tryptic peptide, the
crosslinking site and chemistry were further supported by the presence of several ions generated from cleaving the bond connecting the two crosslinked histidine residues, such as the doubly charged ion at m/z 880.67, the singly charged ion at m/z 1773.77, and the singly charged ion at m/z 992.52 (b8*) shown in Figure 4-5 (all highlighted in blue).

Figure 4-5. CID MS/MS spectrum of the triply charged precursor ion m/z 1178.77 of the crosslinked S215-E229/S215-E229 peptide generated from combined trypsin and GluC digestion. The b8* ion results from cleavage of the His-His bond while the b8 ion contains the crosslinking site.
Figure 4-6. CID MS/MS spectrum of the quadruply charged precursor ions m/z 1549.53 of the crosslinked peptide D217-K244/D217-K244 from Asp-N digestion.
Figure 4-7. CID MS/MS spectrum of the triply charged precursor ions at m/z 1013.41 of the crosslinked peptide D217-E229/D217-E229 from digestion with Asp-N and GluC.

**ETD MS/MS analysis.** As an alternative fragmentation technique, ETD provides sequence information complimentary to that obtained from CID by cleaving a peptide backbone in a less selective manner than CID [34, 49, 50]. Higher charge state ions usually generate more effective ETD fragmentation[50], therefore formic acid instead of TFA was used in the mobile phase to increase charge state for more effective ETD fragmentation and to minimize ion suppression. All ETD MS/MS spectra were collected with supplemental activation and
dominated by charge reduced species. The charge states of 5, 6, 4, and 4 for the peptide S215-K244/S215-K244, D217-K244/D217-K244, S215-E229/S215-E229, and D217-E229/D217-E229, respectively, offered optimal ETD fragmentation for each crosslinked peptide (Figure 4-8, 4-9, 4-10 & 4-11). While different than those from CID, the fragmentation patterns from ETD also support our proposed crosslinking site and chemistry. For instance, the c5 and c*8 ions in Figure 4-8 narrow the site within the HTC motif; the c4 ion in Figure 4-9, c5 and c6 ions in Figure 4-10, and c3 and z8 ions in Figure 4-11 pinpointed the crosslink at His220.

Figure 4-8. ETD MS/MS spectrum of the precursor ion m/z 1339.70 (z=5) of the crosslinked tryptic peptide S215-K244/S215-K244.
Figure 4-9. ETD MS/MS spectrum of the precursor ion m/z 1033.35 (z=6) of the crosslinked peptide D217-K244/D217-K244 from Asp-N digestion.
Figure 4-10. ETD MS/MS spectrum of the precursor ion m/z 884.33 (z=4) of the crosslinked peptide S215-E229/S215-E229 from digestion with trypsin and GluC.
Figure 4-11. ETD MS/MS spectrum of the precursor ion m/z 760.31 (z=4) of the crosslinked peptide D217-E229/D217-E229 from digestion with Asp-N and GluC.

4.4.5 Mechanism of formation for His-His crosslink.

Photo-oxidation and crosslinking between histidine residues have been studied using both free histidine and model peptides. The commonly accepted mechanism is depicted in Scheme 4-
1. Singlet oxygen (e.g., generated from photoactivated dye rose bengal[51]) reacts with histidine to form a highly reactive and labile endoperoxide intermediate, which converts into a hydroperoxide intermediate and then 2-oxo-histidine (2-oxo-His) and His+32 intermediates. Subsequently, the His+32 intermediate can be attacked by the nucleophilic imidazole of another histidine residue; followed by the elimination of a water molecule to give the final crosslinking product (Scheme 4-1)[16, 27-29, 52, 53]. As discussed below, our results are consistent with this mechanism.

Scheme 4-1. Proposed mechanism for the formation of His-His crosslink via photo-oxidation intermediates.
First, oxygen was present in all buffers and water in which IgG1 was exposed to light irradiation. Second, several photo-oxidation intermediates were observed. The endoperoxide intermediate is unstable and has only been observed by low-temperature NMR study[16, 27], so we are not surprised that it was not detected by our LC-MS analysis. However, the subsequent oxidation intermediates, 2-oxo-His (+14 Da) and His+32 species (+32 Da), were detected. The peptides with masses 14 and 32 Da greater than the unmodified peptide, S215-K244 (SCDKTHTCPPCPAPELLGGPSVFLFPPKPK), were observed in the light stressed samples but not in the control sample (Table 4-4). Tandem mass spectra confirmed their structures to be the peptides modified at His220 (Figures 4-12 and 4-13). Third, the reported model studies showed the crosslinking was favored at higher pH, as the neutral (deprotonated) imidazole in histidine (pKa ~6) is more reactive for nucleophilic attack and thus results in a higher yield of crosslinking[54]. Similar pH dependence was observed in our case as discussed above (Figure 4-1). Lastly, the two His220 residues are juxtaposed in the hinge region, which is highly exposed to solvent and flexible, as illustrated in Figure 4-14. In fact, in most crystal structures, the side chains of residues in the hinge region could not be located, indicating a high degree of flexibility. In this illustrative structure (PDB 1HZH), side chain of only one histidine residue was observed.
Table 4-4. Peptides containing the 2-oxo-His (+14 Da) and His+32 (+32 Da) intermediates observed in the stressed IgG1. They were not detected in the control. His220 residues are labeled in red. All cysteine are alkylated with IAA. Peptides are shown with the amino acid residue position in IgG1 in superscript and the adjoining amino acid residues before cleavage in parenthesis. The level of each peptide is determined by peak area of modified peptide over normal tryptic peptide (T219-K224) with the consideration of all charge states.

<table>
<thead>
<tr>
<th>#</th>
<th>Peptide Name</th>
<th>Peptide Sequence</th>
<th>RT (min)</th>
<th>m/z (Charge)</th>
<th>Obs. Mass (Da)</th>
<th>Theor. Mass (Da)</th>
<th>Mass Error (ppm)</th>
<th>Level by Peak Area (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T219-K224</td>
<td>(K)\textsuperscript{219} THTCPPCPAPELGGPSVFLFPPKP\textsuperscript{244}D</td>
<td>93.98</td>
<td>949.82</td>
<td>2845.422</td>
<td>2845.418</td>
<td>1.3</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>S215-K224</td>
<td>(K)\textsuperscript{215} SCDKTHHTCPPCPAPELGGPSVFLFPPKP\textsuperscript{244}D</td>
<td>92.17</td>
<td>1113.87 (3+)</td>
<td>3336.591</td>
<td>3336.587</td>
<td>1.1</td>
<td>83.1</td>
</tr>
<tr>
<td>3</td>
<td>2-oxo-His</td>
<td>(K)\textsuperscript{215} SCDKTH(+14)TCPPCPAPELGGPSVFLFPPKP\textsuperscript{244}D</td>
<td>93.48</td>
<td>1118.20 (3+)</td>
<td>3350.570</td>
<td>3350.566</td>
<td>1.1</td>
<td>2.3</td>
</tr>
<tr>
<td>4</td>
<td>His+32</td>
<td>(K)\textsuperscript{215} SCDKTH(+32)TCPPCPAPELGGPSVFLFPPKP\textsuperscript{244}D</td>
<td>93.00</td>
<td>1124.20 (3+)</td>
<td>3368.578</td>
<td>3368.577</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>S215-K244</td>
<td>(K)\textsuperscript{215} SCDKTHHTCPPCPAPELGGPSVFLFPPKP\textsuperscript{244}D</td>
<td>112.48</td>
<td>1673.54 (4+)</td>
<td>6687.149</td>
<td>6687.153</td>
<td>0.6</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>S215-K244</td>
<td>(K)\textsuperscript{215} SCDKTHHTCPPCPAPELGGPSVFLFPPKP\textsuperscript{244}D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4-12. CID MS/MS spectrum of the tryptic peptide containing the 2-oxo-His (+14 Da) intermediate (m/z=1118.20, z=3, mass=3350.570 Da).
Figure 4-13. CID MS/MS spectrum of the tryptic peptide containing the His+32 intermediate (m/z=1124.20, z=3, mass=3368.578 Da).
Figure 4-14. (A) Space filling illustration of the hinge region of IgG1 antibody (DKTHTCPCPPC); the underlined and bold residues are shown in color; the atoms are shown in color: oxygen in red, nitrogen in dark blue, carbon in light blue and sulfur in yellow. The image is rendered based on PDB 1HZH using the VMD software (Visual Molecular Dynamics). In this structure, His237 is equivalent to the His220 described in the paper. (B) Three-dimensional (3D) structure of an IgG1 (PDB entry 1HZH[55]). The characteristic hinge region sequences (SCDKHTCPPC) of two heavy chains of IgG1 are circled. The first cysteine is disulfide bonded with the C-terminal cysteine in light chain. The other two cysteines form inter-heavy chain disulfide bridges. Two heavy chain histidines (His237) are located in the hinge region which is very flexible and highly solvent accessible. The model of 1HZH is being used for illustrative purposes only; in this structure, His237 is equivalent to the His220 discussed in our paper.
Based on the reaction pathway and protein structure, crosslinking of lysine with the oxidized histidine via nucleophilic addition is also plausible[16, 27], and Lys218 is in the vicinity of His220. Therefore, great effort was made to determine whether the crosslinking is His220-His220 or Lys218-His220. This is particularly challenging due to the pseudo-symmetry in the crosslinked peptide; in other words, when the two chains share identical sequence (e.g., in Figure 4-3, 4-5, and 4-8), any fragment ion could come from either one chain or both. For example, c4 and c5 ions in Figure 4-8 indicated the existence of unmodified Lys218 and Thr219, but could not unambiguously establish whether they were from one chain or both. To address this issue, an asymmetric crosslinked peptide (i.e., two chains of different length) was generated via limited digestion. IgG1 fully digested by trypsin and GluC was treated with Asp-N for a limited time to obtain a crosslinked peptide with two different chains D217-E229/S215-E229 (Figure 4-15 and 4-16). Its precursor ion m/z 1096.09 (z=3) has an observed mass of 3283.234 Da, which is in agreement with the theoretical mass of 3283.240 Da (Table 4-3). As shown in Figures 4-15 and 4-16, cleavage of the His-His bond resulted in ions m/z 1511 and m/z 1773, indicating that the oxidized His residue is on the long chain highlighted in red. Moreover, the c2 and c3 ions from the short chain (highlighted in blue) together with the c2 to c5 ions from the long chain (highlighted in red) indicate the absence of modification for all residues N-terminal to His220 on both chains, thus ruling out crosslinking between Lys218 and His220. This is not unexpected, as at the pH for our studies, the amine on the lysine side chain is mostly protonated and thus renders it unreactive[56, 57]. And, of course, others factors such as local environment and solvation are known to modulate reactivities in enzymes and antibodies[58-60]. Taken together, our data have firmly established that the crosslinking is between the two heavy chain His220 residues.
Figure 4-15. ETD MS/MS spectrum of the quadruply charged precursor ion m/z 821.09 of the crosslinked peptide D217-E229/S215-E229 generated by limited Asp-N digestion of fully digested IgG1 by trypsin and GluC.
Figure 4-16. CID MS/MS spectrum of the triply charged precursor ions at m/z 1096.09 of the crosslinked peptide D217-E229/S215-E229 from limited digestion by Asp-N of the fully digested IgG1 by GluC and trypsin.

4.4.6 Other Crosslinks.

As reported in literature[9-11], crosslinking via thioether between the heavy chain hinge region and the light chain C-termini (HC:S215-K218/LC:T211-S218, SCDK/TVAPTECS) was also observed in the photo-irradiated IgG1 (Table 4-5). The cleavage and formation of carbon-
sulfur (C-S) bonds may occur via either homolytic (e.g., radical or photo-induced) or heterolytic (e.g., elimination and addition) mechanisms[10, 11, 61, 62]. These additional crosslinks may also account for the multiple non-reducible higher molecular species detected by SDS-PAGE and SEC described above (Figure 4-1). In the SDS-PAGE gel (Figure 4-1), the first band (with an apparent molecular weight about 92 kD) for the sample pH9-3xLight (lane 10) was not observed in the samples pH4.8-3xLight (lane 8) and pH7-3xLight (lane 9). It is likely that this band corresponds to the thioether crosslink of LC-HC, as it is favorable under basic conditions[9, 10]. Although the thioether crosslink between two heavy chains (e.g. Cys222-Cys222) has been reported after a higher dose of photo irradiation[11], it was not detected in our sample by MS. The His-His crosslink of two heavy chains may contribute to the band with an apparent molecular weight about 100 kDa. The bands with apparent molecular weight about 150 and 200 kDa are probably due to the crosslinking of more than two chains.

Table 4-5. Thioether crosslinks detected in IgG1. The second thioether crosslink has two missed trypsin cleavages at Lys218 and Lys242 probably due to nearby thioether crosslink site and proline amino acid residues, respectively.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>RT (min)</th>
<th>m/z (charge)</th>
<th>Obs. Mass (Da)</th>
<th>Theo. Mass (Da)</th>
<th>Mass Error (ppm)</th>
<th>Level (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC:T211-S218/HC:S215-K218</td>
<td>22.11</td>
<td>612.77 (2+)</td>
<td>1223.535</td>
<td>1223.534</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>85TVAPTECS18/86SCDK218</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC:T211-S218/HC:S215-K244</td>
<td>90.97</td>
<td>1351.99 (3+)</td>
<td>4050.947</td>
<td>4050.942</td>
<td>1.2</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>85TVAPTECS18/86SCDKTHTCPPCPAPELGGPSVFLFPPKPK244</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.5 Conclusions

Our XChem-Finder workflow again leads to the discovery of an undefined and novel protein His-His crosslink, demonstrating its broad applicability and utility. Since the His-His crosslink is found in the highly conserved hinge region of IgG1, this modification most likely exists in other IgG1 molecules. As discussed above, a low level of crosslinking was present even without light stress, suggesting protein crosslinking in therapeutic proteins is perhaps more common than we have appreciated. Such drastic modification of proteins is likely to affect product quality, clinical efficacy, and even at low abundance, immunogenicity. And again, to the best of our knowledge, there is no other alternative systematic approach that can be generally used to fully characterize protein crosslinking without a priori knowledge of the chemistry and site. With the rapid advancement in mass spectrometric techniques (e.g. high resolution and complementary fragmentation mechanisms), we expect the discovery and elucidation of other new protein crosslinking by our XChem-Finder approach will be equally successful.

4.6 References


Chapter 5: Conclusion and Future Directions

In this thesis, LCMS-based methods have been developed for the detection and characterization of Asp isomerization and protein crosslinks in monoclonal antibodies. As described in Chapter Two, a protein isoaspartate methyltransferase-mediated $^{18}$O-labeling followed by LC/MS analysis is described to detect one of the most challenging post translational modifications—isomerization of Asp due to the subtle difference between isoaspartic acid and aspartic acid and difficulty to differentiate them. Several isoAsp peptides in IgG1 were characterized and isomerization site were unambiguously identified.

In Chapter Three, a comprehensive methodology for the identification of protein crosslinks without a prior knowledge of chemistry via $^{18}$O-labeling and LC/MS analysis is presented. Due to the intrinsic structural complexity of crosslink, it is very challenging to detect and characterize crosslinks, especially when crosslink chemistry is unknown. The utility of our XChem-Finder work flow has been successfully demonstrated via the detection of thioether crosslinks in IgG2 and the discovery of a novel histidine-histidine crosslink in IgG1 in Chapter Three and Four, respectively. Both thioether crosslinks and histidine-histidine crosslinks are found as degradation compounds in the constant region of IgG, therefore these modifications are most likely common to IgG.

Despite considerable efforts to understand the relevance of post-translational modifications such as Asp isomerization and protein crosslinking in the cellular context, we are still in the process of unraveling the complexity of these modifications and their tremendous impact. Sophisticated technological advances like powerful separation techniques, high
resolution mass spectrometry are now increasingly available for identification and characterization of these site specific protein modifications. In this chapter, some future work to extend the utilities of our methods for each project is listed next. For crosslink project, the improvement of our current XChem-Finder work flow will be discussed in more detail.

5.1 isoAsp Project

**Biological Samples.** The method described in Chapter Two can be useful for monitoring isoAsp formation in not only IgG but also other therapeutic proteins during production and storage to ensuring the quality of the therapeutic proteins. Most importantly, the method can be applied to biological samples such as plasma, serum, urine and tissues to identify potential isoAsp proteins to understand the isoAsp process pathways in vivo and to identify disease-associated biomarkers.

As discussed in the section 1.4.1, artifactual deamidation or isomerization can be significant during sample preparation. The extreme pH exposure and high temperature should be avoided during sample preparation. The sample preparation for biological samples typically takes longer time to remove other interference proteins. Therefore, caution should especially be taken to minimize artifactual deamidation or isomeration. The inherent isoAsp and those introduced by sample preparation can be differentiated by preparing sample in $^{18}$O-water and quantified by b-ion intensity calculation procedure (See detail discussion in the section 1.4.3.4)[1-3].

**D-isoAsp Detection.** It is difficult to identify the peptides containing D-isoAsp. So far no sensitive method for D-isoAsp has been reported. The lack of suitable methods has prevented
its biological study. Antibodies are highly specific against their specifically modified proteins[4, 5], but there is very limited reports on antibody-based method yet. It is worthy to explore more in this area in the future.

5.2 Crosslink Project

Other Crosslinks in Proteins Our XChem-Finder work flow has successfully been used to detect thioethers and histidine-histidine crosslinks in IgG. We expect the discovery and elucidation of other new protein crosslinking in pharmaceutical products and biological samples by our XChem-Finder approach will be equally successful. Dityrosine crosslinks in protein has been proposed and studied since 1980s, but its analysis most often employ reversed phase HPLC with fluorometric detection due to its complicated structure in nature and difficulties on its MS data interpretation[6-8]. Calmodulin contains two tyrosyl residues with no cysteine or tryptophan, so it is a good model protein to study dityrosine crosslink[8]. Our XChem-Finder work flow might be applied to directly detect and characterize dityrosine crosslink in calmodulin. This might shed a light on the formation of dityrosine crosslink in other proteins (for example, proximities of two tyrosine residues in proteins).

Sample Enrichment. Our current XChem-Finder process to detect the crosslinks is to directly use tryptic digest from both $^{16}$O and $^{18}$O-water for LC/MS/MS analysis. A chromatographic fractionation of peptic digests may be necessary sometime to avoid suppression of the ionization of some peptides. Various methods to achieve such enrichment are now explored. One such approach that has so far been employed for enrichment of chemically crosslinked peptides makes use of the generally higher charge state that distinguishes chemically
crosslinked peptides from linear peptides[9]. Chemically crosslinked peptides elute in the late fractions in cation-exchange chromatography[9]. Also peptides with high charge states are selected for fragmentation in MS analysis[9]. Taouatas and his colleague combined Lys-N proteolytic digestion, strong cation exchange enrichment, and mass spectrometry (MALDI-MS/MS by CID and LC-MS/MS by CID or ETD) to achieve an optimal targeted strategy for proteome analysis[10-12]. The lack of an enrichment step in the current work flow means that only the most abundant of crosslinks are likely to be found. In the case of photo-degraded IgG1, only histidine-histidine crosslink was discovered in our study although there are highly likely other crosslinks via radical mechanism in the sample. Development of specific and efficient enrichment strategies may help to discover new crosslinks. For example, Lys-N digestion of SEC (size exclusion chromatography) high molecular fractions followed by CEX (cation exchange chromatography) enrichment for LC/MS analysis will be explored in the future (Scheme 5-1)[10-12].
Scheme 5-1. The use of combining sample enrichment and Lys-N digestion for detection of crosslinks[10-12]. Two enrichment steps via SEC at protein level and CEX at peptide level are used to reduce sample complexity. Lys-N peptides that do or do not contain a basic amino acid (e, g. His or Arg. His is used as an example here.) are shown with the charge in the scheme. The Lys-N peptides result in dominant b-ions in MALDI-CID (matrix assisted laser ionization-collision induced dissociation) and c-ions in ETD facilitating MS/MS data interpretation.

**N-Terminal Labeling.** In this thesis, the incorporation of $^{18}$O at C-termini of each newly created peptide was described to distinguish the crosslinked peptides from linear peptides. However, it failed to detect the crosslinked peptides containing C-termini of proteins. Isotope labeling at N-termini with $^2$H$_3$-2,4-dinitrofluorobenzen ([$^2$H$_3$]NDFB) as described in section 1.5.3.2.2 might be useful. In the original protocol, methylation of ε–amino group of lysine and N-terminal tag was conducted before protease digestion which results in miss-cleavage due to dimethyld lysines resistance to protease digestion[13]. This can be overcome by simply switching sample preparation order—protease digestion first followed protection of ε–amino group of Lys and then specific derivatization of the N-terminal amino group with [$^2$H$_3$]DNFB
This will increase cost/consumption of $^{2}$H$_{3}$DNFB reagent, but can minimize miss-cleavage to get too large crosslinked peptides.

Scheme 5-2. Isotopic labeling at N-termini via 1) trypsin digestion; 2) protection of ε–amino group of lysine by reductive methylation; 3) specific derivatization of N-terminal amino group with a 1:1 mixture of DNFB (2, 4-dinitrofluorobenzene) and $^{2}$H$_{3}$DNFB at pH 7.0. The resulting crosslinked and linear peptides can isotopically be identified (1:2:1 triplet and 1:1 doublet with a space of 3 Da for crosslinked and linear peptides, respectively)[13].

In addition, succinylation is known to modify peptides at N-terminal and ε–amino group[14-17]. In the effort to develop a protein quantification method based on isobaric peptide termini labeling, the use of 2-Methoxy-4,5-dihydro-1H-imidazole reacts first with ε–amino group followed by N-termini labeling with succinic anhydride and teterdeuterated succinic anhydride-d4 (Scheme 5-3)[14, 17]. Koehler at al recently reported succinylation selectively
occurred at N-terminal amino group using sodium acetate buffer at pH 7.6 (Scheme 5-3)[15, 16]. In principal, the site-specific N-terminal succinylation can be used for the detection of crosslink peptides. Succinic anhydride specifically isotope label N-termini which results in mass increase of 100 Da and 200 Da for single peptides and crosslinked peptides, respectively (Scheme 5-3). We plan to exploit the succinylation to improve the detection of crosslink peptides in the future.

Scheme 5-3. N-Terminal Succinylation via two-step chemical derivatizations[14, 17] and site specific N-terminal succinylation in sodium acetate buffer pH 7.6[15, 16] can be used to differentiate crosslinked peptides and linear peptides.
Digestion under Acidic Conditions. Trypsin digestion is often performed at slight basic condition which may make some crosslinks unstable. Protease digestion in acidic condition may be explored.

Deglycosylation. Heterogeneity and relatively poor ionization efficiency of glycopeptides increase difficulty to determine the crosslinks near the site of glycan attachment. Deglycosylation before $^{18}$O-labeling and protease digestion is worth to explore in the future.

Complicated Crosslinks. Proteins are known to be degraded in light exposure by a number of mechanisms, which is of concern for products manufactured for the clinic. Protein degradation in the light can involve multiple amino acid residues and form a combination of multiple degradation pathways. This gives very complicated mass spectra which pose great challenge to current methodologies including our XChem-Finder work flow.

Intra-crosslinks. In our XChem-Finder work flow, $^{18}$O-labeling followed by tryptic digestion and LC/MS analysis is used to differentiate single peptides and crosslinked peptides. This approach may not be suitable for intra-crosslinks when the two crosslinked amino acid residues are close so that no tryptic cleavage between them occurs. This can be addressed with the combination of multiple protease digestion (e.g., GluC/Trpsin) and $^{18}$O-labeling, N-terminal labeling or chemical tag. We are going to explore this area in the future.

Others As mention before, some limitation of XChem-Finder (e.g. crosslinks containing C-terminus of protein, high quality MSMS spectra, large crosslinks, etc.) can be addressed by use of protease with different selectivity (e.g., GluC instead of Trypsin), N-terminal labeling and different ion activation (CID and ETD). Chemical tag may provide some solution as well and might be studied in the future. In addition, MALDI often generates single charge ions while ESI
gives multiple charge ions for large peptides and proteins. As such MALDI instead of ESI may simplify the data interpretation of crosslinks. Next, we are going to exploit MALDI for characterization of crosslinks.

5.3 References


