N-terminal Imine Derivatization for Enhanced De Novo Peptide Sequencing: A Study of the Fragmentation Pattern Generated from CID of Peptide-Imines

Khiry L. Patterson

Duquesne University

Follow this and additional works at: https://dsc.duq.edu/etd

Recommended Citation

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection.
N-TERMINAL IMINE DERIVATIZATION FOR ENHANCED DE NOVO PEPTIDE SEQUENCING: A STUDY OF THE FRAGMENTATION PATTERN GENERATED FROM CID OF PEPTIDE-IMINES

A Dissertation
Submitted to the Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for
the degree of Doctor of Philosophy

By
Khiry L. Patterson

April 2020
N-TERMINAL IMINE DERIVATIZATION FOR ENHANCED DE NOVO
PEPTIDE SEQUENCING: A STUDY OF THE FRAGMENTATION PATTERN
GENERATED FROM CID OF PEPTIDE-IMINES

By

Khiry L. Patterson

Approved April 3, 2020

___________________________________________________________
Michael Van Stipdonk, Ph.D.
Professor of Chemistry and Biochemistry
(Committee Chair)

Michael Cascio, Ph.D.
Associate Professor of Chemistry and Biochemistry
(Committee Member)

Jeffrey D. Evanseck, Ph.D.
Professor of Chemistry and Biochemistry
John V. Crable Thesis Chair
Lambert F. Minucci Endowed Chair
in Engineering and Computational Science
(Committee Member)

Theodore Corcovilos, Ph.D.
Assistant Professor of Physics
Bayer School of Natural and
Environmental Sciences
(Outside Reader)

Ellen Gawalt, Ph.D.
Hillman Distinguished Professor
Chair, Department of Chemistry
and Biochemistry

Phillip Reeder, Ph.D.
Dean, Bayer School of Natural and
Environmental Sciences
ABSTRACT

N-TERMINAL IMINE DERIVATIZATION FOR ENHANCED DE NOVO PEPTIDE SEQUENCING: A STUDY OF THE FRAGMENTATION PATTERN GENERATED FROM CID OF PEPTIDE-IMINES

By

Khiry L. Patterson

April 2020

Dissertation supervised by Dr. Michael Van Stipdonk

Tandem mass spectrometry (MS/MS) and collision-induced dissociation (CID) remain two of the most essential tools used for peptide and protein identification in proteomics. Notably, the unique product ion distributions generated from CID are used, in part, to determine peptide/protein sequence, whether by comparison to known peptide fragmentation patterns, or searching sequence databases that attempt to predict fragmentation. However, an impediment to reliable identification by interpretation of CID mass spectra may be caused by an incomplete understanding of peptide fragmentation mechanisms, including sequence rearrangement pathways (sequence scrambling) at increasingly higher stages of MS/MS. This work reports the product ion distributions and fragmentation mechanisms derived from CID of model imine-modified peptide ions that can inform and enhance MS/MS de novo sequencing efforts via N-terminal derivatization.
In this work, model peptides were derivatized by reaction with salicylaldehyde to create N-terminal imines (Schiff bases). Collision-induced dissociation of the protonated peptide-imines generally produced complete series of the sequence informative a and b-ions, now undoubtedly characteristic of the peptide ion species. However, a novel product ion was also observed, denoted the yǂ ion, determined by IRMPD spectroscopy and density functional theory to be generated from the elimination of the N-terminal amino acid residue despite the N-terminal modification. It was concluded that the pathway involved a nucleophilic attack by an amide N atom and the possible formation of an imidazole-4-one intermediate, which collapses to generate a truncated, protonated peptide-imine with a conserved sequence structure.

N-terminal imine-modification was postulated to eliminate sequence scrambling events, presumably by eliminating the macrocyclic b-ion mechanism implicated in the sequence rearrangements. The fragmentation pattern of imine-modified leu-enkephalin unambiguously revealed that CID of the peptide-imine species does not generate previously determined scrambled ions observed for unmodified YGGFL.

CID mass spectra of Ag-cationized imine-modified peptides were also obtained. Interestingly, an apparent even-electron, [M-H]+ peptide ion was generated by the loss of AgH. Isotope labeling explicitly identified the imine C-H group as the hydrogen source, implicating an imine → nitrile transformation. CID of the [M–H]+ ions generated sequence ions analogous to those produced from the [M+H]+ species of imine-modified peptides, however less extensively.
DEDICATION

I dedicate this work to my grandfather James V. Gilliam Sr., grandmother Loretta Ishman, and aunt Regina Zachery, whose love and support is felt even long after you have gone. Thank you for encouraging me and telling everyone you know how proud of me you are. I also dedicate this work to my mother, Terria; father, Terry; and sister, Imani – your love and laughter have kept me resilient.
ACKNOWLEDGMENTS

Thank you to everyone who has guided me along this path and made this research possible. I want to thank Dr. Michael Van Stipdonk, for accepting me into his research group and challenging me along the way. The atmosphere of community, hard work, and the opportunity he has fostered within his research group have been my great pleasure to study in each day. I would also like to sincerely thank Dr. Jeffrey D. Evanseck – without your encouragement and guidance at the start of this journey; I imagine I would not be here at the end able to share my deepest gratitude.

To my research, groupmates – thank you for all the help over the years – every insightful suggestion, discussion, challenge, laugh, and thoughtful commentary. Thank you to the Duquesne University Department of Chemistry and Biochemistry faculty and staff for every class and every helping hand along the way. I am thankful to my dissertation committee members, Drs. Michael Cascio, Jeffrey D. Evanseck, and Theodore Corcovilos for all your assistance and encouragement. I would also like to express my appreciation to the FELIX laboratory at Radboud University for the use of their facilities to acquire IRMPD spectroscopy data. I also thank Duquesne University, Department of Chemistry and Biochemistry, and the Bayer School for Natural and Environmental Sciences for the incredible opportunity to be one of the first in my family to pursue graduate studies.

To my ever-vast extended family, friends, and church - thank you for every listening ear, call, prayer, textbook, meal, and hug you have given, it has made this journey humbling and reaching the goal sweeter than if I had to do it alone.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xiii</td>
</tr>
<tr>
<td>List of Equations</td>
<td>xiv</td>
</tr>
<tr>
<td>List of Schemes</td>
<td>xv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xvii</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>xxii</td>
</tr>
<tr>
<td>CHAPTER 1: MASS SPECTROMETRIC CHARACTERIZATION OF PEPTIDES AND PROTEINS</td>
<td>1</td>
</tr>
<tr>
<td>1.1 An introduction to proteomics</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The significance of protein structure to protein function</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Mass spectrometric approaches to proteomics</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Protein database searching</td>
<td>6</td>
</tr>
<tr>
<td>1.5 De novo sequencing</td>
<td>8</td>
</tr>
<tr>
<td>1.6 Protonated peptide fragmentation</td>
<td>9</td>
</tr>
<tr>
<td>1.7 Protonated peptide fragmentation pathways</td>
<td>12</td>
</tr>
<tr>
<td>1.7.1 bₓ − yz pathway</td>
<td>12</td>
</tr>
<tr>
<td>1.7.2 Side-chain effect: non-classical b-ion</td>
<td>13</td>
</tr>
<tr>
<td>1.7.3 a₁ − yz pathway</td>
<td>14</td>
</tr>
</tbody>
</table>
1.7.4 Aziridinone pathway..............................................................................................................15
1.7.5 Secondary fragmentation: $b_x - b_{x-1}$ pathway.................................................................16
1.7.6 Secondary fragmentation: internal immonium ions............................................................18
1.7.7 Loss of small neutrals............................................................................................................19
1.8 Example of direct de novo sequencing.....................................................................................21
1.9 Sequence scrambling................................................................................................................24
1.10 Enhancing de novo peptide sequencing..................................................................................26
   1.10.1 Metal-cationization........................................................................................................27
   1.10.2 Peptide derivatization....................................................................................................29
1.11 References.............................................................................................................................31

CHAPTER 2....................................................................................................................................41

TECHNIQUES: GAS PHASE ANALYSIS OF PEPTIDES AND PROTEINS.................................41

   2.1 Mass spectrometry.................................................................................................................41
   2.2 Electrospray ionization..........................................................................................................44
   2.3 Mass analyzer......................................................................................................................46
      2.3.1 Introduction..................................................................................................................46
      2.3.2 Linear ion trap..............................................................................................................46
   2.4 Tandem mass spectrometry....................................................................................................50
      2.4.1 Fragmentation by collision-induced dissociation.........................................................53
      2.4.2 Fragmentation by infrared multiple photon dissociation..........................................58
      2.4.3 Theoretical vibrational IR spectra...............................................................................59
CHAPTER 3

METHODS

3.1 N-terminal amine to imine derivatization

3.1.1 Sample preparation

3.1.2 CID experiments

3.1.3 CID energy/time-resolved experiments

3.2 Characterization of N-terminal amino acid elimination from protonated peptide-imines

3.2.1 Sample preparation

3.2.2 CID experiments

3.2.3 ESI FT-ICR mass spectrometry

3.2.4 Infrared multiple photon dissociation (IRMPD)

3.2.5 DFT geometry optimization and frequency calculations

3.3 Metal-cationized peptide-imines

3.3.1 Sample preparation

3.1.2 CID experiments

3.4 References

CHAPTER 4

FRAGMENTATION PATTERN OF PROTONATED, N-TERMINAL PEPTIDE-IMINES

4.1 Introduction
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Typical mass spectrometer characteristics</td>
<td>43</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>DFT predicted energies of minima for possible protonated oHB=AGG Structures and product ion structures generated from loss of 71 u</td>
<td>110</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>List of sequence ions generated from CID of protonated benzyl-, hydroxybenzyl, and methoxybenzyl-imine derivatized AGG</td>
<td>125</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Experimentally determined CID profile curve E_{50%} and appearance energies at activation time, t = 1000 ms</td>
<td>131</td>
</tr>
</tbody>
</table>
## LIST OF EQUATIONS

<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equation 1.1</strong></td>
<td>False discovery rate (FDR)</td>
<td>6</td>
</tr>
<tr>
<td><strong>Equation 2.1</strong></td>
<td>Accelerated ion kinetic energy</td>
<td>41</td>
</tr>
<tr>
<td><strong>Equation 2.2</strong></td>
<td>General precursor ion fragmentation reaction</td>
<td>50</td>
</tr>
<tr>
<td><strong>Equation 2.3</strong></td>
<td>General precursor ion fragmentation by photodissociation</td>
<td>58</td>
</tr>
<tr>
<td><strong>Equation 2.4</strong></td>
<td>Frequency-dependent fragmentation ion intensities</td>
<td>59</td>
</tr>
<tr>
<td><strong>Equation 4.1</strong></td>
<td>Low-mass cut off limit of the LIT</td>
<td>83</td>
</tr>
<tr>
<td><strong>Equation 5.1</strong></td>
<td>CID Fragmentation yield (FY)</td>
<td>128</td>
</tr>
<tr>
<td><strong>Equation 5.2</strong></td>
<td>CID Product ion yield (PY)</td>
<td>140</td>
</tr>
</tbody>
</table>
# LIST OF SCHEMES

| Scheme 1.1 | b<sub>x</sub> – y<sub>z</sub> pathway | .......................................................... 12 |
| Scheme 1.2 | Non-classical b-ion | .......................................................... 13 |
| Scheme 1.3 | a<sub>1</sub> – y<sub>z</sub> pathway | .......................................................... 14 |
| Scheme 1.4 | aziridinone pathway | .......................................................... 15 |
| Scheme 1.5 | b<sub>x</sub> – b<sub>x-1</sub> pathway | .......................................................... 16 |
| Scheme 1.6 | b<sub>x</sub> – a<sub>x</sub> – b<sub>x-1</sub> pathway | .......................................................... 17 |
| Scheme 1.7 | immonium ion | .......................................................... 18 |
| Scheme 1.8 | ammonia loss | .......................................................... 19 |
| Scheme 1.9 | water loss | .......................................................... 20 |
| Scheme 1,10 | sequence scrambling | .......................................................... 25 |
| Scheme 1.11 | charge-remote [b<sub>x</sub> + OH]<sup>+</sup> pathway | .......................................................... 29 |
| Scheme 3.1 | amine to imine derivatization (Schiff base) | .......................................................... 65 |
| Scheme 4.1 | Proposed sidechain elimination from protonated, imine-derivatized TGG | .......................................................... 89 |
| Scheme 4.2 | Proposed pathways for elimination of 71 u from protonated oHB=AGG | .......................................................... 91 |
| Scheme 4.3 | Proposed mechanisms for loss of 94 u from protonated oHB=XGG peptide species | .......................................................... 92 |
| Scheme 4.4 | Proposed sequence ion generating intramolecular hydrogen transfer pathways for non-protonated [oHB=PGG]<sup>+</sup> | .......................................................... 99 |
| Scheme 4.5 | Proposed mechanisms for loss of 97 u from non-protonated [oHB=PGG]<sup>+</sup> | .......................................................... 101 |
| Scheme 4.6 | Proposed mechanisms for the elimination of 57 u from product ion m/z 297 generated through elimination of 71 u from protonated oHB=AGGG | .......................................................... 106 |
Scheme 5.1 Proposed mechanism for generating N-terminal immonium ions from protonated peptide-imines

Scheme 5.2 Proposed mechanism for protonated pHB=peptide charge migration mediated by the hydroxybenzyl-substituent

Scheme 5.3 Proposed mechanism for protonated oHB=peptide charge migration mediated by the hydroxybenzyl-substituent, (termination cycle)

Scheme 6.1 Proposed mechanism for AgH elimination from Ag⁺-cationized imine-modified peptides

Scheme 6.2 Proposed mechanism for y⁺ formation from even-electron [M-H]⁺ ions
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Peptide/Protein Structure</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Proteomic Approaches</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Pathways in competition (PIC)</td>
<td>11</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Sequence ions</td>
<td>11</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Model MS^n spectrum</td>
<td>23</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>General mass spectrometer schematic</td>
<td>43</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Electrospray ionization</td>
<td>45</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Ion traps</td>
<td>49</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Tandem Mass Spectrometry</td>
<td>52</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Collision-induced dissociation</td>
<td>54</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Warhaftig diagram</td>
<td>57</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>Potential energy surface</td>
<td>60</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>CID mass spectrum of</td>
<td>77</td>
</tr>
<tr>
<td>Figure 4.1a</td>
<td>protonated YGGFL (MS^2)</td>
<td></td>
</tr>
<tr>
<td>Figure 4.1b</td>
<td>b_4 product ion (MS^3)</td>
<td></td>
</tr>
<tr>
<td>Figure 4.1c</td>
<td>a_4 product ion (MS^3)</td>
<td></td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>CID mass spectrum of</td>
<td>79</td>
</tr>
<tr>
<td>Figure 4.2a</td>
<td>protonated oHB=YGGFL (MS^2)</td>
<td></td>
</tr>
<tr>
<td>Figure 4.2b</td>
<td>b_4 product ion (MS^3)</td>
<td></td>
</tr>
<tr>
<td>Figure 4.2c</td>
<td>a_4 product ion (MS^3)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.3  CID-MS$^2$ mass spectrum of protonated, underivatized XGG peptides .......................................................... 82

Figure 4.3a  GGG  
Figure 4.3b  AGG  
Figure 4.3c  VGG  
Figure 4.3d  LGG  
Figure 4.3e  SGG  
Figure 4.3f  TGG  
Figure 4.3g  FGG  
Figure 4.3h  YGG

Figure 4.4  CID-MS$^2$ mass spectrum of protonated XGG peptides, imine-derivatized by reaction with salicylaldehyde (oHB) ........................................................................................................... 85

Figure 4.4a  GGG  
Figure 4.4b  AGG  
Figure 4.4c  VGG  
Figure 4.4d  LGG  
Figure 4.4e  FGG  
Figure 4.4f  YGG

Figure 4.5  CID-MS$^2$ mass spectrum of protonated XGG peptides, imine-derivatized by reaction with salicylaldehyde (oHB) ........................................................................................................... 88

Figure 4.5a  SGG  
Figure 4.5b  TGG

Figure 4.6  CID-MS$^2$ mass spectrum of protonated GGA .................................................................................................... 95

Figure 4.6a  underivatized  
Figure 4.6b  imine-derivatized by reaction with benzaldehyde (Bz)  
Figure 4.6c  imine-derivatized by reaction with salicylaldehyde (oHB)

Figure 4.7  CID-MS$^2$ mass spectrum of protonated PGG .................................................................................................... 97

Figure 4.7a  underivatized  
Figure 4.7b  imine-derivatized by reaction with salicylaldehyde (oHB)  
Figure 4.7c  imine-derivatized by reaction with benzaldehyde (oHB)

Figure 4.8  Proposed structure of peptides, imine-derivatized by reaction with salicylaldehyde (oHB) ................................................................................................................................. 98

Figure 4.8a  PGG peptide ion, (non-protonated ion)  
Figure 4.8b  AGG peptide ion, (protonated)
Figure 4.9  CID mass spectrum of........................................................................................................104

Figure 4.9a  protonated AGGG imine-derivatized by reaction with salicylaldehyde (MS^2)

Figure 4.7b  product ion m/z 294 generated from loss of 71 u (MS^3)

Figure 4.7c  protonated GGG imine-derivatized by reaction with salicylaldehyde (MS^2)

Figure 4.10  Nomenclature for sequence ions generated from CID of y^‡ ions..........................108

Figure 4.11  Lowest energy structures for protonated oHB=AGG predicted by DFT..........................................................110

Figure 4.12  IRMPD spectrum of........................................................................................................111

Figure 4.12a  protonated oHB=AGG

Figure 4.12b  protonated oHB=AGG compared to DFT predicted spectrum for imine-N protonated oHB=AGG

Figure 4.12c  protonated oHB=AGG compared to DFT predicted spectrum for amide-O protonated oHB=AGG

Figure 4.13  Possible DFT predicted structures of product ions generated by elimination of 71 u from protonated oHB=AGG ........................................................113

Figure 4.14  IRMPD spectrum of........................................................................................................114

Figure 4.14a  product ion m/z 237 generated from loss of 71 u from protonated oHB=AGG

Figure 4.14b  product ion m/z 237 compared to DFT of predicted imidazole-one pathway product

Figure 4.14c  product ion m/z 237 compared to DFT of predicted ketopiperazine pathway product

Figure 4.14d  product ion m/z 237 compared to IRMPD spectrum of protonated oHB=GG

Figure 5.1  CID-MS^2 mass spectrum of protonated AGG, imine-derivatized by reaction with ....................................................................................................................123

Figure 5.1a  Benzaldehyde (Bz)

Figure 5.1b  p-Hydroxybenzaldehyde (pHB)

Figure 5.1c  m-Hydroxybenzaldehyde (mHB)

Figure 5.1d  o-Hydroxybenzaldehyde (oHB)
Figure 5.2 CID-MS$^2$ mass spectrum of protonated AGG, imine-derivatized by reaction with..................................................................................124

Figure 5.2a Benzaldehyde (Bz)
Figure 5.2b p-Methoxybenzaldehyde (pMB)
Figure 5.2c m-Methoxybenzaldehyde (mMB)
Figure 5.2d o-Methoxybenzaldehyde (oMB)

Figure 5.3 Collision energy resolved profiles for protonated benzyl- and hydroxybenzyl- AGG at an activation time of...............................................................132

Figure 5.3a t = 10 ms
Figure 5.3b t = 100 ms
Figure 5.3c t = 1000 ms

Figure 5.4 Collision energy resolved profiles for protonated benzyl- and methoxybenzyl- AGG at an activation time of...............................................................133

Figure 5.4a t = 10 ms
Figure 5.4b t = 100 ms
Figure 5.4c t = 1000 ms

Figure 5.5 Resonance structures for the protonated...............................................................135

Figure 5.5a Benzylimine peptide ion
Figure 5.5b m-Hydroxybenzylimine peptide ion

Figure 5.6 Resonance structures for the protonated...............................................................136

Figure 5.6a o-Hydroxybenzylimine peptide ion
Figure 5.6b p-Hydroxybenzylimine peptide ion

Figure 5.7 Collision energy resolved profiles for the formation of the y$^\parallel$ ion from protonated benzyl- and hydroxybenzyl- AGG at an activation time of........141

Figure 5.7a t = 10 ms
Figure 5.7b t = 100 ms
Figure 5.7c t = 1000 ms

Figure 5.8 Collision energy resolved profiles for the formation of the y$^\parallel$ ion from protonated benzyl- and methoxybenzyl- AGG at an activation time of........142

Figure 5.8a t = 10 ms
Figure 5.8b t = 100 ms
Figure 5.8c t = 1000 ms

xx
Figure 6.1  CID-MS\(^2\) mass spectrum of argentinated

- Figure 6.1a  oMB=GGG
- Figure 6.1b  oMB=VAAF
- Figure 6.1c  oMB=YGGFL

Figure 6.2  CID-MS\(^2\) mass spectrum of argentinated

- Figure 6.2a  Bz=AGG
- Figure 6.2b  d\(_1\)-Bz=AGG

Figure 6.3  CID mass spectrum of the

- Figure 6.3a  [M-H]\(^+\), even-electron ion, generated by AgH elimination from argentinated oMB=GGG (MS\(^3\))
- Figure 6.3b  [M+H]\(^+\), protonated oMB=GGG (MS\(^2\))
- Figure 6.3c  protonated, underivatized GGG (MS\(^2\))

Figure 6.4  CID mass spectrum of the

- Figure 6.4a  [M-H]\(^+\), even-electron ion, generated by AgH elimination from argentinated oMB=AGG (MS\(^3\))
- Figure 6.4b  [M+H]\(^+\), protonated oMB=AGG (MS\(^2\))
- Figure 6.4c  protonated, underivatized AGG (MS\(^2\))

Figure 6.5  CID mass spectrum of the

- Figure 6.5a  [M-H]\(^+\), even-electron ion, generated by AgH elimination from argentinated oMB=VAAF (MS\(^3\))
- Figure 6.5b  [M+H]\(^+\), protonated oMB=VAAF (MS\(^2\))
- Figure 6.5c  protonated, underivatized VAAF (MS\(^2\))

Figure 6.6  CID mass spectrum of the

- Figure 6.6a  [M-H]\(^+\), even-electron ion, generated by AgH elimination from argentinated oMB=YGGFL (MS\(^3\))
- Figure 6.6b  [M+H]\(^+\), protonated oMB=YGGFL (MS\(^2\))
- Figure 6.6c  protonated, underivatized YGGFL (MS\(^2\))
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>alternating current</td>
</tr>
<tr>
<td>AE</td>
<td>appearance energy</td>
</tr>
<tr>
<td>Bz</td>
<td>benzyl</td>
</tr>
<tr>
<td>CID</td>
<td>collision induced dissociation</td>
</tr>
<tr>
<td>DC</td>
<td>direct current</td>
</tr>
<tr>
<td>DDA</td>
<td>data dependent acquisition</td>
</tr>
<tr>
<td>DFT</td>
<td>density functional theory</td>
</tr>
<tr>
<td>DIA</td>
<td>data independent acquisition</td>
</tr>
<tr>
<td>E&lt;sub&gt;50%&lt;/sub&gt;</td>
<td>collision energy required to fragment 50% of the precursor ion intensity</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>FEL</td>
<td>free electron laser</td>
</tr>
<tr>
<td>FELIX</td>
<td>Free Electron Laser for Infrared eXperiments</td>
</tr>
<tr>
<td>FT-ICR</td>
<td>Fourier Transform Ion Cyclotron Resonance</td>
</tr>
<tr>
<td>FY</td>
<td>fragment yield</td>
</tr>
<tr>
<td>HB</td>
<td>hydroxybenzyl</td>
</tr>
<tr>
<td>HRMS</td>
<td>high-resolution mass spectrometry</td>
</tr>
<tr>
<td>ICR</td>
<td>ion cyclotron resonance</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>IRMPD</td>
<td>infrared multiple photon dissociation</td>
</tr>
<tr>
<td>IVR</td>
<td>intramolecular vibrational distribution</td>
</tr>
</tbody>
</table>
LFDR  local false discovery rate
LIT   linear ion trap
LRMS  low-resolution mass spectrometry
[M+H]^+ protonated ion
[M-H]^+ even-electron ion
MALDI matrix-assisted laser desorption ionization
MB    methoxybenzyl
mHB   meta-hydroxybenzyl
MHC   major histocompatibility complex
MS    mass spectrometry
MS/MS tandem mass spectrometry
MS^n  tandem mass spectrometry @ mass analysis stage n = 1, 2, 3…
m/z   mass-to-charge ratio
NCE   normalized collision energy
NIm^+ N-terminal immonium ion
NMR   nuclear magnetic resonance
oHB   ortho-hydroxybenzyl
PA    proton affinity
PES   potential energy surface
PFP   peptide fragmentation pathway
pHB   para-hydroxybenzyl
PIC   Pathways in competition
PMF   peptide mass-fingerprinting method
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSM</td>
<td>peptide-spectrum match</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>PY$_{\text{max}}$</td>
<td>maximum product ion yield</td>
</tr>
<tr>
<td>PY</td>
<td>product ion yield</td>
</tr>
<tr>
<td>QET</td>
<td>quasi equilibrium theory</td>
</tr>
<tr>
<td>Q-q-Q</td>
<td>triple quadrupole</td>
</tr>
<tr>
<td>Sal</td>
<td>salicylaldehyde</td>
</tr>
<tr>
<td>TOF</td>
<td>time-of-flight</td>
</tr>
<tr>
<td>$y^+$</td>
<td>y “double-dagger” ion, representative of product ions generated from eliminating the residue mass of the N-terminal amino acid</td>
</tr>
</tbody>
</table>
CHAPTER 1

MASS SPECTROMETRIC CHARACTERIZATION OF PEPTIDES AND PROTEINS

1.1 An introduction to Proteomics

As so many of the cellular processes in biological systems are governed and dependent upon interactions at the molecular level, it is crucial to characterize the structure, behavior, and function of biomolecules to address the complexities of health, disease, and disease prevention/intervention. Many fields of study are dedicated to understanding biological systems, each with a particular focus on a critical biomolecule. These include the complete set of an organisms’ DNA (genomics), a cell or cell populations set of RNA (transcriptomics), and – germane to this dissertation – the proteins produced and modified by an organism/system (proteomics).

Specifically, proteomics is the large-scale study of protein interaction, expression, function, and structure. Discovery in proteomics arguably yields some of the most significant information for understanding the mechanisms of life, as proteins are a diverse and complex set of biomolecules directly involved in virtually every chemical process present in the cells of living organisms. Proteins work as antibodies binding to specific foreign molecules (i.e., viruses and bacteria), enzymes executing chemical reactions, as transport carrying small molecules in and out of cells, and as messengers transmitting signals to coordinate biological processes. Inescapably, investigating biological mechanisms require in-depth proteomic
studies to illuminate protein function – a feat that must include accurately identifying protein structure.

1.2 The significance of protein structure to protein function

Proteins are made up of a series of amino acids covalently bonded together. Each amino acid residue (a term used to indicate the loss of water during amide bond formation) has an identifiable molecular mass (except isobaric residues leucine and isoleucine) arising from its alpha carbon side chain that varies in size, structure, and electric charge (Figure 1.1a and Figure 1.1b. The sequence of amino acids is known as a protein's primary structure (Figure 1.1c), where intermolecular interactions (i.e., hydrogen bonds, ionic bonds, van der Waals attractions) between residues and hydrophobicity, direct how polypeptides fold into unique higher-level protein structures.9-11 The arrangement of the 20 common amino acids into different combinations/sequences gives rise to considerably different three-dimensional structures, thus properties and functions.

Changes in sequence can alter hydrophobicity and solubility, give rise to disease (i.e., sickle-cell anemia and Duchenne muscular dystrophy), and destroy the ligand binding site on a protein surface.9, 10, 12 Even folding and unfolding rates of two- and three-stage proteins can be predicted directly from sequence.12, 14 Evolutionary relatedness/families can be established from sequence or conserved residues between sequences and used to predict protein function by comparing the discovered sequence to that of previously studied proteins.10, 15 Undoubtedly, accurate sequencing of peptides and proteins is invaluable to efforts for uncovering protein bioinformation.
Figure 1.1 Peptide/protein structure. a) 20 common amino acids b) amino acid residue masses c) stages of protein structure; primary (sequence), secondary (regions stabilized by hydrogen bonding between atoms), tertiary (3D polypeptide shape determined by regions stabilized by side-chain interactions), quaternary (two or more associated polypeptide regions).
Many bioanalytical techniques have been used to progress proteomic studies such as but not limited to gel electrophoresis,\textsuperscript{16} liquid chromatography,\textsuperscript{17} nuclear magnetic resonance (NMR) spectroscopy,\textsuperscript{18} X-ray crystallography,\textsuperscript{18} and arguably one of the most essential and adaptive techniques for elucidating protein/peptide structure and sequence – mass spectrometry.\textsuperscript{19-25}

1.3 Mass spectrometric approaches to proteomics

Mass spectrometry (MS) (discussed in more detail in section 2.1) is an essential analytical method that has been used extensively to determine peptide/protein sequence and structure.\textsuperscript{26-35} Biological MS utilizes soft ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) to produce intact protonated peptide and protein ions (referred to as \textit{precursor ions} or \textit{parent ions}) in the gas phase. The formed $[M+zH]^{z+}$ ions (where $z$ is the number of protons added to give charge to the molecule) are then separated according to their mass-to-charge ratios (m/z) by a mass analyzer.\textsuperscript{26-29} Additionally, tandem mass spectrometry (MS/MS or MS$^n$) (described in more detail in section 2.4) is a technique that uses energetic collisions with a non-reactive gas to fragment a precursor ion into smaller product ions.\textsuperscript{36, 37} The result of both mass spectrometry and tandem mass spectrometry is a mass spectrum constructed by plotting ion intensity as a function of m/z, from which the m/z values of either precursor or products ions are used to determine or confirm peptide/protein identity.\textsuperscript{36, 37}
Figure 1.2 Proteomic approaches. Top-down and Bottom-up proteomics are the two generalized overarching approaches to proteomic studies. Shotgun bottom-up proteomics indicates the specific identification of proteins in complex mixtures using high-performance liquid chromatography (HPLC) MS.38-41

Mass spectrometric studies are typically approached either through “top-down” also known as “protein-centric” and “bottom-up” also known as “peptide-centric” strategies (Figure 1.2).38-41 In top-down proteomics, the m/z values of intact ionized proteins (precursor ions) and the m/z values of the resulting fragment ions (product ions) are used to determine the proteins amino acid sequence.38-40 Bottom-up proteomics classically utilizes the peptide mass-fingerprinting method (PMF) where proteins in complex mixtures are first separated then cleaved into smaller peptide components by site-specific enzymatic digestion (e.g., trypsin, Lys-C, pepsin) before mass analysis.42-44 The smaller peptide fragments are ionized and separated using a mass analyzer (MS), followed by the isolation of a fragment ion and subsequent
fragmentation event in the gas phase (i.e. CID (section 2.4.1)) and mass analysis of the resulting product ions (MS\textsuperscript{n}).\textsuperscript{42-44} The identification of peptides and proteins is completed by matching the unique mass spectra to the mass spectra of known peptides using database search algorithms.\textsuperscript{26}

1.4 Protein database searching

One of the essential aspects of bottom-up proteomics is the protein database software and algorithms (such as SEQUEST and MASCOT) used to identify peptides from mass spectrometry data.\textsuperscript{45-49} As mentioned above, database search algorithms in general attempt to match a combination of precursor ion m/z values and product ion m/z values to those predicted from \textit{in silico} digestion or theoretical fragmentation of the peptides derived from the \textit{in silico} digestion of a protein with a known sequence.\textsuperscript{26} The database software then uses a scoring function to evaluate the “matching” quality of each experimental mass spectrum to one contained in a database, where the highest score is known as the peptide-spectrum match (PSM).\textsuperscript{48, 49}

The validity of a PSM is often established using the \textit{target-decoy method}. This method accompanies a “nonsense” or “decoy” database of similar size and amino acid composition to the database of known, genome-derived proteins typically by reversing each protein’s amino acid sequence in the database.\textsuperscript{50, 51} At a given PSM score a corresponding false discovery rate (FDR, Equation 1.1) is calculated to quantitate the quality of the score in order to avoid a false positive.\textsuperscript{51, 52}

\textbf{Equation 1.1} \textit{FDR @ score,}\(s = \frac{\text{the number of incorrect matches with a score > } s}{\text{Total number of matches with a score > } s}\)
The FDR corresponds to the ratio of two times the number of decoy database PSMs and the total number of PSMs. The probability that a spectrum with a given FDR score is misidentified can be determined by calculating the local FDR (LFDR), defined at a specific score to be the number of incorrect matches with the same score divided by the number of matches with that score.

The quality and validity of database methods can negatively be affected, resulting in a false positive PSM by factors including:

1. Low-quality mass spectra.
2. Enzymatic digestion at abnormal sites.
3. Errors in charge state determination.
5. Incomplete bioinformatics (including missing and undiscovered proteins to an insufficient understanding of the complex fragmentation chemistry).
6. Incomplete (limited) database search space.

The latter factor may arguably be particularly deleterious to peptide/protein identification, given database search space is often restricted by experimenters to limit the number of variable post-translational modifications (PTMs) that can negatively affect peptide sequence matching by generating false-positive identifications. Customization of the database search space in this way has been shown to improve the false discovery rate. However, if the true sequence of the sample includes a PTM that lay outside the limited list, then the scoring algorithm will misidentify the protein/peptide. Additionally, PTMs would alter the mass-to-charge ratio and product ion distribution. Consequently, modified peptides are thought to comprise at least one third of unassigned mass spectra. This hole in bioinformation is particularly problematic, given
many PTMs significantly impact the specificity and activity of proteins and have been implicated in disease processes.\textsuperscript{61, 62} Similar approaches to eliminate false positives caused by abnormally cleaved peptides can also be employed for database sequencing.\textsuperscript{58} However, disregarding PTMs, and abnormal or inadequately understood peptide fragmentation presents a disadvantage of database sequencing methods that can lead to unsuspecting misidentifications, particularly of artificially created, novel, mutated, modified or lesser known peptides/proteins. In these cases, when database searches can be significantly compromised, the next recourse is to employ bottom-up, \textit{de novo sequencing} strategies.

\subsection*{1.5 De novo sequencing}

De novo peptide sequencing is the process of direct interpretation of peptide product ions (aka. fragment ions) generated from tandem (MS\textsuperscript{n}) mass spectrometry experiments.\textsuperscript{63, 64} The method uses the mass difference between fragment ions to calculate the unique mass of amino acid residues from the peptide backbone. The masses and mass differences of the product ions are then used to piece together the correct amino acid sequence.\textsuperscript{26} Computer software is also available that utilizes de novo sequencing to predict and assign peptide fragment mass values and mass differences.\textsuperscript{65-67} Additionally, the generated product ions can be isolated, fragmentated, and the resulting mass spectrum directly interpreted for sequence information – utilizing the power of multiple MS\textsuperscript{n} stages possible with ion-trapping instruments (section 2.3.2).\textsuperscript{26} In this way, peptide sequencing through direct interpretation of mass spectra generated from the fragmentation of specifically novel and/or lesser documented peptides/proteins can be achieved where database methods are unsuccessful. Inevitably, de novo peptide sequencing must be utilized advance new discoveries in proteomics, revealing the central scientific aim of this
research – to enhance de novo peptide sequencing methods. However, it is important to note either by software or by direct interpretation of mass spectra, the task of peptide sequencing is better facilitated by a sufficient understanding of peptide fragmentation and many studies have sought to advance this knowledge.²⁶-³⁵, ⁶⁸-⁷⁰

1.6 Protonated peptide fragmentation

The widely accepted approach to understanding the dissociation of protonated peptides is the pathways in competition model (PIC).²⁶ It explains that under low-energy collision conditions the dissociation of protonated peptides can be described as a competition between charge-remote and charge-directed peptide fragmentation pathways (PFPs).²⁶ The complicated fragmentation process can more easily be broken up into three stages: pre-cleavage, bond cleavage, and the post-cleavage (Figure 1.3).²⁶ During the pre-cleavage phase the ionizing “mobile” proton may be positioned at any competing reactive site along the peptide backbone.²⁶ Intramolecular interactions along the peptide backbone then lead to intermediate structures that are activated by some mechanism (i.e., CID (section 2.4.1), IRMPD (section 2.4.2)) leading to bond cleavage.²⁶ The resulting products are a protonated product ion observable by mass spectrometry and a neutral fragment(s) indicated by the mass difference between the product and precursor ion.²⁶

Charge-remote pathways are distinguished by intramolecular interactions that lead to backbone cleavages that occur physically remote from the position of the ionizing proton.²⁶ The fragment ions generated from charge-remote pathways are self-descriptively termed non-sequence ions.²⁶ For charge-directed pathways however, the intramolecular interactions are facilitated by the migration and position of the ionizing proton to the various reactive sites on
the peptide backbone, leading to bond cleavages physically near the protonation site.\textsuperscript{26, 71-75}

Peptide dissociation as a result of charge-directed pathways primarily lead to product ions termed sequence ions, described by three different types of sequence identifying peptide bond cleavages that keep amino acid side chains intact (Figure 1.4): the “a/x” alkyl carbonyl bond (CHR-CO), the “b/y” peptide amide bond (CO-NH), and the “c/z” aminoalkyl bond (NHCHR).\textsuperscript{26, 29,76} The a/x, b/y, and c/z sequence ions take on the form \( P_x \), where \( P \) is the ion type of the corresponding bond cleavage and the subscript \( x \) represents \( u-1 \) where \( u \) is equal to the number of amino acid residues in the complete peptide sequence. Note the subscript \( x \) is interpreted from the N- to the C-terminus for a-, b-, and c-type ions and from the C- to the N-terminus for x-, y-, and z-type ions. Though each bond cleavage has the potential to generate two types of sequence ions (a|x, b|y, and c|z respectively), the product ion that predominates in the mass spectrum is in part determined by whichever species has the higher affinity for the ionizing proton, as well as the internal energy distribution of the ion population.\textsuperscript{26, 77, 78} As a result, the generation of a classical sequence ion can be described as a dissociation of an intermediate dimer ion.
**Pathways in competition (PIC)**

Figure 1.3 Pathways in competition (PIC). Flowchart of competing dissociation pathways.\(^\text{26}\)

**Sequence ions**

Figure 1.4 Sequence ions. The dotted lines indicate bond cleavage resulting from peptide ion energetic collisions with a nonreactive gas (i.e., helium). a, b, and c ions contain the N-terminus, while x, y, and z ions contain the C-terminus. Not shown is the ionizing proton (hydrogen) that directs peptide bond cleavage.\(^\text{26, 76}\)
1.7 Protonated peptide fragmentation pathways

Determining the dissociation pathway leading to the generation of product and sequence ions from protonated peptides is essential to creating an adequate understanding of the ion types fragmentation chemistry. Additionally, working fragmentation pathways with defined initiation steps and product ion structures enhances the ability to interpret the fragmentation pattern and behavior of both precursor and product ions. Given under low energy CID the commonly observed sequence ions generated from protonated peptides are the a-, b-, and y-type ions, the fragmentation pathways of these common ions have been investigated.26

1.7.1 bₓ – yₓ pathway

![Scheme 1.1 bₓ – yₓ pathway](image-url)
Migration of the ionizing proton from one reactive site to the nitrogen of the amide bond to be cleaved initiates the \( b_x - y_z \) pathway (Scheme 1.1).\(^{26} \) Nucleophilic attack by the neighboring N-side oxygen on the protonated amide bond carbon center leads to the proton bound \( b_x \) oxazolone – \( y_z \) dimer under low energy collision conditions. The formation of either protonated monomer after CID is determined by the thermochemistry of the two possible sequence ion species and each monomers proton affinity.\(^{26} \) Contrastingly, proton migration to an amide oxygen has also been suggested for \( b_x - y_z \) formation.\(^{26} \) Though amide oxygen protonation is energetically favored over protonation at the amide nitrogen, the dissociation channels of the amide oxygen pathway are not thought to be active under low energy CID.\(^{26} \)

**1.7.2 Side-chain effect: non-classical b-ion**

![Scheme 1.2 Non-classical b-ion](image)
Nucleophilic attack by the amino acid side chain nitrogen of Arg, Lys, Asp, Glu, and His (or by the side chain carbonyl oxygen of Gln and Asn) on the neighboring C-side protonated amide bond carbon center also leads to amide bond cleavage but with “non-classical” b-ion formation (Scheme 1.2).\textsuperscript{26} Competition with the b\textsubscript{x} - y\textsubscript{z} pathway can lead to an isomeric mixture of the oxazolone and non-classical b-ion forms.

1.7.3 \textit{a\textsubscript{1} – y\textsubscript{z} pathway}

\textbf{Scheme 1.3 \textit{a\textsubscript{1} – y\textsubscript{z} pathway}}

14
After the migration of the ionizing proton to the second N-terminal amino acid amide nitrogen, a trimer of a protonated N-terminal and C-terminal monomer with a loosely bound CO loses the CO forming the N-side and C-side dimers (Scheme 1.3).\textsuperscript{26} Dissociation of this dimer leads to the formation of either the $a_1$ or $y_z$ ion.

1.7.4 Aziridinone pathway

Scheme 1.4 Aziridinone pathway
The aziridinone pathway is a charge directed pathway that results in the formation of an aziridinone - $y_z$ proton-shared dimer (Scheme 1.4). If dissociation of the dimer results in the nitrogen protonated aziridinone monomer, structural instability leads to $a_1$ ion formation through the loss of CO. In comparison to the $a_1 - y_z$ pathway, theoretical investigations of dipeptides have indicated the aziridinone pathway to be less kinetically and energetically favored.

1.7.5 Secondary fragmentation: $b_x - b_{x-1}$ pathway

Scheme 1.5 $b_x - b_{x-1}$ pathway
Although b-ions are typically generated through dissociation of protonated peptide precursor ions, b-ions can also be generated via breakdown of product ions. Investigation of the $b_x \rightarrow b_{x-1}$ pathway for various Leu-enkephalin derivatives have shown approximately 50 % of the observed $b_3$ ions originate directly from the breakdown of the corresponding $b_4$ ions. However, the mechanistic details of the transformation remain unclear. Scheme 1.5 depicts a proposed one-step charge directed mechanism where the adjacent amide carbonyl oxygen attacks the oxazolone imine carbon to generate an aziridinone derivative and the protonated $b_{x-1}$ ion.

---

Scheme 1.6 $b_x - a_x - b_{x-1}$ pathway
In contrast Scheme 1.6 depicts an indirect charge directed mechanism that first goes through the \( b_x \rightarrow a_x \) pathway transformation by the loss of CO.\(^{26}\) This is followed by an \( a_x \rightarrow b_{x-1} \) transformation via a nucleophilic attack on the C-terminal a-ion carbonyl carbon by the adjacent carbonyl oxygen – losing an immonium ion and forming the oxazolone \( b_{x-1} \) ion. However, investigation of these two pathways has indicated neither mechanism is likely for \( b_x \rightarrow b_{x-1} \) formation under low energy collision conditions, even despite the validity of the \( b_x \rightarrow a_x \) pathway.\(^{26}\)

### 1.7.6 Secondary fragmentation: internal immonium ions

Scheme 1.7 Immonium ion
Often MS^n spectra also contain immonium ions originating from internal positions along the protonated peptide backbone. These internal immonium ions are a result of charge directed secondary fragmentation pathways as depicted in Scheme 1.7, originating from the bₓ – yₓ pathway. Elimination of CO from the generated bₓ ion oxazolone ring results in the formation of the aₓ-1 and immonium ion, while elimination of CO from the generated yₓ ion results in an immonium ion and/or smaller C-terminal containing yₓ product ions.

1.7.7 Loss of small neutrals

Ammonia loss

Scheme 1.8 Ammonia loss
Under low energy collision conditions small neutral losses (i.e., ammonia and water) are a common subset of charge-directed peptide fragmentations, commonly generating [M + H – NH₃]⁺, [M + H – H₂O]⁺, [bₓ – NH₃]⁺, [bₓ – H₂O]⁺, [yₓ – NH₃]⁺, and [yₓ – H₂O]⁺ product ions.⁶ Ammonia loss typically occurs from the side chains of Asn, Gln, Arg, and Lys, and is particularly abundant in the fragmentation of peptides formed from tryptic digests as Lys or Arg occupy the C-terminus.⁶ Scheme 1.8 depicts ammonia loss from an N-terminal lysine side chain, initiated by protonation of the side chain amino group (similarly for Gln and Asn, but followed by tautomerization of the protonated Arg guanidine group).⁶ The N-terminal nitrogen then attacks the adjacent carbon causing the loss of ammonium and the formation of a cyclic group made up of the long amino acid side chain and N-terminal nitrogen. Lys, Arg, Asn, and Glu found at other positions along the peptide backbone can also experience ammonia loss, however by nucleophilic attack by the neighboring C-terminal carbonyl oxygen.⁶

Scheme 1.9 Water loss
Water loss can occur by dehydration at the side chains of Ser, Thr, the carboxylic acid group of the Asp and Glu side chains, and the C-terminus.\textsuperscript{26} C-terminal water loss is depicted in Scheme 1.9 beginning with proton migration to the amide oxygen adjacent to the last amino acid residue.\textsuperscript{26} The next steps include the concerted transfer of the proton to the C-terminal hydroxy group and release of water, followed by the formation of an oxazolone ring by nucleophilic attack of the amide oxygen. Water loss from Asp and Glu positioned at the N-terminus proceeds by mobilization of the proton to the side chain hydroxy group and attack on the side-chain carboxy carbon center by the N-terminal nitrogen.\textsuperscript{26} Ser and Thr lose water by protonation of the side chain hydroxy group followed by an attack by the adjacent amide oxygen on the adjacent side chain carbon forming a stable five- or six-membered ring.\textsuperscript{26}

1.8 Example of direct de novo sequencing

Generating sequence ions is imperative to the success of de novo peptide sequencing. In contrast to high-energy CID which promotes non-sequence ion formation, low-energy CID conditions, as those classically used in linear ion trap mass spectrometers, promote sequence ion formation and specifically mass spectra rich in a, b, and y-ions.\textsuperscript{26} Without question, \textit{the more types of sequence ions present in a mass spectrum enhance de novo efforts.} However, a straightforward residue-by-residue identification of peptide sequence can be achieved if the full y-ion or b-ion series is generated.

Consider Figure 1.5 depicting a model MS\textsuperscript{2} spectrum for model protonated precursor peptide ion [M+H]\textsuperscript{+} = [P-E-P-T-I-D-E]\textsuperscript{+}. The mass difference between the model [M+H]\textsuperscript{+} ion and the m/z 654.77 peak is equal to 146.15 u, indicating the mass loss of the amino acid residue E + 17 u (the C-terminal hydroxy group) - thus making the m/z 654.77 peak the b\textsubscript{6} ion. Similarly,
the mass difference between the m/z 539.68 ion and b₆ ion peak indicates the loss of the next amino acid residue, D, and thus the formation of the b₅ ion. Continuing this process of identification, all the residues and thus, the amino acid sequence can be determined. Additional stages of MSⁿ can be used to verify product ion assignments. However, some factors present a significant challenge to the effectiveness of de novo peptide sequencing:⁵⁵

1. Incorrect assignment of sequence ions
2. Missing sequence ions, specifically within the b and y series
3. Crowding of the mass spectrum with non-sequence ions
4. Presence of isobaric residues (i.e., K/Q and I/L)

Arguably, the most challenging to the de novo approaches to peptide/protein identification is the phenomenon of sequence scrambling which can render MSⁿ spectra uninterpretable.
Figure 1.5 Model MS^n spectrum. [M+H]^+ is the imaginary protonated peptide ion, [PEPTIDE]^+. The model amino acid depicts, within the yellow box, what portion of the amino acid determines the residue mass.

Model MS^n spectrum

Model amino acid

Residue

<table>
<thead>
<tr>
<th>Residue</th>
<th>Residue Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>115.09</td>
</tr>
<tr>
<td>E</td>
<td>129.14</td>
</tr>
<tr>
<td>I</td>
<td>133.17</td>
</tr>
<tr>
<td>P</td>
<td>97.12</td>
</tr>
<tr>
<td>T</td>
<td>101.11</td>
</tr>
</tbody>
</table>
1.9 Sequence scrambling

The structure of b-ions have been identified as linear N-terminus containing fragment ions with conserved primary amino acid sequence, terminated by a five or six membered oxazolone ring at the C-terminus. Recent experiments strongly suggest that after amide bond cleavage a macrocyclic b-ion intermediate can arise through cyclization of the b-ion (Scheme 1.10). The macrocyclic b-ion can then open at various ring positions causing a scrambling of the peptide sequence. As a result, the CID-MS \textsuperscript{n} spectrum at the MS\textsuperscript{3} stage becomes representative of the dissociation of various perturbations of the original peptide sequence, thus cloaking its identity. CID-MS\textsuperscript{0} studies of b-ions YAGFL-NH\textsubscript{2}, AGFLY-NH\textsubscript{2}, GFLYA-NH\textsubscript{2}, FLYAG-NH\textsubscript{2}, LYAGF-NH\textsubscript{2}, and cyclo-YAGFL performed by Bleiholder et al. have demonstrated that these different peptides of similar amino acid makeup yield indistinguishable spectra. The indistinguishable fragmentation patterns are contrary to the driving power of MS\textsuperscript{n} based sequencing methods that rest on each peptide sequence generating a unique mass spectrum upon fragmentation.

The potential impact of sequence scrambled product ions on automated/database peptide/protein identification (specifically the Mascot software) has been studied for a small set of model tryptic peptides by Saminathan et al. These investigations determined “scrambled” ions had apparently little effect on Mascot match scoring when compared to database searching without “scrambled” ions. However, it is important to distinguish the critically negative effect of sequence scrambling is likely most observable at MS\textsuperscript{3} and higher tandem stages. A potentially exhaustive analysis of a more extensive data set is likely required to assess the impact on database searching. Additionally, the direct effect scrambling has on the de novo sequencing of novel and lesser investigated peptides/proteins is more prevalent, as the de novo method relies
on direct interpretation of sequence ion m/z peaks – which consequently are indistinguishable from “scrambled” ion peaks for unknown sequence.

Scheme 1.10 Sequence scrambling
1.10 Enhancing de novo peptide sequencing

From a bioanalytical viewpoint, efforts to enhance MS\textsuperscript{n} de novo peptide sequencing should seek to:

1. Identify the N- and/or C-terminal amino acid
2. Reduce or eliminate the effects of sequence scrambling to provide unambiguous MS\textsuperscript{n} sequencing
3. Increase sequence coverage, sequence ions as a collective and/or the full series of b and/or y ions
4. Reduce the number of uninformative non-sequence ions for enhanced interpretability
5. Distinguish between isobaric residues
6. Advance the understanding of the dissociation chemistry involved

De novo techniques that succeed in addressing even some of the above points will afford more reliable peptide sequence assignments, as well as possible avenues to better de novo sequencing software. As CID-MS\textsuperscript{n} based techniques will likely continue to remain the most critical tools for peptide and protein identification in proteomics, maximizing its effectiveness and underlining chemistry involved is necessary for progressing the field. Though no one method may succeed at addressing each concern, complementary techniques must be sought to facilitate improved bioinformatics. It is the goal of the research presented in this document to address the central scientific problem of enhancing de novo peptide sequencing by investigating peptide derivatization and metal cationization techniques\textsuperscript{86-88} for elimination of sequence scrambling.
and improved sequence coverage. Explicitly the aims of this research to advance efforts to enhance de novo peptide sequencing methods were to:

- N-terminally modify the common research peptide leu-enkephalin ([YGGFL + H]^+) to make a protonated peptide-imine and observe the CID-MS^n spectra for sequence scrambling and sequence ion formation
- Correspondingly, N-terminally modify small protonated peptides to create protonated peptide-imines and evaluate each CID-MS^n spectrum for product ion formation characteristic to the modification
- Identify the observed novel product ion generated from CID of protonated N-terminal peptide-imines presumed to generate from the elimination of the N-terminal amino acid residue mass
- Investigate the affect the N-terminal modification structure has on product ion formation and product ion intensity of protonated peptide-imines
- Evaluate the CID-MS^n spectra of metal-cationized N-terminal peptide-imines to survey metal-cationization of the modified peptides as an additional technique to mitigate sequence scrambling and promote sequence ion generation

1.10.1 Metal-cationization

Cationization of peptides and proteins is the most commonly used ionization approach for CID-MS^n sequencing methods, mainly by protonation.\textsuperscript{89, 90} However, determining the amino acid sequence for protonated peptides is not particularly intuitive in practice, as the sequence ions that contain the N- and C-terminal amino acids are typically unknown before analysis. Thus,
additional sequence information can be invaluable as supplemental to the mass analysis of protonated peptides. In part to fulfill this need and address the desire to identify and understand intrinsic metal binding sites, alkali and transition metals have been used as alternative peptide cationization reagents.\textsuperscript{90-94}

In contrast to protonated peptides, the CID fragmentation chemistry of alkali metal-cationized peptides has been proposed to be predominately charge-remote, involving the binding of the metal ion to amide carbonyl atoms or the deprotonated C-terminal carboxylic acid.\textsuperscript{92, 95-98} Nevertheless, dissociation of metal-cationized peptides commonly produce the b\textsubscript{x}H, a\textsubscript{x}-H, and y\textsubscript{z}-H sequence ions, as well as a prominent b-ion derivative, [b\textsubscript{x} + OH + metal]\textsuperscript{+} (Scheme 1.11).\textsuperscript{87, 92, 94} Multistage MS\textsuperscript{n} experiments of [b\textsubscript{x} + OH + metal]\textsuperscript{+} have demonstrated complete C-side de novo sequencing capabilities as the C-terminal hydroxyl group is transferred to the adjacent amino acid in the sequence followed by the subsequent loss of the C-terminal amino acid residue mass.\textsuperscript{79, 92, 99} The resulting metal-cationized peptide, lacking the original C-terminal amino acid, can then be used to identify the next amino acid in the sequence by the same C-terminal mechanism. Similarly, the [b\textsubscript{n} + OH + metal]\textsuperscript{+} ion has been observed for silver-cationized peptides.\textsuperscript{95, 100} However, an understanding of both alkali- and silver-cationized peptide dissociation chemistry remain currently limited for widespread use or integration into sequencing software platforms.
1.10.2 Peptide derivatization

Peptide sequence coverage (the percent of peptide sequence covered by identified product/sequence ions) is arguably the most crucial aspect of successful peptide de novo sequencing. However, in practice, complete interpretation of protonated peptide MSn mass spectra is often hard to achieve commonly due to the absence of the b1-ion, which identifies the N-terminal amino acid. Considerable importance is conferred on accurate N-terminal amino
acid identification including but not limited to confirming predicted termini of individual biological or pharmaceutical proteins, and to study proteolysis in dynamic living systems.\textsuperscript{15, 102}

N-terminal tagging and derivatization strategies, such as acetylation, have demonstrated N-terminal amino acid identification capabilities.\textsuperscript{103-105} Additionally, several studies have demonstrated conclusively that acetylation at the N-terminus eliminates sequence scrambling\textsuperscript{29}, presumably because the nucleophilic amino group implicated in the cyclization reaction is functionalized. However, large scale use of specific N-terminal derivatization as a technique for peptide/protein sequencing would require a comprehensive understanding of the fragmentation chemistry the modification induces. The remainder of this work, therefore, seeks out to discuss the gas phase methods and observations of an investigation into the dissociation behavior (fragmentation pattern) caused from N-terminal amine to imine (Schiff base) peptide modification.
1.11 References


CHAPTER 2

TECHNIQUES: GAS PHASE ANALYSIS OF PEPTIDES AND PROTEINS

2.1 Mass Spectrometry

Mass spectrometry (MS) is currently the principal method of analysis for peptide and protein characterization and quantification superseding methods such as Edman degradation, N-terminal amino acid analysis, and ultracentrifugation. Ionization followed by mass analysis of simple to complex peptide/protein solutions provide a swift determination of molecular mass with high mass accuracy and sensitivity, asserting the dominance of mass spectrometric methods in the field. Though mass spectrometers may vary from instrument to instrument, be it manufactured by a major company or home-built/modified in the lab, the essential components that constitute these indispensable tools remain the same (Figure 2.1) as well as the general process:

(1) From the sample of interest, the ionization source generates ions having a mass \( m \) and total charge \( q = ze \), where \( z \) is the charge number and \( e \) the elementary charge constant (1.6021766208 x 10^{-19} \text{ C}). The ions obtain a kinetic energy \( KE \) (Equation 2.1) equal to the product of the ion charge and the electric potential \( V_s \), produced from the applied electric field (acceleration voltage).

\[
KE = \frac{mv^2}{2} = qV_s = zeV_s , \text{ where } v = \text{velocity}
\]
(2) As the ions emerge from the ionization source they are extracted, focused, and accelerated into an ion beam directed toward the mass analyzer(s) via several ion lenses positioned just after the source and along the ion path. In a two mass analyzer scheme, static or dynamic electric and magnetic fields separate ions in the beam according to their mass-to-charge ratio (m/z). These fields are combined or used independently, depending on the type of analyzer. In this way, the ions can be detected or selected for fragmentation. If fragmented, the emerging fragment ions are separated by the second mass analyzer and focused toward the detector.

(3) Detection of the ions and measurement of their abundance is achieved by converting the ions into electrical signals that are processed by a computer program. The resulting mass spectrum is generated by plotting the m/z values along the x-axis and ion abundance along the y-axis.

The figures of merit attributed to the type of mass analyzer (Table 2.1), along with characteristics such as the ionization source and method of fragmentation culminate into a variety of mass spectrometers ranging in experimental allure. The following sections will briefly discuss the components used in the research of this dissertation.
Figure 2.1 General Mass Spectrometer Schematic. The blue tags represent instrumentation/hardware, and the black tag represents software.

Typical Mass Analyzer Characteristics

<table>
<thead>
<tr>
<th>Mass Analyzer</th>
<th>Mass Resolution (FWHM)(a)</th>
<th>Mass accuracy (ppm)</th>
<th>Mass range</th>
<th>Dynamic range(b)</th>
<th>Scan rate(c)</th>
<th>Cost / Space Requirement</th>
<th>Tandem Mass Spectrometry Capability(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadrupole Mass Filters</td>
<td>100 – 1,000</td>
<td>100</td>
<td>4,000</td>
<td>(10^7)</td>
<td>Moderate</td>
<td>Low cost, small space requirement</td>
<td>MS to MS/MS (Q-q-Q)</td>
</tr>
<tr>
<td>Quadrupole Ion Trap (QIT)</td>
<td>1,000 – 10,000</td>
<td>50 – 100</td>
<td>4,000</td>
<td>(10^2 - 10^3)</td>
<td>Moderate</td>
<td>Low cost, small space requirement</td>
<td>MS(n)</td>
</tr>
<tr>
<td>Linear Ion Trap (LIT)</td>
<td>1,000 – 10,000</td>
<td>50 – 100</td>
<td>4,000</td>
<td>(10^3 - 10^4)</td>
<td>Fast</td>
<td>Low cost, small space requirement</td>
<td>MS(a)</td>
</tr>
<tr>
<td>Time-Of-Flight (TOF)</td>
<td>1,000 – 40,000</td>
<td>5 - 50</td>
<td>Greater than 100,000</td>
<td>(10^6)</td>
<td>Fast</td>
<td>Moderate cost, moderate space requirement</td>
<td>MS to MS/MS (TOF/TOF)</td>
</tr>
<tr>
<td>Fourier Transform Ion Cyclotron Resonance (FT-ICR)</td>
<td>10,000 – 1,000,000</td>
<td>1 - 5</td>
<td>Greater than 10,000</td>
<td>(10^2 - 10^4)</td>
<td>Slow</td>
<td>High cost and large space requirement</td>
<td>MS(a)</td>
</tr>
<tr>
<td>Orbitrap</td>
<td>10,000 – 150,000</td>
<td>2 - 5</td>
<td>6,000</td>
<td>(10^3 - 10^4)</td>
<td>Moderate - fast</td>
<td>Moderate cost, low space requirement</td>
<td>MS(a)</td>
</tr>
</tbody>
</table>

Table 2.1. Typical Mass Spectrometer Characteristics. \(a\) Full width at half maximum (FWHM), describes the mass resolving power describes as, \(M/\Delta M\), where \(M\) is the mass of the singly charged ions that make up a single peak, and \(\Delta M\) the width of the peak at \(50\%\) of its height; \(b\) linear dynamic range refers to the range over which the ion signal is directly proportional to the analyte concentration; \(c\) scan rate referred to the number of scans per second possible over the m/z range while satisfying the mass resolution; \(d\) MS/MS is an abbreviation for tandem mass spectrometry indicating only two stages of mass analysis, \(n\) corresponds to stages > 2.\(^{10}\)
2.2 Electrospray ionization

Electrospray ionization (ESI) is an ionization method first investigated by Malcolm Dole but advanced in the late 1980s when Fenn et al\textsuperscript{109} obtained multiply charged ions from proteins, subsequently allowing their molecular weight to be determined by instruments with low mass range limits (>2000 u). The process begins as the sample solution is introduced to the ion source through a metal needle connected either to a syringe pump or the exit of a liquid chromatography column. As the solution passes through the capillary tube at a flow rate typically between 1-15 μL/min, a strong electric field (on the order of 10\textsuperscript{6} V/m) is produced at the needle tip by applying a potential difference of 2-6 kV (Figure 2.2).\textsuperscript{112-116} Consequently, charge accumulation is induced on the surface of the analyte resulting in highly charged droplets as the solution is being sprayed.\textsuperscript{107, 111, 117, 118} The spray transitions into a fine mist assisted by a sheath gas (generally nitrogen) injected inner coaxially at a low flow rate.\textsuperscript{107, 111, 117} The auxiliary gas, an outer coaxial nebulizing gas, aids in evaporating the solution.\textsuperscript{107, 111, 117, 118}

As the solvent contained in the droplets evaporate, the shrinkage causes the charge per unit volume to increase.\textsuperscript{107, 118} The strong applied electric field and the accumulation of charge elongates the shape of the droplets to that of cone known as the Taylor cone.\textsuperscript{118} Smaller droplets emerge (Coulombic explosion) as the repulsive forces between the accumulated charges on the droplet surface become equal to or higher than the droplet surface tension, also known as the Rayleigh limit.\textsuperscript{107, 117} Decomposition below the Rayleigh limit may also occur for droplets formed at the ESI needle tip mechanically deformed by the applied electric field.\textsuperscript{107, 118} By either means, these small highly charged droplets continue to lose solvent by evaporation as they pass through a curtain of heated inert gas (nitrogen), or through a heated capillary producing further droplet generations until the electric field on their surface becomes large enough for desorption
of ions from the surface.\textsuperscript{118} As aforementioned, ESI of large biological macromolecules such as proteins tend to form multiply charged ions through protonation \([\text{M} + z\text{H}]^{z+}\), typically one charge per thousand Daltons unless very few basic amino acids.\textsuperscript{11} Small molecules such as peptides less than a thousand Daltons tend to produce monocharged ions \([\text{M} + \text{H}]^+\) which are focused toward the mass analyzer by skimmers and ion lenses.\textsuperscript{11}

**Electrospray ionization (ESI)**

Figure 2.2 Electrospray ionization.\textsuperscript{107, 112-116}
2.3 Mass analyzer

2.3.1 Introduction

Mass spectrometers are typically distinguished by the mass analyzer(s) it contains – the instrument that measures the mass-to-charge ratio of gaseous ions generating the mass spectrum. Regarding MS-based proteomics, mass spectrometry should be considered not just as a method for discovery but rather a vast field consisting of processes and instruments that differ by, and best lend themselves to a specific question. Regarding proteomics, some essential factors to consider before MS studies include the instrument sensitivity, fragmentation method (including multi-stage capabilities), and the overall analysis strategy.108, 119

Until the recent invention of the Orbitrap,120 the affordability and lab space requirement of spectrometers capable of favorable, high-resolution mass spectrometry (HRMS, resolution = 0.01 – 0.001 Da) was out of reach for many laboratories.119, 121, 122 Thus, low-resolution mass spectrometry (LRMS, resolution = 1 – 0.1 Da) became the dominating mass data collection method for early proteomics and the peptide/protein databases.119, 121, 122 Consequently, for its general user-friendliness, relative affordability, “desk-space” size, high sensitivity, rapid speed of analysis, and multistage fragmentation capabilities108, 119 the mass spectral data collected in this research was primarily obtained using an LRMS instrument, the linear ion trap (LIT) mass spectrometer, more specifically the Thermo Fischer Scientific LTQ-XL. The balance of this section will discuss the operation of the LIT.

2.3.2 Linear ion trap

Generally, mass analyzers utilize static and dynamic electromagnetic fields to separate gaseous ions based on the m/z values.123 The fundamental difference between mass analyzers
rests on how these fields are used to achieve separation. In the case of ion trap mass analyzers, separation is based on the same principles employed in quadrupole analyzers – where ion trajectory stability in oscillating electromagnetic fields is manipulated by applying electric currents at varying amplitudes. The two types of current are direct current (DC) consisting of electric charge flowing in only one direction and alternating current (AC) which allows charge to change direction periodically. Classically, ion traps are composed of four rod-shaped electrodes to which DC and AC are applied.

Different conformations of the basic structure give rise to different types of ion traps with their own set of capabilities. Three-dimensional traps - also known as Paul ion traps (Figure 2.3a) - are configured as if each of the four rods of a quadrupole ion trap is “bent in on itself,” having two ellipsoid end cap electrodes (upper and lower) and a central circular electrode. Within 3D ion traps, the ions experience electromagnetic forces in the radial (x and y) and the axial dimensions (z), overlapping direct (DC) and alternative (AC) potentials to tightly confine ions of all masses. Restricted in this 3D trajectory, ions of a given mass are expelled from the trap to the detector by applying resonant frequency (radio range, 3kHz – 300 GHz) along the z-axis, resulting in a mass spectrum.

On the other hand, two-dimensional ion traps are constructed with four hyperbolic rods positioned adjacent and cut into three sections, thus allowing for greater ion capacity/storage spread out across the z-axis when compared to 3D traps (Figure 2.3b). Electromagnetic force in 2D ion traps is applied only in the x and y dimensions by AC potentials at a radiofrequency voltage to each rod pair with equal amplitude but opposite sign (known as the “main RF”, up to ± 5kV, 1 MHz) to confine ions radially, and by DC potentials applied to each sectioned portion to confine ions axially. It is important to note that the “LIT”
nomenclature is used in literature at times to denote the linear entrance and exit path of ions entering the trap as with 3D traps, as well as to denote an elongated trapping space, and/or lack of an axial (z) force as with 2D traps. For clarity, the 2D distinction will be applied when using the LIT abbreviation for the remainder of the discussion of the ion trap functionality. Additionally, the Thermo Fischer LTQ-XL instrument used in this work is a 2D linear ion trap.
Ion Traps

Figure 2.3 Ion traps. a) *3D ion trap*. Direct (DC) and alternating (AC) potentials are applied in the radial (x and y) and the axial (z) dimensions to tightly confine ions.\textsuperscript{107, 108} b) *2D linear ion trap (LIT) with radial ejection*. Only AC potentials are applied in the radial dimensions (Main RF), while DC potentials are applied in the axial dimension.\textsuperscript{107, 123, 124}
While confined inside the LIT ions move in a “spiral type” motion about the trap dependent on the amount of force applied by the main RF and therefore proportional to its amplitude and the mass of the ion, where lighter ions move about more than heavier ones. Ions are ejected from the trap to the detector radially through parallel slits cut into a pair of the quadrupolar rods opposite each other known as the “exit rods”. This action is achieved by applying an additional AC to each exit rod equal in amplitude but opposite in phase. As the frequency of the trapped ion motion induced by the main RF and dependent upon the ion m/z begins to resonate with the frequency of the exit rod AC the trapped ion will begin to eject from the trap in a process known as resonance ejection. Thus, the mass spectrum is obtained by selecting the “scan” function which corresponds to steadily ramping up the main RF so that the ions in the trap, according to their m/z, sequentially enter resonance with the exit rods AC and exit to the detector.

### 2.4 Tandem mass spectrometry

The mass analysis discussed above is descriptive of one mass spectrometric event or stage whose resulting mass spectrum is often denoted as MS or MS\(^1\). Arguably, however, the most powerful function of the LIT is its ability to perform multiple stages of mass analysis known as tandem mass spectrometry. Tandem MS experiments consist of isolating a precursor ion \((m_p^+\)) within the ion trap, that then undergoes dissociation to yield product ions \((m_f^+)\) and neutral fragments: \[ m_p^+ \rightarrow m_f^+ + m_{\text{neutral}} \]
Regarding peptide sequencing, product ions generated from peptide precursor ions are used to piece together the peptides amino acid sequence. Abbreviations used to denote tandem MS experiments include MS/MS, which fundamentally indicates at least two stages of mass analysis. However, the abbreviation MS\(^n\) more accurately symbolizes multiple stages of mass analysis where \(n\) (i.e., 1, 2, 3…) is the number of stages or generations of ions analyzed.

An important performance-based MS\(^n\) distinction to note is between tandem-in-space (Figure 2.4a) and tandem-in-time mass analysis (Figure 2.4b). Tandem-in-space analysis corresponds to the restriction of mass analysis stages to the number of successive mass analyzers the instrument contains.\(^{107}\) Tandem-in-time (Figure 2.4b) analysis corresponds to the processes of consecutive mass analyses occurring in the same physical space allowing for many stages of analysis.\(^{107}\) Specifically, tandem-in-time MS\(^n\) corresponds to the tandem function of the LIT. The process begins with the completion of the first MS stage, which from the pool of m/z values determined an ion of interest (precursor ion) is isolated within the ion trap starting the MS\(^n\) cycle.\(^{107}\) This cycle can be broken down into four operations:

(1) Precursor ion isolation

(2) Fragmentation

(3) Mass Analysis of product ions (described above)

(4) Selection and isolation of a new precursor ion (i.e., repeating step 1)
Figure 2.4 Tandem Mass Spectrometry. a) Tandem-in-space vs. tandem-in-time. b) The general operation of tandem-in-time mass spectrometry using a linear ion trap.
Isolation is achieved by applying a broadband waveform AC voltage to the exit rods where all the frequencies resonant with trapped ions motion are present while excluding the frequency of the ion to be isolated. The ion isolation waveform voltage coupled with the main RF voltage ejects all the ions except the ion for isolation from the trap. Once isolated, the trapped ion can undergo fragmentation.

2.4.1 Fragmentation by collision-induced dissociation

During both MS and MS" experiments, helium gas is bled into the mass analyzer cavity with two functions, as a damping/buffer gas and a collisional gas. As a buffer gas, helium acts to enhance the instrument sensitivity and mass resolution by reducing the kinetic energy of the ions entering the mass analyzer, allowing the main RF to trap them efficiently. While in the mass analyzer cavity, collisions with the helium gas slow the kinetic energy of the ions. The result is a steady-state amplitude of the ion’s oscillation and thus a focusing of the ion ejection time for more resolved detection. As a collisional gas, the energetic collisions between the precursor ion and the helium gas result in the fragmentation of the precursor ion into one or more product ions, a process known as collision-induced dissociation (CID).

CID is the most common form of fragmentation in linear ion traps and for fragmenting peptide ions. During this process, a resonance excitation AC voltage is applied to the exit rods causing the trapped precursor ion to gain more kinetic energy as it oscillates about the collision cell (i.e., the ion trap in LIT mass spectrometers). The ions are driven into the helium gas, causing many energetic collisions as it continues to oscillate around the collision cell (Figure 2.5). The collision between the precursor ion and the helium target is a very fast step (10^{-14} – 10^{-16}s) coinciding with bond vibrational periods, during which the translational kinetic energy
is converted to the internal vibrational energy of the ion.\textsuperscript{107} This energy is then distributed into the molecules vibrational modes (non-linear, $3N - 6$, where $N$ is the number of atoms).\textsuperscript{107, 125} Many of these inelastic collisions (tens – thousands) cause the ions to become vibrationally excited/activated until the energy threshold for dissociation (the activation energy for a particular product ion to be formed, $E_o$) is surpassed leading to fragmentation along the peptide backbone.\textsuperscript{68, 108, 124, 125}

**Collision Induced Dissociation (CID)**

![Collision Induced Dissociation (CID)](image)

**Figure 2.5 Collision-induced dissociation (CID).** During the first step (1) an excitation voltage is applied to the ion trap exit rods causing the trapped, oscillating ions to gain kinetic energy; the second step (2) involves many energetic, inelastic collisions between the ion and the buffer gas (i.e., helium) resulting in vibrational activation of the precursor ions; step (3) involves the fragmentation of the activated precursor ions, beyond the dissociation threshold, into smaller fragment ions.
The extent of fragmentation and product ion formation of protonated peptide ions under these conditions are dependent on a few factors including, \(^{107}\)

1. **Activation time** – the time precursor ions spend experiencing collisions
2. **Excitation AC voltage**
3. **Precursor ion dissociation probability** as governed by the theory of quasi equilibrium (QET)

At a lengthened activation time, more collisions occur between the precursor ion and the collision gas, resulting in the transfer of more energy into the ions’ vibrational modes. \(^{125}\) The transferable energy at each collision also increases by raising the user-defined excitation AC voltage, i.e., the normalized collision energy (NCE). \(^{125}\) At higher voltages, the kinetic energy of the ions oscillating in the trap increases causing each impact with the collision gas to be more energetic. \(^{125}\) Either by changing the activation time or changing the NCE, the amount of energy transferred into an ion can be controlled and thereby the fragmentation rate; \(^{126}\) presumably, the dissociation threshold energy (stability), and the energetic and kinetic dependence on fragmentation or formation of an ion can also be deduced. Experiments monitoring precursor ion dissociation or product ion formation as a function of increasing NCE or activation time are known as “energy-resolved” and “time-resolved” respectively.

QET is a widely accepted approach that describes the rate of unimolecular decompositions of ions with excess internal energy (\(E_{\text{ex}}\)) under mass spectrometric conditions (i.e., under high vacuum), originally adapted to describe electron impact (EI) ionization. \(^{107, 127-129}\) This theory strongly correlates ion internal energy distribution, like that resulting from CID, to the rate constants of ion dissociation in the gas phase. \(^{107}\) Several postulates govern QET: \(^{107, 128-130}\)
(1) All the degrees of freedom of the gaseous molecule participate in the energy distribution with the same probability.

(2) The rate of dissociation is small compared to the rate of the randomized excitation energy redistribution among all the ions’ internal modes.

(3) The probabilities that the randomized energy distribution becomes concentrated explicitly in ways that give rise to several activated complex configurations (for non-linear ions) determine the rate of dissociation.

(4) The observed dissociation products result from a series of competitive and consecutive irreversible reaction pathways governed by the $E_0$ required to form its products.

By these conditions, the fragmentation pathways of an isolated ion are dependent upon the structure of the ion and the amount of energy deposited during activation and not the method of ion activation itself.\textsuperscript{107, 128-130} Consequently, this allows QET not only to be applied to mass spectra produce from EI but also CID. An illustration of QET is depicted by the Wahrhaftig diagram (Figure 2.6), where the x-axis represents the internal energy of the theoretical ion $\text{ABCD}^+$.\textsuperscript{130} The upper diagram y-axis indicates the probability of a product ion being formed while the lower diagram y-axis indicates the logarithmic dissociation rate constant.
Figure 2.6 Warhaftig Diagram. **Upper diagram.** The probability a specific hypothetical ion is generated at a given amount of deposited internal energy is plotted on the surface of the bell curve. **Lower Diagram.** The logarithmic rate constant for product ion formation is plotted as a function of internal energy. $E_s$ (stable ion), $E_m$ (metastable ion), and $E_o$ correspond to the energy threshold (activation energy) required for an ion/pathway to persist. The red (loose ion complex) and blue (tight ion complex) lines depict the effect structure has on the rate constants as well as represent the thermodynamics of each pathway, where $AD^+$ formation is energetically favored, and $AB^+$ kinetically favored.\(^{130}\)
2.4.2 Fragmentation by infrared multiple photon dissociation

Though an indispensable tool for identifying peptide/protein sequence, the m/z data generated from tandem mass spectrometry do not explicitly offer any product ion structural or conformational information. Determining product ion structure, however, not only validates product ion assignments but also aids in deducing the fragmentation pathway that generated the ion for a better understanding of the fragmentation chemistry. To this end, infrared multiple photon dissociation (IRMPD) spectroscopy has proven to be a valued technique that provides direct structural insight for gas phase peptide ions.  

Activation by IRMPD occurs when trapped gaseous ions (by an ion trap or ICR cell) are irradiated with photons generated from a tunable IR laser. Absorption of a photon occurs when the laser frequency producing the photons matches a vibrational mode of the precursor ion, which is followed by rapid intramolecular vibrational distribution (IVR) over all the vibrational modes of the ion. By IVR the energy of each photon dissipates before the ion absorbs the next photon. As multiple photons are absorbed the internal energy of the ion increases toward a dissociation threshold ($E_o$) to excess, leading to product ions in a “slow heating” process like CID that can be described by QET, and depicted by the following mechanism:

\[
\text{Equation 2.3} \quad m_p^+ \xrightarrow{nhv} m_p^{+ \neq} \xrightarrow{k_{diss}} m_f^+ + m_{neutral}
\]

where $m_p^{+ \neq}$ is the activated precursor ion, $hv$ the photon energy, $n$ the number of absorbed photons, and $k_{diss}$ the rate constant for photodissociation. This type of resonant absorption of photons produces a frequency-dependent fragmentation signal. IR-active spectra are
generated by plotting frequency on the x-axis and product ion intensities in terms of fragmentation yield (Equation 2.4) on the y-axis:

\[
\text{Equation 2.4} \quad \frac{\sum \text{fragment ion intensities}}{\sum \text{fragment ion intensities} + \text{parent ion intensity}}
\]

The resulting IR absorption bands can be used to determine the structure of the precursor ion or compared to the experimental or theoretical IR spectra of ions with a known structure for identification.

### 2.4.3 Theoretical vibrational IR spectra

Theoretical vibrational spectra can be obtained for proposed peptide ion sequence by density functional theory (DFT) calculations. Introduced in 1964 by Kohn and Hohenberg, DFT demonstrates that ground state electronic energy can be determined by electron density.\textsuperscript{137} Generation of theoretical IR spectra is accomplished by performing computational geometry optimization and frequency calculations. The process of geometry optimization consists of computing local and global minimizers of the molecules potential energy surface (PES) - the molecules potential energy as a function of its atomic positions in space (Figure 2.7).\textsuperscript{137-139} The atoms are moved little by little, and at each perturbation the energy is calculated.\textsuperscript{140} The lowest energy is obtained at the optimized geometry, considered a stationary point on the PES, where the net-inter-atomic force on each atom is acceptably close to or equal to zero minimizing the strain on the given system.\textsuperscript{137,140} The low-energy, optimized geometry often corresponds to the intrinsic structure of the molecule from which geometry minima and vibrational frequencies are
calculated and used to produce a theoretical vibrational spectrum.\textsuperscript{140} The success of computational vibrational spectra has been shown for small polypeptides by Gaigot et al. with an excellent agreement to spectra obtained experimentally by IRMPD.\textsuperscript{141-143}

**Potential Energy Surface**

---

**Figure 2.7 Potential energy surface.** DFT geometry optimization protocols sample the surface for energy minima, at which the most stable geometries exist. A saddle point is a minimum on one path along the surface and a maximum along another.\textsuperscript{137-139}
2.5 References


CHAPTER 3

METHODS

3.1 N-terminal amine to imine derivatization

3.1.1 Sample preparation

The peptides GG, GGG, GGA, LGG, FGG, YGGFL, and the aldehydes benzaldehyde, m-methoxybenzaldehyde, and p-methoxybenzaldehyde were purchased from Sigma Aldrich (St. Louis, MO, USA). The peptide VGG was purchased from Bachem (Bubendorf, Switzerland). The peptides YGG, SGG, and TGG were purchased from Chem-Impex International Inc. (Wood Dale, Il, USA). The peptide AGG was purchased from TCI-America (Portland, OR, USA) and the aldehydes salicylaldehyde, o-methoxybenzaldehyde, m-hydroxybenzaldehyde, and p-hydroxybenzaldehyde were purchased from Alpha Aesar (Ward Hill, Ma, USA). All peptides were used as received with no further purification. Peptide imines were produced by reacting the selected peptide with an aldehyde (1:1 mole ratio) in methanol for 2-6 h at 37 °C. The reaction mixture was diluted with a 50:50 solution of methanol/water. The expected mass shift observed in the mass spectra (corresponding to the modifier used) confirmed the successful modification of the peptides by Schiff base (condensation) reaction (Scheme 3.1). The final peptide concentrations for ESI-MS studies were ~10^{-4} M. Protonated peptide-imines, [M+H]^+ ions, were generated by acidifying peptide solutions with 10 µL of 1% acetic acid in water and vaporized to the gas phase using ESI.
Scheme 3.1 Imine-derivatization (Schiff base)

3.1.2 CID experiments

Multiple-stage CID experiments were conducted on a ThermoScientific LTQ-XL linear ion trap mass spectrometer (ThermoScientific, San Jose, Ca, USA). The acidified peptide-imine solutions were infused into the ESI-MS instrument using a syringe pump at a flow rate of 5 µL/min. The atmospheric pressure ionization stack settings for the LCQ (e.g., lens voltages, quadrupole, and octupole voltage offsets) were optimized in the positive ion mode to produce the maximum [M+H]^+ transmission to the ion trap mass analyzer using the auto-tune function within the LTQ Tune package. Following the instrument tune, the spray needle voltage was maintained at +5kV, the N_s sheath gas flow at ten units (arbitrary for the ThermoScientific
instruments) and the capillary temperature held at 200 °C. Helium was used as the bath gas for enhanced trapping efficiency and as the ion trap collision gas.

For the initial CID (MS/MS) experiments the protonated peptide-imine ions, [M+H]⁺ were isolated using an isolation width of 1.0-1.5 mass-to-charge units that reflected the best compromise between high [M+H]⁺ abundance and single isotopic peak isolation. The (mass) normalized collision energy (NCE, defined by ThermoScientific) was set between 20% and 25%, which corresponds roughly to 0.80-0.99 V with the current instrument calibration used in this study. The activation Q (defines the frequency of the applied radio frequency potential) was set to 0.30, and the activation time set to 30 ms. The mass spectra obtained represented an accumulation and averaging of at least 30 isolation, dissociation, and detection steps. Following CID stages (MSⁿ) were performed through the isolation of [M+H]⁺ product ions, and subsequent NCE applied, inducing fragmentation.

3.1.3 CID energy/time-resolved experiments

For the energy/time-resolved experiments, [M+H]⁺ ions were isolated using an isolation width 1.0-1.5 m/z units. CID (MS/MS) normalized collision energies ranging from zero to 50 (0-2.5 V) were applied sequentially to the fragmentation of [M+H]⁺ to 0.0-10% ion abundance. Over the NCE range the activation time (corresponding to the time at each data acquisition step [M+H]⁺ is held in the ion trap during CID) was set to 10, 100, and 1000 ms representing 3 separate time-resolved stages, conducting data acquisition at each NCE (at least 30 isolation, dissociation, and detection steps). The activation Q was set to 0.30. Each activation time trial was repeated in triplicate, and the average obtained.
3.2 Characterization of N-terminal amino acid elimination from protonated peptide-imines

3.2.1 Sample preparation

Peptides GG, GGG, AGG, AGGG, VAAF, YGGFL, and the aldehyde benzaldehyde were purchased from Sigma Aldrich (St. Louis, MO. The peptide PGG was purchased from Bachem (Bubendorf, Switzerland). The aldehyde salicylaldehyde and o-methoxybenzaldehyde were purchased from Alfa Aesar (Ward Hill, Ma, USA). The peptides were used as received without further purification. Peptide imines were produced by reacting peptides with salicylaldehyde (1:1 mole ratio) in methanol for 2-6 h at 37 °C. The reaction mixture was then diluted with a 50:50 solution of methanol/water and then acidified with 10 µL of 1% acetic acid in water. Confirmation of peptide modification was achieved by observation of the corresponding derivatization mass shift, the expected change due to the Schiff base reaction. The final peptide concentration for ESI-MS and IRMPD studies were ~ 10⁻⁴ M.

3.2.2 CID experiments

(See section 3.1.2)

3.2.3 ESI FT-ICR mass spectrometry

Previously established methods were used for the generation of ions and the subsequent collection of IRMPD spectra,133 a brief overview has been provided. Ion trapping and mass analysis were carried out with a laboratory-built Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer,133 equipped with a Micromass Z-spray ESI source positioned inside
the magnetic field near the FT-ICR analyzer cell. The instrument was controlled externally by a modular ICR data acquisition system (MIDAS) developed at the National High Magnetic Field Laboratory (NHMFL) Instrument. Operating parameters were optimized to maximize the formation of [M+H]^+ ions, or b_2 ions generated by the in-source CID, and transfer of the species to the ICR cell. Dry nitrogen (~80 °C) was used to assist in the desolvation process.

### 3.2.4 Infrared multiple photon dissociation (IRMPD)

The experiment apparatus used has been described in detail elsewhere, and only a brief outline is given here. These experiments were carried out at the Free Electron Laser for Infrared eXperiments (FELIX) facility at Radboud University Nijmegen, Netherlands. The IRMPD spectroscopy instrument utilized consists of a Paul-type quadrupole ion trap combined with a time-of-flight (TOF) mass spectrometer. To perform IRMPD the [M+H]^+ was isolated using an ion trap then irradiated with a free electron laser (FEL) infrared pulse. The wavelength of the FEL is continuously tunable between 2000 – 40 cm\(^{-1}\) (5 – 250 µm) and the bandwidth can be varied between 0.4% and 7% of the central wavelength. To cover this wavelength range, FELIX utilizes two lasers, FEL-1, and FEL-2 that work at different electron beam energies, 15 – 25 MeV and 25 – 45 MeV respectively, covering the ranges of 25 – 250 µm (FEL-1) and 5 – 30 µm (FEL-2).

Additionally, the wavelength range between 2000-3500 cm\(^{-1}\) is accessible by operating FELIX on the third harmonic, accomplished by replacing one of the gold-coated cavity mirrors with a dielectrically coated mirror. The fragment ions generated were subsequently pulse extracted to a TOF mass spectrometer. Note, fragment ions were observed whenever the tunable
infrared laser imparted enough energy to induce resonance enhanced multiple photon dissociation. The IRMPD spectrum was generated by obtaining the series of [M+H]^+ mass spectra collected as a function of laser wavelength and then plotting fragmentation yield as a function of wavelength. Additionally, IR spectra were obtained by dividing the total fragment yield by the total ion signal as the wavelength of FELIX is scanned. Note that the IRMPD process induces a vibrational-mode-dependent redshift, which was found to be typically 1.5% for a variety of polyaromatic ions.

### 3.2.5 DFT geometry optimization and frequency calculations

Geometry optimizations for [oHB=AGG + H]^+ and the possible structures resulting from the elimination of 71 u from [oHB=AGG + H]^+ were performed using the Gaussian 03 programs. These calculations were completed with the 3-21G* basis set and using the commonly employed hybrid B3LYP functional. The identified minima were then used to perform relaxed scans by rotating the dihedral angles along the peptide backbone in three degree steps through to 360 degrees. Geometry optimization and frequency calculations for the minima identified through these scans were then obtained using the B3LYP functional and the 6-311+G(d,p) basis set, commonly used for peptide structures in literature. The computed vibrational frequencies were scaled by a factor of 0.98, chosen empirically to generate a correlation between the theoretical and experimental IRMPD derived IR spectra. Theoretical IR spectra were generated by plotting the computed intensities as a function of the computed and scaled vibrational frequencies.
3.3 Metal-cationized peptide-imines

3.3.1 Sample preparation

The peptides GGG, YGGFL, and the aldehyde d_{1}-benzaldehyde were purchased from Sigma Aldrich (St. Louis, MO, USA). The aldehydes benzaldehyde and o-methoxybenzaldehyde were purchased from Alfa Aesar (Ward Hill, MA, USA). The peptide VAAF was purchased from Bachem (Bubendorf, Switzerland). The peptide AGG was purchased from TCI-America (Portland, OR, USA). All peptides were used as delivered without further purification. Peptide imines were generated by reacting peptides with an aldehyde (1:1 mole ratio) in methanol for 2-6 h at 37 °C. The reaction mixture was diluted with 50:50 methanol/water. The expected mass shift resulting from the condensation reaction was used to confirm the successful modification of the respective peptides. The final peptide concentrations for ESI-MS studies were ~10^{-4} M.

Generation of protonated [M+H]^+ ions was promoted by acidifying the modified peptide solutions with 10 µL of 1% acetic acid in water. Similarly, metal-cationization ([M+Ag]^+) was promoted by mixing unacidified peptide solutions with a 1 x 10^{-4} M solution of AgNO_3 in water at a 50:50 ratio. [M+H]^+ and [M+Ag]^+ ions were vaporized into the gas phase by infusing the solution to the ESI-MS instrument using the incorporated syringe pump.

3.1.2 CID experiments

Ion trap multiple-stage CID experiments were conducted on a ThermoScientific LTQXL linear ion trap mass spectrometer (Thermo Scientific, San Jose, CA, USA). The atmospheric pressure ionization stack settings for the LCQ (e.g., lens voltages, quadrupole, and octupole voltage offsets) were optimized for maximum [M+Ag]^+ transmission to the ion trap mass.
analyzer using the auto-tune function within the LTQ Tune program. Following the instrument
tune, the spray needle voltage was maintained at +5 kV, the N\textsubscript{2} sheath gas flow at ten units, and
the capillary temperature at 200 °C. Helium was used as the buffer gas to improve trapping
efficiency and as the collision gas for CID experiments. The CID studies (MS/MS and MS\textsuperscript{n}) for
\([\text{M}+\text{Ag}]^+\) were performed as described in section 3.1.2. To reduce spectral overcrowding, only
the \(^{107}\text{Ag}\) isotope was isolated.
3.4 References


CHAPTER 4

FRAGMENTATION PATTERN OF PROTONATED, N-TERMINAL PEPTIDE–IMINES

4.1 Introduction

Under low-energy CID conditions classical b ions have been proven to isomerize into macrocyclic b-ions through an attack by the N-terminal amine on the C-terminal oxazolone ring. Dissociation of the macrocyclic b ion leads to ring opening at various amide bonds causing the peptides’ primary sequence to be “scrambled,” rendering de novo sequencing efforts unsuccessful. N-terminal modifications such as acetylation have been shown however to reduce or eliminate the pathway that leads to b ion macrocyclization by functionalizing the N-terminal amine implicated in the cyclization pathway. As an alternative, the CID-MS$^2$ spectra of N-terminal amine to imine derivatized peptides by condensation reaction with salicylaldehyde (Sal) (Scheme 3.1) were evaluated for mitigating effects on sequence scrambling and to enhance MS$^n$ approaches to direct (de novo) peptide identification. N-terminal amine to imine (Schiff base) peptide derivatization was hypothesized to eliminate sequence scrambled product ions by the functionalization of the N-terminus much like in the way acetylation was shown to do. It is important to note, though the Sal abbreviation for salicylaldehyde appears in published work, to better distinguish the hydroxy substituent position on the benzyl ring moiety, the “oHB” abbreviation has been used in this present document, derived from the interchangeable o-Hydroxybenzaldehyde (where o = ortho) nomenclature. Additionally, the “=”
symbol, as in \([oHB=\text{Peptide} + H]^+\), has been used to indicate the imine transformation at the N-terminus.

4.2 Results and Discussion

4.2.1 Fragmentation pattern of imine-derivatized leucine-enkephalin

Protonated leucine-enkephalin (YGGFL) has been widely used as the model peptide in fundamental studies of peptide fragmentation.\textsuperscript{29,31,34,79,153} Under low energy CID, the dominant peaks generated for \([\text{YGGFL} + H]^+\) is the \(b_4\) and \(a_4\) ions at 100% and 75% relative intensity respectively (Figure 4.1a). Product ions \(b_3\), \(b_2\), \(y_3\), and \(y_2\) were also observed along with peaks consistent with the neutral loss of water to form the \(b_5\) ion at \(m/z\) 538, followed by CO loss to form the \(a_5\) ion at \(m/z\) 510. Rearrangement reactions (i.e., sequence scrambling reactions) for protonated YGGFL, leading to the apparent elimination of internal amino acid residues, have also been studied and reported by Vachet and coworkers.\textsuperscript{31,34} The product ions marked by an asterisk at \(m/z\) 177, 205, 262, and 380 are resultant of these rearrangement reactions, and therefore contain no sequence-specific information.

Considering the negative influence of sequence scrambled product ions is likely highest at \(\text{MS}^3\) and higher tandem stages, the tandem MS spectrum generated from CID of the protonated YGGFL \(b_4\) (\(m/z\) 425) ion and resulting \(a_4\) (\(m/z\) 397) ion are shown in Figure 4.1b and Figure 4.1c, respectively. CID of the \(b_4\) ion at \(m/z\) 425 (\(556\text{ MS}^2\rightarrow 425\text{ MS}^3\rightarrow\) mass spectrum) produced two sequence ions, the \(a_4\) and \(b_3\) at 100% and less than 2% relative intensity, respectively (Figure 4.1b). The “scrambled” product ion peak at \(m/z\) 262, suggesting a loss of 163 u (Tyr residue mass), was also observed. CID of the \(\text{MS}^3\) generated \(a_4\) ion at \(m/z\) 397 (\(m/z\) 556\text{ MS}^2\rightarrow 425\text{ MS}^3\rightarrow\)}
MS$^4$→ mass spectrum) produced the b$_3$ ion at ≈ 25% relative intensity, with the a$_3$ and b$_2$ ions observed at a relative intensity less than 2% (Figure 4.1c). The most abundant peaks were observed at m/z values 380 and 323. Peak m/z 380 appeared to be generated from the neutral loss of NH$_3$ (17 u) followed by the possible internal elimination of a glycine residue (57 u) to create an apparent “scrambled’ product ion at m/z 323. These apparent “scrambled” peaks were in addition to the m/z 177, 205, and 265 peaks observed at the MS$^2$ stage of protonated YGGFL, saturating the a$_4$ fragmentation pattern with non-sequence-specific product ions and demonstrating the sequence scrambling effect continues into later stages of MS$^n$. 
Figure 4.1 CID spectrum of a) protonated YGGFL (MS²), b) b₄ product ion (MS³), c) and a₄ product ion (MS⁴).
Compared to the underivatized version, CID of protonated imine-derivatized YGGFL ([oHB=YGGFL + H]⁺) generated the full series b type ions (Figure 4.2a). The dominant peak observed was the b₄ ion, with the b₁ and a₁ ions observed at a relative intensity less than 2%. The product ion peaks indicative of the b₅ (m/z 642) and a₅ (m/z 614) ions were also observed. Interestingly, a product ion peak at m/z 547 was observed corresponding to the formation of a [b₄ + H₂O]⁺ ion, indicative of a protonated b₄ ion with a classical carboxy terminus. These species of ions are generally observed for protonated peptides that contain one or more arginine residues, or for metal-cationized (Li⁺, Na⁺, or Ag⁺) peptides forming the analogous [bₓ + OH + Metal]⁺ species. Presumably, imine functional group competes favorably for the protonating hydrogen positioning the charge at N-terminus and allowing for a charge remote pathway similar to Scheme 1.11 to persist at the C-terminus.
Figure 4.2 CID spectrum of a) protonated oHB=YGGFL (MS²), b) b₄ product ion (MS³), c) and a₄ product ion (MS⁴)
CID of the $b_4$ ion generated from [oHB=YGGFL + H]$^+$ (m/z 660 $\text{MS}^2 \rightarrow 529$ $\text{MS}^3 \rightarrow$ mass spectrum) produced a fragmentation pattern containing only the $a_4$, $b_3$, and $b_2$ ions (Figure 4.2b). Subsequent CID of the $a_4$ ion generated from CID of the $b_4$ ion (660 $\text{MS}^2 \rightarrow 529$ $\text{MS}^3 \rightarrow 501$ $\text{MS}^4 \rightarrow$ mass spectrum) produced the full remaining $b_x/a_x$ ion series (Figure 4.2c). The $b_3$ ion $\text{MS}^4$ was observed as the dominant product ion, followed by the $b_2$ ion at $\approx 10\%$ relative intensity, and the $a_3$, $a_2$, $b_1$, and $a_1$ ions at relative intensities less the $5\%$.

Expectedly, the internal residue losses as a result of peptide rearrangements (i.e., scrambling) observed during multistage CID of protonated underivatized YGGFL were not observed for [oHB=YGGFL + H]$^+$. Unexpectedly, the fragmentation pattern of imine-derivatized YGGFL contained no $y_z$ ions. Presumably, the N-terminal imine nitrogen conserved along the $b_x$-$y_z$ pathway (Scheme 1.1) favorably retains the ionizing proton following the cleavage step, promoting $b$ ions in the mass spectrum. Interestingly, CID of [oHB=YGGFL + H]$^+$ also generated a more extensive series of N-terminal $b$ and $a$ ions at each tandem stage, including the $b_1$ ion. Assumedly, the secondary fragmentation $b_x$-$b_{x-1}$ and $b_x$-$a_x$-$b_{x-1}$ pathways (Scheme 1.5 and Scheme 1.6) become more competitive as a result of the apparent elimination of the macrocyclic pathways that lead to peptide rearrangements.

More interestingly, CID of protonated oHB=YGGFL (Figure 4.2a) generated a small peak at m/z 497 indicative of losing 163 u, the residue mass of tyrosine (Y). Additionally, at the MS$^4$ stage, CID of the $a_4$ ion created a peak at m/z 338 also suggesting the loss of 163 u. Assumedly, this loss implicates an internal elimination that expels the N-terminal amino acid residue, while unexpectedly conserving the imine modification to the rest of the primary sequence.
4.2.2 Fragmentation pattern of imine-derivatized tripeptides

To further probe the general fragmentation pattern of imine modified peptides and the apparent N-terminal amino acid elimination, the CID mass spectrum of simple tripeptides with the form XGG derivatized by condensation reaction with salicylaldehyde were obtained. The CID-MS² spectrum of underivatized peptides GGG, AGG, VGG, LGG, FGG, YGG – increasing in N-terminal sidechain size, and underivatized SGG and TGG – whose N-terminal sidechains are implicated in water loss reactions,¹⁶⁰ are presented in Figure 4.3. Each spectrum revealed a simple fragmentation pattern, characteristically containing b, y, and a-ions. The dominate ion generated in each case was the b₂ ion, identified by a loss of 75 u (the C-terminal residue mass of Gly). The a₂ ion identified by a loss of 103 u was also observed for each tripeptide at relative intensities less than 4%. The C-terminal containing y₂ ion identified by the m/z 133 peak was generated by CID of each underivatized peptide, except GGG which generated the y₁ ion at m/z 75 (Figure 4.3a). CID of protonated LGG (Figure 4.3d), FGG (Figure 4.3g), and YGG (Figure 4.3h) generated the a₁ ion at approximately 3%, 12%, and 21% relative intensities, respectively. The a₁-yₓ pathway (Scheme 1.3) has been implicated in the formation of the a₁ and corresponding yₓ ions (e.g., the y₂ ion for tripeptides) for small underivatized peptides.¹⁶¹
Figure 4.3 CID mass spectrum of protonated underivatized a) GGG, b) AGG, c) VGG, d) LGG, e) SGG, f) TGG, g) FGG, and h) YGG.
Note, the \( a_1 \) ion formed from protonated GGG, AGG, and SGG fall below the low-mass cut off (LCMO) of the LIT. Defined in part by the activation \( Q \), the LMCO can be determined by Equation 4.1, where \( m \) is the ion mass and 0.908 the RF amplitude necessary to eject ions from the trap:\textsuperscript{162}

\[
\text{Equation 4.1} \quad \text{LMCO} = \frac{(m \times Q)}{0.908}
\]

In the present case, the \( Q \) was set at 0.25, the default for the Thermo LTQ-XL, therefore determining the LMCO as 52.3 for \([\text{GGG} + \text{H}]^+\), 56.2 for \([\text{AGG} + \text{H}]^+\), and 60.6 for \([\text{SGG} + \text{H}]^+\), placing the \( a_1 \) ion below the limit. Losses of 18 u representing C-terminal \( \text{H}_2\text{O} \) loss (Scheme 1.9),\textsuperscript{163} and 45 u indicative of eliminating \( \text{CO} \) and \( \text{NH}_3 \) via a pathway under investigation by Bythell et al.,\textsuperscript{164} were also observed for each peptide.

The CID-MS\textsuperscript{2} spectra of GGG, AGG, VGG, LGG, FGG, and YGG imine-derivatized with oHB are shown in Figure 4.4. As expected, the \( a \) and \( b \) ions were observed at \( m/z \) values 104 u higher than the analogous underivatized versions, indicative of conserving the imine modification at the N-terminus. Notably, CID of the imine modified peptides in each case generated the full series of \( b_x \) and \( a_x \) ions. Notably, imine-derivatization promoted the generation of the \( b_1 \) ion via CID, characteristically absent from the CID spectra of the underivatized tripeptides. It has been suggested the lack of an amide oxygen N-terminal to \( b_1 \) cleavage site hinders \( b_1 \) ion formation via the oxazolone pathway shown in Scheme 1.1 for underivatized peptides.\textsuperscript{165-167} However, the observed \( b_1 \) ion for imine-derivatized peptides implicates a \( b_1 \) pathway despite lacking the conventional nucleophile. A loss of 18 u likely along the C-terminal
water loss pathway (Scheme 1.9) forming the b₃ ion was also observed for each imine-
derivatized peptide.

CID of each oHB tripeptide generated the [b₂ + H₂O]⁺ ion identified by the peak 18 u higher than the corresponding b₂ ion. Interestingly, the [b₁ + H₂O]⁺ ion was also observed for oHB derivatized AGG at m/z 194 (Figure 4.4b), VGG at m/z 222 (Figure 4.4c), FGG at m/z 270 (Figure 4.4e), and YGG at m/z 286 (Figure 4.4d). In contrast, the [bₓ + H₂O]⁺ ion corresponding to b ions smaller than b₄ was not observed for the CID of the imine-derivatized pentapeptide oHB=YGGFL (Figure 4.2a). More interestingly, CID of the derivatized tripeptides produced several novel fragment ions, including the supposed N-terminal amino acid elimination ion peak, termed the “y-double dagger” ion (yǂ). In contrast to the small peak observed for [oHB=YGGFL + H]⁺, the yǂ ion was observed as the dominate ion peak at m/z 237 for protonated oHB=AGG (-71 u, Figure 4.4b), oHB=VGG (- 99 u, Figure 4.4c), oHB=LGG (- 133 u, Figure 4.4d), oHB=FGG (- 147 u, Figure 4.4e), and oHB=YGG (-163 u, Figure 4.4f). Note, for CID of [oHB=GGG + H]⁺ (Figure 4.4a) the dominate m/z 237 ion can be both indicative the yǂ and [b₂ + H₂O]⁺ ions.

CID of protonated oHB=AGG, oHB=VGG, oHB=FGG, and oHB=YGG also generated small (less than 10% relative intensity) fragment ions at m/z 162 and m/z 180, which appeared to be produced from losing 75 u (C-terminal Gly residue mass) and 57 u (Gly residue mass) respectively, presumably from m/z 237 (yǂ). Note, m/z 162 for [oHB=GGG + H]⁺ is the mass indicative of both the b₁ ion and the loss of 75 u from m/z 237 (Figure 4.4a). Interestingly however, the m/z 180 peak proposed to be an apparent loss of 57 u from m/z 237 was also observed for [oHB=GGG + H]⁺ (Figure 4.4a), suggesting the m/z 237 peak observed for protonated oHB=GGG is a mixture of both the yǂ and [b₂ + H₂O]⁺ ions.
Figure 4.4 CID mass spectrum of protonated a) GGG, b) AGG, c) VGG, d) LGG, e) FGG, and f) YGG, imine-derivatized by condensation reaction with salicylaldehyde (oHB).
CID of polar side-chain containing protonated oHB=SGG and oHB=TGG also generated
the y¹ ion at m/z 237 (Figure 4.5a and Figure 4.5b)), indicating the loss of 87 u (Ser) and 101 u
(Thr) respectively. Interestingly, the product ion peak m/z 162 (i.e. the proposed C-terminal
Gly loss from the y¹ ion) was observed at approximately 82% relative intensity for [oHB=TGG
+ H]^+ (Figure 4.5b), greater than that observed for the y¹ ion. More interestingly, however, the
dominant fragment ion generated for CID of [oHB=TGG + H]^+ was observed at m/z 294 (i.e.,
loss of 44 u) presumably indicative of eliminating the N-terminal amino acid side chain residue.

Pathways proposed for the loss of 44 u from protonated oHB=TGG are shown by Scheme
4.1a, Scheme 4.1b, and Scheme 4.1c. Each path is initiated by protonation at the imine nitrogen
followed by nucleophilic attack by the N-terminal amide oxygen on the benzyl carbon,
generating an oxazole intermediate. Scheme 4.1a involves H-transfer from the Thr sidechain
hydroxyl group to the N-terminal nitrogen and loss of acetaldehyde (44 u). In contrast, Scheme
4.1b involves H-transfer from the Thr sidechain hydroxyl group to the second amino acid
nitrogen, leading to loss of acetaldehyde. Scheme 4.1c involves H-transfer from the Thr
sidechain methyl group to the second amino acid nitrogen, leading to the elimination of ethanol
(44 u) to form the oxazole intermediate.

Note, along Scheme 4.1b and Scheme 4.1c an m/z 162 fragment ion may be generated
from the breakdown of the oxazole intermediate to eliminate GlyGly. Presumably, the
unexpected large intensity observed for m/z 162 for CID of protonated oHB=TGG is
representative of contributions from Scheme 4.1b, Scheme 4.1c and elimination of the C-
terminal glycine residue from the y¹ ion. CID of protonated oHB=SGG also generates a peak at
m/z 294 (i.e., 30 u) with a relative intensity of approximately 4%. Lacking a methyl group at the
N-terminal amino acid side-chain, the loss of 30 u from [oHB=SGG + H]^+ is likely generated
through a pathway similar to the oxazole mechanism proposed in Scheme 4.1a and Scheme 4.1b, assumedly to eliminate formaldehyde (30 u).
Figure 4.5 CID mass spectrum of protonated a) SGG and b) TGG, imine-derivatized by condensation reaction with salicylaldehyde (oHB)
Scheme 4.1 (a) Proposed sidechain hydroxyl → N-terminal nitrogen H-transfer fragmentation pathway for generation of m/z 294 by CID of protonated oHB=TGG. (b) Sidechain hydroxyl → amide nitrogen and (c) sidechain methyl → amide nitrogen H-transfer pathways for the generation of m/z 294, and secondary fragmentation pathway to generate m/z 162 for protonated oHB=TGG.
4.2.3 CID-MS\textsuperscript{n} investigations of novel $y^1$ generating fragmentation pathway

Most interesting about the observed peak at m/z 237 generated from CID of protonated oHB=XGG tripeptides and m/z 497 generated from CID of [oHB=YGGFL + H]\textsuperscript{+} was the implication of a pathway that eliminates the N-terminal amino acid residue while conserving the derivatization. Specifically, not only may the N-terminal amino acid be directly determined from the apparent internal elimination pathway, but fragmentation (e.g., MS\textsuperscript{3}) of the $y^1$ may also be protected from sequence scrambling effects if the conserved derivatization remains N-terminal to the remaining conserved primary sequence. Using protonated oHB=AGG as a model peptide ion, possible mechanisms to explain eliminating the residue mass of the N-terminal amino acid (i.e., 71 u for [oHB=AGG + H]\textsuperscript{+}) are shown in Scheme 4.2a and Scheme 4.2b. In Scheme 4.2a, the reaction is initiated by proton mobilization to the second amino acid amide oxygen. The mobilization of the protonating proton is followed by nucleophilic attack of the imine nitrogen on the C-terminal amide carbon, forming a substituted ketopiperazine intermediate. The intermediate breaks down creating an imine product with a backbone hydroxyaziridine ring and neutral aziridinone loss, representing the elimination of 71 u.

Scheme 4.2b, however, begins with protonation at the imine nitrogen atom, followed by nucleophilic attack by the second amino acid amide nitrogen on the N-terminal imine carbon, creating an imidazole-one intermediate. The formation of an imine-derivatized peptide that is one amino acid residue smaller with conserved primary sequence occurs via ring opening of the imidazole-one intermediate and consequential ejection of CO and an imine. Generation of an imidazole-one intermediate via amide nitrogen attack is similar to the mechanism proposed to explain H\textsubscript{2}O elimination from protonated tetracyglycine\textsuperscript{27, 30, 168}. 

27, 30, 168
Scheme 4.2 Loss of N-terminal residue mass via a) ketopiperazine intermediate b) imidazole-one intermediate for model peptide ion [oHB=AGG + H]^+. 
Early evidence nucleophilic attack at the benzyl carbon, implicated in the \( y^+ \) ion Scheme 4.2b, likely occurs for the oHB derivatized systems was a small fragment ion observed possibly indicative of eliminating the phenol moiety (i.e., 93 u + H) of each protonated oHB=XGG peptide evaluated by CID-MS\(^2\) (Figure 4.4). The proposed mechanism to lose 94 u is shown by Scheme 4.3a and Scheme 4.3b. Scheme 4.3a involves protonation at the imine nitrogen, followed by nucleophilic attack by the amide nitrogen on the second amino acid residue on the benzyl carbon generating an imidazole-one intermediate. Hydrogen transfer from the N-terminal nitrogen to the oHB hydroxyl oxygen leads to breaking the benzyl carbon bond, eliminating water (18 u) and neutral benzyne (76 u). The resulting N-terminal imidazole-one ion represents a fragment ion identified by a total loss of 94 u.

Scheme 4.3 Proposed mechanism for loss of 94 u from protonated oHB=XGG peptide species by (a) an imidazole-one intermediate pathway and (b) an oxazolidine-imine intermediate pathway
Scheme 4.3b begins with protonation at the imine nitrogen, followed however by nucleophilic attack by the N-terminal amide oxygen on the benzyl carbon to generate an oxazole intermediate. Resonance rearrangement to generate an oxazolidine-imine intermediate is followed by hydrogen transfer to the oHB hydroxyl oxygen analogous to Scheme 4.3a and resulting in the elimination of water and neutral benzyne. The resulting N-terminal oxazole-imine ion also represents a fragment ion identified by a total loss of 94 u.

To resolve the implicated loss of phenol moiety (i.e., 93 u + H), the CID spectrum of the protonated unmodified GGA (Figure 4.6a) and GGA imine-derivatized by reaction with benzaldehyde (Bz) and oHB were obtained (Figure 4.6b and Figure 4.6c), respectively. By Scheme 4.3a and Scheme 4.3b, the loss of 94 u from oHB derivatized protonated peptides was presumed to involve hydrogen transfer from the protonated imine-nitrogen to the N-terminal hydroxybenzyl OH. It was hypothesized therefore the loss of 94 u would not be observed for CID of protonated peptides imine derivatized by reaction with benzaldehyde due to the lack of an OH substituent.

The CID-MS² spectrum of protonated unmodified GGA generated the fragment ions identified by loss of 18 u (m/z 186) and 45 u (m/z 159), as well as a dominant b₂ ion at m/z 115 (Figure 4.6a). The y₁ and a₂ ions were also observed for [GGA + H]+. CID of [Bz=GGA + H]+ generated a fragmentation pattern with a wider array of sequence ions (Figure 4.6b), consistent with that observed for N-terminal imine-derivatization by oHB. The dominant fragment ion was observed as the b₂ ion (m/z 203). The b₁ and y₂ ions were observed at small relative intensities, and the a₂ ion at approximately 12% relative intensity. Interestingly, the y₁ ion (m/z 235) generated from [Bz=GGA + H]+ identified by loss of 57 u (i.e., Gly) was observed at approximately 22% relative intensity inconsistent with the dominant y₁ ion observed for the
XGG imine-derivatized peptide species. In contrast, the $y^+$ ion (m/z 251) generated from CID of [oHB=GGA + H]$^+$ (Figure 4.6c), was observed at approximately 76% relative intensity, suggesting an imine-modifier/structural effect on the rate of the reaction mechanism generating the $y^+$ ion. CID of [oHB=GGA + H]$^+$ also generated the full series of b and a ions, as well as the $y_2$ ion at m/z 147 (Figure 4.6c). The dominant ion was observed as the $b_2$ ion at m/z 219. Consistent with the CID-MS$^2$ spectra of the protonated oHB=XGG peptides, the loss of 18 u is observed, as well as the [$b_1$ + H$_2$O]$^+$ ion at m/z 180 and the [$b_2$ + H$_2$O]$^+$ ion at m/z 237.

Expectedly, the loss of 94 u presumed indicative of losing water and neutral benzyne was observed for [oHB=GGA + H]$^+$ at m/z 214, however not observed for [Bz=GGA + H]$^+$ to generate m/z 198 (Figure 4.6b). It was assumed, therefore, the 94 u loss from [oHB=GGA + H]$^+$ likely originates from the N-terminus as postulated by Scheme 4.3. The apparent benzyne loss for [oHB=GGA + H]$^+$ also seemed to further implicate the benzyl carbon as a possible electrophilic site as implicated by the proposed imidazole-one pathway for generating the $y^+$ ion. Additionally, no peak at m/z 215 was observed to indicate a comparable benzyne loss from [Bz=GGA + H]$^+$ by some unknown mechanism.
Figure 4.6 CID mass spectrum of protonated a) GGA, and GGA imine-derivatized by condensation reaction with b) benzaldehyde (Bz) and c) salicylaldehyde (oHB).
4.2.3.1 \( y^+ \) generation from unprotonated imine-derivatized PGG

To investigate the viability of the imidazole-one mechanism (Scheme 4.2b) as the likely pathway for \( y^+ \) ion generation, the CID spectra of underivatized protonated PGG (Figure 4.7a) and oHB derivatized PGG (Figure 4.7b) were obtained. Reactions of aromatic aldehydes and proline have been studied by experimenters and determined to generate a benzyl iminium zwitterion species in equilibrium with a benzyamine oxazolidinone isomer.\(^{169}\) It was reported, the benzyamine oxazolidinone species is likely generated by nucleophilic attack by the carboxy oxide ion on the benzyl carbon, converting the imine nitrogen to an amine.\(^{169}\) Interestingly, the reported benzyamine oxazolidinone structure is similar to the imidazole-4-one intermediate formed by amide nitrogen attack proposed for eliminating the N-terminal amino acid residue to create the \( y^+ \) ion. Additionally, imine derivatization of PGG by condensation reaction with salicylaldehyde was thought to generate an oHB=PGG iminium ion similar in structure of the benzyl iminium species reported for proline and aromatic aldehyde reactions.\(^{169}\) Specifically, the structure of \([\text{oHB}=\text{PGG}]^+\) was presumed to be charged at the N-terminal imine nitrogen similarly to the starting structure proposed in the imidazole-one mechanism (Scheme 4.2b), however, absent a conventional mobile proton (Figure 4.8a).
Figure 4.7 The CID-MS$^2$ spectrum of a) protonated PGG, and PGG derivatized with b) salicylaldehyde (oHB) and c) benzaldehyde (Bz)
The fragmentation pattern of protonated underivatized PGG generated a dominant b₂ ion peak, followed by the a₁ peak at approximately 30% relative intensity thought to be generated through the a₁-y₂ pathway (Scheme 1.3). The a₂ and y₂ product ions were also observed at relative intensities less than 4%. Interestingly, in contrast to the protonated imine-derivatized XGG peptides, the CID-MS² spectrum of unprotonated [oHB=PGG]⁺ (Figure 4.7b) generated a smaller number of sequence ions than underivatized PGG. Additionally, the dominant product ion observed for CID of [oHB=PGG]⁺ was the b₁ ion, while the b₂ and a₁ ions were observed at a relative intensity less than 2%. Presumably, the simple fragmentation pattern of [oHB=PGG]⁺ is indicative of having a “fixed” N-terminal charge at the imine nitrogen (Figure 4.8a) in lieu of the added/protonating proton whose mobilization is classically used to describe peptide fragmentation and the generation of sequence ions. Formation of sequence ions from

Figure 4.8 Proposed structure of non-protonated a) PGG and protonated b) AGG imine derivatized by condensation reaction with salicylaldehyde (oHB).
CID of unprotonated \([oHB=PGG]^+\) was therefore thought to be governed by intramolecular hydrogen transfer reactions involving the oHB hydroxy hydrogen, as proposed in Scheme 4.4a and Scheme 4.4b.

**Scheme 4.4** Proposed intramolecular hydrogen transfer a) \(b_1 \rightarrow a_1\) pathway, and b) \(b_2 - y_1\) pathway for \([oHB=PGG]^+\).
More interestingly, CID of [oHB=PGG]^+ generated the y_{76}^+ ion indicative of losing the residue mass of proline (-97 u) at m/z 237. For completeness, an adapted ketopiperazine (Scheme 4.5a) pathway was postulated for the loss of 97 u from [oHB=PGG]^+. The mechanism involves H-transfer from the oHB hydroxyl substituent to the amide oxygen of the second amino acid residue. Nucleophilic attack by the proline nitrogen on the carbonyl carbon forms a (PGG)-ketopiperazine intermediate. The collapse of the intermediate results in an opening of the proline ring and formation of a 3-[3-(aziridin-1-yl)propyl]aziridin-2-one intermediate. Break down of the aziridine-2-one moiety ejects CO. The formed butyl imine cyclizes creating a proton shared 1-pyrroline/hydroxyaziridine dimer. Dissociation of the dimer leads to either a protonated pyrroline ion or protonated hydroxyaziridine ion indicative of a 97 u loss. In contrast, the adapted (PGG)-imidazole-one pathway shown by Scheme 4.5b for loss of 97 u does not require opening the proline ring, but instead involves elimination of an intact proline ring in a mechanism identical stepwise to that proposed for the loss of 71 u from protonated oHB=AGG (Scheme 4.2b).

To better resolve the mechanism for elimination 97 u from imine derivatized PGG, imine derivatization by condensation reaction with benzaldehyde (Bz) was obtained (Figure 4.7c) given Scheme 4.5a relies on intramolecular H transfer from the hydroxybenzyl OH group to a carbonyl oxygen. Interestingly, the b_{76} ion dominant for CID of [oHB=PGG]^+ was not generated for [Bz=PGG]^+, nor the a_{76} ion. Additionally, the a_{76} ion is produced as a small ion peak (<2% ion intensity) and the b_{76} ion at approximately 10% intensity for CID of [Bz=PGG]^+. Assumedly, the small sequence ion peak intensities are a consequence of benzaldehyde lacking the hydroxy substituent proposed to participate in intramolecular H-transfer reactions.
Scheme 4.5 The a) ketopiperazine and b) imidazole-one pathway (boxed) adapted for the loss of 97 u via CID of [oHB=PGG]⁺
More interestingly, the ion peak representative of losing 97 u \((y^1)\) at m/z 221 was observed as the dominant peak for CID of \([\text{Bz=PGG}]^+\) (Figure 4.7c) despite lacking the hydroxy substituent implicated in Scheme 4.5a. The observance of the \(y^1\) peak for imine-derivatized oHB=PGG and Bz=PGG both presumed to have a fixed N-terminal charge, suggested the imidazole-one pathway (Scheme 4.2b and Scheme 4.5b) are the likely mechanisms for eliminating the residue mass of the N-terminal amino acid. Additionally, H-transfer from the protonated nitrogen of the proposed (PGG)-imidazole-one intermediate in Scheme 4.5b to the C-terminal amide nitrogen followed by elimination of the C-terminal Gly implicates a possible pathway for generating the observed \(b_2\) and \(a_2\) ions from \([\text{Bz=PGG}]^+\).

### 4.2.3.2 CID-MS\(^3\) Determination of \(y^1\) Ion Identity

Under low energy CID, the fragmentation pattern and relative intensities of observed product ions depend upon the internal energy and structure of the precursor ion under analysis.\(^{171}\) However, ion structure cannot be implicitly determined from mass spectra. Nevertheless, the uniqueness of the fragmentation pattern obtained from the CID of precursor and product ions is the basis of tandem mass spectrometric techniques for peptide identification. More importantly, an ion with a particular structure and composition under comparable conditions will generate the same fragmentation pattern each time. Note the imidazole-one pathway for the loss of 71 u (i.e., the residue mass of Ala) from the model imine peptide ion \([\text{oHB=AGG} + \text{H}]^+\) (Scheme 4.2b) generates a final product ion \([\text{oHB=GG} + \text{H}]^+\) with a structure identical to GlyGly independently derivatized with oHB. Presumably therefore, if the structure of the product ion generated from the loss of 71 u from \([\text{oHB=AGG}\)
+ H]⁺ is identical to the structure of protonated GlyGly independently derivatized with oHB the CID mass spectrum of both ions will be indistinguishable from each other, further implicating the imidazole-one mechanism and N-terminal amino acid elimination.

To help resolve the structure of and mechanism for generating the y⁺ ion, the CID mass spectra of protonated tetrapeptide AGGG and protonated tripeptide GGG both imine-derivatized with 2-hydroxybenzaldehyde (oHB) were obtained. Note, the proposed structure of a y⁺ product ion generated via the proposed imidazole-one mechanism (Scheme 4.2b) for protonated oHB=AGGG would be the structure of the protonated tripeptide GGG imine-derivatized with oHB. The mass spectrum of a y⁺ ion generated from CID of [oHB=AGGG + H]⁺ was therefore expected to be indistinguishable from the CID mass spectrum of [oHB=GGG + H]⁺, assuming the loss of 71 u is observed and the imidazole-one pathway is the responsible mechanism for the elimination.

Figure 4.9a shows the CID-MS² spectrum for [oHB=AGGG + H]⁺. Consistent with the fragmentation pattern observed for imine modified tripeptides (Section 4.2.2), the fragmentation pattern of the imine-modified tetrapeptide [oHB=AGGG + H]⁺ is abundant in a- and b-type fragment ions. The most abundant product ion peak generated from CID of [oHB=AGGG +H]⁺ (Figure 4.9a) was the b₂⁺ ion at m/z 233. The b₁, b₃, a₁, a₂, and a₃ ions are also observed for CID of [oHB=AGGG + H]⁺, indicating the generation of the complete series of a and b ions. The [b₁ + H₂O]⁺ (at m/z 194), [b₃ + H₂O]⁺ (at m/z 308), y₂ and y₃ ions are also observed. As seen for the oHB derivatized tripeptides studied (Chapter 4.2.3), the loss of 94 u, presumed to be the loss of benzyne and water, was also detected at m/z 271 with an ion intensity less than 1% (Figure 4.9a). The water loss product ion is also observed at m/z 347.
Figure 4.9  

a) CID-MS\(^2\) spectrum of protonated AGGG imine-derivatized with salicylaldehyde (oHB), [oHB=AGGG + H]\(^+\),
b) CID-MS\(^3\) spectrum of product ion m/z 294 generated from the loss of 71 u from [oHB=AGGG + H]\(^+\),
c) CID-MS\(^2\) spectrum of protonated GGG imine-derivatized with salicylaldehyde, [oHB=GGG + H]\(^+\).
CID of [oHB=AGGG + H]^+ also generated a loss of 71 u, consistent with the residue mass of the N-terminal amino acid (alanine), to produce the m/z 294 fragment ion (y^1) with an intensity of approximately 85%. Interestingly, a small peak with an ion intensity less than 1% at m/z 237 (Figure 4.9a) is also observed that appears to be the loss of 57 u consistent with losing the residue mass of glycine from the m/z 294 ion. More interestingly, the CID-MS^3 spectrum of the y^1 ion isolated from CID of [oHB=AGGG + H]^+ (Figure 4.9b) shows m/z 237 as the dominant fragment ion peak, consequently indicating the origin of the small m/z 237 peak observed from CID-MS^2 of [oHB=AGGG + H]^+ (Figure 4.9a) is likely from dissociation of the y^1 ion (m/z 294).

Both the ketopiperazine pathway (Scheme 4.2a) resulting in an imine-modified N-terminal aziridine product ion structure and the imidazole-one pathway (Scheme 4.2b) resulting in an N-terminal imine-modified peptide product ion structure were applied as possible mechanisms for the loss of 71 u from [oHB=AGGG + H]^+ generating m/z 294. Consequently, two pathways were postulated for generating the loss of 57 u from m/z 294 starting from the aziridine (Scheme 4.6a) and primary peptide (Scheme 4.6b) product ion structures.
Scheme 4.6  

**a)** Proposed mechanism for the elimination of 57 u from the imine-modified aziridine peptide structure of product ion m/z 294 generated from CID of [oHB=AGGG + H]^+.

**b)** Proposed mechanism for the elimination of 57 u from the imine-modified primary peptide structure of product ion m/z 294 generated from CID of [oHB=AGGG + H]^+.
Scheme 4.6a for the loss of 57 u from m/z 294 involves an opening of the aziridine ring generating an imine-modified primary GGG structure, protonated at the N-terminal carbonyl oxygen. Nucleophilic attack of the N-terminal carbonyl carbon by the C-terminal nitrogen generates an imidazole-one intermediate. The collapse of the imidazole ring generates the loss of 57 u by eliminating CO and HN=CH₂. Scheme 4.6b for the loss of 57 u from product ion m/z 294 with an imine-modified GGG structure was adapted from the imidazole-one pathway described by Scheme 4.2b, implicating the elimination of the product ions N-terminal amino acid residue. Given the fragmentation of both initial product ion structures can generate a loss of 57 u, the observance of the m/z 237 peak in the CID mass spectrum of m/z 294 could not be used to resolve the structure of m/z 294 implicitly.

Interestingly however, the CID-MS² fragmentation pattern of [oHB=GGG + H]+ (Figure 4.9c), independently modified, was indistinguishable from the CID spectrum of the m/z 294, suggesting the identity/structure of m/z 294 (yǂ ion) is identical to [oHB=GGG + H]+. Additionally, the identical spectra further implicate the imidazole-one pathway as the mechanism for the loss of 71 u from [oHB=AGGG + H]+ representative of eliminating the N-terminal amino acid residue. Furthermore, the loss of 57 u observed from CID of the yǂ ion generated from CID of [oHB=AGGG + H]+ was implicated as the loss of the new N-terminal amino acid residue (i.e., glycine), suggesting the derivatization and the N-terminal elimination mechanism is conserved through multiple stages of MSⁿ. Consequently, new nomenclature (e.g., bǂ, yǂ, and aǂ) (Figure 4.10) was adopted to distinguish the sequence ions originating from the CID of the yǂ ion:

\[ i = 1, 2, 3, \ldots (R – n + 1), \text{ where } R \text{ is the number of amino acid residues in the peptide sequence and } n \text{ the number of MS stages.} \]
Additionally, the loss of the apparent N-terminal amino acid (glycine) from the \( y^i \) ion (m/z 294) is represented by \( y^{(2)i} \) to indicate the loss of the second N-terminal amino acid in the sequence.

**Figure 4.10** Nomenclature for sequence ions generated from CID of \( y^i \) ions.
4.2.4 IRMPD spectroscopy determination of the product ion generated by the loss of 71 u from protonated oHB=AGG.

To more clearly resolve the pathway for eliminating the apparent N-terminal amino acid residue from imine-modified peptides, IRMPD spectroscopy was used to determine the structure of the precursor and product ions using [oHB=AGG + H]^+ as the model. Determining the structure of the precursor ion before elimination of 71 u and after the elimination of 71 u will give a more unambiguous indication of the elimination mechanism as if to model the beginning and end of the reaction pathway. The proposed ketopiperazine (Scheme 4.2a) and imidazole-one (Scheme 4.2b) pathways for eliminating the N-terminal amino acid residue were used as the starting point of this study, beginning with the site of protonation.

For intact, protonated oHB=AGG, several minima were identified by DFT calculations, including structures in which the imine nitrogen atom is protonated (Scheme 4.2b) and those for which the amide oxygen served as the protonation site (Scheme 4.2a). The four lowest energy structures are shown in Figure 4.11 and the relative energies provided in Table 4.1. The lowest energy structure determined for protonated oHB=AGG was structure M1 (Figure 4.11a), featuring a β-strand cationized with the proton positioned at the imine nitrogen atom. In order of increasing energy, structure M2 is also protonated at the imine nitrogen atom, though differed in the angle of a dihedral angle allowing H-bonding interaction between the imine and the C-terminal oxygen atom. Structures M3 and M4 were protonated at the amide oxygen exhibiting energies 9.5-16.2 kcal/mol higher than M1. The experimentally determined IRMPD spectrum of protonated oHB=AGG (Figure 4.12a) was compared to those predicted by DFT for the lowest energy nitrogen-protonated and oxygen-protonated species M1 (Figure 4.12b) and M3 (Figure 4.12c)), respectively. The best agreement to the experimental IRMPD spectrum was observed
for the theoretical spectrum of structure M1, strongly suggesting the imine nitrogen as the intrinsic protonation site for protonated oHB=AGG.

![Image of lowest energy structures predicted by DFT for protonated oHB=AGG.](image)

**Figure 4.11** Lowest energy structures predicted by DFT for protonated oHB=AGG. Nitrogen (blue), oxygen (red), carbon (dark grey), hydrogen (light grey).

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>ZPE</th>
<th>E + ZPE</th>
<th>ΔE (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>-1084.86604</td>
<td>0.32177</td>
<td>-1084.54427</td>
<td>0</td>
</tr>
<tr>
<td>M2</td>
<td>-1084.85711</td>
<td>0.32212</td>
<td>-1084.53499</td>
<td>5.8</td>
</tr>
<tr>
<td>M3</td>
<td>-1084.85038</td>
<td>0.32121</td>
<td>-1084.52917</td>
<td>9.5</td>
</tr>
<tr>
<td>M4</td>
<td>-1084.84045</td>
<td>0.32205</td>
<td>-1084.5184</td>
<td>16.2</td>
</tr>
<tr>
<td>M5</td>
<td>-837.3679685</td>
<td>0.236842</td>
<td>-837.131127</td>
<td>54.4</td>
</tr>
<tr>
<td>M6</td>
<td>-837.4127483</td>
<td>0.254032</td>
<td>-837.17441</td>
<td>27.3</td>
</tr>
<tr>
<td>M7</td>
<td>-837.365689</td>
<td>0.237762</td>
<td>-837.127927</td>
<td>56.4</td>
</tr>
<tr>
<td>M8</td>
<td>-837.4563257</td>
<td>0.238487</td>
<td>-837.217839</td>
<td>0</td>
</tr>
<tr>
<td>M9</td>
<td>-837.45327</td>
<td>0.239226</td>
<td>-837.214044</td>
<td>2.4</td>
</tr>
<tr>
<td>M10</td>
<td>-837.4300069</td>
<td>0.238536</td>
<td>-837.191471</td>
<td>16.5</td>
</tr>
</tbody>
</table>

**Table 4.1.** Electronic, zero-point, and relative energies of minima determined for oHB=AGG (M1→4) and possible products created by N-terminal amino acid residue elimination (M1→10).
Figure 4.12. a) IRMPD spectrum of protonated oHB=AGG compared to the predicted spectra of lowest energy, b) imine-N protonated (structure 1), and c) amide-O protonated (structure 2) protonated oHB=AGG. The experimental IRMPD spectrum from (a) is transposed as a light grey trace in (b) and (c) for comparison.
Using the product ion structures proposed as a result of the ketopiperazine pathway (Scheme 4.2a) and the imidazole-one pathway (Scheme 4.2b) possible structures (Figure 4.13) resulting from the elimination of 71 u were identified using DFT (relative energies provided in Table 4.1). The identified structures M5, M6, and M7, represent product ion structures representative of the ketopiperazine pathway. Structures M5 and M7 both feature an intact hydroxyazridine ring, while structure M6 is resultant of the aziridine ring opening to form an oxygen-protonated oHB=GG structure in a cis conformation.

However, the three low energy structures, identified as 10.8-56.4 kcal/mol lower in energy than those implicated in the ketopiperazine mechanism, belonged to structures M8, M9, and M10. These structures (M8-M10) are all conformational isomers of oHB=GG, possibly generated through the imidazole-one mechanism. The lowest energy structure, M8, determined by DFT, corresponds to a β-strand, imine nitrogen protonated structure. It is important to note the main difference between the structures M8 and M9 separated by 2.4 kcal/mol in energy, is the change in the dihedral angle of product ion structure M9, breaking of the intramolecular hydrogen bond between the protonated imine nitrogen and the orthopositioned hydroxy substituent of the imine modifier. These structures implicated the possible importance of an intramolecular hydrogen bonding effect (shown by structure M8) for promoting the presumed N-terminal amino acid residue loss.

Figure 4.14a shows the IRMPD spectrum obtained from the m/z 237 product ion created by eliminating 71 u from [oHB=AGG + H]^+. The theoretical IRMPD spectra obtained of the hydroxyazridine product ion structure, M5 (Figure 4.14c), as well as the trans amide oHB=GG structure, M8 (Figure 4.14b), was compared to the experimental IRMPD spectrum of product ion m/z 237 in Figure 4.14a. Additionally, the experimental IRMPD spectrum was obtained of
protonated dipeptide GG (Figure 4.14d), independently imine derivatized by condensation reaction with oHB, and compared to the m/z 237 spectrum shown in Figure 4.14a). The best agreement to the IRMPD spectrum of m/z 237 was observed from the theoretical spectrum of structure M8 (Figure 4.14b), as well as the experimental spectrum of [oHB=GG + H] derivatized independently (Figure 4.14d). Near identical agreement between the m/z 237 IRMPD spectrum and that of [oHB=GG + H] strongly implicated the m/z 237 ion identity as oHB=GG. Agreement between the theoretical spectrum of structure M8 indicative of oHB derivatized GG protonated at the imine nitrogen, further strengthened the evidence supporting the 71 u loss generated from CID of [oHB=AGG + H] is indicative of eliminating the N-terminal residue mass of alanine from the protonated imine-derivatized peptide. Additionally, it strongly identified the proposed imidazole-one mechanism (Scheme 4.2b) as the pathway leading to the N-terminal residue loss and generation of the y ion.

**Figure 4.13** Possible structures of product ions generated by the elimination of 71 u (N-terminal amino acid residue) from activation of [oHB=AGG + H].
Figure 4.14. a) IRMPD spectrum of product ion m/z 237 generated by the loss of 71 u from protonated oHB=AGG compared to the predicted spectra of the lowest energy structures formed by the b) imidazole-one pathway (structure M8) and c) the ketopiperazine pathway (structure M5). d) IRMPD spectrum of protonated oHB=GG. The experimental IRMPD spectrum from (a) is transposed as a light grey trace in (b), (c), and (d) for comparison.
4.3 Conclusions

In the presented study, when compared to the CID mass spectrum of protonated unmodified Leu-enkephalin (YGGFL), the CID mass spectrum of protonated YGGFL imine-modified by condensation reaction with salicylaldehyde (oHB) generated a fragmentation pattern with a more extensive array of a and b type sequence ions. Specifically, the full series of b ions (b₁ to b₄) were observed, as well as ions a₄, a₂, and a₁. “Scrambled” product ions generated from CID of unmodified [YGGFL + H]⁺ previously observed and determined by the work of Vachet and coworkers,³¹ were not observed in the CID-MS² spectrum of [oHB=YGGFL + H]⁺ implicating N-terminal imine modification was successful at eliminating sequence scrambling.

More interestingly, CID of [oHB=YGGFL + H]⁺ produced a novel product ion (yǂ ion) at m/z 497 indicative of eliminating the residue mass of the N-terminal amino acid, thus a method to directly identify the N-terminal amino acid. The CID mass spectra of oHB derivatized tripeptides GGG, AGG, VGG, LGG, FGG, YGG, SGG, and TGG also generated the yǂ ion indicative of losing the residue mass of each respective N-terminal amino acid. Additionally, the series of imine-modified tripeptides provided a representation of the general fragmentation pattern generated from N-terminal imine-derivatization. In comparison to the underivatized analogs, the imine-modified peptides produced a more extensive/complete series of N-terminal a and b type product ions to be used for enhanced peptide sequencing, so named using the traditional nomenclature despite the presence of the modification. The generation and ion intensity of y-type product ion series appeared to be reduced, however, presumably due to the N-terminal imine nitrogen retaining the protonating proton following the cleavage step of the b/y pathway.
The pathway resulting in the apparent elimination of the residue mass of the N-terminal amino acid was investigated further using the model imine-derivatized peptide oHB=AGG by IRMPD spectroscopy and density functional theory (DFT). Possible DFT predicted structures for the [oHB=AGG+H]+ precursor ion and the m/z 237 product ion resulting from loss of 71 u were identified and the corresponding theoretical IRMPD spectra compared to those determined experimentally. The IRMPD spectrum of the [oHB=AGG+H]+ precursor ion matched that predicted for a β-strand, imine nitrogen protonated conformer, while the m/z 237 product ion spectrum matched that predicted for a β-strand, imine nitrogen protonated oHBGG conformer.

Additionally, the IRMPD spectrum of [oHB=GG + H]+ derivatized independently from the dipeptide GG, matched near identically that of the m/z 237 product ion. It was concluded, therefore, the apparent N-terminal amino acid residue mass is eliminated through a pathway that involves a nucleophilic attack by the amide nitrogen of the second amino acid residue in sequence on the benzyl carbon N-terminal to the protonated imine nitrogen generating an imidazole-4-one intermediate. The collapse of the imidazole ring generates the β-strand, imine nitrogen protonated oHB=GG conformer with a retained primary sequence.

The confirmed DFT predicted β-strand structures of the precursor ion [oHB=AGG + H]+ and N-terminal amino acid elimination product ion [oHB=GG + H]+ protonated at the N-terminal imine revealed a possible hydrogen bonding interaction between the N-terminal imine and the hydroxybenzyl oxygen. Interestingly, comparison of the mass spectra of protonated Bz=GGA and oHB=GGA suggested N-terminal amino acid elimination becomes less competitive when the hydrogen bonding interaction between the protonating proton and the hydroxybenzyl oxygen at the ortho position of the modifier is interrupted/missing. An investigation was performed into the apparent influence the derivatizing aldehyde has on the novel fragmentation pathway
eliminating the N-terminal amino acid. Particularly considering modifications that eliminate the apparent hydrogen bonding interaction at the protonated N-terminus, can reveal the influence imine modifier structure has on product ion formation from imine derivatized peptides. These studies are presented in Chapter 5.
4.4 References


CHAPTER 5

PRODUCT ION DISTRIBUTIONS FROM CID OF MODEL PROTONATED PEPTIDE, ALANYLGLYCYLGLYCINE, IMINE-DERIVATIZED BY DIFFERENT N-TERMINAL MODIFIERS

5.1 Introduction

Notable influences on CID generated product ion distributions for protonated peptide ions include charge state, the composition of basic amino acid residues, and the identity of the amino acid adjacent to the cleaved bond.\textsuperscript{172-177} For example, the presence of proline in peptide sequence results in fragment ion peaks consistent with bond cleavage N-terminal to the proline residue, known as the proline effect.\textsuperscript{26, 178, 179} However, inadequacy of current peptide sequencing methods is the disregard of observed product ion intensities while focusing wholly on m/z data, or only pragmatically incorporating ion intensity information by indirect application of the most abundant peaks.\textsuperscript{180} However, the current treatment of ion intensity data is a consequence of the inability to accurately and explicitly predict product ions and their abundance.\textsuperscript{181, 182}

Some work has gone forth to computationally simulate and predict peptide MS/MS spectra intensity patterns derived from low-energy CID by incorporating known, and author observed CID spectra within database searches.\textsuperscript{181, 182} Database methods that combine peak intensity information with amino acid composition have also been shown to improved rates of peptide identification.\textsuperscript{183, 184} Yet, these methods presumably would fall short in their predictions of novel
peptides and modifications with unknown and lesser studied/documented fragmentation patterns.

In the present case, IRMPD spectroscopy of model [oHB=AGG + H]⁺ has indicated the imidazole-one pathway (Scheme 4.2b) as the likely charge-mediated mechanism for generating the novel y¹ ion indicative of eliminating the N-terminal amino acid residue. Specifically, IRMPD spectroscopy has indicated the ionizing proton of [oHB=AGG + H]⁺ is located at the N-terminal imine nitrogen, coinciding with the first step of the imidazole-one pathway (Chapter 4). Additionally, structural modeling of [oHB=AGG + H]⁺ revealed a possible stabilizing hydrogen-bonding interaction between the hydroxybenzyl oxygen atom and the protonating H⁺ (Figure 4.11). Assumedly, the apparent H-bonding interaction promotes protonation at the imine nitrogen, and therefore, the charge mediated N-terminal amino acid loss.

To investigate the apparent hydrogen bonding interaction as a feature influencing y¹ ion peak intensity and to observe the general product ion distribution of imine-derivatized peptides, the imine modifier was systematically exchanged for the model peptide AGG and the CID-MS² spectrum obtained. Benzaldehyde, x-methoxybenzaldehyde, and x-hydroxybenzaldehyde (where x = ortho, para, or meta) were used to generate imine derivatized AGG peptides with the hydrogen-bonding substituent (i.e., OH and -OMe) missing or positioned further away from the DFT determined imine nitrogen protonation site. It was hypothesized imine modifications lacking the H-bonding interaction generate the y¹ ion at a lesser ion intensity than observed for [oHB=AGG + H]⁺ (i.e., 100%).
5.2 Results and Discussion

5.2.1 CID-MS$^2$ spectra of protonated benzyl-, hydroxybenzyl-, and methoxybenzyl-imine modified AGG

Figure 5.1a and Figure 5.2a show the CID-MS$^2$ spectra of protonated benzyl imine (Bz) derivatized AGG, and the CID-MS$^2$ spectra of para (p), meta (m), and ortho (o) hydroxybenzyl imine (HB) derivatized AGG are shown in Figure 5.1b, Figure 5.1c, and Figure 5.1d, respectively. Figure 5.2b, Figure 5.2c, and Figure 5.2d show the CID-MS$^2$ mass spectra for the para (p), meta (m), and ortho (o) methoxybenzyl imine (MB) derivatized AGG peptides, respectively. For convenience, the structure and observed sequence ions are listed in Table 5.1. The CID-MS$^2$ spectra of protonated oHB=AGG, oMB=AGG, mHB=AGG, mMB=AGG, pHB=AGG, and pMB=AGG all generated the complete series of $b_x$ and $a_x$ ions. However, the $b_1$ ion was observed at a relative intensity less than 2% for protonated oMB=AGG, mHB=AGG, mMB=AGG, pHB=AGG, and pMB=AGG. The $y_2$ ion was not observed in the CID-MS$^2$ spectra of the investigated imine-derivatized AGG species except for oHB=AGG at a relative intensity less than 1%. Water loss is also observed for each imine-derivatized AGG species. The $[b_2+\text{H}_2\text{O}]^+$ and $[b_1+\text{H}_2\text{O}]^+$ ions were observed for protonated oHB=AGG, oMB=AGG, mHB=AGG, mMB=AGG, and pMB=AGG at relative intensities less than 2%. The $[b_2+\text{H}_2\text{O}]^+$ ion is also observed for protonated pHB=AGG.
Figure 5.1 CID-MS² spectrum of protonated a) Benzaldehyde AGG, b) p-Hydroxybenzaldehyde AGG, c) m-Hydroxybenzaldehyde AGG, d) o-Hydroxybenzaldehyde AGG.
Figure 5.2 CID-MS\textsuperscript{2} spectrum of protonated a) Benzaldehyde AGG, b) p-Methoxybenzaldehyde AGG, c) m-Methoxybenzaldehyde AGG, d) o-Methoxybenzaldehyde AGG.
<table>
<thead>
<tr>
<th>SCHIFF BASE</th>
<th>STRUCTURE</th>
<th>Observed Sequence Ions</th>
<th>$y^+$ Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzyl imine (Bz=AGG)</td>
<td><img src="image" alt="Structure of Benzyl Imine" /></td>
<td>$b_2$, $a_2$, $b_1$</td>
<td>yes</td>
</tr>
<tr>
<td>o-Hydroxybenzyl imine (oHB=AGG)</td>
<td><img src="image" alt="Structure of o-Hydroxybenzyl Imine" /></td>
<td>$b_2$, $a_2$, $b_1$, $a_1$, $y_2$</td>
<td>yes</td>
</tr>
<tr>
<td>o-Methoxybenzyl imine (oMB=AGG)</td>
<td><img src="image" alt="Structure of o-Methoxybenzyl Imine" /></td>
<td>$b_2$, $a_2$, $a_1$, $b_1&lt;2%$, $b_1&lt;2%$,</td>
<td>yes</td>
</tr>
<tr>
<td>m-Hydroxybenzyl imine (mHB=AGG)</td>
<td><img src="image" alt="Structure of m-Hydroxybenzyl Imine" /></td>
<td>$b_2$, $a_2$, $a_1$, $b_1&lt;2%$, $b_1&lt;2%$,</td>
<td>yes</td>
</tr>
<tr>
<td>m-Methoxybenzyl imine (mMB=AGG)</td>
<td><img src="image" alt="Structure of m-Methoxybenzyl Imine" /></td>
<td>$b_2$, $a_2$, $a_1$, $b_1&lt;2%$, $b_1&lt;2%$,</td>
<td>yes</td>
</tr>
<tr>
<td>p-Hydroxybenzyl imine (pHB=AGG)</td>
<td><img src="image" alt="Structure of p-Hydroxybenzyl Imine" /></td>
<td>$b_2$, $a_2$, $a_1$, $b_1&lt;2%$, $b_1&lt;2%$,</td>
<td>yes</td>
</tr>
<tr>
<td>p-Methoxybenzyl imine (pMB=AGG)</td>
<td><img src="image" alt="Structure of p-Methoxybenzyl Imine" /></td>
<td>$b_2$, $a_2$, $a_1$, $b_1&lt;2%$, $b_1&lt;2%$,</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 5.1. Structure of protonated benzyl, hydroxybenzyl, and methoxybenzyl imine-derivatized AGG and list of CID generated sequence ions.
Interestingly, a product ion is observed at m/z 122 for CID of protonated pHB=AGG (Figure 5.1b) and mHB=AGG (Figure 5.1c), with a relative intensity less than 2% that appears to be an N-terminal immonium ion (NI\textsubscript{m}+\textsuperscript{+}) containing the derivatization and the N-terminal nitrogen. A comparable product ion is observed at m/z 136 from CID of protonated pMB=AGG (Figure 5.2b), mMB=AGG (Figure 5.2c), and oMB=AGG (Figure 5.2d). Scheme 5.1 shows a possible mechanism for generating the proposed NI\textsubscript{m}+ ion that includes N-terminal nitrogen protonation followed by nucleophilic attack of the neighboring amide nitrogen on the N-terminal side alpha carbon to produce an aziridine-immonium proton shared dimer. Dissociation of the dimer then leads to either generation of the NI\textsubscript{m}+ ion with an m/z of 122 or 136 for protonated HB derivatized and MB derivatized AGG, respectively, or an aziridine ion that collapses to form a C-terminal containing imine ion with an m/z of 159 - observed at a low relative intensity for protonated Bz=AGG (Figure 5.1a) and mHB=AGG (Figure 5.2c).

Interestingly, the y\textsuperscript{i} ion corresponding to the elimination of the N-terminal amino acid was observed as the most abundant product ion for each protonated AGG peptide derivatized with hydroxybenzaldehyde and methoxybenzaldehyde despite the benzyl substituent position. An ion intensity of 90.48% was observed for protonated Bz=AGG. However, the ion intensity of product ions generated from the different hydroxybenzyl and methoxybenzyl peptide-imine isomers were observed to vary, indicating an apparent effect the modifier structure has on the peptide ion dissociation chemistry.
Scheme 5.1 Proposed mechanism for the formation of the proposed N-terminal immonium ion generated from protonated AGG imine derivatized by reaction with salicylaldehyde.
5.2.2 Energy (NCE) and activation time (t) resolved dissociation profiles

Energy-resolved collision-induced dissociation tandem mass spectrometry, performed by collecting ion intensity data over a stepwise increase of collision energy to obtain a dissociation curve (profile), has been used to investigate the structures of gas-phase ions.\textsuperscript{186-192} These experiments exploit the effect the internal energy distribution of excited ions, the dynamics of activation and the activation time scale have on product ion formation and intensity.\textsuperscript{193} Particularly, differences in the position and shape of obtained dissociation curves have been reported as evidence of isomeric differentiation/identification (e.g., subtle differences in structure)\textsuperscript{194,195} and qualitative analysis of the energy required to dissociate the precursor ion of study.\textsuperscript{193-202}

To elucidate the effect of N-terminal modifier structure on the fragmentation of imine-derivatized peptides, energy-resolved dissociation profiles of protonated benzyl-, hydroxybenzyl-, and methoxybenzyl- imine derivatized AGG were obtained. These dissociation curves (profiles) were generated by plotting fragmentation yield (FY) as a function of collision energy.\textsuperscript{194,195,196,199,203} Equation 5.1 (where $I_{\text{total}}$ is the total sum of product ion intensities and $I_{\text{precursor}}$ the precursor ion intensity) depicts when FY = 0 the precursor ion is unfragmented, and completely dissociated at FY = 1. A sigmoidal fit was determined for each profile curve with an $R^2$ value $\geq 0.98$ and a 99% confidence interval.\textsuperscript{194,195}

\begin{equation}
\text{Equation 5.1 Fragmentation Yield} = \frac{I_{\text{total}} - I_{\text{precursor}}}{I_{\text{total}}} = \frac{\sum \text{product ion intensities}}{I_{\text{total}}}
\end{equation}
Semi-quantitative approaches for profile curve analysis include: determining the collision energy required to bring about a certain percentage of fragmentation yield or product ion intensity (i.e. 5% (E_{5%}), 10% (E_{10%}), or 50% (E_{50%})),\textsuperscript{192, 196, 198, 203} and linear extrapolation\textsuperscript{194, 197, 199, 200} of the profile curve slope through E_{50%} to the baseline (x-axis) resulting in an appearance (threshold) energy for the onset of precursor ion dissociation. The values resulting from each approach have been used to qualitatively address the stability of precursor ions affected by changes in precursor ion structure.\textsuperscript{194, 196-199} Both linear extrapolation and E_{50%} determination of the fitted sigmoidal curve were utilized here for CID of each investigated imine derivatized AGG analog.

It is important to note, consistent with common observations for the fragmentation of large precursor ions,\textsuperscript{204} the appearance energies (AE) and E_{50%} values of each imine-derivatized profile curve (Figure 5.3 and Figure 5.4) was observed to occur at uniquely higher collision energies (kinetic shift) when short activation times are employed. Data interpretation of energy-resolved profiles has been reported to be more reliable at higher activation times due to the reduced kinetic shift.\textsuperscript{198} As a result, comparative analysis, and discussion of each energy-resolved dissociation curve were performed at an activation time, t = 1000 ms.

Interestingly at activation time t = 1000 ms, the position and shape of the energy-resolved profile curves were observed uniquely differentiable for each derivatized peptide-imine allowing for isomeric identification of oHB=AGG, mHB=AGG, and pHB=AGG (Figure 5.3c), as well as for the isomeric oMB=AGG, mMB=AGG, and pMB=AGG ions (Figure 5.4c). Specifically, ortho species (black) may be characterized by a steady, flattened/upshifted curve and protonated pHB=AGG (Figure 5.3c) characterized by a steep dissociation curve for the hydroxybenzyl series. It is important to note though appearing in the dissociation profile plots, comparative
analysis of the [Bz=AGG + H]^+ dissociation curve to the hydroxybenzyl and methoxybenzyl series should be avoided. Energy-resolved experiments of metalloporphyrin-His-containing peptide systems have shown a linear relationship between degrees of freedom and ion stability under CID conditions, resulting in dissociation curves with flatter slopes for larger ions. This suggests a comparative analysis of ion stability for systems with different degrees of freedom exhibit a degree of uncertainty if the size dependence is not known and accounted.

Comparative analysis using experimental appearance energy values in conjunction with DFT calculated appearance energies has been purported to better reduce the apparent “size effect” on energy-resolved dissociation curves. Presumably; however, comparative analysis of the slope/shape of the dissociation curves from isomeric ions is indicative of the underlying thermochemical considerations minus the apparent size effect. Given the isomeric nature of the hydroxybenzyl and methoxybenzyl series precursor ions investigated and in the absence of DFT calculated AE values, the determined E_{50%} values are principally used in the comparative analysis of the obtained dissociation curves. However, both the AE and E_{50%} values are listed in Table 5.2.

The E_{50%} values determined for the dissociation of the hydroxybenzyl imine derivatized series was observed at 7.82 eV for [mHB=AGG + H]^+, 8.30 eV for [pHB=AGG + H]^+, and 8.90 eV for [oHB=AGG + H]^+ implicating the[oHB=AGG+H]^+ ion as the most stable isomer. The E_{50%} values for the methoxybenzyl series was observed at 7.91 eV for [mMB=AGG + H]^+, 11.39 eV for [pMB=AGG + H]^+, and 14.17 eV for [oMB=AGG + H]^+ revealing a meta < para < ortho order of increasing ion stability for both peptide-imine series. Presumably, the hydrogen-bonding interaction between the ortho-positioned -OH or -OCH₃ benzyl substituent and the protonated N-terminal imine nitrogen (implicated by the DFT calculated low energy
[oHB=AGG + H]^+ structure shown in Figure 4.11) acts as a stabilizing effect for the protonated peptide ion. However, no hydrogen-bonding interaction is available for the meta and para hydroxybenzyl and methoxybenzyl isomers. To better assess the observed E_{50\%} trend, the resonance structures of the protonated peptide imine isomers were considered. Note, the threshold energies determined by linear extrapolation revealed a meta < ortho < para stability trend. Calculated AE values are needed to assess the validity of the experimental AE values.

<table>
<thead>
<tr>
<th></th>
<th>R^2 (sigmoid)</th>
<th>99% Confidence Interval</th>
<th>E_{50%} (eV)</th>
<th>R^2 (linear)</th>
<th>Slope (linear)</th>
<th>Appearance Energy (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Bz=AGG + H]^+</td>
<td>precursor</td>
<td>0.9988 ± 0.0479</td>
<td>7.45</td>
<td>0.9966</td>
<td>0.4187</td>
<td>6.25</td>
</tr>
<tr>
<td></td>
<td>y' ion</td>
<td>0.9992 ± 0.0150</td>
<td>7.55</td>
<td>0.9926</td>
<td>0.1483</td>
<td>6.36</td>
</tr>
<tr>
<td>[oHB=AGG + H]^+</td>
<td>precursor</td>
<td>0.9989 ± 0.0411</td>
<td>8.90</td>
<td>0.9995</td>
<td>0.2135</td>
<td>6.56</td>
</tr>
<tr>
<td></td>
<td>y' ion</td>
<td>0.9990 ± 0.0181</td>
<td>9.03</td>
<td>0.9996</td>
<td>0.0965</td>
<td>6.69</td>
</tr>
<tr>
<td>[mHB=AGG + H]^+</td>
<td>precursor</td>
<td>0.9936 ± 0.0965</td>
<td>7.82</td>
<td>0.9999</td>
<td>0.2706</td>
<td>5.97</td>
</tr>
<tr>
<td></td>
<td>y' ion</td>
<td>0.9974 ± 0.0255</td>
<td>8.13</td>
<td>0.9992</td>
<td>0.0998</td>
<td>6.02</td>
</tr>
<tr>
<td>[pHB=AGG + H]^+</td>
<td>precursor</td>
<td>0.9982 ± 0.0557</td>
<td>8.30</td>
<td>0.9992</td>
<td>0.3659</td>
<td>6.93</td>
</tr>
<tr>
<td></td>
<td>y' ion</td>
<td>0.9983 ± 0.0244</td>
<td>8.43</td>
<td>0.9966</td>
<td>0.1461</td>
<td>6.97</td>
</tr>
<tr>
<td>[oMB=AGG + H]^+</td>
<td>precursor</td>
<td>0.9968 ± 0.0569</td>
<td>14.17</td>
<td>0.9963</td>
<td>0.0750</td>
<td>7.49</td>
</tr>
<tr>
<td></td>
<td>y' ion</td>
<td>0.9973 ± 0.0288</td>
<td>14.06</td>
<td>0.9979</td>
<td>0.0423</td>
<td>7.98</td>
</tr>
<tr>
<td>[mMB=AGG + H]^+</td>
<td>precursor</td>
<td>0.9964 ± 0.0746</td>
<td>7.91</td>
<td>0.9993</td>
<td>0.2355</td>
<td>5.77</td>
</tr>
<tr>
<td></td>
<td>y' ion</td>
<td>0.9961 ± 0.0384</td>
<td>8.34</td>
<td>0.9972</td>
<td>0.1025</td>
<td>5.98</td>
</tr>
<tr>
<td>[pMB=AGG + H]^+</td>
<td>precursor</td>
<td>0.9951 ± 0.0813</td>
<td>11.39</td>
<td>0.9966</td>
<td>0.1445</td>
<td>7.93</td>
</tr>
<tr>
<td></td>
<td>y' ion</td>
<td>0.9978 ± 0.0257</td>
<td>11.89</td>
<td>0.9999</td>
<td>0.0646</td>
<td>8.04</td>
</tr>
</tbody>
</table>

**Table 5.2.** Experimentally determined CID profile curve E_{50\%} and appearance energies at an activation time of 1000 ms.
Figure 5.3 Collision energy resolved CID-MS² precursor ion dissociation experiments for protonated Bz=AGG (yellow), pHB=AGG (red), mHB=AGG (blue), and oHB=AGG (black) at activation times a) t = 10 ms, b) t = 100 ms, c) t = 1000 ms, with linear extrapolation (dashed).
Figure 5.4 Collision energy resolved CID-MS\textsuperscript{2} precursor ion dissociation experiments for protonated Bz=AGG (yellow), pMB=AGG (red), mMB=AGG (blue), and oMB=AGG (black) at activation times a) $t = 10$ ms, b) $t = 100$ ms, c) $t = 1000$ ms, with linear extrapolation (dashed).
Benzyl imines protonated at the imine nitrogen are particularly unique, where the protonated imine nitrogen can pull electron density from the benzyl moiety creating a benzyl carbocation that is resonance stabilized (Figure 5.5a). The benzyl imine resonance structures serve as benchmark stabilizing structures available to each the meta- (Figure 5.5b), ortho- (Figure 5.6a), and para- (Figure 5.6b) hydroxybenzyl imine species and are comparable to the methoxybenzyl imine resonance structures. However, an additional resonance structure unavailable at the meta positions was observed when the hydroxy substituent is at the para and ortho positions. This structure is a result of electron density pushed into the benzyl ring from the hydroxy oxygen, placing the ionizing charge on the oxygen further stabilizing the ortho and para hydroxybenzyl- ion structures of N-terminally protonated peptide-imines.
Figure 5.5 Resonance structures of the a) benzylimine; b) meta-hydroxybenzyl imine modification, where R is the peptide (i.e., AGG) starting with the N-terminal amino acid α-carbon.
Figure 5.6 Resonance structures of the a) ortho- b) para- hydroxybenzyl imine modification, where R is the peptide (i.e., AGG) starting with the N-terminal amino acid α-carbon.
The slope of energy-resolved dissociation curves has been used to qualitatively determine the corresponding precursor ion rate of dissociation, where faster (kinetically favored) rates generate curves with steep slopes.\textsuperscript{199} Fast charge-mediated peptide fragmentation has been suggested to persist when fragmentation occurs in non-selective ways – indicative of randomized (entropic) protonation at various sites along the peptide backbone.\textsuperscript{198} Particularly, the probabilities that several activated complex configurations of protonated peptides become explicitly sampled (as purported by QET to determine dissociation rate)\textsuperscript{128-130} is affected by preferential/selective protonation at specific reaction sites, like the sidechain of basic amino acid residues.\textsuperscript{198} Assumedly, differences in the dissociation curve of investigated model peptide-imine isomers can be interpreted by assessing the subtle structural differences that affect proton mobility and thus the energy-resolved dissociation curve slope.

Semi-quantitative dissociation curve slope values are listed in Table 5.2, determined by calculating the slope of the steepest portion of the corresponding energy-resolved plot. The dissociation slope was observed at 0.2135 for [oHB=AGG + H]+, at 0.2706 for [mHB=AGG + H]+, and 0.3659 for [pHB=AGG + H]+; the steepest slope represented by the pHB species indicating fast dissociation. Interestingly, the observed increasing steepness order, ortho < meta < para, for the hydroxybenzyl imine series, was not observed of the methoxybenzyl series, where the [mMB=AGG + H]+ dissociation profile was observed the steepest at 0.2355. Expectedly, the ortho-positioned hydroxybenzyl, and methoxybenzyl modifiers resulted in the flattest dissociation slopes likely because the apparent hydrogen-bonding interaction at the protonated N-terminal imine nitrogen in conjunction with resonance stabilizing effects promote selective protonation at the N-terminus.
However, it was not readily clear the reason the dissociation curve slope for the para-peptide-imine analog was observed as the steepest amongst the hydroxybenzyl series, while shallower than the meta-analog for the methoxybenzyl series (Figure 5.4c). Possibly the hydroxy substituent of the HB series participates in intramolecular hydrogen-transfer reactions that increase the incidence of protonation (proton mobility) at active sites along the peptide backbone, apparently less likely with benzyl methoxy substituents. Scheme 5.2 proposes an intramolecular hydrogen-transfer mechanism for generating a protonated (iminomethyl) phenolate peptide ion species from a protonated pHB derivatized peptide.

The mechanism is initiated by protonation at the imine nitrogen followed by proton transfer from the hydroxy substituent to a reactive site along the peptide backbone, resulting in a phenolate ion with a net positive charge. Resonance stabilization of the intermediate (iminomethyl) phenolate ion species is possible at para position, making the para-hydroxy substituent position assumedly more acidic than the meta position. However, at the ortho position, the phenolate oxygen may abstract the hydrogen from the protonated N-terminal imine nitrogen regenerating the first structure (Scheme 5.3). In this case, it is plausible the entropic mobilization of the ionizing proton is greatest for protonated pHB derivatized AGG enhanced by the apparent acidic hydroxy substituent at the para-position and least at the ortho position due to that apparent N-terminal hydrogen-bonding and resonance stabilization effects promoting N-terminal imine protonation. Likely the hydrogen-bonding and resonance stabilizing effects dominate for the methoxybenzyl series, resulting in the ortho < para < meta observed dissociation slope trend.
Scheme 5.2 (4-Iminomethyl) phenolate proton migration

Scheme 5.3 (2-Iminomethyl) phenolate proton migration termination cycle

The energy-resolved product ion formation profiles shown in Figure 5.7 represent the $y^i$ ion yields from CID-MS$^2$ of benzyl- and hydroxybenzyl imine AGG over the stepwise increase of the collision energy at 10 ms, 100 ms, and 1000 ms activation times. Figure 5.8 shows the $y^i$ profiles for the methoxybenzyl imine series. Product ion profiles were created by plotting product ion yield (Equation 5.2) as a function of collision energy: where $I_p$ is the product ion intensity and $I_{total}$ the total sum of ion intensities.
The position and shape of each $y^\dagger$ formation profile mirror the dissociation profile of the corresponding precursor ion, suggesting the energy and rate characteristics of the formation profiles are governed by the dissociation chemistry and structure of the precursor ion.\textsuperscript{195} Comparative analysis of $y^\dagger$ ion formation curves was therefore evaluated in terms of the ortho, para, and meta-structures of the hydroxybenzyl and methoxybenzyl imine AGG precursor ions. Specifically, $y^\dagger$ ion formation was evaluated using $y^\dagger$ product ion yield at 100% dissociation of the precursor ion known as $PY_{\text{max}}$ represented by the leveling off (tail) of the profile curve.

The $PY_{\text{max}}$ determined for the formation of $y^\dagger$ were observed at $0.3503 \pm 0.0150$ for dissociation of $[\text{Bz}=\text{AGG} + \text{H}]^+$, $0.4200 \pm 0.0255$ for $[\text{mHB}=\text{AGG} + \text{H}]^+$, $0.4279 \pm 0.0244$ for $[\text{pHB}=\text{AGG} + \text{H}]^+$, and $0.4505 \pm 0.01815$ for $[\text{oHB}=\text{AGG} + \text{H}]^+$. The order of increasing $PY_{\text{max}}$ (meta < para < ortho) is also observed for the methoxybenzyl imine series and is consistent with the stability order for the corresponding protonated imine-peptide precursor ions determined from $E_{50\%}$ values. Possibly, the same resonance and hydrogen-bonding effects that appear to influence the observed stability of the protonated N-terminal peptide-imine isomers also promote $y^\dagger$ product ion yield when the modifier structure likely cause selective protonation at the N-terminal imine nitrogen.
Figure 5.7 Collision energy resolved CID-MS² $y^i$ ion formation profiles for protonated Bz=AGG (yellow), pHB=AGG (red), mHB=AGG (blue), and oHB=AGG (black) at activation times a) $t = 10$ ms, b) $t = 100$ ms, c) $t = 1000$ ms.
Figure 5.8 Collision energy resolved CID-MS$^2$ y$^1$ ion formation profiles for protonated Bz=AGG (yellow), pMB=AGG (red), mMB=AGG (blue), and oMB=AGG (black) at activation times a) t = 10 ms, b) t = 100 ms, c) t = 1000 ms.
5.3 Conclusions

The effect imine modifier structure has on the general fragmentation pattern of protonated imine-modified peptides was probed by comparing the CID spectra of model peptide AGG, derivatized by condensation reaction with benzaldehyde (Bz), and ortho (o), para (p), and meta (m) versions of hydroxybenzaldehyde (HB) and methoxybenzaldehyde (MB). CID of [Bz=AGG + H]^+ (having no substituent on the benzyl ring) showed a less extensive series of a and b ions compared to the HB and MB versions. A notable difference observed when comparing the CID mass spectra of the ortho, para, and meta isomers of protonated HB=AGG and MB=AGG were the intensity of the b_2 and a_2 ions.

The y_ǂ ion peak was observed as the most intense peak (100%) for each HB and MB isomer, despite the disrupted hydrogen bonding interaction between the protonated imine nitrogen and the ortho-positioned substituent oxygen for the meta and para isomers. Presumably, eliminating the hydrogen-bonding interaction would reduce the competitiveness of N-terminal imine protonation implicated as the initiating step for an imidazole-one pathway for y_ǂ ion formation. Energy-resolved CID tandem mass spectrometry experiments were performed to further investigate the dissociation behavior of imine-derivatized peptides and the formation of the y_ǂ ion.

At an activation time of t =1000 ms, the energy-resolved CID profiles of the model HB and MB peptide-imines qualitatively revealed the effect the modifier structure has on the peptide precursor ion stability in terms fragmentation yield. Notably, the observed E_{50%} values revealed the energy requirement to fragment 50% of the precursor ion intensity followed a meta < para < ortho trend in increasing order, implicating the meta species as the least stable of the isomers.
Additionally, the dissociation slope values indicated the protonated ortho HB and MB peptide-imine species dissociate at the slowest rates amongst each isomeric series.

Resonance structures of the protonated HB and MB peptide-imine ions revealed stabilization of the cation at the protonated imine nitrogen likely plays a significant role in the rate of dissociation and overall stability of the previously determined intrinsic structure of protonated peptide-imines (Chapter 4). Specifically, resonance, N-terminal hydrogen bonding, and benzyl substituent acidity appear to influence the entropic mobilization of the ionizing proton to various active sites along the peptide backbone qualitatively indicated by the shape and position of the obtained energy-resolved dissociation curves. Comparison of energy-resolved $y^i$ ion formation profiles generated from CID of the protonated HB and MB peptide-imines indicated the formation profile curve shape appears to retain the CID dissociation profile shape of the corresponding precursor ion. Additionally, the observed max $y^i$ product ion yield for each protonated peptide-imine suggested favorable elimination of the N-terminal amino acid is governed by the structural characteristics which influence selective N-terminal imine protonation.

Consequently, this unambiguous observation could not have been explicitly made from direct interpretation of CID mass spectra alone, given each HB and MB isomer generated the $y^i$ ion at 100% relative intensity. Additionally, the unique shape and position of the energy-resolved dissociation curves allowed for isomeric identification within the HB and MB peptide-imine series, respectively, that would have been difficult from direct interpretation CID mass spectra. Undoubtedly, these experiments further implicated the importance of CID ion intensity data for evaluating the underlying dissociation chemistry of peptide ions and isomeric differentiation.
5.4 References


CHAPTER 6

INVESTIGATION OF EVEN-ELECTRON [M-H]^+ IONS GENERATED FROM CID OF METAL-CATIONIZED PEPTIDE-IMINES, [M+Ag]^+

6.1 Introduction

As mentioned, recent studies have demonstrated that peptide identification is negatively influenced by the macrocyclization of b ions which open at various positions resulting in the loss of sequence information, known as sequence scrambling.\textsuperscript{82, 206-213} In response to evidence macrocyclization is initiated by nucleophilic attack of the N-terminal amine upon the b ion oxazolone ring,\textsuperscript{29} N-terminal derivatization strategies have been investigated to diminish or eliminate the effects of sequence scrambling and enhance CID-MS\textsuperscript{n} approaches to direct peptide identification. Chapter 4 and Chapter 5 described an investigation into Schiff base, N-terminal amine to imine derivatization of singly protonated peptides which demonstrated elimination of scrambling and CID mass spectra rich in a and b type ions indicative of charge mediated pathways.

As mentioned, in contrast to protonated peptides, metal-cationized peptide ions have been proposed to fragment along predominately charge-remote pathways.\textsuperscript{92, 95-98} Fragmentation of metal-cationized peptide ions have been observed to commonly generate even-electron b\textsubscript{c}-H, a\textsubscript{c}-H, and y\textsubscript{c}-H ions indicative of hydride abstraction.\textsuperscript{87, 92, 94} As an extension to the Schiff base derivation method, the fragmentation pattern of metal-cationized peptide imines by argentination
were investigated to provide a comparison for the dissociation behavior of the protonated and cationized analogs of imine-derivated peptides.

6.2. Results and Discussion

6.2.1 CID of silver-cationized peptide-imines, [M+Ag]⁺

To probe for the sequence information produced by CID of metal-cationized peptides, the fragmentation patterns of the Ag⁺-cationized oMB=GGG (Figure 6.1a), oMB=VAAF (Figure 6.1b), and oMB=YGGFL (Figure 6.1c) ions ([M + Ag]⁺) were obtained. The dominate product ion peak created by CID of silver-cationized peptide imines was [b₂-H + Ag]⁺ at m/z 339 for oMB=GGG, [b₃-H + Ag]⁺ at m/z 466 for oMB=VAAF, and [b₄-H + Ag]⁺ at m/z 649 for oMB=YGGFL. Also observed were the [b₂+OH + Ag]⁺ and [a₂-H + Ag]⁺ ions at m/z 357 and 311, respectively, for oMB=GGG; the [b₃+OH + Ag]⁺ and [a₃-H + Ag]⁺ ions at m/z 484 and 438, respectively, for oMB=VAAF; and the [b₄+OH + Ag]⁺ and [a₄-H + Ag]⁺ ions at m/z 667 and 621, respectively, for oMB=YGGFL. These product ions are commonly observed during CID of metal-cationized peptides.⁹⁵, ¹⁵⁸, ¹⁵⁹

The novel elimination of the N-terminal amino acid observed for protonated peptide imines was not observed for silver-cationized oMB=VAAF nor silver-cationized oMB=YGGFL. It was unclear from CID of silver-cationized oMB=GGG whether N-terminal amino acid elimination was observed as the m/z 357 peak observed for silver-cationized oMB=GGG (Figure 6.1a) is the m/z value of both the [b₂ + OH + Ag]⁺ and the possible y¹ ion. However, CID of other silver-cationized Bz=AGG did not generate the m/z 327 product ion indicative of losing the
residue mass of alanine (71 u). Presumably, an N-terminal amino acid elimination pathway is not competitive for silver-cationized peptide-imines.

Note the conventional sequence ion nomenclature is used to define amide bond cleavage sites despite the modified N-terminus. Specifically, the \( b_x \)-H and \( a_x \)-H nomenclature is used to indicate a product ion of its type with one less hydrogen atom (extracted during fragmentation). However, note the distinction does not alter the classical mass of the neutral loss creating the specific sequence ions (e.g., an eliminated C-terminal glycine residue is still a loss of 75 u from the precursor ion mass). The \( b_x \)\(+OH\) nomenclature indicates the formation of a b type ion terminated by a carboxy group as described by Scheme 1.11, instead of an oxazolone ring.
Figure 6.1 CID spectra of $[\text{M} + {}^{107}\text{Ag}]^+$ ions where M is defined as a) oMB=GGG, b) oMB=VAAF, and c) oMB=YGGFL.
Interestingly, CID of each silver-cationized peptide imine generated a product ion peak consistent with eliminating the mass of $^{107}\text{AgH}$ (108 u), indicated by product ion m/z 306 for $[\text{oMB}=\text{GGG} + \text{Ag}]^+$, m/z 523 for $[\text{oMB}=\text{VAAF} + \text{Ag}]^+$, and m/z 672 for $[\text{oMB}=\text{YGGFL} + \text{Ag}]^+$. Evidence (not shown here) confirming the loss of 108 u as the elimination of silver hydride, was obtained through independent CID trails of the isolated $[\text{M} + ^{109}\text{Ag}]^+$ isotopic peak where the analogous loss of 110 u was observed indicating the elimination $^{109}\text{AgH}$. More interestingly the elimination of silver hydride implicated the resulting product ion as an even-electron imine-modified peptide cation ($[\text{M-H}]^+$), or more plainly an imine modified peptide ion without an ionizing proton. It is important to note, elimination of AgH has been reported for CID of Ag(I) complexes with 2-aminoanilide and benzimidazole,$^{214}$ and argentinated N-allyl benzamides.$^{215}$ However, the resulting cationized N-terminally modified peptide species resulting from the silver-hydride elimination appear to be a unique peptide ion type for gas-phase studies.

Logically, obtaining the fragmentation pattern of novel $[\text{M-H}]^+$ peptide ions would provide a way to compare the predominantly charge-mediated fragmentation of protonated N-terminally modified peptides to the peptide fragmentation of the same under hydrogen deficient ($[\text{M-H}]^+$) conditions. However, consistent with charge-mediated explanations of protonated peptide fragmentation,$^{26}$ the origin of the hydride abstraction that generates the $[\text{MH}]^+$ ion was obtained to better aide in interpreting the CID spectra of $[\text{M-H}]^+$ peptide ions. Additionally, the site(s) of hydride abstraction was determined to help construct a possible mechanism for the formation of the novel even-electron peptide cation.
6.2.2 Identifying the abstracted hydrogen involved in AgH elimination from silver-cationized peptide-imines

To probe the origin of the hydride eliminated with Ag\(^+\) or whether the hydride origin is distributed over multiple sites leading to the creation of the [M-H]\(^+\) peptide ion, CID of silver-cationized peptide imine analogs where the amide position and C-terminal hydrogens were exchanged for deuterium were examined (not shown). Upon CID of the [M + \(^{107}\)Ag]\(^+\) ions derived from D-labeled imine peptides, the loss of AgH was observed and not mass shifted AgD, signifying hydride elimination had not originated from the peptide backbone at an amide nitrogen nor the C-terminus. To investigate hydride loss from imine modifier, the CID spectrum of Ag\(^+\)-cationized unlabeled AGG derivatized by reaction with benzaldehyde ([Bz=AGG + \(^{107}\)Ag]\(^+\), Figure 6.2a)) and the CID spectrum of Ag\(^+\)-cationized unlabeled AGG derivatized by reaction with d\(_1\)-benzaldehyde – placing the deuterium atom on the benzyl carbon of the final metal-cationized peptide imine – were obtained ([d\(_1\)-Bz=AGG + \(^{107}\)Ag]\(^+\), Figure 6.2b).

Consistent with the types of product ions observed for CID of silver-cationized peptide imines investigated, the product ions observed for the dissociation of [Bz=AGG + Ag]\(^+\) included [b\(_2\)-H + Ag]\(^+\) at m/z 323, [a\(_2\)-H + Ag]\(^+\) at m/z 295, and [b\(_2\)+OH + Ag]\(^+\) at m/z 343 (Figure 6.2a). Expectedly, the analogous N-terminal containing product ions were upshifted by 1 u for the deuterium-labeled [d\(_1\)-Bz=AGG + \(^{107}\)Ag]\(^+\) species (Figure 6.2b), consistent with a deuterium in the place of hydrogen at the benzyl carbon. Likewise, the hydride elimination of 108 u (AgH) observed for dissociation of [Bz=AGG + Ag]\(^+\) was observed at 109 u (AgD) for [d\(_1\)-Bz=AGG + Ag]\(^+\), strongly identifying the benzyl carbon as the site of hydride elimination.
Figure 6.2 CID spectra of $[M + ^{107}\text{Ag}]^+$ ions where M is identified as of a) Bz=AGG and b) $d_1\text{Bz}=\text{AGG}$. 
Note, a decrease in ion intensity was observed for silver-deuterium elimination (AgD) from [d1-Bz=AGG + Ag]+ in Figure 6.2b when compared to AgH loss for unlabeled [Bz=AGG + Ag]+ in Figure 6.2a consistent with a kinetic isotope effect, particularly associated with a decrease in the [M-H]+ production rate as a result of the H to D substitution. Furthermore, loss of 108 u (AgH) was not observed in the fragmentation pattern of [d1-Bz=AGG + Ag]+, suggesting the hydride elimination is a selective mechanism not resultant of contributions from multiple sites, but explicitly abstracting the hydrogen atom from the benzyl carbon. Silver hydride elimination where the benzyl carbon hydrogen is the source of H-abstraction strongly suggests the generated [M-H]+ peptide ion is charged at the N-terminal nitrogen converting the imine modification into a nitrile.

Optimized structures of underivatized Ag+-cationized GGG derived computationally by Hopkinson, Siu, and coworkers using B3LYP/DZVP have suggested its most stable structure as a “multiring” conformation consisting of the silver cation tetracoordinated at the three carbonyl oxygens and the N-terminal nitrogen. A proposed mechanism for the loss of AgH from model silver-cationized Bz=AGG upon CID is shown by Scheme 6.1, similar to the proposed fragmentation pathways of argentinated unmodified peptides proposed by Hopkinson, Siu, and coworkers. The proposed mechanism is initiated by conversion of Bz=AGG tetracoordinated to silver at the carbonyl oxygens and the imine nitrogen, to a conformer dicoordinated to silver at the imine nitrogen and the second residue carbonyl oxygen. The breakdown of the dicoordinated intermediate by attack of the second residue carbonyl by a lone pair on the carbonyl oxygen of the C-terminus, results in the retention of Ag on the second residue carbonyl oxygen and an intermediate structure terminated by an oxazolone ring. Hydride extraction is then accomplished by N-terminal nitrile formation transferring the hydride to silver, breaking
the AgO bond to release AgH and collapse of C-terminal oxazolone ring to generate the [M-H]$^+$ even-electron ion.

Scheme 6.1 Proposed mechanism for AgH loss from Ag$^+$-cationized imine-modified peptides.
6.2.3 CID fragmentation pattern of even-electron [M-H]^+ ions

To investigate the general CID fragmentation pattern of [M-H]^+ ions created by the loss of AgH from silver-cationized peptide imines, the [M-H]^+ ions generated from silver-cationized oMB=GGG and oMB=AGG were obtained. The fragmentation patterns of protonated underivatized and oMB derivatized analogs were also obtained to probe for general differences in the generated sequence information each system provides. The CID spectra generated from the unmodified [M+H]^+, modified [M+H]^+, and [M-H]^+ versions of GGG and AGG are shown in Figure 6.3 and Figure 6.4, respectively. For both GGG and AGG, the dominant product ion generated from CID of the [M-H]^+ species was the b_2-H ion at m/z 231 for [oMB=GGG – H]^+ (Figure 6.3a) and m/z 245 for [oMB=AGG – H]^+ (Figure 6.4a). Generation of the a_2-H ion was also observed at m/z 203 and m/z 217 for CID of [oMB=GGG – H]^+ and [oMB=AGG – H]^+, respectively. Note the [M-H]^+ nomenclature where M is a peptide imine is not indicative of structure but rather the mass equal to the peptide imine mass minus 1.

Possible intramolecular H-transfer interactions that induce charge-mediated amide bond cleavages to form b, a, and y type ions under conditions with no ionizing (mobile) proton or with a fixed protonation site, have been previously reported by others.\textsuperscript{152,216} Presumably, the proposed N-terminal “fixed” charge at the imine to nitrile transformation and the absent ionizing proton suggests the b and a ions from CID of [oMB=GGG – H]^+ and [oMB=AGG – H]^+ may be generated via the conventional oxazolone pathway outlined by Bythell et al. for peptides lacking a conventional “mobile” proton.\textsuperscript{216} Particularly, the formation of the b_x-H and a_x-H ions from CID of the [M-H]^+ analogs suggested even without an ionizing proton, charge-mediated
pathways are competitive via intramolecular H-transfer reactions that generate the necessary protonated amide nitrogen configurations for amide bond cleavage.
Figure 6.3 CID spectra of a) the [M-H]+ species derived from AgH loss from [oMB=GGG + Ag]+, b) the [M+H]+ ion species derived from protonated oMB=GGG, and c) [GGG + H]+.
Figure 6.4 CID spectra of a) the [M-H]+ species derived from AgH loss from [oMB=AGG + Ag]+, b) the [M+H]+ ion species derived from protonated oMB=AGG, and c) [AGG + H]+.
A minor product ion species attributed to the loss of CO and H₂O (-46 u) was also observed for the CID of [oMB=GGG – H]+ and [oMB=AGG – H]+ at m/z 260 and m/z 274, respectively. Additionally, a minor product ion was observed from CID of [oMB=AGG – H]+ (Figure 6.4a) identified as the [b₂+OH]+ ion at m/z 263. Interestingly, the product ion m/z 249 indicative of losing the N-terminal residue mass of alanine (71 u) was observed for [oMB=AGG – H]+ at approximately 18% relative intensity, suggesting a yᵰ formation pathway is available for “hydrogen-deficient” [M-H]+ ions. It is unclear whether the m/z 249 product ion observed for CID of [oMB=GGG – H]+ (Figure 6.3a) is indicative of generating the [b₂+OH]+ and or the corresponding yᵰ ion (N-terminal Gly loss, 57 u).

For CID of protonated peptide imines ([M+H]+), the fragmentation pathway generating the yᵰ ion was identified as a charge directed pathway initiated by protonation at the N-terminal imine nitrogen (Scheme 4.2b). A similar pathway is proposed by Scheme 6.2 for yᵰ ion formation using the model [oMB=AGG – H]+ ion that can also be described as charge directed pathway, presuming a fixed charge at the N-terminal nitrile nitrogen. The proposed pathway begins with an attack of the nitrile carbon by the second residue amide nitrogen resulting in an imidazole-one intermediate similar to that proposed for [M+H]+ ions. Breakdown of the imidazole-one ring leads to the elimination of CO and formation of a benzylimino intermediate with an N-terminal a-type residue retained at the α carbon. Proton transfer leads to the elimination of the N-terminal amino acid a-type reside (HN=CHR, where R is the sidechain) and formation of the N-terminal nitrile yᵰ ion with retained peptide sequence.
Scheme 6.2 Proposed mechanism for $y^i$ formation from even electron $[M-H]^+$ ions.

Compared to the fragmentation pattern of $[M-H]^+$ ions derived from CID of Ag$^+$ cationized peptide imines, CID of the analogous $[M+H]^+$ ions generated from protonated oMB=GGG (Figure 6.3b) and oMB=AGG (Figure 6.4b) produced more sequence ions than the $[M-H]^+$ analogues including the $b_1$ and $a_1$ ions at m/z 176 and m/z 148 for $[oMB=GGG + H]^+$, respectively, and m/z 190 and 162 for $[oMB=AGG + H]$, respectively. The apparent absence of the $b_1$ and $a_1$ ions from the CID spectra of the hydrogen-deficient $[M-H]^+$ peptide ions are
consistent with decreased sequence ion formation observed for CID of peptides with highly basic residues that retain a charge and thus decrease proton mobility. Additionally, the relative intensity of the $y^1$ ion was observed less intense for N-terminal amino acid elimination from the oMB=GGG (Figure 6.3b), and oMB=AGG (Figure 6.3a) [M-H]$^+$ ion species, indicating the elimination pathway is less competitive for the even-electron version.

To further probe the general fragmentation behavior of N-terminal modified [M-H]$^+$ peptide ions, the CID spectra were obtained of [M-H]$^+$ ions created by silver hydride elimination from silver-cationized oMB=VAAF and oMB=YGGFL. The fragmentation patterns of protonated underivatized and oMB derivatized VAAF and YGGFL were also obtained for comparison. The CID spectra generated from the unmodified [M+H]$^+$, modified [M+H]$^+$, and [M-H]$^+$ versions of VAAF and YGGFL are shown in Figure 6.5 and Figure 6.6, respectively.
Figure 6.5 CID spectra of a) [M-H]+ derived from AgH loss from [oMB=VAAF + Ag]⁺, b) [M+H]+ derived from protonation of oMB=VAAF, and c) [VAAF + H]+.
Figure 6.6 CID spectra of a) [M-H]\(^+\) derived from AgH loss from [oMB=YGGFL + Ag]\(^+\), b) [M+H]\(^+\) derived from protonation of oMB=YGGFL, and c) [YGGFL + H]\(^+\).
For the CID spectrum of [oMB=VAAF – H]⁺ (Figure 6.5a), the most abundant product ion observed was the b₂-H ion, followed by the b₃-H ion observed at ~38% relative intensity. The a₂-H, a₃-H, b₃+OH, and yᵢ ion ions were also observed. The ion intensities of the b₂, a₂, and yᵢ ions generated from CID of [oMB=VAAF + H]⁺ (Figure 6.5b) were observed much greater than the analogous b₂-H, a₂-H, and yᵢ ions generated from CID of [oMB=VAAF – H]⁺. Presumably, the rate of production is greater for the b₂ and a₂ ions generated through classical charge-mediated pathways compared to the implicated intramolecular H-transfer pathways thought to generate sequences ions from the [M-H]⁺ precursors. Though missing from the CID spectrum [oMB=VAAF – H]⁺ and [oMB=VAAF + H]⁺, CID of unmodified [VAAF + H]⁺ (Figure 6.5c) generated the y₂ ion as the most abundant product ion peak. Assumedly the N-terminal imine or nitrile modifications, implicated to contain the ion charge, favor the formation of N-terminal containing product ions.

CID of [oMB=YGGFL – H]⁺ (Figure 6.6a) generated the b₂-H ion as the most abundant product ion peak, followed by the b₄-H peak at ~ 70% relative intensity. Interestingly, of the tri, tetra, and pentapeptides investigated, the b₂-H ion was observed as the most abundant peak for each [M-H]⁺ ion, though the mechanism behind the observation is unclear. Additionally, a very small peak less than 1% relative intensity was observed at m/z 509 indicative of losing the N-terminal amino acid residue mass of Tyr (163 u) from [oMB=YGGFL – H]⁺ to generate the yᵢ ion (Figure 6.6a). Though small, the relative intensity of the yᵢ ion for the [M-H]⁺ species of oMB=YGGFL is consistent with the small peak observed for the yᵢ ion generated from the [M+H]⁺ species. Presumably, the small intensity of the yᵢ ion is indicative of a peptide length effect on the competitiveness of the N-terminal amino acid elimination pathway, because more available sites for amide bond cleavages exist for longer sequences.
Consistent with the fragmentation patterns of the oMB modified GGG, AGG, and VAAF peptides, CID of \([\text{oMB}=\text{YGGFL}+\text{H}]^+\) (Figure 6.6b) generated more sequence information than CID of the [M-H]^+ species. Specifically, the \(a_1\) and \(b_1\) ions were observed for the [M+H]^+ oMB=YGGFL species and consistently missing from the CID spectra of all the investigated [M-H]^+ species. However, the fragmentation pattern of both the [M-H]^+ and [M+H]^+ species of oMB=YGGFL was observed to lack the sequence scrambled product ions – identified with an asterisk in the CID spectrum of protonated unmodified YGGFL (Figure 6.6c). The elimination of the sequence scrambled ions further implicated N-terminal modification, either by imine or nitrile derivatization, as an effective method for eliminating the macrocyclization reaction that leads to the scrambling of unmodified peptides.

6.3 Conclusion

In this study, the fragmentation pattern of Ag^+-cationized GGG, VAAF, and YGGFL imine derivatized by reaction with oMB was investigated. The product ions generated by CID of these [M+Ag]^+ ions included a series of \(b_x\) and \(a_x\) type ions in the form \(b_x-H\), \(a_x-H\), and \(b_x+OH\), consistent with the literature. Interesting a novel product ion was observed consistent with losing \(^{107}\text{AgH}\) to produce an apparent novel N-terminally modified even-electron peptide ion ([M-H]^
}). CID of isotopically labeled \([d_1-\text{Bz}=\text{AGG}+\text{Ag}]^+\) unambiguously revealed the hydrogen abstracted with the silver cation originated from the benzyl carbon, strongly suggesting the structure of the resulting even-electron peptide ion lacks a conventional ionizing “mobile” proton but includes a charged N-terminal nitrile group.

CID of these [M-H]^+ ions generated \(b_\alpha-H\), \(b_\alpha+OH\), and \(a_\alpha-H\) type sequence ions. However, the \(b_1-H\) and \(a_1-H\) ions were characteristically not observed in the CID mass spectra of the [M-
H]+ ions, while the corresponding analogs were observed in the CID mass spectra of the [M+H]+ peptide species. The reason remains unclear why CID of the [M+H] peptide imine species characteristically generated b1 and a1 ions, despite lacking the amide oxygen N-terminal to the b1 cleavage implicated in the b_x – y_z pathway (Scheme 1.1) for underivatized peptides. Evidence that a1 ions are generated through a b2 → a1 fragmentation pathway has been reported by others that suggests the a1 formation only occurs when the b2 ion oxazolone ring nitrogen bears the ionizing proton.217, 218, 219 Presumably, the b2-H ion generated from CID of N-terminally derivatized peptide [M-H]+ ion species has an oxazolone structure where the apparent N-terminal nitrile contains the ionizing charge, making breakdown to the a1-H ion unobservable. Similarly, it is unclear why the b2-H ion was generated as the most abundant product ion peak for each [M-H]+ peptide species investigated. DFT and IRMPD spectroscopy studies to identify the structure of the b2-H ion and the [M-H]+ precursor peptide ion would better reveal the underlying chemistry affecting the product ion distribution of even-electron peptide ions. However, the generation of a and b type sequence ions from CID of [M-H]+ peptide species, despite lacking a conventional ionizing “mobile” proton, suggests possible intramolecular H-transfer interactions mediate the creation of protonated amide nitrogen intermediates necessary for amide bond cleavage.

Interestingly, both the [M-H]+ and [M+H]+ versions of the investigated N-terminally modified peptides generated the novel y� ion resulting from N-terminal amino acid residue elimination. The formation of the y� ion was observed as less intense for [M-H] ions when compared to the [M+H] species. Currently, the mechanism for y� ion formation from [M-H]+ species is not defined. However, the fixed charge at the presumed N-terminal nitrile group may
direct a nucleophilic attack at the benzyl carbon much likely the imidazole-one pathway for eliminating the N-terminal amino acid residue from \([M+H]^+\) peptide imine species.

More interestingly, the sequence scrambled product ions observed for CID of protonated underivatized YGGFL were not observed for CID of the \([M-H]^+\) and \([M+H]^+\) oMB derivatized YGGFL species. These observations further implicate N-terminal amine modification, whether by imine or nitrile derivatization, as a successful method for eliminating the macrocyclization event leading to sequence scrambling. However, the sequence ion rich (b, a, and y type ions) CID mass spectra observed for \([M+H]^+\) peptide imines when compared to the CID mass spectra of the \([M-H]^+\) analogues reiterate the importance of evaluating techniques for enhancing de novo peptide sequencing by multiple criteria, including: a) eliminating sequence scrambling and b) generating sequence informative product ions.
6.4 References


CHAPTER 7

FUTURE DIRECTIONS

7.1 Introduction

Evaluation of novel peptide-imine fragmentation chemistry suggests that it may be useful for primary sequence identification, principally in peptidomics. CID-MS$^2$ of small singly protonated N-terminal imine-peptides 3 to 5 residues long showed increased generation of information rich a$_n$/b$_n$ sequence ions when compared to the unmodified species. Additionally, the elimination of deleterious sequence scrambling events was also observed, restoring the power of multiple tandem MS experiments for de novo peptide sequencing.

However, application of N-terminal imine derivatization for enhanced peptide sequencing may be best suited for use in focused data independent acquisition (DIA) approaches. Widely used database methods for peptide identification is largely representative of a data dependent acquisition approach (DDA). DDA describes detection of ionized analyte peptide ions during an MS$^1$ scan, followed by individual isolation of only the most abundant ions to generate CID-MS$^2$ spectra for interpretation by database searching software.$^{220}$ In contrast, DIA approaches essentially fragment all analyte peptide ions to collect CID-MS$^2$ spectra, resulting in a wealth of sequence information for a greater number of identified peptides than DDA approaches.$^{220}$

DIA has been suggested as a favored approach for the characterization of peptides presented on cell surfaces by major histocompatibility complex (MHC) immunoproteins.$^{220}$ However, the fragmentation observed during DIA may still result in complex MS$^2$ spectra
difficult to interpret through database or traditional de novo sequencing methods.\textsuperscript{220} Additionally, accurate identification and analyses of MHC molecules is essential due to the limited amount of sample often collected.\textsuperscript{220} Possibly, the \textit{a}_n/\textit{b}_n sequence ion specific and scrambling eliminating characteristics of N-terminal imine derivatization may improve the interpretability of MS$^2$ spectra generated during DIA for successful sequencing of MHC bound peptides. Undoubtedly, targeted systematic investigations of N-terminal imine-derivatized model peptides indicative of MHC immunopeptides and other non-tryptic peptides is an important next step.

### 7.2 MHC Immunopeptides

MHC proteins are an essential molecule of the immune system. Viral, tumor, and autoimmune epitopes have been identified and sequenced by ESI-MS approaches.\textsuperscript{221} The two major MHC immunoproteins include class I and class II, both of which adaptively function to present peptides on the surfaces of cells for immune system recognition of foreign substances.\textsuperscript{220-224} The presented peptides thus can elicit T cell responses ranging from the coordination and regulation of effector cells, to lysis and infected cell death, making accurate identification of MHC peptides essential for research in immunology.\textsuperscript{221-223}

Comparatively, class II MHC molecules are constrained to macrophages and lymphocytes, while class I MHC molecules encompass almost every cell in an organism.\textsuperscript{224} MHC class I and class II – bond peptides can also be characterized as being 8 to 11 and 14 to 18 amino acid residues in length, respectively.\textsuperscript{221, 223} Using these length attributes, a targeted evaluation of N-
terminal imine-derivatization for sequencing MHC class I and class II, and other longer non-tryptic peptides can be imagined.

In addition to length, evaluating the effectiveness of N-terminal imine-derivatization for enhanced sequencing of immunopeptides should include a systematic investigation of model peptide-imines with charged amino acid sidechains such as lysine (K) and arginine (R). These residues have been observed to not only generate +2 and +3 charge states but also effect the fragmentation chemistry of unmodified peptide ions, mediating bond cleavage through noncovalent interactions near K and R sites which favorable compete for protonation.\textsuperscript{26, 225-228} Assumedly, N-terminal imine-derivatization by reaction with o-methoxybenzaldehyde or salicylaldehyde may result in favored protonation at the N-terminal imine nitrogen due to the presumed resonance and hydrogen-bonding stability observed in the present work. Additionally, recent literature has shown guanidination effective in selectively transforming the amine at lysine sidechains to more basic guanidino groups while leaving the primary amine at the N-terminus intact.\textsuperscript{229} Evaluating the fragmentation pattern of lysine containing peptides modified both by guanidination and imine-derivatization may address the question of multiple lysine imine-derivatization sites. Continued investigation of the fragmentation chemistry of N-terminal imine-derivatized model MHC immunopeptides under these conditions will contribute to the gas-phase chemistry knowledge of these modified species but also push the techniques foreseeable use to a specialized and important area in peptidomics.
7.3 References


