GAS-PHASE REARRANGEMENT CHEMISTRY OF b_n PEPTIDE FRAGMENT IONS PROBED VIA SYNTHETIC STRATEGIES AND MASS SPECTROSCOPIC TECHNIQUES: EFFECT OF SIZE AND RESIDUES

By

MARCUS E. TIRADO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2012
Foremost to God, my beloved wife Delia, parents, brothers, friends and colleagues
ACKNOWLEDGMENTS

My time at the University of Florida has been marked and forged by a plethora of benevolent people. The utmost of thank you’s, I write and articulate to my advisor, Dr. Nicolas Polfer, for his guidance, wisdom, patience, and direction during my stint in his laboratory. I am indebted to you for all your encouragement, munificence and support in attending the national conferences. I also thank you for giving me the opportunity to travel to the FOM institute in The Netherlands to perform the IRMPD experiments, and to the University of Amsterdam for peptide synthesis. Traveling to Holland was a hallmark experience in my life and meant the world to my mother. As a researcher, I want to acknowledge my committee members, Drs. Nicole Horenstein, Ben Smith, Kenneth Sloan, and Weihong Tan for all their advice and helpful suggestions. I wish to express my gratitude to Dr. John Eyler who has been a great mentor and inspiration to me. I shall always cherish and remember the spectra we collected in Holland. Especially, the wonderful bike rides, dinners, and conversations we shared. These have been invaluable moments in my scientific and spiritual sojourn. In addition, I wish to thank Dr. David Powell for being a mentor to me as well. Thank you for allowing me to utilize your FTICR mass spectrometer. I learned so much performing all the SORI-CID & H/D exchange experiments upon that instrument. Thank you to Drs. Jos Oomens and Jeff Steill at FOM institute. Dr. Britta Redlich, I extend a heartfelt thank you to. You always did an excellent job coordinating and arranging the guesthouse us. I especially thank you for allowing me to share my flamenco guitar music and cultural cuisine with my family at the FOM. I also want to thank Dr. Jan van Maarseveen for welcoming me to perform peptide synthesis experiments in his lab. Jochem Rutters and all the labmates who shared coffee with me every day at 10 am. Thank you for teaching me
synthetic procedures and methodologies. (who can ever forget the pseudo dilutions?)
My colleagues in the Polfer group are acknowledged for the discussion and the suggestions on my research. Finally, I want to thank my family and friends, for their love and support through these years. They have always been very supportive for the decisions I have made. I also want to acknowledge my wife, Delia, for being an exceptional human being. You have taught me the meaning of love. Your love, patience, support and encouragement have brought “our” graduate career to fruition. I dedicate this work to my beloved wife. To my dearest mother, your numen shall always dwell within my spirit’s sentiments. I thank thee for being such a beautiful human being-the color of your voice’s laughter and song shall always resonant within me like a fundamental vibrational frequency. I shall always try my best to honor thee by emulating your teachings.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>9</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>10</td>
</tr>
<tr>
<td>LIST OF SCHEMES</td>
<td>14</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION</td>
<td>17</td>
</tr>
<tr>
<td>Proteomics</td>
<td>18</td>
</tr>
<tr>
<td>Mass Spectrometry</td>
<td>19</td>
</tr>
<tr>
<td>Soft Ionization Techniques</td>
<td>22</td>
</tr>
<tr>
<td>Electrospray Ionization Method</td>
<td>23</td>
</tr>
<tr>
<td>Tandem Mass Spectrometry</td>
<td>24</td>
</tr>
<tr>
<td>Nomenclature of Peptide Dissociation Products</td>
<td>25</td>
</tr>
<tr>
<td>CID</td>
<td>26</td>
</tr>
<tr>
<td>Mobile Proton Model</td>
<td>26</td>
</tr>
<tr>
<td>Structures of b Fragment Ions</td>
<td>27</td>
</tr>
<tr>
<td>Peptide Synthesis</td>
<td>29</td>
</tr>
<tr>
<td>FMOC Solid-Phase Peptide Synthesis</td>
<td>29</td>
</tr>
<tr>
<td>Synthesis of Macrocyclic Peptides</td>
<td>30</td>
</tr>
<tr>
<td>Perspective Overview</td>
<td>30</td>
</tr>
<tr>
<td>CHAPTER 2: EXPERIMENT TECHNIQUES</td>
<td>50</td>
</tr>
<tr>
<td>Fourier Transform Ion Cyclotron Resonance Mass Spectrometry</td>
<td>50</td>
</tr>
<tr>
<td>Natural Motion of Trapped Ions</td>
<td>50</td>
</tr>
<tr>
<td>Cyclotron Motion</td>
<td>50</td>
</tr>
<tr>
<td>Trapping Motion</td>
<td>52</td>
</tr>
<tr>
<td>Magnetron Motion</td>
<td>53</td>
</tr>
<tr>
<td>FTICR-MS Experiment</td>
<td>54</td>
</tr>
<tr>
<td>Fourier Transform Ion Cyclotron Resonance Instrumentation</td>
<td>55</td>
</tr>
<tr>
<td>Magnet</td>
<td>55</td>
</tr>
<tr>
<td>Vacuum System</td>
<td>56</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>FMOC protected 20 amino acids for Solid Phase Peptide Synthesis</td>
<td>49</td>
</tr>
<tr>
<td>2-1</td>
<td>Figures of merit for tunable lasers employed in IRMPD experiments</td>
<td>87</td>
</tr>
<tr>
<td>2-2</td>
<td>Evaluated gas-phase basicities and proton affinities</td>
<td>88</td>
</tr>
<tr>
<td>3-1</td>
<td>Position of arginine residue in ( b_n ) ions for arginine whose IRMPD was recorded and putative structure assignment</td>
<td>114</td>
</tr>
<tr>
<td>3-2</td>
<td>Photofragmentation channels for ( b_9/b_6 ) ions of arginated systems</td>
<td>115</td>
</tr>
<tr>
<td>4-1</td>
<td>Peptides synthesized and characteristic ( b ) fragments analyzed</td>
<td>137</td>
</tr>
<tr>
<td>4-2</td>
<td>Gradual red shift in Amide I/II bands with size increase</td>
<td>137</td>
</tr>
<tr>
<td>4-3</td>
<td>Assignment of product ions</td>
<td>138</td>
</tr>
<tr>
<td>5-1</td>
<td>HPLC retention times and masses of products from cyclization of QWFGLG.</td>
<td>175</td>
</tr>
<tr>
<td>5-2</td>
<td>Photofragmentation assignments</td>
<td>176</td>
</tr>
<tr>
<td>5-3</td>
<td>Theoretical IR band positions for (pyroglutamyl)-R and pyroglutamic acid-R</td>
<td>177</td>
</tr>
<tr>
<td>5-4</td>
<td>Comparison IRMPD spectra of (pyroglutamic)WFGLG, (pyroglutamyl)WFGLG, and cyclo (QWFGLG)</td>
<td>177</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Formation of peptide bond in dipeptide.</td>
<td>33</td>
</tr>
<tr>
<td>1-2</td>
<td>Reactive groups in amino acid.</td>
<td>34</td>
</tr>
<tr>
<td>1-3</td>
<td>Short hand notations and structures of twenty naturally occurring amino acids.</td>
<td>35</td>
</tr>
<tr>
<td>1-4</td>
<td>Venn diagram of grouping amino acids according to properties.</td>
<td>36</td>
</tr>
<tr>
<td>1-5</td>
<td>Chemical structure of an alpha amino acid. Fisher projection for L-amino acid serine and D-amino acid serine.</td>
<td>37</td>
</tr>
<tr>
<td>1-6</td>
<td>Snapshot of “Top-down” and “Bottom-up” proteomics approaches.</td>
<td>38</td>
</tr>
<tr>
<td>1-7</td>
<td>Edman degradation mechanism.</td>
<td>39</td>
</tr>
<tr>
<td>1-8</td>
<td>Typical proteomics workflow.</td>
<td>40</td>
</tr>
<tr>
<td>1-9</td>
<td>Schematic for proposed electrospray ionization process.</td>
<td>41</td>
</tr>
<tr>
<td>1-10</td>
<td>Typical tandem (MS/MS) experiment.</td>
<td>42</td>
</tr>
<tr>
<td>1-11</td>
<td>Schematic representation of scan modes employed in (MS/MS).</td>
<td>43</td>
</tr>
<tr>
<td>1-12</td>
<td>Roepstorff-Fohlmann-Biemann nomenclature for peptide fragment ions.</td>
<td>44</td>
</tr>
<tr>
<td>1-13</td>
<td>General SPPS procedure for linear synthesis.</td>
<td>45</td>
</tr>
<tr>
<td>1-14</td>
<td>Peptide synthesis flow chart.</td>
<td>46</td>
</tr>
<tr>
<td>1-15</td>
<td>Proposed b ion structures.</td>
<td>47</td>
</tr>
<tr>
<td>1-16</td>
<td>Proposed mechanism for sequence permutation/scrambling for b ions.</td>
<td>48</td>
</tr>
<tr>
<td>2-1</td>
<td>Ion cyclotron motion undergone by an ion in the presence of a magnetic field.</td>
<td>76</td>
</tr>
<tr>
<td>2-2</td>
<td>Conventional FTICR cell.</td>
<td>76</td>
</tr>
<tr>
<td>2-3</td>
<td>Ion motions which arise in ICR cell.</td>
<td>77</td>
</tr>
<tr>
<td>2-4</td>
<td>Pulsed sequences utilized in FTICR-MS experiment.</td>
<td>78</td>
</tr>
<tr>
<td>2-5</td>
<td>Uses of ion cyclotron excitation.</td>
<td>78</td>
</tr>
<tr>
<td>2-6</td>
<td>Illustration of the processing of raw data:</td>
<td>79</td>
</tr>
</tbody>
</table>
2-7 Schematic of commercial FTICR mass spectrometer (Bruker Daltonics) 4.7 Tesla magnet.

2-8 Various ICR cell geometries.

2-9 FTICR performance as a function of magnetic field strength.

2-10 Dependence of resolution on acquisition time or number of data points.

2-11 Schematic representation of IVR mechanism.

2-12 Schematic representation of undulator magnets.

2-13 Schematic representation of FELIX couples to FTICR.

2-14 Temporal evolution of ion cyclotron radius with varying ICR techniques.

2-15 Gas-phase H/D exchange mechanisms proposed by Beauchamp and coworkers.

3-1 Overlay of IRMPD spectra of QWFGLMPG b₆ and protonated cyclo(QWFGLM).

3-2 Overlay of IRMPD spectra of QPFGLMPG b₇, QPFGLMPG b₆ and protonated cyclo(QPFGLM).

3-3 Overlay of IRMPD spectra for b₆ fragments with different 4AMB substitutions in the sequence motif QWFGLM.

3-4 Reaction scheme showing impeded b₆ ion formation for QWFGL(4AMBz)PG.

3-5 Resonantly-stabilized π-electron oxazolone ring structure for QWFG(AMBz)MPG b₆.

3-6 Comparison of IRMPD spectra of (4AMBz)WFGLMPG b₆ (top) and (4ABz)WFGLMPG b₆ (bottom).

3-7 Comparison of SORI CID and IRMPD mass spectra for (4Abz)WFGLMPG b₆ fragment.

3-8 Fragmentation mass spectra for b₆ fragments with different 4AMBz substitutions in the sequence motif QWFGLM under SORI CID and IRMPD.

3-9 Overlay of IRMPD spectra for b₆ fragments with different Arginine substitutions in the sequence motif QWFGLM.

3-10 Overlay of IRMPD spectra for b₅ fragments with different Arginine substitutions in the sequence motif QWFGLM.
Proposed mechanism for locked oxazolone structure. ........................................... 113

IRMPD spectra of the b4 ions with the sequence motifs for TyrAlaGly, AlaTyrGly, and GlyAlaTyr. ............................................................................................................. 128

IRMPD labeled bands multiple b4 oxazolone structures. ........................................... 129

IRMPD spectra of the b6 ions with the sequence motifs for TyrAlaGly, AlaTyrGly, and GlyAlaTyr. ............................................................................................................. 130

Comparison of IRMPD spectra of cyclo(YAGYAG)H+, b6 with sequence motif YAG, and b6 from the N-terminally acetylated peptide .............................................. 131

IRMPD spectra of a series of b ions with the sequence motif GlyAlaTyr: b4, b6, b9, b12, and N-acetylated b14................................................................. 132

IRMPD spectrum for b15 from (GlyAlaTyr)3ProGly. ................................................. 133

Sori CID mass spectra of b6 ions with sequence motifs (AlaTyrGly)2, (GlyAlaTyr)2, and (TyrAlaGly)2................................................................. 134

“Geminal” vs. “fraternal” structural hypothesis for macrocyclization. .................... 135

IRMPD spectra for b5 from (AlaTyrGly)2ProGly & (TyrAlaGly)2ProGly ................. 136

HDX data for b5 ion of from (GlyAlaTyr)2ProGly. .................................................... 136

Chemical stick figures for various fragment ions of QWFGLG............................... 160

LCMS chromatogram and mass spectrum of linear peptide QWFGLG. ............. 163

LCMS chromatogram and mass spectrum of cyclization of QWFGLG. .............. 164

Coupling reagents employed to try to cyclize Glutamine from the chain. ............ 165

IRMPD spectra of m/z 689 ions fractions 1 and 2..................................................... 166

Comparison of IRMPD spectra of head-to-tail cyclic (QWFGLG) and HPLC fraction 1 .............................................................................................................. 167

Chemical structures for N-terminal pyroglutamyl-WFGLG and N-terminal pyroglutamic acid –WFGLG ................................................................. 168

Comparison of IRMPD spectra of (Pyroglutamic) WFGLG-OH and LC fraction 2 .............................................................................................................. 169

Appearance spectrum of mass channels for (cyclo) QWFGLG. ......................... 170

Chemical structure of BOC-pyroglutamic acid......................................................... 171
5-10 Chemical structure of (cyclo)KWFGGLG............................................................ 171

5-11 IRMPD spectra of (cyclo)KWFGGLG of head-to-tail and side cyclization of lysine. ........................................................................................................ 172

5-12 Structures of (pyroglutamyl)NHCH₃ and (pyroglutamic)NHCH₃ optimized...... 173

5-13 Predicted IRMPD spectra for (pyroglutamyl)NHCH₃ and (pyroglutamic)NHCH₃.............................................................................................................. 174

6-1 Proposed oxazolone formation from arene substitution patterns.................... 181
## LIST OF SCHEMES

<table>
<thead>
<tr>
<th>Schemes</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Schematic for formation of $b_6$ fragment, initially leading to oxazolone structure.</td>
<td>103</td>
</tr>
<tr>
<td>5-1</td>
<td>Glutamine’s pathways in both solution phase and gas phase.</td>
<td>157</td>
</tr>
<tr>
<td>5-2</td>
<td>Amino acid residue pathways leading to formation of pyroglutamic acid.</td>
<td>158</td>
</tr>
<tr>
<td>5-3</td>
<td>Dehydration and deamination channels of glutamine when situated at N-terminal position.</td>
<td>159</td>
</tr>
<tr>
<td>5-4</td>
<td>Proposed reaction pathways to rationalize dehydration from linear peptide QWFGLG, yielding m/z 689 product ion</td>
<td>160</td>
</tr>
<tr>
<td>5-5</td>
<td>Proposed reaction pathways of deamidation from linear peptide QWFGLG, yielding m/z 690 product ion</td>
<td>161</td>
</tr>
<tr>
<td>5-6</td>
<td>General mechanism for the formation of an amide bond formation using a carbodiimide.</td>
<td>162</td>
</tr>
</tbody>
</table>
Collision-induced dissociation (CID) of protonated peptides has become the hallmark analytical tool for identifying amino acid sequences and proteins in proteomics. Often, in mass spectrometric experiments, the dissociation chemistry is subject to the formation of linear “oxazolone” or “head-to-tail macrocycle” b fragments. The occurrence of head-to-tail macrocycles rationalizes sequence permutation processes in CID, which can complicate sequence analysis. The primary aim of this thesis is to determine how prevalent head-to-tail macrocyclization is, and what chemical trends govern this phenomenon.

Infrared multiple photon dissociation (IRMPD) spectroscopy was employed as a structural technique to identify the product ion structures based on diagnostic vibrations. In addition, head-to-tail macrocyclic peptides were made synthetically to serve as reference compounds. Systematic studies were performed on a number of peptide systems, which were all made in-house by solid-phase synthesis.

The results from the structural studies of b ions demonstrated that head-to-tail macrocyclization is highly prevalent, occurring over wide ranges of size, from mid-sized to larger b ions (b₆-b₁₂). In contrast, smaller b ions (e.g. b₄) were found to often favor
linear oxazolone structures. For mid-sized b ions, only a few factors were found to decrease the propensity for forming macrocyclic structures. For instance, the presence of the rigid residue proline in position 2 on the backbone had a marked effect on disfavoring head-to-tail cyclization. In addition, the presence of a basic arginine residue was found to suppress head-to-tail cyclization in some cases.

The IRMPD spectroscopy technique was also used to identify dehydration products from peptides made in solution, showing that head-to-tail and head-to-side-chain (i.e., pyroglutamyl) cyclizations occurred. These reaction products could be separated by high-performance liquid chromatography (HPLC), but could not be structurally verified by nuclear magnetic resonance (NMR), (presumably) due to the presence of other impurities. By further “purifying” these extracts by mass isolation, the products could be structurally confirmed by IRMPD spectroscopy. This proof-of-concept experiment demonstrates the potential usefulness for IRMPD spectroscopy to characterize analytes from complex mixtures, as the combined separation capability from HPLC and mass spectrometry allows isolation of a particular analyte, while IRMPD spectroscopy permits structural verification.
CHAPTER 1
INTRODUCTION

Amino acids play central roles both as building blocks of proteins and as intermediates for a wide range of biological processes. Amazingly, the 20 amino acids that are found within proteins of eukaryotic species exhibit a plethora of chemical versatility and diversity. Amino acids form proteins, biopolymers, through a nucleophilic attack by the amino group of an amino acid at the electrophilic carbonyl carbon of the carboxyl group of another amino acid. The resulting linkage between the amino acids is referred to as an amide or peptide bond, resulting in an elimination of water. (Figure 1-1). Figure 1-2 illustrates the various functional moieties and charge states, zwitterions, associated with the 20 standard amino acids. All amino acids have a central or alpha carbon, to which are bonded 4 groups:

- a hydrogen
- an amino group
- a carboxyl group
- a unique side chain, also known as a R-group

There are many different ways to represent the structure of a polypeptide or protein. Figure 1-3 conveys short hand notations for describing amino acids and illustrates the chemical side chains. Figure 1-4 shows a Venn diagram grouping amino acids. According to their chemical properties, Glycine is the simplest of all the amino acids. It should be noted that glycine lacks a side chain and subsequently lacks a chiral center. The rest of the natural compendious amino acids are chiral in nature, allowing for steroisomerism to exist. Figure 1-5 presents the various facets of information for depicting amino acid structures, such as the stereo representations and Fischer projection formulas for the enantiomers of serine. All amino acids found in proteins
occur in the L-configuration. D-amino acids are not naturally found in proteins are important in the structure and metabolism of bacteria. For example, D-glutamic acid and D-alanine are structural components of certain bacterial cell walls.¹

**Proteomics**

Proteins are key components of the physiological metabolic pathways of cells. In 1975, O’Farell and Klose simultaneously described the usage of two-dimensional polyacrylamide gel electrophoresis (2-DE).² This technique allowed for the separation of 1,100 protein components from Escherichia coli on a 2-D gel.³ The high-throughput capability of MS, coupled with 2-DE separations of proteins, has made proteomic analysis practical. The name ‘Proteomics’, coined by the Australian scientist Marc Wilkins, encompasses the study of the structures and functions of proteins.⁴ Proteomic information provides a foundation for investigating the functional components participating in various intra- and intercellular events.⁵ For example, proteomics affords researchers the opportunity to identify and study protein properties such as: expression level, post-translational modification, interactions *et cetera*.⁶

Edman degradation was the first direct analytical method employed for peptide/protein sequencing. Edman degradation, developed by Pehr Edman, is a method of sequencing amino acids in a peptide.⁷ Figure 1-7 depicts the Edman degradation reaction scheme. The reaction is comprised of an iterative series of steps. First, phenylisothiocyanate is reacted with uncharged terminal amino group under mildly alkaline conditions. This reaction results in the formation of a cyclical phenylthiocarbamoyl derivative. Secondly, this thiazolinone derivative of the terminal amino acid is cleaved via acidic conditions. Thirdly, isolation of the phenylthiohydantoin (PTH)-amino acid derivative is obtained by organic solvent extraction and acidification.
Lastly, identification is achieved chromatographically. This procedure can then be repeated again to identify the next amino acid. An advantage of the Edman degradation is that it only requires 10-100 pico-moles of peptide for the sequencing process. Edman degradation reaction is automated to expedite the process. An advantageous aspect of this iterative process is that the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues. However, the Edman degradation was a rather tedious and laborious process. It has the disadvantage that it requires a free amino terminus for the Edman degradation to work. For example, if the protein is acetylated at its amino terminus, this will impede the method. In addition, peptides being sequenced are limited to 60 residues, as the cyclical derivatization does not always go to completion. In addition, amino acid analysis following acid hydrolysis fails to detect this modification because of the rapid decomposition of the modified residues at low pH and upon high temperature conditions. However, one manner to circumvent this problem is to cleave large peptides into smaller peptides prior to proceeding with the reaction. This enzymatic approach utilizes enzymes called proteases. These proteases catalyze the hydrolytic cleavage of peptide bonds. For example, trypsin, a protease, converts proteins to peptides. Trypsin is commonly utilized to specifically cleave proteins on the carboxy-terminal side of arginine and lysine residues.

**Mass Spectrometry**

Mass spectrometry (MS) has become an essential analytical investigatory tool for proteomics research and drug discovery. In contrast to Edman degradation, mass spectrometry offers greater sensitivity, more abbreviated analysis time, and an enhanced ability to cope with complex mixtures. The researcher is enabled to identify
and to quantify known and unknown compounds by revealing their structural and chemical properties. Peptide/protein sequencing by mass spectrometry has become an indispensable technique in the identification of proteins. Two strategies are employed for MS characterization of proteins. In the “bottom-up” approach, purified proteins or complex protein mixtures are subjected to proteolytic cleavage, prior to MS analysis. The workflow of “bottom-up” proteomics presents a number of advantages for MS analysis, including efficient extraction of enzymatic digest peptides from gels, facile separation by liquid chromatography, efficient ionization by either electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI), and convenient gas-phase fragmentation for sequence confirmation. A limitation of this approach is the often limited protein sequence coverage, due to ion suppression and varying ionization efficiencies of peptides. In “top-down” proteomics, intact protein ions or large protein fragments are subjected to gas-phase fragmentation for MS analysis. The “top-down” methodology is limited to proteins of smaller molecular weights (<30 kDa) and results in mass spectra that are richer and more complex in regards to sequence information. Fourier transform ion cyclotron resonance (FTICR) mass spectrometer with its ultra-high mass resolution, mass accuracy, and sensitivity are essential to deconvolute the dense spectral information generated by the "top-down" approach. While proteomics research has greatly advanced because of FTICR-MS capability to identify proteins and protein-protein interactions based on accurate mass measurements, the majority of mass-spectrometry-based proteomics is carried out on lower-specification instruments, such as ion traps and time-of-flight instruments. Chapter 2 contains a section devoted to discussing FTICR-MS and how it has been utilized in all the research presented in this
dissertation. Figure 1-6 graphically differentiates “top-down” from “bottom-up” proteomics. Note that the “bottom-up” approach is more relevant in the context of this thesis, as the gas-phase fragmentation chemistry of peptides, under conditions of collision-induced dissociation (CID), is studied. Figure 1-8 displays a scheme of a typical proteomics workflow. There exists no universal method for preparing protein samples for MS analysis because of the intrinsic complex nature of the proteome. However, there exists a repertoire of analytical techniques for confronting these challenges. Protocols differ depending on the sample type, analysis method, and experiment. For example, preparing samples from a biological fluid involves a different set of procedures than preparing samples from cell culture or tissue. Many factors must be considered, including the source, type, physical properties, abundance, complexity, matrix effects and cellular location of the target proteins. Conventionally, the analyst separates the proteins of interest by 2-DE. Secondly, one then proceeds to visualize the proteins and excise the spots of interest. Sequence information may be garnered by employing LC-MS/MS or MALDI-MS/MS. Shotgun proteomics is an alternative approach employed for large-scale protein identification. First, proteins are digested thereby resulting in the generation of peptide sequences. Subsequently, a liquid chromatographic separation is performed upon the peptides. Next, peptide identification is obtained via tandem experiments. Lastly, a computer algorithm is utilized to perform the identification. An advantage of this technique lies in that it obviates the need to perform a 2-DE separation of the protein. Mass spectrometry (MS), has supplanted Edman degradation, because it is much more sensitive and can fragment the peptides in an abbreviated time frame, seconds, compared to inordinate
time scale of hours or days. Moreover, MS obviates the need for proteins or peptides to be purified to homogeneity. In addition, MS has no problem identifying blocked or otherwise modified proteins. In the last few years, further breathtaking technological advances in the field have established MS as the mainstay tool to study the primary structure of proteins. In order to possess a more vivid picture of protein structure and sequence, proteomics imposes that MS provides intact molecular ion information prior to fragmentation chemistry taking place. With the advent of soft ionization methods such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) area of protein sequencing has burgeoned.

**Soft Ionization Techniques**

In 2002, John Fenn and Koichi Tanaka were awarded part of the Nobel Prize in Chemistry for their work in the development of soft ionization techniques in mass spectrometry. In 1989, John Fenn revolutionized the field of mass spectrometry by demonstrating the utility of electrospray to ionize large bio-molecules for mass spectrometric detection. A forte of mass spectrometry is its capability to easily be coupled with other separation techniques. For example, HPLC/ESI-MS affords one the ability to analyze both small and large molecules with varying degrees of polarities in complex biological samples. Furthermore, the inception of matrix-assisted laser desorption ionization (MALDI), and its application to large bio-molecules further bolsters the significance and applicability in mass spectrometry. Other soft ionization processes include fast atom bombardment (FAB), (ESI), and (MALDI). Consequently, they generate a paucity of fragmentation in the ion source. This limits the structural information obtainable in a single-stage mass spectrum. Tandem mass spectrometry
addresses this issue by incorporating extra steps for ion mass isolation and activation/fragmentation.

**Electrospray Ionization Method**

Electrospray ionization (ESI) has become one of the most ubiquitous ionization techniques in mass spectrometry. It permits the mass spectrometric characterization and sequencing of peptides, proteins and other biopolymers of great importance to the biosciences. In addition to being a soft ionization technique, ESI is capable of transferring very large complexes into the gas phase, such as entire viruses.\(^{27-29}\) In contrast to MALDI, ESI generally gives rise to multiply charged ions. This ability brings very high molecular weight species to lower mass-to-charge (m/z) ranges, where the mass resolution and accuracy of most mass analyzers is maximized (e.g. quadrupoles, FTICR).\(^{30}\) The ESI process can be broken down in the following series of steps.\(^{(Figure 1-9)}\) The analyte solution is pushed through an ESI emitter needle (made e.g. of stainless steel) that is floated at defined electrical potential. As the solution exits the emitter needle, it is subjected to a high electric field, resulting in a Taylor cone, from which charged droplets are emitted. A coaxial flow of nitrogen nebulizing gas is often employed to assist stability of the droplet formation.\(^{31}\) A counter-flow of heated drying gas causes solvent evaporation from the droplet. When the charge density at the droplet surface exceeds the surface tension that holds the droplet together, the critical Rayleigh is reached, resulting in Coulombic explosion into smaller droplets. The process of solvent evaporation, droplet contraction and Coulombic explosions is repeated until the molecular adducts are released from the final droplet. There remains some controversy as to the final step in the desolvation process. In the ion evaporation model (IEM),\(^{32}\) field desorption of intact molecules is observed from droplets. Conversely, in
the charge residue model (CRM)\textsuperscript{33} the remaining solvent molecules are evaporated from the ions. Note that either positively charged (e.g. [M+H]\textsuperscript{+} and [M+Na]\textsuperscript{+} adducts), or negatively-charged ions (e.g. [M-H]-) are formed, based on the polarity of the electric field.\textsuperscript{31} ESI is the softest ionization technique and is well suited for plethora of analytical applications ranging from: bio-molecules, gas phase reactivity studies, organometallic compounds, non-covalent complexes, etcetera.\textsuperscript{31}

**Tandem Mass Spectrometry**

Tandem mass spectrometry or MS/MS is one of the most widely used techniques to improve specificity and analyze individual components in complex mixtures. Tandem mass spectrometry is required to obtain more structural information. Tandem experiments involve multiple stages of isolation and fragmentation of precursor ions. Typical MS/MS experiments involve mass selection of the ion of interest in the first MS stage and excitation of the ion followed by its dissociation and mass analysis of the resulting fragment ions in the second MS stage. Figure 1-10 shows a typical MS/MS experiment. Fragmentation of gas-phase ions in tandem mass spectrometry occurs between these different stages of mass analysis. There exist several methods to induce fragmentation of the precursor ions. For example, in-source fragmentation process, ions are generated in the source by entering the vacuum region where they are accelerated by applying voltages. This prompts collisions with other ion’s surrounding the species which can then produce sufficient energy to yield product ions.\textsuperscript{34} Nozzle-skimmer fragmentation refers to the purposeful induction of in-source fragmentation by increasing the nozzle-skimmer potential on usually electrospray based instruments.

Tandem mass spectrometers operate either in space or in time. Tandem MS in space involves the physical separation of the instrument components (e.g. QqQ or
QTOF). Tandem MS in time involves the use of ion traps (e.g. Penning or Paul). The fragmentation mass channels obtained by MS/MS depend on several parameters such as, amino acid composition, size, and charge. In addition, the excitation method of the mass spectrometer as well as the time scale of fragmentation and detection are crucial for determining the fragmentation pattern obtained. Several scan types can be used for recording the spectra of CID products. A schematic of scan modes employed in (MS/MS) experiments is shown in Fig 1-11. It corresponds to the precursor scan, product scan, neutral loss scan and reaction monitoring experiment. One of the disadvantages associated with the tandem MS process is the signal to noise (S/N) ratio. One sequentially dilutes the ion signal at the expense of the ion being fragmented. MS$^4$ and higher MS$^n$ have been successfully been carried out.

**Nomenclature of Peptide Dissociation Products**

When peptides are activated in mass spectrometers, bond cleavage can occur on the backbone. Roepstorff-Fohlmann-Biemann devised a systematic nomenclature for naming peptide fragment ions. In this method, the fragments ions are labeled consecutively from the original amino terminus a$_m$, b$_m$, and c$_m$, in which m represents the number of amino-acid residues that these ions contain. Complementarily, the z(n–m), y(n–m) and x(n–m) ions are numbered from the C-terminus, where (n–m) equals the number of residues that these ions contain. Note that n is the total number of residues in the peptide. Figure 1-12 depicts as an example of a tetra-peptide named according to this nomenclature. Peptides may be fragmented by either low energy or high energy pathways.

Several manners exist for fragmenting peptides in the gas phase such as: blackbody infrared radiative dissociation (BIRD), collision-induced dissociation
(CID),\textsuperscript{26,40} electron capture dissociation (ECD),\textsuperscript{41-44} electron transfer dissociation (ETD),\textsuperscript{45} infrared multiple photon dissociation (IRMPD),\textsuperscript{46,47} and surface-induced dissociation (SID).\textsuperscript{48} The purpose of MS-based peptide fragmentation is generally to derive the primary structure of the peptide and thus confirm its identity.

**CID**

Collision-induced dissociation (CID) is the most common method to fragment peptide ions, where the ions are excited by collisions with a target neutral gas, such as helium. These collisions cause a vibrational excitation of the peptide ions in analogy to 'heating' the ion. At elevated internal energies, bond dissociation transpires. In low-energy CID, fragmentation occurs commonly at amide bonds along the backbone. Sequence-informative b- and y-ions arise as a consequence of these backbone cleavages, which can be identified by tandem mass spectrometry experiments.\textsuperscript{35}

CID methods can be categorized broadly into either "slow" or "fast" regimes\textsuperscript{35} based on the rate of energy deposition and the timescale for mass detection. In ion traps, the ions are “slowly” excited by multiple collisions on millisecond timescales, followed by mass analysis. In QTOF instruments, the ions are activated with a more energetic collision in a collision cell, followed by rapid mass detection on a timescale of microseconds. It is clear that the method of ion activation also plays a role in the type of (rearrangement) chemistry that can take place.

**Mobile Proton Model**

The ‘mobile proton model’ accounts for some of the trends observed in the dissociation of protonated peptides.\textsuperscript{49,50} The fragmentation process in the gas phase is described by charge-remote \textsuperscript{51,52} and charge-directed mechanisms.\textsuperscript{49,53} Upon excitation, the ionizing proton becomes ‘mobile’, and is transferred from a basic site to less energetically
favored protonation sites, such as amide oxygen’s, nitrogen’s and side chain groups. Protonation of the amide nitrogen leads to weakening of the amide bond, which can lead to dissociation at this site. As the protons are mobile along the backbone, dissociation can occur at different positions along the peptide backbone. This does not signify that the dissociation is random. Some amino acids have an influence on the proton localization. In the “pathways in competition” (PIC) model, developed by Paizs, the different stages of peptide fragmentation pathways (PFPs) and their interaction are characterized computationally. PIC considers the dissociation of peptides to undergo three well-defined phases: (1) the pre-cleavage phase, (2) the (amide) bond-cleavage phase, and (3) the post cleavage phase. This involves calculating the potential energy surfaces (PES) and corresponding transition states at the density functional theory (DFT) level.

**Structures of b Fragment Ions**

Upon cleavage of an amide bond, complementary N-terminal b and C-terminal y fragments are formed. The chemical structures of b ions have been a long-standing question in mass spectrometry. Early Bronsted acid chemical ionization studies showed that protonated carboxylic acids eliminated H₂O to form stable acylium ions. However, protonated amino acids, did not form these stable α-amino-acylium ions. Rather than forming acylium structures, as shown in Figure 1-15, Harrison and co-workers suggested that oxazolone structures were formed via nucleophilic attack from a backbone carbonyl (structure shown in Figure 1-15B). Polfer et al. provided the first direct evidence for oxazolone structures. They recorded infrared multiple photon dissociation (IRMPD) spectra of b₄ ions generated from Leu-enkephalin (YGGFL). The spectra revealed a characteristic signature band which corresponded to an oxazolone
C=O stretch band, while the diagnostic acylium CO stretch mode was absent.\textsuperscript{65} A third possible involves a “head-to-tail” cyclization from the N-terminus, as shown in Figure 1-15. For the smallest b ions, b\textsubscript{2}, this structure is known as the diketopiperazine structure, as first proposed by Wesdemiotis and co-workers.\textsuperscript{66} For smaller b\textsubscript{2} ions, the evidence so far has shown that diketopiperazine structures are rarely formed.\textsuperscript{67-71} For larger b ions, however, there is increasing spectroscopic evidence suggesting that “head-to-tail” macrocycle structures (Figure 1-15D) are formed. This will be discussed in detail in this thesis. In CID, the peptides have substantial internal energies, resulting in possible rearrangement processes. As shown in Figure 1-16, it is thus possible that oxazolone structures isomerize to a “head-to-tail” macrocycle structure. A reopening of the macrocycle can lead to oxazolones with the original and/or permutated sequences, depending on where the macrocycle is opened. Sequential fragmentation of permutated oxazolones leads to non-direct/permutated/scrambled sequence ions, which could complicate sequence analysis of unknown peptides. This phenomenon is in principle problematic for MS-based peptide sequencing, because the algorithms of current software for peptide sequencing do not take this into consideration, in worst cases leading to false identification.\textsuperscript{72} There is mounting evidence now through systematic studies that as the length of the peptide sequence increases those b ions adopt macrocyclic structures in the gas phase. As a result, these systems are more prone to re-arrangement processes.\textsuperscript{70,73,74} In order to better investigate this phenomenon, it requires reference compounds as a comparison. Therefore, peptide synthesis of both linear and cyclic peptides becomes an indispensable technique to characterize these systems.
Peptide Synthesis

As discussed earlier, peptides are biopolymers composed of amino acids. During the course of peptide synthesis, unwanted side reaction can occur. For example, to synthesize the dipeptide Gly-Ala, one reacts two amino acids Gly and Ala together with a coupling reagent and would expect to obtain the dipeptide. However, beside the desired dipeptide, Gly-Ala, other by-products will result in the synthesis such as Ala-Gly, Ala-Ala, and Gly-Gly. In order to direct synthesis, it is paramount to employ protecting groups, based upon an orthogonal approach.

FMOC Solid-Phase Peptide Synthesis

Solid-phase peptide synthesis (SPPS), pioneered by Merrifield, revolutionized peptide synthesis. This monumental contribution to science resulted in Merrifield being awarded the Nobel Prize in 1984. Solid phase peptide synthesis consists of assembling amino acids from the C-terminal to the N-terminal direction. The alpha-carboxyl group is attached via an acid labile linker to a solid support, "resin" (Figure 1-13). The most common resins employed in SPPS are Wang and trityl resins. They are tethered to polystyrene beads. The great advantage of this anchoring of the peptide onto the resin beads permits the facile washing, extending, deprotecting, and cleaving of the peptide from the resin. In addition, these resins can easily be functionalized with the desired amino acid. The amino terminal end of the amino acid is protected by a base-labile FMOC (9-fluorenylemethoxycarbonyl) protecting group. This group is readily removed by 20% base, such as piperidine. The side chains are protected by acid-labile groups such as tertiary-butyl (tBu). After the first amino acid is loaded onto the resin, the FMOC group is removed using piperidine (i.e., deprotection step). A Kaiser test is then performed to confirm that all of the FMOC protecting groups are removed. The next
FMOC amino acid is then attached to the growing peptide by activation of its carboxyl group (i.e., coupling step). A Kaiser test confirms that complete coupling has occurred on all the free amines on the resin. The synthesis then proceeds through cycles of 1) deprotection of FMOC amino terminus groups and 2) coupling of the next amino acid residue, until the peptide is completely synthesized. The completely synthesized peptide is then cleaved from the resin and side chain protection groups are removed using 95% trifluoroacetic acid with various scavengers depending upon the sequence (i.e., cleavage step). Table 1-1 lists all 20 FMOC amino acids with their corresponding molecular weights. Figure 1-14 depicts this reiterative process for SPPS synthesis.

**Synthesis of Macro cyclic Peptides**

Macro cyclic peptides are cyclic polypeptides, whose amino and carboxyl termini are linked together via a peptide bond in a circular chain. Cyclic peptides often play important roles in biological processes, such as for instance gramicidin S, which is an antibiotic agent. This explains the interest for synthetic approaches in making cyclic peptides. Cyclization of linear peptides with side-chain protected in the solution phase is the most widely-used method. Synthesis of intramolecular head-to-tail cyclic peptides is normally performed in dilute conditions \((10^{-3}–10^{-4} \text{ M})\) to minimize the competing polymerization of linear peptides. Given the dilute solution conditions and competing reactions, cyclization reactions are typically slow and give low yield.

**Perspective Overview**

This dissertation describes a systematic study of the effect of primary structure on the formation of oxazolone/macrocycle structures under CID conditions. In Chapter 2, discusses the experimental techniques employed For performing these experiments, in specific, FTICR mass spectrometry, IRMPD spectroscopy, HDX, and SORI-CID are
discussed. In Chapter 3, a systematic study was performed to investigate if by impeding the torsional angle of a system would thwart the formation of macrocyclic structure. The residues proline and 4-aminomethylbenzoic acid (4AMBz) were chosen because of their intrinsic rigidity, in the expectation that limited torsional flexibility may impede “head-to-tail” macrocycle formation. In addition, the effect of charge sequestration by arginine has upon restricting macrocyclization was investigated for a series of b ions generated by CID. Chapter 4 investigates the size effect on the formation of the macrocycle structure. The sequence motif GlyAlaTyr, was chosen and the related motifs TyrAlaGly and AlaTyrGly, to study the formation of larger macrocycles in the gas phase. A series of peptides, including (GlyAlaTry)$_2$ProGly, (GlyAlaTry)$_3$ProGly, (GlyAlaTry)$_4$ProGly, (GlyAlaTry)$_5$ProGly, cyclo(TryAlaGlyTyrAlaGly), and other sequence variants, were made by solid-phase synthesis, as described previously. Chapter 5 sets out to compare the similarities of condensed phase vs. gas phase chemistry. A well documented phenomenon which readily occurs in the solution phase namely the deamination or dehydration of glutamine on the N-terminus was investigated to address the question if isobaric species would obscure correct mass spectroscopic identification. Furthermore, we set out to address what are the limitations of IRMPD in correctly characterizing isobaric species. In specific, the model systems studied for this research were cyclo (QWFGLG) and cyclo (KWFGLG). Also, calculations were performed to extract theoretical IR spectra and to provide a qualitative picture of the possible structures recorded. Lastly, Chapter 6 discusses the future work required to further probe and better understand the fundamental chemistry governing the re-arrangement chemistry of b ions in the gas phase. All experiments were conducted on FTICR
instrumentation where the peptide of interest was generated by ESI. IRMPD experiments were performed at the Free Electron Laser for Infrared eXperiments (FELIX) facility employing their magnificent tunable free electron laser. HDX experiments were performed in Dr. Dave Powell’s laboratory at the University of Florida. Computations were conducted at the High-Performance Computing (HPC) center at the University of Florida.
Figure 1-1. Formation of peptide bond in dipeptide.
Figure 1-2. Reactive groups in amino acid.
Figure 1-3. Short hand notations and structures of twenty naturally occurring amino acids. [Adapted with permission of http://www.vivo.colostate.edu/hbooks/molecules/aminoacids.html].
Figure 1-4. Venn diagram of grouping amino acids according to properties. [Adapted with permission from Livingstone & Barton, CABIOS, 9, 745-756, 1993].
Figure 1-5. Chemical structure of an alpha amino acid. Fisher projection for L-amino acid serine and D-amino acid serine.
Figure 1-6. Snapshot of “Top-down” and “Bottom-up” proteomics approaches. [Figure is adapted with permission from Wehr, T. LCGC North America 2006, 24].
Figure 1-7. Edman degradation mechanism.
Figure 1-8. Typical proteomics workflow. [Figure is adapted with permission from http://www.piercenet.com/browse.cfm?fldID=3B750AB6-5056-8A76-4E71-83B42F51BE91].
Figure 1-9. Schematic for proposed electrospray ionization process. [Figure is adapted http://www.chm.bris.ac.uk/ms/theory/esi-ionisation.html]
Figure 1-10. Typical tandem (MS/MS) experiment.
Figure 1-11. Schematic representation of scan modes employed in (MS/MS). [Figure is modified from Hoffmann, E. d.; Stroobant, V. Mass spectrometry: principles and applications; 2nd ed.; Wiley: Chichester ; New York, 2001].
Figure 1-12. Roepstorff-Fohlmann-Biemann nomenclature for peptide fragment ions.
Figure 1-13. General SPPS procedure for linear synthesis.

X = Temporary amino protecting group
Y = Permanent side-chain protecting group
A = Carboxy activating group
Figure 1-14. Peptide synthesis flow chart.
Figure 1-15. Proposed b ion structures.
Figure 1-16. Proposed mechanism for sequence permutation/scrambling for b ions. Formation of oxazolone structures is followed by isomerization to macrocyclic structure and re-opening at a different amide bond. [Figure is modified with permission from Harrison, A. G.; Young, A. B.; Bleiholder, B.; Suhai, S.; Paizs, B. J. Am. Chem. Soc. 2006, 128, 10364].
<table>
<thead>
<tr>
<th>Amino Acids</th>
<th>FMOC Protected</th>
<th>g/mole</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>FMOC-Ala-OH</td>
<td>311.30</td>
</tr>
<tr>
<td>C</td>
<td>FMOC-Cys(Trt)-OH</td>
<td>585.70</td>
</tr>
<tr>
<td>D</td>
<td>FMOC-(Asp(OtBu))-OH</td>
<td>411.50</td>
</tr>
<tr>
<td>E</td>
<td>FMOC-(OtBu)-OH</td>
<td>425.50</td>
</tr>
<tr>
<td>F</td>
<td>FMOC-Phe-OH</td>
<td>387.00</td>
</tr>
<tr>
<td>G</td>
<td>FMOC-Gly-OH</td>
<td>297.30</td>
</tr>
<tr>
<td>H</td>
<td>FMOC-His(Trt)-OH</td>
<td>619.70</td>
</tr>
<tr>
<td>I</td>
<td>FMOC-Ile-OH</td>
<td>353.40</td>
</tr>
<tr>
<td>K</td>
<td>FMOC-Lys(Boc)-OH</td>
<td>468.55</td>
</tr>
<tr>
<td>L</td>
<td>FMOC-Leu-OH</td>
<td>353.40</td>
</tr>
<tr>
<td>M</td>
<td>FMOC-Met-OH</td>
<td>371.50</td>
</tr>
<tr>
<td>N</td>
<td>FMOC-Asn(Trt)-OH</td>
<td>596.70</td>
</tr>
<tr>
<td>P</td>
<td>FMOC-Pro-OH</td>
<td>337.40</td>
</tr>
<tr>
<td>Q</td>
<td>FMOC-Gln(Trt)-OH</td>
<td>610.70</td>
</tr>
<tr>
<td>R</td>
<td>FMOC-Arg(Mtr)-OH</td>
<td>608.70</td>
</tr>
<tr>
<td>S</td>
<td>FMOC-Ser(tBu)-OH</td>
<td>383.40</td>
</tr>
<tr>
<td>T</td>
<td>FMOC-Thr(tBu)-OH</td>
<td>397.30</td>
</tr>
<tr>
<td>V</td>
<td>FMOC-Val-OH</td>
<td>339.40</td>
</tr>
<tr>
<td>W</td>
<td>FMOC-Trp(BOC)-OH</td>
<td>526.60</td>
</tr>
<tr>
<td>Y</td>
<td>FMOC-Tyr(tBu)-OH</td>
<td>466.00</td>
</tr>
</tbody>
</table>
CHAPTER 2
EXPERIMENTAL TECHNIQUES

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry

The inception of the Fourier transform approach by Comisarow and Marshall in the early 1970s\textsuperscript{76} to the ion cyclotron resonance (ICR) technique led to Fourier transform ion cyclotron resonance mass spectrometry\textsuperscript{77} (FTICR-MS), as we know it today. FTICR-MS provides the highest mass resolving power and mass accuracy of any mass spectrometric techniques. As a result, FTICR-MS is very well suited for analytical problems in petroleomics, metabolomics, and proteomics, which need to distinguish between a very large number of constituents\textsuperscript{78}. In addition to the ultra-high resolution and accuracy, the ultra-high vacuum operation and electromagnetic trapping of the ions provided by the FTICR. It further allows for more advanced experiments to be carried out, such as photodissociation with lasers and ion-molecule reactions. The following sections provide a background on ion motions in an FTICR cell, and critical components that comprise an FTICR mass spectrometer.

Natural Motion of Trapped Ions

Cyclotron Motion

Fundamentally, the behavior of ions in the Penning trap of an FTICR mass spectrometer is governed by the interaction and behavior of charged particles in the presence of a uniform homogenous magnetic field. When the ions pass into the magnetic field they are bent into a circular motion in a plane perpendicular to the magnetic field by the Lorentz force Equation 2-1. Figure 2-1 depicts the cyclotron motion that an ion undergoes in a magnetic field. This equation can be derived from first principles, applying Isaac Newton’s Second law of motion.
\[ F = \text{mass} \cdot \text{acceleration} = m \left( \frac{dq}{dt} \right) = qE + q \mathbf{v} \times \mathbf{B} \]  \hspace{1cm} (2-1)

Where \( m \) is the ion mass, \( q \) is the ion charge, \( v \) is the ion velocity, \( E \) is the electric field, and \( B \) is the magnetic field. In order for an ion to undergo periodic circular motion, the Lorentzian force must be equal to the ion's centripetal force. This phenomenon is depicted in Figure 2-1. Neglecting the presence of an electric field, one obtains the following mathematical Expression 2-2.

\[ qv_{xy}B = m \frac{v_{xy}^2}{r} \]  \hspace{1cm} (2-2)

Where \( v_{xy} \) is the velocity of the ion in the \( x-y \) plane is defined as (orthogonal to \( B \)). The angular acceleration \( (dv_{xy}/dt) \) is expressed as \( v_{xy}^2/r \). In addition, one may rearrange Equation 2-2 and obtain the radius of cyclotron motion.

\[ r_c = \frac{mv_{xy}}{qB} \]  \hspace{1cm} (2-3)

Several consequences can be associated with this ion cyclotron orbital radius expression. This equation can further be re-arranged to the following expression.

\[ \omega_c = \frac{qB}{m} \]  \hspace{1cm} (2-4)

Since \( \omega = 2\pi/t = 2\pi f \), an expression for the linear cyclotron frequency is given by:

\[ f_c = \frac{\omega_c}{2\pi} = \frac{qB}{2\pi m} \]  \hspace{1cm} (2-5)
The significance of this equation lies in the fact that ions with different m/q can be sorted based on their individual cyclotron frequencies. It should be noted that because $\omega_c$ is inversely proportional to m/q, the larger the mass-to-charge ratio, the lower the cyclotron frequency. For example, at 7 T ion of m/q=1000 will have a cyclotron frequency of 107.4 kHz.

\[ \frac{1.602 \times 10^{-19} \, C \times 7.00T}{2\pi \times 1000 \, u \times 1.661 \times 10^{-27} \, kg \, u^{-1}} = 107.4 \, kHz. \] (2-6)

Several corollaries may be drawn from the aforementioned equations. Firstly, one can determine the mass of ion by measuring an ion’s cyclotron frequency. Secondly, observed cyclotron frequencies are between the ranges of tens of kilohertz to megahertz. Lastly, cyclotron frequency is determined by certain physical parameters including the mass of the ion, the magnetic field, and the charge of the ion.

**Trapping Motion**

The FTICR mass spectrometer relies upon an electromagnetic ion trap or cell to confine ions. Figure 2-2 shows the radial (x-y dimension) confinement is accomplished by applying a magnetic field parallel to the z-axis of the ion trap. Axial restriction of the ions, along the magnetic field, is accomplished by applying DC potentials to the trapping electrodes of the ICR cell. Thus, a potential well is created that traps the ions, causing periodic ion motion similar to the classical harmonic oscillator. Furthermore, it can be shown that Equation 2-7 can be derived. This equation establishes that the force on the ion applied by the electric field would accelerate the ion along the z axis in a motion parallel to the magnetic field, which is uniform along the z-axis of the laboratory reference frame. An oscillating frequency similar to a simple
harmonic motion is produced in the z direction. The frequency can be expressed by the following equation.\(^{81}\)

\[
\omega_z = \sqrt{\frac{2qV_{trap}\alpha}{ma^2}}
\]

(2-7)

**Magnetron Motion**

A three dimensional ion trap is created by the presence of both magnetic and electric fields within the ICR cell. The electric field and the resulting axial harmonic motion reduces the cyclotron frequency and introduces a second radial motion called the magnetron motion that occurs at the magnetron frequency. Based upon this effect, one may further derive the following two natural rotational frequencies in Equation 2-8.

The first frequency corresponds to the “reduced” cyclotron frequency, \(\omega_+\), also corresponds to the “observed” cyclotron frequency. The second frequency Equation 2-8, \(\omega_-\) is the magnetron frequency that belongs to a circular motion superimposed on the cyclotron motion. The term \(\omega_c = qB/m\) pertains to the unperturbed cyclotron frequency.

\[
\omega_+ = \frac{\omega_c}{2} + \sqrt{\left(\frac{\omega_c}{2}\right)^2 - \frac{\omega_z^2}{2}}
\]

\[
\omega_- = \frac{\omega_c}{2} - \sqrt{\left(\frac{\omega_c}{2}\right)^2 - \frac{\omega_z^2}{2}}
\]

(2-8)

Typically, in ICR experiments, the relation is such that \(\omega_+ >> \omega_z >> \omega_-\). The three natural ion motions are depicted in Figure 2-3.
FTICR-MS Experiment

A great advantage of the ICR cell is that ions can be stored for extended periods of time (seconds–hours), during which multiple experiments can be conducted on the ions. A typical event sequence is shown in Figure 2-4. A simple FTICR experiment is comprised of the following series of events: quenching, delays, ionization, excitation, and detection. In most modern FTICR mass spectrometers, ions are produced externally to the magnetic field (e.g. by ESI). The ions need to be transferred into the magnetic field, and pass through a series of pumping stages to increasingly high vacuum. The ions are guided by multi-poles and electrostatic optics. Once in the cell, ions are confined in the radial direction by the application of a magnetic field. Trapping in the direction of the magnetic field is amenable to either positive or negative ions by adjusting the polarity of the voltage applied to the trapping electrodes. In order to measure ion signal as an image current, the ion packets must be excited to larger radii as coherent ion packets. The frequency of this “image current” can be detected, amplified, digitized, and related to the m/z value of the ions in the cell. In order to minimize collisions, and thus de-phasing of the ion packet, the pressures within the cell is kept at an ultra-high vacuum (< 10^{-9} mBar). Excitation of the ions to larger radii is achieved by applying a dipolar excitation on the excitation plates of the same frequency as the cyclotron frequency. This increases the radius of motion and generates a coherently moving packet of ions. This process is shown in Figure 2-5. In order to excite a wide range of m/z’s to larger radii, a broadband or chirp excitation is applied. The image current detected from a multitude of ions represents an ensemble of different cyclotron frequencies. A Fourier transformation deconvolutes this complex time-domain
spectrum into the frequency spectrum that it is composed of. Subsequently, one converts the frequency domain spectrum into a m/z spectrum, as depicted in Figure 2-6.

Dipolar excitation of the ions allows other detailed manipulations of the ions prior to sequential events. If particular ions are excited beyond the volume of the cell, these will be ejected, resulting in mass isolation of the remaining ions. Similarly, if an ion is excited at a frequency slightly off-set of from its cyclotron frequency, it will undergo periodic acceleration and deceleration cycles. Simultaneous pulsing of a collision gas into the vacuum chamber results in collisional excitation of the accelerated ions, potentially leading to dissociation. The latter scheme is referred to sustained off-resonance irradiation collision-induced dissociation (SORI-CID), and shall be further elaborated upon. Both structural and mechanistic insights governing peptide fragmentation pathways may be gained by SORI-CID.

**Fourier Transform Ion Cyclotron Resonance Instrumentation**

Four general components, a magnet, an analyzer cell, an ultra-high vacuum, and a data system, comprise the FTICR instrument. A schematic diagram of the commercial FTICR mass spectrometer (Bruker Daltonics, Billerica, MA) equipped with a 4.7 T superconducting magnet that is employed in this research is shown in Figure 2-7.

**Magnet**

A magnet is required to create a strong magnetic field in order to induce the cyclotron motion in the ICR cell, which is situated inside the bore of the magnet. The cell is located inside a spatial uniform static superconducting high field magnet (typically 4.7, 7, 9.4, or 12 Tesla.) It should be noted that the National High Magnetic Field Laboratory (NHMFL) in Tallahassee, Florida, a 21T magnet is in construction. The coils of a superconducting magnet are cooled by liquid helium, which is enclosed by another
jacket cooled by liquid nitrogen. Figure 2-8 provides a survey of the various cells designs introduced over the years. FTICR–MS performance increases with magnetic field strength. Figure 2-9 depicts the correlation of magnetic field strength and resolving power. Superconducting magnets for FTICR-MS applications now routinely achieve 9.4 and 12 T. It should be noted that the National High Magnetic Field Laboratory (NHMFL) is currently working on the next-generation 21 Tesla ICR superconducting magnet. This represents a 45 percent increase in magnetic field over the current world-record 14.5 Tesla Fourier transform ICR instrument at the NHMFL. Disadvantages with high-field magnets are their inherent cost, and the cost associated with maintaining cryogens, particularly helium.

**Vacuum System**

Low pressures are a requisite for optimal operation of an FTICR instrument. An ultra-high vacuum (UHV) (10^{-8} to 10^{-10} Torr) regime minimizes collisions of the ions with background molecules in the cell. With external ion sources, differential pumping using turbo-molecular pumps allow bridging of high-pressure ions sources, such as ESI, with an FTICR instrument. Multipoles and ion optics are incorporated to guide ions from a higher to a lower pressure region. In order to leak particular gases into the UHV chamber, FTICR instruments are often equipped with pulsed solenoid valves or leak valves. A pulsed valve was employed in this research to momentarily introduce a collision gas (e.g. N_{2}) to carry out SORI-CID experiments. Moreover, a Varian leak valve, equipped with a sapphire crystal, was used to introduce CH_{3}OD vapor into the ICR cell to do H/D exchange with the trapped ions of interest.
Analyzer Cell

Many events can be performed in an ICR cell, including storage, mass-selection, activation, excitation, and detection. FTICR-MS experiments are based on ion confinement in a Penning trap, namely, a spatially homogeneous static magnetic field $B$, and an electrostatic trapping potential. Over the years, various cell designs have been introduced. Cylindrical cell geometries are preferred because they are more amenable with cylindrical magnet bores. Moreover, cylindrical cells have increased image ion detection over cubic cells. The ICR cell typically consists of six electrode plates. Two of these electrodes are positioned perpendicular to the direction of the magnetic, and the other four electrodes are arranged parallel with the magnetic field lines. The trapping plates restrict the ions in a direction parallel to the magnetic field. The excitation plates excite trapped ions towards larger radii. The detection plates serve as a detection platform for the image current from ions which have been excited to larger radii.

Data System

FTICR instruments require a sophisticated data system. By and large, this is attributed to the demand for speed of data collection as well analysis. In addition, the sizes of datasets generated are often larger, thereby taxing the computer memory. Some of the components of an FTICR system include the following: frequency synthesizer, delay pulse generator, broadband r.f., amplifier and preamplifier, a fast transient digitizer and a computer to coordinate all of the electronic devices during the acquisition of data, as well as to process and analyze the data. In addition, tandem mass spectrometry experiments can be performed simply by editing the experimental sequence.
Mass Resolution

FTICR-MS instruments are distinguished for their phenomenal high resolution which they can achieve when performing mass analysis. For ions of m/z < 5000, FTICR offers the highest mass resolving power and highest mass accuracy of all the mass spectrometers. Mass resolution may also be defined as the ability of separating two narrow mass spectral peaks. Mass resolution \( (m_2 - m_1 \geq \Delta m_{50\%}) \) where \( m_1 \) and \( m_2 \) are the closest masses that can be resolved) is defined as the point where one valley begins to appear between peaks of equal shape and height and is separated by \( \Delta m_{50\%} \). Commercial TOF instruments have a resolving power of \( 10^4 \). FTICR mass spectrometry can routinely achieve resolving powers of hundreds of thousands in broadband mode operation. This facilitates resolving the components in a complex mixture. In principle, it becomes feasible to discern various components in a mixture without chromatographic separation. However, coupling liquid or gas chromatography to an FTICR-MS has proven to be synergistic. One of the great advantages which ICR possesses over magnetic sectors, linear quadrupoles, or TOF’s is that the cyclotron frequency is affected less by ion kinetic energy and position during measurement. Compared to quadrupole ion traps, the ICR frequency is relatively more stable. This is attributed to the fact that the magnetic field of a superconducting magnet is more stable than the magnitude of the radio frequency voltage applied to the ion trap. The following example drawn upon from the literature serves as an excellent example: A fullerene derivative that contains nitrogen can be identified with a 9.4T FTICR mass spectrometer, as \(^{12}\text{C}_{59}\text{N}^+\) (theoretical m/z 722.002525) and \(^{12}\text{C}_{58}\text{^{13}C}_2^+\) (theoretical m/z 722.006161). Without the usage of high-resolution, one cannot confirm for the
presence of $^{12}\text{C}_{59}^{14}\text{N}^+$. The analyst would assume that only one signal was present and therefore assign it as being $^{12}\text{C}_{58}^{13}\text{C}_2^+$. The following equation defines resolving power:

$$ R = \frac{m}{\Delta m} $$  \hspace{1cm} (2-9)

$m$ is the $m/z$, $\Delta m$ is the width of the peak, w.r.t. $m/z$.

Long time domain signals, transients, are required in order to obtain high mass resolution. The number of data points collected during the experiment is set by the user before the transient is collected. In FTICR-MS, mass spectra are recorded as a finite number of data points which can be selected prior to acquisition. The number of data points available in the time-domain spectrum is obtained at a rate determined by the sampling frequency used. The Nyquist theorem states that the sampling frequency must be at least twice that of the highest frequency being recorded. Consequently, acquisition time is determined by the number of data points, and the sampling frequency determined by the low $m/z$ cut-off. The acquisition time may be expressed by the following formula:

$$ T_{\text{acq}} = \frac{N}{S} $$  \hspace{1cm} (2-10)

$T_{\text{acq}}$ is the acquisition time in seconds, $N$ is the dataset size, and $S$ is the sampling frequency (determined by the low $m/z$ cutoff) in Hz. Figure 2-10 shows the dependence of resolution upon acquisition time or number of data points. The attainable resolution in FTICR-MS is limited by the cyclotron frequency of the ion and the length of the time-domain signal that is recorded. Equation 2-11 clearly outlines this relationship:
\[ R = \frac{f_c T}{2} \]

In Equation 2-11, \( R \) is the resolving power, \( f_c \) is the cyclotron frequency and \( T \) is the duration of the transient. Both resolution and transient duration are directly proportional. As the transient time increases, so too does resolution increase. However, there exists an intrinsic property of the FTICR cell which hampers the achievement of higher mass resolution. Just as in all Penning traps\(^{84}\), the ICR cell requires an electric field to axially confine the ions inside the trapping cell during excitation and detection. As a result of axial confinement of the ions, a magnetron frequency is observed, that has a deleterious effect upon the true cyclotron frequency of the ions of interest. These frequency variations result in dephasing of the ion packet which translates into a decreased transient signal that ultimately leads to signal decay and decreased resolution. Great effort has been expended in recent years to offset the deleterious consequences associated with the magnetron motion. The pioneering work by Gabrielse introduced the electrically compensated ICR cell, that relies on approximating to a pure quadrupolar electric field using a segmented cell design\(^{85}\). However, this approach only allows for the ions to be in the center of the trap. Recently, Dr. Nikolaev presented a novel FTICR cell design, that makes it possible to achieve dynamic harmonization of the electric field in the entire volume of the cylindrical Penning trap\(^{86,87}\). As a result, ultra-high resolution of 24,000,000 is reported for reserpine. Clearly, this impressive work not only expands the “armamentarium” to mass spectroscopists, but
opens a new realm of understanding in proteomics, petroleomics, and other areas of science.

**Space Charge Effects**

Resolution of FTICR-MS is adversely affected by space charge effects. Space charge arises from the influence of the electric field of ions in the trapped analyzer cell upon each other. This has been quantified in the equation for the observed frequency of motion. Jeffries and co-workers developed the theory of space charge-induced frequency shifts in Penning cells with different geometries. Space charge effects can also limit mass accuracy measurements in FTICR mass spectrometry. When a large number of different ion packets are present the cell this causes Coulombic interactions between ions. The space charge-induced frequency shift is a systematic error and can be treated through proper calibration.

\[
\omega_\pm = \frac{\omega_c}{2} \left\{ 1 \pm \left[ 1 - 4 \left( \frac{2qV_T G_T}{m} + \frac{\rho q^2 G_i}{\varepsilon_0 m} \right) / \omega_c^2 \right]^{1/2} \right\}
\]

(2-12)

Angular frequency, \(\omega_c\) is the cyclotron frequency, \(q\) is the ions charge, \(m\) is the ion mass, \(V_T\) is the trapping voltage, \(G_T\) and \(G_i\) are geometry factor of the ICR cell, and \(\rho\) is the charge density. An approximated mass calibration can then be derived at the point \(m=0, V_T=0,\) and \(\rho =0\), and Equation above reduces to

\[
\omega_{\text{obs}} = \frac{qB}{m} - \frac{2\alpha V}{a^2 B} - \frac{q\rho G_i}{\varepsilon_0 B}
\]

(2-13)
Terms of the Equation: $q$ is the coulombic charge on the ion, $B$ is the strength of the magnetic field, and $m$ is the mass of the ion, $a$ is the diameter of the cell, $\alpha$ is a constant determined by the cell geometry, and $V$ is the voltage applied to the trap plates, where $\rho$ is the ion density, $G_i$ is a constant that is related to the geometry of the ion cloud, and $\varepsilon_0$ is the permittivity. Thus, the frequency difference between two ions is:

$$\omega'_{\text{eff}} - \omega''_{\text{eff}} = qB(1/m' - 1/m'')$$

(2-14)

Since the magnetic field $B$ is known, the mass of an unknown ion can be determined measuring the values of $\omega'_{\text{eff}}$ and $\omega''_{\text{eff}}$.

**Conclusion**

FTICR-MS has burgeoned into an indispensable analytical tool for bioanalytical studies, including proteomics, glycomics, and natural products. The advent of soft ionization techniques, ESI & MALDI, for bio-molecules and the integration of tandem, MS/MS methodology has placed FTICR-MS as being a premier research tool. FTICR MS offers higher mass resolving power ($10^6$) and mass accuracy (< 0.5ppm) than any other mass spectrometer, which makes it very useful for unequivocal mass assignment. In seeking even higher performance, the most obvious direction seems to be even higher magnetic fields. FTICR instruments are now available with superconducting magnets up to 12 T, and the trend is to continue increasing the field strength.

**Infrared Multiple Photon Dissociation (IRMPD) Spectroscopy**

Vibrational motion may be depicted as position changes of atoms within the molecule. A classical expression of molecular vibrational motion pertains to two atoms
attached and undergoing harmonic oscillation. Equation 2-15 represents the diatomic oscillation frequency ($v_{freq}$) as

$$v_{freq} = \frac{1}{2\pi} \sqrt{\frac{k_f}{\mu}}$$

(2-15)

Where $k_f$ is the bond force constant, $\mu$ the reduced mass of two atoms: one with the mass $m_1$, the other with mass $m_2$ such that

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

(2-16)

The low densities of ions in ion traps ($< 10^8 \text{ cm}^{-3}$) make photon absorbance measurement an arduous task. Action spectroscopy addresses this challenge. Infrared multiple photon dissociation (IRMPD) spectroscopy is a form of action spectroscopy, where one monitors the fragmentation of the precursor ion and the emergence of photofragments. The subsequent dissociation leads to an ion’s change in mass-to-charge ratio. This change is detected by a mass spectrometer. The photodissociation yield is maximized at particular irradiation wavelengths that are resonant with a vibrational mode of the molecule. In addition to producing photofragment information deriving from the fragments precursor ion, wavelength-specific information of chemical moieties in the precursor ion is obtained. The vibrational frequencies, IR absorptions, involving functional groups such as CO, NH, and OH groups, are generally very sensitive to hydrogen bonding. As a result, these vibrational modes provide detailed structural information. For a detailed structural analysis of the results, a comparison to
theoretical linear absorption spectra computed by quantum chemical methods (e.g. density functional theory) is often necessary.

As infrared photons are not very energetic (0.001-1.7ev), absorption of multiple (i.e., tens to hundreds) IR photons are needed to induce dissociation in a molecule. This requires usage of a light source with a high photon flux, such as a laser. In the late 1970s, Beauchamp and co-workers demonstrated that trapped ions in an ICR cell could be dissociated with a low-power CO\textsubscript{2} laser in an ICR trap.\textsuperscript{91,92} The CO\textsubscript{2} laser’s narrow tuning range (9.2–10.8 mm), which is limited to the ro-vibrational states of CO\textsubscript{2}, constrains the amount of vibrational information that can be gleaned. Recent developments of other tunable light sources, such as free electron lasers (FELs) and optical parametric oscillators/amplifiers (OPO/As), have given rise to a renaissance of this technique.

**Infrared Multiple Photon Dissociation Mechanism**

Infrared light is absorbed through the interaction of molecule’s transition dipole with the oscillating electric field of light. The radiant energy is converted to vibrational energy in the ion. The quantized nature of energy stipulates that the transition between energy levels must match the energy of the photon absorbed. Each of the evenly spaced levels correspond to the quantized vibrational energy levels with the ground state labeled by \(v_i=0\). Towards to the bottom of this well, the potential-energy is considered to behave similarly to that of a harmonic oscillator, resulting in energy levels that are distributed evenly. However, the potential energy curve becomes less steep at higher vibrational states, leading to a reduction in the energy gap between each state. If the energy was absorbed in a ladder-climbing process (i.e., \(v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow \ldots\))\textsuperscript{89} an “anharmonic” bottleneck would be reached, where the laser frequency is no longer in
resonance with the normal mode of the molecule. Consequently, anharmonicity will restrict the absorption of additional photons by the molecule. The process of intramolecular vibrational redistribution (IVR) circumvents this bottleneck. Upon absorption, IVR the energy is quickly dissipated into the bath of vibrational degrees of freedom of the molecule. Thus, photon absorption is followed by statistical energy redistribution over all vibrational degrees of freedom. Since the ground vibrational state v0 is recycled, a subsequent photon can be absorbed at the same (fundamental) transition. This process is repeated until the energy stored in the bath of vibrational background states exceeds the dissociation limit of the molecule. As a result, unimolecular dissociation or multiple dissociation channels are exhibited by the molecule. Figure 2-11 depicts the IVR mechanism.

An FTICR mass spectrometer provides a near collisionless environment, minimizing collisional relaxation. This is important, as the vibrational energy that is absorbed, is not lost from the ions in the form of collisions. IRMPD yield does not strictly abide by a linear relationship with either absorption cross-section or laser power, but rather is a kinetically controlled process. Oomens et al. have investigated this by modeling the internal energy distributions of ions following laser irradiation. They were able to reproduce specific IRMPD spectral features, such as a general broadening and red-shifting of bands.

**Free Electron Laser**

The past decade has ushered in IRMPD spectroscopy using free electron lasers (FELs). Currently, there are only two FELs for scientific research set up as user facilities. The Free Electron Laser for eXperiments (FELIX), located at the FOM institute Rijnhuizen in the Netherlands, and Centre Laser Infrarouge Orsay (CLIO), located in
France, are the only two user facilities where visiting scientists can conduct IRMPD spectroscopy experiments on mass-selected complexes with an FTICR-MS\textsuperscript{94,95} or other ion trap mass spectrometers.\textsuperscript{96,97}

**Mechanism of FELIX**

As the FELIX free electron laser was employed in these experiments, the operation of FELIX will be described hereafter. Electron bunches are produced by an electron gun. These electron bunches are accelerated to relativistic speeds (MeV) in a linear accelerator and then injected into a magnetic structure composed of undulating magnetic fields. Figure 2-12 provides a schematic of undulator magnets. When a charged particle traveling at relativistic speeds is subjected to a changing magnetic field, there is a chance that light is emitted. While the electron bunches are deviated by the magnetic field, the emitted light travels in a straight line. The key requirements to coherent emission of radiation in the FEL cavity involve a lagging of the electron bunch by one (or multiple) wavelength(s) per undulator period, as well as extremely short electron pulses (to achieve phase coherence). Intrinsically, FELs are continuously tunable and have large output wavelength ranges. FELIX is tunable between 2.0–250 \( \mu \text{m} \) range. Table 2-1 provides a snapshot of the figures of merit for tunable lasers employed in IRPMD experiments.\textsuperscript{89} The wavelength of the emitted light is dependent on the magnetic field strength and, importantly, the kinetic energy of the electrons. The kinetic energy determines the extent of the Doppler shift that the emitted light is subject to. At a particular kinetic energy, the wavelength of the light can be tuned by a factor of 2 to 3 by adjusting the magnetic field strength in the undulator (by changing the distance of the magnets with stepper motors). The pulse structure of an FEL reflects the injection timing of the electron bunches in the form of a pulse train. For FELIX, this macropulse
pulse train consists of short (0.1–10 ps) micropulses separated by 1 ns, and totaling 5 ms in length. The repetition rate of the macropulses 5 or 10 Hz.

**IRMPD-FTICR-MS Experimental Procedure**

All experiments at the FELIX facility were conducted using a home-built FTICR instrument equipped with an external commercial ESI source (Z-spray, Micromass, UK) and a superconducting 4.7 Tesla magnet. Figure 2-13 shows a schematic of FELIX coupled to this instrument. Peptide solutions at 100 µM were prepared, composed of 70/30/1.5 (v/v/v) water/methanol/formic acid. In other more hydrophobic systems 70/30/2 (v/v/v) water/methanol/formic acid ionized better. The fragment ions were generated by “nozzle-skimmer” dissociation in the ESI source by adjusting the voltage drop between the skimmer and hexapole. Following accumulation in the hexapole, the ions were transferred and guided by the octopole into the ICR cell. SWIFT$^{98,99}$ excitation was used to isolate a specific b fragment ion.$^{98}$ This b ion was subjected to irradiation by the tunable output from FELIX. The remaining precursor ion and FELIX-induced photofragments were excited by an RF broadband pulse. All ions were simultaneously detected by an image current induced at the detection plates of the ICR cell, which was Fourier transformed and converted into a mass spectrum. The laser was then tuned to a different wavelength, so that the photofragmentation process could be repeated for a new ion packet. FELIX was scanned from 1200–2000 cm$^{-1}$. Typically, the mass-selected ion of interest was irradiated with 20-30 macropulses to induce abundant photodissociation. It should be noted that each 5µs macropulse is composed of a train of micropulses at a 1-GHz repetition rate. The energy per macropulse amounts to approximately 50 mJ, of which some 30 mJ finally makes it to the ion cloud in the ICR
trap, due to reflective and absorptive losses in the transfer line. The IRMPD spectrum was obtained by monitoring the IRMPD yield as a function of wavelength. The yield is given by the following equation: \[ \text{Yield} = -\ln\left[1 - \frac{\sum \text{Photofragments}}{\sum \text{All Ions}}\right]. \] The yield is further normalized linearly with the relative FELIX laser power at each wavelength step.\(^{89}\)

**SORI**

Sustained off-resonance irradiation collision-induced dissociation SORI-CID \(^{100}\) is a method for carrying out fragmentation in FTICR cell. It consists of applying a dipolar excitation at the excitation plates that is off-set from the cyclotron frequency of the ion to increase the ion kinetic energy, while simultaneously pulsing in a collision gas to increase the probability of a collision. SORI-CID is generally a low-energy tandem MS technique capable of achieving fragmentation energies on the order of 10 eV,\(^{101}\) even though at higher magnetic field higher collision energies can be achieved. Senko et al. demonstrated that for small molecules (SORI) CAD, very low energy (VLE) CAD, and multiple excitation collisional activation (MECA) provide the additional benefit of producing product ions closer to the center of the trapped ion cell. In SORI-CID, ions are excited by an off-resonance frequency, causing their kinetic energies to increase and decrease repeatedly with time. This results in less-energetic collisions with background molecules over a longer time period when compared to CID.

Experimentally, exciting at a frequency below the resonance frequency (higher mass) gave better results. These collisions can nonetheless result in fragmentation of an isolated ion of interest. Generally, a radio frequency pulse is applied slightly above or below the resonant frequency of the precursor ion. This results in the ion’s kinetic energy to oscillate with time. Figure 2-14 shows the various temporal evolution of ion
cyclotron radius with varying ICR techniques. The internal energy of the ion will slowly increase until the ion undergoes unimolecular dissociation. Generally, the products of SORI are further off resonance and so are not excited by the radio frequency field. Instead, they are collisionally relaxed. As a consequence, low-energy fragmentation pathways may be deduced by off-resonance excitation. In addition to adjusting the duration of SORI, the amplitude of the SORI pulse can be optimized for a given precursor ion and is generally on the order of 10 V(b-p). As a comparison, other methods are shown in the figure to stress how changes in the cyclotron frequency with time effects xy-trajectory. An excitation pulse applied for a time much longer than the time between collisions to ensure multiple collisions during the process. The Internal energy of the system slowly increases by allowing multiple collisions translationally excited ion with the target gas. Eventually, this results in the dissociation of the molecule. For monoatomic gases the CID efficiency is expected to increase with increase in the mass of collision gas because of the corresponding increase in the CM collision energy. Collision gas can be admitted into the ICR cell using a leak valve or a pulsed valve. The velocity of the ion, the collision frequency, and the mass of the collision gas determine the practical range of static pressures for multiple collisions in CID. For example, 1x10⁻⁷ Torr of argon is required to obtain 20 collisions per second with protonated penta-alanine for the CM collision energy of 5 eV (even higher pressure is required for activation of slower ions). However, as the static pressure in the cell increases, both sensitivity and resolution of the ICR degrade rapidly because of collisional damping of the ICR signal during the detection.
There are some similarities between SORI-CID and IRMPD. Both are a slow activation method which explains an overlap in the fragmentation mass channels that are observed. However, dissimilarities are present between the two techniques. For example, SORI-CID is intrinsically mass selective, in that only ions that have a cyclotron frequency close to the excitation frequency are accelerated, whereas in IRMPD all ions are excited that reside in the optical path of the laser and that display absorptivity at the laser frequency. IRMPD generally produces richer fragmentation spectra, which is attributed to the nonselective nature of IRMPD, depositing energy into both the precursor ion and subsequent product ions.

Hydrogen Deuterium Exchange (HDX)

An FTICR mass spectrometer is ideally suited to carry out gas-phase reactions. As an ion trapping mass spectrometer that is operated at an ultra-high vacuum, relatively fast reactions can be observed by employing ultra-low concentrations. Conversely, the long trapping times (i.e., hours) and higher pressures allow relatively slow processes to be studied. Gas-phase hydrogen/deuterium exchange (HDX) involves leaking a deuterating reagent into the vacuum chamber and monitoring the mass increase as a function of time. HDX can be considered as a pseudo-first-order reaction process. This assumption is valid because the extent of deuterating reagent is in great excess relative to the analyte presence inside the cell. Furthermore, quantitative kinetic information can be extracted from the experiment. To determine pseudo-first-order HDX kinetics, the natural logarithm of the relative depletion of the undeuterated peak, ln[d0/∑dn], is plotted as a function of the HDX time.

In peptides, hydrogens bonded to oxygens, nitrogens, and sulfurs are labile to exchange, whereas hydrogens bound to carbons are not. In solution HDX, proteins are
incubated in D$_2$O buffer. The degree of deuterium exchange is ascertained via determined nuclear magnetic resonance or mass spectrometry. The isotopic exchange of these labile hydrogens represents an important probe for determining protein structures, and to study dynamic processes such as protein folding. Factors such as temperature, pH, and neighboring side chains effect observed kinetic exchange rates. Due to the diluted environment in the ICR cell, gas-phase HDX reaction rates are much lower than in solution. The proton affinity may be defined as the enthalpy of reaction by Equation 2-18. The proton affinities of the deuterating reagents and analytes play an important role in HDX kinetics. This is presumably due to the mechanism of HDX, which is thought to involve a proton/deuteron transfer between the analyte and deuterating reagent. The calculated gas-phase proton affinities for deuterating reagents are as follows: D$_2$O, CD$_3$OD, CD$_3$COOD, and ND$_3$, are 166.5, 181.9, 190.2, and 204.0 kcal/mol. The fastest exchange rates are provided by ND$_3$ when exchanging with peptides or amino acids. The understanding to the mechanism of gas-phase HDX remains still inadequate. Several groups have contributed to more accurately account of the gas-phase proton affinities for amino acids through experimental or theoretical modeling. Faster exchange rates correspond to deuterating reagents with higher proton affinities. The evaluated proton affinities of amino acids are listed in Table 2-2. Proton affinities of peptides depend on the proton affinity of every single amino acid residue, as well as its chain length. For example, the proton affinities for the following oligo-glycine systems has been computed: Gly$_2$, Gly$_3$, Gly$_4$ and Gly$_5$ are 210.0, 213.0 218.1, and 218.4 kcal/mol 129,133. These variations
in proton affinities may be correlated to the accessible sites for exchange. Moreover, this equates to dynamic picture of peptide structure in the gas phase.

Gas-phase HDX has also been applied to the structural characterization of peptide b ions generated from CID. "Fast" and "slow"-exchanging structures have been distinguished in this way by a number of groups. The relative differences in H/D exchange rates were ascribed to different ion structures. For example, quantification of oxazolone vs. macrocycle structures can performed by gas-phase hydrogen/deuterium exchange (HDX), by taking into account differences in HDX kinetics between oxazolone ("fast") and macrocycle ("slow") structures. Generally, their rate constants differ by an order of magnitude. Solouki has utilized HDX to investigate how peptide confirmations is affected when complexation between alkali metals and peptides occur in gas the phase. Also, they performed a multiplex HDX exchange experiment where they monitored the gas-phase H/D exchange reactions of the WHWLQL hexapeptide and its b\textsubscript{4}\textsuperscript{+} and b\textsubscript{5}\textsuperscript{+} isobaric fragment ions. These methods allow the researcher the distinct possibility of probing peptide fragmentation pathways under different low energy conditions. For example, their results demonstrated that the parent molecular ion exhibits one ion population whereas the b\textsubscript{4}\textsuperscript{+} and b\textsubscript{5}\textsuperscript{+} fragment ions exhibit a bimodal distribution. The bimodal ion population is attributed to a fast and slow exchanging structures. Consequently, one can discern between an open confirmations vs. a more compact structure. Their rate constants differ by an order of magnitude. Therefore, oxazolone and cyclic structure are ideal candidate structures to study via HDX. A study on b\textsubscript{2}-b\textsubscript{4} for Leu-enkephalin showed similar trends in the increase of macrocycle structures as a function of b fragment length. As discussed by Yu, chain length
becomes an issue because the propensity for sequence permutation increases with length of tryptic peptides. In a study performed by Molesworth, significant formation of nondirect sequence ions were observed for larger b_n ions in size range n=4-10. Clearly, better insights regarding the question of scrambling allows both the proteomics community to make more confident sequence assignments and scientists to possess a firmer grasp on the fundamentals governing peptide fragmentation mechanisms in gas phase.

**HDX Mechanism**

In a simplified form, an HDX reaction may be considered to be a gas phase proton transfer reaction expressed by Equation 2-17. Thermodynamically, this ion-molecule process reaction may be expressed by the following equation. In addition, the free energy and enthalpy if the reaction may be extracted.\(^{118,119}\)

\[
B_{(g)} + H^+_{(g)} \rightarrow BH^+_{(g)}
\]

\(-\Delta G^\circ = GB(B), \ -\Delta H^\circ = PA(B)\) \hspace{1cm} (2-17)

\(-\Delta G^\circ\) (negative of the free energy), \(-\Delta H^\circ\) (negative of the enthalpy), and \(PA(B)\) (proton affinity).

HDX is affected by several factors: the difference in PAs between the analyte and the deuterating reagent, the structure and size of the analyte molecule, and the degree of deuterons comprising the deuterating reagent. If the proton affinity of the sample and deuterating reagent are similar, an exchange becomes feasible. Figure 2-15 depicts several proposed mechanisms for HDX in the gas phase. For low-basicity reagents, such as CD\(_3\)OD or D\(_2\)O, a relay mechanism is proposed for low-basicity reagent such as D\(_2\)O. In this mechanism, the deuterating molecule forms two hydrogen bonds: one
with the protonated site on the N-terminus, and one with amide oxygen. A concerted movement of the N-terminal proton to the deuterating molecule, and movement of a deuteron to a carbonyl oxygen are enacted. For exchanges to occur, it was proposed that the energy gained by forming hydrogen bonds must be larger than the difference in the proton affinities of the exchange sites and the energy lost by opening the internally solvated structure.\textsuperscript{128} On the other hand, ND\textsubscript{3} the most basic deuterating exchanges via an onium' mechanism. Here, the ND\textsubscript{3} solvates the proton as an intermediate.

**Conclusions**

HDX can also be combined with activation experiments, such as in-source CID or SORI-CID, to glean further insight into gas-phase structure and conformational states of peptides. By combining SORI-CID and IRMPD measurements both a richer spectral information such as fragmentation mass channels and structural information can be obtained. Gas-phase hydrogen/deuterium exchange (HDX) measures the reaction rate of the CID product with a deuterating agent (e.g., CH\textsubscript{3}OD, ND\textsubscript{3}).\textsuperscript{113} “Fast” and “slow”-exchanging b fragment structures have been distinguished in this way by a number of groups.\textsuperscript{70,71,114,116,120,121} The relative differences in H/D exchange rates are correlated to different ion structures, namely oxazolone (“fast”) and macrocycle (“slow”) structures in the case of b product ions. Quantification of oxazolone vs. macrocycle structures can be performed by an analysis of the kinetic rates. While HDX can thus give important insights into make-up of structural isomers of CID product ions, the interpretation of HDX results can often be less straightforward.

In IRMPD spectroscopy, the product ion of interest is photo-dissociated with a tunable infrared laser. The infrared spectrum of an ion reveals information on the chemical structure by means of diagnostic vibrational frequencies. Thus, the oxazolone
structure can be confirmed via the presence of a lactone-type C=O stretch band, which appears in the high-frequency region (1770-1950 cm⁻¹) of the mid-infrared spectrum.⁶⁴,¹²²
Figure 2-1. Ion cyclotron motion undergone by an ion in the presence of a magnetic field. [Figure is adapted with permission from Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Mass Spectrom. Rev. 1998, 17, 1].

Figure 2-2. Conventional FTICR cell. Axial and radial trapping occur in this cell and give rise to various ions motions. [Figure is adapted with permission from http://www.chm.bris.ac.uk/ms/theory/esi-ionisation.html]
Figure 2-3. Ion motions which arise in ICR cell. Motions which arise in ICR cell due to axial and radial confinement. [Figure is adapted with permission from Schmid, D. G.; Grosche, P.; Bandel, H.; Jung, G. Biotechnology and Bioengineering 2000, 71, 149.]
Figure 2-4. Pulsed sequences utilized in FTICR-MS experiment. A quench pulse is used at the on-set of the experiment. Ion storage is achieved by incorporating timing delays between ionization and excitation steps.

Figure 2-5. Uses of ion cyclotron excitation. Left. Acceleration of ions to form a spatially coherent packet at detectable orbital radius. Middle: increase ion kinetic energy to above the threshold for CID. Right Expulsion of ions of a given mass-to-charge ratio.
Figure 2-6. Illustration of the processing of raw data: conversion time domain signal $\rightarrow$ frequency spectrum $\rightarrow$ mass spectrum. [Figure taken with permission from Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. Analyst 2005, 130, 18].

Figure 2-7. Schematic of commercial FTICR mass spectrometer (Bruker Daltonics) 4.7 Tesla magnet.
Figure 2-8. Various ICR cell geometries. [Reprinted with permission from John Wiley & Sons, Ltd. Marshall, A.G.; Hendrickson, C.L.; Jackson, G.S. 1998. Fourier Transform Ion Cyclotron resonance mass Spectrometry: A Primer Mass Spectrom. Rev. (Volume 17, page 12, Figure 12)].
Figure 2-9. FTICR performance as a function of magnetic field strength. Left side displays parameters that increase linearly. Right side displays parameters that increase quadratically. [Reprinted with permission from John Wiley & Sons, Ltd. Marshall, A.G.; Hendrickson, C.L.; Jackson, G.S. 1998. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. Mass Spectrom. Rev. (Volume 17, Page 22, Figure 19)].

Figure 2-10. Dependence of resolution on acquisition time or number of data points. [Figure taken with permission from Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. Analyst 2005, 130, 18].
Figure 2-11. Schematic representation of IVR mechanism. [Figure adapted with permission from Polfer, N. C.; Oomens, J. Mass Spectrom. Rev. 2009, 28, 468].
Figure 2-12. Schematic representation of undulator magnets.[ Courtesy of Dr. Jos Oomens the FOM-Institute].
Figure 2-13. Schematic representation of FELIX couples to FTICR. [Reproduced with permission from John Wiley and Sons, Inc. from Eyler, J.R Mass Spectrom. Rev. 2009, 448-467].
Figure 2-14. Temporal evolution of ion cyclotron radius with varying ICR techniques. [Reprinted with permission from John Wiley & Sons, Ltd. Marshall, A.G.; Hendrickson, C.L.; Jackson, G.S. 1998. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. Mass Spectrom. Rev. (Volume 17, Page 21, Figure 18)].
<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onium ion mechanism</td>
<td><img src="image" alt="Onium ion mechanism diagram" /></td>
</tr>
<tr>
<td>Relay mechanism</td>
<td><img src="image" alt="Relay mechanism diagram" /></td>
</tr>
<tr>
<td>Salt bridge mechanism</td>
<td><img src="image" alt="Salt bridge mechanism diagram" /></td>
</tr>
<tr>
<td>Flip-flop mechanism</td>
<td><img src="image" alt="Flip-flop mechanism diagram" /></td>
</tr>
<tr>
<td>Tautomer mechanism</td>
<td><img src="image" alt="Tautomer mechanism diagram" /></td>
</tr>
</tbody>
</table>

Figure 2-15. Gas-phase H/D exchange mechanisms proposed by Beauchamp and co-workers. [Figure is adapted with permission from Campbell, S.; Rodgers, M. T.; Marzluff, E. M.; Beauchamp, J. L. *J. Am. Chem. Soc.* 1995, 117, 12840].
<table>
<thead>
<tr>
<th>Laser system</th>
<th>Type</th>
<th>Range/cm</th>
<th>Peak power</th>
<th>Pulse length</th>
<th>Energy per pulse</th>
<th>Energy per second/J</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO2</td>
<td>Gas discharge</td>
<td>925-1085</td>
<td>5-15 W</td>
<td>cw</td>
<td>N/A</td>
<td>5-15</td>
</tr>
<tr>
<td>CO</td>
<td>Gas discharge</td>
<td>1600-1900</td>
<td>1-10 W</td>
<td>cw</td>
<td>N/A</td>
<td>1-10</td>
</tr>
<tr>
<td>FELIX</td>
<td>Free electron laser</td>
<td>500-2000</td>
<td>10 MW</td>
<td>GHz pulse train of micropulses</td>
<td>Micropulse: 2-12 µJ</td>
<td>0.01-0.6 (10Hz)</td>
</tr>
<tr>
<td>CLIO</td>
<td>Free electron laser</td>
<td>600-2500</td>
<td>20 MW</td>
<td>25 MHz train of micropulses</td>
<td>Micropulse: ≤ 25µJ</td>
<td>≤ 0.6 (25Hz)</td>
</tr>
<tr>
<td>OPO/A</td>
<td>Non-linear optics</td>
<td>2500-4000</td>
<td>0.5-2 MW</td>
<td>4-10 ns</td>
<td>6-10 mJ</td>
<td>0.06-0.1 (10 Hz)</td>
</tr>
<tr>
<td>OPO/A</td>
<td>Non-linear optics</td>
<td>2500-4000</td>
<td>50-100</td>
<td>cw</td>
<td>N/A</td>
<td>0.05-0.1</td>
</tr>
</tbody>
</table>

αFELIX can also be operated on the 3rd harmonic to give a tuning range of 2000-4000 cm⁻¹. Additionally, the Free Electron Laser for IntraCavity Experiments (FELICE) extends IRMPD experiments into the far IR range (i.e., 50-500 cm⁻¹) [Table adapted with permission from Polfer, N. C. *Chem. Soc. Rev.* 2011, *40*, 2211].
Table 2-2. Evaluated gas-phase basicities and proton affinities.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>GB (Kcal mol⁻¹)</th>
<th>PA (Kcal mol⁻¹)ᵃ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>202.7</td>
<td>210.5</td>
</tr>
<tr>
<td>Cys</td>
<td>206.2</td>
<td>214.0</td>
</tr>
<tr>
<td>Ala</td>
<td>206.4</td>
<td>214.2</td>
</tr>
<tr>
<td>Ser</td>
<td>207.6</td>
<td>215.2</td>
</tr>
<tr>
<td>Asp</td>
<td>208.6</td>
<td>216.4</td>
</tr>
<tr>
<td>Val</td>
<td>208.7</td>
<td>216.5</td>
</tr>
<tr>
<td>Leu</td>
<td>209.6</td>
<td>217.4</td>
</tr>
<tr>
<td>Ile</td>
<td>210.8</td>
<td>218.6</td>
</tr>
<tr>
<td>Thr</td>
<td>211.7</td>
<td>219.5</td>
</tr>
<tr>
<td>Phe</td>
<td>212.1</td>
<td>219.9</td>
</tr>
<tr>
<td>Asn</td>
<td>212.8</td>
<td>220.6</td>
</tr>
<tr>
<td>Tyr</td>
<td>213.1</td>
<td>220.9</td>
</tr>
<tr>
<td>Met</td>
<td>213.3</td>
<td>221.1</td>
</tr>
<tr>
<td>Gln</td>
<td>214.2</td>
<td>222.0</td>
</tr>
<tr>
<td>Pro</td>
<td>214.3</td>
<td>222.1</td>
</tr>
<tr>
<td>Glu</td>
<td>215.6</td>
<td>223.4</td>
</tr>
<tr>
<td>Trp</td>
<td>216.1</td>
<td>223.9</td>
</tr>
<tr>
<td>Lys</td>
<td>221.8</td>
<td>235.6</td>
</tr>
<tr>
<td>His</td>
<td>223.7</td>
<td>231.5</td>
</tr>
<tr>
<td>Arg</td>
<td>237.0</td>
<td>244.8</td>
</tr>
</tbody>
</table>

[Table adapted with permission from Harrison, A. G. *Mass Spectrom. Rev.* 1997, 16, 201].
CHAPTER 3
DISFAVORING MACROCYCLE B FRAGMENTS BY CONSTRAINING TORSIONAL FREEDOM AND BY CHARGE SEQUESTRATION

Various systematic studies have brought to light that \( b_n \) fragments are prone to isomerize upon low-energy activation. Specifically, oxazolone structures can undergo a complex rearrangement, involving a nucleophilic attack from the N-terminus, resulting in “head-to-tail” macrocycle structures. The occurrence of macrocycle structures is a central tenet to the scrambling hypothesis by Harrison and Paizs\(^\text{123}\), as the macrocycle may not always re-open where it was originally formed. Subsequent fragmentation from these permuted sequences leads to non-direct/permuted/scrambled CID product ions in tandem mass spectra. The appearance of such permuted sequence ions has been verified for select peptide systems.\(^\text{73,74,124-126}\) A deeper comprehension of the fragmentation chemistry in CID can be obtained by employing structural techniques that confirm the ion structures and mechanistic pathways for bio-molecules in the gas phase. As has been discussed in the experimental techniques chapter, gas-phase hydrogen/deuterium exchange (HDX),\(^\text{113}\) infrared multiple photon dissociation (IRMPD) spectroscopy,\(^\text{46,89}\) and ion mobility mass spectrometry\(^\text{127-129}\) have all been applied to this task.

Analysis of tandem MS studies have shown significant formation of non-direct sequence ions for larger \( b_n \) ions.\(^\text{73,117}\) These findings are consistent with enhanced macrocycle formation for larger \( b_n \) fragments, as well as a higher proclivity for opening up at a different residue than where they were originally fused together. Nonetheless, it appears that some residues, such as arginine, can disrupt the scrambling chemistry, as demonstrated by Van Stipdonk and co-workers.\(^\text{130}\) In their tandem mass spectroscopic experiments, only direct sequence ions for permuted positions of Arginine in \( b_5 \) ions are
observed. Based upon their findings, they deduce that arginine inhibits the formation of macrocyclic $b_n$ ions regardless of its sequence position. Another study by the same group showed that the reopening of macrocycle structures is influenced by the amino acid side chains. For a series of permuted isomers with glutamine, $b_5^+$ ions showed nearly identical MS$^n$ spectra, which suggested that macrocycle structures tend to reopen at the position of glutamine. Conversely, Bythell et al. employed tandem mass spectrometry and calculations to study the effect of Histidine upon cyclization. Their results indicate that cyclization/reopening is not as pronounced for $b_n$ ions containing his residue than for those with only aliphatic residues.

In this Chapter, a systematic study of the effect of specific residues on head-to-tail cyclization is presented for the $b_6$ sequence motif QWFGLM, where all amino acid residues are different, thus in complete contrast to the oligoglycine studies. In addition, this peptide contains no basic amino acids (i.e., arginine, lysine, or histidine), and hence no proton will be located at the side chain. The experimental methods employed include IRMPD and gas-phase HDX. Commercially available cyclo (QWFGLM) is a reference peptide in the study of the $b_6$ motif QWFGLM. For the investigation of the proline effect on the head-to-tail cyclization, a synthetically made peptide, cyclo (QPFGLM) is utilized. From solution-phase cyclopeptide synthesis, it is well known that the conformational flexibility of cyclopeptides can be controlled by the introduction of rigid, non-natural amino acids in the ring. For instance, Kubik synthesized rigid cyclic hexapeptides composed of alternating L-glutamic acid and 3-aminobenzoic acid. Similarly, many cyclic peptides, such as gramicidin S, include proline residues. Work by van Maarseveen et al. further corroborates that through more aggressive chemistries such
as CuI–catalyzed cycloaddition that the seemly challenging synthesis of cyclo-[Pro-Val-Gly-ψ(triazole)-Tyr], a triazole analog of cyclo-[Pro-Val-Pro-Tyr], can readily be achieved in solution phase chemistry. This study aims at determining how amino acid residues with limited torsional flexibility affect macrocycle formation in the b6 fragment sequence motif QWFGLM, which has a very high propensity to form macrocycle structures. The residues proline and 4-aminomethylbenzoic acid (4AMBz) were chosen here in the expectation that their rigid structures may restrict “head-to-tail” cyclization from the N-terminus, as shown in Scheme 3-1. In addition, utilizing the same sequence motif, QWFGLM, the effect of charge sequestration by a basic residue arginine (R) may be evaluated for the presence of macrocyclic or oxazolone pathways.

**Experimental**

**Sample Preparation**

All peptides were synthesized with conventional FMOC solid-phase synthesis methods using 9-fluroenylmethoxycarbonyl (FMOC)-Gly loaded resin at the University of Florida. A CEM microwave Discover System (Matthews, NC) was utilized to synthesize 0.1 mmol scale of the corresponding peptides. Amino acid residues were purchased from Advanced Chem Tech (Louisville, KY) and used as received without further purification. Briefly, upon completion of the synthesis, the peptide was cleaved from the resin using a 14mL cocktail solution made up of 95% trifluoroacetic acid (TFA), 2.5% water, and 2.5% triisopropylsilane. Three rounds of purification and centrifugation were performed. The peptide was precipitated from the TFA solution with ice cold ether. After decanting the last amount of diethyl ether, a gentle stream of nitrogen gas was applied to the surface of the peptide to assist drying. All products had a brown color. High-performance liquid chromatography (HPLC) and mass spectrometry was
subsequently employed to assess product presence and purity. Typically, yields of greater than 90% were obtained. The cyclic peptides were synthesized at the University of Amsterdam. First, the linear peptides QPFGLMPG and QPWFGLMPG were made with solid phase synthesis on a trityl resin. After cleavage and purification of the linear peptide, the peptides were dissolved in THF at concentrations of less than $10^{-3}$ mol/L, with 4.4 equivalent of N, N-diisopropylethylamine (DIPEA), 2.2 equivalent of 2-(1H-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate methanaminium (HATU) and 3H-1,2,3-triazolo[4,5-b]pyridin-3-ol (HOAT) were added into the solution. The reaction solution was kept at room temperature while stirring for 24h. The aliquot from the reaction solution was characterized by LC/MS to ensure that the reaction went to completion. The THF was evaporated off from the reaction solution, and the remaining solid was then dissolved in ethyl acetate. 1M KHSO$_4$ aqueous solution was added to dissolve the unreacted coupling reagents. The organic phase was collected and solid Na$_2$SO$_4$ was added to remove the remaining water. The organic phase was lyophilized to obtain solid products. The crude peptides were purified by reversed-phase HPLC (RF-HPLC) on a C18 column using a gradient of 0-80% B (Buffer A: water/0.05%TFA; Buffer B: 90% acetonitrile/10% water/0.045% TFA) over 30 min.

**Mass Spectrometry and Infrared Photodissociation Spectroscopy**

IRMPD spectra of b ions were recorded using the free electron laser FELIX$^{139}$ located at the FOM institute for Plasma Physics Rijnhuizen. The ions were generated by electrospray ionization (ESI) in a home-built Fourier transform ion cyclotron resonance (FTICR) mass spectrometer described previously.$^{140,141}$ Peptide solutions at 100 μM were prepared, composed of 50/50/2 (v/v/v) water/methanol/formic acid. The fragment ions were generated by “nozzle-skimmer” dissociation in the ESI source. SWIFT
excitation was used to isolate a specific $b_6$ fragment ion. Following accumulation in the hexapole, the ions were transferred and guided by the octopole into the ICR cell. The mass-selected ion of interest was irradiated with 20-30 macropulses from the free electron laser FELIX. Each 5-μs macropulse is composed of a train of micropulses at a 1-GHz repetition rate. The energy per macropulse amounts to approximately 50 mJ of which some 30 mJ finally makes it to the ion cloud in the ICR trap. The IRMPD spectrum was obtained by monitoring the IRMPD yield as a function of wavelength (1300–1975 cm$^{-1}$). The yield is given by the following equation: 

$$\text{Yield} = -\ln\left[1 - \frac{\sum \text{Photofragments}}{\sum \text{All Ions}}\right].$$

The yield is further normalized linearly with the relative FELIX laser power at each wavelength step.

**SORI CID Experiments**

Complementary sustained off-resonance irradiation collision-induced dissociation (SORI CID) were carried out in a commercial FTICR mass spectrometer (4.7 T actively shielded APEX II, Bruker Daltonics, Billerica, MA). The $b$ CID product ions were generated by “nozzle-skimmer” dissociation in the ESI source. These $b$ ions were then subjected to SORI CID in the ICR cell using a frequency offset of -2.0 kHz (relative to the precursor ion’s cyclotron frequency) and a nitrogen gas pulse (<10$^{-7}$ Torr).

**Results and Discussion**

The protonated precursor octapeptide QWFGLMPG was chosen to yield an abundant $b_6$ fragment, due to facile cleavage on the N-terminal side of proline. The IRMPD spectrum of this $b_6$ fragment is shown in Figure 3-1. Clearly, no bands are observed >1770 cm$^{-1}$, which indicates an absence of oxazolone structures. A comparison to the synthetically made cyclic peptide reference system, protonated cyclo(QWFGLM), which had been reported previously$^{70,136}$ shows that both spectra are
nearly identical. There are minor differences in the 1600 cm\(^{-1}\) region, which may suggest some differences in the conformeric structures that are present. Nonetheless, the close similarity to the spectrum of the synthetic cyclic peptide provides compelling evidence that b\(_6\) from QWFGLMPG exclusively adopts a macrocycle configuration with no presence of oxazolone structures. This is in marked contrast to previous studies where mid-sized b fragments were found to be comprised of a mixture of oxazolone and macrocycle structures.\(^{70,71}\) Sequence analogs of QWFGLMPG, incorporating proline and 4-aminomethylbenzoic acid (4AMBz),\(^{142,143}\) were synthesized to investigate their effect on macrocycle formation in the corresponding b fragment.

**Proline**

Due to its secondary amine structure, the proline residue adopts a fixed dihedral angle (\(\sim 60^\circ\)) in peptides on its N-terminal side.\(^{144}\) In addition, proline is an endocyclic five-membered ring which introduces rigidity because no bond rotation is possible over the C (alpha)-N bond. Consequently, this rigidity may hamper closure due to unfavored preorganization of the N- and C-termini in the linear precursor. This strained geometry rationalizes a weakening of the amide bond, and hence enhanced fragmentation. A proline residue was substituted and inserted at position 2 in the sequence motif QWFGLM, to generate Q\(_{\text{P}}\)FGLM \(b_6\) and Q\(_{\text{P}}\)WFGLM \(b_7\). In Figure 3-2, the experimental IRMPD spectra for the proline results are shown spanning the mid-IR range (1200-2000 cm\(^{-1}\)). At the higher frequency portion of the spectrum, \(b_7\) exhibits a broad and sizeable band position from 1770–1870 cm\(^{-1}\), consistent with an oxazolone C=O stretch. For \(b_6\), the presence of weaker oxazolone bands is manifested in a similar range (i.e., 1755–1870 cm\(^{-1}\)). A comparison to the IR spectrum of the synthetically made cyclic structure analog, protonated cyclo-QPFGLM, confirms that \(b_6\) parallels some
spectroscopic features of b⁷ more closely than those of its cyclic structural analogue, notably, at frequencies around 1700 cm⁻¹ for the amide I band (i.e., backbone C=O stretch) and at ~1450 cm⁻¹. In addition, the absence of spectral features around 1600 cm⁻¹ is similar for b⁶ and b⁷ fragment ions. The weaker oxazolone band for b⁶, in contrast to the more intense feature for b⁷, hints at the presence of additional structures for b⁶. Note that the maximum IRMPD yields obtained for the b⁶ and b⁷ ions were 37% and 96%, respectively. The appearance of oxazolone structures for proline-containing peptides is consistent with restricted “head-to-tail” cyclization. Unfortunately, a detailed investigation of the effect of the position of proline on head-to-tail cyclization is complicated by prevalent cleavage on its N-terminal side. While proline in position 2 yields no abundant y₇ ions from QPFGLMPG (due to an inability to form b₁), proline substitutions in other position induce abundant cleavages substantially reducing the abundance of the ion of interest here.

4-Aminomethylbenzoic Acid

Similarly to proline, incorporation of the non-natural amino acid 4AMBz also leads to reduced torsional freedom in a peptide backbone. However, in contrast to proline, 4AMBz does not give rise to redundant dissociation, such as (much) enhanced cleavage on its N-terminal side. Figure 3-3 shows the IRMPD spectra for the b₆ fragments with sequence motif QWFGLM, with different 4AMBz substitutions. With the exception of 4AMBz in the sixth position, all the corresponding b₆ ions were generated.

Despite an intense precursor ion signal for QWFGL(4AMBz)PG, its b₆ fragment ion could not be produced. This is most likely attributed to the fact that 4AMBz sterically hinders the nucleophilic attack from a backbone carbonyl, as depicted in Figure 3-4.
The lack of a QWFGL(4AMBz) b6 fragment is in accordance with the proposed mechanism that oxazolone structures are formed prior to isomerization to macrocycle structures. In Figure 3-3, there is no evidence for oxazolone structures, with the exception of a weak feature between 1765-1860 cm⁻¹ for 4AMBz in position 5. In the corresponding oxazolone structure in Figure 3-5, the oxazolone ring and the benzene ring form a delocalized, resonantly-stabilized π-electron system. Such an extended π-electron system is not possible for any of the other b fragments. It is hence likely that the small fraction of oxazolone structures for b6 QWFG(4AMBz) M is due to a lowering in energy of the oxazolone structure, due to this extended π-electron system, rather than an ability of 4AMBz in obstructing a head-to-tail cyclization.

Similarly, Van Stipdonk and co-workers have noted an enhancement of b-type product ions when the aromatic amino residue 4AMBz was located in the penultimate position from the residue where amide cleavage takes place. This phenomenon is attributed to a highly conjugated and stable oxazolone structure stemming from the aromatic ring substituent 4AMBz. As a control experiment, a poor nucleophile analog of 4AMBz, 4-aminobenzoic acid (4Abz), was used to prevent head-to-tail macrocycle formation. In Figure 3-6, (4Abz)WFGLM b6 exhibits an intense oxazolone band from 1865–1970 cm⁻¹, in stark contrast to the absence of such a band for (4AMBz)WFGLM b6. While the absence of oxazolone bands does not directly confirm the (exclusive) presence of macrocycle structures, the IRMPD spectra for 4AMBz inserted in positions 2-5 display similar amide I band positions compared to the protonated cyclic peptides, cyclo(QWFGLM)H⁺ and cyclo(QPFGLM)H⁺, even if the amide II (i.e., NH bending) positions are slightly red-shifted (10-15 cm⁻¹). The IRMPD
spectrum for (4AMBz)WFGLM b₆ differs from the other 4AMBz analogs, in that it displays an obvious splitting of the amide I stretching modes. The same effect is seen for (4Abz)WFGLM b₆, which suggests that this is related to the benzoic acid residue CO stretch.

**Photofragmentation and SORI CID:**

In addition to the vibrational information from IRMPD spectra, the photofragmentation mass channels can offer insights whether the cyclic peptides in fact give rise to consecutive fragment ions with scrambled sequences. A comparison to SORI CID mass spectra can establish whether these fragmentation patterns are more general for low-energy activation methods. The SORI CID and IRMPD mass spectra for the control b₆ fragment from (4ABz)WFGLMPG are shown in Figure 3-7. IRMPD of (4ABz) WFGLM b₆ exhibits a b ion series arising from cleavage propagating along the backbone. In the SORI CID experiment, many of the same fragments are confirmed (dashed lines), but additional fragments are observed when the precursor b₆ is depleted fully. In particular, the appearance of the internal fragment FGLM confirms that two backbone cleavages can take place under these conditions.

The SORI CID and IRMPD mass spectra for the b₆ CID products from (4AMBz)-substituted peptides are shown in Figure 3-8. In the IRMPD results, despite considerable evidence for macrocycle formation in these b fragments, only one scrambled sequence ion could be identified. For (4AMBz) WFGLM b₆, the photofragment at m/z 552 could correspond to a scrambled (4AMBz) WFL a₄ (marked in red in Figure 3-8) requiring internal elimination of a glycine residue. The photodissociation of these b₆ CID fragments display extensive b-NH₃ and b-H₂O photoproducts, but the corresponding b ions are often not observed. This is surprising,
as ammonia and water losses would be expected to be sequential from b ions. While many similar fragments are observed for IRMPD and SORI CID, many additional fragments are seen in the corresponding SORI CID mass spectra. Some of these additional fragments are assigned to internal sequence losses, scrambling, and/or complex re-arrangements. For (4AMBz)WFGLM b_6, the fragment at m/z 675 is assigned to an internal elimination of Gly-H_2O; the fragment at m/z 621 is compatible with an internal elimination of phenylalanine. These sequential fragment ions likely arise from the formation of a macrocylic structure that has re-opened and undergone the loss of the aforementioned residues. Q(4AMBz)FGLM b_6 provides a lot of evidence for internal fragments where the (4AMBz) residue is retained (marked in blue). On the other hand, the SORI CID and IRMPD mass spectra for QW(4AMBz)GLM b_6 closely reflect the original sequence. The high prevalence of b-NH_3 and b-H_2O fragments could be related to the presence of glutamine. It has been observed previously that when glutamine is on the N-terminus of a peptide chain, either water loss (-18 Da) or ammonia loss (-17 Da) can take place.\textsuperscript{147,148} In addition, Tabb et al. have noted in their statistical analysis of tandem mass spectra from tryptic peptides that the presence of glutamine leads to preferential ammonia loss.\textsuperscript{149}

**Arginine**

It has been hypothesized by Molesworth et al. that the basic nature of arginine can inhibit scrambling chemistry.\textsuperscript{130} They as well as Paizs argue that arginine’s side chain leads to sequestering of the proton, resulting in amide proton mobilization, rather than conventional charge-directed fragmentation (to form the b_5 ions).\textsuperscript{131,150,151} Molesworth et al. investigated permutation processes in b_5 ions for the sequence motif YARFL, and variants thereof. They utilize two criteria to assess b-type scrambling. Both the
elimination of internal amino acid residues and similarity of CID product ion spectra derived from the related sequences are utilized as gauges of $b_n$ type scrambling.\textsuperscript{57,73} A logical question to pose is if arginine residues can inhibit of macrocyclization in other sequence motifs, and whether this phenomenon is restricted to $b_5$ ions, or rather is observed more generally. To address these questions, we utilize the same sequence motif from the proline and 4AMBz work, QWFGLMPG. This serves as an excellent reference system, as most of these $b_6$ ions have a high propensity for head-to-tail macrocycle formation, with the exception of the proline-containing peptides.

Figure 3-9 shows IRMPD spectra for $b_6$ fragments with different arginine substitutions (1, 2, 3 and 4) in the sequence motif QWFGLM. The IRMPD spectra for arginine in all of the varying six positions could not be measured because of time and signal-to-noise constraints. The $b_6$ ion with arginine in position 2, QRFGLMPG $b_6$, exhibits the most distinct oxazolone CO stretch band ($1772−1828$ cm$^{-1}$). On the other hand, the IRMPD spectra for the corresponding $b_6$ ions for RFGLMPG and QFRLMPG reveal no presence of bands in this region. For QRGGLMPG $b_6$, with arginine in the third position, a much weaker band is observed at this frequency. This indicates that oxazolone structures are abundant for QRFGLMPG $b_6$, but are much sparser (QRGGLMPG $b_6$) or even absent (RFGLMPG and QFRLMPG $b_6$) for other sequence motifs.

In Figure 3-10, the corresponding IRMPD spectra for $b_5$ ions are shown, with arginine in positions 2, 3 and 4. Once again, for the sequence QRFGLMPG, with arginine in position 2, a discrete oxazolone band is apparent. However, neither of the other two sequences display a band in this region. Table 3-1 summarizes the
interpretation of the IRMPD results for arginated peptide b ions. The results suggest that the presence of arginine can in fact impede head-to-tail cyclization, but this effect is sequence- and/or position-dependent. Moreover, the absence of an oxazolone band for QWRGLMPG b5, in contrast to the (weak) presence of such a band for QWRGLMPG b6, is consistent with a size dependence. Yet this trend runs counter to previous studies, which had shown that head-to-tail macrocyclization is favored for larger b ions.

The observation that the position of the arginine residue plays a key role in whether oxazolone or head-to-tail macrocycle structures are observed indicates that the chemistry is dependent on the conformations of the peptide, and notably the interaction of the arginine side chain with the remainder of the peptide. In CID chemistry, it is supposed that oxazolone structures are generated prior to isomerization to a head-to-tail macrocycle structure. This head-to-tail nucleophilic attack can only occur if the oxazolone ring remains protonated. If the proton were to migrate from the oxazolone ring to another site in the molecule, the head-to-tail macrocyclization would be precluded. As the guanidine side chain is the most basic naturally occurring moiety in peptides, one would expect the proton to migrate there. The kinetics of this proton migration is determined by the proximity of the guanidine side chain to the oxazolone ring moiety. In Figure 3-11, this hypothesis is shown for the arginated peptide systems investigated here. It is postulated that the proton transfer in QRFGLM b6 occurs rapidly, whereas the transfer in notably RWFGLMPG and QWFRLMPG b6 is much slower. This hypothesis could be tested by rigorous molecular dynamics computations.
The photofragment channels for these b ions are tabulated in Table 3-2. In regards to photofragmentation information, mostly direct sequence ions and a tendency especially for \((b_n\text{-NH}_3)\) formation is seen. There appears to be no evidence for internal fragment losses or unaccounted masses. On the other hand, it should be noted that few photofragments are generated here, compared to other peptide systems. The scarcity of photofragment masses also makes it harder to detect permuted sequence ions.

**Conclusions**

In this study, we employed synthetic chemistry, IRMPD spectroscopy and SORI CID to carry out a systematic study on peptide b fragments with the sequence motif QWFGLM. The aim was to determine if amino acid residues with constrained torsional freedom affect macrocycle formation in the corresponding b fragment. The presence of proline in position 2 resulted in appearance of sizeable oxazolone bands for \(QP\text{WFGLM} b_7\) and to a lesser degree, in \(QP\text{FGLM} b_6\). This supports the hypothesis that constraining the flexibility of the backbone can suppress “head-to-tail” cyclization from the N-terminus. On the other hand, insertion of 4-aminomethylbenzoic acid (4AMBz) had little or no effect on macrocycle formation for \(b_6\), with the exception of a weak oxazolone band for \(QWFG(4AMBz)M b_6\), with 4AMBz in the fifth position. It is probable that this is mainly due to a lowering in energy of the oxazolone structure, as a result of a conjugated \(\pi\)-electron system between the benzene and oxazolone rings. An analysis of the photofragmentation channels provided scant evidence for scrambling phenomena, and rather abundant \(\text{NH}_3\) and \(\text{H}_2\text{O}\) losses from b fragments are observed. The SORI CID mass spectra show slightly more evidence for scrambling products, as well as internal fragments, even if this is dependent on the sequence. The vibrational spectra
demonstrate that the formation of macrocycle b structures can be surprisingly robust
and is difficult to disrupt even using residues with limited torsional degrees of freedom.
On the other hand, the prevalent formation of macrocycle structures does not
necessarily result in scrambled sequence ions. This underlines the importance of the
ring opening chemistry, as the barriers to ring opening are dependent on residue and
ring configuration. The substitution of arginine in the model sequence motif QWFGLM
establishes that the presence of arginine can disrupt head-to-tail macrocyclization;
however, this depends on the peptide sequence and position of the arginine residue. It
is postulated here that head-to-tail macrocyclization is impeded due to rapid proton
transfer from the oxazolone ring to the arginine side chain, as the oxazolone ring only
remains labile to nucleophilic attack from the N-terminus as long as it remains
protonated.
Scheme 3-1. Schematic for formation of b₆ fragment, initially leading to oxazolone structure. Hypothesis of study: Limited torsional flexibility of e.g. 4AMBz residue restricts “head-to-tail” isomerization to macrocycle. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. J. Am. Soc. Mass Spectrom. 2012, 23, 475].
Figure 3-1. Overlay of IRMPD spectra of QWFGLMPG b$_6$ and protonated cyclo(QWFGLM). The chemically diagnostic regions are indicated by color-coding: oxazolone CO stretch (red), amide backbone CO stretch (blue), and amide backbone NH bending (green). [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. J. Am. Soc. Mass Spectrom. 2012, 23, 475].
Figure 3-2. Overlay of IRMPD spectra of QPWFGLMPG b7, QPFGLMPG b6 and protonated cyclo(QPFGLM). The chemically diagnostic oxazolone CO stretch region is highlighted in red. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. J. Am. Soc. Mass Spectrom. 2012, 23, 475].
Figure 3-3. Overlay of IRMPD spectra for b_6 fragments with different 4AMB substitutions in the sequence motif QWFGLM. The inset shows a 3x vertical zoom of the 1780-1870 cm\(^{-1}\) region. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. J. Am. Soc. Mass Spectrom. 2012, 23, 475].
Figure 3-4. Reaction scheme showing impeded b₆ ion formation for QWFGL(4AMBz)PG due to bulky 4AMBz residue. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. J. Am. Soc. Mass Spectrom. 2012, 23, 475].

Figure 3-5. Resonantly-stabilized π-electron oxazolone ring structure for QWFG(AMBz)MPG b₆. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. J. Am. Soc. Mass Spectrom. 2012, 23, 475].
Figure 3-6. Comparison of IRMPD spectra of (4AMBz)WFGLMPG b₆ (top) and (4ABz)WFGLMPG b₆ (bottom). The chemically diagnostic oxazolone CO stretch region is highlighted in red. The chemical structures for 4-aminomethylbenzoic acid and 4-aminobenzoic acid are shown. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. J. Am. Soc. Mass Spectrom. 2012, 23, 475].
Figure 3-7. Comparison of SORI CID and IRMPD mass spectra for (4Abz)WFGLMPG b₆ fragment. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. *J. Am. Soc. Mass Spectrom.* 2012, 23, 475].
Figure 3-8. Fragmentation mass spectra for b₆ fragments with different 4AMBz substitutions in the sequence motif QWFGLM under SORI CID and IRMPD. [Figure taken with permission from Tirado, M.; Rutters, J.; Chen, X.; Yeung, A.; van Maarseveen, J.; Eyler, J.; Berden, G.; Oomens, J.; Polfer, N. *J. Am. Soc. Mass Spectrom.* 2012, 23, 475].
Figure 3-9. Overlay of IRMPD spectra for $b_6$ fragments with different Arginine substitutions in the sequence motif QWFGLM. The peach highlighted region pertains to the oxazolone region from 1750-1835 cm$^{-1}$. 
Figure 3-10. Overlay of IRMPD spectra for b5 fragments with different Arginine substitutions in the sequence motif QWFGLM. The peach highlighted region pertains to the oxazolone region from 1750-1825 cm\(^{-1}\) region.
Figure 3-11. Proposed mechanism for locked oxazolone structure.
Table 3-1. Position of arginine residue in $b_n$ ions for arginine whose IRMPD was recorded and putative structure assignment.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Arginine Position</th>
<th>$b_5$</th>
<th>$b_6$</th>
<th>Oxazolone</th>
<th>Macrocycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>RWFGLMPG</td>
<td>1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>QRFGLMPG</td>
<td>2</td>
<td>✔</td>
<td>✔</td>
<td>✔, ✔</td>
<td>✔</td>
</tr>
<tr>
<td>QWRGLMPG</td>
<td>3</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔, ✔</td>
</tr>
<tr>
<td>QWFRGMPG</td>
<td>4</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>QWFGRMPG</td>
<td>5</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
Table 3-2. Photofragmentation channels for $b_5/b_6$ ions of arginated systems.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Photofragments $b_5$</th>
<th>Photofragments $b_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fragment mass (Daltons)</td>
<td>Fragment mass (Daltons)</td>
</tr>
<tr>
<td>RWFGLMPG</td>
<td>791</td>
<td>$b_6$</td>
</tr>
<tr>
<td></td>
<td>326</td>
<td>($b_2$-$NH_3$)</td>
</tr>
<tr>
<td></td>
<td>298</td>
<td>($a_2$-$NH_3$)</td>
</tr>
<tr>
<td></td>
<td>173</td>
<td>$Y_2$</td>
</tr>
<tr>
<td>QRFGLMPG</td>
<td>602</td>
<td>$b_5$</td>
</tr>
<tr>
<td></td>
<td>733</td>
<td>$b_6$</td>
</tr>
<tr>
<td></td>
<td>716</td>
<td>($b_6$-$NH_3$)</td>
</tr>
<tr>
<td></td>
<td>715</td>
<td>($b_6$-$H_2O$)</td>
</tr>
<tr>
<td>QWRGLMPG</td>
<td>641</td>
<td>$b_5$</td>
</tr>
<tr>
<td></td>
<td>613</td>
<td>$a_5$</td>
</tr>
<tr>
<td></td>
<td>772</td>
<td>$b_6$</td>
</tr>
<tr>
<td></td>
<td>755</td>
<td>($b_6$-$NH_3$)</td>
</tr>
<tr>
<td></td>
<td>641</td>
<td>$b_5$</td>
</tr>
<tr>
<td></td>
<td>624</td>
<td>($b_5$-$NH_3$)</td>
</tr>
<tr>
<td></td>
<td>613</td>
<td>$a_5$</td>
</tr>
<tr>
<td>QWFRGMPG</td>
<td>862</td>
<td>$b_6$</td>
</tr>
<tr>
<td></td>
<td>845</td>
<td>($b_6$-$NH_3$)</td>
</tr>
<tr>
<td></td>
<td>601</td>
<td>($b_4$-$NH_3$)</td>
</tr>
<tr>
<td>QWFRGMPG</td>
<td>675</td>
<td>$b_5$</td>
</tr>
<tr>
<td></td>
<td>658</td>
<td>($b_5$-$NH_3$)</td>
</tr>
</tbody>
</table>
CHAPTER 4
DEFYING ENTROPY - FORMING LARGE HEAD-TO-TAIL MACROCYCLES IN THE GAS PHASE

Background

Chapter 3 embarked upon addressing the question if torsional restriction of a peptide chain can impede macrocyclization. This Chapter addresses the question of whether extending the chain length and varying residue position affects macrocyclization and scrambling. Earlier work by our group has demonstrated that for a series of oligoglycine b fragments (b₂-b₈) generated by collision-induced dissociation (CID), whereas smaller b fragments prefer oxazolone structures. Larger b fragments exhibited a progressively higher propensity to form macrocycles by “head-to-tail” cyclization. In addition, Van Stipdonk and co-workers investigated the influence of peptide size upon scrambling. The size ranges examined were from tetrapeptides to decapeptides. Their findings indicate that as the length of the peptide chain increases, there is a higher susceptibility for scrambling.

These trends have been verified for select peptide systems, such as the b₉ ions of Tyr(Ala)₉, (Ala)₄, Tyr(Ala)₅, and (Ala)₅TyrAla by Harrison, who’s data demonstrates that these systems are considered to be exclusively macrocyclic, as well as for tandem mass spectra from proteomics data. It is not surprising that mid-sized b fragments more readily adopt “head-to-tail” macrocycles than smaller b fragments. Analogously, in solution-phase synthesis of cyclic peptides, mid-sized cyclic peptides of 5-7 amino acid residues are generally more facile to form than smaller cyclic peptides. Nonetheless, larger cyclic peptides are also difficult to form, due to the entropic penalty of cyclic vs. linear structures. Similarly, in the gas phase there should be a limit to the size of “head-to-tail” macrocycle structures that can be generated. This Chapter aims to
address the question whether the trends for macrocycle formation as a function of sequence length are also observed for other sequence motifs, such as TyrAlaGly, AlaTyrGly, and GlyAlaTyr, for a series of $b_n$ ($n=4$-15) fragments. Importantly, this study will address whether there is a “reverse” transition from macrocycles to oxazolones for larger $b$ fragments. Structural characterization is performed by infrared multiple-photon dissociation (IRMPD) spectroscopy using the free electron laser FELIX. These results are complemented by gas-phase HDX and SORI-CID experiments to deduce if macrocyclization necessarily implies scrambling of the sequence.

**Sample Preparation**

All peptides were synthesized with conventional FMOC solid-phase synthesis methods using 9-fluroenylmethoxycarbonyl (FMOC)-Gly loaded resin at the University of Florida.\textsuperscript{137,138} This same method employed as discussed in the previous Chapter. However, in order to realize this systematic experiment, increasing permuted triads sequence motif, (GlyAlaTyr)$_n$, $n=3$-21 were synthesized using an the 432 A Synergy Peptide Synthesizer. Table 4-1 corresponds to the peptides synthesized and characteristic $b$ fragments analyzed. The peptide analyzer afforded us the ability to synthesize 25-$\mu$mol scale of peptides. This corresponds to a quarter the scale of what is typically synthesized on the CEM microwave discover system. Yields of greater than 90 percent were obtained utilizing this peptide synthesizer. The cyclic peptides were synthesized with the following modifications. First, the linear peptides YAGYAG was made with solid phase synthesis on a Wang resin. The linear peptide was dissolved in DMF there were solubility issues with the THF at concentrations of less than $10^{-3}$ mol/L. Next, 2.2 equivalents of 2-(1H-7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyl uronium hexafluorophosphate methanaminium (HATU) and 2.2 equivalents of 3H-1,2,3-
triazolo[4,5-b]pyridin-3-ol (HOAT) were added into the solution. Then, an 10 equivalents of N,N-diisopropylethylamine (DIPEA) was added to the solution. Upon addition of DIPEA, a color change to yellow was observed. The reaction solution was kept at room temperature with stirring for 24h. The DMF was lyophilized to obtain a final solid yellow product. The crude product was not purified. The power of mass isolation of the FTICR was exploited to isolate the cyclic peptide. A robust ion signal of 300 counts was obtained from ESI on the FTICR.

**Mass Spectrometry and Infrared Photodissociation Spectroscopy**

As described in Chapter 3, Infrared spectra were obtained by IRMPD spectroscopy using the FTICR mass spectrometer coupled to the beamline of the free-electron-laser user facility (FELIX) in The Netherlands. Ions were produced by electrospray ionization (ESI) and accumulated in a hexapole ion trap. Peptide solutions at 100 μM were prepared, composed of 50/50/2 (v/v/v) water/methanol/formic acid. Then, ions were injected into a homebuilt FTICR mass spectrometer. CID fragments were generated by nozzle-skimmer dissociation in the high pressure region at the inlet of the mass spectrometer. Ions were accumulated in a linear hexapole trap held at a background pressure of about 10^-3 Torr, where the trapping voltages can be adjusted to further enhance the formation of a specific CID ion. The ion of interest was isolated and then irradiated with FELIX for 4 s at a 5 Hz repetition rate. In these experiments, the wavelength of FELIX is tuned between 5.0μm and 8.3μm and the macropulse energy is approximately 30mJ. After irradiation, parent and fragment ion intensities were recorded. Four mass spectra were averaged at each wavelength point. The IRMPD spectrum was obtained by monitoring the IRMPD yield as a function of wavelength (1300–1975 cm⁻¹). The yield is given by the following equation: Yield = -\ln [1-
\[ \frac{\sum \text{Photofragments}}{\sum \text{All Ions}} \]. The yield is further normalized linearly with the relative FELIX laser power at each wavelength step. Data acquisition and instrument control was accomplished using a modified version of the data system and software of Mize et al.\textsuperscript{155}

**SORI CID Experiments**

Complementary sustained off-resonance irradiation collision-induced dissociation (SORI CID) were carried out in a commercial FTICR mass spectrometer (4.7 T actively shielded APEX II, Bruker Daltonics, Billerica, MA). The b CID product ions were generated by “nozzle-skimmer” dissociation in the ESI source. These b ions were then subjected to SORI CID in the ICR cell using a frequency offset of -2.0 kHz (relative to the precursor ion’s cyclotron frequency) and a nitrogen gas pulse (<10\textsuperscript{-7} Torr).

**Mass Spectrometry and Hydrogen/Deuterium Exchange**

The hydrogen/deuterium exchange (HDX) experiments were carried out at the University of Florida using a commercially available Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (4.7 T actively shielded APEX II, Bruker Daltonics, Billerica, MA) in Dr. David Powell’s lab. The protonated peptide precursor ions were activated by —nozzle-skimmer dissociation in the ESI source region by adjusting the voltage drop between the metal-plated glass capillary and the first skimmer. The CID product ions were then accumulated in the hexapole (3–4 s), prior to transfer to the ICR cell. A frequency sweep was performed in the FTICR experimental sequence to eject other ions and to mass-isolate each desired b product ion (b\textsubscript{5}). The mass-to-charge ratios of b\textsubscript{5} are 526 respectively. The monoisotopic peak of each species was then subjected to gas-phase hydrogen/deuterium exchange (HDX) with CH\textsubscript{3}OD in the ICR cell for different time periods. CH\textsubscript{3}OD was leaked into the vacuum chamber using a Varian leak valve to attain a constant pressure of 1x10\textsuperscript{-8} Torr. Note
that CH$_3$OD was degassed using several freeze-thaw cycles before introduction into the mass spectrometer. Mass spectra with different HDX times were recorded. The data are represented here by plotting the natural logarithm of the ratio of d$_0$ divided by the sum of all ions, ln[d$_0$/Σdn], as a function of the HDX time, as described previously.$^{156}$ The depletion of d$_0$ is also represented as a percentage, which is plotted on a natural logarithm scale. Because the exchange reagent is considered to be in great excess of the analyte, the H/D exchange reaction can be approximated as a first-order reaction in the analyte concentration. Linear fits were performed to determine the reaction kinetics.

**Results and Discussion: Rationale for Sequence Motifs**

Previous results on oligoglycine b$_2$-b$_8$ had shown a size effect for macrocycle formation, culminating in b$_8$, which was found to be exclusively macrocyclic.$^{70}$ Initial experiments aimed to generate larger bn product ions (n > 8) in this series from deca and dodecaglycine; however, these studies were not successful, as the corresponding b ions could not be formed, instead resulting in dehydrated b ions, such as b$_9$-H$_2$O and b$_{10}$-H$_2$O. Thus, other sequence motifs were investigated that would allow a wider series of b$_n$ ions. Peptides with repeat units of (GlyAlaTyr)$_n$ (TyrAlaGly)$_n$ and (AlaGlyTyr)$_n$ (where n=2-4), and terminated by ProGly at the C-terminal side, were found to behave well in CID conditions. The insertion of proline in the sequence motif resulted in abundant amide bond cleavage on its N-terminal side, and hence abundant formation of b$_6$, b$_9$ and b$_{12}$ product ions. The smaller fragment b$_4$ could also be generated from GlyAlaTyrGlyAlaTyrProGly, and the other corresponding sequence motifs. Solubility of the peptides became an issue with increasing chain length. While b$_{15}$ from (GlyAlaTyr)$_5$ProGly could still be generated at a low signal intensity, it proved quite
arduous to generate larger $b_n$ fragments on the FTCIR, such as $b_{18}$ and $b_{21}$, and due to time constraints these measurements could not be carried out.

**IR Spectroscopy for $b_4$ Ions**

Figure 4-1 shows the IRMPD spectra for the $b_4$ ions with sequence motifs TyrAlaGlyTyr, GlyAlaTyrGly and AlaTyrGlyAla. It is clear from the high-frequency bands in the range from 1780-1950 cm$^{-1}$ that a number of abundant oxazolone structures are present. Previous IRMPD spectroscopy results on $b_4$ ions with the sequence motif TyrGlyGlyPhe had shown that the oxazolone CO stretch position depends on the site of protonation, namely the oxazolone ring N and the N-terminal amino group, and the secondary structure. When the N-terminus is protonated and the oxazolone CO is hydrogen-bonded to it, the observed CO frequency is as low as 1780 cm$^{-1}$. Conversely, when attached to the oxazolone ring is protonated and the oxazolone CO is devoid of hydrogen-bonding, the observed CO frequency appears at 1930 cm$^{-1}$. Lastly, protonation occurs on the N-terminus, but the oxazolone CO is not hydrogen-bonded, the band is predicted around 1890 cm$^{-1}$. Similar band positions are observed here at 1780 and 1930 cm$^{-1}$ strongly suggesting that similar structural assignments can be made. Figure 4-2 highlights these band positions more clearly for the $b_4$ ion of sequence motif GlyAlaTyrGly. Comparing the IRMPD spectra of the $b_4$ systems GlyAlaTyrGly and AlaTyrTyrAla on the one hand, and TyrAlaGlyTyr on the other, one notices that the oxazolone CO stretch bands for TyrAlaGlyTyr are reduced with respect to the other $b_4$ ions. This may pertain to the less efficient photodissociation of TyrAlaGlyTyr, due to a higher dissociation barrier, or the presence of another structure, such as the “head-to-tail” macrocycle. Despite this caveat, it is clear that in all these $b_4$ ions, oxazolone structures are abundantly present.
IR Spectroscopy for b₆ Ions

Figure 4-3 shows the corresponding IRMPD spectra for the b₆ fragment ions with sequence motifs (GlyAlaTyr)ₙ, (AlaGlyTyr)ₙ, and (TyrAlaGly)ₙ. Strikingly, there is no evidence of oxazolone bands in any of the b₆ ions suggesting that “head-to-tail” macrocycle structures are formed. In principle, the IRMPD spectra for the b₆ ion for (GlyAlaTyr)ₙ & (AlaGlyTyr)ₙ should be identical—if the b₆ ion isomerizes to a macrocycle structure. Conversely, the b₆ spectrum for the (TyrAlaGly)ₙ system should bear a disparate resemblance from that of (GlyAlaTyr)ₙ & (AlaGlyTyr)ₙ. Indeed, the IRMPD spectra of these systems reveals that both (GlyAlaTyr)ₙ and (AlaGlyTyr)ₙ exhibit greater structural overlap, compared to (TyrAlaGly)ₙ. There are subtle (but reproducible) differences between the IRMPD spectra for (GlyAlaTyr)ₙ and (AlaGlyTyr)ₙ, b₆ indicates that the gas-phase re-arrangement mechanisms is more complex than we understand at this moment. Similarly, if one re-visits figure 3-1 (Chapter 3), where the cyclic reference system (cyclic) QWFGLM was compared to the b₆ ion from QWFGLMPG, there were subtle differences in 1600 cm⁻¹ region. The IRMPD spectra show presence of these bands. In fact, Figure 3-2 the prolineated system where oxazolone bands are present for b₆ and b₇ ions, there’s an absence of the 1600 cm⁻¹ bands. Tyrosine coupled with oxazolone formation could be working in tandem to bring out this spectral feature at 1600 cm⁻¹. Similarly, the 1600 cm⁻¹ band is displayed in b₂-b₄ ions for Leu-Enkephalin where Chen et al. investigated these structures via HDX and IRMPD spectroscopy.⁷¹

In order to test the hypothesis that macrocycle structures were formed for these b₆ ions, control experiments were carried out for TyrAlaGly b₆, which are shown in Figure 4-4. A comparison between the IRMPD spectra of TyrAlaGly b₆ and the synthetically made cyclic peptide shows that both look almost identical, strongly supporting that
“head-to-tail” macrocycle structures are formed. Moreover, the IRMPD spectrum for N-terminal acetylated TyrAlaGly $b_6$, Ac-TyrAlaGlyTyrAlaGlyProGly $b_6$, displays an intense band at 1900 cm$^{-1}$, characteristic for an oxazolone structure CO stretch. This demonstrates that protection of the N-terminus results in the expected oxazolone structure. In summary, the vibrational spectra indicate that $b_6$ ions, in marked contrast to $b_4$ ions, adopt macrocycle structures. This occurs independently of the exact sequence order of the residues glycine, alanine and tyrosine, but rather depends on the sequence length.

**Infrared Spectroscopy Results for GlyAlaTyr $b_4$-$b_{12}$ Ions**

For the sequence motif GlyAlaTyr, the structures of larger b fragments were also interrogated by IR spectroscopy. Figure 4-5 compares the mid-IR IRMPD spectra for the previously shown $b_4$ and $b_6$, as well as the larger $b_9$ and $b_{12}$, and the control experiment, N-acetylated $b_{14}$. Similarly to $b_6$, neither $b_9$ nor $b_{12}$, show any bands in the region diagnostic for oxazolone CO stretches. To show this more clearly, the region from 1780-1970 cm$^{-1}$ is magnified 5x showing no evidence even for weaker features. Conversely, in the control experiment for Ac-$b_{14}$, a weak, yet distinct, band is visible at 1900 cm$^{-1}$. It is clear that the relative intensity of the oxazolone CO stretch decreases with increasing length of the b fragment, as this represents a single IR moiety in a fairly large molecule. Nonetheless, the successful detection of this band for the even larger Ac-$b_{14}$ suggests that this should not constrain the photodissociation of $b_9$ and/or $b_{12}$. In other words, if oxazolone structures had been present for $b_9$ and $b_{12}$, these should have been apparent in the IR spectra.

Other insightful trends are seen when going from $b_4$ to $b_{12}$. Table 4-2 shows that as the sequence size increases a gradual red shift is observed for the Amide I/II
vibrational frequencies, as well as a broadening of these bands. This may be related to the conformational flexibility of the large macrocycle structures (i.e., $b_{12}$), which are likely greater than for smaller "head-to-tail" macrocycles (i.e., $b_6$). Another intriguing observation concerns the decreasing intensity of the 1600-1625 cm$^{-1}$ band upon going to larger $b$ ions. This band is presumably associated with tyrosine side-chain deformation modes. It is possible that these modes couple with the amide I modes for larger $b$ ions, due to the red shift of those latter modes.

The IRMPD spectrum of the even larger $b_{15}$ fragment ion is shown in Figure 4-6. Here, only the region important for oxazolone CO stretches was scanned, again showing no evidence for any spectral features. For $b_{15}$, it is not clear at this time whether the failure to detect a band in this region is due to an absence of oxazolone structures or an inability to photodissociate the ion, since no control acetylation experiment was performed on a larger $b$ ion.

Collectively, the IRPMPD spectra paint a counterintuitive picture of what one would expect for larger $b_n$ ions. It appears that very large macrocycles, with up to 12 (and possibly 15) residues, are readily formed in the gas phase in CID conditions. Such a behavior is not observed in solution. One important difference in both experiments is that in the gas phase the ions are heated up to substantial internal energies (600-800 K), allowing the ions to explore a large range of conformational space. Moreover, the isomerization chemistry between oxazolone and macrocycle structures is specific to the gas phase, and thus the trends may be completely different to what is expected from solution.
In summary, macrocycles seem to be formed abundantly over wide ranges of b ion sizes, from 6-12 residues. For the purpose of peptide sequencing by “bottom-up” proteomics, this covers most of the range of interest. The most common tryptic digest peptides have chain lengths of 12-16 residues, thus often giving rise to b ions in the size range of 6-12 residues.

**Sequence Scrambling/Permutations from SORI-CID**

The spectroscopic results clearly established and corroborated the presence of macrocyclic b structures. We also investigated whether the presence of macrocycles necessarily implies that non-direct/permuted/scrambled sequence ions are observed. Figure 4-7 compares the MS³ SORI-CID mass spectra for the three b₆ fragments. Clearly, the data exhibits much overlap in the sequential product ions that are made. Moreover, scrambling is directly confirmed by the fact that internal residue losses are observed for all three b₆ ions. The obvious consequence of these findings is that macrocyclization may lead to scrambling thereby obfuscating correct sequence.

Sustained off-resonance irradiation collision-induced dissociation (SORI CID) for these b₆ ions in the ICR cell also confirms sequence scrambling identification. Recently, our group has evaluated the occurrence of sequence scrambling statistically for a proteomics study involving >4500 identified precursor ions matched to tryptic digest peptides. It was shown that the propensity for sequence permutation increases with the length of the tryptic peptide. This published data further bolsters the fact that with size scrambling increases.¹¹⁷

In principal, the spectra and fragmentation data for the macrocyclic structures of (GlyAlaTyr)_n & (AlaGlyTyr)_n should have greater overlap than that of the (TyrAlaGly)_n system. When juxtaposing the amide II band around 1530 cm⁻¹ for (GlyAlaTyr)_n &
(AlaGlyTyr)\textsubscript{n}, there exists greater similarity to that of the (TyrAlaGly)\textsubscript{n} system. Figure 4-8 better illustrates our “geminal” vs. “fraternal” structural hypothesis for macrocyclization. Interestingly, the product ion spectra from the SORI-CID experiments for permuted \( b_5 \) ions of (GlyAlaTyr)\textsubscript{n} (Figure 4-7) shows greater structural similarity exists between (GlyAlaTyr)\textsubscript{n} & (AlaGlyTyr)\textsubscript{n} at the higher masses (> 318 Da) than the (TyrAlaGly)\textsubscript{n} system. Whereas at the lower masses (< 318 Da) there exists great overlap between (TyrAlaGly)\textsubscript{n} & (GlyAlaTyr)\textsubscript{n}. This might possibly be ascribed to the fact that the precursor ion was completely depleted away for (TyrAlaGly)\textsubscript{n}.

**IR Spectroscopy for \( b_5 \) Ions**

A number of previous studies had shown evidence for macrocyclic structures for \( b_5 \) ions. In an IR spectroscopic study by Maitre and co-workers, the \( b_5 \) fragment from (Gly)\textsubscript{5}R was shown to exclusively give rise to macrocycle structures.\textsuperscript{157} In a complementary IRMPD spectroscopy and HDX study on oligoglycine \( b \) fragments, Chen et al. showed that the same \( b_5 \) ion as above, generated from octaglycine, was composed of a mixture of oxazolone and macrocycle structures.\textsuperscript{70} For the related sequence \( b_5 \) from YAGFL-NH\textsubscript{2} macrocycle structures were confirmed.\textsuperscript{120}

Figure 4-9 displays a side by side comparison of the \( b_5 \) IRMPD spectra for the (AlaGlyTyr)\textsubscript{n} and (TyrAlaGly)\textsubscript{n} system. Surveying the range between 1770 and 2000 cm\(^{-1}\), associated with the strongly active C0 stretch of the oxazolone structure shows that no bands are observed. Moreover, multiple scans were performed in this oxazolone region with full power and extended irradiation times (4-10sec) to ensure that no oxazolone bands are present. Two main experimental features centered at 1500 cm\(^{-1}\) and 1700 cm\(^{-1}\) are observed for these systems. In the (AlaGlyTyr)\textsubscript{n} system the presence of the vibrational frequency around 1600 cm\(^{-1}\) is blue shifted 50 cm\(^{-1}\) and better resolved.
for the \( b_5 \) ion. In the \((\text{TyrAlaGly})_n\) system there’s an absence of this spectral feature around 1600 cm\(^{-1}\). An HDX exchange experiment was performed for the \( b_5 \) ion of the \((\text{GlyAlaTyr})_n\) system. (Figure 4-10) The presence of two ion populations may be inferred from the two slow kinetic rates obtained from the mass isolated \( b_5 \) ions. Collectively, the IRMPD data and HDX data are in agreement with earlier results from our group. Namely that the two slow exchanging kinetic rates could correspond to different protonation sites or structural variants (e.g. cis/trans)\(^{70,136}\).

**Summary**

In this investigation, a series of \( b_n \) (\( n=4\text{-}15 \)) fragments were studied. The smaller \( b_4 \) ions display sizeable oxazolone bands at ~1900 cm\(^{-1}\), as well as at other characteristic oxazolone positions, confirming the abundant presence of oxazolones. The larger \( b \) ions (\( b_6, b_9, b_{12}, b_{15} \)) are characterized by an absence of these diagnostic bands. A control experiment for a synthetically made cyclic peptide confirms the observation of the macrocycle structure for \( b_6 \), and strongly suggests the exclusive presence of macrocycle structures for larger \( b \) ions (up to 15 amino acid residues). These results, in addition to previous findings, indicate that oxazolone \( b \) fragment structures are mainly confined to smaller \( b \) ions, whereas larger \( b \) fragments more readily adopt macrocycles.
Figure 4-1. IRMPD spectra of the $b_4$ ions with the sequence motifs for TyrAlaGly, AlaTyrGly, and GlyAlaTyr.
Figure 4-2. IRMPD labeled bands multiple b₄ oxazolone structures.
Figure 4-3. IRMPD spectra of the $b_6$ ions with the sequence motifs for TyrAlaGly, AlaTyrGly, and GlyAlaTyr.
Figure 4-4. Comparison of IRMPD spectra of cyclo(YAGYAG)H+, b6 with sequence motif YAG, and b6 from the N-terminally acetylated peptide with sequence motif YAG. The two CID products were made from the precursor ions (YAG)2ProGly and Ac-(YAG)2ProGly, respectively.
Figure 4-5. IRMPD spectra of a series of b ions with the sequence motif GlyAlaTyr: b₄, b₆, b₉, b₁₂, and N-acetylated b₁₄.
Figure 4-6. IRMPD spectrum for $b_{15}$ from (GlyAlaTyr)$_5$ProGly.
Figure 4-7. SORI CID mass spectra of $b_6$ ions with sequence motifs (AlaTyrGly)$_2$, (GlyAlaTyr)$_2$, and (TyrAlaGly)$_2$. 
Figure 4-8. “Geminal” vs. “fraternal” structural hypothesis for macrocyclization.
Figure 4-9. IRMPD spectra for $b_5$ from $(\text{AlaTyrGly})_2\text{ProGly} \& (\text{TyrAlaGly})_2\text{ProGly}$.

Figure 4-10. HDX data for $b_5$ ion from $(\text{GlyAlaTyrGlyAla})_2\text{ProGly}$. 
Table 4-1. Peptides synthesized and characteristic b fragments analyzed.

<table>
<thead>
<tr>
<th>Peptide Sequence</th>
<th>Precursor Ion [M+H]^+</th>
<th>b Fragment (m/z)</th>
<th>b Fragment size</th>
</tr>
</thead>
<tbody>
<tr>
<td>TyrAlaGlyProGly</td>
<td>464</td>
<td>292</td>
<td>3</td>
</tr>
<tr>
<td>(YAG)_2PG</td>
<td>755</td>
<td>583</td>
<td>6</td>
</tr>
<tr>
<td>(YAG)_3PG</td>
<td>1046</td>
<td>874</td>
<td>9</td>
</tr>
<tr>
<td>(YAG)_4PG</td>
<td>1337</td>
<td>1165</td>
<td>12</td>
</tr>
<tr>
<td>(YAG)_5PG</td>
<td>1628</td>
<td>1456</td>
<td>15</td>
</tr>
<tr>
<td>(YAG)_6PG</td>
<td>1919</td>
<td>1747</td>
<td>18</td>
</tr>
<tr>
<td>(YAG)_7PG</td>
<td>2210</td>
<td>2038</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 4-2. Gradual red shift in Amide I/II bands with size increase.

<table>
<thead>
<tr>
<th>b Ions</th>
<th>Amide II (cm(^{-1}))</th>
<th>Amide I (cm(^{-1}))</th>
<th>Unassigned (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1528</td>
<td>1729</td>
<td>1623,1593,1613</td>
</tr>
<tr>
<td>6</td>
<td>1522</td>
<td>1693</td>
<td>1617</td>
</tr>
<tr>
<td>9</td>
<td>1515</td>
<td>1680</td>
<td>-------------------------</td>
</tr>
<tr>
<td>12</td>
<td>1513</td>
<td>1678</td>
<td>-------------------------</td>
</tr>
</tbody>
</table>
Table 4-3. Assignment of product ions. Note that $a^* = a$-$NH_3$. Internal eliminations of amino acid residues are labeled with parenthesis to indicate potential sequence permutation.

<table>
<thead>
<tr>
<th>Fragment Mass (Daltons)</th>
<th>(AlaTyrGly)$_2$</th>
<th>(GlyAlaTyr)$_2$</th>
<th>(TyrAlaGly)$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>136</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>148</td>
<td></td>
<td>a$_3$-45-Ala</td>
<td>a$_3$-45-Ala</td>
</tr>
<tr>
<td>188</td>
<td>unassigned</td>
<td>unassigned</td>
<td>unassigned</td>
</tr>
<tr>
<td>219</td>
<td></td>
<td>a$_5$*-CO</td>
<td>a$_5$*-CO</td>
</tr>
<tr>
<td>247</td>
<td>a$_3^*$</td>
<td>a$_3^*$</td>
<td>a$_3^*$</td>
</tr>
<tr>
<td>314</td>
<td>385-Ala</td>
<td>385-Ala</td>
<td></td>
</tr>
<tr>
<td>318</td>
<td>a$_4^*$</td>
<td>a$_6^*$-Tyr</td>
<td>a$_6^*$-Tyr</td>
</tr>
<tr>
<td>328</td>
<td>491-Tyr</td>
<td>491-Tyr</td>
<td>491-Tyr</td>
</tr>
<tr>
<td>349</td>
<td>b$_6$-Ala-Tyr</td>
<td>b$_4$</td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>unassigned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>410</td>
<td>a$_6^*$-Ala-Gly</td>
<td>a$_6^*$-Ala-Gly</td>
<td>a$_4^*$</td>
</tr>
<tr>
<td>420</td>
<td>b$_6$-Tyr</td>
<td>b$_5$</td>
<td>b$_6$-Tyr</td>
</tr>
<tr>
<td>463</td>
<td>a$_5^*$-CO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>467</td>
<td>a$_6^*$-Ala</td>
<td>a$_6^*$-Ala</td>
<td></td>
</tr>
<tr>
<td>481</td>
<td></td>
<td></td>
<td>a$_5^*$</td>
</tr>
<tr>
<td>491</td>
<td>b$_6$-92</td>
<td>b$_6$-92</td>
<td></td>
</tr>
<tr>
<td>510</td>
<td>a$_6^*$-CO</td>
<td>a$_6^*$-CO</td>
<td>a$_6^*$-CO</td>
</tr>
<tr>
<td>512</td>
<td>b$_6$-Ala</td>
<td></td>
<td></td>
</tr>
<tr>
<td>538</td>
<td>a$_6^*$</td>
<td>a$_6^*$</td>
<td>a$_6^*$</td>
</tr>
<tr>
<td>547</td>
<td>unassigned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>a$_6$</td>
<td></td>
<td>a$_6$</td>
</tr>
<tr>
<td>565</td>
<td>b$_6$-H$_2$O</td>
<td>b$_6$-H$_2$O</td>
<td></td>
</tr>
<tr>
<td>566</td>
<td>b$_6$-NH$_3$</td>
<td>b$_6$-NH$_3$</td>
<td>b$_6$-NH$_3$</td>
</tr>
<tr>
<td>583</td>
<td>b$_6$</td>
<td></td>
<td>b$_6$</td>
</tr>
</tbody>
</table>
CHAPTER 5
ANALYSIS OF SOLUTION PHASE PEPTIDE PRODUCTS BY IRMPD SPECTROSCOPY

Background

Thus far, this dissertation has discussed how synthetic strategies can aid in the characterization of gas-phase fragmentation of peptides. Another viewpoint is that IRMPD spectroscopy coupled with mass spectrometry is a sophisticated platform to investigate analytes in complex mixtures. The role of the mass spectrometer is to separate many of the constituents by mass(-to-charge) and allow sensitive detection, while IRMPD spectroscopy permits chemical characterization. In cases, where reaction products in solution cannot be purified sufficiently, the technique may hence prove to be a useful alternative to nuclear magnetic resonance (NMR). In this Chapter, we focus on the characterization of solution-phase reaction products, using coupling reagents to induce cyclization from the N-terminus and from side chains. The products are first analyzed by liquid chromatography mass spectrometry (LCMS). Isolated LC fractions are further characterized by IRMPD spectroscopy. The sequence motif that is considered here is QWFGLG, which incorporates an N-terminal glutamine residue. The presence of this residue introduces other side-chain-mediated reactions, which will be reviewed hereafter.

Reaction Pathways Involving Amide and Acid Side Chains

Amide and acid side chains can undergo a number of cyclization reactions with the concomitant elimination of water or ammonia. Some of these processes have been observed in solution and the gas phase. For example, Jonsson et al. discuss how Gln or Asn residues can undergo deamidation. The deamidation reaction proceeds via cyclic imide intermediates, i.e., glutarimide for Gln, succinimide for Asn (Scheme 5-1). The
imides are formed via nucleophilic attack by the peptidyl amide nitrogen on the side-chain carbonyl group with concomitant elimination of NH₃. They propose the involvement of a cyclic imide/isoimide in the gas-phase cleavage of the Gln-Gly peptide bond. In addition, they suggest that the bacterial degradation of acidic proline-rich proteins (PRPs) occurs via a similar solution phase mechanism. Glutamine also can undergo other deamidation pathways in the gas phase. Baldwin et al. reported the unusual collision-induced fragmentation of peptides having N-terminal glutamine. One of these glutamine-containing peptides was isolated from an endoproteinase Lys-C digest of the scrapie prion protein (PrP 27-30) with the predicted sequence QHTVTITTK. Their mass spectra show predominant N-terminal a, b, and, to a lesser extent, c, sequence ion series at 17 u less than the predicted masses. They interpreted this result as glutamine losing ammonia to give pyroglutamic acid, a reaction that parallels the commonly observed solution-phase process. (Scheme 5-2) N-terminal glutamine residue and pyroglutamic acid residue could be identified by the presence of two closely spaced chromatographic peaks which give MH⁺ separated by 17 u by fast atom bombardment or liquid secondary ion mass spectrometry. Moreover, they discuss that identification of N-terminal glutamine and N-terminal pyroglutamic acid peptides would aid in interpretation of mass spectra because the N-terminal fragments would exhibit 17-u shifts whereas C-terminal fragments would have identical masses. However, when glutamine or glutamic acid is situated at the N-terminus other chemistries have been reported in the gas phase. Gaseous protonated glutamic acid readily undergoes dehydration and gaseous protonated glutamine undergoes deamination. Peptides with glutamine as the N-terminal amino acid are known to
undergo partial deamination in solution, which converts the terminal glutamine residue into the cyclic pyroglutamyl residue (2-pyrrolidinone-5-carboxylic acid).\textsuperscript{162} Also, these protonated pyroglutamic acid (H—Pyr—OH, 2-pyrrolidone-5-carboxylic acid) are similar to the non-enzymatic and enzymatic reactions observed in solution.\textsuperscript{161} Harrison studies the gas phase behavior of several dipeptide and tripeptide ions with glutamine on the N terminus. He discusses how protonated glutamine fragments largely by the ammonia elimination channel via the lowest internal energies. Subsequently, this loss leads to the cyclization of the protonated pyroglutamic acid. Moreover, he discusses how the breakdown graphs for protonated glutamic acid exhibit dominant dehydration channel at low collision energies from the parent ion.\textsuperscript{163} Obviously, discerning whether pyroglutamic acid arises from the dehydration mass channel of glutamic acid or deamination mass channel of glutamine presents an interesting question not only regards to the fundamentals but correct residue assignment of the sequence of interest.\textsuperscript{163} Identification of these peptides becomes complicated. Wysocki discusses how MS/MS analysis of glutamine and glutamic acid requires derivatization of these amino acids. Moreover, the fragmentation of glutamine during CID gives a fragment peak at the same m/z value as pyroglutamic acid, making impossible the distinction of glutamine and pyroglutamic acid.\textsuperscript{164} Glutamine, glutamic acid and pyroglutamic acid can be distinguished by running the MS/MS in multiple-reaction monitoring (MRM) mode with a neutral loss of 73 Da for glutamine and 102 Da for glutamic and pyroglutamic acid.\textsuperscript{165} However, Wysocki discusses that the fragmentation mechanism for the neutral loss of 73 Da was nebulous. Wysocki studies the fragmentation mechanism involved by disseminating the CID spectra of derivatized glutamine, \textsuperscript{15}N-glutamine, \textsuperscript{2}H\textsubscript{5}-glutamine...
and $^2$H$_6$-isobutanol derivatized glutamine. In addition, other chemistry can be manifested when glutamine is situated on the N terminal position. Stein et al. discuss dehydration versus deamination pathways of N-terminal glutamine in collision induced dissociation. They investigate the CID spectra by changing the collision energy in a triple-quadrupole mass spectrometer. They perform a systematic analysis ions; by studying longer peptides which vary the size of the peptides from 2 to 15 amino acids by using leucine residues as representative medium size amino acid with no reactive functional groups. In addition, they demonstrate for the sequence QLLLPOLLKL and QLLLPOLLLR show that the singly protonated ions undergo loss of ammonia and to a smaller extent loss of water, whereas the doubly protonated ions undergo predominant loss of water. They confirm this trend by analyzing an MS/MS spectra library of peptides derived from tryptic digests of yeast. In the absence of mobile protons, glutamine deamination is the most rapid neutral loss process. Scheme 5-3 presents the two elimination channels for glutamine. In this Chapter, a systematic study of the effect of glutamine on the N terminus of an octapeptide is discussed. Head-to-tail cyclization is presented for the b$_6$ sequence motif QWFGLG. The experimental methods employed include IRMPD and synthesis. One goal of the present work was to determine the extent to which these dehydration/deamination reactions are carried over glutamine-containing peptides in solution phase synthesis. In addition, we set out to address the question if IRMPD spectroscopy could differentiate between head to tail cyclization of the b$_6$ sequence motif QWFGLG (m689), pyroglutamate structure (m689), or pyroglutamic acid structure(m690). In addition, other techniques such molecular
mechanics calculations, and synthetic strategies are utilized to better elucidate and corroborate the correct chemical structures.

**Synthesis of Linear Peptides**

The linear peptide Q(Trt)W(Boc)FGLG was synthesized with conventional FMOC solid-phase synthesis methods using 9-fluorenylmethoxycarbonyl (FMOC)-loaded trityl resin at the University of Florida.\(^{137,138}\) The side chain protecting groups of glutamine and tryptophan were retained on the linear peptide in order to ensure that these side chains would not participate in the cyclization step of the synthesis. The peptide resin was cleaved with AcOH/TFE/DCM (2:2:6).\(^{166-168}\) Post-cleavage, the peptides were dissolved in acetonitrile and freeze-dried overnight. The same procedure was performed for the synthesis of N-terminal and side chain protected K(Boc)W (Boc)FGLG and (Boc)KW(Boc)FGLG. (Pyr)WFGLGPG was synthesized with (FMOC)-Gly loaded resin on the CEM Discover system. Iridescent purple crystals were obtained as product. The products were confirmed by electrospray ionization (ESI) in Fourier transform ion cyclotron resonance (FTICR) mass spectrometer in Dr.Powell's laboratory. Peptide solutions of 100 μM were prepared, composed of 50/50/1 (v/v/v) water/methanol/acetic acid.

**Synthesis of Head-to-Tail Cyclic Peptides**

DIPEA (19.3 mg, 0.3mmol, 25uL), HATU (57 mg, 0.15 mmol, 1.5eq) and HOAt (20 mg, 0.15 mmol, 1.5 eq) were dissolved in 100 ml THF. Peptide Q(Trt)W(Boc)FGLG (105mg, 0.1 mmol, 1eq) was added and the reaction mixture was stirred for 18h. Upon confirmation from LCMS (Thermofinnigan LXQ linear ion trap) that linear product converted to cyclic peptide, the reaction mixture was concentrated in vacuo. The crude product was redissolved in 50 ml EtOAc and the organic solution was washed three
times with 1M KHSO₄ (50 ml), three times with sat. NaHCO₃ (50 ml) and once with brine. The organic solution was dried over Na₂SO₄ and concentrated. This yielded cyclo[Q(Trt)W(Boc)FGLG] as a white solid which was used without further purification. To remove the side chain protecting groups, cyclo[Q(Trt)W(Boc)FGLG] was dissolved in 100ml of a 5:5:1 mixture of dichloromethane, TFA and water. The mixture was stirred for 1h and concentrated in vacuo. Co-evaporation with EtOAc resulted in cyclo[QWFGLG] as a white/brown solid, and this was confirmed by LCMS. The same procedure was performed for the synthesis of head-to-tail and side-chain cyclo[KWFGLG]. Crude peptides were purified by reversed-phase HPLC (RF-HPLC) on a C18 column using a gradient of 0-80% B (Buffer A: water/0.05%TFA; Buffer B: 90% acetonitrile/10% water/0.045% TFA) over 30 min.

**Mass Spectrometry and Infrared Photodissociation Spectroscopy**

IRMPD spectra of ions of interest were recorded using the free electron laser FELIX located at the FOM institute for Plasma Physics Rijnhuizen. As in previous experiments, the peptides were ionized by electrospray ionization (ESI), followed by IR spectroscopic interrogation with the free electron laser FELIX in the Penning trap of a Fourier transform ion cyclotron resonance (FTICR) mass spectrometer. The IRMPD spectrum was obtained by scanning the IRMPD yield as a function of FELIX wavelength. Note that this yield is defined as yield = -ln[1-(ΣIntphotofragments/ΣIntAllIons)], and that it is normalized with the relative laser power at each wavelength step.

**Results and Discussions: The Effect of Glutamine on Solution Phase Head-to-Tail Cyclization**

Synthesis of the linear peptide QWFGLG was found to be fairly efficient, and the peptide could be purified by LCMS, as shown in Figure 5-1. The mass spectrum at
retention time 5.51 minutes displays an abundant peak at m/z 707. After a 24-hour cyclization reaction of this peptide, using HATU/HOAT as a coupling reagent, a mixture of cyclized products was apparently generated. Scheme 5-4 depicts the various dehydration pathways which yield m/z 689 product ion. In Figure 5-2, the mass spectra at retention times 5.61, 6.03, 6.29, 6.76, and 6.98 are shown. The peak observed at 5.61 corresponds to (cyclo) QWFGLG. For both 6.03 and 6.29, an abundant peak at m/z 689 is seen, whereas for 6.98 a peak at m/z 690 is observed. The mass peaks at 689 indicate an elimination of H₂O (i.e., dehydration), whereas the mass peak at 690 suggests an elimination of NH₃ (i.e., deamination). The retention time of 6.76 corresponds to a (cyclo) QWFGLGQWFG (m1208). These results are summarized in Table 5-1.

In principle, there are a number of putative structures that need to be considered for the dehydration reaction for the linear peptide: 1) the 18-atom-ring head-to-tail cyclo (QWFGLG), 2) the 21-atom-ring glutamine side-chain (cyclo) QWFGLG, and 3) (pyroglutamyl) WFGLG. These structures are depicted in Scheme 5-4 along with the pathways that lead to these products. The proposed deamination product, N-terminal (pyroglutamic) WFGLG, is shown in Scheme 5-5. The side-chain-mediated products are hypothesized based on previous studies described in the introduction above.

Fractions 1 and 2 were isolated and ionized by electrospray ionization (ESI) to record the IRMPD spectra of their respective m/z 689 ions. Both IRMPD spectra are compared in Figure 5-4. The vibrational features exhibit considerable differences. Particularly the amide I (CO stretch) and amide II (NH bending) are shifted, suggesting
considerable differences in their secondary structures. This finding supports the hypothesis for competing reaction pathways in solution.

As a control experiment, the glutamine side chain was protected by a trityl group, to disable pathways 2 and 3 (in Scheme 5-4). Following the cyclization reaction, the glutamine side chain was deprotected, yielding a mass spectral peak at 689. It is expected that this reaction product corresponds to the head-to-tail cyclic peptide, (cyclo) QWFGLG. The recorded IRMPD spectrum for this control system bears striking resemblance to the IRMPD spectrum for fraction 1, confirming that fraction 1 in fact represents the head-to-tail (cyclo) QWFGLG product. This control experiment also indirectly confirms that fraction 2 corresponds to a reaction product involving the glutamine side chain. The question to be answered hence becomes whether fraction 2 is product 2 or 3, as hypothesized in Scheme 5-4.

In another control experiment, the N-terminus was protected with an FMOC group prior to the cyclization step. A range of coupling reagents, are shown in Figure 5-3. They were investigated, in an attempt to cyclize from the glutamine side chain. Employing diisopropylcarbodiimide (DIC) resulted in ten percent cyclization. However, cyclization required more than 24 hours and upon reaching the threshold of 10% conversion; the cyclization reaction desisted even with further addition of DIC. Unfortunately, this amount of side-chain mediated cyclic peptide was not enough to isolate upon work-up and measurement on the FTICR for spectroscopic interrogation. Note that this finding is also compatible with the fact that the glutamine side chain amide NH$_2$ is expected to be a poor nucleophile. Carbodiimide based reagents are commonly utilized in peptide chemistry for the formation of amide bonds. Their use in combination
with additive 1-hydroxybenzotriazole (HOBt) in the suppression of the racemization.\textsuperscript{169} DIC is an excellent coupling agent for use in solid-phase peptide synthesis, where urea byproduct insolubility has proved to be problematic.\textsuperscript{170,171} The general mechanism for the formation of amides or peptides using carbodiimides as a dehydrating agent is shown in Scheme 5-5, and proceeds via the formation of an O-acylisourea intermediate. This activated ester subsequently reacts with an amine nucleophile to form the amide bond and diisopropylurea. The major side-reaction is a rearrangement to an N-acylurea, but this is effectively suppressed by the use of solvents such as dichloromethane or chloroform.\textsuperscript{172}

The challenging production of the side-chain-mediated cyclic peptide in the control experiment, as opposed to the facile formation of the product in fraction 2, suggests that the product in fraction 2 involves an unprotected N-terminus. In fact, the separate protection of the glutamine side chain and the N-terminus in both control experiments indicate that the reaction product relies on the presence of an unprotected N-terminus and glutamine side chain, and is hence in principle compatible with the hypothetical product 3 in Scheme 5-4, namely N-terminal pyroglutamyl-WFGLG. In order to confirm this hypothesis, a similar synthetic product was made as a reference compound, N-terminal pyroglutamic acid-WFGLG (m/z 690), using BOC-pyroglutamic acid as a building block. Both structures are almost identical, but importantly differ on their N-terminal side by the presence of a CNH\textsubscript{2} moiety for pyroglutamyl-WFGLG, in contrast to a CO moiety for pyroglutamic acid-WFGLG, as depicted in Figure 5-6. The IRMPD spectrum for pyroglutamic acid-WFGLG (m/z 690) is compared to fraction 2 (m/z 689) in Figure 5-7. While the relative intensities differ somewhat, all bands in both spectra are
reproduced, with the notable exception for the high-frequency band at 1774 cm\(^{-1}\). This band is most likely ascribed to a CO stretch mode, which is also the difference in structure between pyroglutamyl and pyroglutamic acid, as pointed out above.

**IRMPD Results (cyclo) QWFGLG and (Pyroglutamyl) WFGLG Peptide**

The IRMPD spectra of (cyclo) QWFGLG and (Pyroglutamyl)WFGLG peptide are shown in Figure 5-4. Clearly, the isobars exhibit very disparate IRMPD spectra.

The region \((1350-1750 \text{ cm}^{-1})\) for head-to-tail peptide is blue shifted relative to the (N-terminal pyroglutamyl)WFGLG structure. Moreover, the spectral features are much broader than those of the (cyclo) QWFGLG. Comparing the two IRMPD spectra, the amide II (i.e., N-H bending) band is much broader (i.e., \(1450-1550 \text{ cm}^{-1}\)), which is consistent with a more dynamic oxazolone, as opposed to a more constrained macrocycle structure. This is also supported by the broader amide II (i.e., N-H bending) band in the N-acetylated fragment ions as well as the \(b_6\) ion control \((4\text{Abz})WFGLG\).\(^{75,136}\) The N-terminal pyroglytamyl WFGLG structure possess two additional unresolved bands in the higher frequency region \((1700-1800\text{cm}^{-1})\). The presence of these bands is most likely attributed to the N-terminal pyroglutmaly moiety stretching mode since it is higher in energy. Since this spectral feature is manifested in a less congested area at higher energy this facilitates differentiating between the linear and cyclical structures. It is interesting to note, that in Chapter 3 it was discussed how, the IRMPD spectrum for \((4\text{AMBz})WFGLM\) \(b_6\) differs from the other 4AMBz analogs, in that it displays an obvious splitting of the amide I stretching modes. This phenomenon was ascribed to the incorporation of 4AMBz in the first position which suggested that it was related to the benzoic acid residue CO stretch. In this case, one sees that the incorporation of glutamine in the first position not only gives rise to different products in
solution but also to different IRMPD spectra. In addition, a structural parallelism may be drawn between the IRMPD spectra from (cyclo)QWFGLM and (cyclo)QWFGLG. Both spectra are nearly identical whereas the IRMPD spectra N-terminal pyroglutamyl lacks the presence of many of the course spectral features exhibited for both cyclical spectra. Also, one may extract from Figure 5-4 that the region (from about 1500 to 1200 cm\(^{-1}\)) contains a very complicated series of absorptions consisting of bending vibrations within the cyclical molecule. However, there is an absence of these complex spectral features in the N-terminal pyroglutamyl structure. As a direct comparison, the head-to-tail cyclic peptide QWFGLG was synthesized. The IRMPD spectra was recorded and compared to the HPLC fraction 1. Figure 5-5 clearly demonstrates that the HPLC fraction 1 is indeed the head-to-tail cyclization. Characterizing the N-terminal (pyroglutamyl)WFGLG presents more of a challenge.

**IRMPD Results (N-terminal Pyroglutamic)WFGLG**

In solution phase cyclization of QWFGLG, the conditions are optimum for obtaining various mixtures of products such as (cyclo)QWFGLG, linear (Pyroglutamyl) WFGLG-OH and (Pyroglutamic) WFGLG peptides. However, directly synthesizing (Pyroglutamyl) WFGLG presents a synthetic conundrum. Sophisticated orthogonal synthetic approaches would be required to obtain the pyroglutamyl structure. Moreover, it cannot be guaranteed with certainty that the (Pyroglutamyl) WFGLG could be obtained. However, one can directly synthesize (Pyroglutamic) WFGLG peptide. One can obtain commercially available, N-terminal protected, *tert*-Butoxycarbonyl (Boc), pyroglutamic acid, Boc-Pyr-OH (Figure 5-9). One could simply synthesize the linear peptide WFGLG and then couple Boc-Pyr-OH. The Boc protecting groups would then be removed from the N-terminus simultaneously with the linear peptide from the resin
during the TFA cleavage step. In principle, if one were to measure the IRMPD spectra of (Pyroglutamic) WFGLG-OH and compared it to the IRMPD spectra of (Pyroglutamyl) WFGLG-OH they should not differ to a great extent. One would expect both structures to possess great spectral overlap because there exists a constant region in the peptide sequence namely $\sim\sim\sim$WFGLG-OH. However, the IRMPD spectra for the N-terminal position of (Pyroglutamic) structure would contain a diagnostic vibrational frequency. This moiety would correspond to the CO stretch whereas, there would exist an absence of this moiety in the (Pyroglutamyl) structure. Figure 5-7 shows a comparison between the IRMPD spectra of (Pyroglutamic) WFGLG-OH and the LC fraction 2 which corresponds to the (Pyroglutamyl) WFGLG-OH. It is glaringly obvious that both structures are nearly identical from the spectral region (1200 to 1750 cm$^{-1}$). As expected, there’s a marked presence of a diagnostic vibrational frequency in the high-frequency region (1755–1830 cm$^{-1}$) of the spectrum. As discussed earlier, there is strong evidence that LC fraction 2 is in fact linear N-terminal (pyroglutamyl)WGLG-OH. This argument has been developed based upon a comparison of the IRMPD spectral features of (cyclo)QWFGLG and (N-terminal pyroglutamyl)WGLG-OH. The fact that (Pyroglutamic) WFGLG-OH and the LC fraction 2, (Pyroglutamyl) WFGLG-OH, have nearly identical IRMPD spectra with the exception of a blue shifted CO stretch band further bolsters the case that in fact a linear molecule was synthesized in solution. There seems to exist some complex intramolecular interaction that blue shifts this CO band. As previously discussed, In IRMPD spectroscopy, the ion of interest is photodissociated with a tunable laser, indirectly yielding the infrared spectrum of the ion.$^{154}$ In addition, one can gain structural information from the photofragments
observed upon photodissociation of the molecule. Table 5-2 details the photofragments observed for the structures discussed thus far. The (Pyroglutamyl)WFGLG-OH photofragments may be derived as amide bond cleavages along the peptide backbone. One may rationalize that this photofragmentation is logical because the pyroglutamyl structure is affixed as a tertiary carbocation. As a result, this immobilizes the proton and the charge on the N-terminus. Therefore, amide cleavages would propagate along the spin of the peptide. By limiting the mobile proton, one can better predict and characterize both the formation of direct sequence ions as well as chemical species in the gas phase. It is interesting to note that in Table 5-2 that the photofragments recorded for the (cyclo)QWFGLG were performed on three separate times. As a result, laser intensity as well as the laser frequency which was resonant with a particular vibration of the molecule varied. The variation in these experimental parameters dictates the emergency or absence of specific dissociation channels and observed photofragments of the molecule. Consequently, different photofragments will be observed. However, the IRMPD spectra in all three measurements are nearly identical. Figure 5-8 shows a comparison the appearance spectra for the various mass channels observed for (cyclo)QWFGLG and HPLC fraction 2 (cyclo)QWFGLG. The ammonia loss channel (m672) predominates for all three spectra recorded (cyclo)QWFGLG. Stipdonk discusses how in the CID spectrum of $b_5^+$ derived from YAQFLG the dominant fragment ion is a result of the loss of NH$_3$ for the permuted sequences of glutamine. This phenomena arises because there appears to exist a macrocycle structure which is formed in all the cases and selectively opens at glutamine residue.$^{131}$ Therefore, when one has a mid-size b ion which contains a glutamine residue there the ion will cyclize.
and its most dominant fragmentation channel will pertain to ammonia lose. It should be noted that in his CID spectra the loss of glutamine is the second most intense mass channel; however, in Table 5-2 the glutamine loss is only manifested in one of the three recorded spectra. This is most likely attributed to the experimental conditions such as source conditions or other experimental parameters as already discussed. From the three separate measurements performed of (cyclo)QWFGLG; one better ascertains that collectively, the photofragment information in part sketches the chemical structure. One set of measurements of internal fragment losses such as m244 (a5-Phe-Leu) and m143 (Gly-Leu-28) correspond to scrambling of the sequence. In addition, immonium ions such as F (m120) and W (m159) are observer for the (cyclo)QWFGLG.

**Calculations**

Through density functional theory (DFT) (B3LYP/6-31 g*) methods, further evidence may be gleaned as to the putative structure of these molecules. Currently, elaborate calculations in collaboration with Dr. Roitberg's group are being carried out here at the High-Performance Computing (HPC) Center at the University of Florida. Since, these molecules are all hexapeptides, they present computational challenges and require time. During the interim time period, one can perform simple calculations and make certain assumptions to decomplexify the problem. This approach can provide one with a rough idea if certain diagnostic bands are present or absent from these structures. The first approach entails decoupling the (WFGLG) portion of the peptide from the problem. Both N-terminal pyroglutamic acid-WFGLG and N-terminal pyroglutamyl-WFGLG share this same sequence. Their IRMPD spectra are nearly identical in the spectral region (1200 to 1750 cm\(^{-1}\)) with the exception of an unknown band in the high-frequency region (1755–1830 cm\(^{-1}\)) for N-terminal pyroglutamic acid-
WFGLG. By simply performing energy minimization and molecular mechanics on pyroglutamyl and pyroglutamic acid with a methyl group appended on the C-terminus computation time is dramatically abbreviated. In addition, these results can reveal diagnostic bands or an absence of particular vibrational modes. In order to assign the 1774 cm\textsuperscript{-1} band to the pyroglutamic acid CO stretch, computations were carried out on related model systems shown in Figure 5-12.

The molecules were initially built with Chem3D. Structure optimization and frequency calculations were conducted at the density functional theory level, using the B3LYP functional and a 6-31g* basis set in the Gaussian suite of programs. The output from the frequency calculations is shown in Figure 5-13, using a scaling factor of 0.965 and a Gaussian line-broadening function (fwhm = 10 cm\textsuperscript{-1}). Of the various predicted structures, the pyroglutamic acid structure protonated on the backbone carbonyl O confirms a pyroglutamic acid CO stretch at 1795 cm\textsuperscript{-1}. Figure 5-8 shows the various structures obtained after molecular mechanics calculations. Table 5-3 shows the theoretically predicted IR bands for these structures. Based upon the calculations, the harmonic spectrum is produced for the N-terminal pyroglutamic acid-CONHCH\textsubscript{3}. It predicts a CO bending mode at 1795 cm\textsuperscript{-1} for this structure. In the recorded IRMPD spectra a very prominent sharp band is observed at 1774 cm\textsuperscript{-1}. The band is red shifted by 21 cm\textsuperscript{-1}. Even though, one may argue that this is band does not correspond to the one predicted theoretically, one must take consider the following variable in this assignment. First, this calculation provides a qualitative picture of the structure. Molecular mechanics was not performed upon the entire hexapeptide. Therefore, pertinent conformations of the molecule as well as the global minimum were not
obtained. For the N-terminal pyroglutamy-CONHCH₃ has the charge delocalized between the two nitrogen atoms. The predicted spectrum shows a prominent mode at 1657/1636 cm⁻¹. The experimental IRMPD spectrum shows an unresolved band at 1731/1707 cm⁻¹. A couple of factors need to be considered when thinking of this system. First, the molecule is a hexapeptide in nature, whereby hydrogen bonding will red shift peaks. Secondly, in this calculation the charged is affixed on the pyroglutamyl moiety. The predicted spectra will be affected with regards to where the charge is localized. In addition, the molecule participates in hydrogen bonding thereby red shifting the position of the observed band. Lastly, if one examines Table 5-4 it is apparent that great overlap exist between (pyroglutamic)WFGLG and (pyroglutamyl) WFGLG. This is because there structures are quite similar with the exception of the N-terminal position. Also, as a comparison HPLC fraction 2 and cyclo(WFGLG) are compared to emphasize precise spectral accord. Currently, more definitive calculations are being performed to completely characterize the structure of these bio-molecules. The close match between the IRMPD spectrum for fraction 2 (m/z 689) and the control compound, pyroglutamic acid-WFGLG (m/z 690), lends strong credence to the claim that the product in fraction 2 corresponds to the pyroglutamyl-WFGLG structure. The pyroglutamyl structure had already been proposed in gas-phase deamination reactions for N-terminal glutamine-containing peptides, but was not yet known for solution-phase chemistry.

In summary, two competing cyclization reactions are confirmed to take place in the solution-phase coupling reaction for QWFWGLG: 1) head-to-tail cyclization to give rise to a head-to-tail macrocycle, and 2) glutamine side chain reaction with the N-terminus to give rise to a linear pyroglutamyl structure. Both structures could be conveniently
separated by HPLC. In addition, both molecules display very different IRMPD spectra, as evidenced by their amide I and amide II features.

**Lysine Side-Chain-Mediated Chemistry**

Side-chain-mediated macrocyclization was not found to be a facile synthetic route for QWFGLG. In hindsight, it is not surprising that the glutamine side chain amide NH$_2$ is a poor nucleophile. A lysine side chain, on the other hand, incorporates a primary amine NH$_2$, and is hence chemically much more analogous to the N-terminus. The solution-phase cyclization chemistry of the related peptide system KWFGLG was hence investigated to determine head-to-tail vs. side-chain cyclization. The corresponding 18-atom and 22-atom ring structures are shown in Figure 5-10.

The synthesis was directed by separate protection of the lysine side chain and N-terminus to produce either exclusively head-to-tail (cyclo)KWFGLG and side-chain (cyclo)KWFGLG both at m/z 689. The IRMPD spectra for both of these species are shown in Figure 5-11. Due to the close similarity of both spectra, two independent measurements are shown for both. With the exception of a minor difference around 1600 cm$^{-1}$, both IR spectra are found to be almost identical. This difference may relate to NH$_2$ bending modes, which are affected by the cyclization or non-cyclization of the lysine side chain. For these closely related macrocycle structures, it is thus found that IRMPD spectroscopy is effectively incapable to distinguish between them.

**Summary**

This Chapter focused on the cyclization chemistry of peptides in solution, as opposed to gas-phase cyclization in previous chapters. The complex mixture of reaction products was studied by combining off-line HPLC separation and mass spectrometric analysis, followed by structural investigation of some HPLC fractions by IRMPD
spectroscopy. For the N-terminal glutamine-containing peptide QWFGLG, it is confirmed that competing dehydration reactions take place due to head-to-tail cyclization, as well as glutamine N-terminal cyclization to form a pyroglutamyl structure. These two products were successfully separated by HPLC. Moreover, their IRMPD spectra, in combination with control experiments and computations, allow a delineation between cyclical and linear structures. It should be noted that this analysis was not possible by NMR, as the HPLC fractions were insufficiently pure for structural study. This demonstrates the potential usefulness of IRMPD spectroscopy as a structural probe of analytes in heterogenic samples, where HPLC and mass spectrometry are employed as purification methods, while IRMPD spectroscopy can confirm the chemical structure.

In order to explore the limitations of IRMPD spectroscopy in distinguishing chemically distinct structures, the related head-to-tail (18-atom ring) and side-chain (22-atom ring) macrocycle structures were made synthetically for KWFGLG. The IRMPD spectra were found to be too similar to permit their differentiation. As an analytical technique, IRMPD spectroscopy of room-temperature ions faces limitations in distinguishing mid-sized peptide systems that have only minor differences in their vibrational spectra. The band broadening due to room-temperature ions and multiple-photon absorption render these differences even smaller. One approach may hence involve cooling of the ions to maximize the inherent differences in the vibrational spectra of isomers.
Scheme 5-1. Glutamine’s pathways in both solution phase and gas phase. (a) Spontaneous formation of a peptide glutarimide and products of its hydrolysis. (b) Degradation of peptides with the formation of a glutarimide. [Figure reprinted with permission from Jonsson, A. P.; Bergman, T.; Jörnvall, H.; Griffiths, W. J.; Bratt, P.; Strömberg, N. J. Am. Soc. Mass Spectrom. 2001, 12, 337].
Scheme 5-2. Amino acid residue pathways leading to formation of pyroglutamic acid. [Figure is adapted with permission from Baldwin, M. A.; Falick, A. M.; Gibson, B. W.; Prusiner, S. B.; Stahl, N.; Burlingame, A. L. J. Am. Soc. Mass Spectrom. 1990, 1, 258].
Scheme 5-3. Dehydration and deamination channels of glutamine when situated at N-terminal position. [Figure is adapted with permission from Neta, P.; Pu, Q.-L.; Kilpatrick, L.; Yang, X.; Stein, S. E. J. Am. Soc. Mass Spectrom. 2007, 18, 27].
Scheme 5-4. Proposed reaction pathways to rationalize dehydration from linear peptide QWFGLG, yielding m/z 689 product ion: Head-to-tail cyclization is labeled black, glutamine side chain cyclization is labeled in blue, and N-terminal pyroglutamyl is labeled in dotted red.
Scheme 5-5. Proposed reaction pathways of deamidation from linear peptide QWFGLG, yielding m/z 690 product ion.
Scheme 5-6. General mechanism for the formation of an amide bond formation using a carbodiimide. The acid 1 will react with the carbodiimide to produce the key intermediate: the O-acylisourea 2. The O-acylisourea will react with amines to form amide 3 and urea 4. The O-acylisourea 2 can react with an acid 1 to give a carboxylic anhydride 5, which can react further to give the desired amide 3. The main undesired reaction pathway involves the rearrangement of the O-acylisourea 2 to the stable N-acylurea 6. [Figure reproduced with permission from http://www.chempep.com/ChemPep-Generic-Term_Reagents_Reactions.htm]
Figure 5-1. LCMS chromatogram and mass spectrum of linear peptide QWFGLG.
Figure 5-2. LCMS chromatogram and mass spectrum of cyclization of QWFGLG.
Figure 5-3. Coupling reagents employed to try to cyclize Glutamine from the chain. [Figure reproduced with permission from http://www.aapptec.com/broch/Coupling%20Reagents.pdf]
Figure 5-4. IRMPD spectra of m/z 689 ions fractions 1 and 2.
Figure 5-5. Comparison of IRMPD spectra of head-to-tail cyclic (QWFGLG) and HPLC fraction 1.
Figure 5-6. Chemical structures for N-terminal pyroglutamyl-WFGLG and N-terminal pyroglutamic acid –WFGLG.
Figure 5-7. Comparison of IRMPD spectra of (Pyroglutamic) WFGLG-OH and LC fraction 2.
Figure 5-8. Appearance spectrum of mass channels for (cyclo) QWFGLG.
Figure 5-9. Chemical structure of BOC-pyroglutamic acid.

Figure 5-10. Chemical structure of (cyclo)KWFGLG.
Figure 5-11. IRMPD spectra of (cyclo)KWFGGLG of head-to-tail and side cyclization of lysine.
Figure 5-12. Structures of (pyroglutamyl)NHCH₃ and (pyroglutamic)NHCH₃ optimized.
Figure 5-13. Predicted IRMPD spectra for (pyroglutamyl)NHCH$_3$ and (pyroglutamic)NHCH$_3$
Table 5-1. HPLC retention times and masses of products from cyclization of QWFGLG.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Retention Time (minutes)</th>
<th>Mass (Da)</th>
<th>Intensity of MS peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>QWFGLG</td>
<td>5.51</td>
<td>707</td>
<td>1.39 E7</td>
</tr>
<tr>
<td>(cyclo)QWFGLG</td>
<td>5.61</td>
<td>519</td>
<td>1.75 E5</td>
</tr>
<tr>
<td>Fraction 1</td>
<td>6.03</td>
<td>689</td>
<td>6.23 E5</td>
</tr>
<tr>
<td>Fraction 2</td>
<td>6.29</td>
<td>689</td>
<td>3.05 E5</td>
</tr>
<tr>
<td>(cyclo) QWFGLGQWFGLG</td>
<td>6.76</td>
<td>1208</td>
<td>2.51 E5</td>
</tr>
<tr>
<td>(Pyroglutamic Acid)WFGLG</td>
<td>6.98</td>
<td>690</td>
<td>7.95 E4</td>
</tr>
</tbody>
</table>
Table 5-2. Photofragmentation assignments.

<table>
<thead>
<tr>
<th>Fragment Mass (Daltons)</th>
<th>(Pyroglutamyl)WFGLG (HPLC Fraction 1) Recorded April 2010</th>
<th>(Cyclo)QWFGLG (HPLC Fraction 2) Recorded April 2010</th>
<th>(Cyclo)QWFGLG Reference Compound Recorded Aug. 2010</th>
<th>(Cyclo)QWFGLG Reference Compound Recorded Nov. 2010</th>
<th>(Pyroglutamic)WFGLG Recorded Dec. 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>690</td>
<td>[M+H]⁺</td>
<td>[M+H]⁺</td>
<td></td>
<td></td>
<td>[M+H]⁺</td>
</tr>
<tr>
<td>689</td>
<td>[M+H]⁺</td>
<td>[M+H]+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>672</td>
<td>[M+H-NH₃⁺]</td>
<td>[M+H-NH₃⁺]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>671</td>
<td></td>
<td>[M+H-NH₃⁺]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>654</td>
<td></td>
<td>[M+H-H₂O⁺]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>615</td>
<td>b₆</td>
<td>b₆</td>
<td></td>
<td></td>
<td>b₅</td>
</tr>
<tr>
<td>587</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a₅</td>
</tr>
<tr>
<td>561</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>502</td>
<td>b₅</td>
<td>b₅</td>
<td></td>
<td></td>
<td>b₄</td>
</tr>
<tr>
<td>445</td>
<td>b₄</td>
<td>b₄</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>413</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>397</td>
<td>unassigned</td>
<td>unassigned</td>
<td>unassigned</td>
<td></td>
<td></td>
</tr>
<tr>
<td>364</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>280</td>
<td>b₃-18</td>
<td>b₃-18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>244</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>a₅-Phe-Leu</td>
</tr>
<tr>
<td>173</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>159</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>Tyrosine side chain loss</td>
<td>Tyrosine side chain loss</td>
<td>Tyrosine side chain loss</td>
<td>Tyrosine side chain loss</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>
Table 5-3. Theoretical IR band positions for (pyroglutamyl)-R and pyroglutamic acid-R

<table>
<thead>
<tr>
<th>Structure</th>
<th>cm⁻¹</th>
<th>cm⁻¹</th>
<th>cm⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-terminal Pyroglutamyl-R</td>
<td>1565</td>
<td>1636</td>
<td>1657</td>
</tr>
<tr>
<td>CO on ring-protonated Pyroglutamic acid-R</td>
<td>1513</td>
<td>1656</td>
<td>1696</td>
</tr>
<tr>
<td>Backbone CO Pyroglutamic acid-R</td>
<td>1517</td>
<td>1676</td>
<td>1795</td>
</tr>
<tr>
<td>N-protonated Pyroglutamic acid-R</td>
<td>1537</td>
<td>1627</td>
<td>1717</td>
</tr>
</tbody>
</table>

R=(CONHCH₃)

Table 5-4. Comparison IRMPD spectra of (pyroglutamic)WFGLG, (pyroglutamyl)WFGLG, and cyclo (QWFGGLG).

<table>
<thead>
<tr>
<th>(Pyroglutamic)WFGLG (cm⁻¹)</th>
<th>(Pyroglutamyl)WFGLG HPLC Fraction 1 (cm⁻¹)</th>
<th>Cyclo(WFGLG) HPLC Fraction 2 (cm⁻¹)</th>
<th>Cyclo (WFGLG) Reference Compound (cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1234</td>
<td>1236</td>
<td>1230</td>
<td>1232</td>
</tr>
<tr>
<td>1277</td>
<td>1282</td>
<td>1290</td>
<td>1291</td>
</tr>
<tr>
<td>1337</td>
<td>1334</td>
<td>1328</td>
<td>1328</td>
</tr>
<tr>
<td>1408</td>
<td>1400</td>
<td>1386</td>
<td>1386</td>
</tr>
<tr>
<td>1492</td>
<td>1495</td>
<td>1425</td>
<td>1424</td>
</tr>
<tr>
<td>1593</td>
<td>1595</td>
<td>1453</td>
<td>1454</td>
</tr>
<tr>
<td>1667</td>
<td>1663</td>
<td>1537</td>
<td>1537</td>
</tr>
<tr>
<td>1708</td>
<td>1707/1731</td>
<td>1605</td>
<td>1605</td>
</tr>
<tr>
<td>1774</td>
<td>--------------</td>
<td>1641</td>
<td>1641</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1684</td>
<td>1684</td>
</tr>
</tbody>
</table>
CHAPTER 6
CONCLUSIONS AND FUTURE DIRECTIONS

The results in this thesis have provided systematic insights governing the rearrangement chemistry governing cyclization of $b_n$ ions in collision-induced dissociation (CID). It has been shown through a complementary approaches combining infrared multiple photon dissociation (IRMPD) spectroscopy, synthetic chemistry, SORI-CID and HDX the fragmentation pathways and mechanisms of gas-phase chemistry. The first project on this pertains to the influence of the torsion strain of peptide backbone length on $b$ ion structures. The aim was to determine if amino acid residues with constrained torsional freedom affect macrocycle formation in the corresponding $b$ fragment. The presence of proline in position 2 resulted in appearance of sizeable oxazolone bands for $QPWFGLM_{b_7}$ and, to a lesser degree, in $QPFGLM_{b_6}$. This supports the hypothesis that constraining the flexibility of the backbone can suppress “head-to-tail” cyclization from the $N$-terminus. On the other hand, insertion of 4-aminomethylbenzoic acid (4AMBz) had little or no effect on macrocycle formation for $b_6$, with the exception of a weak oxazolone band for $QWFG(4AMBz)M_{b_6}$, with 4AMBz in the fifth position. It is probable that this is mainly due to a lowering in energy of the oxazolone structure, as a result of a conjugated $\pi$-electron system between the benzene and oxazolone rings. The vibrational spectra demonstrate that the formation of macrocycle $b$ structures can be surprisingly robust, and is difficult to disrupt, even using residues with limited torsional degrees of freedom. An interesting experiment would be to utilize (ortho)AMBz stead of the (para)AMBz in the sixth position. Most likely, utilizing (ortho)AMBz would not result in a sterically hindered nucleophilic attack from a backbone carbonyl. Moreover, if the $b_6$ ion were generated then this would further
bolster the proposed mechanism that oxazolone structures are formed prior to isomerization to macrocycle structures. As a control experiment, one could also study the effect that (meta)AMBz would have upon macrocyclization and the formation of the $b_6$ ion in the sixth position. Figure 6-1 better shows these oxazolone structures.

The second project addressed the question of whether or not extending the chain length and varying residue position results in macrocyclization and scrambling. In this investigation, larger $bn$ ($n=4-15$) fragments were studied. The sequence motif GlyAlaTyr, and the related motifs were chosen. In all cases for $b_4$, sizeable bands are observed. The current results lend credence to the claim that head-to-tail cyclization in $b$ ions is highly prevalent in low-energy CID of peptides. These results strongly indicate the presence of head-to-tail macrocycles for all of the larger $b$ ions, $b_6$, $b_9$ and $b_{12}$, as opposed to oxazolone structures for $b_4$, yet again confirming a size effect in the formation of macrocycle structures. A couple of additional studies still remain to better comprehend macrocyclization. First, larger systems still need be explored to ascertain if there exists a size where macrocyclization reverts to an oxazolone structure. Also, the influence of acidic, basic, and amide side chains on the formation of macrocyclization still remains to be explored.

The third project focused upon the cyclization chemistry of peptides in solution, as opposed to gas-phase cyclization in previous chapters. The complex mixture of reaction products was studied by combining off-line HPLC separation and mass spectrometric analysis, followed by structural investigation of some HPLC fractions by IRMPD spectroscopy. For the N-terminal glutamine-containing peptide QWFGLG, it is confirmed that competing dehydration reactions take place due to head-to-tail
cyclization, as well as glutamine N-terminal cyclization to form a pyroglutamyl structure. These two products were successfully separated by HPLC. Moreover, their IRMPD spectra, in combination with control experiments and computations, allow delineation between cyclical and linear structures. It should be noted that this analysis was not possible by NMR, as the HPLC fractions were insufficiently pure for structural study. This demonstrates the potential usefulness of IRMPD spectroscopy as a structural probe of analytes in heterogenic samples, where HPLC and mass spectrometry are employed as purification methods, while IRMPD spectroscopy can confirm the chemical structure.

In order to explore the limitations of IRMPD spectroscopy in distinguishing chemically distinct structures, the related head-to-tail (18-atom ring) and side-chain (22-atom ring) macrocycle structures were made synthetically for KWFGLG. The IRMPD spectra were found to be too similar to permit their differentiation. As an analytical technique, IRMPD spectroscopy of room-temperature ions faces limitations in distinguishing mid-sized peptide systems that have only minor differences in their vibrational spectra. The band broadening due to room-temperature ions and multiple-photon absorption render these differences even smaller. One approach may hence involve cooling of the ions to maximize the inherent differences in the vibrational spectra of isomers.
Figure 6-1. Proposed oxazolone formation from arene substitution patterns.
LIST OF REFERENCES

(7) Edman, P. H., Erik; Sillén, Lars Gunnar; Kinell, Per-Olof Acta Chem. Scand. 1950, 4, 283.
(12) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Science 1989, 246, 64.
(20) http://www.piercenet.com/browse.cfm?flId=3B750AB6-5056-8A76-4E71-83B42F51BE91.


(31) [www2.unine.ch/files/content/sites/saf/files/shared/](http://www2.unine.ch/files/content/sites/saf/files/shared/).


(83) Barrow, M. P.; Burkitt, W. I.; Derrick, P. J. Analyst 2005, 130, 18.


(172) [http://www.carbosynth.com/carbosynth/website.nsf/(w-productdisplay)/85BE0FA6A2C7802180256A6B003595EC](http://www.carbosynth.com/carbosynth/website.nsf/(w-productdisplay)/85BE0FA6A2C7802180256A6B003595EC).


(174) [http://www.aapptec.com/broch/Coupling%20Reagents.pdf](http://www.aapptec.com/broch/Coupling%20Reagents.pdf)


(176) [http://www.vivo.colostate.edu/hbooks/molecules/aminoacids.html](http://www.vivo.colostate.edu/hbooks/molecules/aminoacids.html)

(177) [http://www.russelllab.org/aas/](http://www.russelllab.org/aas/)
BIOMETRICAL SKETCH

Marcus Tirado was born in Jersey City; New Jersey. He has two brothers Domenico and David. Marcus’s entire family is musically and linguistically privileged because of their inherited background. His mother, Rosetta Gramponni-Esposito was from a remote village in the southern part of Italy. She would sing the most melodious and mellifluous Italian operas in short burst in her native Italian language. Marcus’s father from Barranquillia, Colombia would both play La Malaguena from Ernesto Leucona on his Spanish guitar to his children. After graduating from Rutgers College with a degree in Chemistry and English literature, Marcus went West and resided in beautiful San Diego, California for eight years. Here, in San Diego, Marcus worked in eclectic areas of science. In addition, Marcus continued diligently playing classical and flamenco guitar as he had done as a child. After years of playing year, Marcus moved to Spain to continue studying both Classical and Flamenco with some of the greatest guitarist in the world. It was in Jerez, Spain that Marcus found his true love and passion his beloved wife Delia Huanca Calderon. Marcus returned to the US and returned to work at the University of Florida Racing lab. Here, he pursued his studied in Forensic Drug Chemistry and upon advice from the lab Director Dr. Cindy Kollis-Baker applied for the Doctoral program in chemistry. He joined Dr. Charles Cao’s research group and investigated various synthetic avenues of quantum dot synthesis. In addition, his research branched off into nanoparticle applications such as FRET and bio-applications. After three years, Marcus switched groups for personal reasons joined Dr. Nicolas Polfer’s research group, and embarked upon his work on investigating the influences on structures of peptide b fragment ions with infrared multiple photon dissociation, gas-phase hydrogen/deuterium exchange, and Fourier Transform Ion
Cyclotron Resonance Mass Spectrometry, and SORI-CID. He received his Doctor of Philosophy from the University of Florida in August 2012.