NEGATIVE ION MASS SPECTROMETRY STUDIES OF NON-TRADITIONAL DEPROTONATION SITES IN PEPTIDES: GAS-PHASE ACIDITIES OF AMINO ACID ANALOGS AND FRAGMENTATION OF MODEL PHOSPHOPEPTIDES

by

CHELSEA ELIZABETH PLUMMER

CAROLYN J. CASSADY, COMMITTEE CHAIR
DAVID A. DIXON
CAROL DUFFY
SHANE C. STREET
STEPHEN A. WOSKI

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of The University of Alabama

TUSCALOOSA, ALABAMA

2016
ABSTRACT

Studying biomolecule deprotonated anions (amino acids, amino acid amides, phosphorylated amino acids, phosphorylated amino acid amides, and phosphorylated peptides) can further negative ion mode mass spectrometry work in the field of proteomics, the sequencing of human proteins. Traditionally, amino acids are considered to have acidic sites on their C-terminus and carboxylic acid-containing side chains. However, several non-traditional sites capable of deprotonation should be considered when examining peptides and proteins by mass spectrometry.

A new deprotonated tyrosine conformer was found by electrospraying (ESI) tyrosine from aprotic solvents. Ion/molecule (I/M) reactions determined that tyrosine made by ESI from protic solvents is less acidic. The gas-phase acidity (GA) value for the low-energy structure is 324.7 ± 3.6 kcal/mol experimentally and 330.4 kcal/mol computationally. The ESI conditions and hexapole trapping can affect the deprotonation site of tyrosine.

The GAs for the common amino acid amides, phosphorylated amino acids, and phosphorylated amino acid amides were determined experimentally and computationally. Two ion populations were observed via I/M reactions for the amino acid amides deprotonating on the C-terminal amide group that vary in energy by ~5 kcal/mol. Tyrosine, cysteine, tryptophan, and histidine amides undergo side chain deprotonation and are more acidic. Phosphorylated amino acids and their amides were found to be ~20 kcal/mol more acidic than their non-phosphorylated counterparts. Phosphorylated compounds deprotonate on the phosphate side chain, except
phosphotyrosine (pTyr), which deprotonates at the C-terminal carboxylic acid. The acidity of pTyr can be affected by ESI conditions.

Collison-induced dissociation (CID) was used to examine model phosphorylated peptides in the negative ion mode. The CID of singly charged deprotonated precursor ions produced the least amount of sequence informative fragmentation when the phosphorylated residue was located centrally in the peptide. Diagnostic ions and losses were found indicating the phosphate group. Phosphothreonine (pThr) and phosphoserine (pSer) undergo the loss of the phosphate group and side chain aldehyde to produce unique marker ions. Fragmentation of doubly charged precursor anions yielded little sequence informative fragmentation, however; diagnostic product ions indicating loss of the phosphate group allowed for differentiation between pThr, pSer, and pTyr.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>alternating current</td>
</tr>
<tr>
<td>Ala</td>
<td>alanine (A)</td>
</tr>
<tr>
<td>Arg</td>
<td>arginine (R)</td>
</tr>
<tr>
<td>AFTP</td>
<td>4-amino-2,3,4,6-tetrafluoropyridine</td>
</tr>
<tr>
<td>Asp</td>
<td>aspartic acid (D)</td>
</tr>
<tr>
<td>bar</td>
<td>unit of pressure (not an abbreviation)</td>
</tr>
<tr>
<td>Cys</td>
<td>cysteine (C)</td>
</tr>
<tr>
<td>cal</td>
<td>calorie</td>
</tr>
<tr>
<td>c</td>
<td>centi (prefix)</td>
</tr>
<tr>
<td>CH$_3$CN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>CH$_3$OH</td>
<td>methanol</td>
</tr>
<tr>
<td>CID</td>
<td>collision-induced dissociation</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DC</td>
<td>direct current</td>
</tr>
<tr>
<td>DFT</td>
<td>density functional theory</td>
</tr>
<tr>
<td>DIC</td>
<td>1,3-diisopropylcarbodiimide</td>
</tr>
<tr>
<td>ΔG</td>
<td>Gibbs free energy change</td>
</tr>
<tr>
<td>ΔH</td>
<td>enthalpy change</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>ΔS</td>
<td>entropy change</td>
</tr>
<tr>
<td>eV</td>
<td>electron volt</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>ECD</td>
<td>electron capture dissociation</td>
</tr>
<tr>
<td>ETD</td>
<td>electron transfer dissociation</td>
</tr>
<tr>
<td>ExD</td>
<td>electron-based dissociation</td>
</tr>
<tr>
<td>Fmoc</td>
<td>9-fluorenylmethoxycarbonyl</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>FID</td>
<td>free inductance decay</td>
</tr>
<tr>
<td>FWHM</td>
<td>full width half maximum</td>
</tr>
<tr>
<td>g</td>
<td>grams</td>
</tr>
<tr>
<td>GB</td>
<td>gas-phase basicity</td>
</tr>
<tr>
<td>GA</td>
<td>gas-phase acidity</td>
</tr>
<tr>
<td>Gln</td>
<td>glutamine (Q)</td>
</tr>
<tr>
<td>Glu</td>
<td>glutamic acid (E)</td>
</tr>
<tr>
<td>Gly</td>
<td>glycine (G)</td>
</tr>
<tr>
<td>HCT</td>
<td>high capacity trap</td>
</tr>
<tr>
<td>HGP</td>
<td>human genome project</td>
</tr>
<tr>
<td>His</td>
<td>histidine (H)</td>
</tr>
<tr>
<td>Hz</td>
<td>hertz</td>
</tr>
<tr>
<td>HOBT</td>
<td>1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>HPP</td>
<td>human proteome project</td>
</tr>
<tr>
<td>hr</td>
<td>hour</td>
</tr>
<tr>
<td>ICR</td>
<td>ion cyclotron resonance</td>
</tr>
<tr>
<td>Ile</td>
<td>isoleucine (I)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>I/M</td>
<td>ion/molecule</td>
</tr>
<tr>
<td>IRMPD</td>
<td>infrared radiation multiphoton dissociation</td>
</tr>
<tr>
<td>J</td>
<td>joule</td>
</tr>
<tr>
<td>k</td>
<td>kilo (prefix)</td>
</tr>
<tr>
<td>KE</td>
<td>kinetic energy</td>
</tr>
<tr>
<td>L</td>
<td>liter</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>Leu</td>
<td>leucine (L)</td>
</tr>
<tr>
<td>Lys</td>
<td>lysine (K)</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix-assisted laser desorption ionization</td>
</tr>
<tr>
<td>Met</td>
<td>methionine (M)</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry (or mass spectrometer)</td>
</tr>
<tr>
<td>µ</td>
<td>micro (prefix)</td>
</tr>
<tr>
<td>m</td>
<td>milli (prefix)</td>
</tr>
<tr>
<td>m</td>
<td>meter</td>
</tr>
<tr>
<td>M</td>
<td>moles/liter (concentration)</td>
</tr>
<tr>
<td>MeOH</td>
<td>methanol</td>
</tr>
<tr>
<td>mol</td>
<td>mole(s)</td>
</tr>
<tr>
<td>MS/MS</td>
<td>tandem mass spectrometry</td>
</tr>
<tr>
<td>MS²</td>
<td>tandem mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>NETD</td>
<td>negative electron transfer dissociation</td>
</tr>
<tr>
<td>–NH₂</td>
<td>amide</td>
</tr>
<tr>
<td>NH₄OH</td>
<td>ammonium hydroxide</td>
</tr>
<tr>
<td>NMP</td>
<td>N-methylpyrrolidinone</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PES</td>
<td>photoelectron spectroscopy</td>
</tr>
<tr>
<td>Phe</td>
<td>phenylalanine (F)</td>
</tr>
<tr>
<td>PIP</td>
<td>piperidine</td>
</tr>
<tr>
<td>Pro</td>
<td>proline (P)</td>
</tr>
<tr>
<td>pSer</td>
<td>phosphoserine (pS)</td>
</tr>
<tr>
<td>pThr</td>
<td>phosphothreonine (pT)</td>
</tr>
<tr>
<td>PA</td>
<td>proton affinity</td>
</tr>
<tr>
<td>PTM</td>
<td>post-translational modification</td>
</tr>
<tr>
<td>QIT</td>
<td>quadrupole ion trap</td>
</tr>
<tr>
<td>RE</td>
<td>reaction efficiency</td>
</tr>
<tr>
<td>rf</td>
<td>radio frequency</td>
</tr>
<tr>
<td>s</td>
<td>second</td>
</tr>
<tr>
<td>Ser</td>
<td>serine (S)</td>
</tr>
<tr>
<td>S/N</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SORI</td>
<td>sustained off-resonance irradiation</td>
</tr>
<tr>
<td>SPPS</td>
<td>solid-phase peptide synthesis</td>
</tr>
<tr>
<td>T</td>
<td>tesla</td>
</tr>
<tr>
<td>TFA</td>
<td>trifluoroacetic acid</td>
</tr>
<tr>
<td>Thr</td>
<td>threonine (T)</td>
</tr>
<tr>
<td>TIPS</td>
<td>triisopropylsilane</td>
</tr>
<tr>
<td>TMSN₃</td>
<td>trimethylsilyl azide</td>
</tr>
<tr>
<td>TOF</td>
<td>time-of-flight</td>
</tr>
<tr>
<td>Trp</td>
<td>tryptophan (W)</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine (Y)</td>
</tr>
</tbody>
</table>
Val  valine (V)
V  volt
ω  angular frequency
ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Dr. Carolyn J. Cassady for mentoring me over the last five years. Her chemistry knowledge, advice, support, and patience have been second to none and have allowed me to succeed here at The University of Alabama. She is the best teacher I have ever had.

I would also like to thank my committee members, Dr. David Dixon, Dr. Carol Duffy, Dr. Shane Street, and Dr. Stephen Woski. Their contributions are very much appreciated. I also thank Dr. Michele Stover for all of the computational work she has done in collaboration with me. I always enjoyed meeting with you discuss our results. Thank you also to Dixon group members John Davis and Sean Miller for help with calculations.

I have been fortunate to have been funded throughout the majority of my time in graduate school. Thank you to the NSF and the DoEd GAANN program for funding me to do research in the field of mass spectrometry.

A special thanks to the Cassady group members, past and present. I will always cherish my memories in graduate school thanks to Dr. Changgeng Feng, Juliette Commodore, Chelsea McMillen, and Ranelle Duke. I cannot believe how many places we have got to visit together in the last few years traveling the country for conferences and how much fun we’ve had. I appreciate all of the support and discussions you all have given me throughout the years. I also have to thank my other fellow UA mass spectrometrist, Thomas Hester, for being such a great friend and colleague throughout our time in the chemistry department.
My family has been especially supportive throughout my time in graduate school. My parents have always been just a phone call away whenever I need to talk. My Mom always tells the worst best jokes to boost my morale and it’s been an honor to see my Dad sporting Alabama gear instead of the Chicago Bears whenever I come home to visit. Thanks to my sister, Courtney, my brother-in-law, Taylor, and my best friend Chloie for always being supportive in their own little ways. I would like to add a special dedication to my late uncle, Tony Wright. His love for chemistry was inspiring and got me interested in science at a young age.
CONTENTS

ABSTRACT ............................................................................................................. ii

LIST OF ABBREVIATIONS AND SYMBOLS ................................................................ iv

ACKNOWLEDGEMENTS ......................................................................................... ix

LIST OF TABLES ..................................................................................................... xv

LIST OF FIGURES .................................................................................................. xvi

LIST OF SCHEMES .................................................................................................. xx

CHAPTER 1. OVERVIEW OF DISSERTATION ............................................................ 1

References .............................................................................................................. 11

CHAPTER 2. EXPERIMENTAL INSTRUMENTATION AND THEORY ...................... 22

  2.1 Overview ....................................................................................................... 22
  2.2 Electrospray Ionization .................................................................................. 22
  2.3 Mass Analyzers ............................................................................................. 25
    2.3.1 Fourier Transform Ion Cyclotron Resonance ......................................... 25
    2.3.2 Quadrupole Ion Trap ............................................................................. 31
  2.4 Tandem Mass Spectrometry (MS/MS) ........................................................... 32
    2.4.1 Ion/Molecule Reactions ....................................................................... 32
    2.4.2 Sustained Off-Resonance Irradiation Collision-Induced Dissociation ... 36
  2.5 Peptide Synthesis and Structures .................................................................. 38
    2.5.1 Fmoc Solid-Phase Peptide Synthesis ..................................................... 38
    2.5.2 Peptide Sequencing Nomenclature ......................................................... 40
2.5.3 Amino Acid, Amino Acid Amide, and Phosphorylated Amino Acid and Amide Structures ..............................................42

References .................................................................................................45

CHAPTER 3. A NEW GAS-PHASE DEPROTONATED TYROSINE STRUCTURE: AN EXPERIMENTAL AND COMPUTATIONAL STUDY ....................................................52

3.1 Overview ........................................................................................................52
3.2 Introduction .....................................................................................................52
3.3 Experimental ................................................................................................57
  3.3.1 Mass Spectrometry Methods ..................................................................57
  3.3.2 Computational Methods .........................................................................60
3.4 Results and discussion ..................................................................................62
  3.4.1 Calculated GAs and Structures for Tyrosine .........................................62
  3.4.2 Experimental GAs and Structures Produced from a Protic Solvent System ..............................................................63
  3.4.3 Effects of the Solvent on Deprotonated Tyrosine: A New Carboxylate Structure .........................................................70
  3.4.4 Deprotonated Tyrosine Structure as a Function of Hexapole Accumulation Time .................................................................76
3.5 Conclusions ................................................................................................81

References .......................................................................................................82

CHAPTER 4. GAS-PHASE ACIDITIES OF THE COMMON AMINO ACID AMIDES ......88

4.1 Overview ........................................................................................................88
4.2 Introduction ...................................................................................................88
4.3 Experimental ................................................................................................92
  4.3.1 Mass Spectrometry Methods ..................................................................92
  4.3.2 Computational Methods .........................................................................93
### Chapter 4. Results and Discussion

#### 4.4 Amino Acid Amides with Aliphatic Side Chains

4.4.1 Amino Acid Amides with Aliphatic Side Chains ................................................. 96

4.4.2 Amino Acid Amides with Amide, Basic, Hydroxyl, and Thioester Side Chains .................................................. 104

4.4.3 Amino Acids with Acidic Side Chains ................................................................. 111

#### 4.5 Conclusions

Conclusions .......................................................................................................................... 117

References .......................................................................................................................... 118

### Chapter 5. Gas-Phase Acidities of the Phosphorylated Amino Acids and Their Corresponding Amides

5.1 Overview ...................................................................................................................... 124

5.2 Introduction .................................................................................................................. 125

5.3 Experimental ............................................................................................................... 129

5.3.1 Mass spectrometry methods .................................................................................. 129

5.3.2 Computational methods ....................................................................................... 130

5.4 Results and Discussion .............................................................................................. 131

5.4.1 Phosphoserine (pSer) and Phosphothreonine (pThr): IR spectra, GA and Hydrogen Bonding ............................................................... 131

5.4.2 Phosphotyrosine (pTyr): IR spectra, GAs and Hydrogen Bonding ............... 144

5.4.3 pSer, pThr, and pTyr Amide GAs ......................................................................... 151

5.4.3 Comparison of Acidities of Phosphorylated Species with Non-Phosphorylated Species .............................................................. 155

5.5 Conclusions ............................................................................................................... 156

References ......................................................................................................................... 157

### Chapter 6. Fragmentation of Model Phosphorylated Peptides

6.1 Overview ..................................................................................................................... 165

6.2 Introduction .................................................................................................................. 166
6.3 Experimental ....................................................................................................................... 168
6.4 Results ................................................................................................................................ 169
  6.4.1 CID of \([M - H]^-\) for pThr-Containing Peptides ......................................................... 169
  6.4.2 CID of \([M - H]^-\) for pSer-Containing Peptides ......................................................... 171
  6.4.3 CID of \([M - H]^-\) for pTyr-Containing Peptides ......................................................... 172
  6.4.4 CID of \([M - 2H]^2^-\) for Phosphorylated Peptides ....................................................... 175
6.5 Discussion ............................................................................................................................ 176
  6.5.1 Product ions observed for CID of \([M - H]^-\) ............................................................... 176
  6.5.2 Effect of the Position of the Phosphorylated Residue on \([M - H]^-\) Fragmentation ................................................................................................................................. 182
  6.5.3 Diagnostic Ions Resulting from CID of \([M - H]^-\) Precursor Ions ...................... 186
  6.5.3 Diagnostic Ions Resulting from CID of \([M - 2H]^2^-\) Precursor Ions ............. 190
6.6 Conclusions ......................................................................................................................... 194
References ..................................................................................................................................... 198

CHAPTER 7. CONCLUDING REMARKS .................................................................................... 204
References ..................................................................................................................................... 208
LIST OF TABLES

3.1 Experimental, theoretical, and literature gas-phase acidities for tyrosine ........................................64

3.2 Reaction efficiencies for the proton transfer reaction of tyrosine with reference compounds .........................................................67

3.3 Reactions of trimethylsilyl azide with deprotonated tyrosine electrosprayed from various solvents..........................................................71

4.1 Experimental and computational values for the GAs of the amino acid amides ..................95

4.2 Reaction efficiencies for the proton transfer reactions of proline, valine, isoleucine, lysine, phenylalanine, methionine, and arginine amides with neutral reference compounds..........................................................99

4.3 Reaction efficiencies for the proton transfer reactions of arginine, threonine, serine, glutamine, and asparagine amides with neutral reference compounds .........................106

4.4 Reaction efficiencies for the proton transfer reactions of tryptophan, cysteine, tyrosine, and histidine amides with neutral reference compounds .................................113

5.1 Lowest energy gas-phase acidities (ΔG) and the corresponding enthalpies at 298 K at the G3(MP2) level in kcal/mol........................................132

5.2 Reaction efficiencies for the proton transfer reactions of phosphorylated amino acids and their amides ........................................................................137

5.3 Calculated acidities in kcal/mol of phosphoserine at different computational levels with respect to the most stable neutral conformer ...........................................140

6.1 Masses of common losses and diagnostic ions ........................................................................171

6.2 A summary of the diagnostic ion for specific phosphorylated residues of model phosphorylated peptides .................................................................190
LIST OF FIGURES

2.1 Schematic of ESI source for the Bruker 7e BioApex FT-ICR MS.................................24
2.2 Diagram of the Bruker BioApex 7 Tesla FT-ICR MS......................................................26
2.3 Ion transfer optics of the Bruker BioApex 7e FT-ICR MS with voltage profile for positively charged ions.........................................................................................27
2.4 Diagram of the “infinity” cylindrical analyzer cell for FT-ICR from side and front view.....................................................................................................................28
2.5 Fourier transform of FID to a peak typical for an FT-ICR experiment..........................29
2.6 Semi-logarithmic plot of the relative precursor ion intensity versus time of the reaction of deprotonated phosphothreonine with 1,1,1,5,5,5-hexa-2,4-pentadione ...........................................................................35
2.7 Peptide sequencing nomenclature....................................................................................41
2.8 Amino acids/amides with aliphatic side chains ...............................................................42
2.9 Amino acids/amides with basic side chains .....................................................................43
2.10 Amino acids/amides with aromatic side chains ..............................................................43
2.11 Amino acids/amides with acidic side chains ...................................................................43
2.12 Amino acids/amides with amide side chains ..................................................................44
2.13 Amino acids/amides with hydroxyl side chains.............................................................44
2.14 Amino acids/amides with sulfur containing side chains ...............................................44
2.15 Phosphorylated amino acids/amides .............................................................................45
3.1 Mass spectra of the disappearance of deprotonated tyrosine over time while reacting with neutral ethyl cyanoacetate at the pressure of 7.6 x 10^{-8} mbar.................................61
3.2 G3(MP2) results showing the structures of the most stable neutral acid and lowest energy anions resulting from proton loss.................................................................65
3.3 Reactant loss curve for the reaction of deprotonated tyrosine with ethyl cyanoacetate, which is present at a constant pressure of \(8.9 \times 10^{-8}\) mbar. The solvent is 49.5:49.5:1 (v/v/v) \(\text{CH}_3\text{OH:H}_2\text{O:NH}_4\text{OH}\) .................................................. 68

3.4 Mass spectrum resulting from the ion/molecule reaction of deprotonated tyrosine electrosprayed from the protic solvent system 49.5:49.5:1 (v/v/v) \(\text{CH}_3\text{OH:H}_2\text{O:NH}_4\text{OH}\) and trimethylsilyl azide at a constant pressure of \(6.5 \times 10^{-8}\) mbar with a reaction time of 3 seconds .................................................. 72

3.5 Reactant loss curve for the reaction of deprotonated tyrosine with 4-amino-2,3,5,6-tetrafluoropyridine (ATFP), which is present at a constant pressure of \(5.6 \times 10^{-8}\) mbar. The solvent is 99:1 (v/v) \(\text{CH}_3\text{CN:H}_2\text{O}\) .................................................. 74

3.6 Percentage of product ions from the reaction of deprotonated tyrosine with trimethylsilyl azide as a function of time that the ions accumulate in the hexapole. The solvent systems are (a) 74.5:24.5:1 (v/v/v) \(\text{CH}_3\text{OH:H}_2\text{O:NH}_4\text{OH}\), (b) 98:2 (v/v) \(\text{CH}_3\text{CN:H}_2\text{O}\), and (c) 99:1 (v/v) \(\text{CH}_3\text{CN:H}_2\text{O}\) .................................................. 80

4.1 Structure of an amino acid, amino acid amide, and peptide bonds................................. 89

4.2 Reactant loss curve for the reaction of proline amide (a) 1,2,4,5-tetrafluorobenzene (GA = 353.3 kcal/mol) at a constant pressure of \(7.8 \times 10^{-8}\) mbar and (b) 3-methylpyrazole (GA = 348.3 kcal/mol) at a constant pressure of \(9.0 \times 10^{-8}\) mbar ........................................................................ 98

4.3 The most stable neutral and two anions for proline amide at the G3(MP2) level. ........101

4.4 The most stable neutral and two anions for the aliphatic amino acid amides at the G3(MP2) level. .................................................................................................................. 102

4.5 Reactant loss curve for the reaction of glutamine amide with 4-trifluoroaniline at a constant pressure of \(9.9 \times 10^{-8}\) mbar ................................................................. 107

4.6 The most stable neutral and two anions for lysine, serine, glutamine, and asparagine amides at the G3(MP2) level .................................................................................................. 108

4.7 The most stable neutrals and anions for arginine amide at the G3(MP2) level ..........111

4.8 Reactant loss curve of deprotonated histidine amide reacting with ethyl cyanoacetate at a constant pressure of \(6.3 \times 10^{-8}\) mbar ........................................................................ 112

4.9 The most stable neutrals and anions for the amino acid amides with acidic side chains at the G3(MP2) level .......................................................................................................... 115
5.1 Reaction ion loss plot of \([\text{pThr} - \text{H}]^-\) with neutral 1,1,1,5,5,5-hexa-2,4-pentadione 
\(\text{GA} = 310.4 \text{ kcal/mol}\) .................................................................133

5.2 The most stable neutral and two ions for pSer and pThr at the G3(MP2) level .................134

5.3 Infrared spectra of deprotonated phosphoserine anions. IRMPD spectrum
is on top. Anions [A] – [F]: calculated spectra at the B3LYP/aug-cc-pvdz level.
Relative energies of isomers at the G3(MP2) level of theory..................................142

5.4 Infrared spectra of deprotonated phosphothreonine anions. IRMPD spectrum
is on top. Anions [A] – [F]: calculated spectra at the B3LYP/aug-cc-pvdz level.
Relative energies of isomers at the G3(MP2) level of theory.................................143

5.5 The most stable neutral and two ions for pTyr at the G3(MP2) level ........................145

5.6 Infrared spectra of deprotonated phosphothreonine anions. IRMPD spectrum
is on top. Anions [A] – [F]: calculated spectra at the B3LYP/aug-cc-pvdz level.
Relative energies of isomers at the G3(MP2) level of theory.................................146

5.7 Reactant ion loss plot for deprotonated phosphotyrosine reacting with
bisfluoroacetamide (GA= 307.5 kcal/mol) at a constant pressure of
3.0 x 10^-8 mbar .........................................................................................150

5.8 Optimized structures of phospho-serine, -threonine, and -tyrosine amide
neutrals and corresponding anions at the G3(MP2) level .....................................152

5.9 Reactant ion loss plots of (a) pSer, (b) pSer(amide), (c) pTyr, and
(d) pTyr(amide) anions reacting with neutral 1,1,1,5,5,5-hexafluoro-2,4-pentadione
at a constant pressure of 3.2 x 10^-8 mbar ..................................................155

6.1 SORI-CID spectra of \([\text{M} - \text{H}]^-\) from the pThr-containing peptides (a) pTAAAAA,
(b) AApTAAA, and (c) AAAApTA ...................................................................170

6.2 SORI-CID spectra of \([\text{M} - \text{H}]^-\) from the pSer-containing peptides (a) pSAAAAA,
(b) AApSAAA, and (c) AAAApSA .................................................................173

6.3 SORI-CID spectra of \([\text{M} - \text{H}]^-\) from the pTyr-containing peptides (a) pYAAAAA,
(b) AApYAAA, and (c) AAAApYA ..................................................................174

6.4 Structures of diagnostic and internal ions observed in CID mass spectra ..........175

6.5 CID spectra of \([\text{M} - 2\text{H}]^2-\) for (a) pSAAAAA, (b) pTAAAAA, and
(c) pYAAAAA .................................................................................................177
6.6 CID spectra of $[M - 2H]^2-$ for (a) AApSAAA, (b) AApTAAA, and 
(c) AApYAAA

6.7 CID spectra of $[M - 2H]^2-$ for (a) AAAApSA, (b) AAAApTA, and 
(c) AAAApYA

6.8 Ion structures for peptide fragment ions
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Reactions of deprotonated tyrosine with trimethylsilyl azide</td>
<td>56</td>
</tr>
<tr>
<td>6.1</td>
<td>Loss of neutral H$_3$PO$_4$ from phosphopeptides</td>
<td>185</td>
</tr>
<tr>
<td>6.2</td>
<td>Loss of metaphosphoric acid from phosphorylated peptide anions</td>
<td>185</td>
</tr>
<tr>
<td>6.3</td>
<td>Phosphoric acid migration to C-terminus</td>
<td>188</td>
</tr>
<tr>
<td>6.4</td>
<td>Loss of 79 and 97 from doubly deprotonated phosphopeptides</td>
<td>193</td>
</tr>
</tbody>
</table>
CHAPTER 1: OVERVIEW OF DISSERTATION

In the early 1990’s, the introduction of electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) completely changed the way biomolecules could be analyzed.\textsuperscript{1-3} Prior to the implementation of these ionization techniques, it was thought that large molecules would never be able to be ionized for analysis by mass spectrometry (MS). With ESI and MALDI available, understanding sequences of large biopolymers became a major research goal worldwide. In 2003, the Human Genome Project (HGP), the sequencing of human genetic code, was officially declared to be completed, opening the doors for the “omics” movement.\textsuperscript{4, 5} The Human Proteome Project (HPP), which involves the sequencing of all proteins and peptides in the human body, is well underway and uses MS as a necessary tool to decipher the coding of polypeptides.\textsuperscript{6-14} In the most recent HPP report, 3564 of 20123 protein coding genes are left to be sequenced.\textsuperscript{11}

Sequencing helps to clarify the energetic properties, structure, and function of peptides and proteins by providing information about functional groups contained on amino acid sidechains.\textsuperscript{15-18} Proteins have four levels of structure.\textsuperscript{19} The primary level is the coding of the polypeptide sequence in terms of amino acid residues. A polypeptide generally contains a C-terminal carboxylic acid group and an N-terminal amine. Secondary structure invokes hydrogen bonding to form helices, beta-sheets, or disordered regions. The tertiary level of structure is more complex and involves non-covalent interactions between the secondary structures to fold into low-energy structures. These structures often contain an active site, and the proteins can reside in
a holo or apo form. Finally, the quaternary structure includes the interactions of protein subunits to form large macromolecules, which make up an entire protein. To understand any information about the structure, function, and energetics of these complex biomolecules, the sequence (primary structure) must first be known. Knowledge of protein sequence and structure can be used in medicinal chemistry to better design drugs for particular functions.\textsuperscript{20-23}

Proteins are sequenced by use of MS in two major approaches. Both methods are paired regularly with bioinformatics automated software. In “bottom-up” proteomics, a protein isolated from an organism is subjected to enzymatic digestions.\textsuperscript{24-26} For example, trypsin digests cleave peptides on the C-terminal side of the basic amino acids lysine and arginine.\textsuperscript{26} The cleavage produces small peptides capable of being analyzed by MS \textsuperscript{1-3} and are fragmented by techniques by methods such as collision-induced dissociation (CID) or electron based dissociations (ECD, ETD).\textsuperscript{24, 25, 27-29} After the mass spectra are analyzed, by working backwards, the initial protein sequence can be obtained. Though automated techniques can obtain the original sequence information for some proteins, analysis can be done by the researcher, and is similar to solving a puzzle. Overall, bottom-up proteomics can provide useful information about protein sequences and is the conventional method in protein sequencing. However, disadvantages include cleavage of post-translational modification and the inability to distinguish repeat sequences.\textsuperscript{24, 30} Peptides with the same sequence produced by digestion would not be able to be differentiated by MS. Even with these issues, thousands of papers have been published using the bottom-up approach.

Less established, the “top-down” method involves ionization of a native protein complex with subsequent fragmentation. (CID, ECD, or ETD).\textsuperscript{31-33} Pairing top-down proteomics with ion mobility mass spectrometry (IMS), characteristics of protein structure can be determined.\textsuperscript{34, 35} By ionizing an entire protein, the exact mass can be easily determined and because no digests are
performed, “piecing together” the solution-phase fragments is not necessary. Proteins separated by gel electrophoresis can be directly analyzed. Top-down proteomics also can be useful for small proteins as enzymatic digests may not find available cleavage sites. Quantitation of proteins is more accurately obtained by the top-down method. Yet, there are disadvantages to top-down proteomics including: requirement for high-resolution, limitations with studying complex mixtures, and problems identifying proteins. In general, when performing a top-down proteomics experiment, the bioinformatics programs identify the most proteins in the first rounds of collecting data, but in following rounds, proteins are rarely identified. In automated techniques, after collecting data in the first round, the following rounds do not omit proteins and peptides identified originally by the first round. Already-identified protein ions can dominate spectra making it difficult to collect further data on unknown proteins. Top-down and bottom-up proteomics are complementary techniques used to obtain sequence information about peptides and proteins.

Most work in peptide sequencing involves the study of positive ions. The negative ion mode in MS can offer complementary, or even improved, sequence coverage than the positive ion mode. Some peptides are not capable of being studied in the positive ion mode as they lack basic sites to make protonated cations, [M + nH]⁺. Acidic peptides deprotonate more readily than they protonate, to produce [M – nH]⁻ that lend themselves well to being studied by negative ion mode MS. For example, the Cassady and Vincent groups have previously studied a biological chromium binding peptide, sequence of pEEEEGDD, which contains seven acidic sites including the C-terminus and does not make protonated ions by standard ESI conditions. (glutamic acid = E and aspartic acid = D). This peptide undergoes deprotonation to make negative ions and can be analyzed by CID.
To further the development of automated negative ion mode peptide sequencing, a better understanding of product ions produced by acidic peptides is required to build databases. An understanding of charge-site location is necessary, as CID and other fragmentation techniques are often charge-directed. Acidic peptides are abundant in the human body, and are not limited those to containing aspartic and glutamic acid residues. Phosphorylated peptides are common and are very acidic lending themselves to negative ion techniques. Bowie and coworkers have thoroughly studied many anions of small deprotonated peptides and amino acids. They have observed several unique losses and diagnostic ions for specific amino acids located in peptides capable of being deprotonated. Bowie and coworkers have published several papers investigating deprotonated phosphorylated peptides, with one study specifically examining the cyclization of peptides with multiple phosphorylated residues to create diagnostic product ions. In addition, the Cassady group previously has studied numerous peptides in the negative ion mode. For example, one CID study examined model deprotonated hydroxyl-containing peptides and proposed a mechanism for a unique aldehyde loss for peptides with serine and threonine residues. In the case of acidic phosphopeptides, diagnostic ions and neutral losses are commonly observed in the gas-phase from all phosphorylated amino acids.

All of the aforementioned studies implement CID, which commonly produces neutral losses that can provide limited sequence information. In contrast, techniques such as ETD and ECD (ExD) produce fewer neutral losses as they are non-ergodic, or not dependent on energy, however these are performed on multiply charged positive ions because after the addition of an electron, the resulting charge on the ion is decreased by one (an original charge of +1 would result in a neutral, unable to be analyzed by MS). Due to the need for \([M + nH]^{n+}\), ExD
techniques can be challenging for analyzing acidic peptides, as the production of even singly positively charged ions can sometimes be problematic. Negative electron-based techniques, nExD, are other ways to fragment and analyze negative peptide ions, and involve the addition of an electron to a negative charged precursor. But, the nExD techniques are less-developed and are not routinely performed. In contrast to ExD and nExD techniques, CID occurs on a longer timescale than electron-based techniques. The longer timescale in CID (ms) allows for rearrangements allowing the formation of the most energetically stable ion, or the loss of low-energy neutrals or ions drive the fragmentation to form specific product ions. More information on the types of fragment ions formed by CID of negative peptide ions is a useful contribution to proteomics.

Energetic studies of biomolecules are necessary for the examination of charge sites in biomolecules. Mass spectrometry has been used for years to determine thermodynamic values such as gas-phase basicity (GB) and gas-phase acidities (GA). Of interest for negative ion techniques specifically is the GA, the free-energy change of the deprotonation of a neutral.

\[ AH \rightarrow A^- + H^+ \]

1.1

The formation of negative peptide ions involves deprotonation. The C-terminal carboxylic acid and any available acidic side chain, commonly aspartic and glutamic acids, can deprotonate. But, many peptides created by enzymatic digests do not contain traditionally acidic sites, other than the C-terminus. Out of the 20 common amino acid residues, only two are classified as acidic (aspartic and glutamic acids). Also after digestion by trypsin, a basic residue is left on the C-terminus, making this residue more neutral due to salt-bridge possibilities. The study of amino acid analogues can help to determine alternate sites of negative charge for peptide anions. Amino acid amides, containing an amide at the C-terminus, do not have a C-terminal carboxylic acid to
deprotonate. An amino acid amide is a better representation of an amino acid residue incorporated into a peptide chain. By pairing MS in the form of ion/molecule reactions with computational chemistry, a clearer picture of alternate acidic site deprotonation can be obtained.

Gas-phase acidity can give a clearer picture to the deprotonated sites in a peptide, which helps in the comprehension of charge-directed fragmentation. In addition, understanding the charge site can help to determine mechanisms of fragmentation.

The charge site and structural conformation of biomolecules can be effected by ESI source conditions. For proteins, many studies have shown how temperature dependence and solvent composition change the gas-phase ion conformation. This effect has not been studied significantly for small peptides and amino acids. Techniques such as ion mobility, infrared multiphoton dissociation (IRMPD), and ion/molecule reactions can give information relating to structure and energetics of ions created by ESI. Understanding how ESI conditions can alter deprotonation sites and ion conformation is necessary when studying the structure and energetics of a biomolecule. By having a strong understanding of ion structure, energetics, and fragmentation, data bases can build up information about peptides studied by negative ion mode MS for use in bioinformatics for automated sequencing techniques.

The work presented in this dissertation is a study of negatively charged amino acids, amino acid amides, phosphorylated amino acids, phosphorylated amino acid amides, and model hexa-phosphorylated peptides. Many of the techniques discussed in this overview were utilized in this study. Ion/molecule reactions were used to examine the GA of amino acids and amino acid analogues. Collision-induced dissociation was performed to understand the fragmentation associated with the location and type of specific phosphorylated amino acids when incorporated
into a peptide. The research completed for this dissertation furthers the field of proteomics by expanding the knowledge associated MS research upon negative ions of small biomolecules.

Chapter 2 discusses the experimental and theoretical procedures used for this research. The components of mass spectrometers will be discussed, as the entire process for a MS experiment is described. The ionization source, analyzer, detector, and tandem MS techniques (MS/MS) are incorporated in Chapter 2. Derivations of equations relating to mass/charge, reaction rate, and energetics are given. Experimental procedures relating to ion/molecule reactions and CID are listed. As the hexa-phosphorylated peptides were synthesized in house, the synthetic procedures for solid-phase peptide synthesis are given in detail. The nomenclature for peptide fragmentation is addressed, and the structures for the amino acids, amino acid amides, phosphorylated amino acids, and phosphorylated amino acid amides are given.

Chapters 3-5 deal with the deprotonation of amino acids and amino acid amides. Though previous studies show that the carboxylic acid is the most acidic site on an amino acid,98, 122 amino acid amides are analogues that change to remove the carboxylic acid and are better models of an amino acid residue which has been incorporated into a peptide chain. Some amino acids also have side chains capable of deprotonation. The research in these chapters discusses the GA of tyrosine, the amino acid amides, and the phosphorylated amino acids/amides. Ion/molecule reactions were used to experimentally bracket GA values by using a Bruker BioApex 7e Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS). High-level G3(MP2) calculations were completed by Dr. David Dixon’s group at The University of Alabama and are in good agreement with the experimental work. Computational studies are helpful as they predict sites of charge and gas-phase ion conformation.
Chapter 3 specifically discusses the gas-phase acidity of tyrosine. Tyrosine is an amino acid commonly found in peptides in proteins which has a sidechain phenolic functionality. The GA of tyrosine has been previously assigned by several methods.\textsuperscript{98, 122, 123} Tyrosine has been found to deprotonate on its sidechain in former studies by Tian and Kass;\textsuperscript{120, 123} and in a separate study by Oomens,\textsuperscript{124} no evidence of side chain deprotonation was observed. The formation of both a phenolic and carboxylate anion are observed by Tian and Kass depending on the solvent condition used for ESI. In this study, the lowest energy structure observed was a phenoxide anion. The research in this chapter addresses how solvent conditions can effect conformation of anions, which relates to their energetic properties. A novel probe reaction developed by Kass was used to determine the site of deprotonation from tyrosine.\textsuperscript{123} In the work of Chapter 3, twenty solvent systems were used to electrospray tyrosine; and it was found that aprotic solvents yielded two energetically differentiable carboxylate anions by performing ion/molecule reactions. The two ion structures were differentiated by rate plots with two distinguishable reaction rates that were fit to the sum of two exponential decays. A new, lower energy carboxylate structure was found that agrees with computational predictions by the Dixon group. After the isolation of the lower energy carboxylate was obtained, the GA values for low- and high-energy carboxylate structures could be assigned. In addition, the ESI conditions relating to the hexapole used for ion storage showed evidence for a gas-phase conversion between carboxylate to phenolic anions over time. The conversion is likely a result of reactive collisions within the hexapole.

Chapter 4 involves the study of the common amino acid amides. As previously discussed, these amino acid analogues have an amide in place of a carboxylic acid C-terminus. Amino acid amides are more representative than amino acids of residues in peptides. Except in the case of aspartic and glutamic acid amides, an amino acid amide is lacking a traditionally acidic site for
deprotonation. Both aspartic and glutamic acid have been previously studied by the Cassady and Dixon groups.\textsuperscript{88} For amino acid amides with aliphatic (glycine, alanine, proline, valine, leucine, isoleucine, and phenylalanine), basic (lysine and arginine), thioester (methionine), hydroxyl (serine and threonine), and amide (glutamine and asparagine) side chains, two ion populations were observed in ion/molecule reactions varying in energy by ~5 kcal/mol. Computations predicted structures at this energy difference attributed to the conformation of the amide anions. Deprotonation occurs on the C-terminal amide. Cis- and trans- like structures yielded different GA values. In addition, some amino acid amides gave only one ion population, and computations predict side chain deprotonation. These amino acid amides include: tyrosine (phenolic side chain), cysteine (thiol side chain), tryptophan (indole side chain), and histidine (imidazole side chain). For amino acids containing one ion population, these can be considered as potential sites of negative charge when analyzing peptides by negative ion mode MS. Experimental GA values were assigned for 18 amino acid amides, where 14 amino acid amides had two conformations of varying GA value.

Chapter 5 discusses a study of the GAs of the three common phosphorylated amino acids and their corresponding amides. The most common amino acids to be phosphorylated have hydroxyl side chains and are serine (pSer), threonine (pThr), and tyrosine (pTyr). Since phosphates are highly acidic, they readily provide a deprotonation site for creating negative ions. In addition, the carboxylic acid can provide an additional acidic site for phosphorylated amino acids. Ion/molecule reactions of both deprotonated phosphorylated amino acids and amides yielded linear rate plots indicating one major ion population. This contrasts to the work discussed in Chapters 3 and 4. Computational results agree well and found that the lowest energy structures were deprotonated at the phosphate group except in the case of pTyr that was a carboxylate.
Structures were compared to infrared multiphoton dissociation (IRMPD) work by Maitre and coworkers,\textsuperscript{115} for which our pSer and pThr calculated anions agreed with experimental IRMPD data. In the case of deprotonated pTyr, the lowest energy structure did not agree with Maitre’s experimental IRMPD. Solvent and conditions used for ESI in Maitre’s work were unusual and were investigated in this dissertation. pTyrosine anions were analyzed by ion/molecule reactions using the Matire conditions, and were found to be slightly less acidic than pTyr anions produced by our experimental conditions. In both Chapter 5 and Chapter 3 the role of the ESI conditions on ion structures and energetics has been investigated in detail.

Chapter 6 discusses the fragmentation of deprotonated ions produced by ESI on model hexa-phosphorylated peptides. As mentioned earlier in this chapter, understanding fragmentation pathways associated with negative peptide ions is important in the proteomics movement. Also, CID readily cleaves phosphate groups making it difficult to assign where this post-translational modification original resided in a peptide with multiple sites capable of phosphorylation. Diagnostic ions specific for individual phosphorylated amino acids can help to clarify CID data obtained in the negative ion mode. Nine model hexapeptides, XAAAAA, AAXAAA, and AAAAXA (where X = phosphorylated residue, pSer, pThr, pTyr) were synthesized in house and subjected to sustained off-resonance irradiation (SORI) CID, as well as low-energy CID on a quadrupole ion trap. Diagnostic ions were found for pSer- and pThr- containing peptides, as well as diagnostic marker ions for the phosphate group itself. Both location and type of the phospho-residue, as well as the charge on the ion subjected to CID have an effect on fragmentation observed.
Chapter 7 summarizes the important findings in this dissertation research and reflects upon the impact this work has on the fields of MS and proteomics. Additional experiments are proposed that may expand this work for future researchers.

References


(8) Michalski, A.; Cox, J.; Mann, M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 1785-1793.


(11) Omenn, G. S.; Lane, L.; Lundberg, E. K.; Beavis, R. C.; Nesvizhskii, A. I.; Deutsch, E. W. Metrics for the Human Proteome Project 2015: Progress on the Human Proteome and


(23) Williams, J. P.; Brown, J. M.; Campuzano, I.; Sadler, P. J. Identifying Drug Metallation Sites on Peptides using Electron Transfer Dissociation (ETD), Collision Induced


(69) Waugh, R. J.; Bowie, J. H.; Gross, M. L.; Vollmer, D. Collision Induced Dissociations


(81) Molina, H.; Matthiesen, R.; Kandasamy, K.; Pandey, A. Comprehensive Comparison of


CHAPTER 2: EXPERIMENTAL INSTRUMENTATION AND THEORY

2.1 Overview

This chapter discusses the experimental techniques and theory behind the research completed for this dissertation. Electrospray ionization (ESI) was the ionization source for all work and is described. Multi-stage tandem mass spectrometry procedures such as collision-induced dissociation (CID) and ion/molecule reactions (I/M) were utilized to obtain peptide fragmentation and gas-phase acidity (GA) values. Peptides were synthesized in this research, so this chapter concludes with information regarding peptide synthesis, fragmentation nomenclature, and amino acid structures.

2.2 Electrospray Ionization

Electrospray ionization (ESI) is readily used for the study of large, polar, molecules by mass spectrometry. Because these are common characteristics of biomolecules, ESI has become a common-place ionization technique for their analysis. The compatibility of ESI to be paired with separation techniques, such as liquid chromatography (LC)\textsuperscript{1, 2}, has made ESI even more routine. Based on the early work of Dole and coworkers\textsuperscript{3} and later developed by John Fenn,\textsuperscript{1, 4, 5} ESI has made an incredibly large impact on the world of mass spectrometry. Fenn shared the Nobel Prize in Chemistry in 2002 because of the ability of ESI to allow ionization of large molecules and biomolecules.\textsuperscript{6}

Electrospray ionization is a process in which a solution-phase sample undergoes desolvation into the gas-phase.\textsuperscript{5} The mechanism of ESI has been the topic of many studies, with
the ion evaporation model and charge residue model being the most accepted explanations for the formation of ions by ESI.\textsuperscript{3, 5, 7-13} The ESI process is considered a “soft” ionization technique, meaning that only quasi-molecular ions with little to no fragmentation (example \([M–H]^–\)) are produced. To produce ions by ESI, a solution must be prepared and injected into the ESI source. In the work presented in this dissertation, the concentration of the solution was \(~60\, \mu\text{M}\). The solvent used to dissolve the molecule of interest generally contains some ratio of methanol, water, and acetonitrile. Addition of ammonium hydroxide (1-5\%) is common for producing deprotonated ions, because if the molecule in the solution-phase prior to ESI can be pre-ionized, greater signal intensity is observed in the mass spectrum. Once a solution is prepared, it is loaded into a syringe and is slowly pushed at a constant rate \((\sim 100\, \mu\text{L/hr})\) into the ESI needle with a syringe pump.

After entering the source region, desolvation of the sample occurs by use of a heated drying gas set at a constant temperature and pressure. The desolvation process creates the quasi-molecular ions. For negative ion formation, the preferred drying gas is air. The drying gas temperature is usually between 200-250\(^\circ\text{C}\), and the drying gas flow rate is \(~5\, \text{L/min}\). A nebulizing gas is used to direct the ions formed by the ESI process to the ESI capillary region. The nebulizing gas is also air for the Bruker 7e BioApex FT-ICR MS (Fourier transform ion-cyclotron resonance mass spectrometer). After desolvation, the negative ions produced are attracted by a positive 3.5-4.0 V to the capillary and endcap electrode. The glass capillary is coated on both sides with platinum to allow two separate voltages to be applied to each end. Once in the capillary region, the ions are moved through two skimmers and a pre-hexapole to where they are stored in a hexapole. The hexapole can be altered to store ions for various amounts of time before they are pulsed into the FT-ICR’s ion optics where they are directed into
the FT-ICR cell. Other instruments, such as quadrupole ion traps, implement octopoles instead of hexapoles for ion storage. This dissertation demonstrates how the hexapole accumulation time can have a significant effect on ion structure (Chapter 3). A diagram of the ESI source is shown in Figure 2.1.

![Figure 2.1. Schematic of ESI source for the Bruker 7e BioApex FT-ICR MS.](image)

During the ESI process, charged droplets are formed during the nebulization and desolvation process. Large amounts of charge build up in these droplets by addition or subtraction of protons and creates quasi-molecular ions. When the charge has reached the Rayleigh limit, Coloumbic explosions occur which creates the ions formed during ESI. The charge of these ions is usually indicated by the amount of basic or acidic sites a molecule has. For example, in hirudin, an acidic peptide containing many high-acidity glutamic acid residues (E) and one aspartic acid residue (D) can create several quasi-molecular ions because of its abundance of acidic sites. The sequence is GDFEEIPEEYLQ. Hirudin has the ability to deprotonate at each acidic amino acid and the C-terminus (six acidic sites) allowing for quasi-molecular ion production of \([M - nH]^{-}\) where \(n = 6, 5, 4, 3, 2, \text{ or } 1\).
2.3 Mass Analyzers

2.3.1 Fourier Transform Ion Cyclotron Resonance

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) was developed in 1974 by Comisarow and Marshall.\textsuperscript{14} By introducing the Fourier transform to ICR MS, data acquisition became faster, allowed for more data points, and the \textit{m/z} range of observed ions in a given spectrum could be expanded. An exceptional figure of merit for the FT-ICR is the high resolution it is capable of attaining. For example, in an FT-ICR experiment (using a 9.4-T magnet), two peptides only 0.00045 Da can be distinguished from one another; Marshall\textsuperscript{15} discusses that this is less than the mass of an electron (0.00055 Da). FT-ICR can study isotopes of atoms that are not routinely differentiated by mass spectrometry, such as oxygen and nitrogen.\textsuperscript{16} In addition to high resolution, FT-ICR can perform many types of experiments by modifying the pulse sequences available to the instrumental software. Low pressures in the ultra-high vacuum range ($10^{-9}$-$10^{-10}$ mbar) allow ions to remain trapped for extensively long times. The low pressure keeps ions from undergoing unwanted reactions with neutrals in the cell. Ion/molecule reactions can be used to obtain kinetic data by trapping ions for various amounts of time (from ms to days). Collision-induced dissociation (CID) and electron capture dissociation (ECD) can give sequence informative fragmentation about the structures of molecules and biomolecules. The FT-ICR is compatible with many types of ion sources. Having a choice of ion source allows FT-ICR to analyze nearly any compound. Because of its ability to be used for so many types of experiments, the FT-ICR has been referred to as a “gas-phase laboratory.”\textsuperscript{17}

The FT-ICR used for the experiments in this dissertation is a Bruker (Bruker Daltonics, Billerica, MA) BioApex 7e FT-ICR MS (Figure 2.2). Ions are generated with an Apollo API source. The FT-ICR is equipped with a 7-Tesla refrigerated superconducting magnet. Ions are
moved into the FT-ICR cell by electrostatic focusing. To transfer ions efficiently, a series of plates, lenses, and tubes have applied voltages that move ions to the FT-ICR cell. Ions are accelerated at a high voltage to allow them to overcome the magnetic field produced by the superconducting magnet. The ion transfer optics used in our FT-ICR can be viewed in Figure 2.3. Before entering the cell, the ions must be decelerated. The cell traps ions by using a low voltage, ~1 V. Once in the cell, image current measured on the detection plates (Figure 2.4) is used to detect the ions.

Figure 2.2. Diagram of Bruker BioApex 7 Tesla FT-ICR MS.

The FT-ICR cell is both the analyzer and detector of this instrument. The specific cell on our FT-ICR is a cylindrical “infinity” analyzer cell designed by Carvatti and Alleman\textsuperscript{18} (Figure 2.4). The ions enter the analyzer cell through a small hole after being transferred through the ion optics. The analyzer cell has six plates which have different functions for trapping ions. The
trapping plates have a low voltage of ~1 V used to contain ions. The polarity of this voltage depends on the type of ion produced. The excitation plates have a sinusoidal radio frequency (rf) voltage applied to excite ions into higher orbits to allow them to be detected. An image current is produced by the excited ions moving in an alternating coherent motion, towards and away from the detection plates. By using the Fourier transform, the free-induction decay produced is converted into a mass spectrum (Figure 2.5). The Fourier transform takes time domain data and converts it into the frequency domain. The relationship between frequency and mass-to-charge \((m/z)\) ratio is further described below in Equation 2.8.

In the FT-ICR cell, ions undergo three types of motion: cyclotron, trapping, and magnetron motion. Cyclotron frequency is the most important type of motion in regards to how ions move in the cell. Cyclotron frequency stems from the magnetic force, or Lorentz force, that an ion experiences when entering the magnetic field, as well as the centrifugal force that causes ions to move in a circular orbit. When the Lorentz and centrifugal forces are equal, ions will undergo cyclotron motion and orbit in the FT-ICR cell. The cyclotron frequency is dependent
Figure 2.4. Diagram of the “infinity” cylindrical analyzer cell for FT-ICR from side and front view.
Figure 2.5. Fourier transform of FID to a peak typical for an FT-ICR experiment.

\[ F_1 = \text{magnetic force} = Bzv \]

\[ F_2 = \text{centrifugal force} = \frac{mv^2}{r} \]

\[ Bzv = \frac{mv^2}{r} \]

\[ \frac{Bz}{m} = \frac{v}{r} \]

\[ \omega = \frac{v}{r}, (\omega = \text{angular orbital frequency}) \]

\[ \frac{B}{\omega} = \frac{m}{z} \]

\[ \omega = 2\pi f, (f = \text{ion cyclotron frequency in Hz}) \]

\[ \frac{m}{z} = \frac{B}{2\pi f} = \frac{1.537 \times 10^7 B}{f} \]

Where \( B = \text{magnetic field}, z = \text{charge of ion}, v = \text{velocity}, m = \text{mass of ion}, r = \text{radius}. \)
only on the field strength of the superconducting magnet and the mass-to-charge ratio of the ion of interest as derived by the equations in 2.1-2.8. As seen in equation 2.8, the \( m/z \) of any ion can be determined by measuring the cyclotron frequency of an ion while the magnetic field is held constant. For example, an ion with an \( m/z \) of 200 in a 7 Tesla magnetic field will have a cyclotron frequency of \( \sim 538 \) kHz. Trapping motion is also significant because it keeps the ions in the cell moving back and forth between the trapping plates. Magnetron motion does not play a role in mass analysis.

An experimental set-up for FT-ICR includes use of a pulse program. Before inserting the ions generated by the ion source into the analyzer cell, first a voltage pulse occurs to dispel any ions remaining in the cell. This is known as quenching. The voltage applied during the quenching step is opposite in polarity to the ions that will be experimentally analyzed. The opposite polarity attracts these ions to the plates where they are neutralized and not detected. Next, the ions enter the cell by the ion transfer optics. Once in the cell, to perform tandem mass spectrometry experiments on a specific ion, isolation of a specific \( m/z \) of an ion is performed. Isolation places correlated frequency sweeps and shots to eject unwanted ions by exciting them to higher orbits where they collide with ICR plates and are neutralized.\(^{19}\) For example, a sweep would generate a radiofrequency (rf) pulse that is placed on all unwanted ions cyclotron frequencies. The rf sweep places unwanted ions into resonance where their velocity and orbital size increase so that they are ejected from the cell, leaving only the ion with the \( m/z \) of interest. After isolation, experiments such as ion/molecule reactions or fragmentation techniques can be performed on an ion population of a single isolated \( m/z \).
2.3.2 Quadrupole Ion Trap

The quadrupole ion trap (QIT) is another type of pulsed mass analyzer that relies on a
type of ion frequency to detect ions. It was developed by Paul and Demelt, who were awarded
the Nobel Prize in Physics for its invention. There are both linear and three-dimensional ion
traps, both of which are used to study molecules and biomolecules. The specific instrument
used for this dissertation is the Bruker HCTultra PTM Discovery system equipped with ESI. This
is a high-capacity QIT, which is especially efficient at trapping ions. The research using this
instrument was performed as a comparison to sustained off-resonance irradiation (SORI) CID
experiments as described in Section 2.4.2.

Ions are transmitted from electrospray source to the QIT by movement through two
octopoles and a split lens. Ions are then pulsed into the QIT after a storage time in the first
octopole. The QIT employs three electrodes: two endcaps and a ring electrode. The QIT traps
ions by the application of a DC potential on the endcap electrodes while maintaining an rf
potential on the ring electrode. The potentials create an oscillating quadrupolar field that traps
and excites ions.

The ions experience a type of frequency which makes QIT most similar to motion
observed in the FT-ICR. However, the frequency of ions in a QIT is erratic and complicated.
Ions do not orbit in coherent motion. Ions of interest can be isolated by frequency-based ejection
techniques and subjected to fragmentation experiments such as collision-induced dissociation.
Ions are detected by a photomultiplier tube.
2.4 Tandem Mass Spectrometry (MS/MS)

2.4.1 Ion/Molecule Reactions

The determination of thermodynamic properties such as gas-phase acidity (GA) or gas-phase basicity (GB) can provide insight into the ability of a molecule to deprotonate or protonate in the gas phase.\textsuperscript{24, 25} Energetic properties are increasingly useful when studying biomolecules such as peptides, as many amino acids have functional groups that are acidic or basic. Fragmentation in mass spectrometry is usually charge-directed, so understanding locations of charge is especially useful for proteomics analysis.\textsuperscript{24, 26}

There are several methods for determining GA or GB values. The equilibrium method was established by Kebarle and coworkers in 1963,\textsuperscript{27} and since then has been used to determine equilibrium measurements for a wide array of organic and inorganic species.\textsuperscript{28-31} The equilibrium method works by generating ions from a field-free high-pressure region and accelerating ions into a low-pressure region where analysis occurs (such as an ICR cell). The equilibrium method can be performed with ICR when both compounds are volatile and can have neutrals in the cell region, which does not involve high and low pressure regions. Flowing afterglow sources have also been utilized to obtain equilibrium measurements, where ions generated are moved into a helium flow system and eventually are introduced to neutral reactants. The problem with the equilibrium method is that both the sample of interest and the neutral reactant must be volatile enough to be in the gas phase. Amino acids and peptides are not volatile, and thus the equilibrium method is not ideal for determining properties of biomolecules. A separate method for determining energetic properties was developed by Cooks and is known as the kinetic method (later expanded to the extended-kinetic method).\textsuperscript{32-35} This method uses the relative rates of dissociation for proton-bound dimers of an analyte of interest and a neutral reference compound.
of known basicity. The kinetic method has been used to determine proton affinities (PA), GB, and GA values for amino acids and small peptides. The limitation of the kinetic method is the formation of the dimer species. If the analyte of interest does not form a proton-bound dimer with a given reference compound, the kinetic data associated with dissociation cannot be obtained. Also, the structure of an ion as part of a dimer may not adequately represent its isolated structure. If the structure is altered, it may have an effect on the experimentally determined GA or GB values.

A third method, the thermokinetic method, was used in this research. This method can be used with non-volatile compounds, such as biomolecules which readily ionize by ESI. After ions are produced, they are moved into the FT-ICR cell where a reaction between the ion of interest occurs with a neutral reactant gas present in the cell at a constant pressure. The neutral reactant is pressure controlled by use of a sapphire-sealed leak valve. The research presented in this dissertation used the thermokinetic method to determine GA values for amino acids, amino acid amides, and phosphorylated amino acids.

The gas-phase acidity is defined as the negative free energy change (\(\Delta G^\circ\)) for the proton loss of \(M\) (molecule of interest):

\[
M \rightarrow [M - H]^- + H^+ \quad -\Delta G^\circ = GA
\]

2.9

The proton transfer reaction between \(M\) and neutral reactant, \(A\), of known GA, in the FT-ICR is as follows:

\[
\]

2.10

To assign GA values by the thermokinetic method, reaction efficiencies are used to determine when the reaction has become sufficiently exoergic. Reaction efficiency is obtained by taking the ratio of the experimentally determined rate constant (\(k_{\text{exp}}\)) and a theoretical rate constant (\(k_{\text{cap}}\)).
\[ RE = \frac{k_{exp}}{k_{cap}} \]  \hspace{1cm} 2.11

The theoretical rate constant, \( k_{cap} \) is determined by collision theory:

\[ k_{cap} = k_L K_{cap} \]  \hspace{1cm} 2.12

The components of \( k_{cap}, k_L \) and \( K_{cap} \) are derived in Equations 2.13-2.20:

\[ k_L = 2\pi q \sqrt{\frac{\alpha}{\mu}} = 2342 \sqrt{\frac{\alpha \times 10^{-24}}{\mu}} \times I \]  \hspace{1cm} 2.13

Where \( q = \) charge of an electron, \( \alpha = \) polarizability of the neutral, \( \mu = \) reduced mass, and \( I = \) moment of inertia.

\[ K_{cap} = K(T_R I^*) \]  \hspace{1cm} 2.14

\[ T_R = \frac{2\alpha K_B T}{\mu^2 D} \]  \hspace{1cm} 2.15

\[ I^* = \frac{\mu D}{ae\mu} \]  \hspace{1cm} 2.16

\[ K_{cap} = \frac{(x+0.5090)^2}{10.526} + 0.9754, \ x \leq 2 \]  \hspace{1cm} 2.17

\[ K_{cap} = 0.4767 + 0.9754, \ 2 \leq x \leq 3 \]  \hspace{1cm} 2.18

\[ K_{cap} = 0.4767 + 0.9754, \ 3 \leq x \leq 35 \]  \hspace{1cm} 2.19

\[ x = \frac{1}{T_R} = \sqrt{\frac{\mu D^2}{2\alpha K_B T}} \]  \hspace{1cm} 2.20

Where \( T_R = \) reduced temperature (as defined in Equation 2.15), \( \mu_D = \) dipole moment of neutral, \( K_B = \) Boltzmann constant, \( T = \) temperature, and \( e = \) elementary charge.

Experimental rate constants are determined by obtaining a reactant ion decay plot. This plot (Figure 2.5) displays the relative ion intensity of the precursor ion over time. The kinetics of this reaction are pseudo-first order, as the reactant neutral is present in overwhelming abundance compared to the ion intensity produced by ESI. The slope of the line is used to calculate the experimental rate constant (Equation 2.21):
The relative sensitivity of the ion gauge is $R_x$ which is specific to the neutral reactant. The experimental pressure, $P_{\text{exp}}$ is read directly from the pressure meter. The pressure is converted between mbar and Torr by a factor of 1.333. A correction factor for the positioning of the Bayerd-Alpert ion gauge, $IG_{\text{corr}}$ is used (the gauge cannot be located in the FT-ICR cell). To convert between pressure and number of molecules, a conversion factor is used. There are $3.23 \times 10^{16}$ molecules/cm$^3$ in one Torr. The units of $k_{\text{exp}}$ are cm$^3$·molecule$^{-1}$·s$^{-1}$. Once the slope of the line is obtained, the ratio of $k_{\text{exp}}/k_{\text{cap}}$ indicates whether the reaction has reached its “break point.”

Gas-phase acidity (and basicity) values are assigned at an RE = 0.269, where the reaction has become exoergic. Energetic values are bracketed between neutrals of known GA (or GB) value. A lower numerical GA value indicates a more highly acidic compound.

$$k_{\text{exp}} = \frac{slope(1.333)R_x}{(3.23\times10^{16})P_{\text{exp}}IG_{\text{corr}}}$$

Figure 2.6. Semi-logarithmic plot of the relative precursor ion intensity versus time of the reaction of deprotonated phosphothreonine with 1,1,1,5,5,5-hexa-2,4-pentadione.
The ion gauge was calibrated to obtain IG$_{corr}$ by proton transfer reactions of protonated glycine and N,N-dimethylformamide (DMF):

\[
[Gly + H]^+ + DMF \rightarrow [DMF + H]^+ + Gly
\]  

The well-established experimental reaction rate for this reaction\textsuperscript{38} is $8.19 \times 10^{-10}$ cm$^3$·molecule$^{-1}$·s$^{-1}$ and the average experimental reaction rate was $6.77 \times 10^{-10}$ cm$^3$·molecule$^{-1}$·s$^{-1}$. This yields an IG$_{corr}$ value of 1.21.

In many cases (Chapter 3 and 4), the experimental rate plots obtained by ion/molecule reactions did not yield one ion population. In a pseudo-first-order kinetic plot, a straight line of one ion population reacting at one rate is observed. However, for plots containing two ion populations reacting at sufficiently different rates, this is not the case. Here, to obtain $k_{exp}$ values for each ion population, data was fit to the sum of two exponential decays.

Chapter 3 discusses I/M reactions specific to a well-developed probe reaction, between trimethylsilyl azide (TMSN$_3$), present at a constant pressure, and deprotonated tyrosine, produced by ESI and $m/z$-selected with resonance frequency ejection.\textsuperscript{52, 53} For this reaction, when deprotonated tyrosine was reacted with TMSN$_3$, two product ions were observed which indicated the site of deprotonation. This reaction was not used to determine GA values, but provided valuable insight to the gas-phase structure of tyrosine anions.

### 2.4.2 Sustained Off-Resonance Irradiation Collision-Induced Dissociation

Collision-induced dissociation (CID) has a long history of use in sequencing and energetic studies of ions by mass spectrometry.\textsuperscript{54} In CID, ions are excited and undergo collisions with a neutral gas, such as argon or helium. Kinetic energy is internalized by the precursor ion which results in fragmentation. In an FT-ICR, sustained off-resonance irradiation (SORI) CID
uses a low-energy, off-resonance frequency while pulsing in “high” pressures (10⁻⁵ mbar) of a collision gas.

Collision-induced dissociation experiments (including SORI) have been major tools in sequencing biomolecules, especially peptides and proteins.⁵⁵⁻⁵⁷ However, because CID is energy-directed²⁴ and the lowest energy process generally occurs. This results in b- and y-type ions (Section 2.5.2) or cleavages from side chains of peptide ions.²⁴,²⁶ The CID process is a longer timescale than electron-based fragmentation techniques, and often ions undergo rearrangements to lose an energetically favorable neutral, or to make an especially stable ion. These rearrangements can be specific to a particular amino acid, resulting in diagnostic ions for a specific amino acid residue.³⁸⁻⁷¹ The research presented in this dissertation focuses on CID of negative peptide ions, an area much more underdeveloped than positive ion studies.

A SORI-CID experiment requires the isolation of a precursor ion, as discussed in Section 2.3.1. Once isolated, a low-energy, rf pulse is placed on the excitation plates at a frequency near the cyclotron frequency of the precursor ion. Generally, the frequency is ± 700 Hz from away from the precursor ion. The excitation frequency alternates from being in-phase and out-of-phase with the precursor ion, causing acceleration and deceleration, resulting in excitation.

The excited ions collide with a neutral collision gas. In the experiments in this dissertation, the inert gas used was argon. Argon is a good choice for a collision gas, as it is heavier than helium or neon, which allows for more energy transfer in collisions, and it is readily available. By using a pulsed valve, argon is introduced into the cell. The analyzer pressure is raised from 10⁻⁹ mbar to 10⁻⁵ mbar, and all argon is pumped away within ~2 seconds. Collisions
occur between the inert gas and the precursor ion. The kinetic energy of the precursor ion after collisions is represented by the equation:

$$E_{lab} = \frac{\beta^2 q^2 V_{pp}^2}{16 md^2 (\omega_c - \omega_c^2)^2} (1 - \cos(\omega - \omega_c)t)$$

where \(m\) is the mass and \(q\) is the charge of the ion, \(\omega_c\) is the cyclotron frequency of the precursor ion, \(\omega\) is the frequency of the excitation field, \(V_{pp}\) is the peak-to-peak excitation voltage, \(\beta\) is the geometry factor of the FT-ICR cell, \(t\) is the excitation time, and \(d\) is the diameter of the cell. By pulsing argon into the cell, the excited precursor ion readily undergoes multiple collisions. The equation for the center-of-mass-collision energy \(E_{cm}\) of is described in Equation 2.24:

$$E_{cm} = \left(\frac{M_N}{M_N + M_I}\right) E_{lab}$$

\((M_N = \text{neutral gas molecule, } M_I = \text{Mass of Ion})\)

The kinetic energy of the ion is dependent on the peak-to-peak voltage excitation \((V_{pp})\) and the offset frequency. The peak-to-peak voltage can be determined by measuring the actual voltage output of our instrumental power amplifier, \(V_{p-p}^{amp}\) and the associated excitation attenuation, \(AT\).

$$V_{p-p} = V_{p-p}^{amp} \times 10^{-AT/20}$$

During the SORI-CID process, the ions collide with hundreds of argon atoms and receive enough energy to dissociate. The SORI-CID process works well because it results from internal energy transfer to the precursor ion by use of a long activation time with several low-energy collisions with an inert gas.

2.5 Peptide Synthesis and Structures

2.5.1 Fmoc Solid-Phase Peptide Synthesis

A common way of synthesizing peptides for analysis by mass spectrometry is solid-phase peptide synthesis (SPPS). It was first introduced by Merrifield in the 1960s, for which he won
Solid-phase peptide synthesis is commonly automated. In this dissertation research, an automated peptide synthesizer, Advanced ChemTech Model 90 (Louisville, KY, USA), was used to synthesize model hexapeptides.

In SPPS, peptides are built from C-terminus to N-terminus. The C-terminal amino acid is attached to a solid support, known as a Wang resin. Synthesis adds one amino acid residue at a time to the Wang resin in a series of washing, deprotection, coupling, and final washing steps in a reaction vessel. Solvents used for washing include methanol (MeOH) and dimethylformamide (DMF). The deprotection is completed by using a solvent 20% by volume piperidine (Pip) in DMF. The amino acid to be coupled is dissolved to form a 0.5 M solution 0.5 M 1-hydroxybenzotriazole (HOBt) in N-methyl-2-pyrrolidinone (NMP). Once the amino acid solution is introduced to the reaction vessel, a 0.5 M solution of 1,3-diisopropylcarbodiimide in DMF is added which initiates the coupling reaction. Rinsing using DMF and MeOH occurs, and the next residue is ready to be added. The synthesis can also complete at this step. The internal and N-terminal amino acids are protected on their N-terminus by an Fmoc or Boc group. Each amino acid takes approximately two hours to effectively couple to the chain. This time frame is significantly less than liquid-phase peptide synthesis.

After the synthesis is complete, a cleavage solution is prepared of 92:5:3 (v/v/v) trifluoroacetic acid (TFA): water: trisopropylsilane (TIPS). The solid formed from the peptide synthesis contains the peptide attached to a solid support (Wang resin). The cleavage solution removes the peptide from the support resin. The removal of the resin undergoes stirring for 2.5 hours after which point it is filtered into cold ethyl ether (-78 °C). The solid Wang resin is left on the filter, while the peptide forms a gel in the ethyl ether. The product is allowed to cool for 2
hours and is centrifuged for 20 minutes at 3000 rpm and is decanted. Another portion of cold ethyl ether is added to the peptide gel mixture, and is cooled again for 2 hours. A final centrifugation and decantation is performed, and the gel is left to dry in the dessicator. After several hours, all that is left is a solid peptide.

The peptide synthesis creates a product that is easy to isolate by mass spectrometry. The entire process for synthesizing a peptide is completed in approximately one day. Synthesis is readily customizable and allows for the study of peptides not commercially available.

2.5.2 Peptide Sequencing Nomenclature

The peptide sequencing nomenclature used in this dissertation was first introduced by Roepstroff and Folmann. Nomenclature is important to clearly communicate types of ions produced by various bond cleavages in the peptide backbone.

The major types of ions produced by backbone cleavage include N-terminal containing fragment ions, a-, b-, and c-type ions, and C-terminal containing ions, x-, y-, and z-type ions. An example of peptide cleavage is displayed in Figure 2.7. The bond that is broken by the fragmentation process indicates which ion type is formed. When the charge is retained on the C-terminus, it is a C-terminal fragment ion, likewise, a fragment retaining charge on the N-terminal end is an N-terminal ion. In a singly charged ion, the charge will be formally retained on either one of the two cleavage products. In the negative ion mode, the common ion types formed by CID are a-, b-, c-, and y-type ions. Both x- and z-type ions are rarely observed. Each type of negative ion fragment is broken at a different bond. For example, as shown in Figure 2.6, a bond cleavage between residue 1 and 2 at the C-N bond (counting from the N-terminus) would result in an N-terminal b₁ ion and a C-terminal y₃ ion. A secondary fragmentation from b-type ions, a-
ions, are formed by the loss of C=O from b-ions while the charge is retained on the N-terminal side of the peptide. The last type of ion commonly observed in the negative ion mode, c-ions, are formed by N-Cα bond cleavage, still retaining the charge on the N-terminus.

To retain information about mobile hydrogens, an addition to the nomenclature uses primes to indicate the number of hydrogens either gained or lost relative to the structure from the cleavage site. In the negative ion mode, there are two series of b-type ions. The data discussed in Chapter 6 contains one of these series, the "b-series. The two primes located on the left side of the b-ion indicate that this fragment has lost two hydrogens. In the case of ions with additional hydrogens, the primes are on the right side of the fragment name (ex: c_3”). In the negative ion mode, the addition of hydrogens is not usually observed, as the ions are already proton deficient from the deprotonation process.

Additionally, sometimes a peptide will cleave to yield internal ions (a few are discussed in Chapter 6). An example of an internal ion from a deprotonated peptide, AASAAA, would be

Figure 2.7. Peptide sequencing nomenclature.
SA⁻. Normal backbone cleavage including the SA residues would yield N-terminal a-, b-, or c-ions with the sequence AASA, or C-terminal y-ions including SAAA residues. Internal ions are usually formed by an energetically favorable rearrangement to form a stable ion.

2.5.3 Amino Acid, Amino Acid Amide, and Phosphorylated Amino Acid and Amide Structures

The molecules used in this dissertation included amino acids, amino acid amides, phosphorylated amino acids, phosphorylated amino acid amides, and small hexapeptides. The structures for the 20 common amino acids are given below with an −NH₂ included under the C-terminal carboxylic acid to demonstrate the amide functionality. The structures of the three commonly phosphorylated amino acids are also shown.

![Amino acids/amides with aliphatic side chains](image)

Figure 2.8. Amino acids/amides with aliphatic side chains
Figure 2.9. Amino acids/amides with basic side chains

Figure 2.10. Amino acids/amides with aromatic side chains

Figure 2.11. Amino acids/amides with acidic side chains
Figure 2.12. Amino acids/amides with amide side chains

Asparagine  Glutamine
X = OH (acid) or NH$_2$ (amide)

Figure 2.13. Amino acids/amides with hydroxyl side chains (Tyrosine also has a hydroxyl sidechain but is included with aromatic amino acids in Figure 2.9)

Serine  Threonine
X = OH (acid) or NH$_2$ (amide)

Figure 2.14. Amino acids/amides with sulfur-containing side chains
Figure 2.15. Phosphorylated amino acids/amides

References


(48) Bokatzian, S. S.; Stover, M. L.; Plummer, C. E.; Dixon, D. A.; Cassady, C. J. An Experimental and Computational Investigation into the Gas-Phase Acidities of Tyrosine and


(70) Waugh, R. J.; Bowie, J. H.; Hayes, R. N. Collision-Induced Dissociations of


CHAPTER 3: A NEW GAS-PHASE DEPROTONATED TYROSINE STRUCTURE: AN EXPERIMENTAL AND COMPUTATIONAL STUDY

3.1 Overview

Using mass spectrometry and correlated molecular orbital theory, three deprotonated structures were revealed for the amino acid tyrosine. The structures were distinguished experimentally by ion/molecule reactions involving proton transfer and trimethylsilyl azide. Gas-phase acidities from proton transfer reactions and from G3(MP2) calculations generally agree well. The lowest energy structure, which was only observed experimentally using electrospray ionization from aprotic solvents, is deprotonated at the carboxylic acid group and is predicted to be highly folded. A second unfolded carboxylate structure is several kcal/mol higher in energy and primarily forms from protic solvents. Protic solvents also yield a structure deprotonated at the phenolic side chain, which experiments find to be intermediate in energy to the two carboxylate forms. G3(MP2) calculations indicate that the three structures differ in energy by only 2.5 kcal/mol, yet they are readily distinguished experimentally. Structural abundance ratios are dependent upon experimental conditions, including solvent and accumulation time of ions in a hexapole. Under some conditions, carboxylate ions may convert to phenolate ions. This agrees with G3(MP2) calculations that find the folded and unfolded carboxylate forms to differ by 0.3 kcal/mol.

3.2 Introduction

Tyrosine is one of the twenty common amino acids found in nature. As an aromatic phenol, tyrosine readily absorbs UV light, allowing tyrosine residues in biomolecules to be.
Tyrosine is involved in many cellular processes, including signal transduction, and is one of three amino acids which are commonly capable of being phosphorylated. Tyrosine possesses two acidic functionalities capable of deprotonation, the carboxylic acid group and the phenolic hydroxyl group. In the gas phase, most amino acids and peptides deprotonate at the C-terminal carboxylic acid group (if such a group is present). Exceptions include aspartic acid and glutamic acid, which have carboxylic acid groups located on the side chain and can deprotonate readily on either the C-terminus or the side chain. Exceptions include aspartic acid and glutamic acid, which have carboxylic acid groups located on the side chain and can deprotonate readily on either the C-terminus or the side chain.\(^1\) Another exception is cysteine, which has been shown to deprotonate on the side chain,\(^2,4\) with theory and experiment\(^4\) indicating that a proton is shared between the deprotonated sulfur at the side chain and the deprotonated C-terminal carboxylate group. An infrared spectroscopy study suggested that the carboxylic acid group was deprotonated in cysteine, but this still can allow for the shared hydrogen between the two anionic sites.\(^5\) In the context of a peptide backbone, the carboxylic acid functionality becomes part of the amide linkage except at the C-terminus, thus removing a likely deprotonation site. However, deprotonation of peptides lacking highly acidic sites has been observed experimentally.\(^6,7\) Thus, alternative deprotonation sites are accessible using common mass spectrometry ionization techniques such as electrospray ionization (ESI).

Tyrosine is synthesized from phenylalanine in the human body.\(^8\) Because it can be synthesized from a precursor, tyrosine is not an essential amino acid. However, tyrosine plays many roles in neurotransmission. For example, catecholamines are synthesized from tyrosine.\(^9\) Catecholamines are neuromodulators that incorporate benzene rings with two hydroxyl groups and an amine. Tyrosine can be converted to L-DOPA, a precursor for dopamine, by tyrosine hydroxylase.\(^10\) L-DOPA is further enzymatically adapted to dopamine by aromatic L-amino acid decarboxylase.\(^10\) Dopamine is one of the major neurotransmitters in mammals and is responsible
for several biological roles, such as motor function, and feelings of reward, the latter of which has been thoroughly examined in studies of addiction.\(^\text{11}\)

Tyrosine has a unique functionality as it is capable of deprotonation in two sites, the phenolic side chain and the C-terminal carboxylic acid. Of all common amino acids, aspartic and glutamic acid have acidic carboxylic acid side chains capable of deprotonation, and several theoretical and experimental studies have focused on the deprotonation of the thiol located on the side chain of cysteine. In proteomics studies, side chain deprotonation can be of increased value. When forming a peptide bond, the carboxylic acid of an amino acid is lost, unless that amino acid resides on the C-terminus. With the loss of the carboxylic acid, side chain deprotonation can allow for the formation of quasi molecular anions by electrospray ionization mass spectrometry (ESI-MS).

Kass and coworkers\(^\text{12, 13}\) have studied deprotonated tyrosine experimentally and theoretically, with a focus on determining if the deprotonated ion has a carboxylate or phenoxide structure. These researchers developed a useful ion/molecule reaction involving trimethylsilyl azide (TMSN\(_3\)) to distinguish between carboxylate and phenoxide ions.\(^\text{12, 13}\) They employed this chemical probe to study the effects of the solvent system on deprotonated tyrosine structures produced by ESI and found that phenoxide is the favored when the solvent includes methanol, while carboxylate dominates when acetonitrile or acetonitrile/water solvents are used.\(^\text{12}\) Disparate results from photoelectron spectroscopy (PES) experiments indicate that the carboxylate ion is dominant.\(^\text{13}\) Kass and coworkers attributed this inconsistency between their TMSN\(_3\) and PES experiments to different ESI source configurations. The PES experiments, which found carboxylate ions, used a home-built ESI source in which ions leaving the skimmer region pass through an ion guide for 100 ms. The ion/molecule reactions with TMSN\(_3\), which
found primarily phenoxide ions, employed a commercial ESI source where ions leaving the skimmer are accumulated in a hexapole trap for 1-5 seconds prior to their introduction into the mass analyzer. A schematic of the probe reaction with TMSN$_3$ can be viewed in Scheme 3.1. Using density functional theory (DFT) calculations (B3LYP/aug-cc-pVDZ), Kass and coworkers$^{13}$ found a carboxylate structure to be energetically favored by 0.2 kcal/mol over a phenoxide structure. This very small difference would suggest the existence of both structures in the gas phase. These researchers interpreted their results as deprotonated tyrosine undergoing a structural change as the compound goes from a solution-phase neutral to a gas-phase anion.$^{12}$

Oomens and coworkers$^5$ performed infrared multiphoton dissociation (IRMPD) on several deprotonated amino acids, including tyrosine. The IRMPD spectra were compared to calculated spectra at the DFT B3LYP/6-31++G** level of theory. The amino acids were dissolved in 80:20 (v/v) ratio of methanol (CH$_3$OH) and water (H$_2$O), which is similar to the solvent conditions where Tian and Kass$^{12}$ observed increased phenoxide production. The ESI source used by Oomens and coworkers$^5$ had a very similar commercial design to the source used by Kass and coworkers$^{13}$ for the TMSN$_3$ reactions; both sources employed hexapole accumulations times of several seconds. However, the experimental IRMPD results show that deprotonated tyrosine is a carboxylate in the gas-phase, with no evidence of phenoxide.

The Dixon group$^3$ previously predicted that tyrosine is deprotonated at the C-terminal carboxylate site using a variety of correlated molecular orbital (MO) theory methods up through CCSD(T)/aug-cc-pVTZ.$^{14,15}$ These higher level calculations show that the carboxylate ion is more stable than the phenoxide ion by $\sim$2 kcal/mol, which is consistent with the IRMPD experimental results. Using lower level DFT, Li et al.$^{16}$ have found that the experimental IRMPD
Scheme 3.1. Reactions of deprotonated tyrosine with trimethylsilyl azide.

spectra of Oomens and coworkers\textsuperscript{5} are best matched to the calculated IR spectra for the carboxylate ion.

Gas-phase acidity (GA or $\Delta G_{\text{acid}}$) values are an important tool in understanding structure, reactivity, and fragmentation behavior of compounds in mass spectrometry. The GA is the Gibbs free energy change ($\Delta G$) for the reaction: $\text{AH} \rightarrow \text{A}^- + \text{H}^+$. GAs or deprotonation enthalpies ($\Delta H_{\text{acid}}$) have been determined for the amino acids by O’Hair and coworkers\textsuperscript{17} and Poutsma and coworkers\textsuperscript{18} using the kinetic and extended kinetic methods, respectively, which involve collision-induced dissociation (CID) of a proton-bound dimer containing the analyte and a reference compound of known acidity. Poutsma and coworkers\textsuperscript{18} also calculated $\Delta H_{\text{acid}}$ for the common amino acids using a hybrid DFT approach with B3LYP functional combinations. Dixon and Cassady\textsuperscript{19} determined the GAs of glutamic acid and aspartic acid with DFT and molecular
orbital (MO) computational approaches, as well as experimentally using the thermokinetic method, which involves measurement of the rates of ion/molecule reactions between the deprotonated ion and reference compounds of known GA. In addition, Stover et al. predicted the GAs for the twenty common and five rare amino acids using the G3(MP2) method and obtained results consistent with experiment and with other calculations. Kass and coworkers found the GA of tyrosine to be 332.5 ± 1.5 kcal/mol using equilibrium GA measurements in a dual cell Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR).

In our recent study of the gas-phase and aqueous acidities of the amino acids, tyrosine was studied at eight computational levels involving both DFT and MO theory. In all cases, a gas-phase deprotonated carboxylate structure was predicted to be more stable on the free energy scale by 1.7 to 2.7 kcal/mol than a deprotonated phenoxide structure. However, as discussed above, gas-phase equilibrium proton transfer reactions and ion/molecule reactions involving TMSN₃ have found the phenoxide structure to be produced in the greatest abundance from protic solvents using ESI. This difference was the motivation for the current study. Our goal was to experimentally find the lowest energy, most acidic carboxylate form of gas-phase deprotonated tyrosine.

3.3 Experimental

3.3.1 Mass Spectrometry Methods

All experiments were performed with a Bruker Daltonics (Billerica, MA, USA) BioApex 7T FT-ICR mass spectrometer. The amino acids were in the L-stereoisomer. Samples were prepared at 60 µM in solvents containing various ratios of CH₃OH, ultrapure H₂O, acetonitrile (CH₃CN), acetone, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF), and N-dimethylformamide (DMF). All of the organic solvents were HPLC or LC-MS grade, except
DMSO, which was ACS grade. For some solutions, 1% (by volume) of ammonium hydroxide (NH₄OH) was added to promote deprotonation. For experiments in which it was important to have a low water content, the solvents were also dried with 3A pore size molecular sieves.

Analyte solutions were introduced into an Apollo API source using a syringe pump set to deliver ~90 μL/hr. Electrospray ionization (ESI) employed dried air as a heated (225°C) counter and parallel current drying gas. (The air was dried with Labclear (Oakland, CA, USA) refillable gas filter employing RK-400 molecular sieves with a 13x pore size; the sieves also contained Drierite as a color changing indicator to denote the presence of water. The sieves were dried for at least 15 hours prior to experimental use in an oven at 200 °C.) The ESI needle was grounded, while the capillary entrance and end plate were at a potential of 3.5-4.0 kV for negative ion mode analysis. Unless otherwise noted, ions produced by ESI were accumulated in a hexapole for 600-700 ms before being transported to the reaction cell by electrostatic focusing. The time used to transport ions into the cell was 10 ms.

Deprotonated ions, \([M - H]^–\), were isolated using correlated frequency ion ejection techniques.\(^{20}\) Because the ions are exposed to constant pressures of neutral reactants during this isolation period, the time involved in ion isolation was kept to a minimum and was never more than 50 ms. The isolated precursor ions were then allowed to react with a reference compound introduced to the ICR cell through a leak valve at a constant pressure. Each of the ions selected for study were reacted with a series of reference compounds of known GA.\(^{21}\) The reference compound pressures were measured using an ion gauge that was calibrated as described in Chapter 2. Pressures were corrected for reactant gas ionization efficiency,\(^{22}\) which involved polarizabilities calculated by atomic hybrid parameter procedures.\(^{23}\) Reference compound pressures were in the range of \((1 - 10) \times 10^{-8}\) mbar. For the reactions of each precursor ion with
each reference compound, at least one pressure used was from the upper half of this range and at least one pressure was from the lower half of this range; the higher and lower pressures differed by at least a factor of four (×4).

Reaction rate constants, \( k_{\text{exp}} \), were obtained by observing the pseudo-first-order decay in reactant ion intensity as a function of trapping time. Figure 3.1 displays the reaction of tyrosine anions reacting with ethyl cyanoacetate over time for the following reaction:

\[
\]

where \([M - H]^−\) indicates deprotonated tyrosine and \(A\) is the neutral reactant, in this case, ethylcyanoacetate.

In cases where deprotonation was in competition with proton-bound dimer formation, \( k_{\text{exp}} \) was obtained by fitting the experimental ion intensity data as discussed previously. For experiments in which non-linear pseudo-first-order kinetics plots (bimodal plots) indicated the presence of two ion structures reacting at two different rates, the data was fit to the sum of two exponential decays using the program Sigma Plot by Systat Software Inc. (San Jose, CA, USA). The fraction that each exponential contributes to the fit directly relates to the relative abundance of that ion structure, while the slope of each exponential decay is used to calculate the reaction rate constant (in the same manner that this information is used to calculate rate constants from a unimodal kinetics fit). This procedure has been used in the past to obtain rate constants and gas-phase basicities for systems containing two or three ion structures reacting at different rates. This procedure has been used in the past to obtain rate constants and gas-phase basicities for systems containing two or three ion structures reacting at different rates.

The ratio of the experimental rate constant to the thermal capture rate constant yields a reaction efficiency (RE). A RE of 0.269 was used as a break point, where a reaction is
considered to become exoergic and a GA value is assigned. This is described more in more detail in Chapter 2. This method provides excellent agreement between experimental and theoretical GAs for several amino acids and small peptides.\textsuperscript{1,6}

### 3.3.2 Computational Methods

The calculations were performed at the DFT and correlated MO theory levels with the program Gaussian-09.\textsuperscript{35} These calculations were performed by Dr. Michele Stover, a member of the Dixon group. The geometries were initially optimized at the DFT level with the B3LYP exchange-correlation functional\textsuperscript{36,37} and the DZVP2 basis set.\textsuperscript{38} Vibrational frequencies were calculated to show that the structures were minima. A range of structures were optimized to determine the most stable conformers chosen by sampling many conformations with and without hydrogen bonds. A substantial number of low energy conformers were found for the neutrals as discussed below. In previous work on predicting the GAs of amino acids,\textsuperscript{1,3,39} and organic acids,\textsuperscript{39} the Dixon and Cassady groups showed that the high level G3(MP2) correlated MO method\textsuperscript{40} gave agreement for the acidities with the experimental values to within about ±1 kcal/mol and also agreement with higher level CCSD(T) calculations extrapolated to the complete basis set limit with additional corrections.\textsuperscript{41-45} G3(MP2) has an additional advantage over DFT methods in terms of reliable predictions for these types of compounds because the correlated MO methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely-used DFT exchange-correlation functionals. Thermal corrections to the enthalpies and the free energies were calculated in the scaled harmonic oscillator, rigid approximation\textsuperscript{46} using the geometries and frequencies obtained in the G3(MP2) calculations (Hartree-Fock/6-31G* level).
Figure 3.1. Mass spectra of the disappearance of deprotonated tyrosine over time while reacting with neutral ethyl cyanoacetate at the pressure of $7.6 \times 10^{-8}$ mbar.
3.4 Results and Discussion

3.4.1. Calculated GAs and Structures for Tyrosine

To understand experimentally determined energetic data, computational chemistry gives useful information about structure and conformation. For deprotonated tyrosine, the Dixon group has previously predicted at the G3(MP2) level that the carboxylate anion is more stable than the phenoxide by 2.5 kcal/mol.\(^3\) The structure of the carboxylate is folded to maximize hydrogen bonding. The question exists as to whether there are additional higher energy carboxylate conformers with a similar elongated structure to that of the phenoxide. Figure 3.2 shows the optimized structures of the lowest energy conformer of tyrosine, and low energy conformers for the deprotonated tyrosine. Unless noted, all energies refer to free energies at 298 K. The multiple low energy conformers of the neutral amino acids are given in the Supporting Information, and in all cases, the lowest energy folded and unfolded structures differ by less than 0.5 kcal/mol. Table 3.1 shows the experimental and G3(MP2) calculated GAs for these molecules.

There are 22 conformers for tyrosine with energies within 2 kcal/mol of the lowest energy structure. Weak hydrogen bonding between the amine and two carbons on the phenolic ring plays an important role in the neutral, leading to a folded structure. For the tyrosine anion, 10 different structures were studied. Three low energy structures (Figure 3.2) were found that differ by only 2.5 kcal/mol, two carboxylate conformers and one phenoxide isomer, in contrast to the large number of low-lying structures in the neutral. A folded structure is predicted to be the lowest energy structure for the carboxylate anion with ring C-H’s interacting with the CO\(_2\)\(^-\) group. The carboxylate anion has a higher energy unfolded conformer 1.0 kcal/mol less stable than the folded carboxylate conformer. The lowest energy conformer for the phenoxide anion is unfolded, is 2.5 kcal/mol less stable than the lowest energy folded carboxylate conformer, and is...
1.5 kcal/mol less stable than the unfolded carboxylate conformer on the free energy scale at 298 K. In terms of $\Delta H_{298}^{\text{gas}}$, the energy difference between the unfolded anions is smaller, only 0.3 kcal/mol.

A variety of Gx methods and CCSD(T) theory have been used previously to predict the energy difference between the folded carboxylate and phenoxide sites of tyrosine. These same methods were used to calculate the relative energy values for the unfolded carboxylate anion. In all cases, the folded carboxylate anion is predicted to be 1.7 to 2.7 kcal/mol more stable than the phenoxide and the unfolded carboxylate is between the folded carboxylate and the unfolded phenoxide. Higher energy deprotonation sites include the aromatic ring, the alpha carbon, and the amine group.

### 3.4.2 Experimental GAs and Structures Produced from a Protic Solvent System

Experimental studies began with proton transfer ion/molecule reactions to bracket the GAs of tyrosine. This work was performed using a protic solvent mixture of 49.5:49.5:1 (v/v/v) CH$_3$OH:H$_2$O:NH$_4$OH, which is a routine ESI solvent system that often generates intense negative ions. Table 3.3 shows the reference compounds used in this study, their GAs, and the measured reaction efficiencies. From these results, experimental GAs were assigned. Our experimental and computational GA values are summarized in Table 3.1, along with literature values from other studies on these compounds. Initial protic solvent studies were obtained by Cassady group researcher Dr. Samantha Bokatzian.

Tyrosine is an interesting case because there are two potential sites of deprotonation. Using the CH$_3$OH:H$_2$O:NH$_4$OH solvent system, two ion populations were observed in the reactions with ethyl cyanoacetate (GA = 333.6 ± 2.0 kcal/mol$^{21}$) and 4-amino-2,3,5,6-tetrafluoropyridine (ATFP, GA = 332.8 ± 2.0 kcal/mol$^{21}$). To illustrate this bimodal reactivity,
Table 3.1. Experimental, theoretical, and literature gas-phase acidities in kcal/mol for tyrosine

<table>
<thead>
<tr>
<th>Method</th>
<th>Experimental Method</th>
<th>Deprotonation Site</th>
<th>Conformation</th>
<th>Experimental ΔG&lt;sub&gt;298&lt;/sub&gt; gas (GA)</th>
<th>ΔH&lt;sub&gt;298&lt;/sub&gt; gas</th>
<th>Computational Method</th>
<th>Theoretical ΔG&lt;sub&gt;298&lt;/sub&gt; gas</th>
<th>ΔH&lt;sub&gt;298&lt;/sub&gt; gas</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tyrosine</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present Work</td>
<td>Thermokinetic</td>
<td>carboxylate</td>
<td>Unfolded</td>
<td>333.5 ± 2.4</td>
<td>−&lt;sup&gt;a&lt;/sup&gt;</td>
<td>G3(MP2)</td>
<td>331.4</td>
<td>340.2</td>
</tr>
<tr>
<td></td>
<td>Bracketing</td>
<td>phenolate</td>
<td>Unfolded</td>
<td>332.4 ± 2.2</td>
<td>−</td>
<td>332.9&lt;sup&gt;b&lt;/sup&gt;</td>
<td>340.5&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Folded</td>
<td>324.7 ± 3.6</td>
<td>−</td>
<td>330.4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>338.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>G3B3</td>
<td>331.1</td>
<td>339.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phenolate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>332.6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>340.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Folded</td>
<td></td>
<td></td>
<td>330.5&lt;sup&gt;b&lt;/sup&gt;</td>
<td>338.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>G3</td>
<td>331.2</td>
<td>340</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phenolate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>332.9&lt;sup&gt;b&lt;/sup&gt;</td>
<td>340.5&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Folded</td>
<td></td>
<td></td>
<td>330.2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>338.1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>G4</td>
<td>331.2</td>
<td>339.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phenolate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>332.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>340.1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Folded</td>
<td></td>
<td></td>
<td>330.6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>338.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>CCSD(T)/aT//MP2/aT</td>
<td>331.3</td>
<td>339.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phenolate</td>
<td>Unfolded</td>
<td></td>
<td></td>
<td>332.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>340.1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Folded</td>
<td></td>
<td></td>
<td>330.6&lt;sup&gt;b&lt;/sup&gt;</td>
<td>338.2&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>Kass&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Equilibrium</td>
<td>carboxylate</td>
<td>Unfolded</td>
<td>332.5 ± 1.5</td>
<td>340.7 ± 1.5</td>
<td>B3LYP</td>
<td>−</td>
<td>339.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phenolate</td>
<td>Unfolded</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>338.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>carboxylate</td>
<td>Folded</td>
<td>−</td>
<td>−</td>
<td>G3B3</td>
<td>−</td>
<td>339.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phenolate</td>
<td>Unfolded</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>339.0</td>
</tr>
<tr>
<td>Poutsma&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Extended Kinetic</td>
<td>−</td>
<td>−</td>
<td>337.7 ± 2.6</td>
<td>−</td>
<td>B3LYP/6-31+G&lt;sup&gt;*&lt;/sup&gt;</td>
<td>−</td>
<td>339.2</td>
</tr>
<tr>
<td>O'Hair&lt;sup&gt;e&lt;/sup&gt;</td>
<td>Kinetic</td>
<td>−</td>
<td>−</td>
<td>329.5 ± 3.0</td>
<td>336.4 ± 3.1</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

<sup>a</sup> Kass; <sup>b</sup> Poutsma; <sup>c</sup> O'Hair
Figure 3.2. G3(MP2) results showing the structures of the most stable neutral acid and the lowest energy anions resulting from proton loss. Energies are in kcal/mol. Hydrogen bond distances in Å. The yellow arrows show the deprotonation site.

Figure 3.3 gives a semi-logarithmic plot of the intensity of deprotonated tyrosine, [Tyr – H]−, as a function of reaction time with ethyl cyanoacetate. The experimental data (black circles) are an excellent fit to an equation involving the sum of two exponentials (black line). From the reaction efficiency data of Table 3.1, the less acidic deprotonated tyrosine structure accounts for ~27% of the ions and has a GA of 333.5 ± 2.4 kcal/mol, while the more acidic structure accounts for ~73% of the ions and has a GA of 332.4 ± 2.2 kcal/mol. This is the first time that experimental GAs have been separately obtained for the different deprotonated tyrosine structures. Using an equilibrium method, Kass and coworkers reported a single GA value of tyrosine of 332.5 ± 1.5 kcal/mol,13 which is in excellent agreement with our GA value relating to the more acidic structure. Using the kinetic method of CID on a proton bound dimer, O’Hair et al.17 obtained a
value of 329.5 ± 3.0 kcal/mol, which agrees to within experimental error. O’Hair’s value is in excellent agreement with our calculated G3(MP2) value for the folded carboxylate.

The proton transfer reaction data clearly shows two structures for deprotonated tyrosine, but does not distinguish between carboxylate and phenoxide ions. To identify these structures, the ion/molecule reaction with TMSN₃ developed by Kass and coworkers.¹²,¹³ was employed. Using a series of model compounds to characterize reactivity, they found that deprotonated tyrosine with the carboxylate structure reacts with TMSN₃ to form the azide ion, N₃⁻ (m/z 42), and a neutral tyrosine trimethylsilyl ester (TyrOTMS); in contrast, the phenoxide structure of deprotonated tyrosine silylates at the phenol group, producing TMSOTyrCO₂⁻ (m/z 252) and the neutral HN₃. (A diagram of this process is located earlier in this chapter in Scheme 3.1.) In our work, for ion generation from a protic CH₃OH:H₂O:NH₄OH solvent system, the TMSN₃ reaction showed that the phenoxide structure accounts for 67(±7)% of the ions, with the remaining 33(±7)% of the ions being carboxylate.

Taken together, the proton transfer and the TMSN₃ reactivity data indicate that the more abundant and more acidic deprotonated tyrosine species (~70% of the ions) has a phenoxide structure and yields an experimental GA of 332.4 ± 2.2 kcal/mol. This corresponds to structure C of Figure 3.2, unfolded phenoxide, which involves a calculated GA of 332.9 kcal/mol. The less abundant and less acidic deprotonated tyrosine species (~30%) has a carboxylate structure and an experimental of GA of 333.5 ± 2.4 kcal/mol. This corresponds to structure B of Figure 3.2, unfolded carboxylate, which involves a calculated GA of 331.4 kcal/mol. These structures were generated from a polar solvent system of 49.5:49.5:1 (v/v/v) CH₃OH:H₂O:NH₄OH. Kass and coworkers¹²,¹³ found this same phenoxide/carboxylate ratio of 70:30 when tyrosine ions were produced from a solution of 75:25 (v/v) CH₃OH:H₂O. Thus, our results and those of Kass both
Table 3.2. Reaction efficiencies for the proton transfer reactions of tyrosine with reference compounds.

<table>
<thead>
<tr>
<th>Reference Compound</th>
<th>GA(^a) (kcal/mol)</th>
<th>Average Reaction Efficiency (± Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tyrosine (Protic Solvent(^b))</td>
<td>Tyrosine (Aprotic Solvent(^g))</td>
</tr>
<tr>
<td>Chloroform</td>
<td>349.9 ± 2</td>
<td>−(^c)</td>
</tr>
<tr>
<td>4-Trifluoromethyl aniline</td>
<td>346.0 ± 2</td>
<td>−</td>
</tr>
<tr>
<td>Phenol</td>
<td>342.3 ± 2</td>
<td>−</td>
</tr>
<tr>
<td>Acetic acid</td>
<td>341.1 ± 2</td>
<td>−</td>
</tr>
<tr>
<td>Formic acid</td>
<td>339.2 ± 2</td>
<td>−</td>
</tr>
<tr>
<td>Isovaleric acid</td>
<td>338.5 ± 2</td>
<td>−</td>
</tr>
<tr>
<td>Trimethylacetic acid</td>
<td>337.6 ± 2</td>
<td>NR(^d)</td>
</tr>
<tr>
<td>p-Chlorophenol</td>
<td>336.2 ± 2</td>
<td>NR</td>
</tr>
<tr>
<td>Ethyl cyanoacetate</td>
<td>333.6 ± 2</td>
<td>0.23 ± 0.06 (24±9%)(^e)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.03 ± 0.04 (76±9%)(^e)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.028 ± 0.019 (56 ± 7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR (44 ± 7%)</td>
</tr>
<tr>
<td>4-Amino-2,3,5,6-tetrafluoropyridine</td>
<td>332.8 ± 2</td>
<td>0.53 ± 0.25 (30±8%)(^e)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.07 ± 0.01 (70±8%)(^e)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.033 ± 0.011 (61 ± 11%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.005 ± 0.004 (39 ± 11%)</td>
</tr>
<tr>
<td>3-Trifluoromethyl phenol</td>
<td>332.4 ± 2</td>
<td>0.29 ± 0.02</td>
</tr>
<tr>
<td>3,3,3-Trifluoropropionic acid</td>
<td>326.9 ± 2</td>
<td>0.75 ± 0.09</td>
</tr>
<tr>
<td>Difluoroacetic acid</td>
<td>323.8 ± 2</td>
<td>−</td>
</tr>
<tr>
<td>Pentafluorophenol</td>
<td>320.8 ± 2</td>
<td>0.45 ± 0.10 (46 ± 13%)</td>
</tr>
</tbody>
</table>

\(^a\) All reference compound GAs were obtained from reference 17. \(^b\) Protic solvent system of 49.5: 49.5: 1 (v/v/v) CH\(_3\)OH:H\(_2\)O:NH\(_4\)OH. \(^c\) "−" indicates that no reaction was performed. \(^d\) "NR" indicates that no reaction was observed. \(^e\) Two reaction efficiencies and the relative abundances of each are listed for reactions in which bimodal kinetics plots indicate the existence of two ion populations reacting at two different rates. \(^f\) "BREAK" indicates the point where experimental GA was assigned. \(^g\) Aprotic solvent system of 99: 1 (v/v) CH\(_3\)CN: H\(_2\)O.
Figure 3.3. Reactant loss curve for the reaction of deprotonated tyrosine with ethyl cyanoacetate, which is present at a constant pressure of $8.9 \times 10^{-8}$ mbar. The solvent is 49.5:49.5:1 (v/v/v) CH$_3$OH:H$_2$O:NH$_4$OH. The logarithm of [Tyr – H]$^-$ intensity is plotted as a function of reaction time. The experimental data points (black circles) are fitted to an equation involving the sum of two exponential decays.

Our experiments found the GA involving the phenoxide structure to be 0.5 kcal/mol lower than the GA with the carboxylate structure. This differs from the G3(MP2) ordering, which predicts the unfolded carboxylate structure to yield a lower GA by 1.5 kcal/mol. There are several possible experimental reasons for this deviation. First, the proton transfer reactions involving the two deprotonated tyrosine reactant ion structures may have different activation barrier heights, which could lead to a reaction being under kinetic control as opposed to thermodynamic control. This issue is known to affect proton transfer studies of several organic molecules.\textsuperscript{49-51} In addition, the reaction intermediate is a proton bound dimer involving
deprotonated tyrosine and an acidic reference compound, \([(\text{Tyr} - \text{H})^- \cdot \text{H}^+ \cdot (\text{A} - \text{H})^-]\). This is essentially the same type of complex that is used in the kinetic method of gas-phase basicity (GB) and GA determinations, which involves collision-induced dissociation of the dimer.\(^{52, 53}\)

For several small organic acids, deviations from expected GAs or GBs using the kinetic method have been attributed to the dimer adapting a slightly higher energy structure in situations where double hydrogen bonding exists within the dimer or where the analyte conformation preferred in the dimer differs from that of its monomer form.\(^{54-56}\) For example, Fournier et al.\(^ {55}\) found that the amino acid glutamic acid has bimodal dissociation plots during a GA study using the kinetic method. Using calculations at the G3(MP2) and OLYP/aug-cc-pVTZ levels, they concluded that glutamic acid could exist within the dimer in both zwitterionic and non-zwitterionic forms. Finally, steric hindrance of the reactive site, either from bulky substituents or folded conformations with hydrogen bonding, can lessen the ability of a neutral to access an ion’s reactive site, which lowers the reaction efficiency and results in a lower GA assignment.\(^ {24, 57, 58}\)

As a check of the computational method, other approaches were used to predict the energy difference between the two carboxylate conformers and phenoxide isomer as shown in Table 2. The predicted energy differences are essentially independent of the computational method. Geometries were optimized at the MP2/aug-cc-pVTZ level with diffuse functions and higher order polarization, in contrast to the G3 geometries. Relative energies were calculated at the CCSD(T)/aug-cc-pVTZ level using optimized MP2/aug-cc-pVTZ geometries with thermal corrections, entropies and zero point energies calculated at the MP2/aug-cc-pVDZ level. The CCSD(T) results are good to at least \(\pm 1\) kcal/mol for the relative energies.
3.4.3 Effects of the Solvent on Deprotonated Tyrosine: A New Carboxylate Structure

Because the experiments with the CH$_3$OH:H$_2$O:NH$_4$OH solvent system did not result in our finding the low energy carboxylate structure predicted by theory, additional experiments were performed. Kass and coworkers$^{12,13}$ had reported that the identity of the solvent used in the ESI experiment has a dramatic effect on the phenoxide/carboxylate ratio for deprotonated tyrosine. They found that protic solvent systems containing CH$_3$OH favored phenoxide formation, while aprotic solvent systems containing CH$_3$CN favored the carboxylate structure. In addition, Schröder et al.$^{59}$ found that the phenoxide/carboxylate ratio for deprotonated p-hydroxybenzoic acid depends on the solvent, pH, and concentration of the solution being electrosprayed.

Using the TMSN$_3$ ion/molecule reaction to distinguish between phenoxide and carboxylate ions, tyrosine was electrosprayed from a variety of solvents. These included seven solvent systems used by Tian and Kass,$^{12}$ plus thirteen additional solvent systems. The results of these experiments are shown in Table 3.2. Additionally, Figure 3.4 displays a typical reaction of deprotonated tyrosine with TMSN$_3$. Exclusively (100 %) carboxylate ions were found for the entirely aprotic solvents of pure CH$_3$CN, pure acetone, and a mixture of DMF, THF, and DMSO. An addition of 1% CH$_3$OH or H$_2$O to aprotic CH$_3$CN has almost no effect, yielding 99-100% carboxylate ions. However, 1% addition of the base NH$_4$OH (by volume) results in 35% of the deprotonated tyrosine ions having a phenoxide structure. Addition of 2% CH$_3$OH or H$_2$O also has a noteworthy effect, causing about 50-70% of the ions to be phenoxide. In general, phenoxide ions appear in abundance when NH$_4$OH or the protic solvents CH$_3$OH or H$_2$O are used. Results generally agree with those of Tian and Kass$^{12}$ although they concluded that CH$_3$OH greatly affects the phenoxide/carboxylate ion ratio but H$_2$O has no effect, while we
Table 3.3. Reactions of trimethylsilyl azide with deprotonated tyrosine electrosprayed from various solvents.

<table>
<thead>
<tr>
<th>Solvent A</th>
<th>Solvent B</th>
<th>Solvent C</th>
<th>Ratios of TMSN₃ Product Ions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>%</td>
<td>Name</td>
<td>%</td>
</tr>
<tr>
<td>H₂O</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₂O</td>
<td>99</td>
<td>NH₂OH</td>
<td>1</td>
</tr>
<tr>
<td>H₂O</td>
<td>50</td>
<td>CH₃OH</td>
<td>50</td>
</tr>
<tr>
<td>H₂O</td>
<td>49.5</td>
<td>CH₃OH</td>
<td>49.5</td>
</tr>
<tr>
<td>H₂O</td>
<td>25</td>
<td>CH₃OH</td>
<td>75</td>
</tr>
<tr>
<td>H₂O</td>
<td>24.5</td>
<td>CH₃OH</td>
<td>74.5</td>
</tr>
<tr>
<td>H₂O</td>
<td>49.5</td>
<td>CH₃CN</td>
<td>49.5</td>
</tr>
<tr>
<td>H₂O</td>
<td>50</td>
<td>CH₃CN</td>
<td>50</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>49.5</td>
<td>CH₃OH</td>
<td>49.5</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>50</td>
<td>CH₃OH</td>
<td>50</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>97</td>
<td>H₂O</td>
<td>2</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>98</td>
<td>H₂O</td>
<td>1</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>98</td>
<td>H₂O</td>
<td>2</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>98</td>
<td>CH₃OH</td>
<td>1</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>99</td>
<td>NH₂OH</td>
<td>1</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>99</td>
<td>CH₃OH</td>
<td>1</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>99</td>
<td>H₂O</td>
<td>1</td>
</tr>
<tr>
<td>CH₃CN</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acetone</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMF</td>
<td>45</td>
<td>THF</td>
<td>45</td>
</tr>
</tbody>
</table>

* All solvent compositions are % by volume.

found that H₂O and CH₃OH have a similar large effect on the ratio. For example, as indicated in Table 3.3, ESI from a solution of 50% CH₃CN and 50% of either CH₃OH or H₂O both result in 63% phenoxide ions and 37% carboxylate ions. Because aprotic solvents yield almost exclusively carboxylate ions, proton transfer reactions on deprotonated tyrosine produced by ESI from a solvent of 99:1 CH₃CN:H₂O were performed. The mixture included 1% H₂O solution because the presence of a trace of water greatly increased the solubility of tyrosine; however, the water content was tightly controlled, and the LC-MS grade CH₃CN was dried with molecular sieves to insure that only the 1% water added was present. Reactions with TMSN₃ were
performed multiple times (including in close time proximity to the proton transfer reactions) and always showed that this solvent system water content was tightly controlled, forms 99 to 100% carboxylate ions and 1 to 0% phenoxide ions. The reaction efficiencies for proton transfer reactions of tyrosine anions produced from this aprotic solvent are given in Table 3.3. For all six reference compounds, two deprotonated tyrosine ion populations reacting at different rates were observed. To illustrate this, Figure 3.5 shows the semi-logarithmic plot of the intensity of [Tyr – H] as a function of reaction time with the reference compound ATFP. The plot indicates that ~71% of the ions are reacting with ATFP (albeit slowly), while the remaining ions are almost non-reactive. Using the data of Table 3.2, the faster reacting (less acidic) species was found to account for 61(±11)% of the ions and to involve a GA of 333.5 ± 2.4 kcal/mol. The slower

![Figure 3.4. Mass spectrum resulting from the ion/molecule reaction of deprotonated tyrosine electrosprayed from the protic solvent system 50:50:1 MeOH:H₂O:NH₄OH and trimethylsilyl azide at a constant pressure of 6.5 x 10⁻⁸ mbar with a reaction time of 3 seconds.](image-url)
reacting (more acidic) species includes 39(±11)% of the ions and yields a GA of 324.7 ± 3.6 kcal/mol. Because there is ~9 kcal/mol difference in acidities, these two species are readily distinguished by ion/molecule reactions.

This most acidic structure for deprotonated tyrosine, yielding an experimental GA of 324.7 ± 3.6 kcal/mol, was not produced from the protic solvent system. This ion only forms in aprotic solvent and the TMSN₃ reactions indicate that it has a carboxylate structure. Therefore, this is the lowest energy deprotonated tyrosine with a carboxylate structure that our calculations had previously predicted. The G3(MP2) calculated GA involving the folded carboxylate, structure A of Figure 3.2 is 330.4 kcal/mol. This value is 5.7 kcal/mol higher than the experimental GA, which is a greater deviation than we typically find between experiment and theory. However, the calculations indicate that this structure is highly folded with the deprotonated carboxylate group participating in two hydrogen bonds.

Due to its highly folded structure, the experimental GA value relating to this structure is likely to have more error than the other experimental GA values obtained this study because, as noted above, steric hindrance of the reactive site can yield erroneously low reaction efficiencies. Still, reactions with TMSN₃ show that carboxylate ions are produced almost exclusively from aprotic solvents, while proton transfer reactions find a carboxylate structure to be the lowest energy gas-phase conformer or isomer of deprotonated tyrosine.

The different deprotonated tyrosine ion structures generated from protic versus aprotic solvents may relate to a different tyrosine neutral structure in the two types of solvents. In aqueous solution, amino acids exist in their zwitterionic form in ~10,000 times greater abundance than their non-zwitterion form. In general, the zwitterionic forms of amino acids dominate in protic solvents, where they are stabilized by strong hydrogen bonding to
Figure 3.5. Reactant loss curve for the reaction of deprotonated tyrosine with 4-amino-2,3,5,6-tetrafluoropyridine (ATFP), which is present at a constant pressure of $5.6 \times 10^{-8}$ mbar. The solvent is 99:1 (v/v) CH$_3$CN:H$_2$O. The logarithm of [Tyr – H]$^-$ intensity is plotted as a function of reaction time. The experimental data points (black circles) are fitted to an equation involving the sum of two exponential decays.

Therefore, in our protic solvent system of CH$_3$OH:H$_2$O:NH$_4$OH neutral tyrosine is a zwitterion. Consequently, during the ESI process, the site that loses a proton to form deprotonated tyrosine is the ammonium group of the zwitterion. In contrast, for aprotic solvents there is no consistency with regard to whether a neutral amino acid exists in zwitterionic or non-zwitterionic form. The form relates to both the solvent and the amino acid, and is difficult to predict. In general, the larger the side chain of an amino acid, the less likely it is to exist as a zwitterion in aprotic solvents because steric hindrance will limit solvation. Due to the low solubility of tyrosine in most solvents, almost nothing is known about the form of its neutral in
non-aqueous solvents. However, tyrosine has one of the largest amino acid side chains, which suggests that its neutral exists as a non-zwitterion in aprotic solvents. Therefore, it is likely that in an aprotic solvent, the ESI process is deprotonating the most acidic site of the non-zwitterion of tyrosine, which is the carboxylic acid group (pKₐ = 6.15 for carboxylate deprotonation and pKₐ = 10.37 for phenolate deprotonation³).

To test the premise that the neutral form of tyrosine differs in protic and aprotic solvents (and that this might account for different gas-phase ion structures), attempts were made to study the solution-phase structure(s) of tyrosine using FT-IR, FT-Raman, and NMR. Solutions saturated with tyrosine were analyzed in solvent systems ranging from 100% aprotic to 100% protic with 0-5% acid or base additive. Unfortunately, no usable data was able to be obtained with these spectroscopic techniques because of the low solubility of tyrosine in all solvents tested (including water⁶²). Titrations at various pH values of tyrosine in several solvent systems was also performed to probe the solution-phase structure of tyrosine which was also unsuccessful. The fact that tyrosine can be studied by mass spectrometry in these solvents is testimony to the fact that mass spectrometry can analyze very dilute solutions.

The second, faster reacting tyrosine anion population generated from aprotic solvent is also interesting. Reactions with TMSN₃ indicate that the aprotic solvent produces 99% carboxylate ions, but the experimental GA from this second ion population exactly matches the experimental GA relating to the phenoxide ion found in protic solvent. This could mean that the less acidic (more elongated) carboxylate structure is converting to the phenoxide due to interactions with the references compounds. In particular, AFTP is only ~1 kcal/mol more acidic than deprotonation of tyrosine to generate the unfolded carboxylate. During collisions of AFTP neutrals with tyrosine anions a proton bound dimer intermediate, [(Tyr – H)⁻·H⁺·(AFTP – H)⁻],
could form. This dimer might dissociate with conversion of a tyrosine ion from carboxylate to phenoxide. Such a process would not yield a change in mass and, therefore, would not have been apparent in the mass spectra.

Additional support for a conversion between structures is provided by the ATFP reaction data for tyrosine ions produced from protic solvent. As seen in Table 3.2, for tyrosine anions produced from protic solvent reacting with ATFP the measured reaction efficiency is $0.53 \pm 0.25$ (mean ± standard deviation). The high standard deviation illustrates that this data is much less reproducible than all of the other reaction efficiencies that were measured. For this reaction only, the efficiency varied with pressure and was a relatively poor fit to the sum of two exponentials. This data suggested that the ion population was changing in composition as we were studying it. Thus, data for the reactions of ATFP with deprotonated tyrosine ions produced from both protic and aprotic solvents suggest that a gas-phase conversion between carboxylate and phenoxide ion structures can occur.

### 3.4.4 Deprotonated Tyrosine Structure as a Function of Hexapole Accumulation Time

The ratio of carboxylate to phenoxide ions was studied as a function of accumulation time for ions in the hexapole that immediately follows our ESI source. Previously, Kass and coworkers\textsuperscript{13} had attributed the inconsistency between their TMSN$_3$ and PES experiments to different ESI source configurations, with shorter accumulation times (100 ms) favoring a carboxylate structure and longer times (1-5 s) favoring a phenoxide structure. Our mass spectrometer has a hexapole for ion accumulation and its timescale can be easily manipulated, allowing us to study accumulation times similar to those for the two instruments used by Kass in the PES and TMSN$_3$ experiments. Our proton transfer reactions that bracketed GAs in both protic and aprotic solvent systems used a hexapole accumulation time of 0.6 to 0.7 s. These times
were selected because they generally provided the optimal signal intensity for deprotonated tyrosine. In ESI, ions are produced continually but the FT-ICR mass analyzer requires a pulsed packet of ions. To convert a continuous stream of ions into a pulse, ions leaving the ESI source are trapped in the hexapole where they remain until moved into the FT-ICR cell. During the entire accumulation time, the hexapole can accept new ions. Thus, some ions are trapped in the hexapole for almost the entire accumulation time while other more recently formed ions only spend a few milliseconds in the hexapole. Because proton transfer reactions with ethyl cyanoacetate can readily distinguish the carboxylate and phenoxide tyrosine structures generated from protic solvent systems, reactions with this reference compound were used to monitor the ion populations as a function of hexapole accumulation time. With an extremely short accumulation time of 1 ms, the faster reacting species (i.e., carboxylate) nearly disappears, resulting in 7% carboxylate ions and 93% phenoxide ions. With an accumulation time of 0.7 s (from the data of Table 3.1), there is 24% carboxylate and 76% phenoxide. For a very long accumulation time of 5 s, the ratios of the two populations remained at ~24:76.

The effects of accumulation time for ions produced from various solvents were also studied using TMSN$_3$ reactions to measure the carboxylate and phenoxide abundances. Data obtained for three solvent systems is shown in Figure 3.6. Figure 3.6(a) involves a protic solvent system of 74.5:24.5:1 (v/v/v) CH$_3$OH:H$_2$O:NH$_4$OH. At very short accumulation times the ion population is 55% phenoxide and 45% carboxylate, but this stabilizes at ~70% phenoxide and ~30% carboxylate after ~0.7 s. Similar affects were observed with other protic solvent systems. Figure 3.6(c) shows data obtained with an aprotic solvent of 99:1 (v/v) CH$_3$OH:H$_2$O. At short accumulation times, the ions are almost exclusively carboxylate; at a long time of 5 s, ~3% of the ions are phenoxide. Figure 3.6(b), involving 98:2 (v/v) CH$_3$OH:H$_2$O, is especially interesting
because within less than 1 second of accumulation time the ion ratio goes from ~55% carboxylate/45% phenoxide to the reverse. This experiment was repeated several times over a period of months, and the result was always the same.

The general trend is that as the ions spend more time in the hexapole, the relative abundance of phenoxide ions increases and carboxylate ions decreases. This is especially pronounced when the ESI solvent system contains at least 2% of a protic solvent such as H₂O or CH₃OH. A possibility is a preferential retention of phenoxide ions or a preferential loss of carboxylate ions, although it is difficult to envision a reason for either of these events to occur. Another possibility is that a conversion between structures is occurring. Presumably, this would be the less stable carboxylate ion (experimental GA of 333.5 kcal/mol) converting to the slightly more stable phenoxide ion (experimental GA of 332.4 kcal/mol), just as these ions appear to convert during the reactions with the reference compound ATFP.

Hexapole ion guides are generally known for collisional cooling of ions but a slight heating of accumulated ions can occur due to a radial stratification effect. This effect is most pronounced for ions of high charge and high mass-to-charge ratio. Deprotonated tyrosine ions are singly charged and of relatively low mass, making addition of energy during the hexapole trapping process less likely. However, very little energy would be required to convert between the deprotonated tyrosine isomers and conformers.

If a conversion is occurring, it may be facilitated by solvent interactions with ions in the hexapole. The pressure in the hexapole is in the 10⁻³ mbar range, with the primary component being dried air (the ESI drying and nebulizer gas) along with an unknown amount of solvent. The fact that the change in carboxylate/phenoxide ratio is most pronounced in the presence of H₂O supports the involvement of protic solvent (which can hydrogen bond to tyrosine) in an
isomerization process. Tian and Kass\textsuperscript{12} proposed that there was conversion between the two isomeric structures in the presence of CH\textsubscript{3}OH. They attributed this effect to a gas-phase relay mechanism\textsuperscript{67, 68} in which CH\textsubscript{3}OH simultaneously coordinates to the deprotonated carboxylic acid and neutral phenol sites, protonating the carboxylate while deprotonating the phenol.

Structures for the starting point of the proposed relay mechanism at the DFT level (B3LYP/ aug-cc-pVDZ) for the folded and unfolded carboxylate and phenoxide structures with the addition of 2, 3, or 6 waters were studied. Relative energies were calculated for $\Delta G_{\text{gas}}$ and $\Delta G_{\text{aq}}$ and are given in Table 5. The Gibbs free energy for deprotonation in aqueous solution ($\Delta G_{\text{aq}}$) was calculated from the gas-phase free energy and the aqueous solvation free energy obtained from self-consistent reaction field calculations.\textsuperscript{69} The solvation energy was calculated as the sum of the electrostatic energies (polarized solute - solvent) and the non-electrostatic energies using the COSMO parameterization\textsuperscript{70} at the B3LYP/aug-cc-pVDZ level using the gas-phase geometries obtained at this level. Solvents used include H\textsubscript{2}O, CH\textsubscript{3}OH and CH\textsubscript{3}CN with respective dielectric constants of 78.39, 32.63, and 36.64.\textsuperscript{71} In the gas phase, the folded carboxylate and unfolded phenoxide anions are ~1 kcal/mol higher in energy than the unfolded carboxylate structure. Without the addition of any explicit waters of solvation, the phenoxide anions are still higher in energy than the carboxylate anions independent of the solvent used. When using a 2 or 3 water relay mechanism, the folded carboxylate anion is more stable in the gas phase whereas the unfolded carboxylate is more stable in aqueous solution. In both cases, the phenoxide anions are higher in energy than the carboxylate anions by ~4 to 9 kcal/mol. Extra water molecules were added to better solvate the $\text{O}^-$ and $\text{COO}^-$ sites on the anions. The addition of 6 waters resulted in the unfolded and folded carboxylate structures having approximately the same energy in the gas phase whereas in aqueous solution the unfolded carboxylate was still
Figure 3.6. Percentage of product ions from the reaction of deprotonated tyrosine with trimethylsilylazide (TMSN₃) as a function of time that the ions accumulate in the hexapole. The presence of m/z 252 indicates a phenoxide deprotonated ion structure, while m/z 42 indicates a carboxylate structure. The solvent systems are: (a) 74.5:24.5:1 CH₃OH:H₂O:NH₄OH, (b) 98:2 CH₃CN:H₂O, and (c) 99:1 CH₃CN:H₂O.
more stable. The unfolded phenoxide structure is more stable with the additional waters and has approximately the same relative energy as the folded carboxylate anion, ~4 kcal/mol. The stability in solution of the unfolded structure is likely due to an effective larger dipole moment than for the folded structure. These results show that there is additional chemistry occurring leading to the experimental observations.

3.5 Conclusions

Gas-phase deprotonation of the amino acid tyrosine produces three structures: a species deprotonated at the phenolic side chain and two conformers deprotonated at the C-terminal carboxylic acid group. To the best of our knowledge, this is the first time that three distinct structures have been observed when removing a proton from the same small organic precursor molecule by mass spectrometry. The lowest energy, most stable structure is deprotonated at the carboxylic acid group, and G3(MP2) calculations indicate that the structure is highly folded with extensive hydrogen bonding. This structure was only observed experimentally during ESI from aprotic solvents, where tyrosine may exist in a non-zwitterionic form. The second lowest energy structure found experimentally, which is several kcal/mol higher in energy, is deprotonated at the phenol group and is present when the solvent system contains at least 2% of a protic compound. In protic solvent, neutral tyrosine exists as a zwitterion. The third structure is a carboxylate ion, which G3(MP2) calculations indicate to be unfolded. This carboxylate ion is only ~1.5 kcal/mol higher in energy than the phenolate ion and the two structures appear to be converting during ion/molecule reactions with some acidic reference compounds and also during ion accumulation in a hexapole. G3(MP2) calculations found that the three structures differ in energy by ~2.5 kcal/mol, yet all three are readily distinguished experimentally by proton transfer ion/molecule
reactions. Experimental conditions, such as solvent and time that ions are accumulated in a hexapole, greatly affect the abundance ratios of the deprotonated tyrosine structures.

References


(8) Cotton, R. G. H.; Gibson, F. The Biosynthesis of Phenylalanine and Tyrosine; Enzymes Converting Chorismic Acid into Prephenic Acid and their Relationships to Prephenate Dehydratase and Prephenate Dehydrogenase. Biochimica et Biophysica Acta (BBA) - General Subjects 1965, 100, 76-88.


CHAPTER 4: GAS-PHASE ACIDITIES OF THE COMMON AMINO ACID AMIDES

4.1 Overview

Using proton-transfer reactions in a Fourier transform ion cyclotron resonance mass spectrometer and correlated molecular orbital theory at the G3(MP2) level, gas-phase acidities (GAs) and the associated structures for amides corresponding to the common amino acids have been determined for the first time. These values are important because amino acid amides are models for residues in peptides and proteins. For compounds whose most acidic site is the C-terminal amide nitrogen, two ions populations were observed experimentally with GAs that differ by 4–7 kcal/mol. The lower energy, more acidic structure accounts for the majority of the ions formed by electrospray ionization. G3(MP2) calculations predict that the lowest energy anionic conformer has a cis-like orientation of the [–C(=O)NH]− group whereas the higher energy, less acidic conformer has a trans-like orientation of this group. These two distinct conformers were predicted for compounds with aliphatic, amide, basic, hydroxyl, and thioether side chains. For the most acidic amino acid amides (tyrosine, cysteine, tryptophan, histidine, aspartic acid, and glutamic acid amides) only one conformer was observed experimentally and its experimental GA correlates with the theoretical GA related to side chain deprotonation.

4.2 Introduction

Amino acid amides contain an amide group in the place of the carboxylic acid functionality located at the C-terminus of amino acids. For amino acids, the carboxylic acid is
generally the most acidic site on the molecule. However, an amino acid amide should provide a better representation of the energetic values associated with amino acid residues incorporated into a peptide chain. As shown in Figure 4.1, after the formation of a peptide bond, the carboxylic acid group is no longer available at an internal residue of a peptide. When studying deprotonated peptides or proteins, unless an amino acid has an acidic side chain, the site most readily available for deprotonation is the C-terminal carboxylic acid group. Because the majority of amino acid residues are not located at the termini of peptides, understanding the acidity of amino acid amides gives insight when evaluating potential sites of deprotonation for internal residues.

![Amino Acid and Amino Acid Amide](image)

**Locations of Peptide Bonds**

Figure 4.1. Structure of an amino acid, amino acid amide, and peptide bonds.

The acidities of amino acid residues are important to the field of proteomics. Gas-phase acidity (GA) and gas-phase basicity (GB) values can assist in explaining many attributes of gas-phase behavior, including conformation, reactivity, and fragmentation. The GA, or $\Delta G_{\text{acid}}$, of a molecule is the Gibbs free energy change ($\Delta G$) of the reaction:
Although tandem mass spectrometry (MS/MS) techniques for sequencing proteins and peptides are often conducted in the positive ion mode, the negative ion mode can provide either greater or complementary information on sequence.\textsuperscript{6-12} Acidic peptides often deprotonate more easily than they protonate,\textsuperscript{6, 13-15} and the negative mode is better suited for sequencing acidic peptides. Since peptide fragmentation pathways are usually charge-directed,\textsuperscript{4, 16} amino acid amides can provide valuable acidity and deprotonation site information to assist in understanding the fragmentation patterns in negative ion mode MS/MS techniques for peptide and protein sequencing.

Several of the twenty common amino acids deprotonate readily at their side chains in the gas phase. Glutamic acid and aspartic acid contain side chain carboxylic acid groups that deprotonate and these acidic groups are also present in their amino acid amide forms.\textsuperscript{2} The deprotonated structure of tyrosine has been studied extensively experimentally and computationally by us,\textsuperscript{17} as well as by the Kass\textsuperscript{18-20} and Oomens\textsuperscript{21} groups. Tyrosine deprotonates at both the C-terminal carboxylic acid group and the phenolic side chain. Julian and Brodbelt have also examined deprotonated tyrosine and found it to give enhanced N-C\textsubscript{\alpha} bond cleavage in peptides by negative ion electron transfer dissociation (NETD) and ultraviolet photodissociation (UVPD).\textsuperscript{22} Cysteine has been investigated theoretically and experimentally and has been found to form a thiolate ion by side chain deprotonation with a strong hydrogen bond to it from the carboxylate group.\textsuperscript{1, 21, 23} In addition, Ren and coworkers have observed cysteine side chain deprotonation in experimental and computational GA studies of small cysteine-containing peptide amides.\textsuperscript{24-27} These studies indicate that residue side chains may deprotonate in the gas phase even in the absence of a side chain carboxylic acid group. In addition, amide nitrogens along a peptide backbone can deprotonate in the absence of acidic side chains; Cassady and
Dixon\textsuperscript{3} have demonstrated this in an experimental and computational study of the GAs for glycine- and alanine-containing tripeptide methyl esters.

Additional evidence for the deprotonation of amino acid side chains is found in MS/MS experiments involving collision-induced dissociation (CID) on deprotonated peptide ions. For example, residues with hydroxyl side chains (i.e., serine, threonine, and tyrosine) display characteristic neutral losses when deprotonated peptides are subjected to negative ion mode CID.\textsuperscript{6, 10, 28-32} Proposed mechanisms for these processes involve side chain hydroxyl groups that are deprotonated. Aspartic acid\textsuperscript{10, 12} and glutamic acid\textsuperscript{33} residues induce specific peptide backbone cleavages by negative ion mode CID that suggest deprotonation at their side chains. Another example of side chain deprotonation was observed by Grzetic and Oomens via IRMPD,\textsuperscript{34} as well as Bowie and coworkers\textsuperscript{10, 35} by CID on asparagine-containing peptides. Both groups proposed a succinamide structure that deprotonates at the asparagine side chain and allows charge-directed backbone cleavage.

Although the GAs of amino acids has been the focus of several studies,\textsuperscript{1, 2, 17, 36, 37} the only previous report of GAs for amino acid amides is the work of our groups with glutamic acid amide and aspartic acid amide.\textsuperscript{2} In this prior report, our experimental and computational G3(MP2) level GA values were in excellent agreement, and the calculations confirmed that both amides undergo side chain deprotonation at their carboxylic acid groups. In the current study, we expand our work to include the eighteen other common amino acid amides whose GAs and lowest energy deprotonation sites are currently unknown.
4.3 Experimental

4.3.1 Mass Spectrometry Methods

All experiments employed a Bruker Daltonics (Billerica, MA, USA) BioApex 7T Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. The amino acids were in the L-stereoisomer. All solutions were at 60 µM in a solvent of volume ratio of 49.5:49.5:1 CH$_3$OH:H$_2$O:NH$_4$OH. The 1% ammonium hydroxide (concentrated) was added to facilitate deprotonation of the amides in the solution phase prior to electrospray ionization (ESI). For glycine amide and alanine amide, addition experiments were performed using 1% sodium hydroxide. In the case of cysteine amide, detritylation was performed,$^{37}$ as well as reduction of disulfide linkage.$^{38}$ Amino acid amide solutions were introduced into an Apollo API source using a syringe pump set to deliver ~115 µL/hr. Solutions were ionized by ESI, which utilized air as the drying gas. The ESI needle was grounded, and the capillary entrance and end plate were at a potential of 3.5-4.0 kV for optimal negative ion formation.

Deprotonated amide ions were isolated by correlated frequency resonance ejection techniques$^{39}$ and reacted with a series of neutral reference compounds with known GAs. Autoionization of the anions was not observed in the experiment or in the computations. Reference compounds were introduced through a leak value at constant pressures, which were in the range of (1.0-11) × 10$^{-8}$ mbar. Ion/molecule reaction times in the FT-ICR cell were varied from 0-300 seconds. Pressures were measured by a calibrated ion gauge,$^{17,40}$ and the pressure of each reference compound was corrected for its ionization efficiency,$^{41}$ which is determined by polarizabilities calculated with atomic hybrid parameter procedures.$^{42}$ This is described in further detail in Chapter 2.
The pseudo-first-order decay of precursor ion intensity as a function of reaction time was utilized to obtain experimental rate constants from which GA values were assigned. For experiments involving non-linear pseudo-first order kinetic behavior resulting in bimodal plots, data were fitted to the sum of two exponential decays using SigmaPlot by Systat Software Inc. (San Jose, CA, USA). The fraction that each exponential contributed to the fit directly related to the relative abundance of that ion structure. To assign GA values, the ratio of the experimental rate constant to the thermal rate capture rate constant provided a reaction efficiency (RE). This method is known as the thermokinetic method and is described in Chapter 2. A RE of 0.269 was used as a break point, where a reaction is considered to become exoergic and a GA value was assigned. The thermokinetic method of obtaining GA values has yielded excellent agreement for experimental and computational GAs for small biomolecules.

4.3.2 Computational Methods

The calculations were performed at the density functional theory (DFT) and correlated molecular orbital (MO) theory levels with the program Gaussian-09. Dr. Michele Stover, a Dixon group member performed the calculations with assistance from undergraduate researcher, John T.M. Davis. A range of structures were initially optimized at the DFT level with the B3LYP exchange-correlation functional and the DZVP2 basis set to determine the most stable conformers of the neutrals and anions by sampling many conformations with and without hydrogen bonds. Vibrational frequencies were calculated to show that the structures were minima. A substantial number of low energy conformers were found for the neutrals and anions as discussed below. Our previous predictions of the GAs of amino acids and organic acids, at the correlated G3(MP2) molecular orbital theory level are in agreement with the experimental values to within about ±1 kcal/mol and with higher level CCSD(T) calculations.
extrapolated to the complete basis set limit with additional corrections\textsuperscript{55-59} at optimized MP2/aug-cc-pVTZ geometries and with CCSD(T)/aug-cc-pVTZ calculations at the same MP2 geometries for larger molecules. G3(MP2) has an additional advantage over DFT methods in terms of reliable predictions for these types of compounds because of the important role that hydrogen bonding plays in controlling the lowest energy structures; the correlated MO methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely-used DFT exchange-correlation functionals. Thermal corrections to the enthalpies and the free energies were calculated in the harmonic oscillator, rigid rotor approximation\textsuperscript{60} using the geometries and scaled frequencies obtained at the Hartree-Fock/6-31G* level of theory from the G3(MP2) calculations.

4.4 Results and Discussion

The experimental and G3(MP2) GA values obtained in this study are summarized in Table 4.1. For purposes of comparison and completeness, Table 4.1 also includes our previously published GAs for aspartic acid amide and glutamic acid amide. (Note that a lower GA numerical value means a more acidic compound.) There is excellent agreement between the experimental and theoretical GAs. The values agree to within experiment error except for arginine amide, where the lowest energy calculated GA is only 0.3 kcal/mol below the experimental error range. As discussed below, Table 4.1 provides the G3(MP2) GA for the formation of the most stable anion (most acidic). In addition, G3(MP2) GAs are provided for higher energy conformers and/or isomers that have a slightly higher energy GA (less acidic). For some amino acid amides, the experimental data also support the presence of two deprotonated ion conformers and the assignment of two GAs. The reaction efficiencies for each amino acid amide reacting with the appropriate references compounds (from which the experimental GAs
Table 4.1. Experimental and computational values for the GAs of the amino acid amides

| Amino Acid Amide | Lowest Energy Calc. GA (kcal/mol)
|------------------|-----------------------------------
| Alanine          | 351.1                             |
| Arginine         | 339.4                             |
| Asparagine       | 340.2                             |
| Aspartic acid    | 325.9                             |
| Cysteine         | 332.9                             |
| Glutamine        | 340.2                             |
| Glutamic acid    | 326.4                             |
| Glycine          | 351.5                             |
| Histidine        | 331.3                             |
| Isoleucine       | 348.3                             |
| Leucine          | 349.5                             |
| Lysine           | 347.5                             |
| Methionine       | 345.2                             |
| Phenyl-alanine   | 347.8                             |
| Proline          | 350.8                             |
| Serine           | 342.1                             |
| Threonine        | 341.7                             |
| Tryptophan       | 336.1                             |
| Tyrosine         | 336.5                             |
| Valine           | 349.3                             |

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>355.2</td>
<td>e</td>
<td>358.8</td>
<td>362.5</td>
</tr>
<tr>
<td>Arginine</td>
<td>347.3</td>
<td>344.6</td>
<td>347.4± 3.2</td>
<td>352.2</td>
</tr>
<tr>
<td>Asparagine</td>
<td>346.1</td>
<td>346.5± 2.7</td>
<td>348.0</td>
<td>353.5</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>331.6</td>
<td>332.9</td>
<td>332.9</td>
<td>338.2</td>
</tr>
<tr>
<td>Cysteine</td>
<td>342.6</td>
<td>n</td>
<td>340.1</td>
<td>350.1</td>
</tr>
<tr>
<td>Glutamine</td>
<td>347.4± 3.2</td>
<td>n</td>
<td>343.0</td>
<td>339.1</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>332.0</td>
<td>n</td>
<td>334.0</td>
<td>339.1</td>
</tr>
<tr>
<td>Glycine</td>
<td>359.7</td>
<td>359.7</td>
<td>357.2</td>
<td>357.2</td>
</tr>
<tr>
<td>Histidine</td>
<td>333.7</td>
<td>333.7± 3.3</td>
<td>333.7± 3.3</td>
<td>342.2</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>352.1</td>
<td>352.1± 3.2</td>
<td>355.1</td>
<td>359.2</td>
</tr>
<tr>
<td>Leucine</td>
<td>352.7</td>
<td>352.7± 3.2</td>
<td>357.3</td>
<td>361.1</td>
</tr>
<tr>
<td>Lysine</td>
<td>352.5</td>
<td>352.5± 3.2</td>
<td>355.8</td>
<td>358.8</td>
</tr>
<tr>
<td>Methionine</td>
<td>349.6</td>
<td>347.7± 2.5</td>
<td>352.4</td>
<td>357.2</td>
</tr>
<tr>
<td>Phenyl-alanine</td>
<td>350.1</td>
<td>350.1± 3.0</td>
<td>354.6</td>
<td>360.5</td>
</tr>
<tr>
<td>Proline</td>
<td>355.4</td>
<td>357.7± 4.0</td>
<td>357.8</td>
<td>362.2</td>
</tr>
<tr>
<td>Serine</td>
<td>347.8</td>
<td>347.6± 3.2</td>
<td>349.3</td>
<td>355.0</td>
</tr>
<tr>
<td>Threonine</td>
<td>347.4</td>
<td>347.4± 3.2</td>
<td>349.2</td>
<td>354.7</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>348.2</td>
<td>n</td>
<td>336.1</td>
<td>355.5</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>344.2</td>
<td>n</td>
<td>342.5</td>
<td>354.7</td>
</tr>
<tr>
<td>Valine</td>
<td>350.2</td>
<td>355.0± 4.0</td>
<td>356.4</td>
<td>361.3</td>
</tr>
</tbody>
</table>

a Unless otherwise noted, the deprotonated ion has mc-cis structure. mc-cis = main chain deprotonation with cis-like orientation. Previous work by our group. Reference 2. sc structure. sc = side chain. A higher energy folded sc structure was also predicted: ΔH=344.3; ΔG=336.9. No deprotonated ion signal was obtained experimentally and therefore no experimental GA could be assigned. Unless otherwise noted, the deprotonated ion has mc-trans structure. mc-trans = main chain deprotonation with trans-like orientation.

8 A lower energy sc structure was also predicted: ΔH=348.0; ΔG=340.8. mc-cis/sc shared structure. A higher energy mc-trans/sc shared structure was also predicted: ΔH=342.5; ΔG=336.1. mc-cis/sc shared structure. A mc-cis/sc shared structure was predicted to have similar energies: ΔH=349.7; ΔG=342.9. mc-cis/sc shared structure. A higher energy mc-trans/sc shared structure was also predicted: ΔH=341.8; ΔG=335.0. A higher energy mc-trans structure was also predicted: ΔH=349.7; ΔG=343.1. A higher energy mc-trans structure was also predicted: ΔH=360.4; ΔG=353.0. Only one ion population was observed experimentally and only one experimental GA could be assigned.
were assigned) are given in the appropriate section for the classification of amino acid amides later in this chapter.

4.4.1 Amino Acid Amides with Aliphatic Side Chains

The least acidic amino acid amides have aliphatic side chains that are non-polar and contain only carbon and hydrogen atoms. The aliphatic amino acid amides are those derived from glycine, alanine, proline, valine, leucine, isoleucine, and phenylalanine. For these amino acid amides, only C-terminal amide nitrogen deprotonation is possible because there is no heteroatom on the side chain. The glycine and alanine amides could not be studied experimentally because they did not produce an adequate deprotonated ion signal by ESI. Attempts to improve the anionic signal by addition of sodium hydroxide to the solution undergoing ESI were unsuccessful. Apparently, the ESI environment was insufficient to deprotonate glycine amide and alanine amide, which were the smallest and among the least acidic compounds involved in this study (G3(MP2) GAs of 351.5 kcal/mol and 351.1 kcal/mol, respectively).

The least acidic amino acid amide that could be studied experimentally was proline amide, which has a cyclic aliphatic side chain. Proline amide yielded experimental data showing two ion populations reacting at two different rates. Both experiment and theory provided two GA values that differed by 5-7 kcal/mol. The lowest energy experimental GA of 350.2 ± 2.5 kcal/mol has a G3(MP2) counterpart of 350.8 kcal/mol, and the higher energy experimental GA of 357.7 ± 4.0 kcal/mol correlates with a G3(MP2) value of 355.4 kcal/mol. Rate data for all aliphatic amino acid amides is located in Table 4.2. Figure 4.2(a) provides an example of the bimodal experimental data from the reaction of deprotonated proline amide with the reference compound 1,2,4,5-tetrafluorobenzene, whose GA of 353.3 kcal/mol is between the two GA values for
proline amide, thus allowing the two ion populations to be distinguished. Figure 4.2(b) shows experimental data for 3-methylpyrazole (GA = 348.3 kcal/mol\textsuperscript{61}), which is a more acidic reference compound than either proline amide population. In Figure 4.2(b), both proline amide ion populations react readily and they cannot be distinguished experimentally. For proline amide, the majority of the ions (65 ± 10 %) are involved in the more acidic and lower energy process.

Structures for the most stable neutral and two anions for proline amide are shown in Figure 4.3. The remaining structures for aliphatic amino acid amides are given in Figure 4.4. Important hydrogen bond distances are shown in Å and the two anion structures are designated with capital letters, with [A] being the most acidic structure.

The neutral structure for proline is stabilized by a O···H(N) hydrogen bond within the –C(=O)NH\textsubscript{2} group (2.53 Å) and by a hydrogen bond between the –C(=O)NH\textsubscript{2} and the –NH– in the pyrrole ring (2.17 Å). The calculations predict that there are two C-terminal deprotonated conformers with different orientations of hydrogen bonds. In the lowest energy cis-like anion, the loss of the hydrogen from the –C(=O)NH\textsubscript{2} terminal group results in a shorter internal O···H(N) hydrogen bond (2.43 Å) in the [–C(=O)NH]\textsuperscript{–} group. In addition, the NH on the ring reorients to form a bond with the [–C(=O)NH]\textsuperscript{–} group, N(H)···(N)CO (2.07 Å). In the higher energy trans-like anion, the rotation of the H away from the O results in only one strong hydrogen bond between the pyrrole NH and the [–C(=O)NH]\textsuperscript{–} group, N(H)···(O)CN (1.95 Å).
Figure 4.2. Reactant ion loss curves for deprotonated proline amide reacting with (a) 1,2,4,5-tetrafluorobenzene (GA = 353.3 kcal/mol) at a constant pressure of $7.8 \times 10^{-8}$ mbar and (b) 3-methylpyrazole (GA = 348.3 kcal/mol) at a constant pressure of $9.0 \times 10^{-8}$ mbar. The logarithmic $[\text{M-H}]^-$ intensity is plotted versus reaction time. In (a) the experimental data points (black circles) are fit to an equation involving the sum of two exponentials, while (b) involves a single exponential fit. Hydrogen bond distances in the structures of (a) are given in Å.
Table 4.2. Reaction efficiencies for the proton transfer reactions of proline, valine, leucine, isoleucine, lysine, phenylalanine, methionine, and arginine amides with neutral reference compounds.

<table>
<thead>
<tr>
<th>Reference Compound</th>
<th>GA (kcal/mol)</th>
<th>Proline</th>
<th>Valine</th>
<th>Leucine</th>
<th>Isoleucine</th>
<th>Lysine</th>
<th>Phenylalanine</th>
<th>Methionine</th>
<th>Arginine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclohexanone</td>
<td>358.7 ± 2</td>
<td>0.25 ± 0.02&lt;br&gt;(31 ± 14%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.19 ± 0.02&lt;br&gt;(8 ± 1%)&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NR&lt;sup&gt;d&lt;/sup&gt;</td>
<td>NR</td>
<td>92 ± 5%&lt;sup&gt;d&lt;/sup&gt;</td>
<td>69 ± 9%&lt;sup&gt;d&lt;/sup&gt;</td>
<td>31 ± 14%&lt;sup&gt;d&lt;/sup&gt;</td>
<td>8 ± 1%&lt;sup&gt;d&lt;/sup&gt;</td>
<td>92 ± 5%&lt;sup&gt;d&lt;/sup&gt;</td>
<td>69 ± 9%&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Acetophenone</td>
<td>354.5 ± 2</td>
<td>0.77 ± 0.01&lt;br&gt;(40 ± 7%)</td>
<td>0.34 ± 0.10&lt;br&gt;(25 ± 13%)</td>
<td>0.0001 ± 0.0003&lt;br&gt;(60 ± 7%)</td>
<td>0.0006 ± 0.0006&lt;br&gt;(75 ± 13%)</td>
<td>0.0001 ± 0.0003&lt;br&gt;(60 ± 7%)</td>
<td>0.0006 ± 0.0006&lt;br&gt;(75 ± 13%)</td>
<td>0.0001 ± 0.0003&lt;br&gt;(60 ± 7%)</td>
<td>0.0006 ± 0.0006&lt;br&gt;(75 ± 13%)</td>
</tr>
<tr>
<td>1,2,4,5-Tetrafluoro-benzene</td>
<td>353.3 ± 2</td>
<td>0.814 ± 0.14&lt;br&gt;(34 ± 27%)</td>
<td>0.52 ± 0.08&lt;br&gt;(43 ± 26%)</td>
<td>0.26 ± 0.06&lt;br&gt;(27 ± 16%)</td>
<td>0.19 ± 0.10&lt;br&gt;(25 ± 21%)</td>
<td>0.14 ± 0.08&lt;br&gt;(24 ± 26%)</td>
<td>0.0014 ± 0.0005&lt;br&gt;(73 ± 16%)</td>
<td>0.0013 ± 0.0007&lt;br&gt;(75 ± 21%)</td>
<td>0.0022 ± 0.0020&lt;br&gt;(76 ± 26%)</td>
</tr>
<tr>
<td>Pyrrole</td>
<td>350.9 ± 2</td>
<td>0.088 ± 0.016&lt;br&gt;(26 ±10%)</td>
<td>0.089 ± 0.15&lt;br&gt;(73 ± 10%)</td>
<td>0.009 ± 0.001&lt;br&gt;(80 ± 6%)</td>
<td>0.006 ± 0.001&lt;br&gt;(75 ± 19%)</td>
<td>0.009 ± 0.002&lt;br&gt;(86 ± 11%)</td>
<td>0.0003 ± 0.0002&lt;br&gt;(63 ± 7%)</td>
<td>0.0007 ± 0.0004&lt;br&gt;(73 ± 5%)</td>
<td>-&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Chloroform</td>
<td>349.9 ± 2</td>
<td>0.37 ± 0.13&lt;br&gt;(26 ±10%)</td>
<td>0.36 ± 0.12&lt;br&gt;(73 ± 10%)</td>
<td>0.38 ± 0.09&lt;br&gt;(26 ±10%)</td>
<td>0.38 ± 0.13&lt;br&gt;(43 ± 12%)</td>
<td>0.39 ± 0.14&lt;br&gt;(48 ± 8 %)</td>
<td>0.29 ± 0.06&lt;br&gt;(44 ±25%)</td>
<td>0.005 ± 0.013&lt;br&gt;(66 ±25%)</td>
<td>0.04 ± 0.02&lt;br&gt;(66 ±12%)</td>
</tr>
</tbody>
</table>

Table 4.2., continued.
<table>
<thead>
<tr>
<th>Compound</th>
<th>M/L</th>
<th>M/L</th>
<th>M/L</th>
<th>M/L</th>
<th>M/L</th>
<th>M/L</th>
<th>M/L</th>
<th>M/L</th>
<th>M/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Methylpyrazole</td>
<td>348.3 ± 2</td>
<td>0.41 ± 0.16</td>
<td>0.39 ± 0.13</td>
<td>0.64 ± 0.07 (50 ± 7%)</td>
<td>0.68 ± 0.13 (52 ± 2%)</td>
<td>0.69 ± 0.20 (52 ± 12%)</td>
<td>0.48 ± 0.07 (32 ± 9%)</td>
<td>0.19 ± 0.01 (26 ± 12%)</td>
<td>0.13 ± 0.06 (33 ± 17%)</td>
</tr>
<tr>
<td>1-Butanethiol</td>
<td>347.4 ± 2</td>
<td>0.58 ± 0.20</td>
<td>0.55 ± 0.18</td>
<td>0.32 ± 0.20</td>
<td>0.43 ± 0.12</td>
<td>0.28 ± 0.02</td>
<td>0.09 ± 0.03</td>
<td>0.36 ± 0.08 (35 ± 12 %)</td>
<td>0.21 ± 0.11 (65 ± 12%)</td>
</tr>
<tr>
<td>4-Trifluoromethyl aniline</td>
<td>346.0 ± 2</td>
<td>−</td>
<td>−</td>
<td>0.49 ± 0.23</td>
<td>0.54 ± 0.06</td>
<td>0.36 ± 0.05</td>
<td>0.21 ± 0.06</td>
<td>0.22 ± 0.15 (64 ± 8%)</td>
<td>0.11 ± 0.08 (60 ± 6%)</td>
</tr>
<tr>
<td>Indene</td>
<td>344.6 ± 2</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0.38 ± 0.11</td>
<td>0.28 ± 0.03</td>
</tr>
<tr>
<td>Phenol</td>
<td>342.3 ± 2</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0.51 ± 0.14</td>
<td>0.68 ± 0.30</td>
</tr>
<tr>
<td>Acetic Acid</td>
<td>341.1 ± 2</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>0.74 ± 0.21</td>
<td>0.77 ± 0.23</td>
</tr>
</tbody>
</table>

\(^a\) All reference compound GAs were obtained from reference \(^1\). \(^b\) Two reaction efficiency and the relative abundance of each ion population are listed; these reactions exhibited bimodal kinetics indicating two ion populations reacting at different rates. \(^c\) "−" indicates no reaction was performed. \(^d\) "NR" indicates that no reaction, or a reaction with an efficiency <<< 0 was obtained. \(^e\) "BREAK" indicates point where experimental GA value was assigned.
Figure 4.3. The most stable neutral and two anions for proline amide at the G3(MP2) level. Important hydrogen bond distances are given in Å. mc-cis = main chain deprotonation with cislike orientation and mc-trans = main chain deprotonation with trans-like orientation (Note this orientation is of the N-H bond in the [–C(=O)NH]– group). All energetic values are in kcal/mol.

The other aliphatic side chain amides (valine, leucine, isoleucine, and phenylalanine amides) also yielded kinetic data indicating two ion populations reacting at two different rates. The two GA values for each amide differ by 4-6 kcal/mol. For the valine, leucine, and isoleucine amides, the lower energy (more acidic) GA is 348-350 kcal/mol and the higher energy GA is 352-355 kcal/mol. The two GAs for the aromatic phenylalanine amide are comparable to those for the non-aromatic aliphatic amides.

Both experiment and theory show that as the size of the aliphatic side chain increases, the amide becomes slightly more acidic. The lowest energy (more acidic) process was always observed experimentally in the greatest abundance, generally accounting for ~65 % of the ion population produced by ESI. The lowest energy anionic conformer for the aliphatic side chains and phenylalanine amide have a cis-like orientation of the [–C(=O)NH]– group with a O···H(N) hydrogen bond with distances of ~2.4 Å. The higher energy conformer is generated by rotating...
Figure 4.4. Continued on following pages.
Figure 4.4. Continued on following page.
Figure 4.4. The most stable neutral and two anions for the aliphatic amino acid amides at the G3(MP2) level.

The H away from the O and breaking this hydrogen bond. This rotation results in a trans-like orientation of the [–C(=O)NH]– group and the GAs of the higher energy conformer are 4-6 kcal/mol higher (less acidic) than the cis-like orientation of the lowest energy conformer.

4.4.2 Amino Acid Amides with Amide, Basic, Hydroxyl, and Thioester Side Chains

The amino acid amides discussed in this section have side chains that contain heteroatoms capable of deprotonation, but only deprotonation of the main chain C-terminal amide group was predicted and observed. The experimental and computational results for these amides are comparable to the results for the aliphatic amides. Again, two ion populations were found to react at two rates, providing GA values that differ by 5-6 kcal/mol, with the lowest energy most acidic population accounting for ~65% of the ions formed by ESI. The amino acid amides in this category have side chains that are basic (lysine), hydroxyl (serine and threonine), thioester (methionine), and amide (asparagine and glutamine). Arginine is discussed separately at
the end of this section. The reaction rates for the amino acid amides in this subclass are located in Table 4.3, with the exception of lysine and methionine amides that are less acidic and in Table 4.2.

For these amino acid amides, the calculations predict the presence of both the cis-like (more hydrogen bonded, lower energy) and trans-like (less hydrogen bonded, higher energy) structures. An example of this behavior experimentally can be found in Figure 4.5. This figure displays kinetic rate data for glutamine amide reacting with isovaleric acid. Similar to proline amide (Figure 4.2), when reacted with less acidic neutrals, two ion populations are clearly observed. Because these compounds contain side chain heteroatoms that are capable of hydrogen bonding, they can fold into more stabilized structures. Consequently, these amino acid amides are more acidic than the amides with aliphatic side chains. As seen in Table 4.1, for these amino acid amides, the GA relating to the lowest energy anion is in the range of 340-348 kcal/mol, while the GA for the higher energy process ranges from 346-352 kcal/mol.

The most stable neutrals and two differing energy anions are shown in Figure 4.6 for lysine, methionine, threonine, serine, glutamine, and asparagine amides. The presence of an accessible oxygen or nitrogen on the side chain that is available for hydrogen bonding increases the acidity of the amides by ~8 kcal/mol. Serine, threonine, asparagine, and glutamine were determined to have nearly equal GA values, varying only in a 1.9 kcal/mol range computationally. Asparagine and glutamine amides were found to be slightly more acidic than the hydroxyl-containing serine and threonine amides. The –C(=O)NH₂ group on the side chains of asparagine and glutamine amides allow for a more compact, folded structure that results from 4 to 5 hydrogen bonds in the neutrals and anions. Serine and threonine amides have relatively
Table 4.3. Reaction efficiencies for the proton transfer reactions of arginine, threonine, serine, glutamine, and asparagine amides with neutral reference compounds.

<table>
<thead>
<tr>
<th>Reference Compound</th>
<th>GA$^a$ (kcal/mol)</th>
<th>Average Reaction Efficiency (± Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Arginine</td>
</tr>
<tr>
<td>3-Methyl pyrazole</td>
<td>348.3 ± 2</td>
<td>0.13 ± 0.06&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(33 ± 17%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(77 ± 17%)</td>
</tr>
<tr>
<td>1-Butanethiol</td>
<td>347.4 ± 2</td>
<td>0.27 ± 0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(20 ± 6%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.08 ± 0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(80 ± 6%)</td>
</tr>
<tr>
<td>4-Trifluoromethyl aniline</td>
<td>346.0 ± 2</td>
<td>0.44 ± 0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(40 ± 6%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.11 ± 0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(60 ± 6%)</td>
</tr>
<tr>
<td>Indene</td>
<td>344.6 ± 2</td>
<td>0.20 ± 0.15</td>
</tr>
<tr>
<td>Phenol</td>
<td>342.3 ± 2</td>
<td>0.29 ± 0.03</td>
</tr>
<tr>
<td>Acetic Acid</td>
<td>341.2 ± 2</td>
<td>0.45 ± 0.14</td>
</tr>
<tr>
<td>Formic Acid</td>
<td>339.1 ± 2</td>
<td>0.69 ± 0.10</td>
</tr>
<tr>
<td>Isovaleric Acid</td>
<td>338.5 ± 2</td>
<td>−&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Trimethylacetic Acid</td>
<td>337.6 ± 2</td>
<td>−</td>
</tr>
</tbody>
</table>

$^a$ All reference compound GAs were obtained from reference 1. $^b$Two reaction efficiency and the relative abundance of each ion population are listed; these reactions exhibited bimodal kinetics indicating two ion populations reacting at different rates. $^c$ “BREAK” indicates point where experimental GA value was assigned. $^d$ “−” indicates no reaction was performed.
three. For methionine amide, which contains a thioester functionality, the side chain length is similar in size to the aliphatic amides leucine and isoleucine. However, the neutral methionine amide folds more readily upon itself, increasing the acidity. Lysine amide contains a longer aliphatic chain (four methylenes) with an amino $\text{–NH}_2$ group, while methionine has a shorter aliphatic chain (two methylenes) with an $\text{–SH}$ group. This combination leads to a GA for lysine amide that is 2.3 kcal/mol less acidic than methionine amide. For lysine amide, the lowest energy neutral and anion structures contain very similar hydrogen bonding networks with three hydrogen bonds between the side chain $\text{–NH}_2$, the N-terminal $\text{–NH}_2$, and the C-terminal amide group.
Figure 4.6. Continued on following pages.
Figure 4.6. Continued on following page.
Figure 4.6. The most stable neutral and two anions for lysine, serine, glutamine, and asparagine amides at the G3(MP2) level. See Figure 4.3 caption for additional details.

For arginine amide, even though the two low energy conformers of the neutral differ by the arrangements of the hydrogens on the guanidine functional group, they are essentially isoenergetic. The two stable neutral conformers and three anions of arginine amide are shown in Figure 4. The lowest energy structure for the anion is the result of deprotonation from the C-terminal amide group as a cis-like conformer with a predicted GA of 339.4 kcal/mol. Deprotonation from the side chain guanidine functional group is only 1.4 kcal/mol higher and both values are within the bounds of the lowest energy experimental GA of 340.7 ± 3.0 kcal/mol. Thus, both of these low energy predicted structures are within the experimental error bar. The higher energy experimental GA is 347.4 ± 3.2 kcal/mol and is predicted by theory to correspond to a trans-like amide deprotonated structure. The structures associated with arginine amide neutral and anions are below in Figure 4.7.
4.4.3 Amino Acids with Acidic Side Chains

Six amino acid amides have side chains that could be sufficiently acidic to deprotonate in the gas phase. These side chains have carboxylic acid groups (aspartic and glutamic acids), a phenolic group (tyrosine), a thiol group (cysteine), and cyclic nitrogen-containing groups (histidine and tryptophan). Proton transfer reactions involving these six amino acid amides gave linear pseudo-first order kinetic behavior (Figure 4.8), indicating the presence of one predominant deprotonated ion structure. The experiments also leave open the possibility of two or more structures with near identical GAs.
The most acidic species are glutamic acid amide and aspartic acid amide, with calculated GAs of 326.4 kcal/mol and 325.9 kcal/mol, respectively, which were previously studied experimentally and computationally and found to undergo exclusively side chain deprotonation. The kinetic rate data for all acidic amino acids studied in this work (tyrosine, cysteine, histidine, and tryptophan amides) is located in Table 4.4.

The data clearly show that deprotonation of the –OH group in tyrosine amide and of the –SH group in cysteine amide occurs preferentially over deprotonation of the C-terminal amide functionality. This is consistent with the fact that deprotonation of the –OH group in tyrosine amino acid is slightly less favored that deprotonation of the C-terminal –CO$_2$H and that deprotonation of –SH occurs in the amino acid cysteine. $^{1,17-22}$ Tyrosine amide deprotonation has an experimental GA of 336.4 ± 2.7 kcal/mol, which is in excellent agreement with the calculated
Table 4.4. Reaction efficiencies for the proton transfer reactions of tryptophan, cysteine, tyrosine and histidine amides with neutral reference compounds.

<table>
<thead>
<tr>
<th>Reference Compound</th>
<th>GA$^a$ (kcal/mol)</th>
<th>Average Reaction Efficiency (± Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Acetic Acid</td>
<td>341.2 ± 2</td>
<td>−$^b$</td>
</tr>
<tr>
<td>Formic Acid</td>
<td>339.1 ± 2</td>
<td>0.007 ± 0.004</td>
</tr>
<tr>
<td>Isovaleric Acid</td>
<td>338.5 ± 2</td>
<td>0.08 ± 0.03</td>
</tr>
<tr>
<td>Trimethylacetic Acid</td>
<td>337.6 ± 2</td>
<td>0.15 ± 0.07</td>
</tr>
<tr>
<td>Ethyl cyanoacetate</td>
<td>333.6 ± 2</td>
<td>0.42 ± 0.05</td>
</tr>
<tr>
<td>4-amino-2,3,5,6-</td>
<td></td>
<td>Break</td>
</tr>
<tr>
<td>tetrafluoropyridine</td>
<td>332.8 ± 2</td>
<td>0.71 ± 0.11</td>
</tr>
<tr>
<td>3-Trifluormethyl</td>
<td>332.4 ± 2</td>
<td>−</td>
</tr>
<tr>
<td>phenol</td>
<td></td>
<td>Break</td>
</tr>
<tr>
<td>3,3,3- Trifluoropropionic acid</td>
<td>326.9 ± 2</td>
<td>−</td>
</tr>
</tbody>
</table>

$^a$ All reference compound GAs were obtained from reference $^1$. $^b$−$^c$ indicates no reaction was performed. $^c$ “BREAK” indicates point where experimental GA value was assigned.

GA of 336.5 kcal/mol involving an ion deprotonated at the side chain. G3(MP2) calculations found that deprotonation of tyrosine amide at the C-terminal amide group results in a GA of 347.4 kcal/mol which is ~11 kcal/mol higher in energy. This higher energy deprotonated ion was not observed experimentally. For side chain deprotonation of cysteine amide, the experimental GA is 335.6 ± 3.8 kcal/mol and the G3(MP2) GA is 332.9 kcal/mol. The calculations predicted that C-terminal amide deprotonation results in a GA of 342.6 kcal/mol, which is ~10 kcal/mol more energetic than side chain deprotonation and again was not observed experimentally. As a
consequence of their acidic side chains, these compounds are several kcal/mol more acidic than the amino acid amides discussed earlier.

The most stable neutral conformers and the two anion isomers of the amino acid amides with acidic side chains (tyrosine, cysteine, tryptophan, and histidine amides) are shown in Figure 4.9, listed in order of increasing acidity. (This figure does not include the structures for the amides of aspartic and glutamic acids, published previously.\textsuperscript{2}) For tyrosine amide, the most stable neutral as well as the higher energy anions have extensively folded and hydrogen bonded structures which allow the amide and NH\textsubscript{2} groups (where the NH has a positive charge) to interact with the π cloud on the benzene ring. These cation-π type interactions are well established in the literature for compounds containing benzene rings.\textsuperscript{62} The lowest energy phenoxide anion prefers an unfolded structure with no interactions with the benzene π cloud. However, a folded structure for the phenoxide anion where a NH proton interacts with the benzene π cloud was found to only be 0.4 kcal/mol higher in energy. For cysteine amide, both the lowest energy neutral and anion are folded structures, with the neutral containing four hydrogen bonds while the lowest energy anion only contains three. As noted previously, this ability to make a compact, folded structure may help to stabilize the negative charge as it does in cysteine amide, but clearly is not important in tyrosine amide.

Tryptophan and histidine amides display one experimental deprotonated ion population, suggesting side chain deprotonation. The most stable neutral of tryptophan amide is extensively folded with multiple hydrogen bonds between the ring and the C-terminal amide and N-terminal –NH\textsubscript{2} groups. The lowest energy structure for tryptophan amide is predicted to be deprotonated at the nitrogen of the indole side chain and is 12.1 kcal/mol more stable than the structure involving C-terminal amide deprotonation. The lowest energy anion maintains the extensive
Figure 4.9. Continued on following page.
Figure 4.9. The most stable neutrals and anions for the amino acid amides with acidic side chains at the G3(MP2) level. mc-cis/sc = main chain deprotonation with cis-like orientation and proton sharing between the main and side chains and mc-trans = main chain deprotonation with trans-like orientation and proton sharing between the main and side chains. See Figure 4.2 caption for additional details.
hydrogen bonding network of the neutral. Deprotonation of the C-terminal amide group leads to a higher energy structure with far less hydrogen bonding.

Histidine amide is an interesting case because the side chain is basic and protonates readily in positive ion mode mass spectrometry.\textsuperscript{63-66} In the amino acid, the $\tau$ tautomer is more stable than the $\pi$ tautomer with $\Delta G(298 \text{ K}) = 1 \text{ kcal/mol.}\textsuperscript{1}$ The Dixon group predicts the $\tau$ tautomer neutral amide to be more stable than the $\pi$ tautomer neutral amide by 2.5 kcal/mol. Side chain deprotonation of histidine amide has an experimental GA is $330.8 \pm 4.8 \text{ kcal/mol}$. For both of the histidine amide tautomers $\pi$ and $\tau$, deprotonation of the $\text{–NH–}$ in the imidazole ring generated the same lowest energy anion with three hydrogen bonds. The GA with respect to the less stable $\pi$ tautomer neutral is 328.8 kcal/mol and the GA with respect to the more stable $\tau$ tautomer neutral is 331.3 kcal/mol. Both of the calculated GA values are within the experimental error bar. The higher energy anion resulting from deprotonation of the C-terminal amide in the $\pi$ tautomer leads to a GA that is 4.4 kcal/mol less acidic than the lowest energy anion. The anion derived by deprotonation of the C-terminal amide in the $\tau$ tautomer is 20 kcal/mol higher in energy than the most stable anion. These results differ from those of deprotonation of the amino acids tryptophan and histidine which deprotonate at the carboxylic acid group.\textsuperscript{1,21,35}

4.5 Conclusions

GA values have been assigned for 18 of the 20 common amino acid amides in this study. Previous work completed has experimentally and computationally assigned GA values for the remaining two amides. Computations determined a majority of the amides (glycine, alanine, proline, valine, lysine, leucine, isoleucine, phenylalanine, methionine, arginine, serine, threonine, asparagine, and glutamine) deprotonate at the C-terminal amide group. Proton-transfer reactions for the amides calculated to deprotonate at the amide backbone displayed two distinct ion
populations, with the exception of glycine and alanine which did not deprotonate readily enough to undergo ion/molecule reactions (glycine and alanine did have GA values assigned computationally). Computations agreed and predicted two structures: a higher energy, less acidic conformer (trans-like), and a lower energy, more acidic conformer (cis-like) varying in energy by ~5 kcal/mol. A smaller portion of the amides (tyrosine, cysteine, phenylalanine, tryptophan, and previously studied aspartic and glutamic acid amides) displayed one distinguishable ion population experimentally. One experimental and computational GA value was assigned for these amides. Computational results indicate the lowest energy, and most acidic structure occurs when deprotonation occurs on the side chain. The closest structure in energy for amides displaying only one ion population is on average ~7.5 kcal/mol higher in energy and was not observed experimentally. This study indicates that six amino acid residues incorporated into a peptide chain can readily undergo deprotonation and should be considered as potential acidic sites when evaluating biomolecules.

References


CHAPTER 5: GAS-PHASE ACIDITIES OF THE PHOSPHORYLATED AMINO ACIDS
AND THEIR CORRESPONDING AMIDES

5.1 Overview

Gas-phase acidities and heats of formation have been predicted at the G3(MP2)/SCRF-COSMO level of theory for ten phosphorylated amino acids and their corresponding amides, including phospho-serine (pSer), -threonine (pThr), and -tyrosine (pTyr), providing the first reliable set of these values. The gas-phase acidities (GAs) of the three named phosphorylated amino acids and their amides have been determined using proton transfer reactions in a Fourier transform ion cyclotron mass spectrometer. Excellent agreement was found between the experimental and predicted GAs. The phosphate group is the deprotonation site for pSer and pThr and deprotonation from the carboxylic acid generated the lowest energy anion for pTyr. The infrared spectra were calculated for six low energy anions of pSer, pThr, and pTyr. For deprotonated pSer and pThr, good agreement is found between the experimental IRMPD spectra and the calculated spectra for our lowest energy anion structure. For pTyr, the IR spectra for a higher energy phosphate deprotonated structure is in good agreement with experiment. Additional experiments tested electrospray ionization (ESI) conditions for pTyr and determined that variations in solvent, temperature, and voltage can result in a different experimental GA value, indicating that ESI conditions affect the conformation of the pTyr anion.
5.2 Introduction

Post-translational modifications (PTMs) allow for the function of a protein to be altered through covalent modification, e.g., the transformation of a side chain functionality into another one.\textsuperscript{1,2} Phosphorylation is a common PTM and plays an important role in various cellular processes such as cell signaling.\textsuperscript{3} Phosphorylation involves the addition of a phosphate group on the side chain of an amino acid residue, which adds a site for proton loss and increases its capacity for hydrogen bonding.\textsuperscript{4} Because phosphorylation occurs after the protein has been encoded, the process of phosphorylation enables the cell to meet changes in its environment. In eukaryotes, phosphorylation most commonly occurs at the hydroxyl side chains of serine, threonine, and tyrosine. Phosphoserine (pSer) accounts for \textasciitilde 90\% of the phosphorylated residues in the human proteome, with phosphothreonine (pThr), and phosphotyrosine (pTyr) being estimated at 10\% and 0.05\% respectively.\textsuperscript{5} Additional amino acids including cysteine, histidine, arginine, lysine, aspartic acid, glutamic acid, and glycine have been shown to undergo phosphorylation.\textsuperscript{6} Though there are generally thought to be 20 common amino acids, the phosphorylated amino acids are abundant in nature. The PhosphoNET database has compiled information on over 180,000 human phosphorylation sites and have found that 80\% of proteins encoded by human genes undergo phosphorylation.\textsuperscript{7}

Phosphorylation is a reversible process and can have a very short lifetime, which makes it a challenging process to analyze. Additionally, the multitude of potential sites for phosphorylation and the potential for isoforms to exist in the proteome further complicate the analysis. Mass spectrometry is an important method in the analysis of phosphoproteomes because of its ability to identify and sequence phosphorylated peptides and determine the sites of phosphorylation.\textsuperscript{8}
The use of soft ionization techniques such as electrospray ionization (ESI)\(^9\) and matrix-assisted laser desorption ionization (MALDI)\(^{10}\), allow for the study of proton transfer reactions. Studying the gas-phase acidities (GAs), \(\Delta G\) of \(AH \rightarrow A^- + H^+\), of phosphopeptides provides detailed information on structural conformations and energetics. Experimental conditions, especially those related to ESI, can have an effect on gas-phase conformations, such as the degree of folding, as well as sites available for protonation or deprotonation.\(^{11-24}\) Because phosphopeptides contain a very acidic phosphate group, negative ion mode mass spectrometry could provide the more suitable environment for studying their behavior in the gas phase.\(^{25}\)

The most commonly used tandem mass spectrometry (MS/MS) techniques for identifying phosphopeptides are collision-induced dissociation (CID), electron capture dissociation (ECD), and electron transfer dissociation (ETD).\(^8\) Phosphopeptides create many problems for detection and identification with MS because their strong acidities can lead to low positive ion mode ionization efficiencies by ESI and MALDI.\(^{26-28}\) Additionally, sequence information can be lost as a result of the cleavage or migration of the phosphate group during ion excitation in MS/MS.\(^{29-36}\) A range of studies have been performed using negative ion mode CID to sequence phosphopeptides and have shown that the phosphorylated residues can be identified and characterized based on the types of neutral loss the peptides undergo and the negative ions formed.\(^{25, 26, 30, 37-50}\) Electron-based techniques do not depend on energy, so do not preferentially undergo the low energy cleavage of a phosphate group. This allows for identification of phosphorylated residues more readily. Despite the interest in phosphorylated peptides and amino acids, there has not been much work done to study the gas-phase thermochemical properties of individually phosphorylated amino acids. Ohanessian and coworkers\(^{51}\) used infrared multiple photon dissociation (IRMPD)
spectroscopy together with density functional theory (DFT) at the B3LYP/6-31+G* level to determine the vibrational signatures of protonated [pSer + H]⁺, [pThr + H]⁺, and [pTyr+H]⁺. They suggest that the vibrational modes of phosphate groups between 900 – 1300 cm⁻¹ can be used to identify the site of phosphorylation in phosphopeptides, as the P−O−H and P=O bands directly relate to the amino acid to which the phosphate group is bound. Subsequently, Ohanessian and coworkers⁵² applied these same methods to report the IRMPD spectra of a protonated phosphodipeptide. Because the phosphate group is expected to be quite acidic and deprotonated in the gas phase, Maitre and coworkers¹² used IRMPD to study the IR spectra of deprotonated [pSer − H]⁻, [pThr − H]⁻, and [pTyr − H]⁻. They first measured the IRMPD spectrum of H₂PO₄⁻ to obtain the characteristic frequencies of the deprotonated phosphate bending and stretching modes. They then used these frequencies together with calculated vibrational spectra at the DFT/6-31+G* and DFT/6-311+G** levels of theory to determine that the deprotonation site for all three phosphorylated amino acid anions was the phosphate group. They predicted that phosphoserine and phosphothreonine have a folded structure, whereas phosphotyrosine has an extended structure. Gaigeot and coworkers⁵³,⁵⁴ performed DFT-based Car-Parrinello molecular dynamics (CPMD) simulations at room temperature on deprotonated [pSer – H]⁻ and protonated [pSer + H]⁺ to predict the effects of temperature and to obtain anharmonic frequency corrections. By comparing the spectra of [pSer – H]⁻ and [pSer + H]⁺, they predicted the main vibrational modes that can be used to determine the protonation state of the phosphate, carboxylate, and amino groups. They suggest that temperatures effects will increase as the size of the phosphorylated peptide increases. Additional IRMPD studies⁵⁵ have been performed on monohydrated [pTyr•H₂O + H]⁺ together with calculations at the B3LYP-D/SVP, M06/6-31G(d,p), and MP2/SVP levels of theory to investigate the effect of adding water
to the phosphorylated amino acids on the vibrational spectra. These studies showed that the added water molecule creates a hydrogen bonding bridge between the phosphate and ammonium groups resulting in significant changes to the characteristic frequencies. In 1996, Spiccia and coworkers\textsuperscript{56} performed potentiometric investigations to determine the pK\textsubscript{a} of pSer and found a value of 2.19. In 2010, Smiechowski\textsuperscript{57} performed DFT with the SMD solvation model calculations to predict the pK\textsubscript{a} of pSer to be 1.6 which is in good agreement with experiment.

Building on the previous work on amino acids and peptides,\textsuperscript{15, 58, 59} by the Cassady and Dixon groups, we have extended our studies to the phosphorylated amino acids. The GAs of the common phosphorylated amino acids and their corresponding amides have been predicted at the G3(MP2) level of theory, and have been measured experimentally by the thermokinetic method. Amino acid amides are a better representation of an amino acid residue incorporated into a peptide chain (see Chapter 4 for a detailed explanation). In addition to our work, there are a number of other studies of the gas-phase acidities of the amino acids.\textsuperscript{13, 60} Because of the importance of pSer, pThr, and pTyr in the human proteome and the potential for the formation of a gas-phase zwitterion (a salt bridge formed by an ammonium group between two anionic centers), additional calculations at the MP2 level with augmented correlation-consistent basis sets\textsuperscript{61} up through the triple-ζ were performed to further study the hydrogen bonding present in the anionic structures. Similar to the work presented in Chapter 3, phosphotyrosine proved to have its acidity dependent on the ESI solvent and experimental conditions. When electrospraying pTyr from two different solvent systems, the GA value is changed. This clearly displays that sample preparation and electrospray conditions play a major role in the type of ions formed via ESI.
5.3. Experimental

5.3.1 Mass Spectrometry Methods

All experiments were performed using a Bruker Daltonics (Billerica, MA, USA) BioApex 7T Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. All amino acids and their amides were in the L-stereoisomer. All solutions were at 60 µM in a solvent of volume ratio of 49.5:49.5:1 methanol:water:ammonium hydroxide (CH₃OH:H₂O:NH₄OH). The 1% ammonium hydroxide assisted with solution-phase deprotonation prior to electrospray ionization (ESI). Analyte solutions were introduced into an Apollo ESI source using a syringe pump set to deliver ~125 µL/hr. Solutions were ionized by ESI, with a drying gas temperature of 220 °C. The ESI needle was grounded, and the capillary entrance and end plate were at a potential of 3.5-4.0 kV for optimal negative ion formation.

In the case of phosphotyrosine, additional ESI conditions were examined based upon the work of Maitre and coworkers.¹² Phosphotyrosine solutions were prepared at a concentration of 1 mM in a solvent by volume of 60:20:20 acetonitrile:isopropanol:water (ACN:IPA:H₂O) with an additive of ammonium acetate present at a 1.0 µM concentration. The ESI drying gas temperature was 150 °C and the voltage at the capillary entrance was 4.0 kV in order to optimize phosphotyrosine negative ion formation.

Deprotonated phosphorylated ions were isolated by correlated frequency resonance ejection techniques⁶² and reacted with neutral reference compounds with well-established GAs.⁶³ Reference compounds were introduced through a leak value at constant pressures, which were in the range of (1.0-17) × 10⁻⁸ mbar. Pressures were measured by a calibrated ion gauge (as described in Chapter 2) and the pressure of each reference compound was corrected for its
ionization efficiency,\textsuperscript{64} which is determined by polarizabilities calculated with atomic hybrid parameter procedures.\textsuperscript{65}

The pseudo-first-order decay of precursor ion intensity as a function of reaction time was utilized to obtain experimental rate constants from which GA values were assigned. To assign GA values, the ratio of the experimental rate constant to the thermal rate capture rate constant provided a reaction efficiency (RE). A RE of 0.269 was used as a break point, where a reaction is considered to become exoergic and a GA value can be assigned. The thermokinetic method is described in detail in Chapter 2. In past Cassady and Dixon collaborations, the thermokinetic method of obtaining GA values has yielded excellent agreement for experimental and computational GAs for amino acids and small peptides.\textsuperscript{15, 58, 66, 67} Multiple examples of this agreement are provided in Chapters 3 and 4 of this dissertation.

5.3.2. Computational Methods

Calculations were performed at the density functional theory (DFT) and correlated molecular orbital (MO) theory levels with the program Gaussian09.\textsuperscript{68} Dr. Michele Stover and Sean Miller completed the computational portion of this work. Geometries were initially optimized at the DFT level with the B3LYP exchange-correlation functional\textsuperscript{69, 70} and the DZVP2 basis set.\textsuperscript{71} Vibrational frequencies were calculated to show that the structures were minima and to provide zero point and thermal corrections to the enthalpy and entropies so that free energies could be calculated for direct comparison to experiment. Phosphorylated L-amino acids, their corresponding amides, and the corresponding anions were studied at a range of deprotonation sites. Extensive conformational sampling was performed using density functional theory (DFT) to search the conformational space which is complicated by the flexibility of the amino acid and the OPO$_3$H$_2$/OPO$_3$H$^-$ groups. Dixon group previous predictions of the GAs of amino acids and
peptides \(^{15,58}\) at the correlated G3(MP2) molecular orbital theory level \(^{72}\) are generally in agreement with the experimental values to within about ±1 kcal/mol and with higher level CCSD(T) calculations extrapolated to the complete basis set limit with additional corrections. \(^{73-77}\) G3(MP2) has an advantage over DFT methods with commonly used functionals in terms of reliable predictions for these types of compounds as the correlated molecular orbital methods in G3(MP2) perform better in the prediction of hydrogen bond energies as well as steric non-bonded interactions than do most widely used DFT exchange-correlation functionals.

### 5.4 Results and Discussion

The lowest energy G3(MP2) and experimental GAs are given in Table 5.1 for phosphorylated amino acids, as well as their phosphorylated amides. The previous work done by Dixon and Cassady on the GAs for the L-common amino acids \(^{59}\) and their amides (Chapter 4) has been included for comparison. Excellent agreement is found between the G3(MP2) and the available experimental GAs and will be discussed in further detail below. For phospho-serine, -threonine, and -tyrosine, phosphorylation occurs at the hydroxyl group on the side chain. The amino acids serine and threonine have very similar structures with threonine only differing from serine by the presence of a methyl group on the side chain. Thus, the results of phosphoserine and phosphothreonine will be discussed together.

#### 5.4.1. Phosphoserine (pSer) and Phosphothreonine (pThr): IR spectra, GAs, and hydrogen bonding

Phosphoserine and phosphothreonine have hydroxyl side chains that have been modified to a phosphate ester. When subjected to ESI, the phosphorylated amino acids readily form [M – H]\(^-\) deprotonated ions easily isolatable for ion/molecule reactions. All kinetic data obtained by ion/molecule reactions yielded pseudo-first-order kinetic plot which indicates the presence of one major ion population (or multiple ion populations at
Table 5.1. Lowest energy gas-phase acidities ($\Delta G$) and corresponding enthalpies at 298 K at the G3(MP2) level in kcal/mol.

<table>
<thead>
<tr>
<th>Compound</th>
<th>$\Delta G$ expt</th>
<th>$\Delta H$ calc</th>
<th>$\Delta G$ calc</th>
<th>$\Delta H$ calc</th>
<th>$\Delta G$ calc</th>
<th>$\Delta H$ calc</th>
<th>$\Delta G$ expt</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-Amino Acid $^{57}$</td>
<td>Phosphorylated Amino Acid $^{59}$</td>
<td>Amino Acid Amide $^{83}$</td>
<td>Phosphorylated Amino Acid Amide $^{13}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>serine</td>
<td>325.8±3.0 $^{59}$</td>
<td>332.6</td>
<td>325.7</td>
<td>315.7</td>
<td>309.2</td>
<td>308.5 ± 3.4</td>
<td></td>
</tr>
<tr>
<td>threonine</td>
<td>325.2±3.0 $^{59}$</td>
<td>332.3</td>
<td>324.8</td>
<td>314.2</td>
<td>308.3</td>
<td>308.2 ± 3.4</td>
<td></td>
</tr>
</tbody>
</table>
tyrosine| 329.5±3.0 $^{59}$ | 338.3 | 330.4 | 314.5 | 306.7 | 307.0 ± 4.0 |
arginine| 323.2±2.3 $^{59}$ | 330.1 | 322.0 | 317.8 | 309.1 | |

essentially the same acidity). Figure 5.1 is an example of the kinetic data obtained by ion/molecule reactions. Figure 5.1 displays the reaction of pThr and the neutral 1,1,1,5,5,5-hexafluoro-2,4-pentadione (GA = 310.4 kcal/mol). It is clear from the straight line obtained, that there are not ion populations varying by several kcal/mol in GA. Because the phosphorylated amino acids/amides are increasingly more acidic than the compounds studied in Chapters 3 and 4, locating neutral molecules of established and necessary GA values was a harder task. When acidity increases, often neutral compounds are less volatile, or corrosive, leading to their inability to be used in the FT-ICR leak valves. These properties of highly acidic neutral molecules are the reason behind the larger standard deviations of the experimental GA values (this is the case for all of the phosphorylated compounds studied).
For pSer and pThr, two low energy neutral conformers with basically the same structure were predicted. The lowest energy neutral conformer, pSer-1 (pThr-1), has one strong hydrogen bond (~ 1.8 to 1.9 Å) between the $\text{PO}_3\text{H}_2$ and the $\text{NH}_2$ groups and a weaker hydrogen bond (~ 2.5 to 2.6 Å) between the $\text{NH}_2$ and the $\text{CO}_2\text{H}$ groups. In the higher energy neutral conformer, pSer-2 (pThr-2), the $\text{CO}_2\text{H}$ group forms strong hydrogen bonds with both the $\text{PO}_3\text{H}_2$ and $\text{NH}_2$ groups (~ 1.8 and 1.9 Å respectively). A weaker hydrogen bond (2.4 Å) is formed between the $\text{NH}_2$ and the $\text{PO}_3\text{H}_2$ groups. Deprotonation of pSer and pThr generated six low energy anion conformers/isomers with the sites of deprotonation being the phosphate and/or carboxylate groups. The experimental GAs of pSer and pThr are $308.5 \pm 3.4$ and $308.2 \pm 3.4$ kcal/mol, respectively. For pSer, the G3(MP2) GAs of anion structures pSer-A, pSer-B, pSer-C, and pSer-D all fall within the experimental error, whereas for pThr, only the G3(MP2) calculated GAs of anion structures pThr-A and pThr-B are in agreement with experiment. Structures of the neutral and anionic pSer and pThr are located in Figure 5.2.

Figure 5.1. Reaction ion loss plot of [pThr – H]$^-$ with neutral 1,1,5,5,5-hexa-2,4-pentadione (GA = 310.4 kcal/mol). The presence of one reaction rate strongly suggests one major ion structure.
Figure 5.2. The most stable neutral and two anions for pSer and pThr at the G3(MP2) level. Important hydrogen bond distances are given in Å.
Ion/molecule reaction efficiency data used to obtain the experimental GA values is given in Table 5.2. (Note that a lower numerical GA value indicates a more acidic compound.) In all cases, pseudo-first-order kinetics plots were linear, suggesting the presence of one dominant phosphorylated anion structure (or possibly multiple structures with very similar GAs). Due to the highly acidic nature of these phosphorylated species, few reference compounds were suitable for thermokinetic bracketing experiments. This inability to study a larger number of reference compounds can lead to a relatively wide range in uncertainty when assigning experimental GA values. For example, thermokinetic bracketing indicated that pThr is slightly more acidic than pSer by \(~0.3 \pm 3.4\) kcal/mol. To confirm this acidity trend, mixture analysis was performed in a manner similar to Cassady group past experiments to elucidate GA trends of glutamic and aspartic acid and of their corresponding amides.\(^{58}\) A mixture of equimolar concentrations of pSer and pThr was electrosprayed and both [M– H\(^-\)] were reacted simultaneously with 1,1,1,5,5,5-hexafluoro-2,4-pentadione. pSer anion reacted to 100% completion faster than pThr anion, clearly indicating that pSer is less acidic than pThr and, thus, confirming the acidity trend of the reaction efficiency data. Similar experiments were conducted to confirm that pThr amide is less acidic than pSer amide.

From the calculations, the most stable anion structure (structures are located in Figure 5.2), pSer-A and pThr-A, has partial proton transfer between the phosphate and carboxylate groups with the proton closer to the carboxylate, so this is labelled as ‘phosphate’ deprotonation. The higher energy anion structure, pSer-B and pThr-B, is a more folded structure than pSer-A and pThr-A with partial proton transfer again for ‘phosphate’ deprotonation. Anion structure pSer-C and pThr-C is an unfolded ‘phosphate’ deprotonated structure. As anion structures pSer-A (pThr-A), pSer-B (pThr-B), and pSer-C (pThr-C) are all generated from phosphate
deprotonation, we predict that the anion observed experimentally (formed by ESI) is
deprotonated at the phosphate group. Higher energy anion structures pSer-D (pThr-D) and pSer-
E (pThr-E) are carboxylate deprotonated structures with pSer-D (pThr-D) exhibiting partial
proton transfer between the \(-\text{CO}_2^-\) and \(-\text{PO}_3\text{H}_2\). The highest energy anion structure pSer-F
(pThr-F) is zwitterionic (salt-bridge like) with an \(-\text{NH}_3^+\) group sandwiched between the \(-\text{PO}_3\text{H}^-\)
and \(-\text{CO}_2^-\) groups and held together by strong hydrogen bonds. A detailed discussion of the
hydrogen bonding in these structures is given below.

Due to the unusual occurrence of a small gas-phase zwitterionic anion, structural
optimization of the lowest energy neutral and six anions for pSer and pThr at the MP2/aug-cc-
pVnZ level of theory with \(n = D\) and \(T\) was performed. The energy differences between the six
anion sites are shown in Table 5.3 for pSer. Regardless of the computational method used,
structure pSer-B (pThr-B) is predicted to be higher in energy than structure pSer-A (pThr-A) on
the free energy scale, although this can be reversed for the enthalpy at the MP2 level. In any case
these energy differences at the MP2 level are small, on the order of 0.5 kcal/mol. Structure pThr-
A is predicted to be lower in energy than structure pThr-B in terms of free energy and enthalpy
by 0.8 and 1.3 kcal/mol respectively at the MP2/aT level. Structure pSer-C is predicted to be 1.1
cal/mol higher in energy than structure pSer-A at the MP2/aT level and structure pThr-C is
predicted to be 3.9 kcal/mol higher in energy than structure pThr-A. Thus, only structures pSer-
A (pThr-A), pSer-B (pThr-B), and pSer-C are relevant in terms of their energies. All three anion
structures pSer-A (pThr-A), pSer-B (pThr-B), and pSer-C are ‘phosphate’ deprotonated. The
most interesting result was the decrease in the energy difference between the higher energy
Table 5.2. Reaction efficiencies for the proton transfer reactions of phosphorylated amino acids and their amides.

<table>
<thead>
<tr>
<th>Reference Compound</th>
<th>GA(^{a}) (kcal/mol)</th>
<th>Average Reaction Efficiency (± Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>pThreonine Amide</td>
</tr>
<tr>
<td>Difluoroacetic Acid</td>
<td>323.8 ± 2</td>
<td>NR(^{d})</td>
</tr>
<tr>
<td>Pentafluorophenol</td>
<td>320.8 ± 2</td>
<td>0.07 ± 0.03</td>
</tr>
<tr>
<td>Trifluoroacetic Acid</td>
<td>317.4 ± 2</td>
<td>0.21 ± 0.07</td>
</tr>
<tr>
<td>Heptafluorobutric Acid</td>
<td>314.9 ± 2</td>
<td>0.28 ± 0.11</td>
</tr>
<tr>
<td>1,1,1,5,5,5,-Hexafluoro-2,4, pentadione</td>
<td>310.3 ± 2</td>
<td>0.53 ± 0.15</td>
</tr>
<tr>
<td>Bistrifluoroacetamide</td>
<td>307.5 ± 2</td>
<td>–</td>
</tr>
<tr>
<td>Propanedinitrile</td>
<td>303.6 ± 2</td>
<td>–</td>
</tr>
</tbody>
</table>

\(^{a}\) All reference compound GAs were obtained from reference 63. \(^{b}\) pTyrosine in solvent system of 60:20:20 acetonitrile:isopropanol:water (ACN:IPA:H\(_{2}\)O) with 1 µM NH\(_{4}\)OAc. \(^{c}\) pTyrosine in a solvent system of 49.5:49.5:1 MeOH:H\(_{2}\)O:NH\(_{4}\)OH. \(^{d}\) “NR” indicates that no reaction, or a reaction with an efficiency < 0.001 was obtained. \(^{e}\) “−” indicates no reaction was performed. \(^{f}\) “BREAK” indicates point where experimental GA value was assigned.
zwitterionic anion structure pSer-F (pThr-F) and structure pSer-A (pThr-A), from 4.3 to 2.5 kcal/mol for pSer and from 5.3 to 3.6 kcal/mol for pThr at the MP2/aT levels of theory. The relative energy differences at the MP2/aD only differ from the energy differences already given at the MP2/aT levels of theory by 0.1 kcal/mol. Thus, the use of the lower levels of geometry optimization in the G3(MP2) composite calculations is not significantly biasing the predictions.

Maitre and coworkers\textsuperscript{12} compared their IRMPD spectrum of deprotonated [pSer – H]\textsuperscript{−} and [pThr – H]\textsuperscript{−} with the calculated spectra of their lowest energy PO\textsubscript{3}H\textsuperscript{−}, COO\textsuperscript{−}, and zwitterionic anions. Their three structures correspond to our anion structures pSer-B (pThr-B), pSer-E (pThr-E), and pSer-F (pThr-F), respectively. They did not report the deprotonated phosphate anion structures pSer-A (pThr-A) and pSer-C (pThr-C).

The Dixon group calculated the IR spectra for the six [pSer – H]\textsuperscript{−} and [pThr – H]\textsuperscript{−} anion structures pSer-A (pThr-A) to pSer-F (pThr-F) at the B3LYP/aug-cc-pvdz level. These calculated spectra as well as the experimental IRMPD spectrum for deprotonated [pSer – H]\textsuperscript{−} and [pThr – H]\textsuperscript{−} obtained by Maitre and coworkers are given in Figures 1 and 2 respectively. Maitre and coworkers suggested that the experimental IRMPD spectrum corresponds to their lowest energy PO\textsubscript{3}H\textsuperscript{−} anion, our structure pSer-B (pThr-B). However, the calculated IR spectra for anion structures pSer-A (pThr-A) and pSer-C (pThr-C) also show good agreement with the IRMPD spectrum. Additionally, it is noted that the calculated IR spectra for the zwitterionic anion structure pSer-F (pThr-F) shows good agreement. Because the relative energy differences between anion structures pSer-F and pSer-A is only ~2.5 kcal/mol at the MP2/aT level of theory, the anionic zwitterion structure pSer-F could be present at about 1%. The Dixon calculations agree with Maitre and coworkers that the experimental spectra of deprotonated [pSer – H]\textsuperscript{−} and
[pThr – H]− are composed of mostly phosphate deprotonated anions. However, our calculations show that these PO₃H− anions can be present as several different low energy conformers.

At equilibrium in the gas phase at 298 K, the ratio for structures pSer-A, pSer-B, and pSer-C relative to pSer-A is 1.0:0.36:0.16 at the MP2/aT level whereas the ratio for structures pThr-A, pThr-B, and pThr-C relative to pThr-A is 1.0:0.11:0.01. Thus, the experimental IRMPD spectra of deprotonated [pSer – H]− and [pThr – H]− are probably composed of a mixture of PO₃H− anion conformers. To further study the formation and stabilization of the anionic zwitterion structure pSer-F (pThr-F), the relevant hydrogen bonding was examined at the different computation levels and the bond lengths are given for the six pSer anion structures in Figure 5.2. The atom labels for the anions are given in Figures 5.3 and 5.4 for the pSer and pThr anions, respectively. For the six pThr anion structures, the hydrogen bonding can be seen in Figure 5.2. Because the structures of the pThr anions are basically the same as the pSer anions, the hydrogen bonding results follow the same trends and have approximately the same bond distances. Therefore, only the pSer hydrogen bond distances are discussed in detail. In general, the results show that there is not much change in key bond lengths at the B3LYP, MP2/6-31G(d) from G3(MP2), MP2/aD, and MP2/aT levels of theory. For the pSer anion structures pSer-A, pSer-B, pSer-D, and pSer-E, two hydrogens are bound to the nitrogen (H1 and H18) and both have a distance of 1.01 to 1.03 Å, respectively. A zwitterionic anion structure cannot be formed in any of these cases as the third hydrogen needed to form the zwitterion is shared between the −CO₂− and the −PO₃H− groups. Anion structures pSer-A and pSer-D are similar, as are anions pSer-B and pSer-E with the major difference being the rotation of the phosphate group to allow for hydrogen bonding with the −NH₂ group in pSer-B and pSer-E.
Table 5.3. Calculated acidities in kcal/mol of phosphoserine at different computational levels with respect to the most stable neutral conformer.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Prop</th>
<th>B3LYPa</th>
<th>G3(MP2)</th>
<th>MP2/aDb</th>
<th>MP2/aTc</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSer PO₃H⁺ [A]</td>
<td>ΔH</td>
<td>313.8</td>
<td>315.7</td>
<td>312.6</td>
<td>313.6</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>307.1</td>
<td>309.2</td>
<td>306.0</td>
<td>307.1</td>
</tr>
<tr>
<td>pSer PO₃H⁺ [B]</td>
<td>ΔH</td>
<td>315.6</td>
<td>316.0</td>
<td>312.1</td>
<td>313.2</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>310.6</td>
<td>310.5</td>
<td>306.5</td>
<td>307.7</td>
</tr>
<tr>
<td>pSer PO₃H⁺ [C]</td>
<td>ΔH</td>
<td>315.5</td>
<td>317.9</td>
<td>314.1</td>
<td>315.4</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>308.8</td>
<td>310.7</td>
<td>307.0</td>
<td>308.2</td>
</tr>
<tr>
<td>pSer COO⁻ [D]</td>
<td>ΔH</td>
<td>316.1</td>
<td>317.5</td>
<td>313.6</td>
<td>314.2</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>309.9</td>
<td>311.7</td>
<td>307.8</td>
<td>308.4</td>
</tr>
<tr>
<td>pSer COO⁻ [E]</td>
<td>ΔH</td>
<td>316.7</td>
<td>318.0</td>
<td>313.6</td>
<td>314.4</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>311.5</td>
<td>312.8</td>
<td>308.4</td>
<td>309.2</td>
</tr>
<tr>
<td>pSer zwitt [F]</td>
<td>ΔH</td>
<td>316.3</td>
<td>320.1</td>
<td>315.0</td>
<td>316.2</td>
</tr>
<tr>
<td></td>
<td>ΔG</td>
<td>310.1</td>
<td>313.5</td>
<td>308.4</td>
<td>309.6</td>
</tr>
<tr>
<td>A</td>
<td>ΔΔH</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>ΔΔG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>ΔΔH</td>
<td>1.8</td>
<td>0.3</td>
<td>-0.5</td>
<td>-0.4</td>
</tr>
<tr>
<td>B</td>
<td>ΔΔG</td>
<td>3.5</td>
<td>1.3</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>C</td>
<td>ΔΔH</td>
<td>1.7</td>
<td>2.2</td>
<td>1.5</td>
<td>1.8</td>
</tr>
<tr>
<td>C</td>
<td>ΔΔG</td>
<td>1.7</td>
<td>1.5</td>
<td>1.0</td>
<td>1.1</td>
</tr>
<tr>
<td>D</td>
<td>ΔΔH</td>
<td>2.3</td>
<td>1.8</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>D</td>
<td>ΔΔG</td>
<td>2.8</td>
<td>2.5</td>
<td>1.8</td>
<td>1.3</td>
</tr>
<tr>
<td>E</td>
<td>ΔΔH</td>
<td>2.9</td>
<td>2.5</td>
<td>1.0</td>
<td>0.8</td>
</tr>
<tr>
<td>E</td>
<td>ΔΔG</td>
<td>4.4</td>
<td>3.6</td>
<td>2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>F</td>
<td>ΔΔH</td>
<td>2.5</td>
<td>4.4</td>
<td>2.4</td>
<td>2.6</td>
</tr>
<tr>
<td>F</td>
<td>ΔΔG</td>
<td>3.0</td>
<td>4.3</td>
<td>2.4</td>
<td>2.5</td>
</tr>
</tbody>
</table>

a B3LYP = B3LYP/DZVP2. b aD = aug-cc-pVDZ. c aT = aug-cc-pVTZ.
Anion structures pSer-A and pSer-B are ‘phosphate’ deprotonation and anion structures pSer-D and pSer-E are ‘carboxylate’ deprotonation. In all four cases, there is one hydrogen bond formed between the hydrogen H1 in the −NH₂ group and oxygen O5 in the carboxylate group. For all four anions, there is no O5⋯H18 hydrogen bond (r ≥ 3 Å). The (O5⋯H1) hydrogen bond in pSer-D and pSer-E is ∼ 2 to 2.2 Å showing that the conformational difference between the structures has no effect on the hydrogen bond distance. The (O5⋯H1) hydrogen bond distance for pSer-A and pSer-B is weaker, ∼ 2.3 and 2.6 Å respectively. The slight increase in the hydrogen bond distance in pSer-B is likely due to the presence of a hydrogen bond of O5 with H8. In pSer-C, the −NH₂ group is stabilized between the −CO₂H and the −PO₃H⁻ groups.

Structure pSer-F is formed from the additional hydrogen bond between the nitrogen and H8. Structure pSer-C is similar to structure pSer-F with the only difference being the transfer of the proton H8 from O5 in the carboxylic acid group to the nitrogen to form the −NH₃⁺ group. In both pSer-C and pSer-F structures, H1 is bound to the nitrogen with an undistorted bond distance of ∼ 1.02 Å and is not involved in hydrogen bonding with the −PO₃H⁻ and −CO₂⁻ groups, (r ≥ 3.1 Å). The (N-H18) bond is elongated with respect to the (N-H1) bond, ∼ 1.04 and 1.09 Å for C and F respectively showing partial transfer of the proton to the −PO₃H⁻ group. In structure pSer-C, the (O5-H8) and (N⋯H8) bond distances are ∼ 1.01 and 1.73 Å showing that the proton is bound to the carboxylic acid group whereas in structure pSer-F the (O5⋯H8) and (N-H8) bond distances are ∼ 1.71 and 1.06 Å. This shows that the proton is partially transferred to the −CO₂⁻ group but is much closer to the nitrogen forming the zwitterionic structure.
Figure 5.3. Infrared spectra of deprotonated phosphoserine anions. IRMPD spectrum is on top. Anions [A] – [F]: calculated IR spectra at the B3LYP/aug-cc-pvdz level. Relative energies of isomers at the G3(MP2) level of theory. IRMPD spectrum are from Ref. 12.
Figure 5.4. Infrared spectra of deprotonated phosphothreonine anions. IRMPD spectrum is on top. Anions [A] – [F]: calculated IR spectra at the B3LYP/aug-cc-pvdz level. Relative energies of isomers at the G3(MP2) level of theory. IRMPD spectrum are from Ref. 12.
5.4.2. Phosphotyrosine (pTyr): IR spectra, GAs, and hydrogen bonding

Phosphotyrosine has a phenolic side chain that has been modified to incorporate a phosphate group. As discussed in Chapters 3 and 4, tyrosine is capable of sidechain deprotonation and is one of the more acidic amino acids (and amino acid amides). Phosphotyrosine was found to be slightly more acidic than phosphoserine and phosphothreonine, has an experimental GA value of 307.0 ± 4.0 kcal/mol, and a theoretical GA of 306.7 kcal/mol. Phosphotyrosine was bracketed between 1,1,1,5,5,5-hexafluoro-2,4, pentadione (GA = 310.3 ± 2 kcal/mol) and bistrifluoroaceteamide (GA = 307.5 ± 2 kcal/mol). All pTyr kinetic plots gave pseudo-first order kinetic plots indicative of one ion population.

For pTyr, two low energy neutral conformers and six corresponding low energy anion conformer/isomers were found (Figure 5.5 and 5.6) The lowest energy neutral, pTyr-1, has only one strong hydrogen bond between the −CO₂H and the −NH₂ groups (r(N···H(O)) = 1.9 Å). In the higher energy neutral, pTyr-2, the hydrogen in the −CO₂H group is rotated away from the nitrogen, so the −NH₂ group rotates to form a weaker hydrogen bond with the −CO₂H group (r(N(H)···O) = 2.6 Å).

Only the lowest energy anion structure, pTyr-A, is in agreement with the experimental value. The next lowest energy anion structure pTyr-B has a calculated GA of 310.7 kcal/mol at the G3(MP2) level (Figure 3). The most stable anion structure pTyr-A is a carboxylate deprotonated structure. The benzene ring distorts to allow the −CO₂− group to hydrogen bond with the −PO₃H₂ group. Both of the hydrogens in the phosphate group form strong hydrogen bonds with one oxygen in the carboxylate group (r(O(H)···O) = 1.6 and 1.7 Å). Additionally, this oxygen forms a weaker hydrogen bond with the NH₂ (r(N(H)···O) = 2.4 Å). Anion structure pTyr-B is similar to pTyr-A except that the phosphate group is deprotonated. The −CO₂H group
forms one strong hydrogen bond with the \( -\text{PO}_3\text{H}^- \) group (1.7 Å) and maintains the weaker hydrogen bond (2.4 Å) with the \( -\text{NH}_2 \) group found in anion structure pTyr-A. Anion structure pTyr-C is an unfolded phosphate deprotonated structure, so there is no hydrogen bonding present between the \( -\text{CO}_2\text{H} \) and \( -\text{PO}_3\text{H}^- \) groups. Both \( -\text{PO}_3\text{H}^- \) anion structures pTyr-B and pTyr-C are 4.0 kcal/mol higher in energy than the lowest energy structure pTyr-A. Anion structures pTyr-D and pTyr-E are zwitterionic with \( -\text{NH}_3^+ \), \( -\text{PO}_3\text{H}^- \), and \( -\text{CO}_2^- \) groups. Anion structure pTyr-D is a folded zwitterionic structure stabilized through hydrogen bonding with the \( -\text{PO}_3\text{H}^- \) and \( -\text{CO}_2^- \)

![Phosphotyrosine](image)

**Figure 5.5.** The most stable neutral and two anions for pTyr at the G3(MP2) level. Important hydrogen bond distances are given in Å.
Figure 5.6. Infrared spectra of deprotonated phosphotyrosine. IRMPD spectrum is on top. Anions [A] – [D] & [F]: calculated IR spectra at the B3LYP/aug-cc-pvdz level. Relative energies of isomers at the G3(MP2) level of theory. IRMPD spectrum are from Ref. 12.
groups and is predicted to be 11.0 kcal/mol higher in energy than anion structure pTyr-A. Anion structure pTyr-E is an unfolded zwitterionic anion that is only stabilized through hydrogen bonding with the –CO$_2^-$ group. This structure could only be optimized at the B3LYP/6-31+G(d) and HF/6-31G(d) levels of theory. Thus, the G3(MP2) method had to be adjusted to obtain the calculated GA value shown in Figure 3. Since this structure could not be optimized at the MP2/6-31G(d) level of theory, the optimized HF geometry was used to perform the single point QCISD(T) and MP2/Large calculations. The approximate G3(MP2) GA value for anion pTyr-E is 19.3 kcal/mol higher in energy than anion pTyr-A. Anion structure pTyr-F is an unfolded carboxylate anion and is 20.3 kcal/mol higher in energy than pTyr-A at the G3(MP2) level.

Maitre and coworkers$^{12}$ measured the IRMPD spectrum of deprotonated phosphotyrosine. They compared the experimental spectrum to their calculated spectra for their lowest energy PO$_3$H$^-$, COO$^-$, and zwitterionic anions. These three structures correspond to our anion structures pTyr-C, pTyr-E, and pTyr-F respectively. They did not find either of the folded anion structures pTyr-A or pTyr-B, which have distorted benzene rings and hydrogen bonding between the phosphate and carboxylate groups. Anion structure pTyr-A is the only structure that matches the experimental GA obtained in our proton transfer reactions involving pTyr anions produced from a common ESI solvent system. The calculated IR spectra for the [pTyr – H]$^-$ anion structures pTyr-A to pTyr-D and pTyr-F at the B3LYP/aug-cc-pvdz level as well as Maitre and coworkers experimental IRMPD spectrum are given in Figure 5.7. Anion structure pTyr-E is not shown because it could not be optimized at this level of theory. A table containing the position of the experimental bands, the calculated frequencies of the anion structures pTyr-A, pTyr-B, pTyr-C, and pTyr-D at the B3LYP/aug-cc-pvdz, the ratio between the experimental and calculated frequencies, and the previously proposed assignments of the IRMPD bands by Maitre and
coworkers for [pTyr-H]− is given in the Supporting Information. Maitre and coworkers suggest that the experimental IR spectrum corresponds to their lowest energy PO3H− anion, which is our anion structure pTyr-C. While their calculated spectrum does match the IRMPD spectrum relatively well, we have shown that anion structures pTyr-B to pTyr-F do not match our experimental GA. This suggests that the IRMPD experiment is missing the ground state distorted benzene structure. In previous work58 on the amino acid acidities, we showed that the IRMPD experiments11 did not yield the lowest energy structure for cysteine. Our calculated G3(MP2) results as well as experiments using the extended kinetic and gas-phase equilibrium methods13 found that the ground state anion resulted from deprotonation of the side chain –SH group. In contrast, the IRMPD results showed that the lowest energy anion was generated from deprotonation of the carboxylic acid. We suggested that the complications in the interpretation of the IRMPD spectra could originate from the partial proton transferring we predicted between the side chain thiol and carboxylic acid groups. In the case of phosphotyrosine, the conditions used to prepare the anion in the IRMPD experiment are likely to be biased against the folded anion. Sample preparation and ESI conditions varied between our work and that of Maitre and coworkers. Our standard solvent system was 49.5: 49.5: 1 MeOH:H2O:NH4OH, while Maitre and coworkers used a solvent of ~60:20:20 ACN:IPA:H2O with ammonium acetate present at 1.0 µM.12 Our ESI drying and nebulizer gas temperature was 220º C, while Maitre and coworkers used 150º C. Due to the higher concentration of water our solvent was significantly less volatile than that of Maitre and coworkers and, consequently, required a higher drying gas temperature to produce sufficient ion signal for study. Although less likely to significantly impact the results, our concentration of 60 µM pTyr and ESI high voltage of 3.5-4.0 kV were also lower than the Maitre conditions of 1 mM pTyr and 4.0 kV.
To investigate the impact of these varied experimental parameters, we performed ion/molecule reactions to bracket the GA involving pTyr anions produced with the solvent and ESI conditions of Maitre and coworkers. Anions from the Maitre parameters yielded a GA of 307.8 ± 3.4 kcal/mol for pTyr, while a GA of 307.0 ± 4.0 kcal/mol was obtained with our standard solvent system and ESI conditions. The more acidic nature of ions formed from our solvent/conditions is clearly illustrated by the pseudo-first-order kinetics plots of Figure 5.7, which were obtained for reactions pTyr anions with neutral bistrifluoroacetamide (GA= 307.5 ± 2 kcal/mol\textsuperscript{63}) at a constant pressure of 3.0 x 10\textsuperscript{-8} mbar. To facilitate comparisons, all of the data in Figure 5.7 was obtained on the same day with no major experimental changes other than those noted above. In addition, removal of the ammonium acetate additive from the Maitre solvent system had no effect on the reaction rate.

The two sets of experimental conditions undoubtedly yield different predominant deprotonated pTyr ion structures. This is most likely to be a consequence of either the change in solvent, which can impact the solution-phase conformation of pTyr neutrals, or the change in temperature, which can affect ESI desolvation. Several studies have found that solvent and ESI source conditions can influence sites of ion protonation/deprotonation and conformation.\textsuperscript{13-19, 78} For example, work by us and by Tian and Kass\textsuperscript{13,14} on tyrosine anions has previously shown that changes in the ESI solvent (protic vs. aprotic) affect deprotonation site and conformation (Chapter 3). In the current study of pTyr, both solvents are protic but their different volatilities required in a 70º C difference in ESI drying and nebulizer gas temperatures. For proteins, an increase in temperature at the ESI source can result in a conformational change that unfolds the molecule.\textsuperscript{20, 21, 24, 79-81} However, we have found no reports of the effect of ESI source temperature on the structures of small organic ions. We postulate that the increased temperature with a
Figure 5.7. Reactant ion loss plot for deprotonated phosphotyrosine reacting with bisfluoroacetamide (GA= 307.5 ± 2 kcal/mol) at a constant pressure of 3.0 x 10^{-8} mbar. In red, pTyr was electrosprayed from a 49.5:49.5:1 by volume solvent of MeOH:H₂O:NH₄OH. In blue, pTyr was electrosprayed from a 60:20:20 ACN:IPA:H₂O by volume solvent with an addition of 1 µM NH₄OAc.

methanol:water solvent may have provided the energy to overcome an activation barrier allowing formation of the lower energy folded pTyr anion structure. The folding process to convert pTyr anion C into p-Tyr anion A is complicated involving at least 5 bond rotations which can occur in different orders. Thus we have not attempted to find the highest energy transition state for this complex combinatorial problem. We have examined one such set of rotations without optimization of each one and, there is a barrier between A and C for the single process examined.
5.4.3. pSer, pThr, and pTyr amide GAs

The phosphorylated amides were formed by converting the carboxylic acid functional group (CO\(_2\)H) into an amide. We have used this substitution previously to better understand side chain acidity, as well backbone acidity, as the proton can no longer be lost from the carboxylic acid to form the acid (CO\(_2^-\)) as described in Chapter 4. The optimized geometries of the low energy neutrals and corresponding anions for pSer, pThr, and pTyr amides and their relative energies at the G3(MP2) level are given in Figure 5.8. The experimental and lowest energy calculated GA for each phosphorylated amino acid amide are given in Table 5.1. Because the CO\(_2^-\) anion can no longer be formed, the lowest energy GAs result from deprotonation of the phosphate group.

For pSer amide, the lowest energy neutral, pSer(amide)-1, the amide group is involved in two hydrogen bonds. It forms one hydrogen bond with the NH\(_2\) group (2.18 Å) and one strong hydrogen bond with the phosphate group (1.75 Å). In the higher energy neutral structure, pSer(amide)-2, the −NH\(_2\) group is involved in two hydrogen bonds, one with the −C(O)NH\(_2\) group (2.15 Å) and one with the −PO\(_3\)H\(_2\) group (1.70 Å). Both of the PO\(_3\)H\(_-\) anion structures pSer(amide)-A and pSer(amide)-B are stabilized by two hydrogen bonds with A being ~3 kcal/mol more stable than pSer(amide)-B. The experimental GA of pSer amide is 314.0 ± 4.3 kcal/mol and is in excellent agreement with the predicted value for our lowest energy anion structure pSer(amide)-A. Anion structure pSer(amide)-C is the zwitterionic anion and is higher in energy by 16 kcal/mol. Thus, a low energy zwitterionic anion is only seen with phosphoserine as there is no second anionic site in the amide.
Figure 5.8. Continued on following page.
Figure 5.8. Optimized structures of phospho-serine, -threonine, and -tyrosine amide neutrals and corresponding anions at the G3(MP2) level. GAs and relative energies are in kcal/mol. Important hydrogen bond distances are given in Å.
For pThr amide, a single low energy neutral and three anions were found. The neutral and higher energy anion structures contain three hydrogen bonds whereas the lowest energy anion structure only contains two. The experimental GA of pThr amide is $315.3 \pm 4.3$ kcal/mol and is in excellent agreement with both of our calculated anion structures pThr(amide)-A and pThr(amide)-B, which differ in energy by 0.6 kcal/mol. As for pSer amide, anion structure pThr(amide)-C is only shown to further demonstrate that the zwitterionic anion will not form with the amides.

For pTyr amide, two low energy neutrals and only one low energy anion were found. The lowest energy neutral pTyr(amide)-1, is an unfolded structure with only one hydrogen bond between the $-\text{C(O)NH}_2$ and $-\text{NH}_2$ groups (2.17 Å) whereas the higher energy neutral structure, pTyr(amide)-2, is folded with two hydrogen bonds, one between the $-\text{C(O)NH}_2$ and $-\text{NH}_2$ groups (2.24 Å) and one between the $-\text{C(O)NH}_2$ and $-\text{PO}_3\text{H}_2$ groups (1.82 Å). The experimental GA of pTyr amide is $307.2 \pm 4.0$ kcal/mol, consistent with our calculated GA of 307.0 kcal/mol.

The experimental reaction rates of pSer(amide) and pThr(amide) are significantly different than their corresponding acids. However, in the case of pTyr(amide,) the amide and acid react at essentially the same rates. Figure 5.9 displays this behavior. In (c) and (d), it is clear that the reaction of pTyr and pTyr(amide) with 1,1,1,5,5,5-hexafluoro-2,4-pentadione (GA = 310.4 kcal/mol) show that the acidities are similar, as they are reacting at nearly the same rates. Both the acid and amide react completely within 300 seconds. In the case of pSer and pSer(amide) (also for pThr and its amide, not shown), in (a) and (b) the rates are significantly different. The reaction with 1,1,1,5,5,5-hexafluoro-2,4-pentadione takes 100 seconds for pSer to completely react, however, pSer(amide) is less acidic and only takes 5 seconds to react to completion.
Phosphoserine is more acidic than its amide which is reflected in the experimental and computational GA values.

5.4.4. Comparison of Acidities of Phosphorylated Species with Non-Phosphorylated Species

In summary, the GAs of the phosphorylated amino acids are 13 to 27 kcal/mol more acidic than their corresponding non-phosphorylated amino acids. The largest change is predicted for Tyr/pTyr pairs (~23 kcal/mol). pSer and pThr, are ~17 kcal/mol more acidic than their corresponding non-phosphorylated amino acids. The phosphorylated amino acids are also 15 to 20 kcal/mol more acidic than phosphoric acid (GA = 322.2 kcal/mol) and more similar to sulfuric acid (GA = 304.6 kcal/mol), a strong gas-phase acid. In general, the phosphorylated amino acids are 1 to 7 kcal/mol more acidic than their corresponding amides. For pTyr vs. pTyr(amide), the acids and amides have essentially the same acidities within < 0.5 kcal/mol. For
pSer and pThr vs. their amides, this is not the same. Experimental rate data displaying the
difference in reactivities of acids vs. amides is located in Figure 5.9. The phosphorylated amino
acid amides are 22 to 30 kcal/mol more acidic than their corresponding non-phosphorylated
amino acid amides.

5.5. Conclusions

The gas-phase acidities for the common phosphorylated amino acids and their amides
were predicted and experimentally determined, providing the first reliable set of these values.
The calculated GAs at the G3(MP2) level for pSer, pThr, and pTyr and their amides were
compared to GAs determined by proton transfer reactions in a mass spectrometer. Excellent
agreement was found between the predicted GAs and the experimental values. The lowest energy
anions for pSer and pThr are generated from phosphate deprotonation. For pTyr, the most stable
anion is a carboxylate deprotonated structure in which the benzene ring is distorted to allow for
strong hydrogen bonding between the carboxylate and phosphate groups. The infrared spectra of
the anions of pSer, pThr, and pTyr were calculated and compared to experimental IRMPD
spectra. For \([M - H]^-\) deprotonated \([pSer - H]^-\) and \([pThr - H]^-\) good agreement is found
between the experimental spectra and the calculated spectra for our lowest energy anion
structure. For pTyr, good agreement is not found between the calculated and experimental
spectra with our lowest energy anion and instead the IRMPD spectra correspond to a higher
energy phosphate deprotonated structure. However, the experimental and predicted GAs are in
good agreement with each other for the lowest energy structure pTyr-A of the anion and the
calculated GA for structure pTyr-C is not in good agreement with the experimental GA. This
suggests that the conditions used to generate the ion \([pTyr - H]^-\) from which the IRMPD
spectrum was obtained did not generate the lowest energy structure. Our experiments involving
similar solvent and ESI conditions to those used in the IRMPD experiments likewise yielded a higher energy [pTyr-H]− structure, providing a clear indication that ESI solvent and temperature affect the deprotonation of pTyr.

In general, the phosphorylated amino acids are 13 to 27 and 1 to 7 kcal/mol more acidic than their corresponding non-phosphorylated amino acids and phosphorylated amino acid amides respectively. This energy difference between the phosphorylated amino acids and their amides differs from what was found for the simple amino acids and shows the importance of strong hydrogen bonding between the phosphate group and the carboxylate groups in the anions. The phosphorylated amino acids are 15 to 20 kcal/mol more acidic than phosphoric acid and are more similar in acidity to sulfuric acid, a strong gas phase acid.

References


CHAPTER 6: FRAGMENTATION OF MODEL PHOSPHORYLATED PEPTIDES

6.1 Overview

Nine model phosphorylated hexapeptides (XAAAAA, AAXAAA, AAAAXA, where X = phosphorylated serine, threonine, and tyrosine) were synthesized and their deprotonated ions were subjected to collision-induced dissociation (CID). For singly charged precursor ions, the location of the phosphorylated amino acid was observed to affect dissociation patterns. Phosphorylation of the residue at the N-terminal position yields the highest amount of fragmentation, followed by near-C-terminal, and central positions. There were also multiple signature losses observed for the dissociation of phosphopeptide anions. The diagnostic ions at \( m/z \ 97 \) (\( \text{H}_2\text{PO}_4^- \)) and \( m/z \ 79 \) (\( \text{PO}_3^- \)) indicate the presence of a phosphate group on the peptide. In addition, characteristic losses of formaldehyde and acetaldehyde were observed in the CID spectra for phosphoserine- (pSer) and phosphothreonine- (pThr) containing peptides. Backbone fragment ions produced from phosphotyrosine- (pTyr) containing peptides commonly undergo elimination of HPO\(_3\) (80 Da). In addition, singly charged phosphopeptide precursor ions readily lose H\(_3\)PO\(_4\) (98 Da), as do the backbone cleavage ions from pSer and pThr peptides. When performing CID on doubly charged precursor ions, the relative intensities of the product ions, \([M–2H–\text{PO}_3^-]\) and \([M–2H–\text{H}_2\text{PO}_4^-]\), yield information about the type of phosphorylated residue present in the peptide.
6.2 Introduction

Phosphorylation is a common post-translational modification (PTM) in peptides and proteins.\textsuperscript{1-3} As discussed in Chapter 5, when an amino acid residue is phosphorylated, the acidity of the peptide increases significantly. The three common phosphorylated amino acids are pSer, pThr, and pTyr, all of which are phosphorylated on their hydroxyl side chains. Phosphorylation is a dynamic process biologically; phosphates are added enzymatically by kinases and removed by phosphatases.\textsuperscript{1,2} Phosphorylation mediates many cellular functions such as enzyme inhibition and cell signaling.\textsuperscript{1,2}

Because phosphorylated peptides are prevalent in nature, there have been several attempts to sequence phosphopeptides for proteomic studies by CID. Though many of these studies have been in the position ion mode,\textsuperscript{4-23} because phosphates are acidic and readily deprotonate, they lend themselves well to negative ion mode studies. There have been fewer studies of peptide anions; however, the literature in this field discusses phosphopeptide signal enhancement, relative fragmentation ion energetics, and neutral losses commonly observed for phosphopeptide anions.\textsuperscript{4-17} For example, several studies by Bowie and coworkers\textsuperscript{7,9-12,14} have examined peptides in both the negative and positive ion modes and have completed theoretical work on common losses, HPO\textsubscript{3} (80 Da) or H\textsubscript{3}PO\textsubscript{4} (98 Da), from pSer, pThr, and pTyr. Lehmann and coworkers have also extensively studied negatively charged phosphopeptides.\textsuperscript{6,13,15,16,18} Important to the findings of the work in this chapter is the migration of phosphoric acid in pTyr-containing peptides. Lehmann and coworkers\textsuperscript{13} studied several pTyr-containing peptides and found that the phosphate group can migrate from pTyr to other side chains. Their work even suggests the C-terminus as a possible phosphate group acceptor. Lehmann and coworkers\textsuperscript{16} have
also studied the loss of 79 Da (PO$_3^-$) from doubly charged phosphopeptide anions, and have discussed the possibility of different deprotonation sites affecting fragmentation by CID.

Nearly every mass spectrometer has CID capabilities. It logically follows that the most widely used form of fragmentation to study phosphopeptides is CID. However, CID does not always produce complete sequence coverage for peptides and often low-energy rearrangements cause specific ions to dominate spectra. For phosphopeptides, a common problem in identifying the original site of phosphorylation stems from the loss of the phosphate group itself.$^{8-10, 15, 16, 19}$ This is observed in both positive and negative ion mode CID.$^8$ In biological peptides containing several amino acids capable of phosphorylation, the loss or migration of the phosphate group is a commonly observed low-energy process that can complicate CID spectra.$^{20-26}$ To provide complementary data to CID, both negative and positive ion electron-based techniques have also been utilized to sequence phosphorylated peptides.$^{27-33}$

Collision-induced dissociation can produce diagnostic ions or neutral characteristic losses. Diagnostic ions can indicate the presence of a particular amino acid residue incorporated in the peptide chain. Negative ion studies by Bowie and coworkers$^{7, 9, 11, 12, 14, 19, 34-36}$ on deprotonated peptides and phosphopeptides have resulted in the identification of characteristic fragments and neutral losses, such as the elimination of the phosphate group from phosphopeptides, the loss of side chain aldehydes from peptides containing hydroxyl side chains, and diagnostic cyclic di- and tri-phosphate anions from peptides with multiple phosphorylated residues. Lehmann and coworkers$^{15, 16}$ have discussed strategies for identification of pThr and pSer by neutral loss of H$_3$PO$_4$ for negative peptide ions. Additionally, several diagnostic ions have been examined for protonated phosphopeptides. In a positive ion mode electron capture dissociation (ECD) study, Cooper and coworkers$^{32}$ performed experiments where the neutral loss
of H$_3$PO$_4$ from CID of protonated phosphopeptides was used as a diagnostic indicator. The loss of H$_3$PO$_4$ instructed their data acquisition program to perform additional ECD experiments to obtain more sequence information on peptides that lose a phosphate group. Reid and coworkers, in a different positive ion mode study, discuss an interesting diagnostic loss of acetaldehyde and formaldehyde from pThr and pSer-containing peptides, respectively. Though some characteristic losses and diagnostic peaks have been discovered for phosphopeptides in both the negative and positive ion modes, the generally observed trend is that diagnostic ions can indicate the presence, but not necessarily the position, of a phosphate group.

This chapter discusses the search for diagnostic ions from the CID of deprotonated model phosphorylated hexapeptides to assist in phosphoproteomic studies. The synthesis involved varying the location of the phosphorylated amino acid residue. The peptide sequences are XAAAAA, AAXAAA, and AAAAXA where X = pSer, pThr, or pTyr. Alanine is used as a residue with a neutral side chain because its methyl group will not accept a phosphate group via migration, and also will not undergo rearrangements or reactions. That is, alanine is considered “inert.” The phosphorylated residue could not be placed directly on the C-terminus because no Wang resins are available for peptide synthesis with phosphorylated amino acids residues.

6.3 Experimental

Experiments were performed using a Bruker Daltonics (Billerica, MA, USA) BioApex 7T Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometer. All peptide solutions were 60 µM in a solvent of volume ratio of 49.5:49.5:1 CH$_3$OH:H$_2$O:NH$_4$OH. As described in previous chapters, 1% ammonium hydroxide was added to facilitate deprotonation of the amides in the solution phase prior to electrospray ionization (ESI). Electrospray conditions were used as described in Chapter 2.
In low-energy CID experiments, the precursor anions were isolated by correlated frequency resonance ejection techniques and activated by sustained off-resonance irradiation (SORI). An optimized \( V_{pp} \) pulse at 700 Hz off-resonance (lower frequency) from the precursor ion was employed for a duration of 100-150 ms. The collision gas was argon. When the collision gas was pulsed in using a pulse valve, the pressure reached \( 10^{-5} \) mbar and was pumped away for 2-3 seconds prior to detection. This is further discussed in Chapter 2.

As a comparison, some low-energy CID experiments were performed on a quadrupole ion trap (QIT), a Bruker Daltonics HCTultra PTM Discovery System equipped with ESI. The electro spray conditions were comparable to the FT-ICR experiments. The collision gas was helium for CID.

All peptides were synthesized in-house with an Advanced ChemTech Model 90 peptide synthesizer (Louisville, KY, USA) using procedures described for Fmoc solid state peptide synthesis as described in Chapter 2.

6.4 Results

6.4.1 CID of [M – H]− for pThr-containing Peptides

Peptides with pThr residues were generated by ESI and the [M – H]− precursor ions were investigated using CID (Figure 6.1). The phosphorylated residue of the studied pThr peptides is located at three distinct positions: N-terminal (Figure 6.1(a)), central (Figure 6.1(b)), and the penultimate residue to the C-terminus (subsequently referred to as “near-C-terminal”) (Figure 6.1(c)). The four series of backbone product ions, "a-, "b-, c-, and y-ions, case of the near C-terminal pThr peptide. The peptide, AAAApTA, dissociates to yield several c- ions that do not contain the phosphorylated residue. A known diagnostic ion\(^{19} \) corresponding to the loss of both the phosphate group (98 Da) and acetaldehyde (44 Da) is present in all pThr peptide CID spectra,
Figure 6.1. SORI-CID spectra of $[M-H]^{-}$ from the pThr-containing peptides (a) pTAAAAA, (b) AApTAAA, and (c) AAAApTA.
Previously reported diagnostic ions for the phosphate group, H$_2$PO$_4^-$ and PO$_3^-$, are intense in the CID spectra. Table 6.1 lists the formulas, names, and nominal masses for all of the common neutral losses and ion fragments from the CID spectra of pThr-, pSer, and pTyr-containing peptides.

<table>
<thead>
<tr>
<th>Chemical Formula</th>
<th>Name</th>
<th>Nominal Mass (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_3$PO$_4$</td>
<td>Phosphoric Acid</td>
<td>98</td>
</tr>
<tr>
<td>HPO$_3^-$</td>
<td>Metaphosphoric acid</td>
<td>80</td>
</tr>
<tr>
<td>H$_2$PO$_4^-$</td>
<td>Dihydrogen Phosphate Ion</td>
<td>97</td>
</tr>
<tr>
<td>PO$_3^-$</td>
<td>Phosphite Ion</td>
<td>79</td>
</tr>
<tr>
<td>C$_2$H$_4$O</td>
<td>Acetaldehyde</td>
<td>44</td>
</tr>
<tr>
<td>CH$_2$O</td>
<td>Formaldehyde</td>
<td>30</td>
</tr>
<tr>
<td>H$_3$PO$_4$ + C$_2$H$_4$O</td>
<td>Phosphoric Acid + Acetaldehyde</td>
<td>142</td>
</tr>
<tr>
<td>H$_3$PO$_4$ + CH$_2$O</td>
<td>Phosphoric Acid + Formaldehyde</td>
<td>128</td>
</tr>
</tbody>
</table>

**6.4.2 CID of [M – H]$^-$ for pSer-containing Peptides**

The [M – H]$^-$ precursor ions for pSer peptides were fragmented by CID (Figure 6.2). Like pThr peptides, the phosphorylated residue is placed in three positions: N-terminus (Figure 6.2(a)), central (Figure 6.2(b)), and near-C-terminus (Figure 6.2(c)). The typical loss of H$_3$PO$_4$ from the "a-, "b-, c-, and y-ions was observed. The same diagnostic ions, indicating the presence of a phosphate group, seen in the CID spectra for pThr peptides were also observed in the CID spectra of pSer peptides. The diagnostic ion associated with the combined loss of the phosphate group and an aldehyde, corresponds to loss of formaldehyde and H$_3$PO$_4$ in the pSer peptides.
Dissociation of deprotonated pSer peptides yields internal fragment ions, including an internal pSA$^-$ ion with the same $m/z$ as the "$b_2" ion from pSAAAAA (labeled as Int. pSA, Figure 6.2(b)) and a dialanine anion (AlaAla$^-$). Structures for the dialanine anion, pSA$^-$ ion, and the diagnostic ions for the phosphate group are located in Figure 6.4.

**6.4.3 CID of [M - H]$^-$ for pTyr-containing Peptides**

The CID spectra for deprotonated pTyr peptides are located in Figure 6.3. The pTyr peptides produce different fragments by CID than the pSer and pThr peptides. The product ions include members of the "$a\text{-}, "b\text{-}, c\text{-}, \text{ and y-ion series, but these ions lose the phosphate group commonly in the form of HPO}_3\text{. Diagnostic ions and common losses associated with the phosphate group are present in the spectra. Figures 6.4(a) and (c) show that the peptides with pTyr at the N-terminal and near-C-terminal positions yield more intense backbone fragment ions than AApYAAA, which is phosphorylated at the central position. Product ions resulting from backbone cleavage are often the most intense in the CID spectra; in Figure 6.3(a) and 6.3(c) $[c_2 - 80]^- \text{ and } c_4^-$ are the base peaks of the spectra, respectively. Overall, the pTyr peptides generally yield fewer product ions by CID than pThr or pSer-containing peptides.
Figure 6.2. SORI-CID spectra of $[M - H]^- from the pSer-containing peptides (a) pSAAAA, (b) AApSAAA, and (c) AAAApSA.
Figure 6.3. SORI-CID spectra of \([M–H]^–\) from the pTyr-containing peptides (a) pYAAAAA, (b) AApYAAA, and (c) AAAApYA.
Figure 6.4. Structures of diagnostic and internal ions observed in CID mass spectra.

6.4.4 CID of $[M - 2H]^2-$ for Phosphorylated Peptides

Figures 6.5-6.7 contain the CID spectra of the doubly deprotonated phosphopeptides. Figure 6.5 shows CID spectra for the XAAAAA peptides, whereas Figure 6.6 shows CID spectra for the AAXAAA peptides, and Figure 6.7 shows CID spectra for the AAAAXA peptides. The overall amount of fragment ions in CID spectra is greatly diminished for $[M - 2H]^2-$ compared to $[M - H]^-. The fragmentation of $[M - 2H]^2-$ results in predominantly two product ions, $[M - 2H - 79]^-$ and $[M - 2H - 97]^-$, for all model phosphopeptides and does not depend on the location of the phosphorylated residue. The base peak for pThr peptides is $[M - 2H - 79]^-$, whereas the base peak for pSer peptides is $[M - 2H - 97]^-. For the pTyr peptides, the product ion, $[M - 2H - 79]^-$,
dominates the spectra. In addition, some low intensity backbone cleavage ions of the "b-, c-, and y-series are present in CID spectra.

Included at the end of this chapter, Supplemental Figures 6.1-6.3 show CID spectra obtained using the HCTultra ESI/QIT. The CID spectra obtained from the QIT agree well with the SORI-CID spectra. Some diagnostic ions and small backbone fragments were not observed due to the low m/z discrimination of the QIT.40

6.5 Discussion

6.5.1 Product Ions Observed for CID of [M – H]–

For all model phosphopeptides, common negative ion mode backbone cleavages are observed upon CID of the precursor [M – H]–. The c- and y-ion series are frequently observed fragments for negative ion mode CID.41-47 Additionally, "a- and "b-ions, have been observed for tripeptides with alkyl side chains.41 The a- and b-ions in this chapter are "a- and "b-series ions, where the primes indicate that two hydrogen atoms have been lost from the backbone fragment ion as discussed in Chapter 2. The peptides in this work contain mostly alanine residues, so observation of "b-ions agrees with Harrison.41 Possible ion structures for common dephosphorylated peptide backbone fragments are given in Figure 6.8. The "a-ions shown are two of the most probable structures as described by Oomens,48 the linear "b-ions and y-ions are those described by Bowie,49 while Harrison50,51 and Oomens proposed the cyclic oxazalone "b-ions.52 The c-ions are deprotonated peptide amides, as studied by Grzetic and Oomens53 and Pu and Cassady.47

In addition to the traditional backbone cleavage ions, several neutral losses were observed from nearly all of the phosphorylated peptides including: H3PO4, HPO3, and H2O.
Figure 6.5 CID Spectra of \([M - 2H]\) for (a) pSAAAAA, (b) pTAAAAA, and (c) pYAAAAA.
Figure 6.6 CID spectra of $[M - 2H]^- \times 0.8$ for (a) AApSAAA, (b) AApTAAA, and (c) AApYAAA
Figure 6.7 CID spectra of $[M - 2H]^-$ for (a) AAAApSA, (b) AAAApTA, and (c) AAAApYA
Peptides containing pSer lost formaldehyde (CH$_2$O) and pThr peptides lost acetaldehyde (C$_2$H$_4$O). These are discussed in Section 6.5.3.

The loss of the phosphate group from deprotonated phosphorylated peptides is a well-known diagnostic loss.\textsuperscript{23, 38, 39} The product ions corresponding to elimination of 98 Da (H$_3$PO$_4$) from [M – H]$^-$ always have high intensity in the CID spectra. The loss of 98 Da has been studied by Lehmann and coworkers\textsuperscript{13} and is attributed to the elimination of neutral H$_3$PO$_4$, rather than sequential loss of HPO$_3$ and H$_2$O. The ion corresponding to HPO$_3$ elimination is not commonly observed in the CID spectra, so sequential loss is unlikely. The pSer and pThr peptides can lose either H$_2$PO$_4$ or HPO$_3$ from backbone fragment ions, but pTyr peptide backbone fragment ions predominantly lose HPO$_3$. Bowie and coworkers\textsuperscript{7, 10} have investigated the energetics involved with losing HPO$_3$ and H$_3$PO$_4$ from pTyr tetrapeptides using CID and calculations at the HF/6-31+G(d)//AM1 level of theory. They observed that elimination of the phosphate group involves an S$_{N2}$ process between the C-terminus and the pTyr side chain. In addition, the reactions to form H$_2$PO$_4^-$ and [M – H – H$_3$PO$_4$]$^-$ are endothermic, requiring 43 and 77 kcal/mol, respectively.\textsuperscript{14} The observation of the [M – H – H$_3$PO$_4$]$^-$ ion from pTyr peptides (Figure 6.4) indicates this endothermic process has occurred. The mechanism for elimination of H$_3$PO$_4$ will be further discussed in Section 6.5.2. In addition, the lower energy generation of PO$_3^-$ from pTyr peptides\textsuperscript{7, 10} results in a higher intensity peak at m/z 79 than observed for pSer (Figure 6.3) and pThr (Figure 6.1) peptides. Conversely, pSer and pThr peptides have formation energies for [M – H – H$_3$PO$_4$]$^-$ and H$_2$PO$_4^-$ that are more favorable than the energies associated with formation of [M – H – HPO$_3$]$^-$ and PO$_3^-$.\textsuperscript{12} This trend is reflected in the relative intensities of the diagnostic ions for the phosphate group in Figures 6.1-6.3 for all model phosphopeptides.
Peptides containing pSer often have CID spectra with a low signal-to-noise ratio (S/N). This is due to the benchtop synthesis for pSer peptides yielding many byproducts, complicating isolation of precursor ions. The peptides synthesized for this work were not purified in solution. Previous studies by another Cassady group researcher, Suma Kavati, also showed low S/N in the spectra for synthesized pSer peptides.
6.5.2 Effect of the Position of the Phosphorylated Residue on [M – H]$^-$ Fragmentation

In the CID spectra of [M – H]$^-$ for all phosphorylated peptides (XAAAAA, AAXAAA, and AAAAXA) many recurring ion types and neutral losses are observed. Backbone cleavage produces "a-, "b-, c-, and y-ions. The phosphate group is eliminated from backbone fragments in the form of H$_3$PO$_4$ for pThr and pSer peptides, and HPO$_3$ for pThr and pTyr peptides. Diagnostic ions for the phosphate group are always intense products at $m/z$ of 79 and 97. The phosphorylated residue does not enhance backbone cleavage at adjacent sites, but many backbone fragment ions contain the phosphorylated residue. In general, more N-terminal backbone product ions are present in the spectra than C-terminal ions.

For XAAAAA peptides, many sequence informative product ions are observed upon CID of [M – H]$^-$. The base peak in the CID spectrum is specific to the peptide being studied. The pSAAAAA peptide has [M – H – H$_3$PO$_4$]$^-$ as the base peak, whereas pTAAAAA has [M – H – H$_3$PO$_4$ – C$_2$H$_4$O]$^-$ as the base peak, and the base peak for pYAAAAA is [c$_4$ – HPO$_3$]$^-$. The pYAAAAA peptide produces more intense backbone fragment ions than the pSer and pThr peptides. For all three peptides, some y-ions without the phosphorylated residue are observed in the CID spectra. All pXAAAAA peptides contain fragments corresponding to abundant cleavage along the peptide backbone. All three N-terminal phosphorylated peptides have a nearly complete series of "b-ions, as well as several "a-ions, intense c-ions, and sparse low intensity y-ions. Because most product ions contain the N-terminus, they also contain the phosphorylated residue. This can be attributed to the high acidity of the phosphorylated amino acid amides (Chapter 5) that makes the phosphorylated residue the likely deprotonation site in ESI.

In the CID spectra of [M – H]$^-$ for AAXAAA, less sequence informative fragmentation is obtained, and backbone product ions often contain the phosphorylated residue. Each spectrum
for the three AAXAAA peptides is dominated by phosphate group marker ions: \([M - H - H_3PO_4]^-\), \(PO_3^-\) at \(m/z\) 79, and \(H_2PO_4^-\) at \(m/z\) 97. These diagnostic ions have the highest intensities in the CID spectra, Figures 6.1-6.3(b). The "a-, "b-, c-, and y-ions are present in all AAXAAA peptide CID spectra, but occur with lower intensity than in the spectra for XAAAAA and AAAXAA. Because the backbone fragment ions are much lower in intensity for peptides with centrally located phosphorylated residues, product ions may be harder to identify by the automated sequencing techniques\(^5\) used for unknown peptides from biological samples. For AApSAAAA, the CID spectrum is nearly identical to that for pSAAAAA (Figure 6.3(a) and (b)). The CID spectrum for AApTAAA has similar backbone fragments as seen for AApSAAA. Although there are ten backbone fragments for the AApTAAA anion, this location of the phosphorylated residue results in fewer and lower intensity backbone fragments than the pTAAAAA anion. A unique dissociation result for AApYAAA was observed, where every backbone fragment ion contains the phosphorylated residue. This outcome was not seen in the analogous pSer and pThr peptide CID spectra and could be helpful for identification of pTyr-containing peptides. The lower overall number of backbone fragments for AApYAAA compared to the other peptides provides limited information about the peptide sequence.

When the phosphorylated residue is located near the C-terminus, an intermediate amount of fragmentation is obtained from CID of \([M - H]^-\). Peptides with the sequence AAAAXA produce less sequence informative fragmentation than XAAAAA peptides, but more fragmentation than AAXAAA peptides. Peptides with near-C-terminal phosphorylation confirm the preferential formation of N-terminal ions, even without the incorporation of a phosphorylated residue. A high abundance of N-terminal backbone fragment ions that do not contain a phosphate group are an indication of a phosphorylated residue located far from the N-terminus. For
example, the AAAApSA peptide has many product ions, c1-c4 and "a3, that do not contain the pSer residue. This agrees with the work of Cassady group researcher, Dan Pu, on model non-phosphorylated peptides. Because the phosphorylated residue is located near the C-terminus, most y-ions include the phosphorylated residue. None of the backbone cleavage ions that contain the phosphorylated residue have significantly high intensity; therefore, it does not seem likely that cleavage is enhanced by the phosphorylated residue. This is surprising because CID fragmentation of peptides is generally charge-directed. This suggests that a proton may become mobile upon ion activation in CID and neutralized the deprotonated phosphate group.

Though peptide synthesis cannot be performed for an AAAAAX peptide, the AAAAXA peptide may be more representative of digested peptides by the bottom-up approach used for peptide sequencing. Most proteins are commonly digested with the enzyme trypsin, which cleaves on the C-terminal side of arginine and lysine residues. Therefore, the C-terminus of a biological sample would most likely not contain a phosphorylated amino acid. Both chymotrypsin and pepsin digest on the carboxyl side of aromatic amino acids, so there is a possibility that peptides with pTyr at the C-terminus may be generated from this enzymatic digestion, but pSer and pThr C-terminal peptides are not likely.

There are some notable trends associated with the product ions from CID of the model phosphorylated hexapeptides. In general, c-ions tend to be more abundant when the phosphorylated residue is not included, see Figures 6.1-3(c). When examining y-ions, in nearly every case, the phosphorylated residue is included. Also, the unusual occurrence of "b-ions as small as one residue is observed (Figure 6.1, 6.2(a)). In studies of negative b-ions, it is known that at least two residues must be present in an oxazalene structure, and in protonated peptide CID, b1"+ ions are rarely observed.
The loss of \( \text{H}_3\text{PO}_4 \) is common for backbone cleavage ions of pThr and pSer peptides, but is uncommon for pTyr peptides. Because of this, several researchers have referred to the loss of \( \text{H}_3\text{PO}_4 \) as diagnostic for peptides with pSer and pThr residues.\(^{15, 16, 19, 23, 38, 39}\) Lehmann and coworkers\(^ {38}\) proposed a six-membered ring transition state that explains the elimination of \( \text{H}_3\text{PO}_4 \) from peptides with pSer and pThr, but not pTyr (Scheme 6.1).

![Scheme 6.1](image.jpg)

Scheme 6.1 Loss of neutral \( \text{H}_3\text{PO}_4 \) from phosphopeptides by six-membered ring. This mechanism is modified from Lehmann and coworkers.\(^ {38}\)

The inability of pTyr to form this six-membered ring transition state is attributed to the unfavorable steric hindrance of the tyrosine side chain.\(^ {38}\) The pTyr peptides instead fragment by elimination of \( \text{HPO}_3 \) (Scheme 6.2), which is also observed for pThr peptides. The loss of \( \text{HPO}_3 \) leaves a negative charge on the side chain of the dephosphorylated residues. The bulky tyrosine side chain can stabilize that negative charge. Chapter 3 discusses the ability of tyrosine to deprotonate on its side chain, and Chapter 4 discusses the side chain deprotonation of tyrosine amide.

![Scheme 6.2](image2.jpg)

Scheme 6.2. Loss of metaphosphoric acid from phosphorylated peptide anions.
6.5.3 Diagnostic Ions Resulting from CID of \([M - H]^-\) Precursor Ions

Diagnostic CID product ions are useful in determining specific structural information about precursor ions. When examining phosphopeptides by negative ion mode CID, Bowie\(^8\), Reid\(^10\), and their coworkers have stated that characteristic losses alone cannot determine the site of phosphorylation. However, some of the diagnostic ions from CID that we report in this section are able to give insight to the type of phosphorylated residue incorporated in a peptide.

In agreement with previous literature for negative ions,\(^{23, 38, 39}\) we observe the diagnostic marker ions for the phosphate group at \(m/z\) 97 (\(\text{H}_2\text{PO}_4^-\)) and \(m/z\) 79 (\(\text{PO}_3^-\)). These ions are a major indicator that a phosphate group resides on an amino acid residue within a peptide. These are not observed in positive ion CID; however, the diagnostic loss of 98 and 80 Da occurs in both negative and positive ion CID of phosphopeptides.\(^8\)

In pThr peptides, the loss of 142 Da yields an important diagnostic ion. This mass corresponds to the combined loss of the phosphate group (as \(\text{H}_3\text{PO}_4\)) and an additional 44 Da, corresponding to \(\text{C}_2\text{H}_4\text{O}\). In negative ion mode CID studies of threonine-containing peptides, acetaldehyde is eliminated by either a charge-induced or charge remote mechanism.\(^{45, 59, 60}\) Bowie and coworkers report the signature losses of \(\text{CH}_2\text{O}\) from pSer peptides and \(\text{C}_2\text{H}_4\text{O}\) from pThr peptides.\(^{19}\) Previous work in the Cassady group also discusses the loss of neutral aldehydes from threonine- and serine-containing peptides.\(^{45}\) The combined loss of an aldehyde and a phosphate group from pSer and pThr peptides was also observed in the positive ion mode by Reid and coworkers.\(^{22}\) The loss of \(\text{H}_3\text{PO}_4\) and \(\text{C}_2\text{H}_4\text{O}\) (142 Da, for pThr) or \(\text{H}_3\text{PO}_4\) and \(\text{CH}_2\text{O}\) (128 Da, for pSer) from the precursor ion can be diagnostic for which residue is phosphorylated.

As previously discussed in Section 6.5.4, the CID spectrum of the AApSAAA anion is similar to that of the pSAAAAA anion. An internal fragment ion, \([\text{pSA} - 98]^-\), is observed for
AApSAAA with the same exact $m/z$ as the $[^\text{b2-98}]^-$ fragment in the negative ion CID of pSAAAAA (Figure 6.3). An internal fragment ion is not formed by cleavage of the peptide backbone on only one side; cleavage must have occurred on both sides of two internal amino acid residues. Only this phosphopeptide, with a centrally located pSer residue, produces an internal fragment ion containing the phosphorylated residue. However, when the pSer is located near the C-terminus in AAAApSA, the spectra do not contain the pSA$^-$ diagnostic ion.

Phosphoserine has been reported to have diagnostic internal ions in a positive ion mode study by Medzihradszky and Trinidad. In their work, peptides containing sequential proline-pSer residues showed a diagnostic internal “y-like” ion 10 $m/z$ greater than the normal y-ion observed upon cleavage of the proline-pSer peptide bond. The internal pSA$^-$ ion observed in this work is a potentially useful diagnostic ion as its formation appears to be dependent on the position of pSer in the peptide chain.

Phosphotyrosine is reported in the literature to characteristically lose HPO$_3$ in the negative ion (and positive ion) mode. However, this loss is not common in the CID spectra of this work. In pYAAAAA, no ion is seen at [M – H – HPO$_3$]$^-$. For AApYAAA and AAAApYA, the peak indicating loss of HPO$_3$ is low intensity but still diagnostic for a pTyr residue. Backbone cleavage ions often lose HPO$_3$, but the precursor ion mostly eliminates H$_3$PO$_4$. The loss of H$_3$PO$_4$ has been attributed to the possible migration of the phosphate group to the C-terminus to yield [M – H – H$_3$PO$_4$]$^-$.

As seen in Scheme 6.3, a mechanism is derived from Lehmann and coworkers involving zwitterionic [M – H]$^-$. Due the partial positive charge on the phosphorus, it is likely [M – H]$^-$ may be non-zwitterionic; this structure is also included in the mechanism. Upon CID activation of the precursor ion, the C-terminus no longer contains a carboxylic acid moiety.
because the phosphate group migrates to that position. This process could help explain the lack of C-terminal backbone product ions from pTyr peptides. To study how the position of the pTyr residue influences phosphate migration, Lehmann and coworkers\textsuperscript{13} monitored the intensity of the \([M - H - H_3PO_4]^-\) product ion from the isomeric peptides, IYIQSpYR and pYIYQIQSR. They observed an increased intensity of \([M - H - H_3PO_4]^-\) for the IYIQSpYR peptide. Lehmann and coworkers\textsuperscript{13} noted that phosphate migration to the C-terminus can take place when the phosphorylated residue is several amino acids away from the C-terminus. In our work, when the phosphate group is located at the N-terminal and central positions, \([M - H - H_3PO_4]^-\) is an intense product ion, suggesting migration occurs. The intensity of \([M - H - H_3PO_4]^-\) is

![Scheme 6.3. Phosphoric acid migration to C-terminus, based on Lehmann and coworkers.\textsuperscript{13} Zwitterionic precursor ion was proposed by Lehmann and coworkers.\textsuperscript{13}](image-url)
suppressed for peptides with a pTyr at the C-terminus, suggesting that the backbone flexibility may enhance migration. Phosphate migration can also be used to explain the loss of $\text{H}_3\text{PO}_4$ from pSer and pThr peptides. A better understanding of the energetics for $\text{H}_3\text{PO}_4$ migration versus elimination by a six-membered ring intermediate (Scheme 6.1) for the pSer and pThr peptides could reveal which pathway is more favorable. Both processes may be contributing to the observed intensity for $[\text{M} - \text{H} - \text{H}_3\text{PO}_4]^{-}$ in the CID spectra. As observed for pTyr peptides, the pSer and pThr peptides also have decreased intensity for $[\text{M} - \text{H} - \text{H}_3\text{PO}_4]^{-}$ when the phosphorylated residue is located near the C-terminus. This suggests that poor flexibility of the backbone inhibits the loss of $\text{H}_3\text{PO}_4$.

In CID spectra of phosphopeptides, the simultaneous observation of $[\text{M} - \text{H} - \text{H}_3\text{PO}_4]^{-}$ and backbone cleavage ions that have eliminated $\text{HPO}_3$ indicates the presence of a pTyr residue. In contrast, pSer and pThr peptides lose $\text{H}_3\text{PO}_4$ from backbone cleavage ions. The intensity of the marker ion, $[\text{M} - \text{H} - \text{H}_3\text{PO}_4]^{-}$, from pTyr peptides can indicate the position of the pTyr residue relative to the C-terminus. $^{13}$ Our experiment supports this idea since the pTyr peptides, pYAAAAA and AAPYAAA, have an intense $[\text{M} - \text{H} - \text{H}_3\text{PO}_4]^{-}$ product ion and do not contain a pTyr residue near the C-terminus.

Based on the work presented in this section, the diagnostic ions observed in the CID spectra of singly deprotonated phosphopeptides can potentially be used to indicate the type of phosphorylated residue present in an unknown peptide sample. Table 6.2 summarizes the ions and combined losses that are diagnostic for pThr, pSer, and pTyr peptides. The diagnostic ions for $[\text{M} - 2\text{H}]^{2-}$, are discussed in the next section.
Table 6.2 A summary of the diagnostic ions for specific phosphorylated residues of model phosphopeptides.

<table>
<thead>
<tr>
<th></th>
<th>$[M - H]^{-}$</th>
<th>$[M - 2H]^{2-}$</th>
</tr>
</thead>
</table>

$^a$Ions shown in parentheses may provide information about the position of the corresponding phosphorylated residue.

6.5.4 Diagnostic Ions Resulting from CID of $[M - 2H]^{2-}$ Precursor Ions

Collision-induced dissociation experiments were performed on doubly charged precursor ions, $[M - 2H]^{2-}$. Figures 6.5-6.7 (a) and (b) show pSer and pThr peptides yield two major product ions: $[M - 2H - PO_3]^{-}$ and $[M - 2H - H_2PO_4]^{-}$. Figures 6.5-6.7(c) show pTyr peptides dissociate nearly 100% to the product ion $[M - 2H - PO_3]^{-}$. Table 6.2 shows that the relative abundance of these peaks is diagnostic for specific phosphorylated residues. The relative intensities of these diagnostic product ions did not change significantly even when increasing the collision energy. The pSer peptides were the most effected by an increase in collision energy; the $[M - 2H - PO_3]^{-}$ and $[M - 2H - H_2PO_4]^{-}$ intensities became closer to equal as collision energy was increased, but this is still a ratio that is not observed for pThr- and pTyr-containing peptides. Only pSer peptides have a base peak, or a peak rivaling the intensity of the base peak, that indicates $[M - 2H - H_2PO_4]^{-}$. Additionally, fragmentation of these peptides does not provide useful information on the amino acid sequence because backbone fragmentation is negligible. It is important to note that the observed relative ion intensities in this work are reliable because multiple trials were performed on different days with reproducible results.
Figure 6.4 displays CID spectra for three doubly charged phosphopeptides on a center-of-mass collision energy scale, ranging from ~17-23 eV. In Figure 6.5(a), the pSAAAAA peptide produces \([M - 2H - H_2PO_4]^-\) in abundance, and \([M - 2H - PO_3]^-\) to a lesser extent (~33% of the base peak). Figure 6.5(b) shows the pTAAAAA peptide has an inversion of the relative intensities for those product ions, with \([M - 2H - PO_3]^-\) as the base peak, and \([M - 2H - H_2PO_4]^-\) at approximately 33% of the intensity of the base peak. The pTyr peptides yield a different spectrum entirely, dissociating almost 100% to \([M - 2H - PO_3]^-\). Table 6.2 lists these as parameters for identification of phosphorylated residues in \([M - 2H]^{2-}\).

Increased Coulombic repulsion in peptide ions of higher charge states usually results in more sequence informative fragmentation than for peptide ions of lower charge states.\(^{56, 63-65}\) However, the model phosphopeptides presented in this work show a surprising lack of sequence informative fragmentation for the doubly charged ions. The precursor ion, \([M - 2H]^{2-}\), has two possible locations for deprotonation, the phosphate group at the side chain and the carboxylic acid at the C-terminus (shown as \([M - 2H]^{2-}\) in Scheme 6.4). Multiply charged side chains are unique; for example, there are no amino acid residues that are expected to doubly protonate in the positive ion mode. If both negative charges are located on the phosphate group, peptides may not undergo characteristic backbone cleavage.\(^{66}\) Lehmann and coworkers\(^{16}\) propose a doubly deprotonated phosphorylated side chain is necessary for the elimination of \(PO_3^-\). This loss is observed for all of the model phosphopeptides in this work. Lehmann and coworkers\(^{16}\) propose a mixture of peptide deprotomers (i.e., ions with differing deprotonation sites) is formed upon CID activation of \([M - 2H]^{2-}\). To test this theory, they varied the collision voltage and found that at low-to-moderate collision activation energy, the formation
of PO$_3^-$ from a pThr-containing peptide is prevalent. They attribute the observation of PO$_3^-$ to the cleavage of an activated anion with a doubly deprotonated sidechain. Additionally, the observation of H$_2$PO$_4^-$ confirms the presence of the other deprotomer that contains separate charge sites (phosphate group and C-terminus). To determine whether the mixture of ion populations is formed by the CID or ESI processes, Lehmann and coworkers methyl esterified the C-terminus and saw no evidence of a doubly deprotonated phosphate group as fragmentation did not produce PO$_3^-$ as a product from [M – 2H]$^{2-}$. This indicates that the doubly deprotonated side chain must form via charge migration from a deprotonated carboxylic acid group. Because the deprotomers of the two ion populations have the same $m/z$, they fragment simultaneously in CID, leading to the observation of both [M – 2H – H$_2$PO$_4$]$^-\text{and [M – 2H – PO}_3$]$^-$ for pThr and pSer peptides, as suggested by Scheme 6.4. The absence of a peak corresponding to a loss of H$_2$PO$_4^-$ suggests two ion populations are not formed upon CID activation for the doubly deprotonated pTyr peptides. Side chain double deprotonation from the ESI process is likely for pTyr peptides because pTyr is the most acidic phosphorylated amino acid, as reported in Chapter 5. In addition, the larger side chain of pTyr may stabilize the negative charges better than the smaller pSer and pThr side chains. As discussed in Chapters 3 and 5, the structures, (i.e, deprotonation sites), of tyrosine and phosphotyrosine can be influenced by ESI parameters.
Scheme 6.4 Loss of 79 and 97 Da from doubly deprotonated phosphopeptides. The mechanism for loss of 79 Da from a phosphopeptide with a doubly deprotonated side chain is adapted from work by Lehmann and coworkers.\textsuperscript{16}
6.6 Conclusions

Negative ion CID was performed on nine model hexapeptides containing the common phosphorylated residues, pThr, pSer, and pTyr. For $[\text{M} - \text{H}]^-$, the common negative ion mode backbone fragments, a-, b-, c-, and y-ions, were observed. The backbone fragment ions often eliminated the phosphate group as $\text{H}_3\text{PO}_4$ or $\text{HPO}_3$. When the phosphate group is located at the N-terminus, the largest amount of sequence informative fragmentation is produced, whereas the central location generates the least fragmentation. Backbone cleavage does not preferentially yield ions that contain the phosphorylated amino acid. In fact, backbone fragment ions are usually N-terminal, regardless of the location of the phosphorylated residue. For example, when the phosphorylated amino acid is located near the C-terminus, many intense N-terminal CID product ions are formed that do not contain the phosphate group. Diagnostic ions for the phosphate group were observed at $m/z$ 79 ($\text{PO}_3^-$) and at $m/z$ 97 ($\text{H}_2\text{PO}_4^-$). The well-established neutral losses of CH$_2$O and C$_2$H$_4$O$_{16}$ accompanied by loss of the phosphate group (as H$_3$PO$_4$) can be used to identify peptides containing pSer and pThr, respectively. Simultaneous observation of $[\text{M} - \text{H} - \text{H}_3\text{PO}_4]^-$ and a-, b- and c-ions that only lose HPO$_3$ indicates the presence of a pTyr residue. In addition, the location of a pTyr residue may influence the relative intensity of $[\text{M} - \text{H} - \text{H}_3\text{PO}_4]^-$.

For $[\text{M} - 2\text{H}]^{2-}$, the relative intensities of $[\text{M} - 2\text{H} - \text{PO}_3]^-$ to $[\text{M} - 2\text{H} - \text{H}_3\text{PO}_4]^-$ (pSer, pThr), or the absence of $[\text{M} - 2\text{H} - \text{H}_2\text{PO}_4]^-$ (pTyr), can give useful information about the type of phosphorylated amino acid present in a peptide.

The work presented in this chapter discusses how the location and identity of a specific phosphorylated amino acid residue can influence observed product ions in the CID spectra of singly and doubly deprotonated phosphopeptides. This is the first reported compilation of diagnostic ions that make identification of pThr, pSer, and pTyr in phosphorylated peptides...
possible using only negative ion mode CID. The identification of diagnostic ions enhances the information obtained from CID spectra and may benefit proteomic studies by encouraging more routine sequence analysis to be performed in the negative ion mode.
Supplemental Figure 6.1 QIT CID spectra [M − H]− for (a) pTAAAAA, (b) AApTAAA, (c) AAAApTA.
Supplemental Figure 6.2 QIT CID spectra of [M – H]⁻ for (a) pSAAAAA, (b) pTAAAAA, (c) pYAAAAA.
References


CHAPTER 7: CONCLUDING REMARKS

The research in this dissertation helps to answer questions about the deprotonation of biomolecules. The site of negative charge was important for understanding thermodynamic properties of amino acids, amino acid amides, phosphorylated amino acids, and phosphorylated amino acid amides. Tyrosine was found to have deprotonation at two sites, as well as two conformers of varying energy where the C-terminal carboxylate group was deprotonated. Also of importance was how the solution-phase structure and ESI conditions affected ions present in the gas phase. In the case of the amino acid amides, the C-terminal amide group resembles a structure more closely representing a peptide backbone. It was found that six amino acid amides had acidic side chains capable of deprotonation: tyrosine, cysteine, tryptophan, histidine, and from previous work,\textsuperscript{33} aspartic and glutamic acid amides. These six amino acids should have side chains considered as potential charge sites when evaluating MS and MS/MS data. For phosphorylated amino acids and their amides, the phosphate group, a common PTM, readily deprotonated by ESI. Only in case of pTyrosine was a carboxylate anion found computationally as the lowest energy structure. In addition, studying pTyrosine also reflected that the solution and ESI conditions can change the conformation of ions in the gas phase. As most fragmentation is charge directed, knowing the location of the charge site can help to interpret product ions made in the CID process.\textsuperscript{34,35} When studying phosphorylated model peptides, the loss of the phosphate group caused charge migration to other parts of the peptide, commonly resulting in product

204
anions and a neutral phosphate loss. The location of the phosphate group, the type of amino acid residue, and the charge on the precursor significantly impacted the fragment ions observed by CID. By understanding common losses and diagnostic ions from phosphorylated peptides in the negative ion mode, more information can be built up to assist proteomic data bases identify peptides and proteins. Negative ion CID studies of phosphorylated peptides can be beneficial for sequencing. For phosphorylated peptides, understanding negative ion fragmentation is especially useful because the acidity of the phosphate group allows for the facile formation of negative precursor ions.

The examination of tyrosine gave energetic information about anions that we¹, Tian and Kass,², ³ and Oomens⁴ have studied. The ability of tyrosine to deprotonate on both its sidechain and its backbone provides an interesting charge site to consider when tyrosine is incorporated into peptides and proteins. This is the first report of multiple tyrosine anionic conformers, in contrast to isomers. The observation of three different tyrosine anionic structure is novel for a small molecule, i.e, an amino acid. In addition, ESI of tyrosine from various solvents in combination with the probe reaction of trimethylsilyl azide gave useful information about deprotonation site. This allowed for the isolation of pure carboxylate anions in later experiments. Searching for other probe reactions that could determine side chain vs C-terminal deprotonation in amino acids such as cysteine, histidine, and tryptophan (all discussed in Chapter 4 to side chain deprotonate) could be useful for determining the charge site. It is interesting to consider if certain amino acids retain solution-phase zwitterionic structures after the ESI process, or if there is some conversion between structures upon entering a mass spectrometer. Especially in the field of peptide and protein supercharging (the addition of extra protons), the gas-phase/solution-phase interface in ESI is being investigated,⁵, ⁶ and ion/molecule reactions where deprotonation sites
could be probed may assist in obtaining more information about solution-phase structures being observed in the gas phase. Tian and Kass\textsuperscript{2, 7} also studied p-hydroxylbenzoic acid and cysteine and found that deprotonation occurs on sites other than the carboxylic acid group in these systems. It would be interesting to use these compounds (or other compounds) to perform spectroscopic measurements in solution phase where data about the deprotonation sites in various solvent systems could give additional confirmation of this phenomenon. In this dissertation research, such experiments were attempted by us for tyrosine, but were not successful due to its poor solubility. In addition, ion mobility experiments may be helpful to examine the conformations and deprotonation sites of tyrosine.\textsuperscript{8, 9}

In regards to the studies of amino acid analogues (amide, phosphorylated amino acids, and phosphorylated amino acid amides) other modifications to common amino acids could provide more insight into the acidity of biological systems. For example, there are several amino acids that are prevalent in nature but are not among the 20 amino acids used as building blocks for peptides and proteins. There are many post-translational modifications (PTM) in nature that were not examined in our work such as: acetylation, methylation, and glycosylation. Another analogue example is selenocysteine, where the sulfur is replaced with selenium, a trace element found in the human body.\textsuperscript{10} If modified amino acids (and their amides) are readily obtainable, performing ion/molecule reactions with the thermokinetic method could yield GA (or GB) values for these compounds. This could be another collaboration with the Dixon group, and could provide more insight into potential side chain deprotonations in peptides and proteins. Additionally, ion mobility mass spectrometry experiments may provide further confirmation of computational data, as the cross-sectional area and drift time give information about ion conformation.
The study of phosphopeptides by negative ion mode mass spectrometry can yield useful and complimentary information not obtained in the positive ion mode. However, when performing CID experiments on negatively charged precursor ions, there are prevalent neutral losses that dominate the spectra. Though these can be useful in identifying the presence of a phosphate group, it is not obvious as to the location of the residue originally containing the phosphate group. The diagnostic ions discussed in Chapter 6 can resolve some ambiguity, but having other straightforward ions to identify phosphorylated residues would help in phosphoproteomic analysis. As a good comparison for negative ion mode CID, the positive ion mode CID should also be used to provide as much information as possible. Some comparison was done to positive ion spectra in this work; however, future experiments should take this into account. In addition, electron based techniques (ECD, ETD) retain post-translational modifications, such as phosphate groups. Because CID and ETD are complementary in the backbone cleavages they produce, positive ion mode studies should be done on any phosphorylated peptide capable of producing a doubly protonated ion. There have been several studies to date that have examined electron-based spectra of phosphopeptides. Future researchers in the Cassady group have instruments available to complete ETD or ECD experiments. In the case that a doubly protonated ion could not be obtained, another method would be to adduct metals to obtain a higher charge state. Metal-adducted peptide ions have been studied in the Cassady group and by other researchers, and should be pursued with phosphorylated peptides. Also, negative electron based techniques such as NETD (negative electron transfer dissociation), EDD (electron detachment dissociation), and niECD (negative ion electron capture dissociation) could be used to see how phosphate groups are retained. The negative electron based techniques generally fragment peptides to yield a- and x- type ions.
With the combination of CID in both negative and positive ion mode, ECD or ETD, and negative electron-based techniques, multiple ion types can be obtained and can be used for optimal validation of a peptide sequence. Also for future studies, the structures of ions generated by MS/MS can be analyzed by computational methods to give information about charge location and energetics.

By establishing how amino acids, amino acid analogues, and model peptides deprotonate and undergo gas-phase reactions and fragmentation, we can more thoroughly understand negative ion gas-phase chemistry for biomolecules. By gaining information about acidity and charge sites, negative ion mode mass spectrometry can be used to its full potential in the study of proteomics.

References


