An Experimental and Theoretical Study of Proton Affinity in Proline-Containing Dipeptides: Exploring the “Proline Effect”

Anton Luke Lachowicz
College of William and Mary

Follow this and additional works at: https://scholarworks.wm.edu/honorstheses

Part of the Analytical Chemistry Commons

Recommended Citation

This Honors Thesis is brought to you for free and open access by the Theses, Dissertations, & Master Projects at W&M ScholarWorks. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of W&M ScholarWorks. For more information, please contact scholarworks@wm.edu.
An Experimental and Theoretical Study of Proton Affinity in Proline-Containing Dipeptides: Exploring the “Proline Effect”

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Chemistry Department from The College of William and Mary

by

Anton Luke Lachowicz

Accepted for Honors

Professor John C. Poutsma, Director

Professor Tyler K. Meldrum

Professor Christopher J. Abelt

Professor David S. Armstrong

Williamsburg, VA
Abstract
Mass Spectrometry is a very useful technique for proteomics studies. Currently Bottom-up proteomics uses peptide-sequencing databases to identify peptides from fragmentation spectra. However, these databases lack information about selective fragmentation of proline containing peptides, resulting in the failure of that peptide being sequenced. The selective cleavage proline causes during low-energy dissociations in the gas-phase is known as the “proline effect.” In order to better understand the proline effect, the proton affinity of proline-containing dipeptides is obtained theoretically using B3LYP and compared to experimental values from an extended kinetic method experiment on a triple quadrupole mass spectrometer. Pro-Pro, Pro-Phe and Phe-Pro were found to have proton affinities of 990, 979 and 974 kJ/mol.
Table of Contents

Chapter 1: Introduction 1
  1.1 Proteomics 1
    1.1.1 Top-down 1
    1.1.2 Bottom-up 2
  1.2 Proline 4
  1.3 Mobile Proton Mechanism 5
  1.4 Proline effect 6
  1.5 Kinetic Method 9
  1.6 Mass Spectrometers 12
  1.7 Computational Chemistry 14
    1.7.1 Hartree-Fock 14
    1.7.2 B3LYP 15
Chapter 2: Methods 17
  2.1 Experimental 17
  2.2 Computational 19
Chapter 3: Results and Discussion 22
  3.1 Prolyl-Proline 24
  3.2 Prolyl-Phenylalanine 30
  3.3 Phenylalanyl-Proline 36
  3.4 Structural features 42
  3.5 Proton Affinities 45
  3.6 Experimental Studies 46
Chapter 4: Conclusion and future work 48
References 50
Figures, Charts and Tables

Figure 1: proline structure 4
Figure 2: ion nomenclature 5
Figure 3: mobile proton 6
Figure 4: N-bias 7
Figure 5: AAPAA spectrum 8
Figure 6: kinetic method diagram 10
Figure 7: ESI 12
Figure 8: quadrupole diagram 13
Figure 9: ionization sites 22
Table 1: starting conformers 23
Table 2: end conformers 23
Figure 10: propron 24
Figure 11: propron 2,3 25
Figure 12: proproha 25
Figure 13: proproha 2,3 26
Figure 14: proprohb 26
Figure 15: proprohb 2,3 27
Figure 16: proprohc 28
Figure 17: proprohc 2,3 28
Figure 18: prophen 29
Figure 19: prophen 29
Figure 20: prophen 2,3 30
Figure 21: prophen 30
Figure 22: prophen 2,3 31
Figure 23: prophen 31
Figure 24: prophen 2,3 32
Figure 25: prophen 32
Figure 26: prophen 2,3 33
Figure 27: prophen 33
Figure 28: prophen 2,3 34
Figure 29: phepron 34
Figure 30: phepron 2,3 35
Figure 31: phepron 35
Figure 32: phepron 2,3 36
Figure 33: phepron 36
Figure 34: phepron 2,3 37
Figure 35: phepron 37
Figure 36: phepron 2,3 38
Figure 37: phepron 38
Table 3: proton affinities 45
Table 4: kinetic method data 46
Chart 1: kinetic method plot 1 47
Chart 2: kinetic method plot 2 47
Acknowledgements

I would like to thank Dr. Poutsma and the ionlab for teaching me analytical and computational chemistry, as well as, supporting my thesis endeavors.

I would also like to thank the College of William and Mary and the Charles Center for making my honors project possible through the fellowship program.

I would like to thank the National Science Foundation for providing the funding to ionlab.
Chapter 1: Introduction

1.1 Proteomics

Proteomics is a branch of biotechnology concerned with applying the techniques of molecular biology, biochemistry, and genetics to analyzing the structure, function, and interactions of the proteins produced by the genes of a particular cell, tissue, or organism, with organizing the information in databases, and with applications of the data.¹ There are several different methods for studying proteomics; these include structural and differential proteomics and the top-down and bottom-up methods. The top-down and bottom-up methods are very commonly used for protein sequencing in gas-phase using mass spectrometers.

1.1.1 Top-down proteomics

Top-down analysis involves the analysis of intact proteins. This method is particularly useful for better understanding post-translational modifications (PTM). PTMs are modifications made to proteins after assembly, and can help determine the protein’s location in the cell and whether or not it is activated. From these data the actual function of a protein can be better understood. Traditional proteomic methods can require prior knowledge of these PTMs and of the protein on which they reside. Not only can the top-down method determine what PTMs might be on the protein, it can also identify where the protein was modified. Another
important feature of the top-down method in a mass spectrometer is the mass spectrometer’s ability to detect very small quantities of ions. This ability can be used to discover rare proteins and PTMs that may not show up in high enough quantities during traditional proteomics methods. Because other methods using mass spectrometry, namely bottom-up, involve cleaving the protein before analysis, these PTMs may be lost from chemical interactions before analysis. Another benefit to the top-down approach is its quickness since proteins do not need much preparation before hand. Because of this, the method has very high throughput for single protein analysis. However, this method comes with some problems. There is more complex chemistry required to maintain these proteins in the gas-phase, especially since many charges on a single molecule can lead to Coulombic repulsion that denatures the protein and cause subunits to dissociate. This can be remedied partly by using very high mass-range and very high-resolution instruments such as the ion cyclotron resonance (ICR) mass spectrometer or the orbitrap. These instruments are very expensive to buy and maintain, so scientists often turn to a less costly approach that yields consistent results.

1.1.2 Bottom-up proteomics

Bottom-up proteomics can be used to determine the identity, quantity or primary structure of proteins by enzymatically cleaving proteins prior to introduction to the instrument. Reverse-phase high-pressure liquid
chromatography (HPLC) is used to separate the resulting peptide chains, and electrospray ionization tandem mass spectrometry, usually in an ion trap or triple quadrupole instrument, is used to analyze the peptides.\(^5\) One of the common enzymes used to cleave proteins is trypsin, which cleaves after arginine and lysine residues unless the residue is followed by a proline.\(^6\) This helps ensure reproducible results from protein cleavage and consequent spectra. Bottom up proteomics is one of the fastest ways to identify a large-scale mixture of proteins. Using the bottom-up method with HPLC-MS/MS has the potential to identify thousands of proteins in a single run.\(^7\)-\(^8\) Like top-down approaches, it can also be used to identify the location of PTMs on a protein,\(^9\) though because of the use of identification programs, prior knowledge of the PTM is beneficial for accurate identifications. After isolation and fragmentation of peptides in the mass spectrometer, the data is run through a protein-sequencing database for identification. The database determines the peptide sequence from the fragmentation pattern and then discovers which proteins are known to contain that particular peptide.\(^10,11\) This method does run into some issues, including missing peptide fragments from ion-suppression, low ionization efficiencies, and low peptide quantities. However there are two more sources of error, poor instrumentation and poor database information.\(^12\) Improving databases (SEAQUEST, Mascot, Peptidesearch) is an easy way to improve peptide-sequencing results, since changing parameters for the other issues may help certain peptide sequences but be detrimental to the overall results. As peptides are isolated
at their mass/charge ratio (m/z) and then fragmented via collision-induced
dissociation (CID) the database simulates the fragmentation of all peptide
sequences that match the mass to charge ratio and compares the actual
spectrum to the simulated spectra.\textsuperscript{10} This method relies on the program to
accurately predict the fragmentation patterns, which includes the m/z and
relative intensities of each m/z.

1.2 Proline

![Proline structure](image)

Figure 1: Proline structure.

Proline is one of the 20 protein amino acids. It differs from all of the others
by having its R-group, which is a 3-carbon alkyl chain, bound to its N-terminus
forming a five-membered ring.\textsuperscript{13} The rigidity of the ring causes proline to have
certain characteristics, which are highlighted when it is incorporated into peptide
chains.\textsuperscript{13, 14} One of proline’s dihedral angles is locked at -65 degrees causing it to
disturb a secondary structure, such as, β-sheets.\textsuperscript{14, 15} Proline also disturbs another
secondary structure, the α-helix, by interrupting the hydrogen bonding used to
stabilize the helix.\textsuperscript{16} These disruptive characteristics carry over to the gas-phase
where it causes selective cleavage that can confound peptide identification programs.\textsuperscript{22}

1.3 Mobile Proton Mechanism

In order to better identify proteins in the gas-phase, tandem MS/MS is used to determine tryptic peptide sequences. The method for MS/MS first selects a precursor ion, which is subsequently dissociated, and the fragments are analyzed in the secondary MS. Peptides can be cleaved anywhere along the backbone. Depending on which bond is cleaved the resulting fragments are designated as a, b, c or x, y, z ions. Figure 2 shows the possible places a peptide can cleave along its backbone and which letter designations are associated with a particular bond dissociation. CID mostly creates b and y ions from the dissociation of the amide bond. In CID a ions can be formed from the loss of CO from a b ion, but are not directly formed through this mechanism. The number next to the ion type indicates how many amino acids are included in that ion.

\[ \text{Figure 2: Naming scheme for fragment ions from peptide dissociation.}\textsuperscript{17} \]
Identification from this method works best if the precursor dissociates in a reproducible way. While high-energy (100-1000V) dissociation sees fragmentation at sites not related to the position of the ionizing proton, at low energies (<100V), cleavage is dependent on the charge site of the peptide, making it a fairly predictable fragmentation pathway. The mobile proton mechanism is the mechanism by which peptide cleavage occurs at low-energy dissociation in the gas-phase. It favors the creation of b and y ions as shown below.

![Protein cleavage mechanism](image)

**Figure 3:** The mobile proton mechanism for formation of b and y ions.

The ionizing proton on a peptide chain will most likely be initially on the N-terminus or on basic side chains. When the peptide is subjected to low-energy CID in the gas-phase it obtains enough energy for the proton to be moved along the peptide chain, where it will come to rest on one of the amide carbonyls. This leads to the formation of an oxazolone ring and dissociation of the peptide bond. After the
bond dissociation the N-terminus of the y fragment may induce proton transfer from the b ion neutralizing it, and ionizing itself.\textsuperscript{22}

1.4 Proline effect

During peptide fragmentation every amino acid influences the cleavage of its adjacent amide bonds, however proline exhibits the most preference for cleavage of its N-terminus amide bond.\textsuperscript{22}

![Figure 4: N-bias peptide cleavage for amino acids.\textsuperscript{22}

Figure 4 details each amino acid’s likelihood to cleave on the N-terminus, and whether that cleavage tends to form a y ion or b ion. The 25\textsuperscript{th} and 75\textsuperscript{th} percentiles for each amino acid are shown as the top and bottom of the bar the middle line is the mean bias. A high score indicates preferential cleavage on the N-terminus side, while a low number indicates preferential cleavage of the C-terminus peptide bond.
Zero represents no preference. For both ion types proline is heavily N-biased. The “proline effect” is the term used to describe the selective cleavage proline exhibits on peptide fragmentation. Not only does proline show a bias for N-terminal cleavage, depending on the other amino acids in the peptide, it can cause the dominant cleavage in the mass spectrum.

Figure 5: Fragmentation spectrum for AAPAA. The \( y_3 \) ion, from AAPAA shown in Figure 5, is the ion formed from the cleavage of the peptide bond directly to the N-terminus of the proline residue. It is the dominant ion from this peptide, while \( b_3 \) is barely visible and there are no peaks for \( b_2 \), \( y_2 \) or \( y_4 \). The N-terminus of proline is very basic and has been shown by theoretical calculations to stabilize the protonation of proline’s N-terminus peptide carbonyl. However, studies have been done using proline analogs, which have similar gas-phase basicities, and some of the analogs created ‘b’ ions and some ‘y’ ions. The analogs were azetidine-2-carboxilic acid (Aze, four-membered ring), proline (Pro, five-membered ring), and pipecolic acid (Pip, six-membered ring). Proline was also compared to N-methylalanine (NMe). These have proton affinities (PAs) of 933, 937,
944$^{25}$ and 931$^{26}$ respectively. They were substituted into a pentapeptide chain such that the sequence was AAXAA, where A is alanine and X is either Aze, Pro, Pip or NMe. The spectra for Aze and Pro show high preference for cleavage at the N-terminal and proton transfer to form $y_3$ ions while Pip and NMe preferentially cleave at their C-terminal peptide bonds to form $b_3$ ions. All analytes exhibit $b_4$ ion creation or cleavage before the last alanine. Also none of the analytes showed appreciable amounts of complementary ions ($b_2$ and $y_3$, $y_2$ and $b_3$ etc) making it seem that cleavage would occur at the mobile proton site and then one of the two possible product ions at that site would be highly favored. Substitution of the alanines with other alkyl chain residues had little to no effect on fragmentation $^{20}$. However substitution of the amino acid on the N-terminal side of proline with basic amino acids will increase the amount of $b$ ions created from cleavage at the N-terminal amide bond of proline. Dominant cleavage of that bond is subject to the composition of the peptide.$^{20,23}$ These studies suggest that something in the structure of proline and its analogs create the difference in the cleavage site type of ion created, rather than just the basicity.

1.5 The kinetic method

In the gas-phase there are several techniques that can be used to determine thermochemical information. Equilibrium, bracketing and the kinetic method are the three most popular methods. The kinetic method requires generation of a proton-bound heterodimer. On one side of the dimer is the analyte and the other half is a reference base, of known basicity. The dimer ion is subjected to collision-
induced dissociation and the relative amounts of the two product ions depend on their basicity.\textsuperscript{27} This method only works if a couple of assumptions are true, namely that there are no reverse activation energy barriers and there is little or no entropy difference between the reference and analyte. The heterodimer is introduced to the gas phase where it is mass selected and then undergoes CID and competitively dissociates to two sets of products, as shown in figure 6. The abundance of protonated base and analyte are measured and the ratio can be used to determine the difference in gas-phase properties.

\[ \text{B}_1\text{H}^+ + \text{B}_2 \quad (k_1) \]

\[ \text{B}_2\text{H}^+ + \text{B}_1 \quad (k_2) \]

\[ \Delta \epsilon_0 \]

\[ \epsilon_0(1) \quad \epsilon_0(2) \]

\[ \text{B}_1\text{H}^+ \text{B}_2 \]

\[ \text{B}_1 + \text{B}_2\text{H}^+ \quad (k_3) \]

Figure 6: Schematic representation of the kinetic method for thermochemical determinations.\textsuperscript{28}

The use of similarly structured species for the references can help simplify these determinations, because the entropy should be very similar leaving the
difference in ion concentration entirely up to the enthalpy differences, or difference in proton affinity ($\Delta(\Delta H)$) between the reference base and the analyte. Since there are no entropy effects $\Delta G = \Delta H - T\Delta S$ will simplify to $\Delta G_{\text{analyte}} = \Delta H_{\text{analyte}}$ and $\Delta H_{\text{analyte}} = \Delta H_{\text{reference}} + \Delta(\Delta H)$. However, modifications to this method have been made to acquire more accurate results. Performing CID on the heterodimers at different collision energies will yield data that can determine the entropy and enthalpy. At each energy level of CID there is a ratio of reference to analyte for each reference ($\ln(I_B/I_A)$) where I is the intensity of ions in the mass spectrometer, and A and B refer to analyte and reference base, respectively. The ratio of reference base to analyte in this case will be proportional to $\Delta G_{\text{apparent}}/T_{\text{eff}}$. These ratios are plotted against the difference in proton affinity between each reference and the average of all the references ($\Delta H_B - \Delta H_{\text{average}}$). The slope of the line this plot creates is proportional to $1/T_{\text{eff}}$ where $T_{\text{eff}}$ is the effective temperature, which is an indication of the internal energy of the heterodimer after excitation by CID. Each collision energy will have a line and all of the lines must cross at a single isothermal point. This point can be used to determine the enthalpy and entropy of the reaction. The lines from the raw data rarely cross at a single point due to spread in the experimental data. Consequently, the crossing point is determined via orthogonal distance regression (ODR) methods. ODR optimizes the shifting to keep the lines at their best fit to the raw data, while still creating the isothermal point. Alternatively, the $T_{\text{eff}}$ of each collision energy can be used to calculate $\Delta G_{\text{apparent}}$ for each collision energy. A plot of $T_{\text{eff}}$ vs $\Delta G_{\text{apparent}}$ will yield a line with a slope is $\Delta (\Delta H_A)$ and the intercept is $-T_{\text{eff}} \Delta(\Delta S)$. It is also worth noting that instead of increasing the collision
energy, the pressure of the collision gas can also be increased imparting more energy through more collisions instead of more energy per collision or the collision gas itself can be changed to a one of higher molecular or atomic weight to impart more energy per collision without increasing the total kinetic energy. The addition of the entropy calculation earned this technique the name the extended kinetic method.²⁸, ³⁰, ³¹

1.6 Mass Spectrometers

A Thermo TSQ triple quadrupole mass spectrometer equipped with an electrospray ionization (ESI) source was used to collect the experimental data used in this thesis. This study is done on biomolecules, which are not particularly volatile and do not ordinarily carry charge in the gas-phase.

![Diagram of ESI](https://example.com/esi_diagram.png)

Figure 7: Diagram of ESI.

ESI is a technique commonly used to accomplish the goal of moving a biomolecule from solution and introducing it to the gas-phase with one or more charges. This is accomplished by flowing the solution through a capillary needle with an applied voltage of usually 4-5 kV. This causes the charged droplets to
nebulize. Often nitrogen gas is used as a drying gas to help desolvation along with a heated capillary. As the solvent evaporates the charged particles gather on its surface until the Rayleigh limit (the maximum amount of charge a droplet can carry) is reached, at which time Coulombic fission occurs and the droplets split into smaller charged droplets, as shown in Figure 7. This is continued until the analytes are completely desolvated.\textsuperscript{32,33}

![Figure 8: Quadrupole diagram](image)

This mass spectrometer uses quadrupoles to mass select ions of interest. A quadrupole is simply four parallel rods set up with two on an x-axis and two on a y-axis. Rods on each axis are given an alternating direct current (DC) or radio frequency (RF) potential, which are tuned to create a resonance for a specific m/z ratio. Figure 8 shows a quadrupole with applied DC and RF voltage guiding ions through the cell and causing ions of different m/z ratios out to be pushed out. In a triple quadrupole, three quadrupoles are set in a row with the first (Q1) and third (Q3) set up for m/z selection, which allows tandem mass selection where the first
and third can select for different m/z ratios. The second quadrupole (Q2) is an RF only ion guide that is used as a collision cell, a neutral gas, often argon, is leaked into this cell to induce CID. There are several scan types for MS/MS, but for this experiment only one is used, the product scan. This uses Q1 to select an ion, which undergoes CID in Q2 followed by a m/z scan for products in Q3. Finally after Q3 ions are attracted to a charged dynode that leads to detection via an electron multiplier.

1.7 Computational Chemistry

1.7.1 Hartree-Fock

The Hartree-Fock calculation for approximating atomic and molecular structures, allows for a relatively simple calculation of a complex problem. After the discovery of the Schrödinger equation, the behavior of small particles could be accurately predicted. This allowed scientists to start modeling atomic and molecular systems. Unfortunately, they ran into a problem after trying to calculate systems with more than two, interacting particles. The Hartree-Fock equation makes several approximations, one of which is the Born-Oppenheimer approximation, which separates the nuclei and electrons into separate sets of calculations. This approximation works very well because the mass difference between the nuclei and electrons is so great that the positive nuclei rearrange at a much slower rate than electrons. Another approximation is the mean field theory or the self-consistent field approximation, which determines the charge density of a system and then isolates each electron and calculates its position based on the charge density that
The calculations for this experiment used restricted Hartree-Fock (RHF), which is used for closed shell systems, because it assigns electrons to the lowest energy orbital with alternating spins.\textsuperscript{36,37} While RHF can calculate an exact electron exchange, it still falls short of accurately predicting the energy of a system. One of the biggest problems of the Hartree-Fock method is its failure to take into account the fact that the motion of an electron is affected by, or correlates to, the motion of all other electrons in a system, which led to the origins of hybrid density functional theory (DFT). Ultimately, the difference between the energy determined by the Hartree-Fock method and the true energy of a system came to be known as correlation energy, even though this difference does not stem entirely from electron correlation. It actually includes error generated from all of the approximations used in the method.\textsuperscript{37}

1.7.2 B3LYP and density functional theory

The B3LYP method takes the molecular geometry calculation a few steps further. B3LYP is one of the most popular functional combinations and gives very accurate approximations for organic molecular systems. While DFT and HF contain a lot of the same theory, DFT has some differences from HF. The energy density is determined by Kohn-Sham (KS) equations, which effects the potential and Coulombic energy of a system. Also, when using DFT based on KS, the electron exchange energy can no longer be exactly determined.\textsuperscript{35,37} This led Becke to search for an approximation for this value. He concluded that the most accurate approximation was the gradient combination of Local Spin Density Approximation.
(LSDA), the exact exchange from the Hartree-Fock method and his own gradient correction to LSDA. Originally, he used the 1991 Perdew-Wang approximation for electron correlation, however the Lee, Yang, Parr (LYP) functional for electron correlation has often been used for that approximation. B3LYP uses Gaussian functions to determine molecular orbitals, which are denoted as 6-31+G* where 6 is the basis set of six Gaussian functions that are used to determine the inner shell orbitals while the numbers after the dash represent the valence orbitals which are split into two, the inner valence with 3 Gaussians and the outer with 1. The * represents adding a d orbital function to atoms other than hydrogen, allowing for more polarizability. Adding a second * denotes adding a p orbital function to hydrogen. Other diffuse functions are represented by the + which allows for greater electron diffusion.
Chapter 2: Methods

2.1 Experimental methods

Kathy Huynh, a previous member of the lab studied proline-containing dipeptides using the Thermo TSQ triple quadrupole mass spectrometer. The dipeptides were bought premade and then dissolved in a 50% methanol/50% water solution which was then diluted to $10^{-4}$ M. Reference bases were selected from NIST, solvated and diluted in a similar manner. 1:1 Aliquots of the solutions were then mixed together and formic acid was added to make up 1% of the total solution to make conditions favorable for heterodimer formation. The solutions were then run through the TSQ at 7-10 μL/min and the reference-dipeptide heterodimer was mass selected and fragmented at increasing collision energy intervals of 3 V from 0-30 V. Three separate days of data were taken per each heterodimer.\textsuperscript{41}

Prolyl-phenylalanine (prophe) was studied in a new Thermo TSQ triple quadrupole mass spectrometer for this experiment. The methods were very similar to the old experiment. The reference bases were the same as the previous experiment pyrrolidine, piperidine, 1-methylimidizole, 2,4-lutidine and diisopropylamine, which have proton affinities of 948.3 954 959.6 962.9 and 971.9 kJ/mol respectively. Each of the references and prophe were measured out to either $10^{-4}$ or $10^{-5}$ mol then solvated in 10 mL of 50:50 methanol water to make $10^{-2}$ or $10^{-3}$ M solutions. These were the stock solutions and aliquots were diluted to $10^{-4}$ M for mixtures. The dimer solutions were a 1:1 ratio of reference to analyte, usually 1mL.
of each, which resulted in an end concentration of $5 \times 10^{-5}$ that was run through the triple quadrupole. Formic acid was added to increase dimerization at a concentration of 1% of total volume. Each reference heterodimer was run on three separate days in order to reduce statistical error and so that residual ions did not interfere with another data set. For each run I used 300 uL in a 500 uL gas-tight syringe at a flow rate of 5 uL/min and a capillary temperature of 373 K. The run would consist of taking data at collision energies 0-30 V in intervals of 3 using argon as the collision gas. The data was acquired via scans of m/z 10-400 over a 10 second interval. After each run a flush solution was run through the triple quadrupole to try and eliminate residual ions. Because the higher collision energies can cause secondary fragmentation of the analyte, each of the references were fragmented at a collision energy of 30 and the fragments were documented. After all the data was collected the average signal over the 10 seconds of scan time was analyzed for reference and analyte fragments. a ratio of IB/IA was established where IB is the intensity of the reference base and all of its fragments and IA is the intensity of the analyte and all of its fragments, (fragments were only added if the ion count was high enough to not be considered noise). All of the ratios for every reference, at each collision energy were averaged over all three days, The natural log of each average ratio for each reference base was plotted versus the proton affinity of the reference minus the average proton affinity of all the references. This gives a chart with a line for each of the collision energies and every point in the line is a reference base. The second chart uses data from the first, It plots the slope of the lines vs the negative intercepts which is equal to $1/RT_{eff}$ vs $(H_A-H_{avg})-T_{eff}\Delta S/RT_{eff}$ The slope of this
graph is the difference between the proton affinity of the analyte and the average proton affinities of the reference bases.

2.2 Computational methods

The initial step in running calculations was to create a series of structures on which theoretical calculations can be run. These structures were created using PCmodel. Each dipeptide was drawn first by using the amino acid creator, which allows the user to create a chain of amino acids connected by peptide bonds. The neutral was created this way and then the ions were protonated manually by adding hydrogen atoms onto each of the sites of interest. These sites were the N-terminus, the oxygen of the amide carbonyl, and the nitrogen of the amide and the oxygen of the carbonyl in the carboxylic acid of the C-terminus. These structures were then energetically minimized and a GMMX force field search was used with 15 kcal/mol for the first cycle and 10 kcal/mole for the second cycle. These searches were done with all rings included as well as all rotatable bonds including amides and esters. The search was set for the generation of 30,000 random conformers. The first cycle of the search would start at a randomly generated conformer, and will keep all structures within 15 kcal/mol of the starting structure. Identical conformers are eliminated. The second cycle takes the conformers generated from the first cycle, re-orders them, and then only accepts conformers within 10 kcal/mol of the lowest-energy conformer found in cycle 1.

After the search generated structures, they were then uploaded to a server equipped with a fortran compiler. For the first two molecules prolyl-proline
(propro) and prolyl-phenylalanine (prophe) that we studied, we were limited to 99 structures per site (neutral, a, b, c, and d), so the lowest 99 energies were chosen if more than 100 structures were created. We modified our input file generator to accept up to 999 structures for Phenylalanyl-proline (phepro) so all generated structures were examined.

The computations were run on five different servers over the period of 9 months. These servers were all installed with Gaussian 09. At first only 3 servers were at William and Mary; two additional servers were added in January 2016. The new servers made running calculations much quicker as they each had four quad processors. The computations were run at several levels of complexity with the first level being a restricted Hartree-Fock at 3-21G with geometry optimization, the second level is B3LYP at 3-21G geometry optimization, level 3 is B3LYP at 6-31+G* geometry optimization, level 4 is a vibrational frequency at B3LYP 6-31+G* and finally a single point energy at B3LYP 6-311++G** is run. At levels one, two and three energies were compared to each other and any structures that were equivalent in energy to other structures were removed from further calculations. After the single-point energy level calculations the structures were examined in Gaussview to assess whether they had isomerized. For protonated structures this examination also determined if the proton stayed on its initial ionization site or if it jumped to one of the other sites. Often times the protons would move which led to the energetics data for those structures being transferred to be with the other data for the corresponding end site.
The first step of the data workup was adding the thermal correction from the vibrational frequency calculation to the single point energy, which brings the single point conformers to 298 K. The sum was called $H_{298}$. Then the difference in energy of each conformer from the lowest conformer in kJ/mol is established which can be used in the Boltzmann equation (the gas constant in kJ/mol K was used instead of the Boltzmann constant since we had already converted to kJ/mol) to create a Boltzmann distribution. The population of each conformer was then determined as a fraction of the whole population. All of the conformers’ corrected energies and population fractions were used to determine the Boltzmann weighted energy of a particular protonation site and the neutral molecule. A parallel calculation was carried out in which the single point energy was corrected for Gibbs free energy at 298 K. The steps are the same until the end when the Gibbs corrected energies were multiplied by the Gibbs population fractions, but they were also used with the enthalpy population fractions to Gibbs and enthalpy-weighted Boltzmann energies.

Next the all of the ions were put together, and their population fractions were determined, as were the three weighted energies. These were used to determine the proton affinity by subtracting the ion Boltzmann-weighted energies from the neutral Boltzmann-weighted energies. The proton affinity was then converted into kcal/mol where a correction of 1.48 kcal/mol was added for the proton $^4$ and then the sum was converted to kJ/mol. Raw proton affinities for each site were determined by subtracting the lowest energy conformer of the neutral by the lowest energy conformer of the protonated site.
Chapter 3: Results and Discussion

In PCmodel and Gaussview oxygen atoms are represented as red and nitrogen atoms are represented as blue. In PCmodel only carbons are cyan and hydrogens are gray. In Gaussview carbons are gray and hydrogens are off-white. For the ionized molecules the added proton is identified by an arrow.

Figure 9: labeling scheme for dipeptides

Figure 9 shows prolyl-proline whereas, the other two dipeptides studied had a phenylalanine substituted for each proline separately. Phenylalanine has a different structure and a free N-terminus but the protonation sites are the same.
<table>
<thead>
<tr>
<th>site</th>
<th>propro</th>
<th>prophe</th>
<th>phepro</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>99</td>
<td>99</td>
<td>507</td>
</tr>
<tr>
<td>ha</td>
<td>94</td>
<td>99</td>
<td>183</td>
</tr>
<tr>
<td>hb</td>
<td>99</td>
<td>99</td>
<td>420</td>
</tr>
<tr>
<td>hc</td>
<td>56</td>
<td>99</td>
<td>188</td>
</tr>
<tr>
<td>hd</td>
<td>99</td>
<td>56</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 1: Starting conformers (2234 total).

<table>
<thead>
<tr>
<th>site</th>
<th>propro</th>
<th>prophe</th>
<th>phepro</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>54</td>
<td>30</td>
<td>154</td>
</tr>
<tr>
<td>ha</td>
<td>68</td>
<td>77</td>
<td>112</td>
</tr>
<tr>
<td>hb</td>
<td>40</td>
<td>47</td>
<td>54</td>
</tr>
<tr>
<td>hc</td>
<td>35</td>
<td>16</td>
<td>51</td>
</tr>
<tr>
<td>hd</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Ending conformers (734 total).

Tables 1 and 2 show the number of conformers before the start of calculations and after the completion of calculations. The starting conformers are the ones produced by the GMMX search, many of which get eliminated after the first level of calculations. The end conformers have gone through the entire set of calculations and are sorted by conformers that end in that protonation site regardless of where they started.

In the description of the following figures there is a Boltzmann population (Bpop) for each conformer, this will help describe the relative energy differences between conformers and the relevance of each conformer in the Boltzmann weighted proton affinity. The “raw score” proton affinity (PA) is also present for the protonated lowest energy conformers. This value is calculated by subtracting from the lowest energy neutral. All theoretical values have an associated error of ± 10 kJ/mol.
3.1 Prolyl-Proline

Figure 10: a) PCModel drawing of initial conformation of propro b) neutral propro lowest energy conformer (LEC) Boltzmann population (Bpop) = 45.9%.

Figure 10 shows the initial structure for propro neutral, as created by PCModel, which has a hydrogen bond between the N-terminus and the amide carbonyl, as well as the proline rings being twisted up to avoid steric strain. The final geometry (Figure 10b) for the lowest energy conformer shows that not much has changed from the original PCModel conformer other than hydrogen bonding is definitely
occurring between the C-terminus and the amide carbonyl. Figure 11 shows the second and third lowest energy conformers for propro, as compared to the LEC.

![Figure 11: Neutral propro 2nd (left) and 3rd (right) lowest energy conformers Bpop = 31.1, 8.8% (for the 2nd and 3rd lowest energy conformers, respectively).](image)

Figure 11 shows the prolyl residue is hinging farther away from the amide carbonyl and lengthening the hydrogen bond between them. While this may account for the increase in energy, there also seems to be slightly different conformations in the N-terminus prolyl ring for each conformer that will also contribute to the energy differences.

![Figure 12: Proproha lowest energy conformer Bpop = 25.1% proton affinity (PA) = 988 kJ/mol.](image)
The lowest energy conformer for proproha (Figure 12) has the ionizing proton at the N-terminus forming a hydrogen bond with the oxygen of the amide carbonyl.

Figure 13: Proproha 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers Bpop = 15.3, 10.4\% PA= 987, 986 kJ/mol.

The 2\textsuperscript{nd} and 3\textsuperscript{rd} conformers (Figure 13) show the same hydrogen bond, these conformers are very similar to the LEC however 2 has a different envelope structure in the N-terminal proline residue and 3 has its C-terminal prolyl residue is flipped (cis amide).

Figure: 14 Proprohb lowest energy conformer Bpop= 1.00\% PA= 971 kJ/mol.
For isomer "b" the ionizing proton is on the amide carbonyl forming a hydrogen bond with the N-Terminus similar to proproha. This is allowing the resonance structure of the amide, which involves the double bond from the carbon alternating between the oxygen and nitrogen. The C-terminus carboxylic acid and the prolyl residue are perpendicular to each other reducing steric strain.

![Figure 15: Proprohb 2nd (left) and 3rd (right) lowest energy conformers](image)

| Bpop = 0.00, 0.00% PA = 923, 922 kJ/mol. |

The 2nd and 3rd conformers have rotated N-terminus prolyl residues, which caused them to form hydrogen bonds between the ionizing proton and the C-terminal carbonyl oxygen instead of the N-terminus. This conformation led to a difference in energy so great that the Boltzmann distribution shows the LEC to be essentially 100% (actually 99.999976%) of the population, meaning none of the other conformers contribute statistically relevant energies to the PA of the “b” site. For the “a” site only about 50% of the population is covered by the 3 lowest energy conformers, though they only differ by PA by 2kJ/mol, while the LEC of the “b” site differs from the next lowest energy conformer by almost 50 kJ/mol.
Figure 16: Proprohc lowest energy conformer Bpop = 33.4% PA = 894 kJ/mol.

In the lowest energy conformer of proprohc (figure 16) the ionizing proton is on the nitrogen of the amide bond and is hydrogen bonding with the N-terminus forcing the rings to be close to each other, which will raise the energy of this structure, both the “a” and “b” site conformers have geometries in which the rings are farther apart reducing interactions.
Figure 17: Proprohc 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers 
Bpop = 32.6, 17.6\% PA = 894, 893 kJ/mol.

Figure 17 shows the energy differences from the LEC seem to stem from the slightly different conformations of both prolyl residue rings, these conformations also have an effect on the hydrogen bond strength between the N-terminus and amide nitrogen.

![Figure 17: Proprohc 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers.](image)

Figure 18: Proprohd lowest energy conformer.

Figure 18 shows the starting structure for a “d” isomer and the resulting optimized geometry. The proton that was added to the C-termin\textsuperscript{al} carbonyl did not stay there for any of the conformers. The amide carbonyl’s basicity is much higher and has pulled the hydrogen closer to it than the C-terminus carbonyl, though it still remains hydrogen bonded to the C-terminus. The movement of the proton occurred during the first step of the calculations, likely due to the amide carbonyl having a much greater basicity than the C-terminus.
3.2 Prolyl-Phenylalanine

Figure 19: Neutral prophe lowest energy conformer Bpop = 22.0%.

The lowest energy neutral conformer, shown above, has the hydrogen on the nitrogen of the amide hydrogen bonded to both the N-terminus and the C-terminus carbonyl.

Figure 20: Neutral prophe 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers Bpop = 15.3 12.9%.
While all 3 exhibit the same hydrogen-bonding scheme there are different prolyl ring conformations for each one as shown in figure 20. Also, the phenylalanyl residue has been rotated away from the N-terminus in conformer 3.

Figure 21: Propheha lowest energy conformer Bpop = 22.1% PA = 980 kJ/mol.

Figure 21 shows the ionizing proton on the N-terminus is forming a hydrogen bond with the amide. This structural motif is common in amino acids without basic side chains.

Figure 22: Propheha 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers.
Bpop = 8.33 3.32% PA = 978, 978 kJ/mol.

All three of the LECs have very similar structure with hydrogen bonding between the N-terminus and amide carbonyl as shown in figure 22. There are changes in the proline envelope and rotation of the phenylalanyl side chain.

![Protein Structure](image)

Figure 23: Propehb lowest energy conformer Bpop = 37.7% PA = 949 kJ/mol.

Figure 23 shows that the ionizing proton on the amide carbonyl is hydrogen bonding with the C-terminus carbonyl, while the hydrogen on the nitrogen of the amide is hydrogen bonding with the N-terminus. The prolyl residue has been rotated about 180° from the “a” site conformations making it so no hydrogen bonding happens between the amide carbonyl and N-terminal.
Figure 24: Prophehb 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers Bpop = 20.2 20.2\% PA = 948, 948 kJ/mol.

The 2\textsuperscript{nd} and 3\textsuperscript{rd} are extremely close in structure and energy and have minor differences from the LEC, such as different conformations for the proline envelope and the phenylalanyl residue moving away from the prolyl residue, as shown in figure 24.

Figure 25: Prophehc lowest energy conformer Bpop = 52.9\% PA = 887 kJ/mol.
The “c” site conformer, shown above, has a hydrogen bond from the amide nitrogen to the N-terminus. This is very similar to the “b” site conformers, except there is no proton to form a hydrogen bond between the amide carbonyl and C-terminus, which will lower the proton affinity comparatively.

Figure 26: Prophec 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers $B_{pop} = 13.1$, 11.8\% $PA = 883, 883 \text{kJ/mol}$.

Figure 26 shows the 2\textsuperscript{nd} and 3\textsuperscript{rd} lowest energy conformers have slightly different prolyl ring conformations from the LEC as well as slight changes in the conformation of the peptide backbone and rotation of the phenylalanyl residue.
Figure 27: Prophed lowest energy conformer, Bpop = 50.7% PA = 889 kJ/mol.

In this case, the C-terminus carbonyl actually managed to retain its ionizing proton, the ionizing hydrogen bonded to the C-terminus has moved between the oxygens of the carboxylic acid and has hydrogen bonded to the other oxygen, forcing the other hydrogen to the far side, as shown in figure 27. The hydrogen on the nitrogen of the amide is hydrogen bonding to the N-terminus, much like in the “c” site conformers. Unlike the “b” site conformers a hydrogen bond does not form between the C-terminus and amide carbonyl.
While the 2\textsuperscript{nd} conformer is very close to the LEC the 3\textsuperscript{rd} conformer, shown in figure 28, veers off with a different prolyl ring conformation and a significant rotation of the phenylalanyl group closer to the oxygens of the C-terminus.

### 3.3 Phenylalanyl-Proline

Figure 29: Neutral phepro lowest energy conformer pop = 30.8%.
Figure 29 shows that the lowest energy conformer of the neutral phepro has a hydrogen bond from the hydrogen of the C-terminus and the hydrogen of the N-terminus both to the amide carbonyl. The proline ring is perpendicular to the phenyl ring reducing steric strain.

Figure 30: Neutral phepro 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers pop =17.6, 10.5%.

Figure 30 shows the N-terminus of the 2\textsuperscript{nd} conformer has been rotated compared to the LEC, which, brings the N-terminal lone pair closer to the amide carbonyl increasing the energy of the molecule. The 3\textsuperscript{rd} conformer has a rotated phenylalanyl residue.
Figure 31: Pheproha lowest energy conformer pop = 34.8% PA = 971 kJ/mol.

Figure 31 shows the ionizing proton forming a hydrogen bond with the carbonyl of the C-terminus, while another proton from the N-terminal is hydrogen bonding to the amide carbonyl. Also, the phenyl and prolyl rings are on opposite sides of the molecule, not interacting with each other.

Figure 32: Pheproha 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers pop = 23.3, 23.3% PA = 970, 970 kJ/mol.

All three of the lowest energy conformers are extremely similar in structure and carry all of the same motifs, as shown in figure 32. They differ by slightly different proline envelope conformations.
Figure 33: Pheprohb lowest energy conformer pop=35.8% PA = 928 kJ/mol.

Figure 33 shows that this conformer does not have hydrogen bonding, but most groups are far away from each other with the C-terminus coming out of the page, while the N-terminus is going into the page and the prolyl ring is above the phenyl ring. These conformers are quite similar in structure to the “a” site conformers but the phenylalanyl residue has been rotated slightly moving the N-terminus away from the amide carbonyl and not allowing the hydrogen bond that shows up in the “a” conformers.
Figure 34: Pheprohb 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers pop =18.2, 11.4\% PA = 926, 925 kJ/mol.

Figure 34 shows that the 2\textsuperscript{nd} conformer is very similar to the first with differences in the prolyl ring conformations. However, the 3\textsuperscript{rd} conformer is was actually originally protonated on the C-terminus. In order for the proton to jump from the C-terminus to the amide carbonyl the prolyl residue had to be flipped from the 1\textsuperscript{st} and 2\textsuperscript{nd} conformers.

Figure 35: Pheprohc lowest energy conformer pop=58.7\% PA = 894 kJ/mol.

The ionizing proton on the amide nitrogen is hydrogen bonding to the N-terminus, as shown in figure 35. (the amide carbonyl is parallel to the page pointing to the upper left, while the C-terminus is going into the page)
Figure 36: Pheprohc 2\textsuperscript{nd} (left) and 3\textsuperscript{rd} (right) lowest energy conformers Bpop = 38.6, 1.1\% PA = 893, 884 kJ/mol.

The differences in all three conformers, shown in figure 36, stem from variations in the prolyl ring conformations and which way the phenylalanyl residue is rotated. These structures are quite similar to the LEC of the “b” site, geometrically. However, the “b” site did not have an extra proton that could be used to hydrogen bond between the amide nitrogen and N-terminus.
Figure 37: Phiprohd lowest energy conformer.

Figure 37 shows that the “d” site conformers failed to retain the ionizing proton, which has jumped over to the amide carbonyl. However, it is still hydrogen bonded to the C-terminus. The proton transfer happened during the first step of calculations, indicating the amide carbonyl is significantly more basic than the C-terminus.

3.4 Structural features

The neutral conformers of each of these series all exhibit a hydrogen bonding motif where a proton on the N-terminus and a proton on the C-terminus both hydrogen bond to the amide carbonyl.

The N-terminal protonated conformers all had significant hydrogen bonding to the amide carbonyl. Phipro also seemed to be hydrogen bonding to the C-terminus though the other two had no interactions between N and C–termini. There were many conformers that started protonated at the amide, both carbonyl and nitrogen that lost their proton to the N-terminus.

The protonated amide carbonyl conformers have different behavior for all three dipeptides. In propro the amide carbonyl seems to hydrogen bond with either the N-terminus or C-terminus while the prophe conformers have the amide carbonyl hydrogen bonding with the C-terminus they also have the N-terminus hydrogen bonded to the amide nitrogen. The phipro conformers do not have any hydrogen bonding except for the 3rd lowest energy conformer, which is hydrogen bonded to the C-terminus, but that is because it was originally protonated on the C-
terminus and isomerized. This may explain why phepro has the lowest amide carbonyl proton affinity (928 kJ/mol vs 949 and 971 kJ/mol).

The amide nitrogen protonated conformers show hydrogen bonding to the N-terminus in all three dipeptides. This is fairly predictable since the N-terminus is often close by even for the other protonated sites while the amide structure will keep the proton away from both the oxygen of the amide carbonyl and the C-terminus.

The protonated C-terminals all showed strong hydrogen bonding to the amide carbonyl, to the point where the proton transferred to the amide carbonyl and was actually only hydrogen bonded to the C-terminus. Some of the proph and prophe protonated C-terminus conformers were the only ones to successfully keep the proton, (but many of the conformers did transfer to the amide carbonyl) and had hydrogen bonding occurring between the N-terminus and the amide nitrogen. This bonding may have reduced the amide basicity, by slightly increasing the charge. This bond was also on the opposite side of the peptide backbone than the carbonyls, if the transfer requires a certain proximity between the carbonyls a bond on the other side of the backbone may have made the molecule too rigid to allow them to be that close. Also both proph and prophe LECs with protonated C-termini had N-termini that were fairly close to the amide nitrogen, but the hydrogens on these atoms did not form hydrogen bonds between the two.

The amino acid proline has been shown in both neutral and protonated states to have hydrogen bonding between the N-terminus and C-terminal carbonyl. However this study only protonated the N-terminus and not the C-terminus.43 A
previously mentioned study by Paizs determined that amide nitrogens on both the N and C termini are about 10-20 kcal/mol higher in energy (depending on conformation) than the N-terminal amide carbonyl. Also, the C-terminal amide carbonyl is 5-10 kcal/mol higher than an N-terminal amide carbonyl. These studies agree fairly well with the data collected here. All of the neutral and the protonated N-terminus propro showed the N-terminus to carbonyl hydrogen bonding seen in the amino acid proline and the amide nitrogens all have significantly lower proton affinities than do the amide carbonyls.

3.5 Proton Affinities

<table>
<thead>
<tr>
<th>Protonated site</th>
<th>propro PA kJ/mol</th>
<th>prophe PA kJ/mol</th>
<th>phepro PA kJ/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-terminus (ha)</td>
<td>988</td>
<td>980</td>
<td>971</td>
</tr>
<tr>
<td>Amide carbonyl (hb)</td>
<td>971</td>
<td>949</td>
<td>928</td>
</tr>
<tr>
<td>Amide nitrogen (hc)</td>
<td>894</td>
<td>887</td>
<td>894</td>
</tr>
<tr>
<td>C-terminus (hd)</td>
<td>---</td>
<td>889</td>
<td>---</td>
</tr>
<tr>
<td>Boltzmann</td>
<td>990</td>
<td>979</td>
<td>974</td>
</tr>
<tr>
<td>Experimental</td>
<td>---</td>
<td>975.1 ±16.0</td>
<td>968.2 ±10.8</td>
</tr>
</tbody>
</table>

Table 3: Theoretical data from prolyl-proline, prolyl-phenylalanine and phenylalanyl-proline with a raw PA for each protonated site (if possible) and the Boltzmann weighted average for PA. (error ≈ ±10 kJ/mol for all theoretical values)

The proton affinities of phepro and prophe were previously determined experimentally in our lab by a past master’s student. Phepro was measured at 968.2 kJ/mol with an error range of 10.8 kJ/mol and prophe was measured to be 975.1 kJ/mol with an error range of 16.0 kJ/mol. The theoretical calculations for phepro and prophe were 979 and 974 kJ/mol respectively and are very close to the experimental values with less than 7 kJ/mol differences.
The mobile proton method requires an energetic proton to be stabilized on a peptide carbonyl in order to induce fragmentation.\textsuperscript{21} Therefore, the b site raw score and the overall proton affinity of the dipeptide can help determine the proline effect on adjacent residues. Propro's amide carbonyl has a PA of 971 kJ/mol and a Boltzmann of 990 kJ/mol, while prophe has a b site PA of 949 kJ/mol and a Boltzmann-weighted average of 979 kJ/mol. Phepro had the lowest PA with a b site PA 928 kJ/mol and a Boltzmann of 974 kJ/mol.

These results are what is to be expected, since proline is the more basic dipeptide two proline residues should be the most basic. Since the N-terminus of peptides is generally the most basic, prophe should be more basic than phepro since the former has the more basic nitrogen at the N-terminus. However, the amide carbonyl of phepro would seem like the more basic amide carbonyl of phepro/prophe, since the basic proline nitrogen is there to help stabilize the protonation. \textsuperscript{23} The lowest energy conformer of pheprohb did not seem to have any hydrogen bonds or exhibit characteristics that would lower energy, but proprohb and pheprohb both show clear hydrogen bonding.
3.6 Experimental Studies

<table>
<thead>
<tr>
<th>CE</th>
<th>pyrrolidine</th>
<th>piperidine</th>
<th>1-methylimidazolone</th>
<th>2,4 lutidine</th>
<th>diisopropylamine</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-12.78</td>
<td>-3.63</td>
<td>-0.75</td>
<td>-2.02</td>
<td>-0.69</td>
</tr>
<tr>
<td>6</td>
<td>-5.15</td>
<td>-3.55</td>
<td>-3.55</td>
<td>-2.19</td>
<td>-1.01</td>
</tr>
<tr>
<td>9</td>
<td>-3.31</td>
<td>-3.30</td>
<td>-2.40</td>
<td>-1.54</td>
<td>-0.52</td>
</tr>
<tr>
<td>12</td>
<td>-3.06</td>
<td>-3.35</td>
<td>-2.34</td>
<td>-1.46</td>
<td>-0.08</td>
</tr>
<tr>
<td>15</td>
<td>-3.85</td>
<td>-2.87</td>
<td>-2.07</td>
<td>-1.29</td>
<td>-0.13</td>
</tr>
<tr>
<td>18</td>
<td>-2.71</td>
<td>-2.79</td>
<td>-1.92</td>
<td>-1.34</td>
<td>0.29</td>
</tr>
<tr>
<td>21</td>
<td>-3.15</td>
<td>-2.27</td>
<td>-1.45</td>
<td>-0.65</td>
<td>0.56</td>
</tr>
<tr>
<td>24</td>
<td>-2.01</td>
<td>-1.77</td>
<td>-1.39</td>
<td>-0.45</td>
<td>0.96</td>
</tr>
<tr>
<td>27</td>
<td>-1.50</td>
<td>-0.75</td>
<td>-1.05</td>
<td>-0.55</td>
<td>0.32</td>
</tr>
<tr>
<td>30</td>
<td>-1.54</td>
<td>-1.27</td>
<td>-0.84</td>
<td>-0.20</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Table 3: ln(I_B/I_A) values

We were able to do a kinetic method experiment on prophe using the new TSQ quantum ultra triple quadrupole instrument. Dimers were created between prophe and the five reference bases as outlined in the experimental section Table 3 shows ln(ratio) values for the five reference compounds at varying collision energies. This table of values should display some trends: the rows should increase in value from left to right and the columns should tend towards 1 from top to bottom. Overall that trend holds true but there are some serious dips and rises among the data. Chart 1 below shows the first kinetic method plot for this data.
Chart 1: Kinetic method plot 1.

Chart 2: Kinetic method plot 2.

Chart 2 shows the second kinetic method plot for this data. The slope of this plot is the difference in proton affinity between the average PA of all the references and the analyte. In this case, the experiment gives the PA of prophe to be 978 kJ/mol, which is quite close to the previously determined PA of 975 ± 16 kJ/mol and a theoretical PA of 979 kJ/mol.
Chapter 4: Conclusion and future work

The proton affinities of three amino acids were determined through theoretical calculations, using Hartree-Fock and B3LYP theory. Pro-pro, prophe and phe-pro were determined to have proton affinities of 990, 979 and 974 kJ/mol respectively. These values were determined to be in good agreement with the experimental data from another study as well as new experimental data, which determined the proton affinity of prophe to be 978 ± 16 kJ/mol. Further work will fully analyze the proline effect in dipeptides.

The energetics of dipeptides containing proline help define when and in which direction preferential cleavage will occur, therefore all dipeptides that contain proline should be calculated and compared to each other and experimental results. However, due to the limitations of calculations, specifically the size of molecules, will hinder theoretical studies on proline structure and behavior in the gas-phase. This can partly be remedied by studying proline homologs, since the pipecolic acid preferentially cleaves towards the C-terminus instead of the N-terminus\textsuperscript{20} the difference between proline dipeptide conformers and pipecolic acid dipeptide conformers may reveal which energy-lowering behaviors are due to structural differences. The use of the mass spectrometer will be very helpful in determining if the calculated proton affinities will be proportional to proline’s preferential cleavage. The difference in structures and proton affinities for these dipeptide homologs may help us understand how the structure of proline actually
affects its proton affinity and the affinity of the carbonyl groups in peptide bounds that surround it. Computers with higher processing power will be available in the future allowing for larger molecules to be studied using theoretical calculations. The study of tripeptides with proline in the middle would be a better indication of proline's behavior in a longer peptide chain that could undergo the mobile proton mechanism of fragmentation. If proline is at the C-terminal of a tripeptide the other two residues can start the formation of an oxazolone ring, which might show why proline structure favors y ions. The theoretical calculations will be very useful for understanding the steric effects of amino acids in a peptide chain better with the possibility of determining the real structures of gas-phase peptides.
References


