


5-2022

Gas-Phase Proton Affinities for Twenty of the Proline-Containing Dipeptides

Henry Cardwell

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

 Part of the [Analytical Chemistry Commons](#), [Computational Chemistry Commons](#), and the [Physical Chemistry Commons](#)

Recommended Citation

Cardwell, Henry, "Gas-Phase Proton Affinities for Twenty of the Proline-Containing Dipeptides" (2022). *Undergraduate Honors Theses*. William & Mary. Paper 1753.
<https://scholarworks.wm.edu/honorstheses/1753>

This Honors Thesis -- Open Access 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.

Gas-Phase Proton Affinities for Twenty of the Proline-Containing Dipeptides

A thesis submitted in partial fulfillment of the requirement for
the degree of Bachelor of Science in Chemistry from William &
Mary

by

Henry Thomas Cardwell

Accepted for Honors
(Honors) ~~High Honors, Highest Honors~~


John C. Poutsma, Advisor


Nathanael Kidwell


Stephen Trefethen

Williamsburg, VA
May 2, 2022

Abstract

Peptide fragmentation plays a crucial role in the analysis of proteins through mass spectrometry-based proteomics. Most proteomics experiments take place in the low-energy regime and are governed by the mobile proton model which predicts random cleavages along the peptide backbone; however, there sometimes arise circumstances where the mobile proton model fails causing sequencing algorithms to misidentify peptides. One such example is noted in the “proline effect” wherein proline-containing peptides preferentially fragment N-terminal. While it has been established that the “proline effect” is due to the rigidity and basicity of the proline N-terminus, a further understanding of the factors influencing the “proline effect” is desired. This paper aims to work towards an understanding of how adjacent amino acid residues aid in enhancing or hindering the “proline effect”.

To this effect, tandem mass spectrometry was used with the extended kinetic method to evaluate the proton affinities of the ProXxx dipeptides. ProArg and ProLys were unable to be quantified, but the remaining dipeptides ranged in proton affinity from 969.2 ± 13 kJ/mol to 1010.8 ± 17 kJ/mol for ProGly and ProGlu respectively. Additional computational work was performed using DFT with the B3LYP basis set to identify proton affinities for each dipeptide, as well as provide structurally resolved information. There is general agreement between the computational and experimental results. Further work is still required to form a more complete understanding of the “proline effect”.

Broader Impacts

Proteins have been a recent target for both clinical and academic based research. Due to their wide varieties of functions, a detailed understanding of the proteome, the complete set of proteins that an organism expresses, is useful. Proteomics is the study of the proteome, typically accomplished using mass spectrometry (MS).

MS based proteomics first relies on digesting proteins, long chains of amino acids, into shorter sections known as peptides. These peptides are then fragmented inside the instrument in a mostly predictable manner. Computer programs can analyze the data from these fragments and work backwards to identify peptides, which can be used to identify whole proteins.

Unfortunately, peptides containing proline, an amino acid, often produce atypical fragments that can confound these computer algorithms. If there are cases where a specific protein involved in disease heavily incorporates proline, it would be more difficult to identify using the typical proteomics tools. To better understand the unusual “proline effect,” it is important to understand how the other amino acids that may be adjacent to the proline in these proteins contribute.

By using tandem MS and computational chemistry, it is possible to identify the thermochemical properties of small units called dipeptides which contain only proline and another amino acid. These thermochemical properties serve to inform us about the “proline effect,” and may pave the way to improving the computer algorithms used in proteomics research.

Table of Contents

Chapter 1: Introduction	1
1.1 Background	1
1.2 The Extended Kinetic Method	4
1.3 Computational Studies	10
1.4 Guide to Thesis	12
Chapter 2: Methods	13
2.1 Experimental Determination of Proton Affinities	13
2.2 Computational Predictions for Proton Affinities	19
Chapter 3: Results and Discussion	22
3.1 Kinetic Method Proton Affinity Examples	22
3.1.1 Example 1: ProVal	22
3.1.2 Example 2: ProSer	25
3.2 Kinetic Method Results Summary	29
3.3 Computational Predictions Structural Motifs	31
3.3.1 ProGly Predictions	31
3.3.2 ProVal Predictions	32
3.3.3 ProSer Predictions	34
3.4 Computational Proton Affinities	37
3.5 Dipeptides Summary	39
Chapter 4: Conclusion	43
Appendix	45
References	53

Acknowledgements

I would like to thank Dr. JC Poutsma for his guidance and expertise over the past three and a half years in the Ionlab. I would also like to thank the Faculty and Staff of the Department of Chemistry at William & Mary that have helped push me forward towards growing as a chemist. Their support and insight have been invaluable.

Also, I would like to thank the members of my lab that have helped me directly on my project including Trinh Ton, Nihal Patel, and everyone else. Thank you to my peers Anna Grace Towler, Daria Moody, and Sophie Messinger for fostering the lab environment during our time together.

This work would not have been possible without contributions from The College of William & Mary, the National Science Foundation, and The Arnold and Mabel Beckman Foundation.

List of Tables

Table 3.1: Reference Bases for ProVal	22
Table 3.2: Reference Bases for ProSer	26
Table 3.3: Summary of Experimental Proton Affinities	30
Table 3.4: Summary of Computational Proton Affinities	39
Table 3.5: Summary of Dipeptides Proton Affinities	40
Table 4.1: Summary of Dipeptides Proton Affinities	43
Table A.1: Reference Bases Usage	45

List of Figures

Figure 1.1: Proline	3
Figure 1.2: Competitive Cleavage of a Proton-bound Dimer	8
Figure 2.1: Kinetic Method Plot 1 for ProGly	15
Figure 2.2: Effective Temperature Plot for ProGly	16
Figure 2.3: Kinetic Method Plot 2 for ProGly	18
Figure 3.1: ProVal Structure	22
Figure 3.2: Kinetic Method Plot 1 for ProVal	23
Figure 3.3: Kinetic Method Plot 2 for ProVal	24
Figure 3.4: Orthogonal Distance Regression for ProVal	25
Figure 3.5: ProSer Structure	26
Figure 3.6: Kinetic Method Plot 1 for ProSer	27
Figure 3.7: Kinetic Method Plot 2 for ProSer	28
Figure 3.8: Orthogonal Distance Regression for ProSer	29
Figure 3.9: Lowest Energy Conformers of ProGly	31
Figure 3.10: Lowest Energy Conformers of ProVal	33
Figure 3.11: Alternative Structures for ProVal	34
Figure 3.12: Lowest Energy Conformers for ProSer	36
Figure 3.13: Alternative Structures for ProSer	37
Figure A.1: ODR Plot for ProAsn	45
Figure A.2: ODR Plot for ProAsp	46
Figure A.3: ODR Plot for ProCys	46
Figure A.4: ODR Plot for ProGln	47

Figure A.5: ODR Plot for ProGlu	47
Figure A.6: ODR Plot for ProGly	48
Figure A.7: ODR Plot for ProHis	48
Figure A.8: ODR Plot for Prolle	49
Figure A.9: ODR Plot for ProLeu	49
Figure A.10: ODR Plot for ProMet	50
Figure A.11: ODR Plot for ProPro	50
Figure A.12: ODR Plot for ProThr	51
Figure A.13: ODR Plot for ProTrp	51
Figure A.14: ODR Plot for ProTyr	52

Chapter 1: Introduction

1.1 Background

When the human genome was sequenced in the beginning of the 21st century through the Human Genome Project, it was clear that biochemistry was entering an unprecedented area of discovery. At the same time, scientists were surprised to discover that the entirety of human existence could be encoded by less than 65,000 genes with only approximately 20,000 of those genes encoding for proteins.¹ In fact, simple organisms such as corn have more genes than humans with 32,000 protein-encoding genes.² Such results were shocking. It was increasingly clear that whatever humans like to pride themselves on as setting them apart, it did not originate from a quantity of genes. Rather, one area that has been the subject of investigation is that of the proteins encoded by these genes. To understand how proteins support life, it is important to understand the key connection between their structure, function, and energetics.

One field of chemistry that has been used to study proteins is mass spectrometry. Mass spectrometry is a highly powerful analytical tool that has been used for chemical analysis since the turn of the 20th century when J.J. Thompson identified non-radioactive isotopes for a variety of elements.³ By separating ions by a mass to charge (m/z) ratio, these devices have myriad applications. With the invention of electrospray ionization (ESI) as a soft ion source for mass spectrometry by John Fenn and Masamichi Yamashita in 1984, the field of mass spectrometry gained ready access to biomolecules.³ It was not long after this that the field of mass spectrometry-based proteomics was established. Proteomics is

the study of the proteome, the total set of proteins that is encoded by an organism's genetic code, often using mass spectrometry. One of the most common forms of proteomics is named bottom-up proteomics wherein a protein is chemically digested into smaller units known as peptides, which are shorter chains of amino acids. These peptides are then ionized using ESI and subjected to collision-induced dissociation which causes the peptides to fragment.⁴ From the fragments, it is possible to work backwards and identify the peptides and, ultimately, the original protein that was digested. It is possible to determine peptide sequence from fragments by hand, but it is more often accomplished using computer algorithms that predict peptide fragmentation spectra.

Most protein sequencing algorithms rely on the mobile-proton model to predict peptide fragmentation from collision-induced dissociation (CID).^{5,6} The mobile-proton model assumes that when a positively charged peptide is collisionally activated, the proton is made mobile and can move to other sites inducing fragmentation; however, this assumption does not always hold true. More specifically, there arise situations where the mobile proton model fails to accurately predict fragmentation. One such irregularity can be seen in the case of proline-containing peptides.⁷⁻⁹ Proline, as depicted in Figure 1.1, has a unique side chain that connects back to the peptide backbone on the amine site. When proline is inserted into a peptide, the peptide preferentially fragments N-terminal to proline residues, an irregularity that has been called "the proline effect". By fragmenting

preferentially, these peptides can be mis-assigned by the sequencing algorithms that assume a more random distribution of fragmentations.

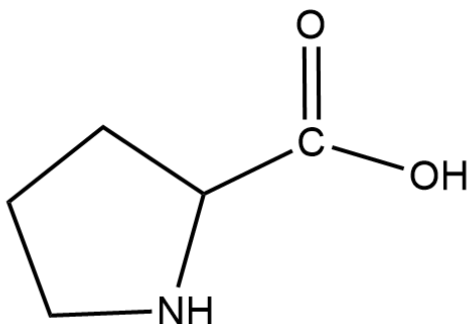


Figure 1.1: Proline

Previous work by Poutsma, Wysocki, and others has shown that the “proline effect” results from the increased basicity of the tertiary amide, as well as the rigid structure of proline’s unique five membered ring.^{10,11} It has even arisen that more than the differing basicity, it is the rigidity that plays the key role in altering the fragmentation chemistry of these sites likely due to restricting the mobility of the proton.^{10,11} Further, a study by Yates and Wysocki looking at the fragmentation intensities of proline-containing peptides identified that certain adjacent residues produced N-terminal fragmentation more often. Specifically, fragmentation of the XxxPro bond was strongest when Xxx was His, Asp, Val, Ile, or Leu, and weakest when Xxx was Gly or Pro; however, the specific mechanism for the influence of the adjacent residues is unknown.¹² To gain a more practical understanding of the causes of the preferential fragmentation, it is important to take a more systematic and rigorous approach that looks at the individual interactions between each residue with proline. In the pursuit of a mechanistic understanding of the “proline effect”, an investigation into how the presence of each amino acid affects the

“proline effect” is necessary. To better grasp the impacts of the “proline effect” in a more complete context, the thermochemical properties of each of the thirty-nine proline-containing dipeptides were probed experimentally, supported by computational structural analysis. By breaking the “proline effect” into discrete units with resolved structures, it becomes possible to establish a better mechanistic understanding of the preferential fragmentation that has been observed.

1.2 The Extended Kinetic Method

When seeking to understand molecular activity, it has become increasingly clear that the structure, thermochemical properties, and function of a given molecule are highly related.^{13,14} One area that holds importance for amino acids and peptides is that of acid-base properties. Such properties are clearly important for the sake of biology and medicine where slight changes in pH can cause severe harm, but they also act on an even more fundamental level. For the sake of peptide fragmentation, acid-base properties can be seen to play a pivotal role due to the dependence of the mobile proton model on the affinity of different sites for the mobile proton.

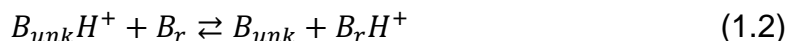


Using the Bronsted-Lowry definition of an acid, an acid-base reaction may take the form of Equation 1.1. The conjugate acid is a protonated species (BH^+)

which loses a proton by dissociation to produce a free proton (H^+) and a neutral base (B) that would act as a proton acceptor. Proton affinity is a gas-phase analog of basicity that corresponds to the energy required to remove the proton from BH^+ .¹⁵ The proton affinity (PA) of B is equal to the enthalpy of reaction (ΔH_{rxn}) for the deprotonation reaction depicted in Equation 1.1. From Hess's Law, ΔH_{rxn} is equal to the sum of the heats of formation (ΔH_f) of the products minus the $\Sigma \Delta H_f$ of the reactants. While aqueous pKa values would be more relevant to the native function and behavior of proteins, which are typically solvated, this specific study is concerned with peptide fragmentation which occurs in the gas phase. The removal of solvation effects can lead to completely different energetic trends.¹⁶ As such, gas-phase measurements and trends cannot be taken directly from an aqueous analog.

PA is a useful property to consider, yet it is highly difficult to accurately measure the ΔH_f of the gas-phase species described in Equation 1.1 using standard calorimetry. To overcome the calorimetric limitations of gas-phase species, several different methods have been established to determine the proton affinity of a gas-phase molecule. The first of these methods, known as the equilibrium method, involves utilizing a reference (B_r) with known proton affinity, as depicted in Equation 1.2.¹⁷ By determining the equilibrium constant (K_{eq}) of the reversible reaction as a function of temperature, it is possible to extract the proton affinity of B_{unk} using a Van't Hoff plot; however, the equilibrium method has several strict requirements. Firstly, the samples must be both highly pure and volatile to determine the concentrations of neutrals. Secondly, there must be an appropriate

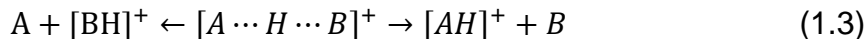
reference base with a PA within 4-8 kJ/mol of the analyte. In the gas phase, reactions need to be almost thermo-neutral for an equilibrium to be observed as endothermic reactions are highly inefficient.



Another potential option for the determination of proton affinities is ion-molecule bracketing.¹³ Rather than quantifying an equilibrium constant, ion-molecule bracketing uses a series of references and simply asks whether the ionized products from the reaction depicted in Equation 1.2 are observed or not. If the reaction proceeds, then the reaction is assumed to be exothermic. When the reaction does not proceed, the reaction is assumed to be endothermic. From the observance and lack thereof of products, it is thus possible to gain an approximate idea of the thermochemical property of interest for the analyte. Ion-molecule bracketing, while less strict than equilibrium, still suffers from some difficulty in reference selection. Rather than one suitable reference, several close references must be identified. Additionally, they must also be relatively pure samples to avoid side product formation.

As an alternative to these two techniques, each with their limitations, Graham Cooks established the kinetic method in 1977.¹⁸ Rather than looking at an equilibrium reaction, the kinetic method uses the competitive cleavage of a proton-bound heterodimer in a tandem mass spectrometer to extract the proton affinity. By starting with a heterodimer, it is possible to obtain a high degree of initial purity

through mass isolation in a mass spectrometer. When the heterodimer receives external energy, the proton can either dissociate with the analyte (A) or with the reference base (B) as shown in equation 1.3. The cleavage will proceed based upon the proton affinities of the two molecules.



Two rate constants are then produced: k_A and k_B . From transition state theory, the two rate constants can be combined to produce Equation 1.4 and Equation 1.5. The simple version of the kinetic method then requires a negligible difference in entropy between the analytes yielding the simplified Equation 1.6.

$$\ln \left(\frac{k_B}{k_A} \right) = \frac{\Delta(\Delta G)}{RT_{eff}} \quad (1.4)$$

$$\ln \left(\frac{k_B}{k_A} \right) = \ln \left(\frac{Q_B^*}{Q_A^*} \right) + \frac{\Delta \varepsilon_0}{RT_{eff}} \quad (1.5)$$

$$\ln \left(\frac{k_B}{k_A} \right) \approx \frac{\Delta(\Delta H)}{RT_{eff}} \quad (1.6)$$

T_{eff} in all four equations is an “effective” temperature that serves as a microcanonical energetic substitute for temperature. As the analytes are under vacuum in a mass spectrometer, they cannot freely exchange energy as required to have a temperature. Because there is no temperature, the microcanonical ensemble, which assumes the number of particles, volume, and energy are held constant, must be used in place of the canonical ensemble, which assumes

temperature is constant rather than energy. The Q terms represent the partition function for the reference (Q_B^*) and the analyte (Q_A^*). $\Delta\varepsilon_0$ is the difference in activation energy between the two dissociation pathways. $\Delta(\Delta H)$ represents the difference in PA for the reference and analyte. Because there are no terms from neutral species in the kinetic method, this technique can be applied to nonvolatile compounds such as peptides that can be ionized via electrospray ionization. A better depiction of the dissociation pathway energetics can be seen in Figure 1.2. The figure shows how the number of available quantum states for each compound depend upon the activation energy.

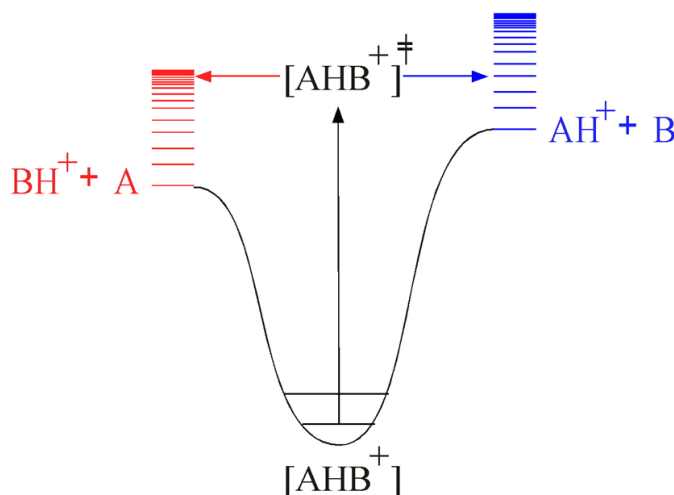


Figure 1.2: Competitive Cleavage of a Proton-bound Dimer, by measuring the dissociation of the complex into the two ions BH^+ and AH^+ , the kinetic method allows for the determination of the difference between the thermochemical properties of these two compounds. The depiction shows a lack of barriers to each dissociation pathway, as well as the role of the partition function in the statistical mechanical dissociation.

Compared to prior techniques, the kinetic method provides a robust experimental technique for the determination of thermochemical properties, yet it still has several limitations. Most significantly, the simplest version of Cooks'

method relies upon the cancelation of entropic effects between the two competing fragments. In the case where entropic effects cancel, the recorded $\Delta(\Delta G)$ can be approximated as the difference in solely the enthalpic contributions which are sought. Requiring the cancelation of entropy is a difficult proposition for an analyte with unknown thermochemical properties. Additionally, many reference bases with similar entropic factors may likewise have an unknown PA rendering them useless as a reference. Given that many amino acid residues have highly entropic side chains like that of lysine, it would be highly suspect to neglect the contribution of entropic factors.

To account for the differences in entropic terms, it is necessary to adopt amendments to the kinetic method as suggested by Wu, Fenselau, Wesdemiotis, and Armentrout.^{19–22} These amendments comprise what is referred to as the extended kinetic method. Instead of the above simplified equations, the extended kinetic method transforms Equation 1.4 into Equation 1.7.

$$\ln \left(\frac{k_B}{k_A} \right) = \frac{\Delta(\Delta H)}{RT_{eff}} + \frac{\Delta(\Delta S)}{T_{eff}} \quad (1.7)$$

With the extended kinetic method, additional measurements are taken under different experimental conditions in which T_{eff} is varied systematically to extract the $\Delta(\Delta S)$ and $\Delta(\Delta H)$. At each of the different T_{eff} , the slope of the corresponding line changes; however, these different lines all cross at a point termed the isothermal point. This point contains the enthalpic and entropic values of the analyte of interest and represents a position where the cleavage of a heterodimer containing

a reference with identical properties would be constant at any T_{eff} . Along with the extended experiment, a statistical analysis proposed by Armentrout and Ervin named orthogonal distance regression (ODR) is then applied to the data.²³ ODR determines an isothermal point of best fit given the experimental values for T_{eff} using Monte Carlo simulations.

1.3 Computational Studies

With the advance of technology in the past 20 years, computational methods have become more accessible and accurate for use in chemistry. Such calculations rely on the appropriateness of the technique, as well as the way they are performed. When used in tandem with experimental results, calculations serve as a highly useful tool. In addition to identifying a fallacy in either the experiment or the calculation, by combining the two a better understanding of the intricacies of inter and intramolecular interactions are made more apparent. The proton affinity of a species can be determined with relative ease by calculating the enthalpy of formation of both the protonated and the neutral species of a molecule in addition to the known value for the proton as depicted in Equation 1.8.

$$PA_{unk} = \Delta H_{rxn} = (\Delta_f H_{298}(B) + \Delta_f H_{298}(H^+)) - \Delta_f H_{298}(BH^+) \quad (1.8)$$

The most crucial aspect of computational studies lies in the choice of method given the myriad options available. From the computationally cheap empirical methods such as molecular mechanics to intensive *ab initio* calculations,

a range of options exists. By combining several techniques, the advantages of each can aid in establishing a more complete picture of the potential energy surface (PES) for the molecules of interest.

The PES for large molecules can often be a complex space with possibly thousands of local minima. To determine the global minima of these PES, it is important that the chosen method sufficiently samples across the space when choosing starting conformers. Starting conformers can be identified using less intensive molecular mechanics (MM) methods.²⁴ MM calculations rely on empirically derived rules to assign energetic penalties to generated structures based on deviations from known low-energy positions. These penalties are assigned using force fields that quantify the extent of the penalty to be applied based on experimental measurements. Many different force fields exist, with each designed for a different collection of test systems.²⁵

Starting structures can then be minimized using more accurate and rigorous methods such as Hartree-Fock (HF) calculations or density functional theory (DFT) calculations which both rely more heavily on quantum calculations of energy.²⁴ As the energy of a system with more than two bodies cannot be solved exactly, there are again several options to approach the correct answer that make different assumptions or assign different cut offs. All three of these techniques are incorporated and discussed in greater detail in the Methods section.

1.5 Guide to Thesis

The next chapter will detail how the extended kinetic method and computational methods are applied to proline-containing dipeptides to answer questions about thermochemistry and the “proline effect”. An example of one experiment with ProGly is included to demonstrate the relationships being discussed. The third chapter will discuss the findings from both sets of methods for each of the ProXxx dipeptides. A discussion of proton affinities will be made, as well as a more specific discussion of structure for the three model dipeptides ProGly, ProVal, and ProSer. Uncertainties and remaining questions will also be explored. The fourth chapter will serve as a conclusion to the findings presented in this paper, followed by an appendix with supplementary data.

Chapter 2: Methods

2.1 Experimental Determination of Proton Affinities

All dipeptides were purchased from the commercial manufacturer AnaSpec in their pure forms and stored appropriately until use. Stock solutions of the dipeptides and the reference bases were prepared at concentrations of 10^{-3}M in 50:50 methanol:water with 1% formic acid. Inclusion of formic acid encourages protonation and the formation of the heterodimer. Experimental samples were prepared in ratios of either 1:1 or 3:2 dipeptide to reference base using the respective stock solutions, with the 3:2 ratio further encouraging the formation of heterodimers.

The prepared samples were then injected into a Finnigan TSQ Quantum Ultra triple quadrupole mass spectrometer in positive ion mode using an external syringe pump at a flow rate of $10\text{ }\mu\text{L/min}$. Samples were ionized using electrospray ionization with the capillary heated to 115°C , the spray voltage set to 4000V , and a sheath gas flow rate of 10-20 arbitrary units. The instrument was operated in product mode allowing the heterodimer to be isolated in the first quadrupole (Q1) with an isolation width of 0.80 Th, fragmented using collision-induced dissociation (CID) in the second quadrupole (Q2) at a partial pressure of argon of 0.5 mTorr, and scanned in the third quadrupole (Q3) at an isolation width of 0.9 Th. Fragmentation product data was collected at collision energies from 3-30 eV in intervals of 3 eV. In addition to each sample, a fragmentation spectrum of each dipeptide and each reference base was also collected to aid in identifying

secondary fragmentation products. Spectral data was time averaged and exported to Microsoft Excel® for data analysis. A ratio of reference base fragments (B) over dipeptide fragments (A) can then be calculated. Data collection was replicated over a minimum of three days and averaged for each sample. The natural log of the averaged ratio at each collision energy was then taken for further analysis.

In the case of a mass spectrometry experiment, the ratio of the two rate constants k_a and k_b from Equation 1.4 can be approximated as the ratio of the signals for the two different ion products. Substituting the ratio of product ion appearance into the equation for the extended kinetic method yields Equation 2.1. This equation can be further amended by subtracting the average PA of the set of reference bases (PA_{avg}) from the PA of the reference base and the analyte giving Equation 2.2. Subtracting the average PA allows for the points of interest to become centered on the y-axis removing unwanted correlation between the slope and the y-intercept as suggested by Armentrout and Ervin.²³

$$\ln\left(\frac{k_B}{k_A}\right) \approx \ln\left(\frac{[BH]^+}{[AH]^+}\right) = \frac{\Delta(\Delta H)}{RT_{eff}} + \frac{\Delta(\Delta S)}{T_{eff}} \quad (2.1)$$

$$\ln\left(\frac{[BH]^+}{[AH]^+}\right) = (\Delta H_{B_i} - \Delta H_{avg}) \frac{1}{RT_{eff}} - \left[\frac{(\Delta H_A - \Delta H_{avg})}{RT_{eff}} + \frac{(\Delta S_A - \Delta S_{B_i})}{R} \right] \quad (2.2)$$

Equation 2.2 demonstrates the linear relationship between the difference in PAs, and the natural logarithm of product ion ratios at a constant T_{eff} . In the triple quadrupole mass spectrometer, the collision energy of the ions is used to adjust the T_{eff} . The PAs of each reference base were obtained from the NIST Webbook and can be found in the Appendix.^{26,27} A linear plot in slope-intercept form, hereafter referred to as Plot 1, is then constructed from Equation 2.2. Figure 2.1 shows an example of Plot 1 for the proline-containing dipeptide ProGly. The experimentally determined natural logarithms are y-values for the plot. The x-values correspond to $\Delta H_{\text{Bi}} - \Delta H_{\text{avg}}$ of the reference bases. The slope of the best-fit lines to the data is $\frac{1}{RT_{\text{eff}}}$ and the term in the square brackets of equation 2.2 is the y-intercept of the best fit lines, which is a constant at a given T_{eff} . Each vertical set of data represents the data from a singular reference base. A linear regression is generated for each collision energy, and these lines then should cross over at the

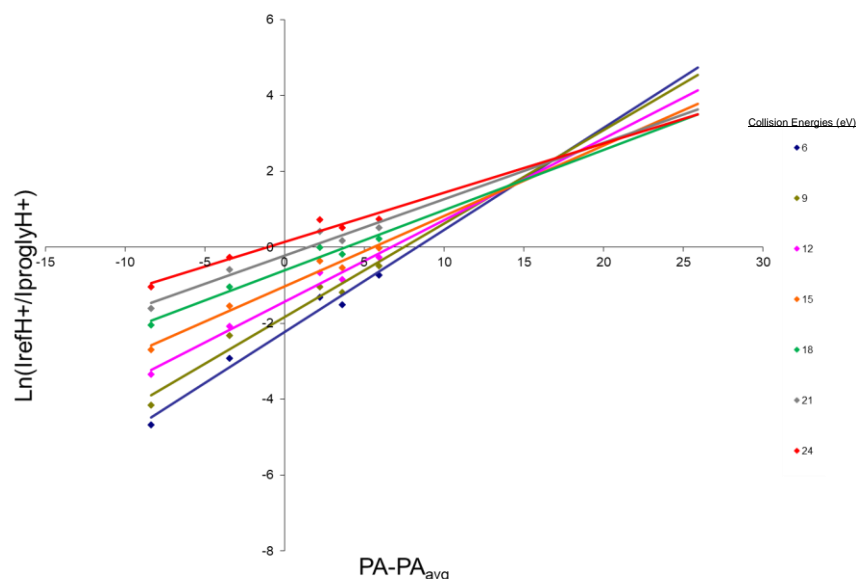


Figure 2.1: Kinetic Method Plot 1 for ProGly representing the changing relationship between product ion formation and the proton affinity of the reference base.

isothermal point. There is uncertainty present in Plot 1 given that the isothermal point is not a single, easily identifiable point that may be due to ion stability efficiencies. The x-value of this point is equal to the $PA_{\text{dipep}} - PA_{\text{avg}}$ and the y-value contains the entropy term of the energy, specifically in the form of $\Delta S/R$.

To check the data for accuracy, it is often useful to construct a plot of the calculated T_{eff} for each collision energy, versus the collision energy. As the collision energy is the means through which energy is added to the heterodimer, these two terms should be highly correlated in a linear manner. Such a plot can be seen below in Figure 2.2 for the case study of ProGly. Any deviations from linearity can often be traced directly back to experimental discrepancies. Any such deviations

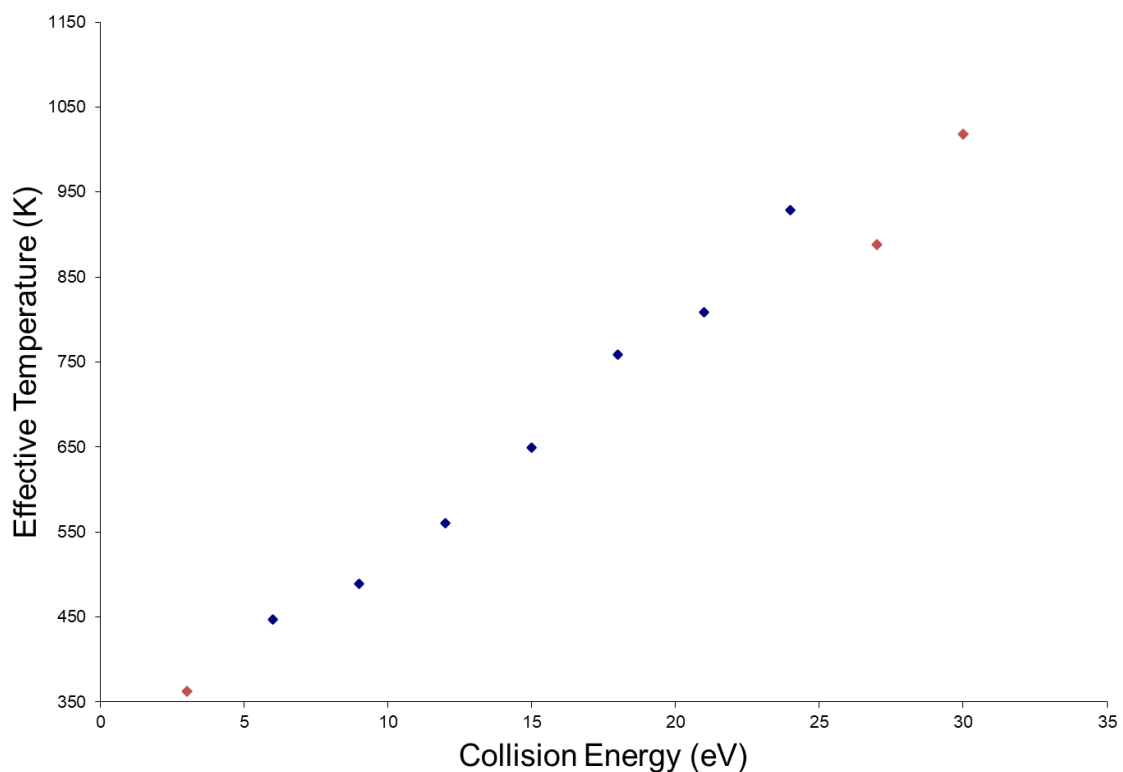


Figure 2.2: Effective Temperature Plot for ProGly demonstrating the linear relationship between T_{eff} and collision energy. Data points at collision energies of 3 eV, 27 eV, and 30 eV were chosen to be omitted from further workup due to deviations from linearity.

that cannot be easily corrected for with additional data collection or by changing the choice of reference base is then excluded from the final calculations. For example, when looking at Figure 2.2, the correlation between T_{eff} and collision energy is approximately linear until collision energies of 27 eV and 30 eV. Additionally, collision energy 3 eV somewhat deviates from linearity. Potential explanations for deviation from linearity are explored in Section 3.1. Because of the loss of linearity, Plot 1 above and any further work only include collision energies from 6 to 24 eV.

An additional plot, Plot 2, is useful for again verifying the choice of collision energies. Plot 2 is obtained by plotting the negative intercept from each T_{eff} line versus their respective slopes. Plot 2 for ProGly can be seen in Figure 2.3. Equation 2.3 represents the relationship being plotted in Plot 2. The left hand of Equation 2.3 is said to be y and is equal to the negative intercept of the line represented by Equation 2.2. The x -values of $\frac{1}{RT_{eff}}$ come from the slopes of lines represented by Equation 2.2. In essence, such a plot verifies the correlation between these two terms which should be highly related. An r^2 correlation coefficient can be generated using a linear regression of Plot 2 providing a qualitative metric as to the fit of the data. Plot 2 also allows for the calculation of the values for PA and entropy of the analyte. The slope of Plot 2 can be added to PA_{avg} to obtain the PA of the analyte, and the entropy of the analyte is obtained by multiplying the y -intercept by R .

$$\frac{(\Delta H_A - \Delta H_{avg})}{RT_{eff}} + \frac{(\Delta S_A - \Delta S_{Bi})}{R} = (\Delta H_A - \Delta H_{avg}) \frac{1}{RT_{eff}} + \frac{(\Delta S_A - \Delta S_{Bi})}{R} \quad (2.3)$$

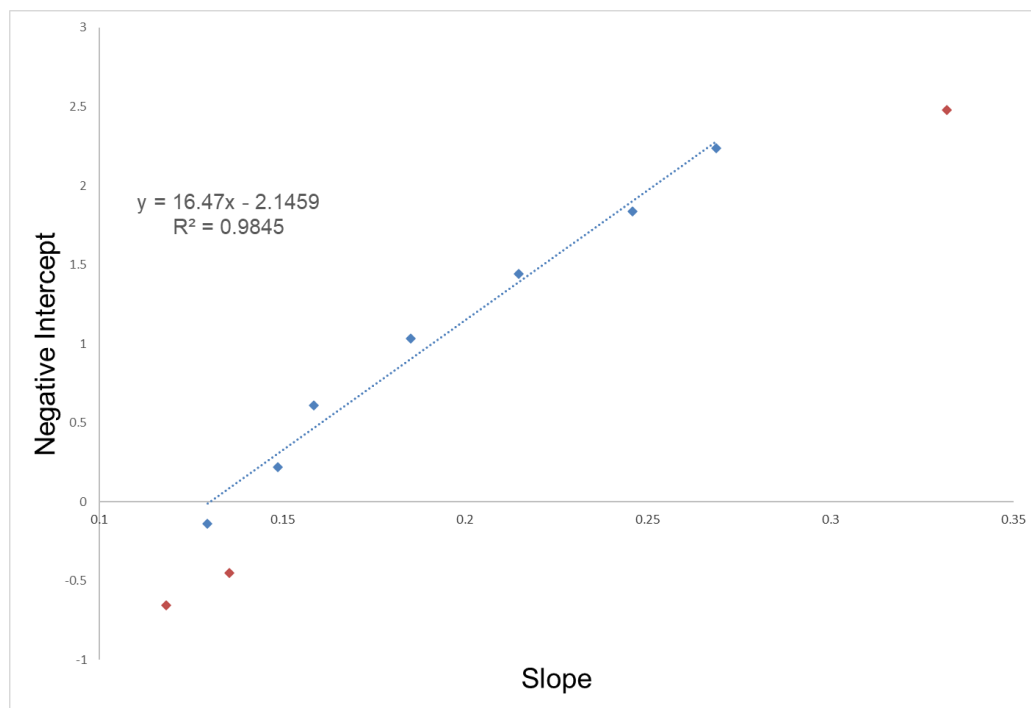


Figure 2.3: Kinetic Method Plot 2 for ProGly demonstrates the linear relationship between the negative intercept and the slope. An R^2 correlation value of 0.9845 and preliminary values of 968.2 kJ/mol and $-17.8 \text{ J mol}^{-1} \text{ K}^{-1}$ for the PA and entropy of ProGly were extracted from Plot 2 respectively.

As Plot 1 and 2 are constructed from experimental values and simple linear regressions, there is often some uncertainty present in the isothermal point. Such uncertainty can clearly be seen in Figure 2.1 where the linear regressions do not cross at a singular point. Due to the inherent uncertainty in the experimental isothermal point, the initial PA and ΔS values derived from Plot 2 are used as starting points for the ODR which utilizes a thousand iterations of Monte Carlo simulations to determine a more thoroughly correct isothermal point with appropriate uncertainties.

2.2 Computational Predictions for Proton Affinities

Calculations for this project were performed at various levels of theory to reduce overall calculation time and to properly sample the PES of each dipeptide. Initial conformers were generated using a GMMX searching algorithm using the MMX force field in the PCModel program.²⁸ Conformers within 42 kJ/mol of the lowest energy conformer were identified with a maximum of 30,000 unique conformers sampled. All further calculations were performed using Gaussian 09.²⁹ Following identification in the MMX search, conformers were further minimized using Hartree-Fock calculations with a RHF/3-21G basis set to identify any redundant conformers. Unique RHF/3-21G conformers were then minimized using DFT calculations and a B3LYP/3-21G basis set followed by optimization at the B3LYP/6-31+G* level.³⁰ Other studies have shown the use of B3LYP as an appropriate functional for the analysis of organic-based biomolecules.³¹ The B3LYP/6-31+G* calculations represent the final geometry optimization for all conformers which were then processed in two ways. The first of these is a single-point energy calculation for each conformer at the B3LYP/6-311++G** level of theory. Secondly, vibrational-frequency calculations were performed at the B3LYP/6-31+G* level of theory from which entropy, enthalpy, and free energy corrections were determined. By starting from cheaper calculations, redundant conformers were removed at the cheaper levels rather than at the more expensive levels.

For each dipeptide, conformers were generated for protonated and neutral species. Protonation sites of N-terminus, the amide carbonyl oxygen, and the

amide nitrogen were used as starting structures for each dipeptide. In the case where side chains could reasonably hold a charge, starting structures were generated using these alternative protonation/deprotonation sites. Likewise, zwitterionic neutrals and cations were investigated as alternative starting structures for some dipeptides. Depending on the dipeptide, anywhere from fifty to five thousand conformers were generated in this way for each state of the molecule.

Results from these calculations were compiled into Excel documents. The free energy for each conformer was generated from the sum of the single-point energy from the B3LYP/6-311++G** calculation and the free energy correction from the B3LYP/6-31+G* calculation. Enthalpies for each conformer were also generated from the single-point energies and the correction to the enthalpy. Many of these conformers are within 20, or even 10, kJ/mol of the lowest energy conformer implying that they can be occupied at room temperature and do contribute to the proton affinities of each dipeptide. As such, the enthalpies of the different conformers were Boltzmann weighted according to 298K free energy to provide a final enthalpy (H^{298}) for the molecule at the designated charge.

PA approximations from raw calculations using Equation 2.4 often feature consistent discrepancies compared to experiment. To correct for these errors inherent in the calculations, reported values were isodesmically corrected using glycine with a PA of 886.5 kJ/mol as the reference as depicted in Equation 2.5.^{26,27} Isodesmic correction adds the difference in calculated values between the analyte and a reference to the experimentally known value for that reference. By basing

the computational values on an experimentally known quantity, it is possible to cancel much of the error that may be present in a specific computational method. In addition to yielding a value for comparison with experiment, these calculated values also serve as an initial estimate for the selection of references in the experimental procedure.

$$PA_{unk} = \Delta H_{rxn} = (\Delta_f H_{298}(B) + \Delta_f H_{298}(H^+)) - \Delta_f H_{298}(BH^+) \quad (2.4)$$

$$PA_{iso} = (PA_{raw}(unk) - PA_{raw}(ref)) + PA_{exp}(ref) \quad (2.5)$$

Chapter 3: Results and Discussion

3.1 Kinetic Method Proton Affinities Examples

3.1.1 Example 1: ProVal

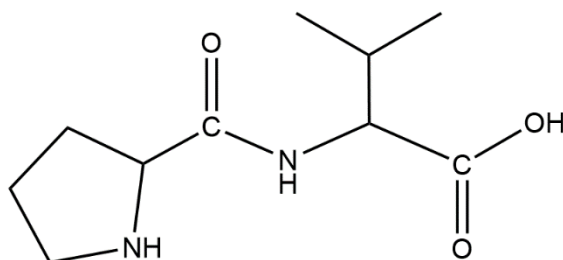


Figure 3.1: ProVal

The structure of ProVal can be seen in Figure 3.1. For a detailed description of which reference bases were used for the ProVal experimental procedure, see Table 3.1. Plot 1 for this data set can be seen in Figure 3.2, plotting the natural logarithm of the ratios of reference base over the analyte versus the relative proton affinity of the reference base ($\Delta H_{\text{avg}} - \Delta H_{\text{ref}}$). Each vertical series of data points represents the data of a single reference base with each color corresponding to a different collision energy. The isothermal point can be identified to the right of the data where the lines of each collision energy cross. While data closer to the

Table 3.1: Reference Bases for ProVal detailing which reference bases were used in the extended kinetic method experiment for ProVal with their respective PAs.

Reference Base	Proton Affinity (kJ/mol)
Pyrrolidine	948.3
Piperidine	954.0
4-Tertbutylpyridine	957.7
2,4-Lutidine	962.9
Diisopropylamine	971.9

isothermal point would be ideal, bases at larger PAs tended to yield inconsistent ratios that deviated at higher collision energies. It is postulated that the inconsistencies result from missing secondary fragmentation products that are not as well detected by the instrumentation at higher energies.

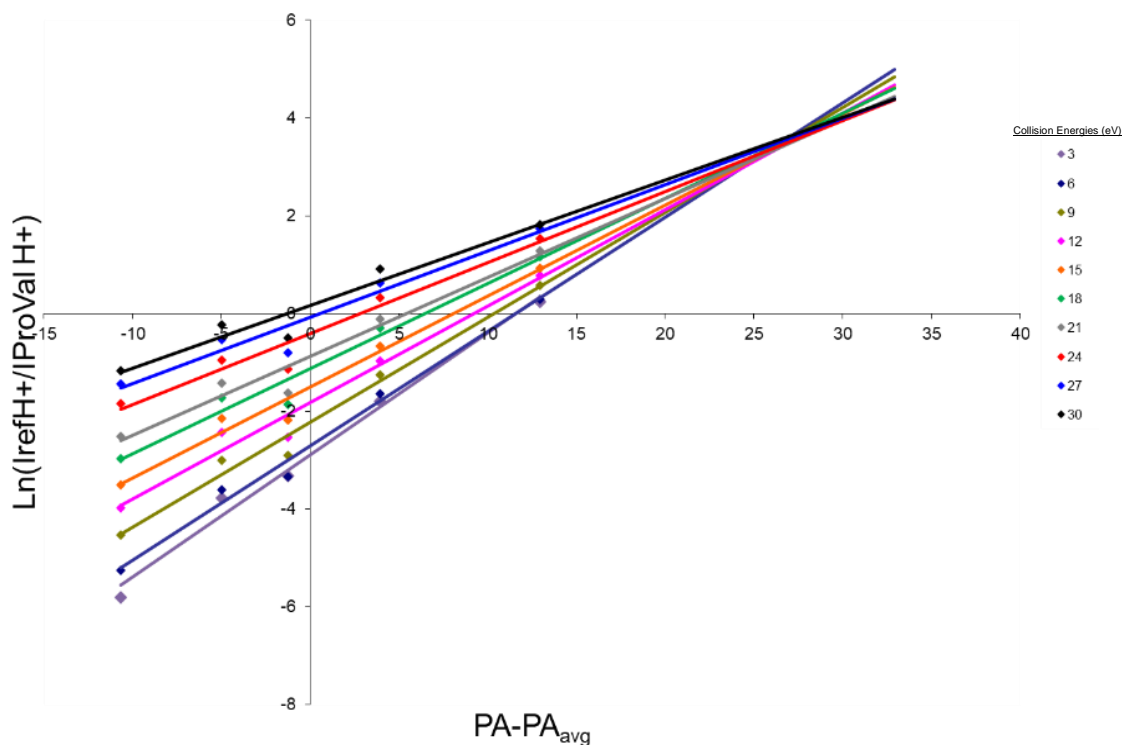


Figure 3.2: Kinetic Method Plot 1 for ProVal showing the inversion of product ion formation as a function of proton affinity. The isothermal point can be approximated to be ~27 kJ/mol above the PA_{avg} .

From Figure 3.2, the slopes and the negative of the intercepts of the individual best-fit lines can be extracted. As both terms depend on $\frac{1}{RT_{eff}}$, Plot 2 is constructed to afford the PA and entropic terms for the analyte. As previously mentioned, Plot 2 allows for a check between the correlation of both terms and provides preliminary estimates for PA and entropic terms. Figure 3.3 shows Plot 2 for ProVal demonstrating an r^2 value of 0.992. The data for ProVal held a strongly

linear correlation across all collision energies such that no data points were omitted. When the values from the slope and the intercept are adjusted using the average values of the references, a preliminary value of 984.7 kJ/mol was predicted for the PA of ProVal.

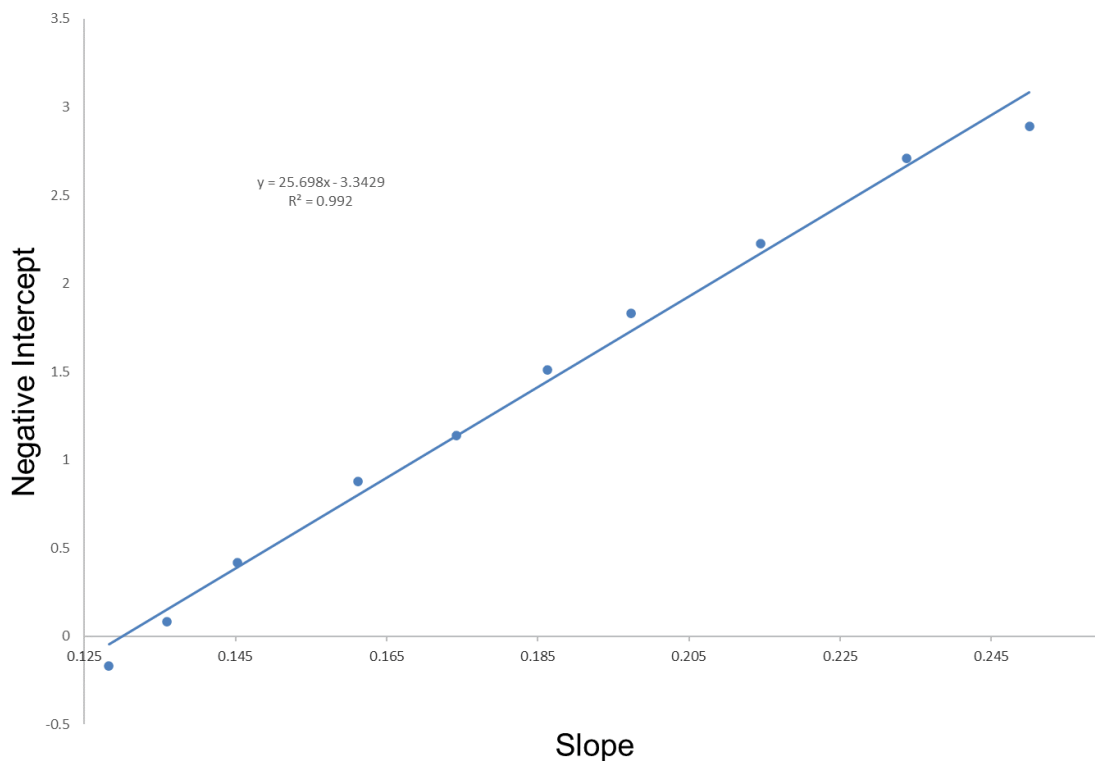


Figure 3.3: Kinetic Method Plot 2 for ProVal demonstrates the linear relationship between the negative intercept and the slope. An R^2 correlation value of 0.992 was calculated for the linear regression showing the strong correlation between these terms. No points were omitted due to general agreement with linearity.

The preliminary values were used as initial guesses for the ODR analysis along with the experimental data and T_{eff} values as initial inputs. Two different analyses were run using uncertainty windows of ± 2 kJ/mol and ± 8 kJ/mol for the PA of the reference bases, which primarily change the uncertainty in the final value

yet rarely change the value itself. Any reported values represent the ± 8 kJ/mol analysis. Final plots utilize effective temperatures generated from the ± 2 kJ/mol data. Figure 3.4 represents the ODR analysis of ProVal with data plotted from the experiment and lines generated from the T_{eff} outputs of the program for the ± 2 kJ/mol uncertainty window. A final value for the experimental PA of ProVal is 987.0 ± 14 kJ/mol from the ODR analysis. The ODR analysis is in excellent agreement with the preliminary values indicating the appropriate choice of reference bases.

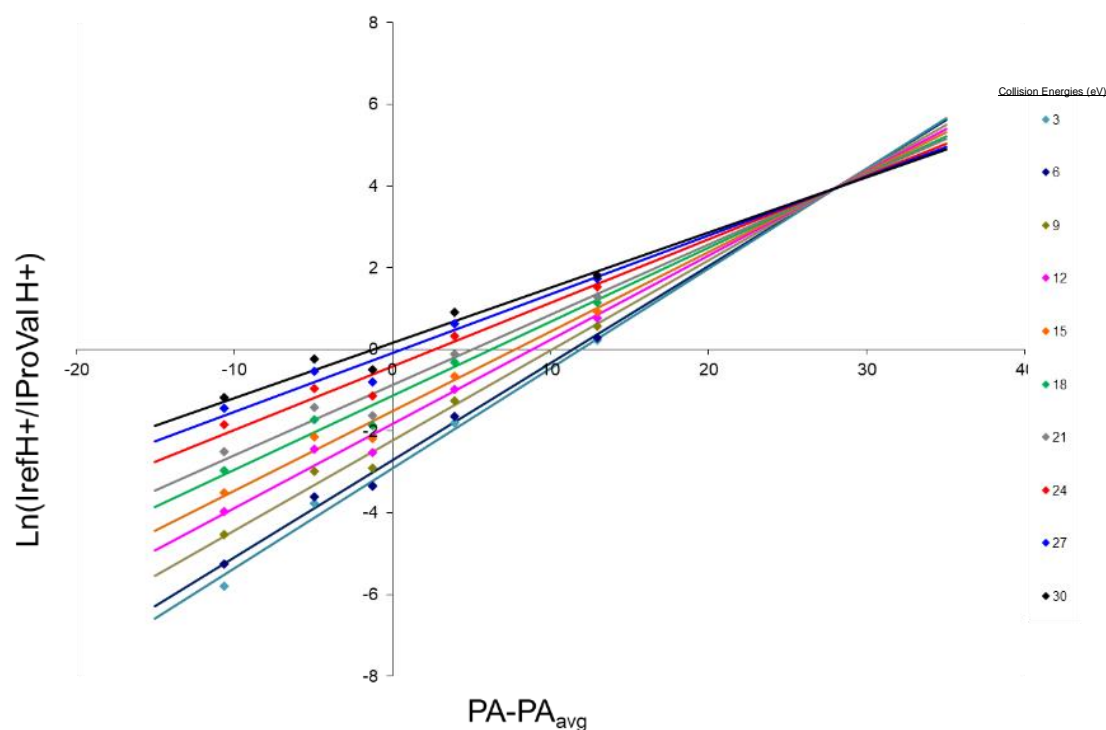


Figure 3.4: Orthogonal Distance Regression for ProVal predicts an ideal crossing point for each T_{eff} line. By plotting the outputs of the program as linear lines, it is possible to examine how the data fits the expected trend.

3.1.1 Example 2: ProSer

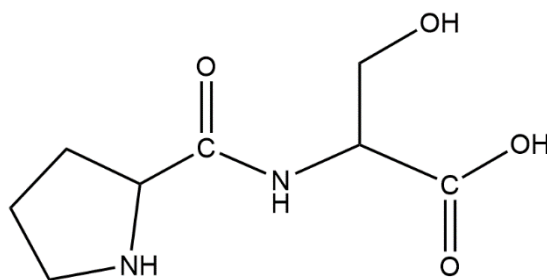


Figure 3.5: ProSer

A similar procedure was followed for the proline-containing dipeptide ProSer, depicted in Figure 3.5. A summary of the reference bases used for ProSer can be found in Table 3.2. Additionally, Plot 1 for ProSer can be seen in Figure 3.6. Higher collision energies for this analyte did not converge, which is attributed to possibly missing secondary fragments unaccounted for in the described ratios. Such fragments would be more abundant at the higher collision energies. These fragments have a small mass to charge that is not efficiently transferred by the region of stability within the quadrupoles to the detector leading to an absence of signal and errant ratios. Because of this, only data from collision energies of 3 to

Table 3.2: Reference Bases for ProSer detailing which reference bases were used in the extended kinetic method experiment for ProSer with their respective PAs.

Reference Base	Proton Affinity (kJ/mol)
Pyrrolidine	948.3
Piperidine	954.0
4-Tertbutylpyridine	957.7
2,4-Lutidine	962.9
1-Methylpyrrolidine	965.6

21 eV were used for analysis. Further justification for the omission of these data points can be seen in Plot 2.

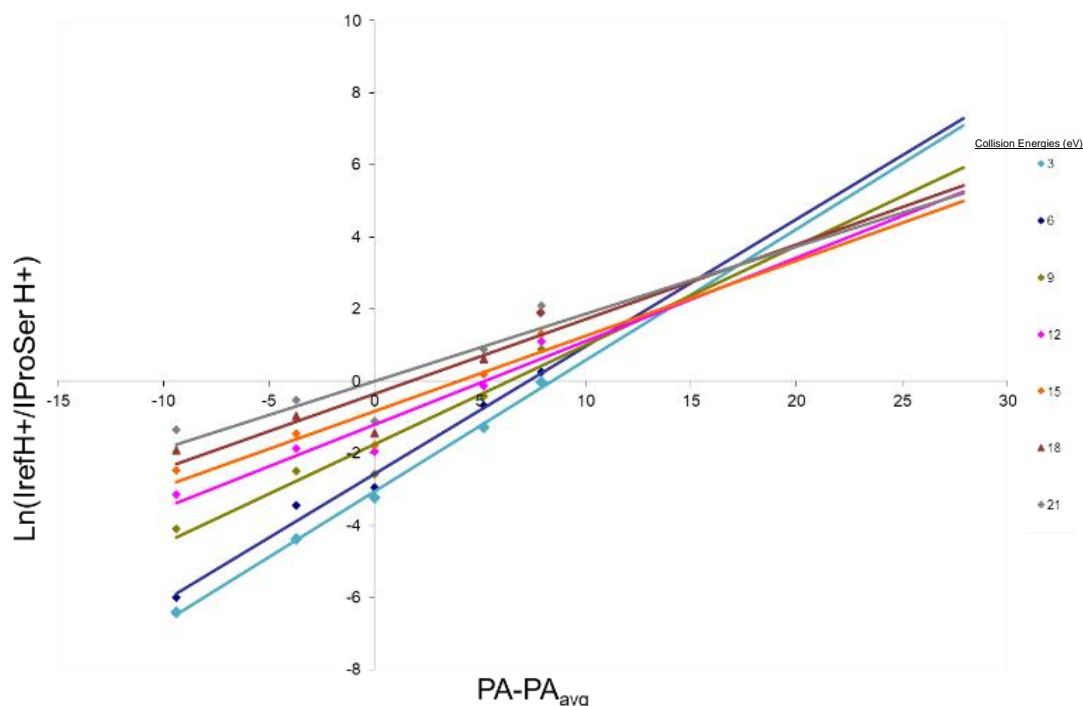


Figure 3.6: Kinetic Method Plot 1 for ProSer shows the experimentally collected data for ProSer. Only collision energies of 3-21 eV are shown.

The corresponding Plot 2 for the data in Figure 3.6 can be seen in Figure 3.7 where the negative y-intercept at each collision energy is plotted versus the slope. An r^2 correlation coefficient of 0.9575 relates these two metrics. A significant deviation from linearity can be observed for the points corresponding to collision energies of 24-30 eV. Because of their deviation, these points were omitted. While the r^2 of Plot 2 for ProSer is lower than that of ProVal, it was still deemed to be acceptable for further calculation. From the adjusted slope and intercept of Figure 3.7, a preliminary value of 973.0 kJ/mol was determined for the PA of ProSer.

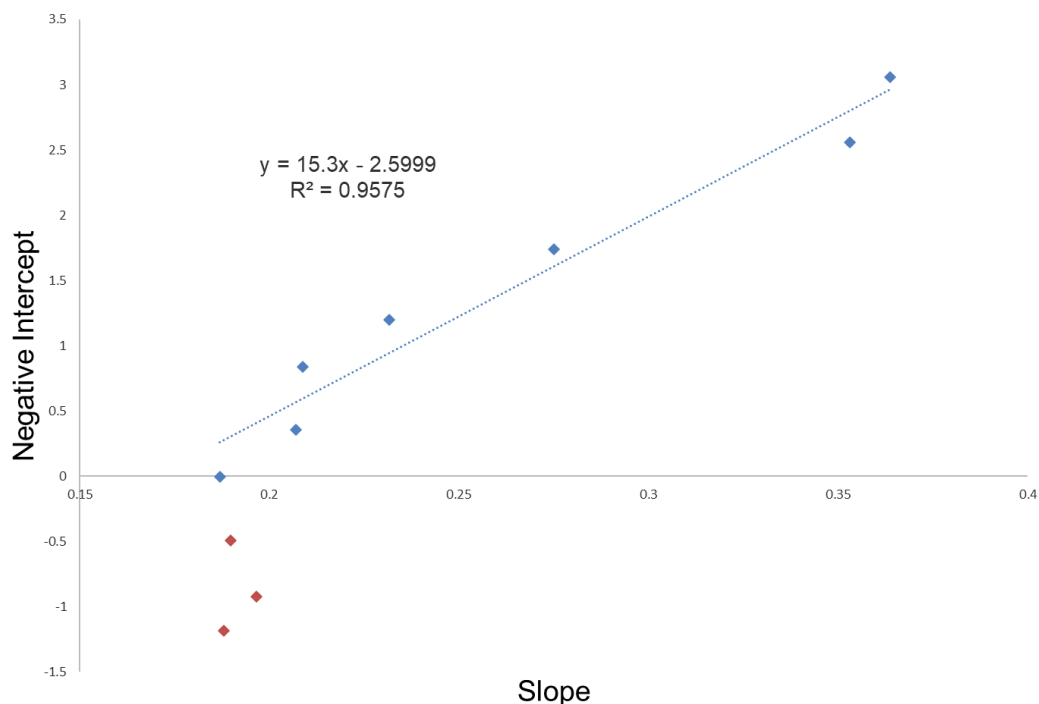


Figure 3.7: Kinetic Method Plot 2 for ProSer demonstrates the linear relationship between slope and the negative intercept at collision energies of 3-21 eV, and then deviations from linearity at higher collision energies. An R^2 of 0.9575 was determined for the selected data.

The preliminary values were then used as starting points for the ODR analysis. Figure 3.8 shows the ODR analysis of the data with the experimental data plotted as points. For this data set, the determined isothermal point is at a relatively high entropy point (y-value) of $-43.5 \text{ J mol}^{-1} \text{ K}^{-1}$ when multiplied by R . This could be due to the potential for hydrogen bonding from the side chain hydroxyl. From the analysis, a final value of $980.9 \pm 12 \text{ kJ/mol}$ for the PA of ProSer is suggested.

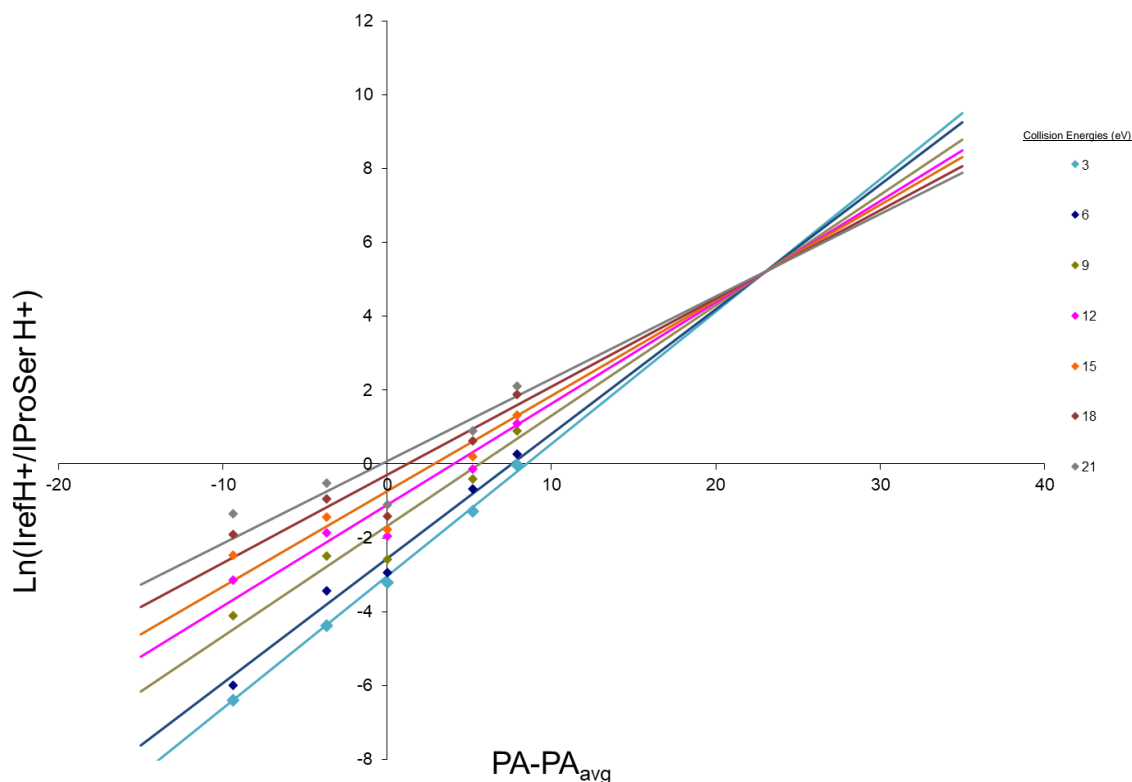


Figure 3.8: Orthogonal Distance Regression for ProSer predicts a highly precise isothermal point. A PA of 980.8 ± 12 kJ/mol is obtained for ProSer. A high entropic term can be seen with the high y-value.

3.2 Kinetic Method Results Summary

Similar procedures and analyses were applied to the other ProXxx dipeptides. A complete summary of the reference bases used for each dipeptide can be found in the Appendix. One consistency among the choice of reference base lies in the appearance of only a single basic site. By using bases that only have one basic site, it is possible to minimize differences in energetic considerations such as hydrogen bonding or other intra- and intermolecular interactions. Such interactions can introduce entropic considerations that may invalidate several assumptions made during the experiment. For example, if

multiple sites on a reference base could be readily protonated, then two different transition states may be measured confounding results. Table 3.3 shows a summary of the experimental values for each of the ProXxx dipeptides analyzed using the extended kinetic method.

Table 3.3: Summary of Extended Kinetic Method Data for the ProXxx dipeptides. Asterisks denote data collected by previous members of the lab.

Dipeptide	Experimental Proton Affinity (kJ/mol)	Dipeptide	Experimental Proton Affinity (kJ/mol)
ProAla*	974.7±15.5	ProLeu	983.5±12.8
ProArg	-----	ProLys	-----
ProAsn	985.8±10.2	ProMet	984.4±14.8
ProAsp	982.7±13.2	ProPhe*	975.1±16.0
ProCys	972.46±17.5	ProPro	996.1±9.5
ProGln	1010.8±16.8	ProSer	980.2±16.6
ProGlu	989.5±12.8	ProThr	981.6±12.7
ProGly	969.2±12.7	ProTrp	1000.3±11.2
ProHis	1003.2±12.6	ProTyr	1003.3±30
Prolle	987.8±16.5	ProVal	987.4±13.2

Two points to note include the missing values for ProLys and ProArg. While both dipeptides were investigated, a full analysis was unable to be completed for either. ProArg is expected to have a very high PA which makes it difficult to find reference bases within an appropriate range. The three reference bases of 1,1,3,3-tetramethylguanidine (TMG), 1,5-diazabicyclo[4.3.0]non-5-ene, and 1,8-diazabicycol[5.4.0]undec-7ene were investigated, but each either yielded insufficient fragmentation ratios for analysis or did not form sufficient heterodimer to isolate. ProLys could also not be sufficiently quantified, yet not from a lack of appropriate reference bases. Proper heterodimer formation was observed for ProLys, yet the fragmentation trends were often inconsistent either internally or

with each other. Additionally, only a small quantity of ProLys was available for analysis making it challenging to perform enough runs to obtain the required consistency of the experiment. It is thought that the challenges in the case of ProLys are due in part to its higher PA compared to other proline-containing dipeptides, as well as the high entropic factors inherent in its unique sidechain.

Lastly, the uncertainty for ProTyr is unexpectedly high which may be due to incomplete data collection. Further bases may need to be run to yield a more conclusive value. The rest of the data points agree with expected trends, and the reported values otherwise have similar degrees of uncertainties.

3.3 Computational Prediction Structural Motifs

3.3.1 ProGly

The lowest energy conformers for protonated and neutral ProGly were determined at the B3LYP/6-311++G**// B3LYP/6-31+G* level of theory. The lowest energy structures for neutral and protonated ProGly are displayed in Figure 3.9. Dark gray, light gray, blue, and red atoms correspond to carbon, hydrogen, nitrogen, and oxygen, respectively.

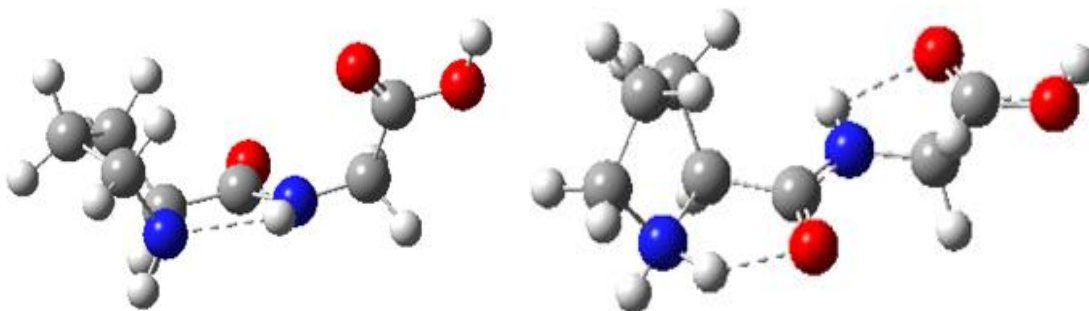


Figure 3.9: Lowest Energy Conformers of Neutral and Protonated ProGly. The neutral (left) conformer is structurally very similar to the protonated (right) conformer, as expected for the simplest dipeptide of the 39.

nitrogen, and oxygen respectively. Solid lines represent full covalent bonds, and dotted lines represent weaker intramolecular bonds such as hydrogen bonding. Generally, only four structural motifs are needed to distinguish ProGly conformers: protonation site, orientation of proline ring, amide bond orientation, and a description of the C-terminal. As the ProGly sidechain is only a hydrogen, there is little to be said about the orientation of this group. For the first motif, the proline ring can be oriented either in toward the amide bond, out away from the rest of the molecule, or in a twisted conformation. The lowest energy neutral structure of ProGly can be seen featuring the proline ring out, the amide bond trans, and the carbonyl group up. The lowest energy protonated structure on a near identical note adopts a proline ring in, a different trans amide bond, and carbonyl up structure. It is not surprising that the only difference between these two structures is a ring flip and the opposite trans amide bond orientation. Given the relative simplicity of ProGly, there are fewer distinct low energy structures to discuss among the conformers within 20 kJ/mol of the lowest energy conformer.

3.3.2 *ProVal*

The lowest energy conformers for protonated and neutral ProVal were determined at the B3LYP/6-311++G**// B3LYP/6-31+G* level of theory. The lowest energy structural motifs for neutral and protonated ProVal are displayed in Figure 3.10. To distinguish the different conformers of ProVal, five structural motifs can be assigned to each conformer in a similar manner as with ProGly. These motifs include the site of protonation for protonated structures, the orientation of the

proline ring, the trans/cis isomerism of the amide bond, a description of the carboxylic acid at the C-terminal, and a description of the isopropyl side chain. Looking at the lowest energy neutral for ProVal, it can be seen to have the proline ring out, a trans amide bond, carbonyl down with a syn proton, and the side chain up anti- to the nitrogen; however, the protonated structure is instead proline ring out, trans, with the C-terminal in plane with the molecule, and the side chain down.

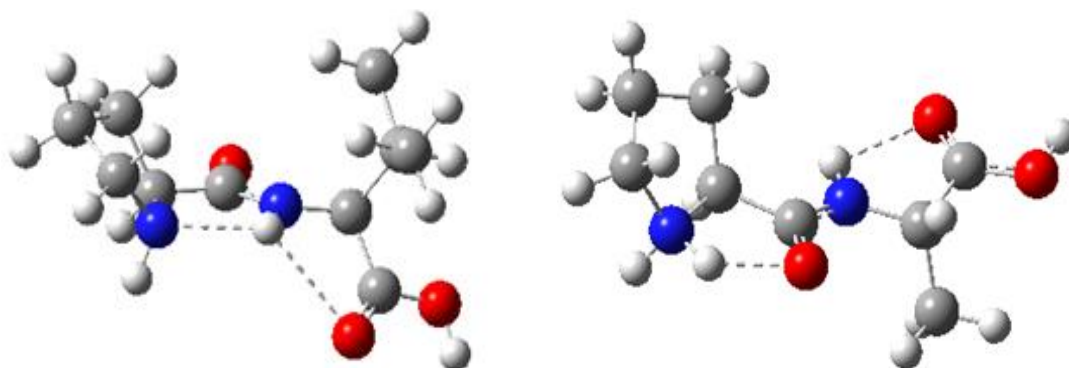


Figure 3.10: Lowest Energy Conformers of Neutral and Protonated ProVal. The neutral (left) conformer positions the side chain up, as compared to the protonated (right) which places the isopropyl group down. Yet, both cases provide a stabilizing hydrogen bond to the backbone nitrogen from the C-terminal.

Several motifs appear repeatedly across the lowest energy conformers. For the protonated ProVal conformers, the trans amide bond tends to dominate orienting the backbone carbonyl into proximity with the protonated N-terminal to form a hydrogen bond. In comparison, the neutral conformers form almost exclusively the opposite trans amide bond which puts the backbone nitrogen into proximity to hydrogen bond instead. The tendency of the protonated conformers to form a hydrogen bond to the carbonyl oxygen may be due to oxygen's increased electronegativity compared to nitrogen. An alternative explanation may be that for the neutral, the size match between the nitrogen to hydrogen to nitrogen bond is

more favorable. The protonated conformers did adopt the cis amide bond, but it is less commonly energetically favorable as it introduces more steric strain in the molecule. With regards to the side chain, the isopropyl group tends to almost always be placed “up” in the same direction as the proline group for the neutrals pushing the C-terminal down. For the protonated structures, the side chain appears to prefer the “down” orientation instead, possibly allowing better interaction between the C-terminal and the backbone nitrogen.

Some more complex hydrogen bonding does appear for higher energy ProVal conformers. One of the neutral conformers adopts a different trans amide bond orientation putting the oxygen closer to the N-terminal. Additionally, the C-terminal proton would adopt an anti- orientation to form a hydrogen bond to the backbone of the dipeptide. Depictions of these hydrogen bonding motifs can be found in Figure 3.11. Both ProGly and ProVal have relatively similar motifs across their various conformers given the nonpolar nature of their sidechains. Without other accessible nonpolar locations to interact with, the isopropyl group of ProVal

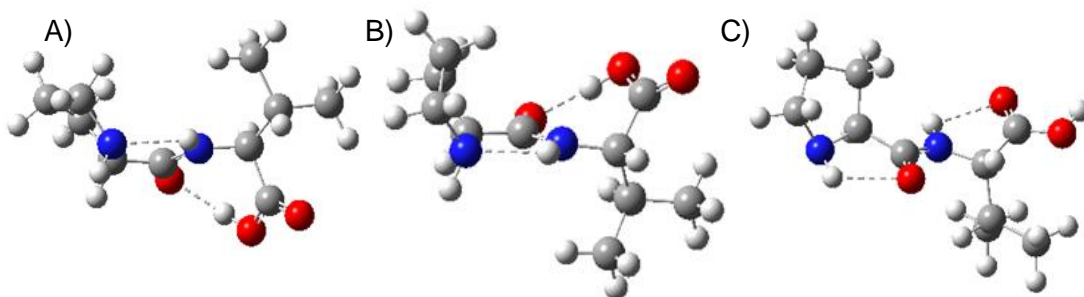


Figure 3.11: Alternative Structures for ProVal. A) Neutral conformer with new hydrogen bond from C-terminal to the backbone carbonyl. B) Neutral conformer with the C-terminal up and the side chain down, providing a less strained hydrogen bond to the backbone carbonyl. C) Neutral conformer with the protonated trans amide bond orientation, placing higher strain on the N-terminal.

and the hydrogen of ProGly can primarily reduce energy by lowering the strain that they may induce in the molecule.

3.3.2 *ProSer*

The lowest energy conformers for protonated and neutral ProSer were determined at the B3LYP/6-311++G**// B3LYP/6-31+G* level of theory to obtain the computational prediction for PA of ProSer. Five major structural motifs serve to differentiate most neutral and protonated ProSer conformers. The first four of these are identical to those of ProGly and ProVal: protonation site, the direction of the proline ring, the isomerism of the amide bond, and the orientation of the carboxylic acid at the C-terminal. The fourth point of demarcation lies in the orientation of the side chain. As compared to ProVal's aliphatic side chain, ProSer's hydroxyl chain can engage in additional hydrogen bonding adding complexity to the description of this dipeptide. A labeled description of the lowest energy protonated and neutral structures for ProSer can be seen in Figure 3.12. From the comparison of the two structures, it's possible to see how the addition of a proton facilitates increased hydrogen bonding. The two trans orientations identified for ProVal appear for ProSer as well, facilitated by a new hydrogen bonding interaction between the hydroxyl side chain and the backbone nitrogen.

For the 20 lowest energy neutral and protonated conformers, several of the motifs appear most often. The proton appears to strongly tend towards the N-terminal. Carbonyl-protonated conformers were generated, but they tended to remain higher in energy. It is likely that carbonyl protonation reduces the hydrogen

bonding potential of the side chain. Additionally, the protonated ProSer conformers tend to place the C-terminal in the “up” position most often. For the neutral there is no clear tendencies for the orientation of the same group. The side chain is almost always gauche to the backbone nitrogen for both neutral and protonated conformers. Often the side chain forms some new hydrogen bond for both sets of conformers as well.

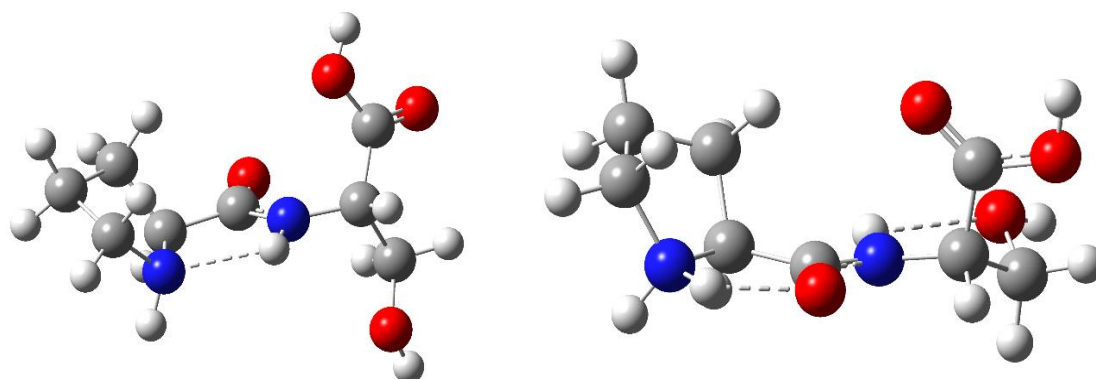


Figure 3.12: Lowest Energy Conformers of Protonated and Neutral ProSer. The neutral (left) conformer can be seen to have the proline ring out, the standard neutral trans bond, C-terminal up with a syn proton, and the side chain gauche to the backbone nitrogen. The protonated (right) conformer has the proton on the N-terminal with the proline ring out, the protonated standard trans bond, C-terminal up with a syn proton, and the side chain gauche to the backbone nitrogen.

There are also some deviations present in these 20 lowest energy conformers. While the trans amide bond still prevails, a different orientation for the trans amide bond has also appeared at these low energies for the neutral. Rather than the neutral forming a hydrogen bond between the proline N-terminal and the central amide, some conformers hydrogen bond to the carbonyl oxygen instead. While this structure was observed for the protonated state where the N-terminal has an additional proton, appearance for the neutral is unusual. The energetic

strain for the new trans bond is likely stabilized by additional hydrogen bonding interactions formed by the side chain. Additionally, some of the proline rings adopted a “twisted” conformer where the N-terminus is out of planarity with the rest of the proline ring. Lastly, some conformers did adopt anti dihedral angles that are only stabilized by increased intramolecular interactions such as a side chain anti to the backbone nitrogen in some neutrals and a C-terminal hydrogen anti to the carbonyl for some protonated conformers. Figure 3.13 depicts many of these deviations.

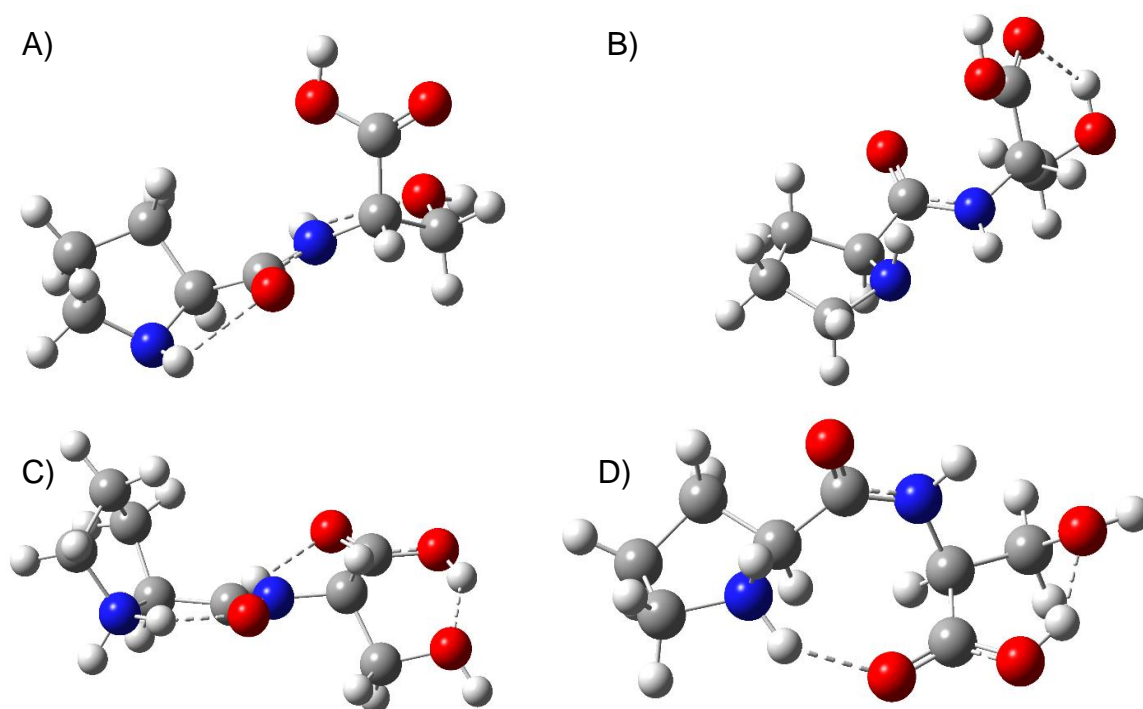


Figure 3.13: Alternative Structures for ProSer. A) Neutral conformer with new trans orientation. The structure is reminiscent of the protonated lowest energy conformer as it is stabilized by the side chain hydrogen bonding. B) Neutral conformer with twisted proline ring, as well as an anti- side chain. The twisted ring destabilizes the standard trans structure. C) Protonated conformer with proline in and C-terminal proton anti to the carbonyl to form a hydrogen bond to the side chain, anti to the backbone nitrogen. D) Protonated conformer with a cis amide bond, forming less common interactions.

3.4 Computational Proton Affinities

The remaining neutral and protonated ProXxx dipeptides were also analyzed using the B3LYP/6-311++G**// B3LYP/6-31+G* level of theory. B3LYP was chosen as the functional due to its suitability for thermochemical analysis. A complete summary of the computational predictions for the PA of the most basic site of each molecule can be seen in Table 3.4. Looking at the trends observed in the data, the predicted PAs correspond to the expected trends for each amino acid. For example, ProArg has an exceedingly high PA that is approximately 80 kJ/mol higher than most of the proline-containing dipeptides. On the other side of the data, ProGly is one of the least basic dipeptides as it features no basicity enhancing or stabilizing side chains. From testing with other basis sets, past molecules, and alternative functionals, the uncertainty in computed values is expected to lie between 6-10 kJ/mol. An exact value for uncertainty for each molecule is difficult to ascertain. Groups of dipeptides with similar side chain compositions tend to be closely grouped. An example of this is dipeptides with aliphatic secondary residues which only range from 966.0 to 986.6 kJ/mol.

Included in Table 3.4 is a summary of the PA of the backbone carbonyl for each dipeptide, as well as the PA of the zwitterionic conformers for dipeptides with basic side chains. The PA of these specific sites is determined from only the lowest energy conformer that had the proton on the relevant location. Understanding the PA of various points along the peptide backbone is important for better understanding the “proline effect” as it is often these sites at which fragmentation

may occur rather than solely the most basic site. Unfortunately, these values cannot easily be measured experimentally as the kinetic method is only sensitive to the most basic site of the molecule. From the table, the backbone carbonyl is not the most basic site for any of the ProXxx dipeptides, yet the trends are different than the standard PA values. For example, ProAsn is not a very basic proline-containing dipeptide with a PA of only 968.8 kJ/mol, yet its carbonyl with a PA of 961.0 kJ/mol is the 6th most basic carbonyl. These trends reveal how the specific intramolecular interactions of each dipeptide influence both the basicity of specific sites, as well as potential fragmentation patterns.

Table 3.4: Summary of Computational Predictions for PA of the ProXxx dipeptides. Asterisks denote that the most basic site of the dipeptide was the side chain. All reported values were isodesmically corrected using glycine as a reference. All values are in kJ/mol.

Dipeptide	PA (Most Basic)	PA (CO)	PA (zwitt)	Dipeptide	PA (Most Basic)	PA (CO)	PA (zwitt)
ProAla	970.0	935.8		ProLeu	974.8	947.0	
ProArg*	1063.5	975.4	1042.5	ProLys*	1014.2	962.1	986.9
ProAsn	968.8	961.0		ProMet	986.6	948.3	
ProAsp	979.2	948.9		ProPhe	979.0	951.5	
ProCys	970.0	927.8		ProPro	990.4	923.1	
ProGln	1004.5	978.0		ProSer	962.0	942.7	
ProGlu	995.8	964.2		ProThr	971.0	939.6	
ProGly	964.0	911.4		ProTrp	991.5	963.3	
ProHis*	1009	955.5	980.5	ProTyr	979.6	950.9	
Prolle	975.8	944.7		ProVal	975.2	943.8	

3.5 Dipeptide Summary

Table 3.5 contains both the computational results and the experimental results for each of the ProXxx dipeptides. Generally, the two sets of data agree with each other, especially once uncertainties are considered. Most PAs are within

± 10 kJ/mol of their counterparts, which is on a similar scale as many reactions. Additionally, as this study intends to primarily understand the role of these adjacent residues on the fragmentation of proline-containing peptides, the agreement is sufficient. Due to the agreement of these two sets of data, it is believed that the chosen basis set and density functional were appropriate for the study. As previously noted, several of the experimental values could not be sufficiently quantized from the experiment. The theoretical values for the ProArg and ProLys PAs do support that their high values may contribute to difficulties in the choice of reference base. Given the exceedingly high value of 1064 kJ/mol for ProArg, there

Table 3.5: Summary of Dipeptide PA for the ProXxx dipeptides. All calculated values are isodesmically corrected. Experimental values could not be determined for ProLys and ProArg.

Dipeptide	Experimental PA (kJ/mol)	Calculated PA (kJ/mol)	Dipeptide	Experimental PA (kJ/mol)	Calculated PA (kJ/mol)
ProAla	974.7 \pm 15.5	970.0	ProLeu	983.5 \pm 12.8	974.8
ProArg	-----	1063.5	ProLys	-----	1014.2
ProAsn	985.8 \pm 10.2	968.8	ProMet	984.4 \pm 14.8	986.6
ProAsp	982.7 \pm 13.2	979.2	ProPhe	975.1 \pm 16.0	979.0
ProCys	972.46 \pm 17.5	970.0	ProPro	996.1 \pm 9.5	990.4
ProGln	1010.8 \pm 16.8	1004.5	ProSer	980.2 \pm 16.6	962.0
ProGlu	989.5 \pm 12.8	995.8	ProThr	981.6 \pm 12.7	971.0
ProGly	969.2 \pm 12.7	964.0	ProTrp	1000.3 \pm 11.2	991.5
ProHis	1003.2 \pm 12.6	1009.0	ProTyr	1003.3 \pm 30	979.6
ProIle	987.8 \pm 16.5	975.8	ProVal	987.4 \pm 13.2	975.2

are few other compounds that are within a range to be used for reference. When the additional criteria of minimal hydrogen bonding locations is added, the list of potential candidates for ProArg is made even smaller. ProLys is also of a relatively high PA, being the second highest theoretical prediction in the data set.

One trend that does appear throughout the data exists in a slight discrepancy between the theory and the experiment. While the two sets are in general agreement, there are several cases where the theoretical numbers are lower than the experiment. The most notable examples of such a trend can be seen in the data of ProSer, ProTyr, and ProThr. One obvious similarity between each of these dipeptides is the presence of an oxygen in the sidechain of the C-terminal residue. One possible explanation as to the discrepancy could be an overcorrection of the calculations. If the selected basis set is overstabilizing the neutral of these dipeptides, then the PA for each would be smaller than expected as the neutrals would lie at a lower energy. Alternatively, if the cation is being understabilized, then the PA would similarly be too low due to the high energy term for the cation. Another possible explanation could lie in the manner through which the calculations are performed. While several different starting structures are tested and put through the minimization process to try to properly sample the PES of these dipeptides, there could be structures that were missed yet contribute significantly to the energy. Missing neutrals would further lower the PA, but missing cation structures could explain the discrepancies. Another potential explanation could lie in the experiment where we may be missing the complete entropic contribution due to the enhanced hydrogen bonding that arises from the side chain.

Lastly, it is possible that the only discrepancy is a matter of uncertainty as there is overlap between the two values when uncertainties are considered. Further studies are necessary to identify the specific cause for the discrepancy between the experimental and computational methods where there otherwise exists general agreement. Additional calculations are currently underway using alternative basis sets as well as alternative post HF methods to verify the accuracy of the computational predictions.

A full picture of how this data relates to the “proline effect” is still a work in progress. Additional data on the XxxPro dipeptides will be needed to compare how the N-terminal of proline, the site of preferential fragmentation, influences the proton affinity of these dipeptides with otherwise identical molecular composition. As the “proline effect” causes preferential fragmentation at the N-terminal, it is likely that the remaining 19 proline-containing dipeptides which place the proline at the amide bond will be of note. By more directly influencing the carbonyl as well as the sterics at the amide bond, it is likely that a more pronounced effect will be observed. With both the ProXxx and the XxxPro data sets, the differences that appear, and any consistencies that arise, will help to form a more complete understanding of peptide fragmentation.

Chapter 4: Conclusion

In conclusion, computational predictions for the PA of each ProXxx dipeptide were determined at the B3LYP/6-311++G**//B3LYP/6-31+G* level of theory. Additionally, the experimental PA of the same molecules were determined using the extended kinetic method using triple quadrupole mass spectrometry, except for previously determined PAs, ProArg, and ProLys. These values are restated in Table 4.1 with appropriate uncertainties reported. The experimental values range from 969.2 kJ/mol for ProGly to 1010.8 kJ/mol for ProGln, with the computational values for the same compounds ranging from 966.0 kJ/mol to 1004.5 kJ/mol. Computational predictions for ProLys and ProArg which could not be determined experimentally are 1014.0 kJ/mol and 1063.5 kJ/mol respectively. The computational and experimental values are in general agreement with each other.

Table 4.1: Summary of Dipeptide PA for the ProXxx dipeptides. All calculated values are isodesmically corrected. Experimental values could not be determined for ProLys and ProArg.

Dipeptide	Experimental PA (kJ/mol)	Calculated PA (kJ/mol)	Dipeptide	Experimental PA (kJ/mol)	Calculated PA (kJ/mol)
ProAla	974.7±15.5	970.0	ProLeu	983.5±12.8	974.8
ProArg	-----	1063.5	ProLys	-----	1014.2
ProAsn	985.8±10.2	968.8	ProMet	984.4±14.8	986.6
ProAsp	982.7±13.2	979.2	ProPhe	975.1±16.0	979.0
ProCys	972.46±17.5	970.0	ProPro	996.1±9.5	990.4
ProGln	1010.8±16.8	1004.5	ProSer	980.2±16.6	962.0
ProGlu	989.5±12.8	995.8	ProThr	981.6±12.7	971.0
ProGly	969.2±12.7	964.0	ProTrp	1000.3±11.2	991.5
ProHis	1003.2±12.6	1009.0	ProTyr	1003.3±30	979.6
ProIle	987.8±16.5	975.8	ProVal	987.4±13.2	975.2

While it is difficult to make a conclusive assessment regarding the influence of adjacent amino acids on the “proline effect” currently, it is evident that the adjacent amino acids do have an influence on the basicity and rigidity of proline-containing dipeptides. Further work on the remaining XxxPro dipeptides will seek to form a complete picture of the extent that adjacent amino acids influence the “proline effect”. Additional calculations shall also be performed to isolate where any discrepancies between theory and experiment may arise.

Lastly, several structural motifs have been identified for ProXxx dipeptides. Namely, neutral conformers appear to prefer the formation of a hydrogen bond between both the backbone and terminal nitrogen while protonated conformers instead hydrogen bond between the terminal nitrogen and the backbone carbonyl. Further investigation into the specific structures of each of the other dipeptides will identify further motifs that may aid in understanding the mechanism behind the “proline effect”. A holistic understanding of peptide fragmentation may remain on the horizon, but through studies such as this humanity is brought closer every day.

Appendix

Reference Bases														
	3-Picoline	Pyrridine	Piperidine	3,5-Lutidine	4-terbutylpyridine	N,N-diethylaniline	2,4-Lutidine	1-methylpyrrolidine	disopropylamine	triallylamine	disecbutylamine	triethylamine	2,2,6,6-tetramethylpiperidine	tripropylamine
Proton Affinity	943.4	948.3	954	955.4	957.7	959.8	962.9	965.6	971.9	972.3	980.7	981.8	987	991
Dipeptide	ProAsn		X		X		X	X	X		X			
	ProAsp				X		X	X	X					
	ProCys		X	X	X	X	X							
	ProGln								X	X		X	X	X
	ProGlu						X	X	X		X			
	ProGly	X	X	X	X									
	ProHis								X		X	X	X	X
	ProIle			X	X		X	X	X					
	ProLeu			X	X		X	X	X					
	ProMet			X	X		X	X	X					
	ProPro							X	X		X	X	X	
	ProSer		X	X	X		X	X						
	ProThr				X		X	X	X					
	ProTrp						X	X	X			X	X	
	ProTyr			X	X		X	X	X					
	ProVal		X	X	X		X		X					

Table A.1: Choice of reference bases for each of the ProXxx dipeptides

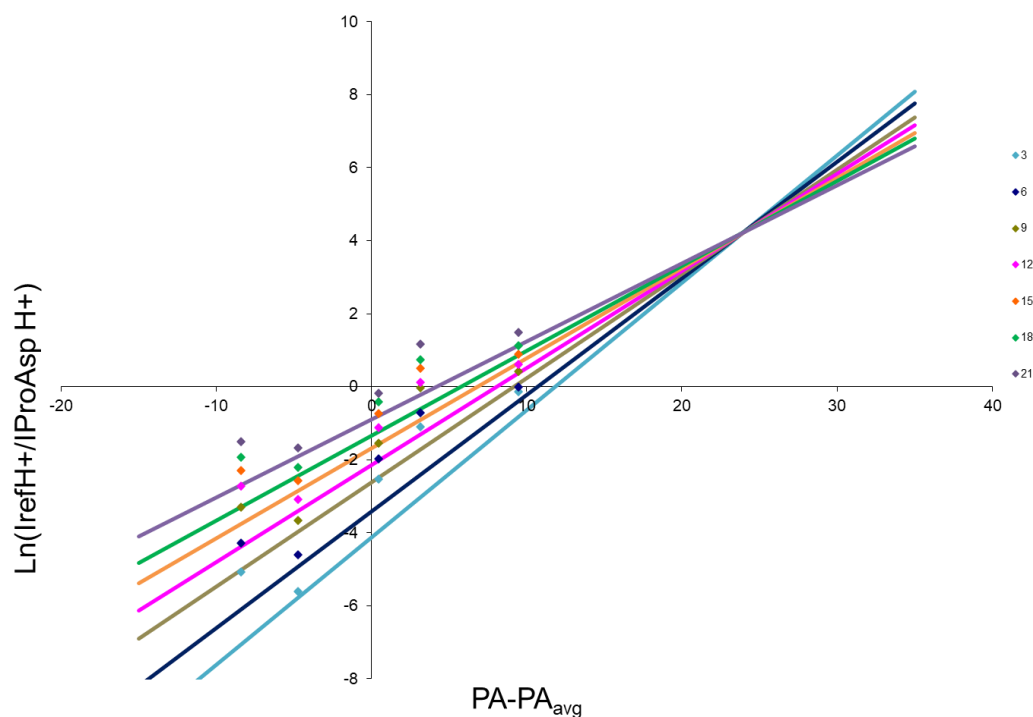


Figure A.1: ODR plot for ProAsp with collision energies 3-21 selected for quantification.

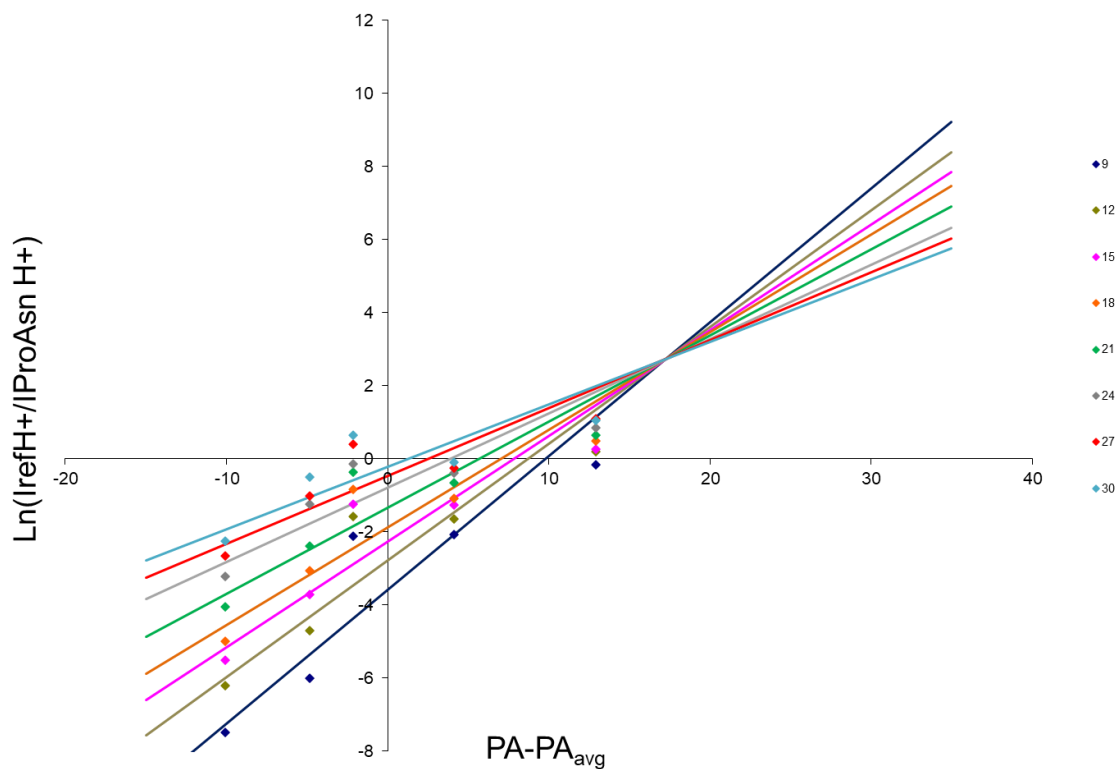


Figure A.2: ODR plot for ProAsn with collision energies 9-30 selected.

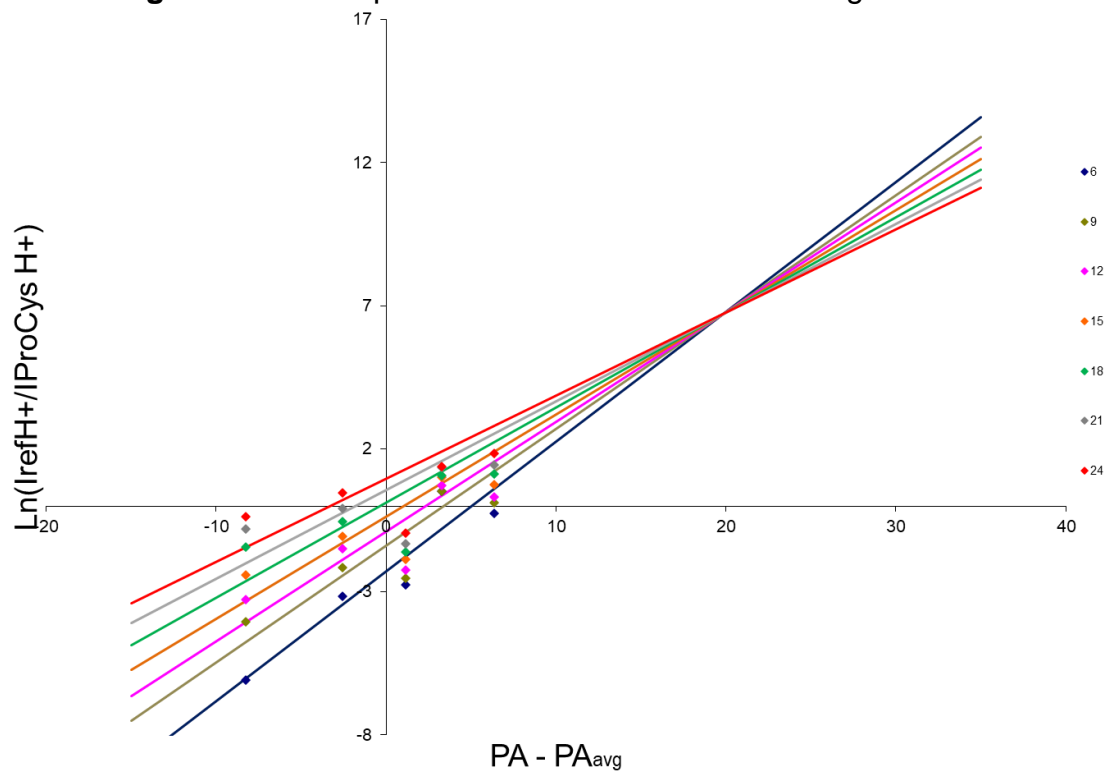


Figure A.3: ODR plot for ProCys with collision energies 6-24 selected.

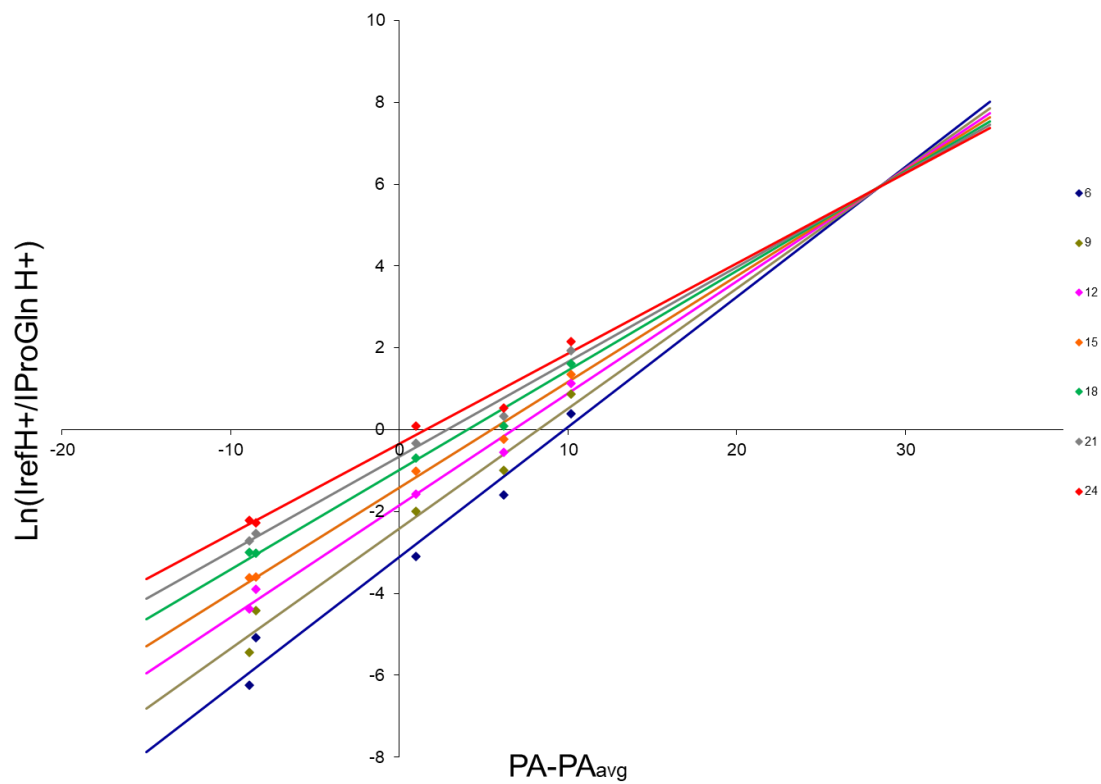


Figure A.4: ODR plot for ProGln with collision energies 6-24 selected.

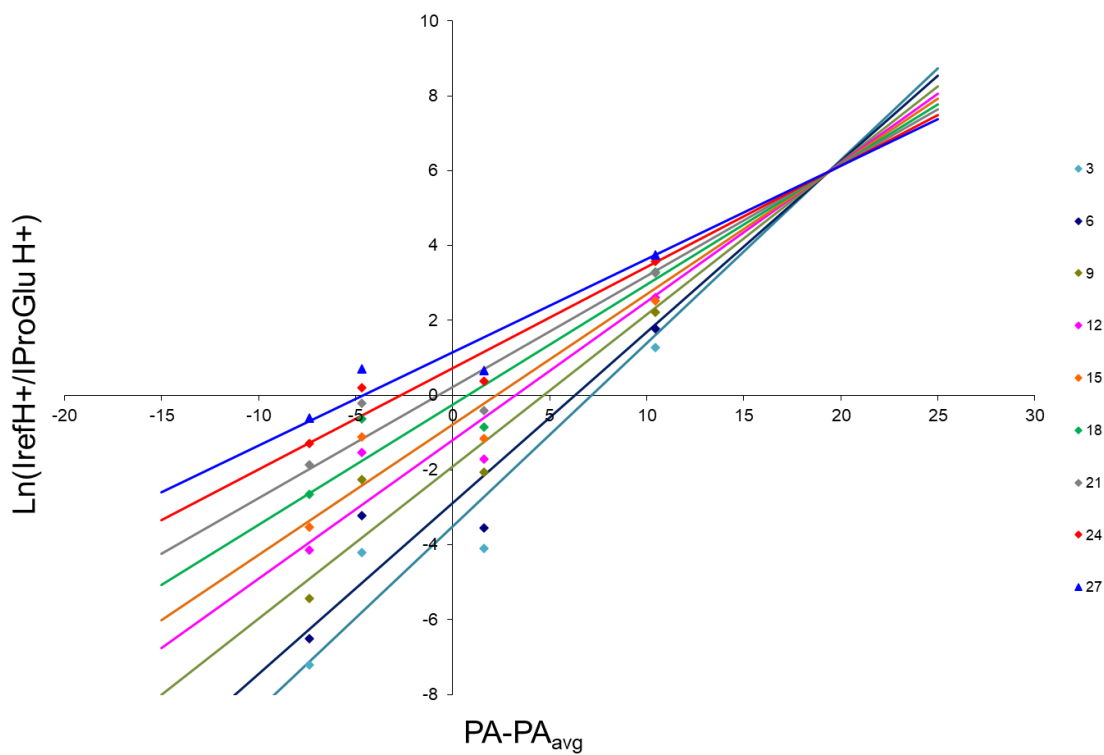


Figure A.5: ODR plot for ProGlu with collision energies 3-27 selected.

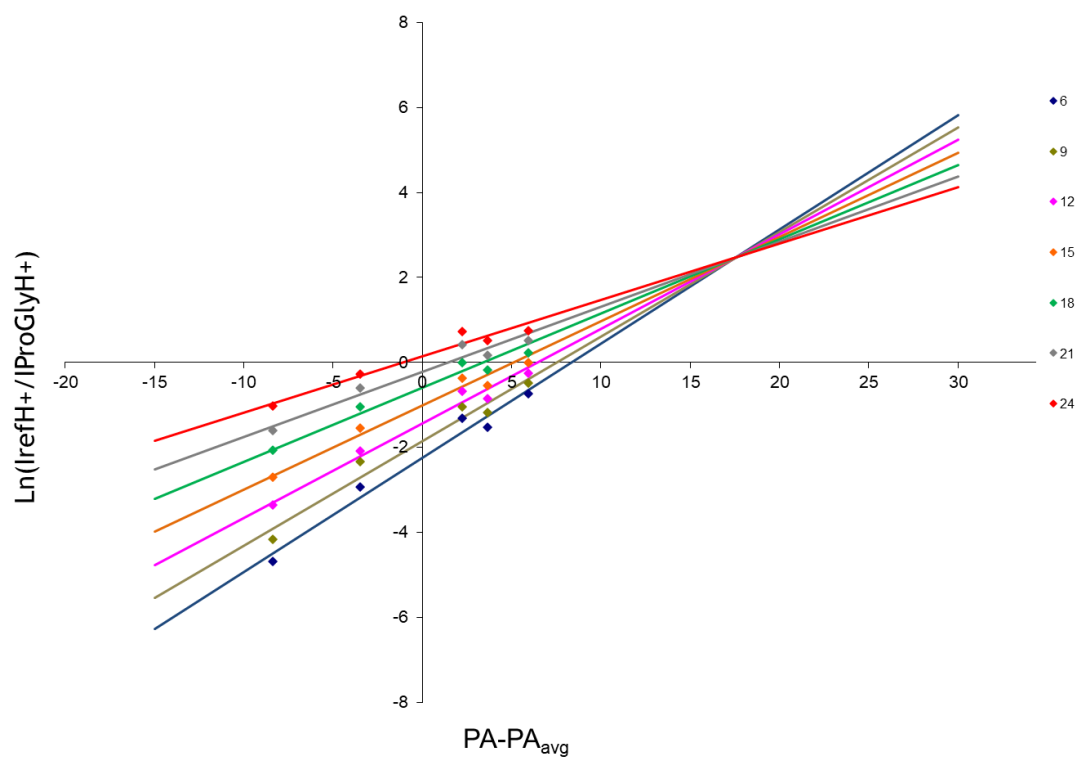


Figure A.6: ODR plot for ProGly with collision energies 6-24 selected.

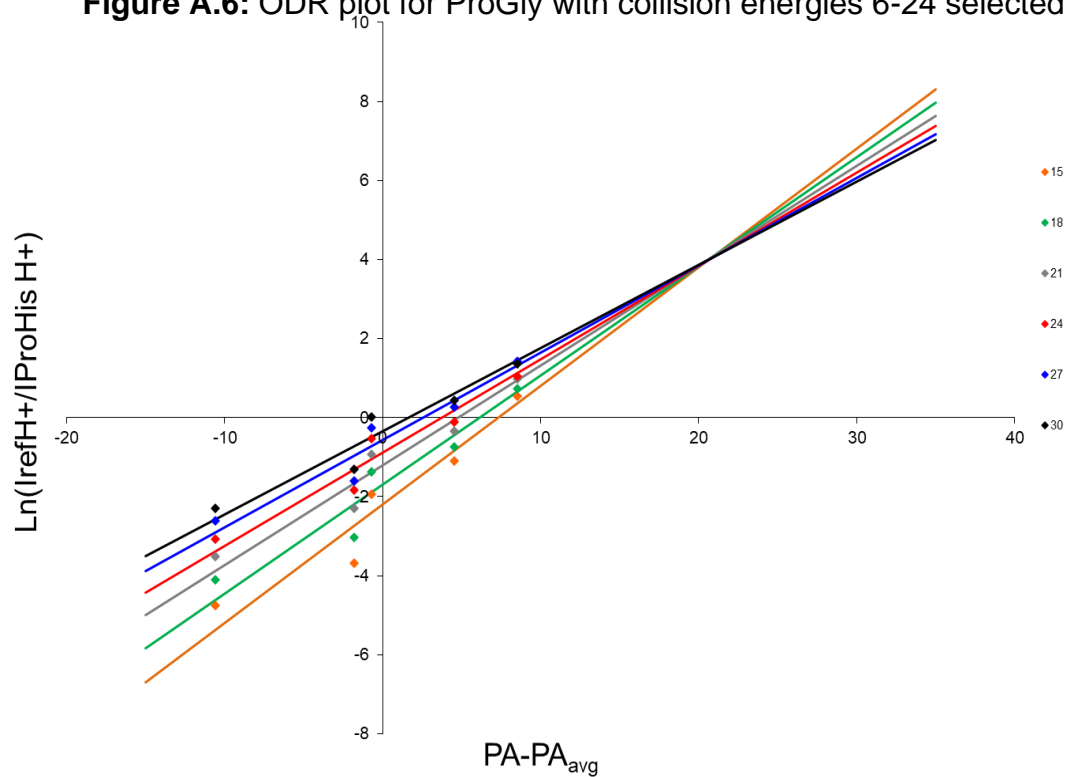


Figure A.7: ODR plot for ProHis with collision energies 15-30 selected.

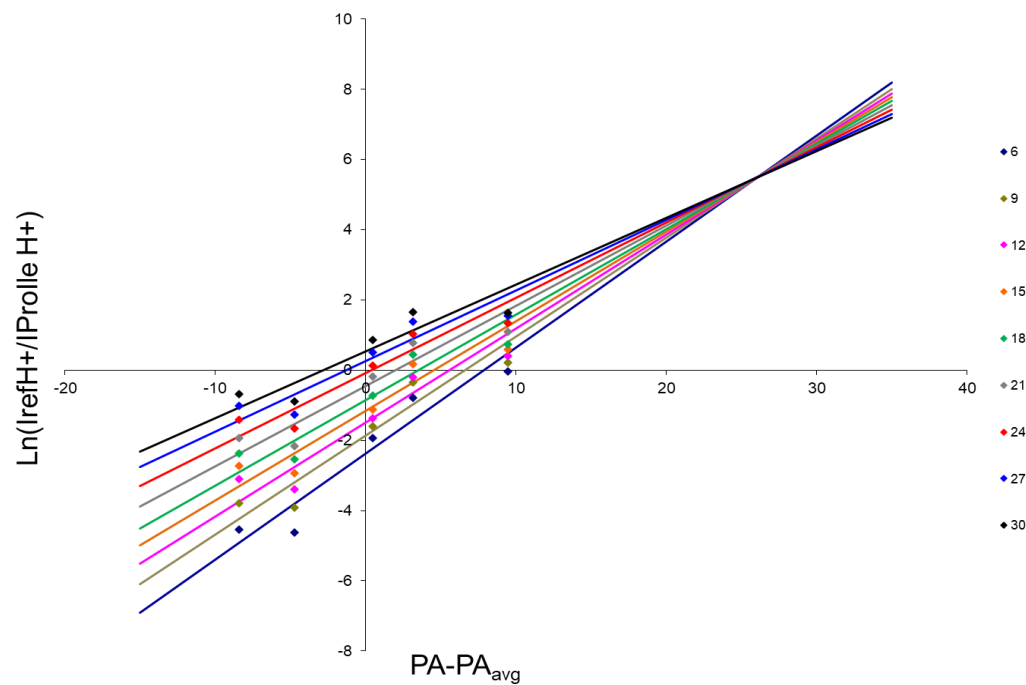


Figure A.8: ODR plot for Prolle with collision energies 6-30 selected.

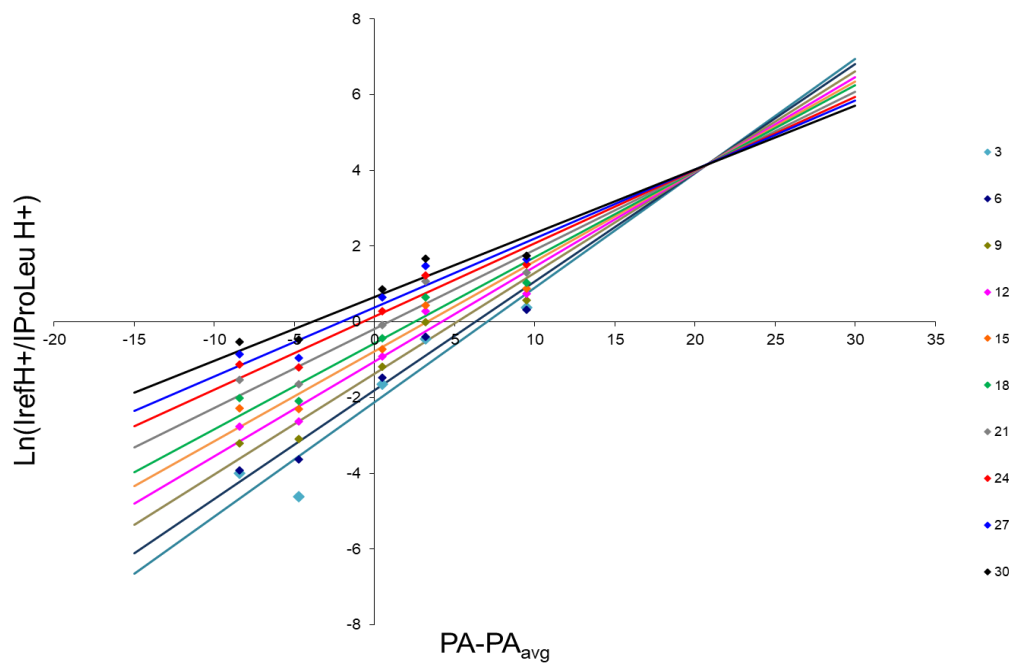


Figure A.9: ODR plot for ProLeu with collision energies 3-30 selected.

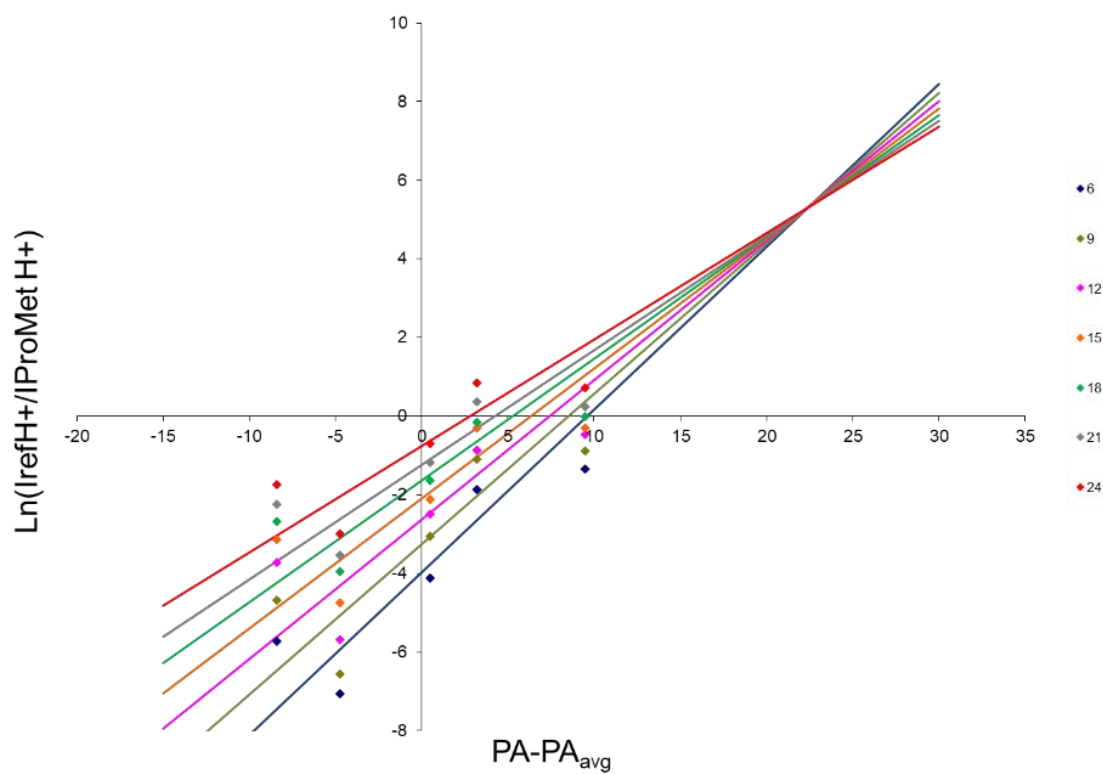


Figure A.10: ODR plot for ProMet with collision energies 6-24 selected.

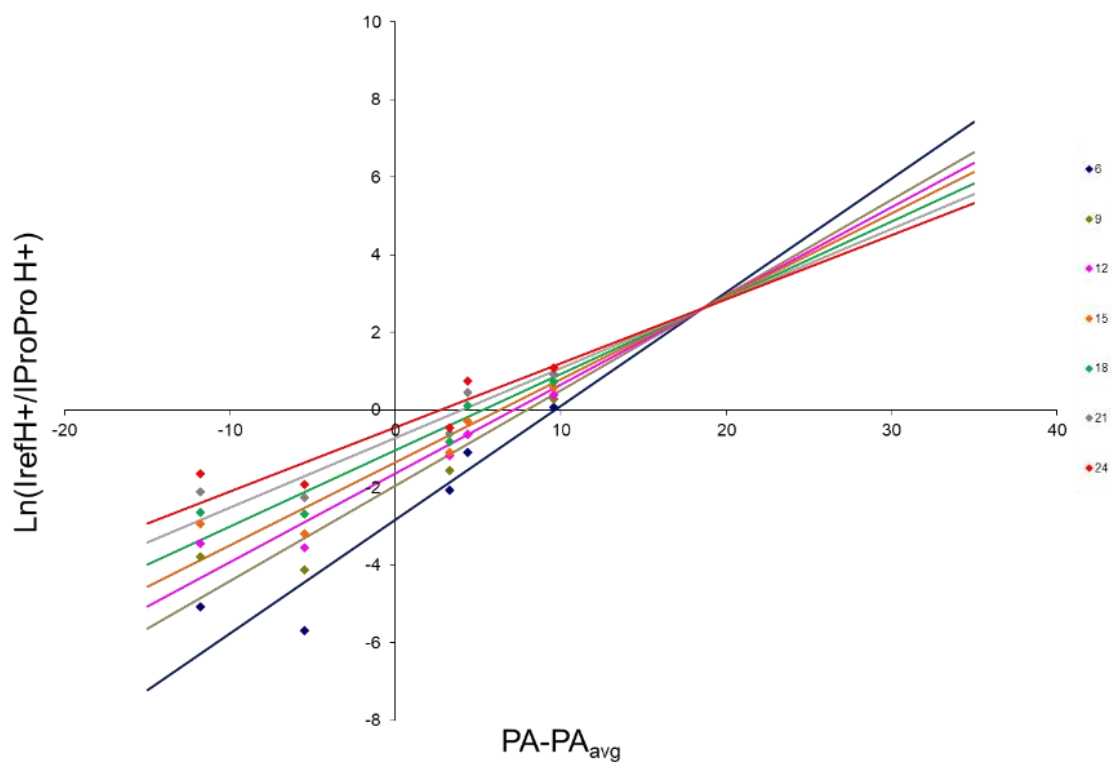


Figure A.11: ODR plot for ProPro with collision energies 6-24 selected.

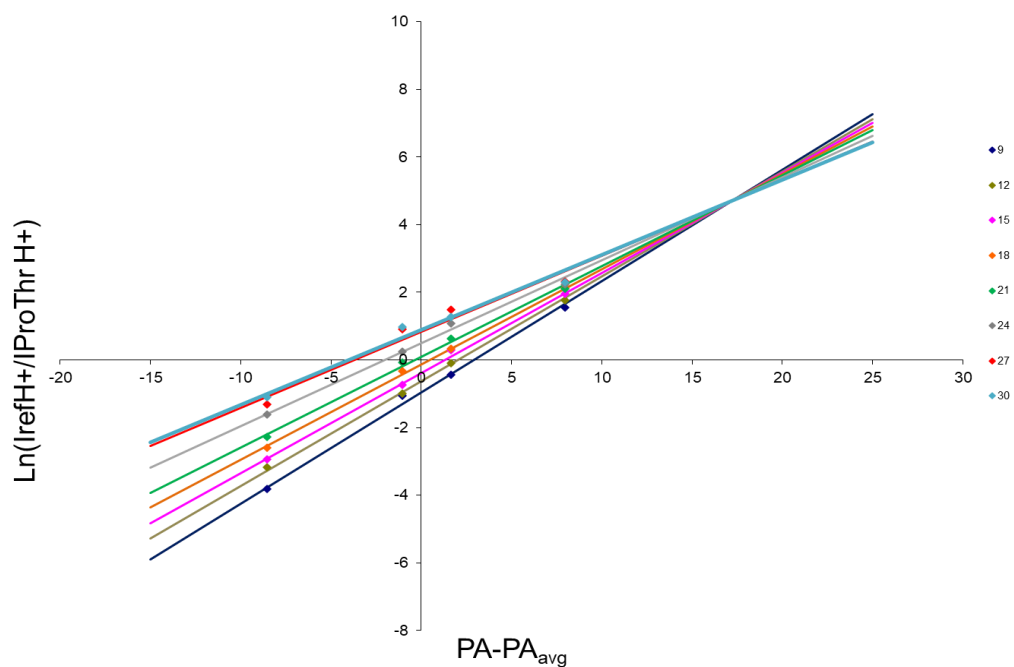


Figure A.12: ODR plot for ProThr with collision energies 9-30 selected.

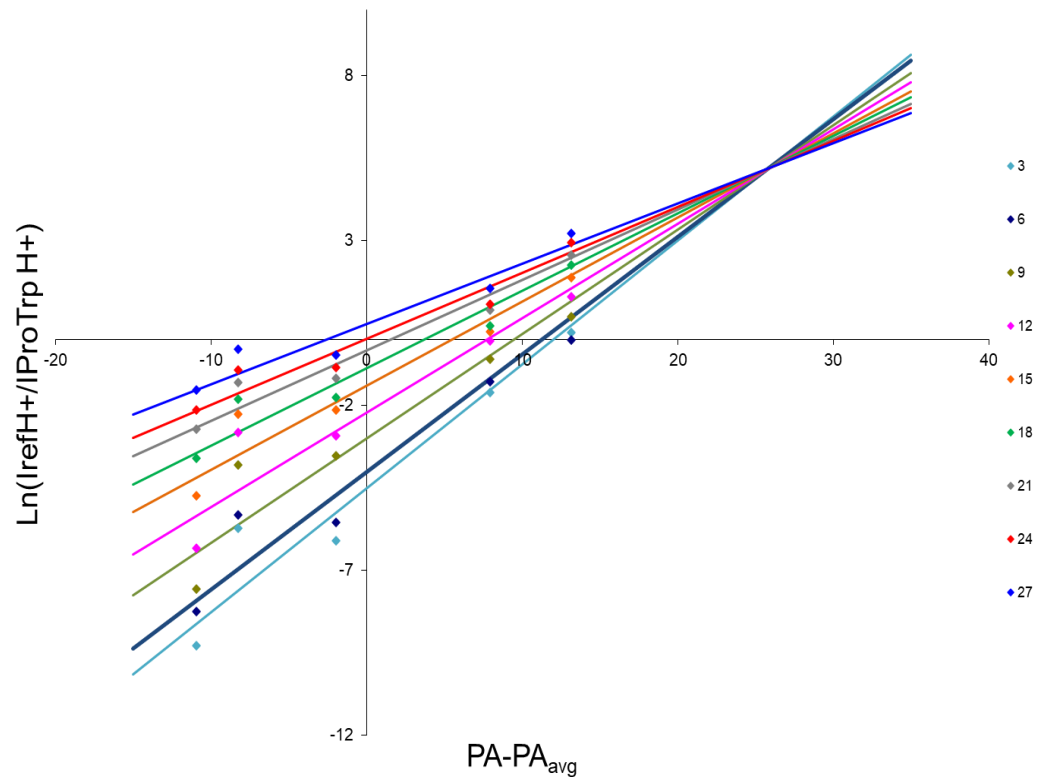


Figure A.13: ODR plot for ProTrp with collision energies 3-27 selected.

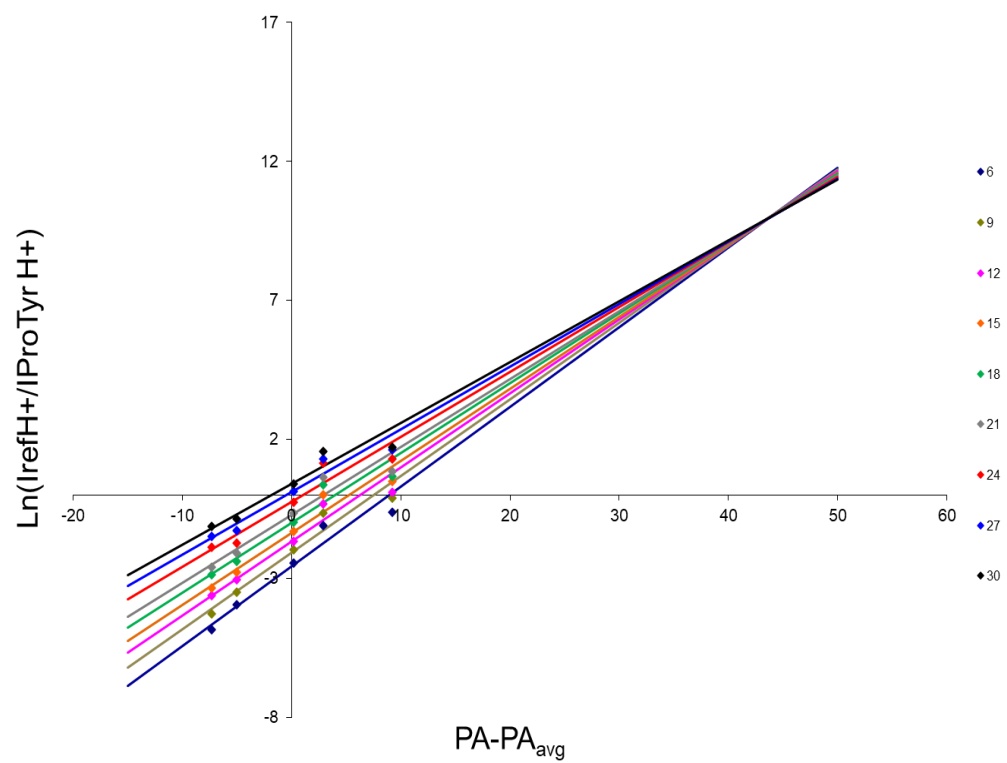


Figure A.14: ODR plot for ProTyr with collision energies 6-30 selected.

References

- (1) Lander, F.; Rogers, E. Finishing the Euchromatic Sequence of the Human Genome. **2004**.
- (2) Schnable, P. S.; Ware, D.; Fulton, R. S.; Stein, J. C.; Wei, F.; Pasternak, S.; Liang, C.; Zhang, J.; Fulton, L.; Graves, T. A.; Minx, P.; Reily, A. D.; Courtney, L.; Kruchowski, S. S.; Tomlinson, C.; Strong, C.; Delehaunty, K.; Fronick, C.; Courtney, B.; Rock, S. M.; Belter, E.; Du, F.; Kim, K.; Abbott, R. M.; Cotton, M.; Levy, A.; Marchetto, P.; Ochoa, K.; Jackson, S. M.; Gillam, B.; Chen, W.; Yan, L.; Higginbotham, J.; Cardenas, M.; Waligorski, J.; Applebaum, E.; Phelps, L.; Falcone, J.; Kanchi, K.; Thane, T.; Scimone, A.; Thane, N.; Henke, J.; Wang, T.; Ruppert, J.; Shah, N.; Rotter, K.; Hodges, J.; Ingenthron, E.; Cordes, M.; Kohlberg, S.; Sgro, J.; Delgado, B.; Mead, K.; Chinwalla, A.; Leonard, S.; Crouse, K.; Collura, K.; Kudrna, D.; Currie, J.; He, R.; Angelova, A.; Rajasekar, S.; Mueller, T.; Lomeli, R.; Scara, G.; Ko, A.; Delaney, K.; Wissotski, M.; Lopez, G.; Campos, D.; Braidotti, M.; Ashley, E.; Golser, W.; Kim, H.; Lee, S.; Lin, J.; Dujmic, Z.; Kim, W.; Talag, J.; Zuccolo, A.; Fan, C.; Sebastian, A.; Kramer, M.; Spiegel, L.; Nascimento, L.; Zutavern, T.; Miller, B.; Ambroise, C.; Muller, S.; Spooner, W.; Narechania, A.; Ren, L.; Wei, S.; Kumari, S.; Faga, B.; Levy, M. J.; McMahan, L.; van Buren, P.; Vaughn, M. W.; Ying, K.; Yeh, C. T.; Emrich, S. J.; Jia, Y.; Kalyanaraman, A.; Hsia, A. P.; Barbazuk, W. B.; Baucom, R. S.; Brutnell, T. P.; Carpita, N. C.; Chaparro, C.; Chia, J. M.; Deragon, J. M.; Estill, J. C.; Fu, Y.; Jeddeloh, J. A.; Han, Y.; Lee, H.; Li, P.; Lisch, D. R.; Liu, S.; Liu, Z.; Nagel, D. H.; McCann, M. C.; Sanmiguel, P.; Myers, A. M.; Nettleton, D.; Nguyen, J.; Penning, B. W.; Ponnala, L.; Schneider, K. L.; Schwartz, D. C.; Sharma, A.; Soderlund, C.; Springer, N. M.; Sun, Q.; Wang, H.; Waterman, M.; Westerman, R.; Wolfgruber, T. K.; Yang, L.; Yu, Y.; Zhang, L.; Zhou, S.; Zhu, Q.; Bennetzen, J. L.; Dawe, R. K.; Jiang, J.; Jiang, N.; Presting, G. G.; Wessler, S. R.; Aluru, S.; Martienssen, R. A.; Clifton, S. W.; McCombie, W. R.; Wing, R. A.; Wilson, R. K. The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* (1979) **2009**, 326 (5956), 1112–1115.
https://doi.org/10.1126/SCIENCE.1178534/SUPPL_FILE/SCHNABLE.SOM.PDF.
- (3) Griffiths, J. A Brief History of Mass Spectrometry. *Analytical Chemistry* **2008**, 80 (15), 5678–5683. <https://doi.org/10.1021/AC8013065>.
- (4) Angel, T. E.; Aryal, U. K.; Hengel, S. M.; Baker, E. S.; Kelly, R. T.; Robinson, E. W.; Smith, R. D. Mass Spectrometry-Based Proteomics: Existing Capabilities and Future Directions. *Chemical Society Reviews* **2012**, 41 (10), 3912–3928.
<https://doi.org/10.1039/C2CS15331A>.
- (5) Cox, K. A.; Gaskell, S. J.; Morris, M.; Whiting, A. Role of the Site of Protonation in the Low-Energy Decompositions of Gas-Phase Peptide Ions. *J Am Soc Mass Spectrom* **1996**, 7 (6), 522–531. [https://doi.org/10.1016/1044-0305\(96\)00019-0](https://doi.org/10.1016/1044-0305(96)00019-0).
- (6) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J Am Chem Soc* **1996**, 118 (35), 8365–8374. <https://doi.org/10.1021/JA9542193>.

- (7) Schwartz, B. L.; Bursey, M. M. Some Proline Substituent Effects in the Tandem Mass Spectrum of Protonated Pentaalanine. *Biological Mass Spectrometry* **1992**, 21 (2), 92–96. <https://doi.org/10.1002/BMS.1200210206>.
- (8) Grewal, R. N.; Aribi, H. el; Harrison, A. G.; Siu, K. W. M.; Hopkinson, A. C. Fragmentation of Protonated Tripeptides: The Proline Effect Revisited. *Journal of Physical Chemistry B* **2004**, 108 (15), 4899–4908. https://doi.org/10.1021/JP031093K/SUPPL_FILE/JP031093K_S.PDF.
- (9) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns. *Analytical Chemistry* **2005**, 77 (18), 5800–5813. https://doi.org/10.1021/AC0480949/SUPPL_FILE/AC0480949SI20050726_085631.PDF.
- (10) Raulfs, M. D. M.; Brechi, L.; Bernier, M.; Hamdy, O. M.; Janiga, A.; Wysocki, V.; Poutsma, J. C. Investigations of the Mechanism of the “Proline Effect” in Tandem Mass Spectrometry Experiments: The “Pipelicolic Acid Effect.” *J Am Soc Mass Spectrom* **2014**, 25 (10), 1705–1715. <https://doi.org/10.1007/S13361-014-0953-5>.
- (11) Muetterties, C.; Touzani, A. D.; Hardee, I.; Huynh, K. T.; Poutsma, J. C. Gas-Phase Acid-Base Properties of 1-Aminocycloalkane-1-Carboxylic Acids from the Extended Kinetic Method. *International Journal of Mass Spectrometry* **2015**, 378, 59–68. <https://doi.org/10.1016/J.IJMS.2014.07.010>.
- (12) Brechi, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. *Analytical Chemistry* **2003**, 75 (9), 1963–1971. <https://doi.org/10.1021/AC026359I>.
- (13) Nielsen, A. D.; Borch, K.; Westh, P. Thermochemistry of the Specific Binding of C12 Surfactants to Bovine Serum Albumin. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **2000**, 1479 (1–2), 321–331. [https://doi.org/10.1016/S0167-4838\(00\)00012-1](https://doi.org/10.1016/S0167-4838(00)00012-1).
- (14) Valley, C. C.; Cembran, A.; Perlmutter, J. D.; Lewis, A. K.; Labello, N. P.; Gao, J.; Sachs, J. N. The Methionine-Aromatic Motif Plays a Unique Role in Stabilizing Protein Structure *. *Journal of Biological Chemistry* **2012**, 287 (42), 34979–34991. <https://doi.org/10.1074/JBC.M112.374504>.
- (15) Hoffmann, E. de; Stroobant, V. Mass Spectrometry : Principles and Applications. 489.
- (16) Brauman, J.; Blair, L. Gas-Phase Acidities of Alcohols. *J Am Chem Soc* **1970**, 92 (20), 5986–5992.
- (17) Ervin, K. M. Experimental Techniques in Gas-Phase Ion Thermochemistry. *Chemical Reviews* **2001**, 101 (2), 391–444. <https://doi.org/10.1021/CR990081L>.
- (18) Cooks, G.; Kruger, T. Intrinsic Basicity Determination Using Metastable Ions. *J Am Chem Soc* **1976**, 99 (4), 1279–1281.

- (19) Cheng, X.; Wu, Z.; Fenselau, C. Collision Energy Dependence of Proton-Bound Dimer Dissociation: Entropy Effects, Proton Affinities, and Intramolecular Hydrogen-Bonding in Protonated Peptides. *J. Am. Chem. Soc* **1993**, *115*, 4844–4848.
- (20) Wu, Z.; Fenselau, C.; Graham Cooks, R. Gas-Phase Basicities and Proton Affinities of Lysine and Histidine Measured from the Dissociation of Proton-Bound Dimers. *Rapid Communications in Mass Spectrometry* **1994**, *8* (9), 777–780. <https://doi.org/10.1002/RCM.1290080923>.
- (21) Cerda, B. A.; Wesdemiotis, C. Binding to the DNA and RNA Nucleobases. Bond Energies and Attachment Sites from the Dissociation of Metal Ion-Bound Heterodimers. **1996**.
- (22) Armentrout, P. B. Entropy Measurements and the Kinetic Method: A Statistically Meaningful Approach. *J Am Soc Mass Spectrom* **2000**, *11* (5), 371–379.
- (23) Ervin, K. M.; Armentrout, P. B. Systematic and Random Errors in Ion Affinities and Activation Entropies from the Extended Kinetic Method. *Journal of Mass Spectrometry* **2004**, *39* (9), 1004–1015. <https://doi.org/10.1002/JMS.682>.
- (24) Chatfield, D. Christopher J. Cramer: Essentials of Computational Chemistry: Theories and Models. *Theoretical Chemistry Accounts* **2002**, *108* (6), 367–368. <https://doi.org/10.1007/S00214-002-0380-8>.
- (25) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals. *Journal of Physical Chemistry A* **2007**, *111* (42), 10439–10452. <https://doi.org/10.1021/JP0734474>.
- (26) Hunter, E.; Lias, S. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD 20899.
- (27) Hunter, E. P. L.; Lias, S. G. Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *Journal of Physical and Chemical Reference Data* **1998**, *27* (3), 413–656. <https://doi.org/10.1063/1.556018>.
- (28) PCModel. Serena Software 2020.
- (29) Frisch, M. J.; et. al. Gaussian. Gaussian, Inc.: Wallingford, CT 2009.
- (30) Becke, A. Density-Functional Thermochemistry III. The Role of Exact Exchange. *Journal of Chemical Physics* **1993**, *98*, 5648.
- (31) Tirado-Rives, J.; Jorgensen, W. L. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *Journal of Chemical Theory and Computation* **2008**, *4* (2), 297–306. https://doi.org/10.1021/CT700248K/SUPPL_FILE/CT700248K-FILE002.PDF.