

Exploring the Effects of Methylation on the CID of Protonated Lysine: A Combined Experimental and Computational Approach

Kenneth Lucas^{a, †}, Amy Chen^{b, †}, Megan Schubmehl,[†] Kristopher J. Kolonko,[‡] and George L. Barnes^{*, †}

[†]*Department of Chemistry and Biochemistry
Siena College
515 Loudon Road
Loudonville, NY 12211*

[‡]*Stewart's Advanced Instrumentation and Technology (SAInT) Center
Siena College
515 Loudon Road
Loudonville, NY 12211*

E-mail: gbarnes@siena.edu

^aPresently at SUNY Upstate Medical University; College of Medicine; 766 Irving Avenue; Syracuse, NY 13210

^bACS Project SEED High School Student; Presently pursuing an undergraduate degree at Northwestern University

Abstract

We report the results of experiments, simulations, and DFT calculations that focus on describing the reaction dynamics observed within the collision-induced dissociation of L-lysine- H^+ and its side-chain methylated analogues, N_ϵ -Methyl-L-lysine- H^+ (Me₁-lysine- H^+), N_ϵ, N_ϵ -Dimethyl-L-lysine- H^+ (Me₂-lysine- H^+), and $N_\epsilon, N_\epsilon, N_\epsilon$ -Trimethyl-L-lysine- H^+ (Me₃-lysine- H^+). The major pathways observed in the experimental measurements were m/z 130 and 84, with the former dominant at low collision energies and the latter at intermediate to high collision energies. The m/z 130 peak corresponds to loss of $N(CH_3)_nH_{3-n}$ while m/z 84 has the additional loss of H_2CO_2 likely in the form of H_2O+CO . Within the time frame of the direct dynamics simulations, m/z 130 and 101 were the most populous peaks, with the latter identified as an intermediate to m/z 84. The simulations allowed for the determination of several reaction pathways that result in these products. A graph theory analysis enabled the elucidation of the significant structures that compose each peak. Methylation results in the preferential loss of the side-chain amide group and a reduction of cyclic structures within the m/z 84 peak population in simulations.

1 Introduction

Post-translational modifications (PTMs) are both common and expand the functionality of proteins and peptides.¹⁻³ PTMs increase chemical complexity, which in turn modifies the interactions present within these species, resulting in important biological implications.⁴ Methylation and acetylation of lysine, in particular, are predominately

observed on histone proteins and can function as a biomarker for gene activation.⁵⁻⁸ Methylation can occur at either the N-terminus or the side-chain and has three distinct forms, namely mono-, di-, and tri-methylation. Identifying methylation via tandem mass spectrometry (MS2), specifically collision-induced dissociation (CID), is a frequently occurring task. Indeed, MS2 spectroscopy is a well-used tool for studying biological systems

using both experimental^{9–12} and theoretical, direct dynamics based studies.^{13–18} That said, trimethylation and acetylation of lysine result in a mass difference of just ~ 0.036 Da and requires high-resolution instruments to distinguish. Direct dynamics studies of methylation and acetylation can provide insight into the dynamics taking place during CID.

Direct dynamics simulations have shown great utility in elucidating atomic level information regarding dissociation mechanisms within MS2 spectroscopy.^{16,17} Typically, these simulations make use of semi-empirical treatments of the intramolecular potential energy surface and yet still yield good agreement with experiment. Moreover, the results from an ensemble of direct dynamics simulations performed using an appropriate sampling method provide an unbiased means of exploring the reaction pathways of interest and often identify unexpected reaction mechanisms. Subsequently, the energetics of these reaction pathways can be obtained via higher-level DFT or *ab initio* calculations, and hence provide significant insight into the dynamics taking place within experiment.

Although PTMs are common, only a few direct dynamics-based studies have focused on their role and influence on the mechanisms seen in CID. In addition, the PTM systems studied experimentally are often too large for a direct dynamics computational study. In this work, we seek to elucidate the impact of methylation on the observed CID mechanisms. To accomplish this goal, we performed both direct dynamics simulations and experimental measurements on L-lysine-H⁺ (lysine-H⁺), N_ϵ -Methyl-L-lysine-H⁺ (Me-lysine), N_ϵ, N_ϵ -Dimethyl-L-lysine-H⁺ (Me₂-lysine), and $N_\epsilon, N_\epsilon, N_\epsilon$ -Trimethyl-L-lysine-H⁺ (Me₃-lysine) to allow for a direct comparison between results.

As mentioned above, direct dynamics simulations are a valuable tool with a significant history of providing information regarding MS2 systems. Recently, several groups have started to use graph theory to analyze these simulations.^{14,19–22} The application of graph theory to chemistry is well documented,²³ and its application within direct dynamics simulations typically begins with the definition of the adjacency matrix, which is closely related to what one of the authors termed the connectivity matrix in previous work.²⁴ There

are two primary means of obtaining the adjacency matrix from simulation data, namely the use of distance parameters to determine bonding^{20–22} or the direct use of the bond order obtained from the quantum mechanical wavefunction used within the dynamics.^{14,19} With the adjacency matrix in hand, a graph analysis allows for a determination of the number and types of fragments obtained, which we employ here to analyze the diversity of structures within each mass spectrum peak. This classification further simplifies the identification of the relevant mechanisms within the simulations.

An outline for the remainder of the paper is as follows: in Section 2, we provide an overview of our computational and experimental methods; in Section 3, we present our results and compare the CID pathways between lysine-H⁺ and Me_n-lysine-H⁺; and in Section 4, we provide an overview of our findings.

2 Methods

Below we present our computational and experimental approach to studying L-lysine-H⁺ (lysine-H⁺), N_ϵ -Methyl-L-lysine-H⁺ (Me₁-lysine-H⁺), N_ϵ, N_ϵ -Dimethyl-L-lysine-H⁺ (Me₂-lysine-H⁺), and $N_\epsilon, N_\epsilon, N_\epsilon$ -Trimethyl-L-lysine-H⁺ (Me₃-lysine-H⁺).

2.1 Computational Approach

The literature contains several examples of the use of direct dynamics simulations to investigate MS2 systems. Moreover, there have been recent perspectives written on its use^{13,16,25} as well as a tutorial review on the methods involved.¹⁶ We refer the interested reader to those works for specific details regarding the approach and provide specific information for our systems of interest below.

2.1.1 Structures and Simulation Method

We generated initial structures for lysine-H⁺ and Me_n-lysine-H⁺ (n=1-3) with the methylation occurring on the side-chain nitrogen using Avogadro.²⁶ The charge of each species was +1, which was accomplished by placing the excess proton

on the side-chain for lysine-H⁺ and Me₁₋₂-lysine-H⁺. Me₃-lysine-H⁺ has a +1 charge without an additional proton. These species are sufficiently large that it would be haphazard to trust human intuition to find the global potential energy minimum. Hence, we employed a simulated annealing approach. In particular, the species were imported into GROMACS²⁷ where an initial geometry optimization was performed, followed by a 300 K equilibration. The simulated annealing procedure commenced using six heat-cool cycles in which the temperature was ramped to 1000 K over 100 ps held at 1000 K for 100 ps and then cooled to 0 K over 100 ps. The resulting 0 K structures were optimized at the RM1 level as implemented in Mopac2016.²⁸ The molecular mechanical potential for GROMACS was based on the Gromos 54A7²⁹ potential and then modified using the Automatic Topology Builder (ATB)³⁰⁻³² to accommodate both the methylation and charged nature of each species.

Direct dynamics simulations were performed using the RM1 semi-empirical method³³ to calculate the potential energy surface. RM1 has been shown to produce good results for numerous previous simulations involving protonated peptides.^{18,24,25,34,35} Although the potential energy parameters for Ar CID given in the literature³⁶ could have been used for the PTM systems, we instead made use of the microcanonical sampling scheme to impart a statistical distribution of internal energy.³⁷ Simulations obtained population in nearly all experimentally observed peaks, which suggests that sudden or fast fragmentation events¹⁹ are not that significant for these systems. A range of internal energies was considered for each system with lysine-H⁺, Me₁-lysine-H⁺, and Me₂-lysine-H⁺ having values of 250 (1046), 300 (1255.2), and 350 (1464.4) kcal/mol (kJ/mol), and Me₃-lysine having values of 300 (1255.2), 350 (1464.4), 400 (1673.6) kcal/mol (kJ/mol). Four hundred trajectories were performed with randomly selected initial conditions for each specified internal energy. The particular internal energy values were chosen to allow for significant reactivity to explore the chemical space without producing too many low-mass products. During the analysis, the products are considered either collectively or as a function of internal energy. In general, re-

activity increased with increasing internal energy within the time frame of the simulations.

Hamilton’s equations of motion were solved using a 6th order symplectic integration scheme,³⁸ for a maximum simulation time of 50 ps with a 1 fs step size and output written every 50 fs using our in-house simulation package tightly coupled with Mopac2012.³⁹ As we were focused on the charged fragments, neutral fragments were removed from the simulation if they were at least 15 Å away from any charged fragment. In addition, a trajectory was halted early if the final charged fragment had an $m/z \leq 60$. Conservation of energy was excellent for all trajectories.

2.1.2 Theoretical Mass Spectra and Reaction Mechanisms

As described previously,^{14,40} our in-house simulation software obtains the bond order between all QM atoms as a function of time throughout the simulation. These bond orders are averaged over a 5 fs window to reduce occurrences of momentary fluctuations. The averaged bond order matrix is converted to a connectivity matrix according to

$$C_{ij} = \begin{cases} 1, & \text{if } B_{ij} \geq B_{cut} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where i and j are atom indexes, C_{ij} is the connectivity matrix, B_{ij} is the averaged bond-order matrix, and $B_{cut} = 0.7$ is the threshold for atoms to be considered connected. Note that diagonal elements of this matrix are zero. C_{ij} allows for an on-the-fly connectivity analysis, which provides an initial preview of the reactive events. In post analysis, each C_{ij} is available and used to form $C_{ij}(t)$, the time-dependent connectivity matrix for the entire trajectory. Analysis of $C_{ij}(t)$ yields the times of each fragmentation event that takes place within a trajectory. Momentary bond stretches can result in a bond order that is below B_{cut} . Such fragmentation events “recombine” quickly – within 150 fs – and are not considered in the subsequent analysis. C_{ij} was also analyzed to determine the times at which protons move within the molecule. Correlating the times for proton motion and fragmentation allows for a determination of which proton transfers influence bond cleavage events. In par-

ticular, a proton hop is considered relevant for the A-B bond cleavage if a proton moved to or from any heavy atom directly bound to either A or B or A and B themselves.⁴⁰

The information from $C_{ij}(t)$ allowed for straight-forward identification of final products for each trajectory. The matrix C'_{ij} , which includes only atoms in the charged fragment, is formed using information from the last time step of each trajectory. We note here that our connectivity matrix is equivalent to the adjacency matrix of a graph, and hence analysis techniques from graph theory can be applied to it. We have adapted the method of Vazquez *et al*^{20,21} and define a modified adjacency matrix as

$$A_{ij} = \begin{cases} 1 + \frac{Z_i}{10}, & \text{if } i = j \\ C'_{ij}, & \text{otherwise} \end{cases} \quad (2)$$

where Z_i is the atomic number of the i^{th} atom. This modification results in a permutation invariant matrix, and the lowest eigenvalue from this matrix identifies unique structures within each final product. With unique structures for each m/z identified, their relative abundance was easily obtained.

Hence, the direct dynamics simulations combined with connectivity analysis and graph theory allow for easy identification of the most important products and the mechanistic steps taken to reach them. With this information on hand, structures and energetics of relevant minima, intermediates, and transition states were obtained for the lysine- H^+ system at the $\omega\text{B97X-D/aug-cc-pVTZ}$ level of theory as implemented in the Gaussian09 software package.⁴¹ We expect that these energy profiles would be similar for the methylated analogues. $\omega\text{B97X-D}$ is a long-range corrected, hybrid density functional that includes atom-atom dispersion correction.⁴² We note here that for several pathways, the proton transfer occurred via a monotonically increasing energy profile, i.e., a definitive transition state could not always be located.

2.2 Experimental Approach

High purity samples of L-lysine, N_ϵ -Methyl-L-lysine, N_ϵ, N_ϵ -Dimethyl-L-lysine, and $N_\epsilon, N_\epsilon, N_\epsilon$ -Trimethyl-L-lysine were obtained from Sigma Aldrich and used without further purification.

Deuterium oxide (≥ 99.96 atom % in D) and L-lysine (α - ^{15}N) was obtained from Cambridge Isotope Laboratories (Cambridge, MA). All samples were prepared in mass spectrometry grade, 1:1 mixture of acetonitrile and water purchased from Sigma Aldrich. High-resolution mass spectrometry data were acquired on an unmodified, commercially available Bruker Maxis Impact HD (Quadrupole-Time of Flight) spectrometer using sodium formate solution as the mass calibrant and the standard heated electrospray ionization (ESI) source. Pseudo MS3 experiments were performed utilizing in-source fragmentation. Typical collision energies considered were 5, 10, 15, 20, 25, and 30 eV using nitrogen within the CID cell. Additional experimental details are provided in the supporting information.

3 Results and Discussion

An overview of all simulated and experimental data is provided in Figure 1, which includes example spectra and break-down curves for all systems. Additional experimental data is provided in the supporting information. The most prominent experimental peaks for each system are provided in Table 1, 16 out of 20 of which are observed in the direct dynamics simulations.

Figure 1 shows that at low and intermediate collision energies, a mixture of significant peaks is observed. At high collision energy, the m/z 84 peak ($(\text{C}_5\text{H}_{10}\text{N})^+$) is dominant in all species. From this experimental observation, it seems likely that a common mechanistic pathway exists for lysine- H^+ and its methylated analogues that lead to the m/z 84 peak. Hence we will begin with an analysis of the lysine- H^+ system since a good understanding of the reaction mechanisms responsible for the significant experimental peaks observed in that system will serve as a framework for understanding the methylated analogues.

3.1 Lysine- H^+

The experimental measurements find that in lysine- H^+ ($(\text{C}_6\text{H}_{15}\text{N}_2\text{O}_2)^+$), the two most significant peaks are m/z 84 ($(\text{C}_5\text{H}_{10}\text{N})^+$) and 130 ($(\text{C}_6\text{H}_{12}\text{NO}_2)^+$). This finding is in agreement with

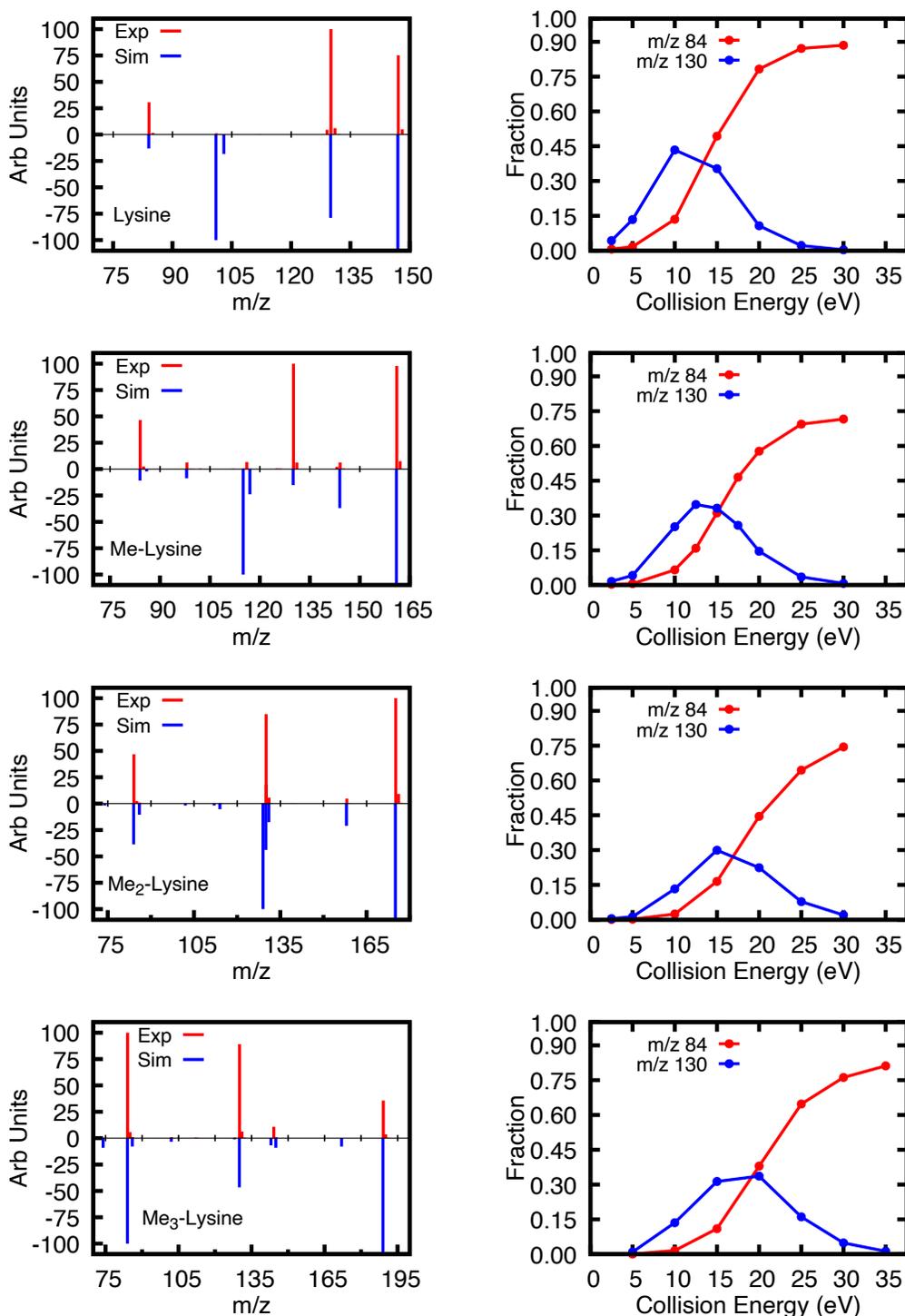


Figure 1: Experimental/simulated mass spectra are presented in the first column with experimental breakdown curves provided in column 2. The first column's experimental m/z spectrum intensity is scaled to the most intense peak for each given spectrum. In contrast, the simulated spectrum is scaled to the most intense fragment peak to ease comparison. The experimental collision energies are, from top to bottom, 10, 12.5, 15, and 20 eV, while the simulated internal energies are 250, 300, 350, and 400 kcal/mol, respectively. In column 2, the experimental breakdown curves are relative to the total ion signal collected rather than the most intense peak.

previous measurements made by Zhang *et al.*⁴³ as well as Dookeran *et al.*⁴⁴ Our direct dynamics simulations also find that m/z 84 and 130 are common products; however, m/z 101 ($(C_5H_{13}N_2)^+$) is the peak with the largest population for internal energies of 300 and 350 kcal/mol.

Table 1: Intensities for all experimental peaks above a fraction of 0.01 normalized intensity. Peaks resulting from less common isotopes, such compounds containing ^{13}C , are removed. All peaks are also observed in the simulations unless indicated.

Peak	Fraction	Loss
Lysine- H^+ ^a		
84.0807	0.135	$NH_3 + H_2CO_2$
129.1024 ^b	0.021	H_2O
130.0864	0.434	NH_3
147.1130	0.328	–
Me ₁ -lysine- H^+ ^a		
84.0806	0.159	$NH_2(CH_3) + H_2CO_2$
98.0965	0.023	$NH_3 + H_2CO_2$
116.1073 ^b	0.022	CH_3NO
130.0863	0.347	$NH_2(CH_3)$
144.1020	0.023	NH_3
161.1286	0.336	–
Me ₂ -lysine- H^+ ^a		
84.0805	0.165	$NH(CH_3)_2 + H_2CO_2$
130.0861	0.299	$NH(CH_3)_2$
130.1222 ^b	0.063	CH_3NO
131.1258	0.005	CO_2
158.1173	0.016	NH_3
175.1439	0.353	–
Me ₃ -lysine- H^+ ^a		
84.0807	0.380	$N(CH_3)_3 + H_2CO_2$
130.0865	0.336	$N(CH_3)_3$
144.1385 ^b	0.042	CH_3NO
189.1601	0.140	–

^a Same experimental collision energies as in Figure 1.

^b Not observed in simulations.

The m/z 84 peak corresponds to the loss of $NH_3 + H_2CO_2$, and our graph theory analysis reveals that nine different structures are seen for this peak within simulations. That said, three structures make up 84.5% of the population within the peak. These three most important structures, accounting

for 53.2%, 18.1%, and 13.3% of the peak population, respectively, are shown in Figures 2 and 3 along with reaction mechanisms that can produce them. The energetics of these mechanisms are shown in Figure 4.

The simulated m/z 84 pathways illustrate that this product can be formed by losing NH_3 as the first or second step with the nitrogen originated from either the side-chain or the N-terminus. Although loss is possible from either location, most simulated reaction pathways show loss of the side-chain, which is the preferred product for all internal energies considered. When considering all internal energies, 84.6% of the m/z 84 population obtained shows loss of the side-chain. This finding is in good agreement with our experimental measurements of lysine- H^+ (α - ^{15}N labeled), which allows for an experimental determination of side-chain vs. N-terminus loss. Examining the m/z 84 and 85 populations within the ^{15}N labeled experiments shows a loss of the side-chain $\sim 90\%$ of the time. Milne *et al.*⁴⁵ used chemical ionization to examine ^{15}N labeling on both the side-chain and N-terminus. They concluded that loss from either location was possible, but the loss of the side-chain was preferred by a $\sim 2:1$ ratio. The method of activation could potentially impact these ratios. After losing the side-chain, the ion must also lose H_2CO_2 to reach the final m/z 84 product. In simulations, this most commonly occurs as $C(OH)_2$ as depicted in the first three pathways in Figure 2. However, some trajectories show loss of either $H_2O + CO$ or HCO_2H . It is likely that in the long time limit of our simulations, $C(OH)_2$ will rearrange to one of these two more stable neutral species, which was observed in some simulations. Moreover, our simulations have a relatively large internal energy in order to allow for reactivity occur within a computationally feasible timeframe. Gregg *et al.* showed that simulations of protonated peptides have an increased reaction rate with increased collision energy while keeping the same mechanism.⁴⁰ That work did not feature competing mechanisms, and it is possible that the distribution among competing pathways could be impacted by internal energy.

O’Hair and co-workers⁴⁶ performed a detailed computational analysis on the preferred path for H_2CO_2 loss in protonated glycine. They compared

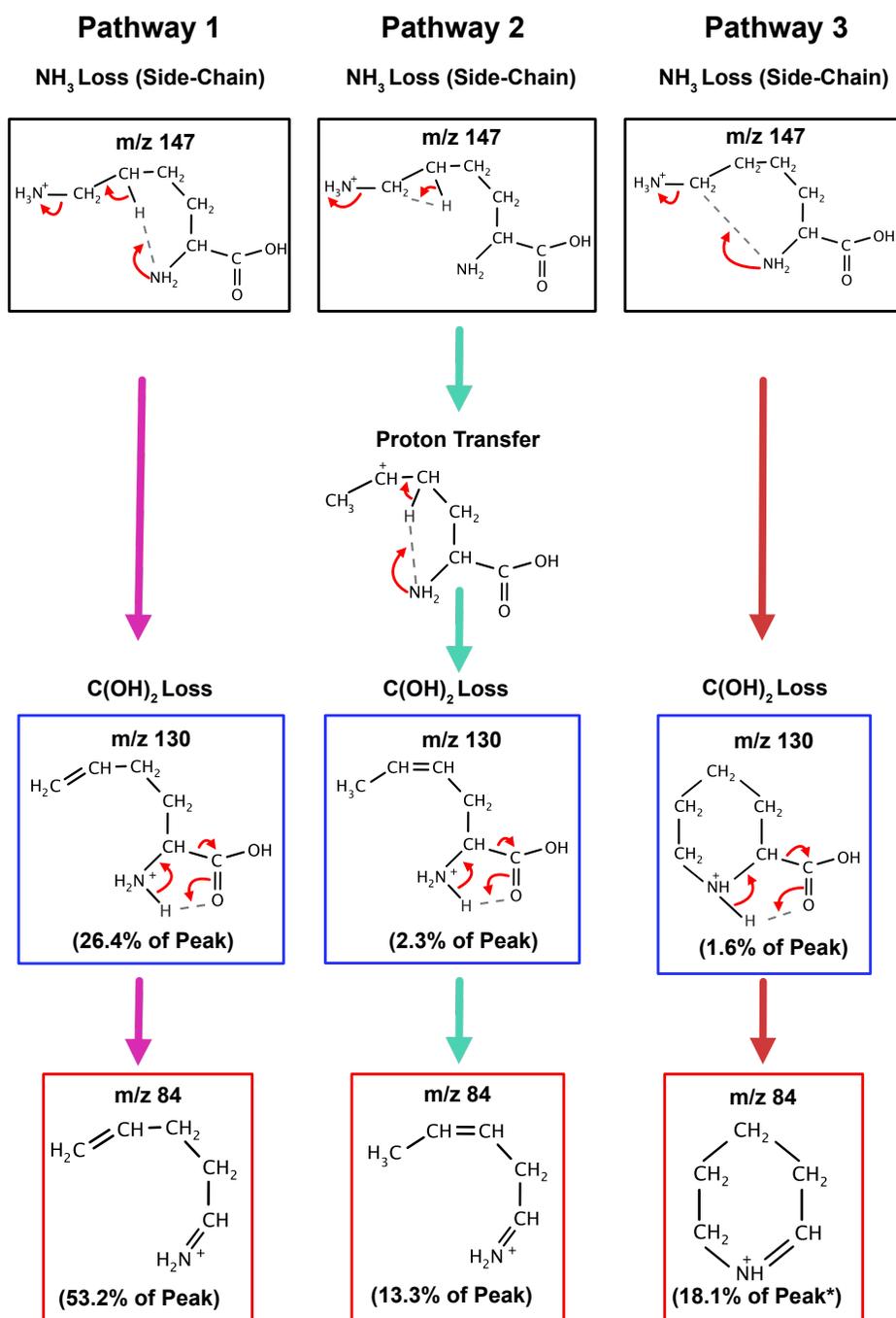


Figure 2: Reaction mechanisms for lysine-H⁺ frequently observed within the simulations that begin with NH₃ loss from the side-chain. The initial protonation state of lysine (black boxed structure) used in the simulations is the starting point for each mechanism. Loss of NH₃ results in *m/z* 130 (blue boxed structures) and leads to *m/z* 84 (red boxed structures) after loss of H₂CO₂. The energetics are provided in Figure 4 with reaction arrows color-coded to match. The *m/z* 84 structure that results from Pathway 3 also results from Pathway 4c.

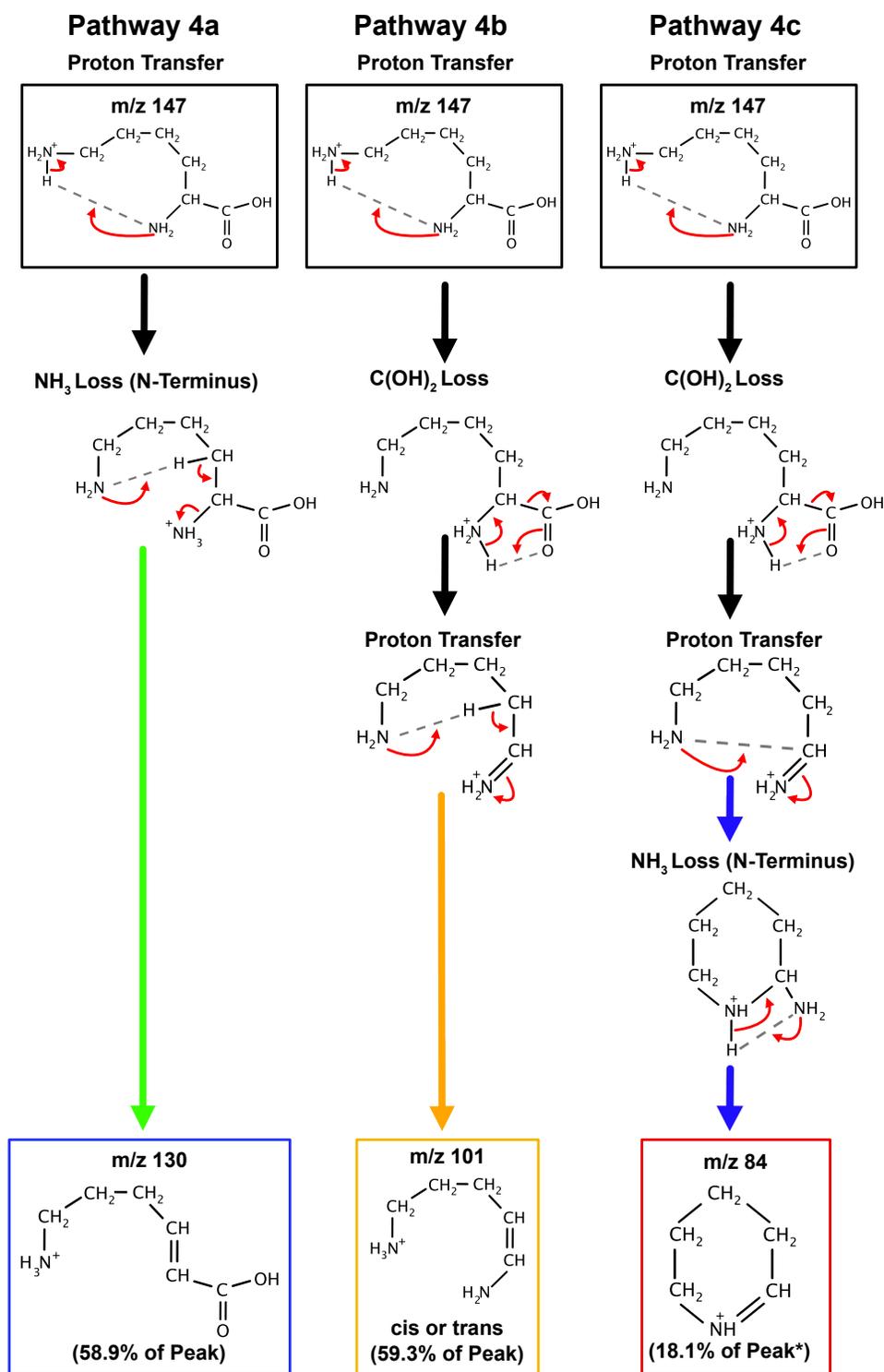


Figure 3: Reaction Pathway 4a-c are branched mechanisms frequently observed within simulations of lysine- H^+ that each begins with a proton transfer from the side-chain nitrogen to the N-terminus. Pathway 4a proceeds with loss of NH_3 from the N-terminus and results in m/z 130. Pathway 4b-c loses H_2CO_2 and leads to either m/z 101 intermediates or m/z 84 after the additional loss of NH_3 from the N-terminus. Black reaction arrows represent shared steps along the pathway. The energetics are provided in Figure 4. The m/z 84 structure that results from Pathway 4c also results from Pathway 3.

the change in free energy for all three possible neutral losses discussed above and concluded that the $\text{H}_2\text{O} + \text{CO}$ exit channel had the lowest barriers and is preferred. With this in mind, we performed additional calculations exploring the energetics of a direct $\text{H}_2\text{O} + \text{CO}$ loss starting from the m/z 130 structure in Pathway 1. Moving the the excess proton from the N-terminus to the -OH group of the carboxylic acid group results in water loss. If one considers the water to be infinitely separated from the results m/z 112 ion, the system would have an energy that is ~ 5 kcal/mol higher than the transition state shown in Figure 4 for loss of $\text{C}(\text{OH})_2$. The subsequent loss of CO to yield the m/z 84 ion is a downhill process and results in a final product with the same structure as that in Pathway 1. The final energy of these products is lower by ~ 28.8 kcal/mol since $\text{H}_2\text{O} + \text{CO}$ is more stable than $\text{C}(\text{OH})_2$. The limit of infinite separation of water prior to CO loss may not take place in either experiment or simulations. Considering the energy of an interacting water, we obtain a value that is 8.8 kcal/mol lower than the transition state shown in Figure 4. These DFT calculations show that direct loss of $\text{H}_2\text{O} + \text{CO}$ is possible but would likely not impact the final ion structure. Lastly, we note that the three structures shown for m/z 84 can interconvert, and at the $\omega\text{B97X-D/aug-cc-pVTZ}$ level of theory, the six-member ring structure is the six-member ring structure that has the lowest energy.

Simulations also revealed Pathway 4c, a unique m/z 84 formation mechanism that shows the initial loss of $\text{C}(\text{OH})_2$ followed by NH_3 . This pathway is the sole source of the $\sim 15\%$ of trajectories that lose the N-terminus and begins with a proton transfer from the side-chain nitrogen to the N-terminus. This type of proton transfer was commonly observed both in the present simulations and in our previous works.^{14,47} No distinct TS between these two protonation sites has been found, but rather the side-chain and N-terminus form short, strong hydrogen bonds.⁴⁸ After the excess proton is transferred to the N-terminus, there are two possible reactive pathways open to the ion. One pathway begins by transferring a proton from the N-terminus nitrogen to an oxygen of the C-terminus, resulting in either $\text{C}(\text{OH})_2$ or $\text{H}_2\text{O} + \text{CO}$ loss depending on the receiving oxygen. The product of this loss has m/z 101 and can subsequently rearrange

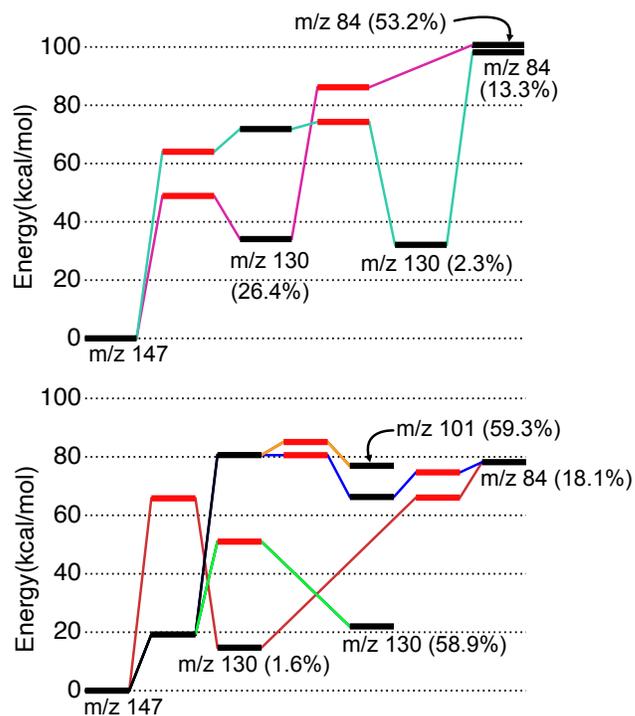


Figure 4: Energies for pathways shown in Figure 2 and 3 calculated at the $\omega\text{B97X-D/aug-cc-pVTZ}$ level of theory are shown above. Minima are labeled with solid black lines, while TSs are labeled with red. Pathways are displayed in two panels and color-coded in agreement with Figures 2 and 3.

to the m/z 84 product through a ring-closure reaction or result in a mixture of protonated cis/trans 1-Pentene-1,5-diamine, which are long-lived on the timescale of our simulations. We note that the final products can also be reached via a direct transfer of the proton from the side chain nitrogen to the oxygen of the C-terminus. Pathways 4b-c show that the m/z 101 population within the simulations is easily explained as an intermediate on the way to m/z 84. Our graph theory analysis shows that nine different structures correspond to m/z 101. However, 59.3% of the products are given by the boxed structure shown in Figure 3 with the other structures representing either intermediates along Pathways 4b-c or structures that can interconvert with the dominant boxed structure in Figure 3.

In addition to addressing the presence of m/z 101, these Pathways 1-3 and 4a show that m/z 130 is an intermediate to m/z 84. While the graph theory analysis finds 14 unique structures

for the m/z 130 peak, eight of these structures occur in just a single trajectory and are likely fleeting intermediates. Only two structures show sizable populations, as shown in Figures 2 and 3, with $\sim 59.3\%$ arising from the Pathway 4a-c which also branches between m/z 130, m/z 101, and m/z 84. Experimentally, as seen in Figure 1, m/z 130 starts as the most dominant peak but is supplanted by m/z 84 as the collision energy increases, which is consistent with the intermediate nature of m/z 130. While m/z 130 and 101 are intermediates for m/z 84, that does not necessarily relate to the time dependence for the formation of these ions at the internal energies considered here. Some of the randomly selected initial conditions quickly react, while others have a significant lifetime in the either the starting structure or intermediates.

While the majority of experimental peaks for lysine- H^+ are seen within simulations, Table 1 identifies water loss as an observed experimental peak that is not observed in simulations. In simulations, if water is lost, it is accompanied by other neutral losses, such as CO. Experimentally, water loss without other neutral loss is slightly above the 0.01 cutoff and accounts for a fraction of 0.021. However, this product does not reach the cutoff for the other systems. While we do not see water as the only neutral loss in the lysine- H^+ simulations, it is observed in simulations for Me_2 -lysine- H^+ . It is likely that if more lysine- H^+ simulations were calculated or if the internal energy was lower, water as the sole neutral loss in lysine- H^+ would also be observed.

With the results from the simulation and DFT calculations in mind, the experimental spectra can be interpreted. Experimental MS measurements show that at high collision energy ($\geq 15eV$), m/z 84 is the dominant feature of the MS spectrum, which is consistent with the computational results. Isomeric fragments with m/z 130 are common intermediates in Pathways 1-4, and MS3 experiments confirm that the fragmentation of m/z 130 yields m/z 84. While both DFT calculated barriers and energy minimum for m/z 130 structures are lower than those for m/z 84, the formation of the m/z 84 fragment from m/z 130 requires the loss of H_2CO_2 , which is likely a non-reversible loss since it is a dissociative step. Thus, given enough

time and energy, the system will eventually reach this product. The formation of the m/z 130 structure from Pathway 4a that accounts for 58.9% of the simulated peak population appears to be a low energy endpoint with a relatively low barrier to formation. However, it is easy to believe that this structure will ultimately lose H_2CO_2 , in some form, resulting in m/z 84. Reaching m/z 84 from this m/z 130 structure requires rearrangements that were not observed within the time frame of our simulations.

Pseudo MS3 experiments on m/z 101 also have m/z 84 as a product, consistent with Pathways 4b-c. The DFT calculations show that the formation of m/z 101 is an apparent endpoint for Pathway 4b, On the other hand, Pathway 4c, which leads to m/z 84, shares several steps with Pathway 4b. Since the unique steps in Pathway 4b are not dissociative, it seems likely that given sufficient time the population in m/z 101 resulting from Pathway 4b will also lead to m/z 84.

3.2 Me_n -lysine- H^+

As seen in Figure 1 and Table 1, the Me_n -lysine- H^+ series all have m/z 130 and 84 as the most critical peaks for the collision energies considered. These peaks are closely related to the same peaks in lysine- H^+ and correspond to side-chain loss of $NH_{3-n}(CH_3)_n$ for m/z 130 and the additional loss of H_2CO_2 for m/z 84. Simulations indicated that similar mechanisms are observed for Me_n -lysine- H^+ as for lysine- H^+ , and in fact, after the loss of $NH_{3-n}(CH_3)_n$ the remaining steps in Pathways 1-3 do not depend on methylation. There are some differences seen in the relative population of final ion structures. The lysine- H^+ simulations found that Pathways 3 and 4c result in an m/z 84 ion with a six-member ring structure and accounted for 18.1% of the peak. This structure is infrequently observed within the timeframe of the simulations for any of the methylated systems. For Me_n -lysine- H^+ , the equivalent of Pathway 4c would no longer result in an m/z 84 as it involves the loss of NH_3 from the N-terminus rather than the side-chain $NH_{3-n}(CH_3)_n$. In contrast, Pathway 3, which involves the direct loss of the side-chain $NH_{3-n}(CH_3)_n$ group, would still result in an m/z 84 ion. When considering all in-

ternal energies, this pathway that results in a six-member ring now accounts for 3%, 2% and 5.5% of the Me_n -lysine ($n=1,2,3$) m/z 84 population, respectively. The addition of methyl groups likely slows ring closure, making six-member ring formation less common in the Me_n -lysine- H^+ simulations. That said, the methylated systems do still exhibit cyclic final products. In Me_2 -lysine- H^+ and Me_3 -lysine- H^+ , a five-member ring structure is the third most common when considering all internal energies and consists of $\sim 10\%$ of the m/z 84 population for each of those systems. An equivalent structure is also observed in Me_1 -lysine- H^+ and lysine- H^+ , though it is not as prevalent. This product occurs via N-terminus attack at the ϵ carbon of the side-chain. This attack results in the loss of $\text{NH}_{3-n}(\text{CH}_3)_n$ along with a proton transfer from ϵ to γ carbon of the side-chain. This is followed by H_2CO_2 loss to form the m/z 84 structure shown in Figure 5. These differences aside, simulations predict that the most common structure for m/z 84 in the methylated systems is the equivalent of the straight-chain structure that results from Pathway 1 of Figure 2. The simulations show that the same mechanism is followed except for the initial loss of $\text{NH}_{3-n}(\text{CH}_3)_n$ rather than NH_3 .

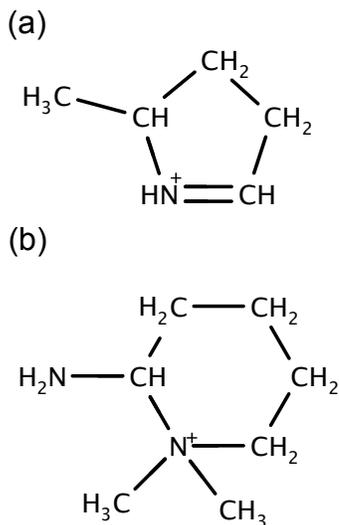


Figure 5: Panel (a) shows the five-member ring structure with m/z 84 while panel (b) shows the six-member ring structure within m/z 129 for Me_2 -lysine.

In considering the pathways that form the

m/z 130 peak, the most important structure seen in lysine- H^+ was formed through Pathway 4a and involved loss of NH_3 from the N-terminus. In the Me_n -lysine- H^+ systems, the equivalent mechanism to that shown in Pathway 4a will not result in a m/z 130 ion, and it is observed that the overall population within the m/z 130 peak is decreased for Me_n -lysine- H^+ compared to lysine- H^+ . Comparing populations between systems is non-trivial since they have different internal degrees of freedom, which could impact the overall reactivity for a given internal energy. In our simulations, for an internal energy of 300 kcal/mol in lysine- H^+ , the total fraction of reactive simulations is 0.585, with the m/z 130 population contributing 0.140 to that total. In comparison, at an internal energy of 350 kcal/mol in Me_1 -lysine- H^+ , the total fraction of reactive simulations is 0.558, with m/z 130 contributing 0.068 to that total, which shows that the m/z 130 population has indeed decreased.

When considering all internal energies, the major product seen in all methylated systems for m/z 130 is the intermediate shown in Pathway 1 that accounts for 26.4% of the lysine- H^+ m/z 130 population. This structure accounts for 64.2%, 72.2%, and 64.6% of the populations in m/z 130 for the Me_n -lysine- H^+ systems ($n=1,2,3$), respectively. Cyclic structures formed primarily from the intramolecular reaction of the N-terminus and the side-chain carbons are observed for all systems. They account for a larger fraction of the population than was seen for lysine- H^+ . However, this observation may be due to the reduced overall population of the peak within the simulations.

While m/z 84 and 130 are the most important peaks for Me_n -lysine and occur due to loss of the side-chain nitrogen, there are possible peaks that correspond to the loss of the N-terminus. In particular, loss of $\text{NH}_3 + \text{H}_2\text{CO}_2$ would correspond to m/z 98, 112, and 126 in Me_n -lysine ($n=1,2,3$), respectively. Any population found in these peaks would be seen within m/z 84 for lysine- H^+ as loss of the side-chain, and N-terminus NH_3 are indistinguishable in that system. These first two peaks are observed in the simulations but are rare events, and only the first peak is observed experimentally above the 0.01 threshold. For Me_1 -lysine- H^+ , this product is observed in 22 trajectories across all internal energies considered, with the most common

structure being analogous to that formed by Pathway 4c in lysine- H^+ , i.e., a six-member ring with a methyl group attached to the nitrogen. For Me_2 -lysine- H^+ the number drops to just three trajectories, all with unique structures, none of which correspond to that seen in the lysine- H^+ system. Loss of the N-terminus alone is not observed for Me_3 -lysine- H^+ . In comparison, loss of the N-terminus occurred in roughly 15% of trajectories that resulted in m/z 84 for lysine- H^+ . One aspect of this effect is that to lose NH_3 from the N-terminus, a proton must be transferred to that position for Me_n -lysine- H^+ ($n=1,2$) and that as more methyl groups are added, the proton affinity of the side-chain increases. However, for Me_3 -lysine, the excess proton is already placed on the N-terminus nitrogen, and NH_3 loss is still uncommon. Methylation of the side-chain has effectively shut down the loss of the N-terminus since $NH_{3-n}(CH_3)_n$ is a better leaving group.

Just as there are possible peaks that involve loss of $NH_3 + H_2CO_2$, there are possible peaks that involve the sole loss of NH_3 . These possible peaks correspond to m/z 144, 158, and 172, for Me_n -lysine ($n=1,2,3$), respectively. None of these products are shown in Table 1, meaning that their experimental intensity is less than 0.01 fraction of the total ion signal. Within the simulations with internal energies of 350, 350, and 400 kcal/mol, these peaks are observed to have an intensity of 0.050, 0.030 and 0.018 for Me_n -lysine- H^+ ($n=1,2,3$), respectively. In comparison, the fraction of m/z 130 for these systems is 0.068, 0.063, and 0.100. Although we note that the fractions in simulations depend on the time window calculated, the N-terminus loss peaks follow the trend of decreasing importance with increased methylation but are seen more commonly than in experiment. These ions have a straight-chain structure with a double bond formed between the α carbon and the first carbon of the side-chain.

In simulations of lysine- H^+ , it was seen that the m/z 101 peak was the most common for intermediate all internal energies and that this ion was an intermediate to m/z 84 as seen in Pathways 4b-c. The analogous peaks in Me_n -lysine- H^+ are m/z 115, 129, and 145. For Me_1 -lysine- H^+ and Me_2 -lysine- H^+ , these peaks are again the most populous peaks within the simulations. When considering all in-

ternal energies, the most common structure observed, making up 57.1% and 57.8% of the of m/z 115 and m/z 129 population, is the analogue of the final product of Pathway 4b. In the Me_2 -lysine- H^+ system, a six-member ring structure is the next most common structure with 14.5% of the population and is shown in Figure 5. However, it would be an intermediate to m/z 112, i.e., loss of the N-terminus, rather than the side-chain loss that would result in an m/z 84 structure. For Me_3 -lysine- H^+ , the m/z 145 peak, which corresponds to the loss of CO_2 rather than H_2CO_2 , is not the most populous peak. Instead, at 400 kcal/mol, this peak makes up a fraction of just 0.020 of the calculated trajectories, while m/z 84 has a fraction of 0.220 and is the most significant fragment peak.

The simulations do not observe some of the peaks outlined in Table 1, namely m/z 116.1073 in Me_1 -lysine- H^+ , m/z 130.1222 in Me_2 -lysine- H^+ , and 144.1385 in Me_3 -lysine- H^+ . These peaks correspond to the loss of CH_3NO with the nitrogen originating from the N-terminus. These peaks are minor contributors to the experimental spectra, though the contribution in the Me_2 -lysine- H^+ does reach $\sim 6\%$ of the total ion signal. To investigate if these peaks originated from sudden fragmentation events following collision with the CID gas, Ar CID simulations for Me_2 -lysine- H^+ were performed with collision energies ranging from 10 to 20 eV, but loss of CH_3NO was not observed. A few trajectories produced CO due to the sudden collision with Ar followed by the formation of NH_4^+ . If the excess proton had transferred back to the larger fragment, it would produce the experimentally observed m/z ratio, i.e., loss of $CO+NH_3$. However, that did not occur in the simulations, and these were rare events making it unlikely that a sudden event leads to this experimental peak.

4 Summary

We have presented results from experiments and direct dynamics simulations aimed at exploring the reaction pathways that occur within lysine- H^+ and its side-chain methylated analogues, Me_n -lysine- H^+ with $n=1-3$. Experimentally, for the lower collision energies considered, the m/z 130 peak ($(C_6H_{12}NO_2)^+$) is the most important, with

the m/z 84 peak ($(C_5H_{10}N)^+$) becoming dominant as collision energy increases. These two peaks correspond to loss of $NH_{3-n}(CH_3)_n$ and $NH_{3-n}(CH_3)_n + H_2CO_2$, respectively.

Direct dynamics simulations show population in final ion structures that correspond to nearly all peaks observed in experiment that have a total ion contribution above 0.01. Multiple reaction pathways were identified from the simulations and quantified through ω B97X-D/aug-cc-pVTZ single point calculations to identify relevant minima and transition states. Simulations show that in the lysine- H^+ system, m/z 84 can be formed from loss of nitrogen at either the side-chain or the N-terminus, though the side-chain is preferred. The loss of NH_3 can occur as either the first or second neutral loss within the mechanisms. Simulations of lysine- H^+ show that the m/z 101 peak ($C_5H_{13}N_2$) is the most important within the time frame considered. In addition, the simulations, the DFT calculations, and pseudo MS3 experiments all show that m/z 101 is an intermediate to m/z 84. Three primary structures are found for m/z 84, all of which can interconvert, with the six-member ring structure being the energetic minimum. The m/z 130 ion is an intermediate along most of the reaction pathways described. The DFT calculations reveal that the formation of m/z 130 has the lowest energy barriers, consistent with the experimental results.

The experiments and simulations of the Me_n -lysine- H^+ systems allow for differentiation of the side-chain from the N-terminus. It is seen that side-chain loss is dominant in both experiments and simulations, which is consistent with $NH_{3-n}(CH_3)_n$ being a better leaving group than NH_3 . The structures observed in simulations of Me_n -lysine- H^+ are analogous to those in lysine- H^+ , though the m/z 84 population shows a de-emphasis of the six-member ring structure.

Our graph theory analysis allowed for an automatic and in-depth view of the structures that made up each peak observed in the simulated mass spectra. These insights allowed for targeted DFT-based calculations that characterized the energy landscape for the lysine- H^+ system. We expect that the lysine- H^+ energy landscape is similar to that of the methylated analogues. Moreover, we also expect that using graph theory based post-

simulation analysis will continue to be developed and become an invaluable tool within the community due to the insight that it provides.

5 ACKNOWLEDGEMENTS

The authors acknowledge the generous support from the National Science Foundation under Grant No. 1763652. GLB is a member of the MERCURY consortium, which receives support through National Science Foundation grant No. CHE-2018427.

References

- (1) Kumar, A.; Narayanan, V.; Sekhar, A. Characterizing Post-Translational Modifications and Their Effects on Protein Conformation Using NMR Spectroscopy. *Biochemistry* **2019**, *59*, 57–73.
- (2) Ryšlavá, H.; Doubnerová, V.; Kavan, D.; Vaněk, O. Effect of posttranslational modifications on enzyme function and assembly. *J. Proteomics* **2013**, *92*, 80–109.
- (3) Boscher, C.; Dennis, J. W.; Nabi, I. R. Glycosylation, galectins and cellular signaling. *Curr. Opin. Cell Biol.* **2011**, *23*, 383–392.
- (4) Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **2003**, *21*, 255–261.
- (5) Hyun, K.; Jeon, J.; Park, K.; Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **2017**, *49*, e324.
- (6) Zhang, X.; Huang, Y.; Shi, X. Emerging roles of lysine methylation on non-histone proteins. *Cell. Mol. Life Sci.* **2015**, *72*, 4257–4272.
- (7) Zhang, X.; Wen, H.; Shi, X. Lysine methylation: Beyond histones. *Acta Biochim. Biophys. Sin. (Shanghai)*. **2012**, *44*, 14–27.
- (8) Ivanov, G. S.; Ivanova, T.; Kurash, J.; Ivanov, A.; Chuikov, S.; Gizatullin, F.;

- Herrera-Medina, E. M.; Rauscher, F.; Reinberg, D.; Barlev, N. A. Methylation-Acetylation Interplay Activates p53 in Response to DNA Damage. Mol. Cell. Biol. **2007**, 27, 6756–6769.
- (9) Cooks, R. G.; Ast, T.; Pradeep, T.; Wysocki, V. H. Reactions of Ions with Organic Surfaces. Acc. Chem. Res. **1994**, 27, 316–323.
- (10) Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. Mass Spectrom. Rev. **1995**, 14, 49–73.
- (11) Williams, E. R. Tandem FTMS of Large Biomolecules. Anal. Chem. **1998**, 70, 179A–85A.
- (12) Wysocki, V. H.; Joyce, K. E.; Jones, C. M.; Beardsley, R. L. Surface-induced dissociation of small molecules, peptides, and non-covalent protein complexes. J. Am. Soc. Mass Spectrom. **2008**, 19, 190–208.
- (13) Martin Somer, A.; Macaluso, V.; Barnes, G. L.; Yang, L.; Pratihari, S.; Song, K.; Hase, W. L.; Spezia, R. Role of Chemical Dynamics Simulations in Mass Spectrometry Studies of Collision-Induced Dissociation and Collisions of Biological Ions with Organic Surfaces. J. Am. Soc. Mass Spectrom. **2020**, 31, 2–24.
- (14) Lucas, K.; Barnes, G. L. Modeling the Effects of O-Sulfonation on the CID of Serine. J. Am. Soc. Mass Spectrom. **2020**, 31, 1114–1122.
- (15) Barnes, G. L.; Shlaferman, A.; Strain, M. Fast fragmentation during surface-induced dissociation: An examination of peptide size and structure. Chem. Phys. Lett. **2020**, 754, 137716.
- (16) Pratihari, S.; Barnes, G. L.; Hase, W. L.; Pratihari, S.; Barnes, G. L.; Hase, W. L. Chemical Dynamics Simulations of Energy Transfer, Surface-Induced Dissociation, Soft-Landing, and Reactive-Landing in Collisions of Protonated Peptide Ions with Organic Surfaces. Chem. Soc. Rev. **2015**, 45, 3595–3608.
- (17) Pratihari, S.; Barnes, G. L.; Laskin, J.; Hase, W. L. Dynamics of Protonated Peptide Ion Collisions with Organic Surfaces. Consonance of Simulation and Experiment. J. Phys. Chem. Lett. **2016**, 7, 3142–3150.
- (18) Frederickson, D.; McDonough, M.; Barnes, G. L. A Computational Comparison of Soft Landing of Alpha-Helical vs Globular Peptides. J. Phys. Chem. B **2018**, 122, 9549–9554.
- (19) Barnes, G. L.; Shlaferman, A.; Strain, M. Fast fragmentation during surface-induced dissociation: An examination of peptide size and structure. Chem. Phys. Lett. **2020**, 754, 137716.
- (20) Vázquez, S. A.; Otero, X. L.; Martínez-Núñez, E. A trajectory-based method to explore reaction mechanisms. Molecules **2018**, 23, 3156.
- (21) Rodríguez, A.; Rodríguez-Fernández, R.; A. Vázquez, S.; L. Barnes, G.; J. P. Stewart, J.; Martínez-Núñez, E. tsscds2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. J. Comput. Chem. **2018**, 39, 1922–1930.
- (22) Rossich-Molina, E.; Salpin, J.-Y.; Spezia, R.; Martínez-Núñez, E. On the gas phase fragmentation of protonated uracil: a statistical perspective. Phys. Chem. Chem. Phys. **2016**, 18, 10000–10008.
- (23) Trinajstić, N. Chemical Graph Theory, 2nd ed.; CRC Press, 1992; p 352.
- (24) Barnes, G. L.; Young, K.; Yang, L.; Hase, W. L. Fragmentation and Reactivity in Collisions of Protonated Diglycine with Chemically Modified Perfluorinated Alkylthiolate-Self-Assembled Monolayer Surfaces. J. Chem. Phys. **2011**, 134, 094106.
- (25) Barnes, G. L.; Hase, W. L. Energy Transfer, Unfolding, and Fragmentation Dynamics in Collisions of N-Protonated Octaglycine with

- an H-SAM Surface. *J. Am. Chem. Soc.* **2009**, *131*, 17185–17193.
- (26) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminform.* **2012**, *4*, 17.
- (27) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- (28) Stewart, J. P. Mopac2016. 2016; <http://openmopac.net>.
- (29) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; Van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.
- (30) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated force field Topology Builder (ATB) and repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7*, 4026–4037.
- (31) Koziara, K. B.; Stroet, M.; Malde, A. K.; Mark, A. E. Testing and validation of the Automated Topology Builder (ATB) version 2.0: Prediction of hydration free enthalpies. *J. Comput. Aided. Mol. Des.* **2014**, *28*, 221–233.
- (32) Canzar, S.; El-Kebir, M.; Pool, R.; Elbassioni, K.; Mark, A. E.; Geerke, D. P.; Stougie, L.; Klau, G. W. Charge group partitioning in biomolecular simulation. *J. Comput. Biol.* **2013**, *20*, 188–198.
- (33) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111.
- (34) Homayoon, Z.; Pratihari, S.; Dratz, E.; Snider, R.; Spezia, R.; Barnes, G. L.; Macaluso, V.; Martin Somer, A.; Hase, W. L. Model Simulations of the Thermal Dissociation of the TIK(H⁺)₂ Tripeptide: Mechanisms and Kinetic Parameters. *J. Phys. Chem. A* **2016**, *120*, 8211–8227.
- (35) Gu, M.; Zhang, J.; Hase, W. L.; Yang, L. Direct Dynamics Simulations of the Thermal Fragmentation of a Protonated Peptide Containing Arginine. *ACS Omega* **2020**, *5*, 1463–1471.
- (36) Meroueh, O.; Hase, W. L. Collisional Activation of Small Peptides. *J. Phys. Chem. A* **1999**, *103*, 3981–3990.
- (37) Hase, W. L.; Buckowski, D. G. Monte carlo sampling of a microcanonical ensemble of classical harmonic oscillators. *Chem. Phys. Lett.* **1980**, *74*, 284–287.
- (38) Schlier, C.; Seiter, A. High-Order Symplectic Integration: An Assessment. *Comput. Phys. Commun.* **2000**, *130*, 176–189.
- (39) Stewart, J. P. Mopac2012. 2012; <http://openmopac.net>.
- (40) Gregg, Z.; Ijaz, W.; Jannetti, S.; Barnes, G. L. The Role of Proton Transfer in Surface-Induced Dissociation. *J. Phys. Chem. C* **2014**, *118*, 22149–22155.
- (41) Frisch, M. J. et al. Gaussian 09, Revision D.01. 2009.
- (42) Chai, J. D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- (43) Zhang, P.; Chan, W.; Ang, I. L.; Wei, R.; Lam, M. M.; Lei, K. M.; Poon, T. C. Revisiting Fragmentation Reactions of Protonated α -Amino Acids by High-Resolution Electrospray Ionization Tandem Mass Spectrometry with Collision-Induced Dissociation. *Sci. Rep.* **2019**, *9*, 6453.

- (44) Dookeran, N. N.; Yalcin, T.; Harrison, A. G. Fragmentation reactions of protonated α -amino acids. J. Mass Spectrom. **1996**, 31, 500–508.
- (45) Milne, G. W.; Axenrod, T.; Fales, H. M. Chemical Ionization Mass Spectrometry of Complex Molecules. IV.1 Amino Acids. J. Am. Chem. Soc. **1970**, 92, 5170–5175.
- (46) O’Hair, R. A.; Broughton, P. S.; Styles, M. L.; Frink, B. T.; Hadad, C. M. The fragmentation pathways of protonated glycine: A computational study. J. Am. Soc. Mass Spectrom. **2000**, 11, 687–696.
- (47) Shaikh, K.; Blackwood, J.; Barnes, G. L. The Effect of Protonation Site and Conformation on Surface-Induced Dissociation in a Small, Lysine Containing Peptide. Chem. Phys. Lett. **2015**, 637, 83–87.
- (48) Humbel, S. Short strong hydrogen bonds: A valence bond analysis. J. Phys. Chem. A **2002**, 106, 5517–5520.