The Contributions of Molecular Framework to IMS Collision Cross-Sections of Gas-phase Peptide Ions

Lei Tao,[a] David B. Dahl,[b] Lisa M. Pérez,[c] and David H. Russell[a]

[a] The Laboratory for Biological Mass Spectrometry, Department of Chemistry, Texas A and M University, College Station, Texas, USA
[b] Department of Statistics, Texas A and M University, College Station, Texas, USA
[c] The Laboratory for Molecular Simulation, Texas A and M University, College Station, Texas, USA

Molecular dynamics (MD) is an essential tool for correlating collision cross-section data determined by ion mobility spectrometry (IMS) with candidate (calculated) structures. Conventional methods used for ion structure determination rely on comparing the measured cross-sections with the calculated collision cross-section for the lowest energy structure(s) taken from a large pool of candidate structures generated through multiple tiers of simulated annealing. We are developing methods to evaluate candidate structures from an ensemble of many conformations rather than the lowest energy structure. Here, we describe computational simulations and clustering methods to assign backbone conformations for singly-protonated ions of the model peptide (NH$_2$-Met-Ile-Phe-Ala-Gly-Ile-Lys-COOH) formed by both MALDI and ESI, and compare the structures of MIFAGIK derivatives to test the 'sensitivity' of the cluster analysis method. Cluster analysis suggests that [MIFAGIK + H]$^+$ ions formed by MALDI have a predominantly turn structure even though the low-energy ions prefer partial helical conformers. Although the ions formed by ESI have collision cross-sections that are different from those formed by MALDI, the results of cluster analysis indicate that the ions backbone structures are similar. Chemical modifications (N-acetyl, methylester as well as addition of Boc or Fmoc groups) to MIFAGIK alter the distribution of various conformers; the most dramatic changes are observed for the [M + Na]$^+$ ion, which show a strong preference for random coil conformers owing to the strong solvation by the backbone amide groups.   (J Am Soc Mass Spectrom 2009, 20, 1593–1602) © 2009 American Society for Mass Spectrometry

The emphasis of mass spectrometry based biological chemistry is shifting from compound identification to structural studies of large biomolecules and biomolecule complexes [1–7], including membrane proteins [8]. The next phase of 'omics' related research must be aimed at obtaining and predicting additional dimensions of information, such as secondary, tertiary, and quaternary structures and linkage-specific information for glycans. Although sophisticated structural characterization tools, such as NMR and XRD, provide the most information, high throughput analysis of complex biological mixtures obtained by using these techniques is an underdeveloped technology. On the other hand, IMS is much more than a separation device, the structural information derived from 2D conformation space afforded by IM-MS is potentially well-suited to both high throughput applications and complex biological samples.

A number of laboratories have focused their research on developing IM-MS for biophysical studies of peptides and proteins [9–14]. In previous work, we showed that a large proportion of singly charged peptide ions (formed by MALDI) appear on a single trendline in 2D mobility-m/z plots, i.e., plots of arrival-time distribution (ATD) or ion-neutral collision cross-section (Ω) versus m/z [15]; however, a few ion signals deviate (>3% to ~20%) from the expected trendline and nonpeptidic ion signals appear on separate, compound class specific trendlines [14, 15]. Ruotolo et al. showed that gas-phase [M + H]$^+$ ions of LLGNVLVVVLAR (derived from bovine hemoglobin) prefer extended (helical) structure(s) resulting in a larger collision cross-section than random coil structures having the same or similar m/z values [12, 13], while some post-translationally modified (PTM) peptide ions (phosphopeptides) tend to pack more tightly than the unmodified protonated peptide ions owing to intra-molecular charge-solvation and/or formation of salt-bridged type structures [16, 17]. In addition, we have used chemical derivatization via acetylation of the N-terminus and internal basic lysine residues and methylation of the acidic glutamic

Address reprint requests to Dr. D. H. Russell, The Laboratory for Biological Mass Spectrometry, 3255 TAMU, College Station, TX 77843-3255, USA. E-mail: russell@mail.chem.tamu.edu

residues to show that the helical propensity of a given peptide can be increased by reducing the number of 'salt-bridge' intramolecular interactions [18].

Molecular dynamics (MD) is the method of choice for correlating the measured collision cross-section (Ω_{meas}) with candidate structures. Ion structure is derived by comparing Ω_{meas} with the calculated collision cross-section (Ω_{calc}) of the lowest energy structure obtained using MD simulations. This approach involves selecting the 'lowest energy structure' from a large pool of candidate structures generated through multiple tiers of simulated annealing [19]. It is important to note, however, the lowest energy structure may not be representative of Ω_{meas} since under the experimental conditions, i.e., long resident times in the drift cell and small amounts of collisional heating can facilitate rearrangement reactions that may be subject to low-energy barriers. The ion population is composed of an ensemble of many conformations at a defined ion effective temperature range as opposed to a single conformation assumed at the lowest energy [20].

The challenge for structure IM-MS is to design gas-phase experiments that critically evaluate the structural assignments [21–23]. The research described here is aimed at bridging the gap between the experimental domain of IM-MS and the MD simulations and bioinformatic tools that are essential to interpretation of the data. In this study, we apply a novel clustering algorithm to a model peptide to identify groups of structural elements from a large pool of diverse candidate structures. The clustering algorithm is similar to that used by Damsbo et al. [24], i.e., have grouped the candidate structures based on the similarity of backbone structure; however, our procedure provides estimates of the uncertainty of the cluster membership and the degree of purity of the cluster.

In this study, we use NH$_2$-Met-Ile-Phe-Ala-Gly-Ile-Lys-COOH (MIFAGIK), residues 80–86 of cytochrome *c*, as our model peptide to develop methods to evaluate the complete candidate structure population. We selected this peptide because the Ω_{meas} for [M + H]$^+$ ions formed by MALDI is about 5% larger than that expected for globular conformation [15] and smaller than that expected for helical structure. The observation raises the question of whether the ion population is composed of a number of very similar conformations and can we use statistical analysis tools to gain information about structural diversity. Although simulated annealing experiments yield a lowest energy structure that contains a helical turn conformation in the region of the residues Ile ^{2}-Phe^3-Ala^4, candidate structures with Ω_{calc} that fall within ±2% of the measured value are comprised of a number of structure types. We have also compared MIFAGIK derivatives, including acetylated, Boc and Fmoc protected N-terminus, C-terminal methyl ester, Fmoc protected Lys^7, and [M + Na]$^+$ ions. The experimental data provides new insight into the role of protecting groups in the N- and C-terminal positions

and the role of the charge site in the conformational preference of small peptides.

Experimental

Materials

The peptide MIFAGIK (MW 778.5) was purchased from Genscript Corp. (Piscataway, NJ, USA) and used without further purification. Bradykinin (RPPGFSPF, MW 1060.2), substance P (RPKPQQFFGLM-NH$_2$, MW 1346.6), anhydrous methanol, acetyl chloride, acetic anhydride Boc anhydride, 9-fluorenylmethyl chloroformate, and dioxane were obtained from Sigma (St. Louis, MO, USA). The peptide derivatives were synthesized using conventional solution phase methods [25, 26]. The acetyl, Boc and Fmoc group were used for amino group protection and the C-terminus was protected by the methyl ester. Deprotection of Boc group was performed using hydrochloric acid. All peptide sequences were confirmed using tandem mass spectrometry (Applied Biosystems 4700 Proteomics analyzer, Applied Biosystems, Foster City, CA, USA).

MIFAGIK-OMe

The methanolic HCl reagent was prepared by dropwise addition of 800 μL acetyl chloride to 5 mL of anhydrous methanol with stirring. After 5 min, 200 μL methanolic HCl was added to 1 mg of dry peptide. The solution was incubated for a period of 2 h at room-temperature and was dried by lyophilization. The resulting O-methyl ester was dissolved in 1 mL distilled water and used for mass analysis without further purification.

N-Ac-MIFAGIK

The stock solution of peptide MIFAGIK was prepared by dissolving 1 mg peptide in 1 mL of 50 mM ammonium bicarbonate. The acetylation reagent was prepared by mixing 10 μL acetyl anhydride with 20 μL anhydrous methanol and then added to 50 μL peptide stock solution. The reaction was allowed to proceed for a period of 15 min at room temperature. The sample was dried by lyophilization. The peptide solution was reconstituted by mixing with 50 μL distilled water and analyzed by mass spectrometry without further purification.

N-Fmoc-MIFAGIK

The Fmoc group was introduced by treatment of 10 μL free peptide MIFAGIK stock solution with 10 μg 9-fluorenylmethyl chloroformate in 10 μL aqueous dioxane. The reaction was allowed to proceed for a period of 60 min at room temperature. The sample was dried by lyophilization. The peptide solution was reconstituted by mixing with 20 μL distilled water and analyzed by mass spectrometry without further purification.

MIFAGIK-N-Fmoc

The MIFAGIK-*N-Fmoc* derivative was synthesized by first protecting the N-terminus with a Boc protecting group and then reacting with Fmoc followed by removable of the Boc group under acidic conditions. A solution of Boc-anhydride (2 μmoL) in 10 μL anhydrous methanol was added to 10 μL free peptide MIFAGIK stock solution. After stirring for 30 min at room temperature, the mixture was concentrated in vacuo. The residue was dissolved in 10 μL aqueous dioxane in the presence of sodium carbonate to which 10 μg 9-fluorenylmethyl chloroformate in 10 μL aqueous dioxane was added. The reaction was allowed to proceed for 60 min at room-temperature. The Boc group was removed by addition of 5 μL of 1.2 M HCl, which was allowed to react for 10 min. The product was then analyzed by mass spectrometry without further purification.

MALDI-IM-TOFMS

Stock solutions of peptide (1 mg/mL) were mixed with re-crystallized α-cyano-4-hydroxycinnamic acid (Sigma) in a molar ratio of ~2000:1, and 1 μL aliquots of the peptide/matrix mixture was deposited on a stainless steel MALDI sample plate.

The MALDI-IM-TOFMS analysis was performed using a home-built instrument previously described [27]. Briefly, MALDI was performed using a frequency tripled (355 nm) Nd:YAG laser (CrystaLaser, Reno, NV, USA) operated at a pulse rate of 200–400 Hz. The resulting ions were introduced into a 15 cm drift tube maintained at a pressure of ca. 2.7 Torr of He maintained at room-temperature. Under these conditions the E/p ratios range from 18 to 34 Vcm^{-1}torr$-^{-1}$, which provides mobility resolution of 15 to 25. The ions eluting the drift cell were extracted and mass analyzed by an orthogonal reflectron-TOF, typical m/z resolution ca. 2000–4000. The mass spectrometer was externally calibrated using two-point calibration on C_{60} (M_r = 720) and C_{70} (M_r = 840) radical cations (Sigma) [5]. The measurements of collision cross sections were externally calibrated with $[M + H]^+$ ions of bradykinin (Ω_{meas} = 245 Å^2 and substance P (Ω_{meas} = 292 Å^2 [28]. The 2D IM-MS data were acquired and processed by using custom software (Ionwerks, Inc., Houston, TX, USA).

The experimental ion-neutral collision cross-sections were determined from the empirical drift times (t_d) by the following equation [29]:

$$\Omega_{meas} = \frac{(18\pi)^{\frac{1}{2}}}{16} \frac{ze}{(k_b T)^{\frac{1}{2}}} \left(\frac{1}{m_1} + \frac{1}{m_B} \right)^{\frac{1}{2}} \frac{t_d E}{L} \frac{760}{P} \frac{T}{273.2} \frac{1}{N_0}$$

where z is the charge of the ion, e is elementary charge, N_0 is the number density of the drift gas at STP, k_b is Boltzmann's constant, m_I is the mass of the ion, m_B is the mass of buffer gas, E is the IM electric field strength, L is the drift tube length, P is the buffer gas pressure, and T is the system temperature.

Molecular Dynamics

Molecular dynamics simulations were performed as described previously [12, 30]. Peptide structures were first constructed using Insight II 2000 (Accelerys, San Diego, CA, USA). Starting structures were limited to four different structures: an extended form, α-helix form, 3_{10}-helix form, and π-helix form. For the peptide $[M + H]^+$ ions, we assumed that the proton was on the most basic residue, Lys and N-terminus, whereas for the $[M + Na]^+$ ions, Na^+ was placed in a central position on the peptide backbone. Simulated annealing was performed with the OFF program in Cerius 4.9 (Accelrys) using the Consistent Force Field (CFF 1.02). During the simulation, the starting structures were gradually heated over the course of 280 ps in a step-wise fashion, (relaxation time is 0.1 ps and time step is 0.001 ps), starting and ending at 300 K and peaking at 1000 K. After each annealing cycle, a minimized structure was produced. Annealing cycles were repeated 300× generating 300 minimized structures for each starting structure. Eight more structures were selected from first stage results as the starting structures of the second stage simulations, creating a total of 3600 minimized structures for each peptide. Ω_{calc} of each minimized structure was calculated using the elastic hard sphere scattering (EHSS) method in the MOBCAL software [31].

Dynamic Clustering Algorithm

We develop a novel clustering procedure as an integral part of our methodology to correlate IMS collision cross-section data with candidate structures. Our approach uses an ensemble of many conformations rather than simply the lowest energy structure. Damsbo et al. [24] also use clustering methods for this purpose, although our approach differs from theirs in several respects. First, the pairwise distance for our clustering procedure is the root-mean-squared-deviation (RMSD) of the backbone coordinates calculated between all pairs of simulated structures that correlate to ±2% of Ω_{meas}, whereas Damsbo et al. [24] defined the pairwise distance in terms of backbone torsion angles. Even more fundamental, however, is the fact that the clustering procedures themselves are quite different. Damsbo et al. use the *k*-means clustering procedure which, depending on the initial starting conditions, may converge to sub-optimal configurations. Even if the optimal clustering is found, however, *k*-means only produces a single clustering estimate. Our novel clustering procedure, while more computationally intensive, is robust to the initial starting configuration. Further, rather than merely giving a single clustering estimate, our procedure provides a distribution of clusterings. This allows us to investigate important aspects of the clustering

distribution, including assessing the proximity of clusters and measures of clustering uncertainty.

Intuitively, our clustering procedure repeatedly re-allocates structures among existing or new clusters based on the proximity (as measured by RMSD) of the structures to each other. For any given structure, the probability that it is allocated to an existing cluster is a function of the RMSD values between this structure and those already in the existing cluster. Rather than being forced into the cluster with the highest probability, however, a structure is randomly allocated to clusters based on these probabilities. Further, there is also a chance that a structure may form a new cluster. A key aspect of the algorithm is that the clustering is dynamic in that structures are reallocated among the existing and new clusters in each iteration. Repeated iterations provide a clustering distribution, from which a point estimate of clustering can be obtained and clustering uncertainly can be assessed. As we describe below, we use least-squares clustering [32] to obtain a point estimate and we assess uncertainty using the resulting pairwise probability matrix for whether two structures occupy the same cluster.

Formally, the dynamic clustering procedure can be explained as follows. A clustering of n structures can be represented by a set partition $\pi = \{S_1, \ldots, S_q\}$ of a set $S_0 = \{1, \ldots, n\}$, where the subsets are nonempty, mutually exclusive, and exhaustive. Two structures k and k' are clustered in cluster S if and only both k and k' are in S. The algorithm uses a Markov chain [32] starting at an initially clustering π and updating according to the following transition rule. For $k = 1, \ldots, n$,

$$p(k \in S \mid \cdot) \propto h_k(S), \text{ for all } S \in \pi$$

$$p(k \in S^* \mid \cdot) \propto \alpha,$$

where S^* is an new cluster not currently associated with any other structures and $h_k(S)$ is defined as follows. Let $d_{kk'}$ be the pairwise RMSD between structures k and k'. Let $w_{kk'} \propto (d^* - d_{kk'})^t$ be a measure of proximity of items k and k', where d^* is the overall maximum pairwise distance plus a small increment to ensure that all weights are strictly positive. Note that t has the effect of dampening or accentuating the proximity measurements. For each $k = 1, \ldots, n$, scale $w_{k1}, \ldots, w_{kn}$ such that $\Sigma_{k' \neq k} w_{kk'} = n - 1$. Finally, define $h_k(S) = \Sigma_{k' \in S} w_{kk'}$. Note that the probability of assigning a structure k to a new cluster is $\alpha/(\alpha + n - 1)$ and the probably to being assigned to an existing cluster is $(n - 1)/(\alpha + n - 1)$.

A key theoretical consideration is whether this procedure has an equilibrium distribution. This Markov chain has a finite state space since the number of possible clustering, given by the Bell number, is finite. Since every clustering is accessible from every other clustering through repeated application of the transition rule, the Markov chain is irreducible. The Markov chain is aperiodic because, for every clustering, possibly an application of the transition rule will lead to the same clustering. Finally, since the weights are strictly positive, there is positive probability of returning to every clustering and the Markov chain is recurrent. Since this Markov chain is irreducible, recurrent, and aperiodic, it indeed satisfies the conditions to have an equilibrium distribution.

Following repeated application of the transition rule, many candidate clusterings have been generated. The pairwise probability $\hat{p}_{kk'}$ that two structures k and k' are clustered is estimated by the relative frequency among the candidate clusterings that k and k' occupy the same cluster. The least-squared clustering [32] selects the candidate clustering closest to the estimated pairwise probabilities in terms of squared distances:

$$\pi^{LS} = \operatorname*{argmin}_{\pi \in \{\pi^{(1)}, \ldots, \pi^{(B)}\}} \sum_{k=1}^{n} \sum_{k'=1}^{n} (\delta_{kk'}(\pi) - \hat{p}_{kk'})^2$$

where $\delta_{kk'}(\pi) = 1$ if k and k' occupy the sample cluster in clustering π, and 0 otherwise. This method minimizes of a posterior expected loss of Binder [33] with equal costs of clustering mistakes.

To get an indication of clustering uncertainty, we propose to quantity we call the c-value. The c-value for a structure k in cluster S is the average pairwise probability of being clustered with the other structures in cluster S, i.e.,

$$c - value_k = \frac{\sum_{k' \in S} \hat{p}_{kk'}}{|S|},$$

where $|S|$ is the size the cluster S. If an item has a high c-value, there is more confidence that structure is clustered appropriately.

The clustering uncertainty can also be assessed by plotting a pairwise probability matrix. Arrange the rows and columns by the clustering as indicated by the least-squares clustering and make the color of each element indicate the value of the estimated pairwise probability. This plot makes it easy to see what clusters are well defined and which clusters are closely related or very different from other clusters. We give an example of this plot later in the paper.

Table 1. Ion-neutral collision cross sections of peptide ions measured by MALDI-IM-TOFMS are determined from empirical drift times (t_d) using a hard sphere approximation

Name	Sequence	m/z	Ω_{meas}* (Å²)
$[M + H]^+$	MIFAGIK	779.45	213
	MIFAGIK-O-Met	793.46	216
	Ac-N-MIFAGIK	821.48	223
	Boc-N-MIFAGIK	879.50	239
	Fmoc-N-MIFAGIK	1001.52	251
	MIFAGIK-Fmoc	1001.52	248
$[M + Na]^+$	MIFAGIK+Na	801.44	217
	MIFAGIK-O-Met+Na	815.45	223
	Ac-N-MIFAGIK-Ac+Na	885.47	237

*Errors for all measured collision cross-section values are around ±2%.

In implementing the procedure in this paper, we set $\alpha = 1$ and t at a value between 15 and 45 so as to provide a few well-separated clustering. We applied the transition rule about 500,000 times. In repeated application of the algorithm to the same dataset, the resulting clustering very similar as measured by the adjusted Rand index [34].

Results and Discussion

$[M + H]^+$ Ions of MIFAGIK

All Ω_{meas} of the $[M + H]^+$ and $[M + Na]^+$ ions of MIFAGIK and the various derivatives are listed in Table 1. The $[M + H]^+$ ions of MIFAGIK ($\Omega_{meas} = 213 \pm 4$ Å^2)

is $\sim 5\%$ larger than that expected for ions that fall on the globular mobility-mass trendlines [15], which suggests that the ion structures are somewhat elongated (partial helix). MD simulations as described above yield a total of 3600 candidate structures for the $[M + H]^+$ ions and 631 fall within $\pm 2\%$ of Ω_{meas}. Figure 1a contains a scatter plot of energy versus Ω_{calc} generated by annealing molecular dynamics.

The dynamic clustering procedure described above was applied to the 631 structures that fall within $\pm 2\%$ of Ω_{meas}. The structural similarities among the candidate structures are revealed by plotting the pairwise probability matrix (Figure 1b), where the color of the element (i, j) indicates the estimated probability that structures i
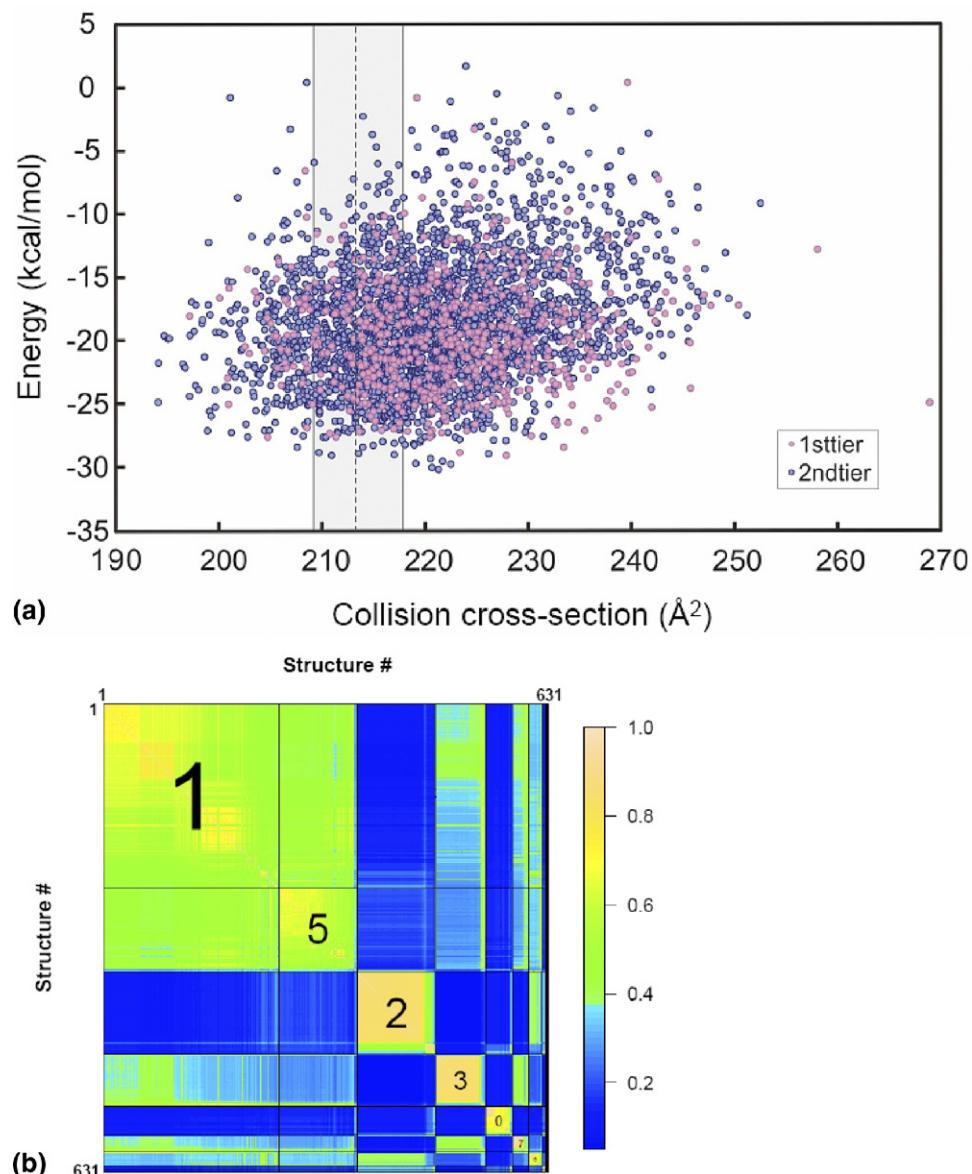
Figure 1. (**a**) Scatter plot of MD energy versus calculated ion-neutral collision cross-section generated by annealing molecular dynamics for [MIFAGIK + H]$^+$ ions. The shaded area indicates $\pm 2\%$ deviation of the measured ion-neutral collision cross-sections; (**b**) the pairwise probability plot of the MIFAGIK peptide ion.

and j have similar backbone coordinates. Possibly structures that fall within a given cluster may in fact share structural elements with another cluster. For example, cluster 1 contains conformations that are similar to each other as well as cluster 5, whereas the structures represented by cluster 2 are very different from those of cluster 5 and cluster 3. (see Figure 2). That is, the tan color indicates high probability of being clustered (i.e., having similar structures), whereas the off-diagonal blocks of dark colors (i.e., dark blue or purple) indicate the clusters do not share structural features. Eleven potential clusters were generated for the 631 candidate structures; however, clusters 1, 5, 2, and 3 contain over 85% of the total structural elements of the ion populations. Backbone projections of all candidate peptide structures within the most populated clusters are overlapped and displayed in Figure 2. The overlaid backbone structures of each cluster show random fluctuations, both in terms of backbone coordinates as well as positioning of the side chains. A representative structure for each cluster is obtained by selecting the structure with the highest c-value in the cluster. The c-value provides an indication of the certainty for each structure in the cluster. That is, a structure with a high c-value indicates a high confidence of being clustered appropriately; thus the structure with the highest c-value should be the best estimate of a cluster.

Note that clusters 1, 5, 2, and 3 represent different conformations owing to coordination of the charge site and specific backbone carboxyl groups. For example, cluster 1 is comprised of structures having a turn in the Met^1-Ile^2-Phe^3-Ala^4-Gly^5 sequence; cluster 5 is best described as a random coil; cluster 2 is characterized by a β-turn over Ile^2-Phe^3-Ala^4-Gly^5 sequence and cluster 3 corresponds to partial helical character over Ile^2-Phe^3-Ala^4 sequence. Note that 'the lowest energy structure' within the Ω_{meas} corresponds to a helical turn conformation over the IFA sequence, similar to cluster 3; however, this structure comprises a minor portion of the total ion population.

Our ultimate goal would be to evaluate the ion structural distribution as a function of energy; however, structure-energy distributions are a function of both energy and reaction dynamics and our current methodologies do not allow for such detailed information. Higher-level calculations are underway to yield meaningful free-energy values for pursuing this important question [20]. Nonetheless, it is interesting to compare the range of energies for each conformation for the four most populated clusters. For example, ~80% of the structures in cluster 3 (α-helix; 11% of the ion population) have energies (Figure 3) that are lower than -20 kcal/mol, which suggests that a greater fraction of the low internal energy ions prefer cluster 3, whereas cluster 1 (α-turn; 39% of the ion population) is comprised of higher energy conformers. These data raise an obvious question, is the structural population among the various clusters dependent on ion internal energy?

Cluster	Backbone Structures	Population Percentage	Representative Structure	Structure Elements	High c-value Structures	C-values
1		39		α-turn (MIFAG)		0.616076 0.616042 0.615796
5		18		Random Coil		0.799788 0.799544 0.799392
2		18		β-turn (IFAG)		0.796034 0.795719 0.795341
3		11		α-helix (IFA)		0.627372 0.626751 0.62523
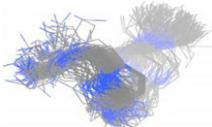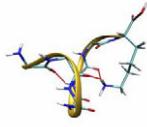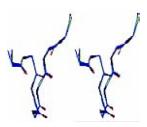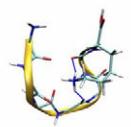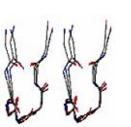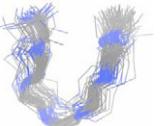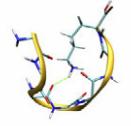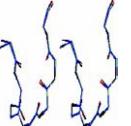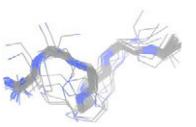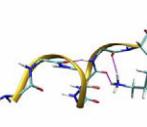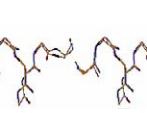
Population: 631 candidate structures

Figure 2. Backbone structures of four most populated clusters for [MIFAGIK + H]$^+$ ions classified by their RMSD values and the structure with highest c-value (representative structure) in each cluster.
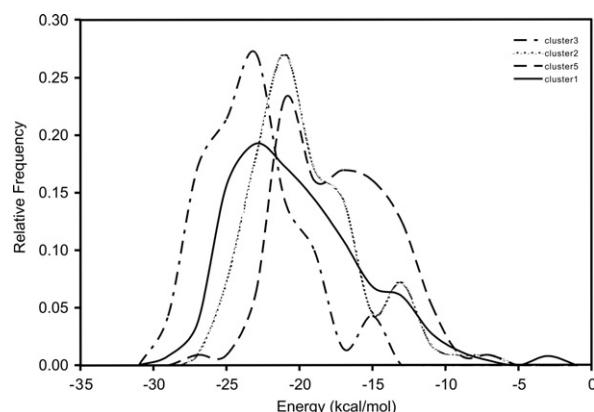
Figure 3. Energy profiles of four most populated clusters for $[MIFAGIK + H]^+$ ions.

Such questions can be probed by producing a greater population of low internal ions or possibly by performing variable-temperature IMS experiments. Now we do not have experimental capabilities for either experiment; however, Clemmer et al. have measured collision cross-sections for MIFAGIK [M + H]+ ions formed by ESI-IMMS, and the Ω_{meas} of 207 ± 4 Å^2 [35] differs by 4% from that measured by MALDI. A total of 458 candidate structures that fall within ±2% of the ESI Ω_{meas} are subjected to cluster analysis (Figure 4 left). Note that the most populated clusters are quite similar in terms of backbone conformations to those formed by

MALDI, i.e., ~34% of the structures have a turn in the Met^1-Ile^2-Phe^3-Ala^4-Gly^5 sequence. The β-turn structure spanning Ile^2-Phe^3-Ala^4-Gly^5 and FAGI sequence are also favored. The partial helical structure over Ile^2-Phe^3-Ala^4 sequence is the least abundant. Thus, apparently the structures of peptide ions formed by ESI and MALDI have similar backbone structures and the difference in collision cross-sections are most likely a result of different projections of side chains; however, it is equally likely that these differences are simply variations in the experimental results. We are currently developing IM-MS instruments that will allow us to perform more extensive comparisons of MALDI and ESI formed ions as well as VT (~90-650K)-IM-MS experiments.

$[M + Na]^+$ Ion of MIFAGIK

The charge site can also affect the conformations of gas-phase peptide ions. Previous studies have shown that the dissociation reactions of $[M + H]^+$ and $[M + Na]^+$ ions differ significantly owing to different charge locations, H$^+$ prefer basic sites and Na$^+$ is very oxyphilic [36]. The MD simulations of MIFAGIK $[M + Na]^+$ ions were performed by placing Na$^+$ ion on the neutral peptide backbone, and the resulting energy versus collision cross-section plot was similar to that shown in Figure 2. The candidate structures that fall within ±2% of the Ω_{meas} (217 ± 4 Å^2 (a total of 654 conformations)

$[MIFAGIK+H]^+$ formed by ESI*

Representative Structure	Structure Elements	Population Percentage
	α-turn (MIFAG)	34
	β-turn (IFAG)	24
	β-turn (FAGI)	14
	α-helix (IFA)	8

Population: 458 candidate structures

$[MIFAGIK+Na]^+$

Representative Structure	Structure Elements	Population Percentage
	Random Coil	66
	Random Coil	18
	Random Coil	11
	Random Coil	5

Population: 654 candidate structures

Figure 4. Backbone structures of four most populated clusters for $[MIFAGIK + H]^+$ ions formed by ESI-IMMS* (left) and $[MIFAGIK + Na]^+$ ions formed by MALDI-IMMS (right). (*Ω_{meas} of $[MIFAGIK + H]^+$ was taken from data published by Clemmer et al. [35].

are subjected to cluster analysis. Cluster analysis yields a total four clusters, and two of these clusters contain ~84% of the total ensemble of representative structures (see Figure 4 right). These structures are characteristic of random coil amide backbones with the Na$^+$ ion surrounded by the amide groups in what appears to be a 'charge-solvated' structure. When superimposed the two conformations comprising clusters 1 and 3 appear to be mirror images. Cluster analysis is performed on [*N-Ac*-MIFAGIK-*N-Ac* + Na]$^+$ and [MIFAGIK-*O-Me* + Na]$^+$ ions as well, the Ω_{meas} are listed in Table 1. These ions prefer random coil conformation that are similar to those for [MIFAGIK + Na]$^+$ ions.

N- and C-terminal Derivatives of MIFAGIK

To test the 'sensitivity' of the cluster analysis method, we introduce subtle variations in peptide ion composition. Protecting groups, which alter the polarity of the N- or C-terminus, may influence the conformational preference of small peptides [37], and such changes are detected by using MD/cluster analysis for MIFAGIK derivatives [MIFAGIK-*OCH*$_3$ + H]$^+$ and [*N-Ac*-MIFAGIK + H]$^+$ ions. The introduction of the methyl and acetyl groups should have minimal effects on ion structure because both groups are relatively small, steric effects should be small relative to charge-solvation by the amide backbone, Lys^7 is the preferred

site of protonation, and the intramolecular interactions between the charged Lys^7 and the backbone carboxyl groups resembles those of MIFAGIK [M + H]$^+$ ions. The Ω_{meas} for [MIFAGIK-*OCH*$_3$ + H]$^+$ and [*N-Ac*-MIFAGIK + H]$^+$ ions are 216 ± 4 Å^2 and 223 ± 4 Å^2, respectively. A total of 768 and 1145 candidate structures that fall within ±2% of the measured collision cross sections were subjected to cluster analysis (Figure 5). The [MIFAGIK-*OCH*$_3$ + H]$^+$ ions adopt similar backbone conformations to MIFAGIK [M + H]$^+$ ions. This result is probably not surprising because the charge carrying site for the methyl ester derivative is the same as that of the underivatized peptide. Note, however, that differences in population size are detected. On the other hand, there are significant differences observed for the [*N-Ac*-MIFAGIK + H]$^+$ ions. For example, the most populated cluster corresponds to random coil structures rather than α-turn structures.

Peptides with Bulky Protecting Group

To further validate the effect of charge site on the conformational preference of peptide MIFAGIK, [MIFAGIK-*N-Fmoc* + H]$^+$ and [*N-Fmoc*-MIFAGIK + H]$^+$ (Fmoc: 9-Fluorenylmethyloxycarbonyl), the peptide MIFAGIK is modified by having the Fmoc protecting group on either N-terminus or Lys side residue. In the case of [*N-Fmoc*-MIFAGIK + H]$^+$, the charge site is

[*N-Ac*-MIFAGIK + H]$^+$

Representative Structure	Structure Elements	Population Percentage
	Random Coil (1)	54
	β-turn (IFAG)	23
	α-helix (IFA)	12
	Random Coil (2)	8.7

Population: 1145 candidate structures

[MIFAGIK-*O-Me* + H]$^+$

Representative Structure	Structure Elements	Population Percentage
	α-turn (MIFAG)	32
	β-turn (IFAG)	30
	Random Coil	17
	α-helix (IFA)	6.5

Population: 768 candidate structures

Figure 5. Backbones of four most populated clusters for [*N-Ac*-MIFAGIK + H]$^+$ ions (left) and [MIFAGIK-*O-Me*+ H]$^+$ ions (right) classified by RMSD values.

on the side-chain of Lys[7], whereas the proton is moved to the N-terminus for [N-Fmoc-MIFAGIK + H]$^+$ ions. Peptide sequences of two peptide ions are confirmed using tandem mass spectrometry and spectra are shown in Supplementary Figure 1, which can be found in the electronic version of this article. As can be seen from the spectra, all the mass shifts are on y ions for [MIFAGIK-N-Fmoc + H]$^+$ ions and all the mass shifts are on b ions in the case of [N-Fmoc-MIFAGIK + H]$^+$.

The ATD plots for the [MIFAGIK-N-Fmoc + H]$^+$ and [N-Fmoc-MIFAGIK + H]$^+$ ions are shown in Supplementary Figure 2. The centroid of ATD of [N-Fmoc-MIFAGIK + H]$^+$ is consistently *ca.* 8 μs longer than that for [MIFAGIK-N-Fmoc + H]$^+$ ions, and the calculated collision cross-sections for [N-Fmoc-MIFAGIK + H]$^+$ ions is *ca.* 3 Å^2 larger than that of [MIFAGIK-N-Fmoc + H]$^+$ ions. It is unclear whether these differences arise as a result of different structures or differences in the spatial projections of the Fmoc group, but more detailed studies on similar effects are currently underway.

Cluster analysis was performed on ~1000 candidate structures extracted from MD simulations for both peptide ions and results are illustrated in Figure 6. For [N-Fmoc-MIFAGIK + H]$^+$, three clusters contain over 100 structures. The dominant conformer represents ~60% of the total ensemble, and this conformer appears to adopt a random coil structure, which is attributed to hydrogen bonding between the protonated Lys[7] and backbone amide groups. The remaining most populated clusters exhibits a turn in the Ile^2-Phe^3-Ala^4-Gly^5 sequence or a helical turn in Ile^2-Phe^3-Ala^4 sequence,

which are similar to the most populated structures obtained for [MIFAGIK + H]$^+$ ions. In the case of [MIFAGIK-N-Fmoc + H]$^+$ ions, the two most populated clusters account for over 70% of the total ensemble. Interestingly, both conformations contain a bridge between Ile^2 and Gly^5 which is quite different from that for peptides with the charge on Lys[7]. It is interesting to note that helical conformations are not found for this ion since the preferred charge site of [MIFAGIK-N-Fmoc + H] is probably the N-terminus and the helix is probably destabilized by the interaction between the charge and the helix macrodipole.

Conclusions

Cluster analysis of IM-MS data is a powerful approach for classification of peptide structures derived from molecular dynamics simulations. Each cluster represents probable conformations observed for gas-phase ions. Specifically for the peptide MIFAGIK, cluster analysis suggests that [M + H]$^+$ ions formed by MALDI or ESI both prefer a α-turn structure; this structural preference is probably the result of stabilization afforded by charge-solvation by the backbone amide groups. That is, there appears to be a significant preference for helical or partial helical conformers by the low-energy ions, but the major fraction of the ion populations exists as α-turn, random coil, and β-turn conformers. Although the derivatives of MIFAGIK that influence charge site and intramolecular interactions alter the distribution of various conformers, the most

[N-Fmoc-MIFAGIK+H]$^+$

Representative Structure	Structure Elements	Population Percentage
	Random Coil	61
	β-turn (IFAG)	17
	β-turn (MIFA)	11
	Helix (IFA)	6.2

Population: 1053 candidate structures

[MIFAGIK-N-Fmoc+H]$^+$

Representative Structure	Structure Elements	Population Percentage
	Random Coil	35
	β-turn (IFAG)	35
	Random Coil	8.5
	β-turn (IFAG)	7.6

Population: 1229 candidate structures
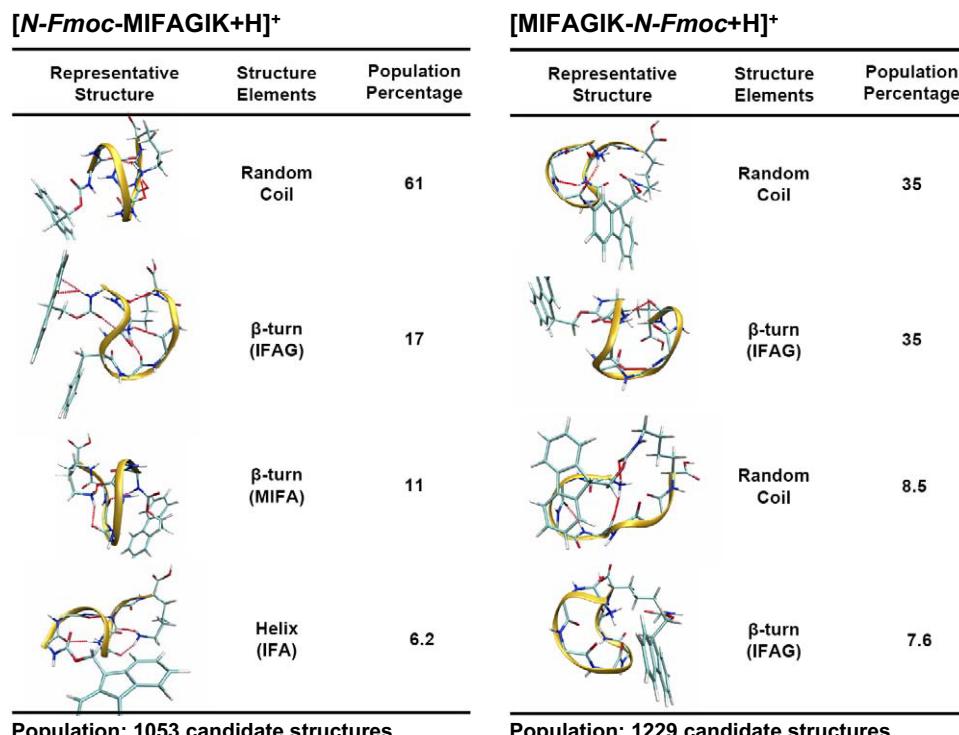
Figure 6. Four most populated clusters of [N-Fmoc-MIFAGIK + H]$^+$ ions (left) and four most populated clusters of [MIFAGIK-N-Fmoc+ H]$^+$ ions (right) classified by RMSD values.

dramatic changes are observed for the $[M + Na]^+$ ion, which show a strong preference for random coil conformers owing to the strong solvation by the backbone amide groups. We tested the 'sensitivity' of the cluster analysis method by introducing subtle variations in peptide ion composition as well. For example, the N-acetyl and methylester derivatives as well as the Boc and Fmoc derivations all share very similar backbone structures. All the results presented here support the hypothesis that difference in collision cross-sections for MIFAGIK peptide ions are related to the different orientation of side chains, i.e., MIFAGIK is a small peptide which has limited degrees-of-freedom for positioning backbone structural elements. We are currently expanding these studies to studies of the results to larger peptide systems in an effort to increase statistical confidence of gas-phase conformation assignment.

Acknowledgments

Appendix A
Supplementary Material

Supplementary material associated with this article may be found in the online version at doi:10.1016/j.jasms.2009.048.018.

References

1. Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005**, *310,* 1658–1661.
2. Kaddis, C. S.; Lomeli, S. H.; Yin, S.; Berhane, B.; Apostol, M. I.; Kickhoefer, V. A.; Rome, L. H.; Loo, J. A. Sizing Large Proteins and Protein Complexes by Electrospray Ionization Mass Spectrometry and Ion Mobility. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 1206–1216.
3. Kaddis, C. S.; Loo, J. A. Native protein MS and Ion Mobility: Large Flying Proteins with ESI. *Anal. Chem.* **2007,** *79,* 1778–1784.
4. Loo, J. A., Kaddis, C. S. Direct characterization of protein complexes by electrospray ionization mass spectrometry and ion mobility analysis. John Wiley & Sons, Inc.; Hoboken, NJ, 2007; pp. 1–23.
5. Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chem. Int. Ed.* **2007,** *46,* 8001–8004.
6. Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protocols* **2008,** *3,* 1139–1152.
7. Uetrecht, C.; Versluis, C.; Watts, N. R.; Wingfield, P. T.; Steven, A. C.; Heck, A. J. R. Stability and shape of hepatitis B virus capsids in vacuo. *Angew. Chem. Int. Ed.* **2008,** *47,* 6247–6251.
8. Barrera, N. P.; Di Bartolo, N.; Booth, P. J.; Robinson, C. V. Micelles Protect Membrane Complexes from Solution to Vacuum. *Science* **2008,** *321,* 243–246.
9. Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Naked Protein Conformations: Cytochrome *c* in the Gas Phase. *J. Am. Chem. Soc.* **1995,** *117,* 10141–10142.
10. Shelimov, K. B.; Jarrold, M. F. Conformations, Unfolding, and Refolding of Apomyoglobin in Vacuum: An Activation Barrier for Gas-Phase Protein Folding. *J. Am. Chem. Soc.* **1997,** *119,* 2987–2994.
11. Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. Design of Helices hat are Stable in Vacuo. *J. Am. Chem. Soc.* **1998,** *120,* 12974–12975.
12. Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Gillig, K. J.; Russell, D. H. Observation of conserved solution-phase secondary structure in gas-phase tryptic peptides. *J. Am. Chem. Soc.* **2002,** *124,* 4214–4215.
13. Ruotolo, B. T.; Russell, D. H. Gas-phase conformations of proteolytically derived protein fragments: Influence of solvent on peptide conformation. *J. Phys. Chem. B* **2004,** *108,* 15321–15331.
14. McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. Ion mobility-mass spectrometry: A new paradigm for proteomics. *Int. J. Mass Spectrom.* **2005,** *240,* 301–315.
15. Tao, L.; McLean, J. R.; McLean, J. A.; Russell, D. H. A Collision Cross-Section Database of Singly-Charged Peptide Ions. *J. Am. Soc. Mass Spectrom.* **2007,** *18,* 1232–1238.
16. Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H. Distinguishing between phosphorylated and nonphosphorylated peptides with ion mobility-mass spectrometry. *J. Proteome Res.* **2002,** *1,* 303–306.
17. Ruotolo, B. T.; Gillig, K. J.; Woods, A. S.; Egan, T. F.; Ugarov, M. V.; Schultz, J. A.; Russell, D. H. Analysis of phosphorylated peptides by ion mobility-mass spectrometry. *Anal. Chem.* **2004,** *76,* 6727–6733.
18. McLean, J. R.; McLean, J. A.; Wu, Z.; Becker, C.; Pérez, L. M.; Pace, C. N.; Scholtz, J. M.; Russell, D. H. Factors that Influence Helical Preferences for Singly-Charged Gas-Phase Peptide Ions: The Effects of Multiple Charge-Carrying Sites. *J. Am. Chem. Soc.,* submitted.
19. Wilson, S. R.; Cui, W. Applications of simulated annealing to peptides. *Biopolymers* **1990,** *29,* 225–235.
20. Fernandez-Lima, F. A.; Wei, H.; Gao, Y. Q.; Russell, D. H. On the structure elucidation using IMS and Molecular Dynamics. *J. Phys. Chem. A,* submitted.
21. Stearns, J. A.; Boyarkin, O. V.; Rizzo, T. R. Spectroscopic signatures of gas-phase helices: Ac-Phe-(Ala)5-Lys-H+ and Ac-Phe-(Ala)10-Lys-H+. *J. Am. Chem. Soc.* **2007,** *129,* 13820–13821.
22. Stearns, J. A.; Guidi, M.; Boyarkin, O. V.; Rizzo, T. R. Conformation-specific infrared and ultraviolet spectroscopy of tyrosine-based protonated dipeptides. *J. Chem. Phys.* **2007,** *127,* 154322.
23. Stearns, J. A., Mercier, S., Seaiby, C., Guidi, M., Boyarkin, O. V.; Rizzo, T. R. Conformation-specific spectroscopy and photodissociation of cold, protonated tyrosine and phenylalanine. *J. Am. Chem. Soc.* **2007,** *129,* 11814–11820.
24. Damsbo, M.; Kinnear, B. S.; Hartings, M. R.; Ruhoff, P. T.; Jarrold, M. F.; Ratner, M. A. Application of evolutionary algorithm methods to polypeptide folding: Comparison with experimental results for unsolvated Ac-(Ala-Gly-Gly)5-LysH+. *Natl. Acad. Sci. U.S.A.* **2004,** *101,* 7215–7222.
25. Carpino, L. A.; Han, G. Y. 9-Fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group. **1970,** *92,* 5748–5749.
26. Reid, G.; Simpson, R.; O'Hair, R. J. A mass spectrometric and ab initio study of the pathways for dehydration of simple glycine and cysteine-containing peptide $[M + H]+$ ions. *J. Am. Soc. Mass Spectrom.* **1998,** *9,* 945–956.
27. Gillig, K. J.; Ruotolo, B. T.; Stone, E. G.; Russell, D. H.; Fuhrer, K.; Gonin, M.; Schultz, J. A. Coupling high-pressure MALDI with ion mobility/orthogonal time-of-flight mass spectrometry. *Anal. Chem.* **2000,** *72,* 3965–3971.
28. Stone, E.; Gillig, K. J.; Ruotolo, B.; Fuhrer, K.; Gonin, M.; Schultz, A.; Russell, D. H. Surface-induced dissociation on a MALDI-ion mobility-orthogonal time-of-flight mass spectrometer: Sequencing peptides from an "in-solution" protein digest. *Anal. Chem.* **2001,** *73,* 2233–2238.
29. Mason, E. A.; McDaniel, E. W. Transport Properties of Ions in Gases; Wiley: New York, 1988; pp 1–29
30. Sawyer, H. A.; Marini, J. T.; Stone, E. G.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H. The structure of gas-phase bradykinin fragment 1-5 (RPPGF) ions: an ion mobility spectrometry and H/D exchange ion-molecule reaction chemistry study. *J. Am. Soc. Mass Spectrom.* **2005,** *16,* 893–905.
31. Shvartsburg, A. A.; Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996,** *261,* 86–91.
32. Dahl, D. B. *Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model*; Cambridge University Press: Cambridge, 2006; pp 201–218.
33. Binder, D. A. Bayesian Cluster Analysis. *Biometrika* **1978,** *65,* 31–38.
34. Hubert, L.; Arabie, P. Comparing Partitions. *J. Classification* **1985,** *2,* 193–218.
35. Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. A database of 660 peptide ion cross sections: Use of intrinsic size parameters for bona fide predictions of cross sections. *J. Am. Soc. Mass Spectro.* **1999,** *10,* 1188–1211.
36. Lee, S.-W.; Kim, H. S.; Beauchamp, J. L. Salt Bridge Chemistry Applied to Gas-Phase Peptide Sequencing: Selective Fragmentation of Sodiated Gas-Phase Peptide Ions Adjacent to Aspartic Acid Residues. *J. Am. Chem. Soc.* **1998,** *120,* 3188–3195.
37. Ganesh, S.; Jayakumar, R. Role of N-t-Boc group in helix initiation in a novel tetrapeptide. *J. Peptide Res.* **2002,** *59,* 249–256.