Assessing Intrinsic Side Chain Interactions between *i* and *i* + 4 Residues in Solvent-Free Peptides: A Combinatorial Gas-Phase Approach†

Catherine A. Srebalus Barnes‡,§ and David E. Clemmer*,‡

Department of Chemistry, Indiana University, Bloomington Indiana 47405, and Eli Lilly and Company, Indianapolis, Indiana 46285

Received: April 28, 2003; In Final Form: September 2, 2003

Ion mobility measurements and molecular modeling techniques have been used to survey the gas-phase structures of a series of alanine-rich peptides. The peptides, examined as $[M + 2H]^{2+}$ ions, have the general forms $NH_2-(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3$ and $Ac-(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3$, where residues 8 and 12 are randomized. In total, 160 different peptide ions (80 related NH_2-terminated and -acetylated sequences) have been studied. Substitutions of residues 8 and 12 permit an assessment of the influence of specific interactions between residues in an adjacent helical turn. The formation of helices and globular structures in the gas phase appears to be sensitive to specific interactions between amino acid side chains. A preliminary discussion of these results in terms of what is currently known about helix formation in the gas phase and in solution is given. Overall, it appears that this combinatorial approach to studying sequence-to-structure interactions that are intrinsic to the peptide is a viable strategy for surveying trends in large numbers of sequences without interference from solvation effects.

Introduction

Central to understanding protein folding is the question of what types of interactions favor helices. From early crystallographic data, Chou and Fasman assessed the propensities of amino acids to be found in different secondary structural motifs.[1] Baldwin and co-workers examined the relative stabilities of helices in alanine-rich peptides having different amino acids introduced at specific positions in the sequence.[2,3] Although these efforts (and many others) provide benchmarks about how variations in an amino acid sequence influence structure, it is still difficult to predict how a given fold will be influenced when a specific residue is substituted into a polypeptide sequence. At least in part, this difficulty comes about because sequence alone does not define the fold; protein structure also depends on environment. That is, variations of sequence in solution will also affect factors other than structure, such as solubility or propensity to form aggregates.

During the last several years, the structures of isolated proteins and peptides in the gas phase have been examined. These studies are possible because of developments in mass spectrometry techniques that allow the gentle production of solvent-free macromolecular ions.[4] Studies of peptide structures in vacuo should help delineate factors that influence structure in solution. Several techniques for studying the structures of peptides in the gas phase are currently under development, including isotopic H/D exchange of trapped ions;[5] measurements of ion collision cross sections[6] and neutral dipole moments;[7] studies of gas-phase ion reactivity[8] and fragmentation;[9] as well as molecular mechanics simulations.[10] In the present paper, we have measured cross sections for a series of 80 related peptide sequences, having the general form NH_2- and $Ac-(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3COOH$ (160 different peptide ions in total). In cases where α-helical motifs arise (that is, the formation of networks of hydrogen-bonding interactions between the amide backbone group of residue *i* and the carbonyl group of residue *i* + 4 along the polypeptide chain), substitution at residues 8 and 12 allows us to assess whether specific interactions are favorable or disruptive to helix formation. The analysis of many related sequences provides information about general characteristics of side chain interactions in the absence of solvation effects.

Factors that influence the secondary structure of peptides in solution have been studied extensively. In a recent review, Baldwin summarized the dominant factors that influence the formation α-helices: (1) the intrinsic helix propensities of the component amino acid residues; (2) interaction of charged residues with the helix macrodipole; (3) N-terminal and C-terminal capping interactions; and (4) side chain interactions between residues in adjacent helical turns.[3] Evidence from structural data for proteins[1] and peptides[11] indicates that the propensity of an amino acid to be found in helices increases as Glu < Gln < Lys < Leu < Arg < Ala. The Gly and Pro residues have low helix propensities and generally disrupt helix formation. The second factor, charge stabilization of the helix macrodipole,[12] involves interactions between charged amino acid side chains and the dipoles that arise from the orientation of NH and CO groups along the backbone. These interactions are typically distinguished from N- or C-capping interactions (factor 3)[13] that involve hydrogen bonding between polar (or charged) side chains with terminal amide or carbonyl groups.

Most of the studies that have examined how side chain interactions between residues (factor 4)[14] influence helix formation use sequence scaffolds that are known to be helical and maintain solubility as substitutions of residues at specific positions are made. A number of different $i \rightarrow i + 4$ residue combinations appear to stabilize α helices, including interactions between nonpolar−nonpolar,[15] aromatic−basic,[16] acidic−basic,[17] and polar−polar[18] residue pairs. Additionally, the

† Part of the special issue "Charles S. Parmenter Festschrift".
* To whom correspondence should be addressed.
‡ Indiana University.
§ Eli Lilly and Company.

10.1021/jp030519s CCC: $25.00
Published on Web 11/14/2003

TABLE 1: Specific Side-Chain Interactions that Stabilize Peptide Helices[a]

interactions	examples
nonpolar−nonpolar	Leu−Ile
	Leu−Leu
	Leu−Val
	Tyr−Leu
	Tyr−Val
aromatic−basic	Phe−His
	Trp−His
acidic−basic	Asp−Lys
	Asp−Arg
	Glu−Lys
	Glu−Arg
	His−Asp
polar−polar	Gln−Asp
	Gln−Glu

[a] Adapted from Chakrabartty, A.; Baldwin, R. L. *Adv. Protein Chem.* **1995**, *46*, 141 and Baldwin, R. L.; Rose, G. D. *Trends Biol. Sci.* **1999**, *24*, 26.

unusual stabilities of the folded states of some thermophilic proteins have been attributed to multiple ion-pairing interactions between the side chains of amino acids at i and $i + 4$ positions.[19] Table 1 provides a summary of various interactions (and example residue combinations) that stabilize helices in solution.

Although fewer studies of factors that influence structure in the gas phase have been conducted, there is now evidence that peptides can form several types of structures, including tightly packed globules,[20,21] extended helical states,[22,23] and a folded state resembling a hinged helix coil.[23] Jarrold's group has studied a series of polyamino acids and reported the relative helix propensities increase as Gly < Val < Ala <Leu,[24] an ordering that deviates from the Gly < Val < Leu < Ala scale reported from solution work. It also appears that the role of side chain entropy in vapor phase peptides differs from that in solution.[11] Our group has derived information about the average sizes,[25] and also volumes,[26] of amino acids in heterogeneous sequences (average results from a database of measurements of tryptic peptides). The volumes of individual residues in the gas phase can be substantially smaller than values for crystals, indicating that globular conformations pack with high efficiency.

In the present paper, we utilize information about general motifs and the sizes of individual residues to assess $i \rightarrow i + 4$ interactions in a series of related sequences that are generated by a combinatorial synthetic strategy. We have chosen NH_2-terminated and N-terminal acetylated forms of 15 residue peptides having the general form $(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3-CO_2H$, where Xxx is varied over the Gly, Ala, γ-Abu,[27] Ser, Pro, Val, Thr, Leu, Asn, Asp, Gln, Lys, Glu, Phg, His, Phe, Cha, Arg, Tyr, and Phe(NO_2) residues and Yyy represents Ser, Leu, Glu, and Lys. This polymer length falls in a range that makes it possible to assess whether ions have roughly globular or helical structures. Additionally, this paper is the first to address the structures of $[M + 2H]^{2+}$ ions in detail. Information about $[M + 2H]^{2+}$ ion structure is of topical interest because of recent attempts to understand peptide dissociation patterns in the emerging area of proteomics.

Experimental Section

Ion Mobility/Time-of-Flight Measurements. Detailed descriptions of the apparatus used to record collision cross sections are given elsewhere.[28,29] Only a brief description is given here. Solutions of N-terminal acetylated and nonacetylated forms of four peptide libraries (NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Ser-(Ala)_3-CO_2H$, NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Leu-(Ala)_3-CO_2H$, NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Lys-(Ala)_3-CO_2H$, and NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Glu-(Ala)_3-CO_2H$) were prepared [$\sim 10^{-5}$ M (per peptide) in 45:45:10 trifluoroacetic acid (TFA):water:acetic acid]. The solutions were electrosprayed directly into the source region of a high pressure drift tube coupled to a time-of-flight mass spectrometer. Ions enter the instrument into a differentially pumped desolvation region and are then pulled to the front of the drift region using a weak electric field. We estimate that ions spend ~20 to 60 ms in the source region of the instrument before reaching the front of the drift region. Mobility experiments are initiated by gating a short packet of ions (200 μs in duration) into the drift region. In these studies, the drift tube was operated with helium buffer gas pressures of ~150 to 170 Torr and a uniform electric field of 171.0 V cm^{-1}. The mobilities of ions depend on both the average collision cross section and charge state of the ions. For a given charge state, compact structures will have higher mobilities than elongated structures. Higher charge state ions experience a greater drift force than low charge states having a similar cross section and thus have higher mobilities.

Ions that exit the drift tube are focused into the source region of a reflectron geometry time-of-flight (TOF) mass spectrometer. High frequency pulses, synchronous with lower frequency mobility pulses, are used to initiate the TOF experimental sequence. Drift times through the high pressure drift region are substantially longer (ms) than flight times (μs) in the evacuated TOF region. This difference in experimental time scales permits the acquisition of hundreds of flight time distributions for each pulse of ions introduced into the drift tube. We refer to this as a nested drift(flight) time measurement and denote peaks using units of ms(μs) in the nomenclature proposed previously.[28] Because arrival times at the TOF detector are known with respect to both the mobility and TOF pulses, drift and flight time information can be accumulated for the mixture of ions without changing any experimental parameters. Thus, even small relative differences in mobilities for different sequences are highly reproducible.

Peptide Synthesis. Mixtures of peptides were synthesized using a combinatorial approach. Four different 15 residue alanine-based peptide libraries were designed to incorporate specific combinations of amino acid side chain functional groups at the i and $i + 4$ positions. Substitution of a single Ser, Leu, Lys, or Glu residue at position twelve ($i + 4$) and randomization over 20 different amino acids at position eight (i) resulted in four libraries of the general form NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Ser-(Ala)_3-CO_2H$, NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Leu-(Ala)_3-CO_2H$, NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Lys-(Ala)_3-CO_2H$, and NH_2/Ac$-(Ala)_7-Xxx-(Ala)_3-Glu-(Ala)_3-CO_2H$ (where Xxx corresponds to Gly, Ala, γ-Abu, Ser, Pro, Val, Thr, Leu, Asn, Asp, Gln, Lys, Glu, Phg, His, Phe, Cha, Arg, Tyr, and Phe(NO_2)). Randomization at the Xxx position in the Leu library was limited to only eighteen amino acids (Phg and Cha residues were omitted).

The Ser, Leu, Lys, and Glu libraries were synthesized separately in nitrogen-agitated reaction vessels using standard mix and split protocols[30] with Fmoc (fluorenylmethoxycarbonyl) peptide chemistry.[31] Following incorporation of the N-terminal amino acid, each library was split into two separate fractions. The first fraction was reserved for peptide cleavage to yield peptides with free amino termini, and the second fraction was subjected to N-terminal acetylation using acetic anhydride. The acetylated and nonacetylated fractions of each library were cleaved from the solid support and side chain protecting groups removed using a trifluoroacetic acid (TFA):phenol:water:

thioanisole:ethanedithiol solution (82.5:5:5:5:2.5 by volume). Resin peptides were precipitated in ether, washed several times with ether, and vacuum-dried. The resulting library peptides were dissolved in 20% aqueous 2-propanol and lyophilized.

Assignment of Experimental Drift(Flight) Time Peaks to Specific Sequences. Peaks in the drift(flight) time distributions are assigned as described in detail previously. Briefly, experimental flight times are converted to mass-to-charge (m/z) ratios using a multipoint calibration. In the present system, the drift-(flight) time distributions are dominated by a series of $[M + 2H]^{2+}$ ions.[32] Comparisons of m/z ratios (derived from flight times) with calculated values for $[M + 2H]^{2+}$ ions of the expected library peptides allow us to unambiguously assign nearly all of the peaks in each mixture. One pair of peaks in each library (corresponding to substitution of Gln or Lys) cannot be unambiguously assigned based on comparisons of experimental and calculated m/z ratios; the m/z difference between Gln and Lys in $[M + 2H]^{2+}$ ions is only ~0.1. From other studies, it appears that the Lys, His, and Arg residues often influence cross sections in a similar fashion. Thus, for the analysis presented below, we distinguish between peaks for sequences containing Lys and Gln by assigning the peak having a drift time that is similar to the analogous His and Arg peptides to the Lys peptide.

We also note that in a number of cases we do not observe peptides containing the γ-Abu and Phe(NO_2) residues. Coupling efficiencies for these amino acids are substantially lower than those for the other residues.[33]

Experimental Collision Cross Sections. Experimental collision cross sections are obtained from the relation[34]

$$\Omega(\mathrm{exp}) = \frac{(18\pi)^{1/2}}{16}\frac{ze}{(k_bT)^{1/2}}\left[\frac{1}{m_I}+\frac{1}{m_B}\right]^{1/2}\frac{t_DE}{L}\frac{760}{P}\frac{T}{273.2}\frac{1}{N} \quad (1)$$

where t_D is the experimental drift time, E is the electric field, T and P are the temperature and pressure of the buffer gas, L is the length of the drift region, N is the neutral number density, k_b is Boltzmann's constant, and m_I and m_B are the mass of the ion and the buffer gas, respectively. The experimental parameters in eq 1 can be measured precisely, and therefore, experimental collision cross section measurements are also expected to be very accurate. Usually, any two measurements of a collision cross section for any ion using different instruments in our lab differ by less than 1% (relative uncertainty). In cases where we can compare with other labs, differences are usually less than ±2%.

Molecular Modeling and Calculation of Cross Sections for Model Conformers. Molecular modeling simulations were performed using the Insight II[35] software package and the ESFF (Extensible Systematic force field) to obtain energy minimized model conformers for representative $[M + 2H]^{2+}$ library peptide ions. The simulations were carried out for a minimum of 250 ps at 300 K using a dielectric of 1.0. A total of 50 structures were obtained for each 250 ps simulation. In some cases, simulated annealing protocols employing temperatures of 600 K were utilized to obtain representative compact random globular structures. A variety of initial peptide conformations were used in the simulations including α helices, π helices, trans-extended structures, and random globules.

Theoretical collision cross sections [Ω(calc)] were calculated for each conformer obtained in the simulations using an exact hard spheres scattering (EHSS)[36] method that has been previously calibrated[37] to values from the more accurate trajectory method[38] using a comparison of cross section values obtained from EHSS and trajectory calculations for globular and helical $[Ala_n + H]^+$ ions.[22,20,39] The calibration was verified by examining a few sequences with both methods, and the calibrated EHSS values are expected to agree with values for the more rigorous approach to within ±1% (relative difference). Structures were considered candidate structures if their calculated collision cross sections were within 2% of the experimental cross section. The lowest-energy candidate structures within this cross section range are considered as structural types. A more detailed discussion of structural assignments is given below.

Grouping of Ions into Structural Types. The experimental cross sections and calculated cross sections for trial conformers have been grouped into structural types by a procedure that is described in detail below. This procedure makes it possible to remove differences in cross sections that arise from the different intrinsic sizes of individual amino acid side chains (which are substituted into the 8th and 12th positions along the polymer chain). When the differences in side chains are accounted for, we find that there are several structural types. As described below, we have divided structural types as helical [a designation that implies an elongated geometry in which the secondary structure is made up of helical turns -hydrogen bonding interactions between i(amide) and $i + 4$(carbonyl) backbone groups], globular (a designation that that implies compact geometries with relatively random structures), and other (meaning that the data do not seem to fall within the limits expected for helical or globular structures).

Results

Drift(Flight) Time Distributions for the NH_2−(Ala)$_7$−Xxx−(Ala)$_3$−Leu−(Ala)$_3$−CO_2H and Ac−(Ala)$_7$−Xxx−(Ala)$_3$−Leu−(Ala)$_3$−CO_2H Libraries. Figure 1 shows drift-(flight) time distributions for the NH_2−(Ala)$_7$−Xxx−(Ala)$_3$−Leu−(Ala)$_3$ and Ac−(Ala)$_7$−Xxx−(Ala)$_3$−Leu−(Ala)$_3$ libraries. Peaks for the NH_2-terminated peptides fall into two distinct conformer families: a high-mobility family observed at relatively short drift times for all of the sequences, and a low-mobility family at longer drift times for the Xxx = Ala, Val, Leu, Gln/Lys, His, Arg, and Tyr sequences. Note that the mass difference between the Gln and Lys residues is too small to be distinguished here. Peaks in the drift(flight) time data for the acetylated peptides occur at slightly longer drift times than the corresponding low-mobility ions observed for the NH_2-terminated peptides. Presumably, within the family, different sequences have similar overall shapes. The slight decrease in the mobility of the low-mobility family that is observed upon acetylation (~2%) is near what is expected from the additional size associated with the acetyl group.

An initial impression about the types of structures associated with these families can be obtained by considering one of the sequences in more detail. Below, we show that when the sizes of different residues are accounted for the structural types appear to be general to most of the residue substitutions that have been examined. The NH_2-terminated Ala(i) → Leu($i + 4$) sequence exhibits peaks at drift times of 15.7 and 17.1 ms, and from eq 1, experimental cross sections of 256.7 and 281.7 Å^2, respectively. The 256.7 Å^2 value is near the value of $\Omega[Ala_{15} + 2H]^{2+}$ = 253.3 Å^2 that we reported previously, which corresponds to a compact globule for a pure polyalanine peptide of the same length.[23] Our previous molecular modeling studies of $[Ala_{15} + 2H]^{2+}$ indicated that this ion assumes a compact globular conformation in which the backbone carbonyl groups solvate protonated sites along the polypeptide chain. Thus, it appears that the large peak at 15.7 ms corresponds to a compact globule.

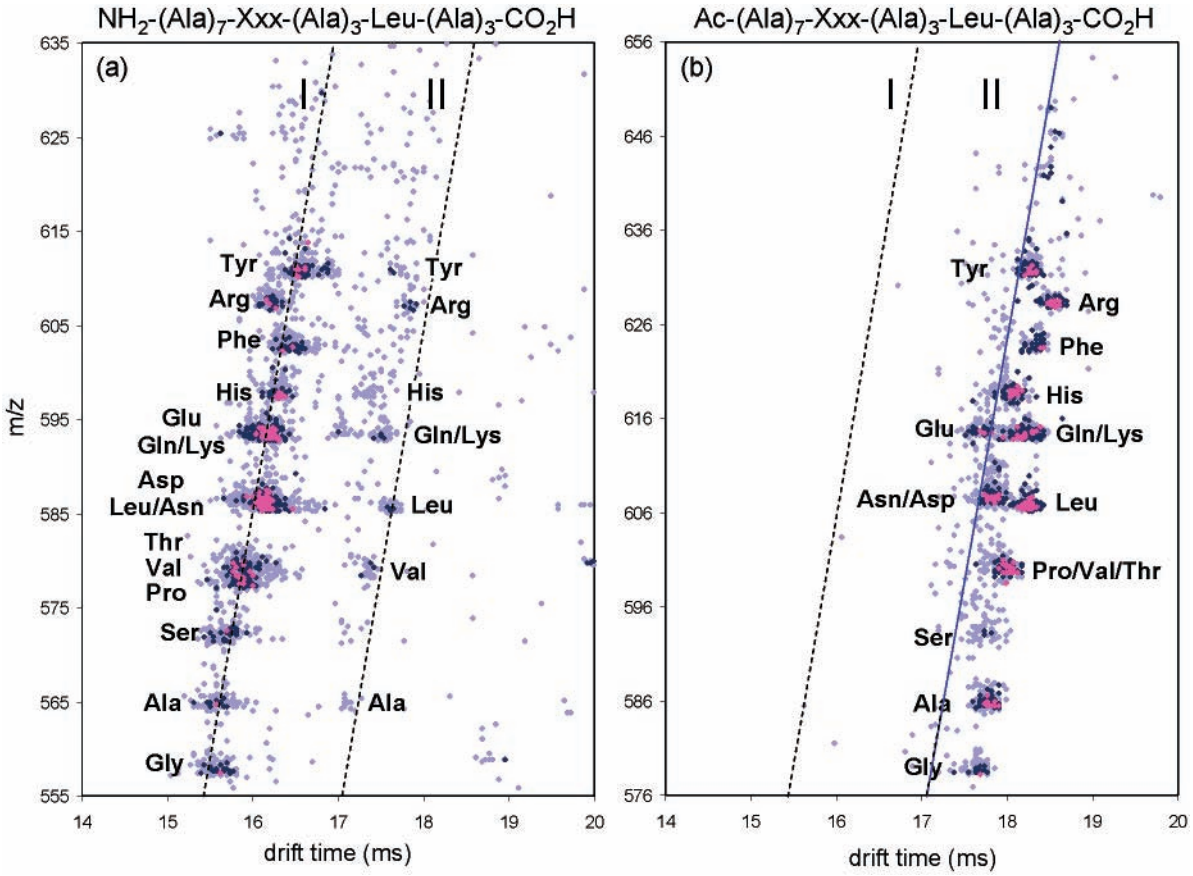

Figure 1. Ion mobility/time-of-flight data for the $NH_2-(Ala)_7-Xxx-(Ala)_3-Leu-(Ala)_3-CO_2H$ peptide library (part a) and the $Ac-(Ala)_7-Xxx-(Ala)_3-Leu-(Ala)_3-CO_2H$ (part b). These data were recorded using helium buffer gas pressures of 157.8 (a) and 158.0 Torr (b) and a drift field of 170.97 V cm^{-1}. Drift times have been normalized to 150.0 Torr to allow direct comparison of experimental drift times for different data sets. Peak intensities are indicated in the plots using a false color scale in which darker colors represent higher intensity features. To represent low intensity peaks in the IMS/TOF distributions, no baseline intensity cutoff was used. Average drift times for high and low mobility families (types I and II, respectively) are indicated using the horizontal dashed lines. Peaks are labeled according to the identity of the Xxx amino acid residue.

The 281.7 $Å^2$ cross section associated with the larger Ala(i) → Leu(i + 4) conformation is near the value of 291.5 $Å^2$ reported by Jarrold and co-workers for the 15-residue Lys-terminated peptide, $[Ala_{14}-Lys + H]^+$, that favors extended helical structures.[20] Thus, from these simple considerations, we determine that the high- and low-mobility families appear to correspond to compact and relatively extended structures.

Information About the Relative Abundances of Different Structural Types. Information about the relative abundances of the different structural types can be obtained by examining ion mobility distributions for specific sequences, obtained by integration of the drift(flight) time data over specific flight time windows. Figure 2 shows examples for NH_2- and Ac-terminated peptides for the Ala(i) → Leu(i + 4), Leu(i) → Leu(i + 4), and Phe(i) → Leu(i + 4) substitutions. The NH_2-terminated sequences are dominated by peaks corresponding to the higher-mobility compact ions. A smaller peak (associated with the arrival of lower-mobility elongated ions) is also observed for the Ala(i)→Leu(i + 4), Leu(i)→Leu(i + 4) sequences; however, this feature is much smaller in the Phe(i)→Leu(i + 4) peptide. As noted above, these low-mobility peaks occur near a large peak that dominates the drift time distributions recorded for the analogous Ac-terminated peptides.

Overall, the trends that are shown in these three sequences appear to be general to all of the sequences that we have studied (including those that are discussed below for the other libraries). That is, nearly all of the amino-terminated sequences show evidence for a compact state [the only exceptions being those sequences that involve two basic substitutions: $NH_2-Lys(i)$ → Lys(i + 4), $NH_2-His(i)$ → Lys(i + 4), and $NH_2-Arg(i)$ → Lys(i + 4)]. Many of the distributions for the amino-terminated sequences also show evidence for a peak at longer times (at drift times that are similar to the corresponding Ac-terminated form), and the relative abundance depends on the peptide sequence. For the analysis that follows, we consider any reproducible feature that comprises more than 5% of the distribution to be a peak. Features that comprise less than 5% of a distribution are often not observed in replicate experiments.

Drift(Flight) Time Distributions for the $NH_2-(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3$ and $Ac-(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3$ Libraries. Figures 3 and 4 show drift(flight) time distributions for the NH_2- and Ac-terminated forms of analogous libraries corresponding to Yyy = Ser and Glu. Although the detailed positions of individual peaks in each library differ from those shown in Figure 1, the overall trends in the data are similar. As shown below, it appears that many of the changes associated with differences in the positions of peaks arise due to the different amino acid sizes. Overall, the distributions for these libraries show that all of the NH_2-terminated sequences form compact globules; some also form elongated structures. As discussed above, the Ac-terminated forms of these sequences favor extended structures.

Figure 5 shows the drift(flight) time distribution for the library of peptides having Yyy = Lys. This distribution displays several distinct differences (compared with the distributions for the Yyy = Leu, Ser, and Glu). In particular, the NH_2-terminated peptides

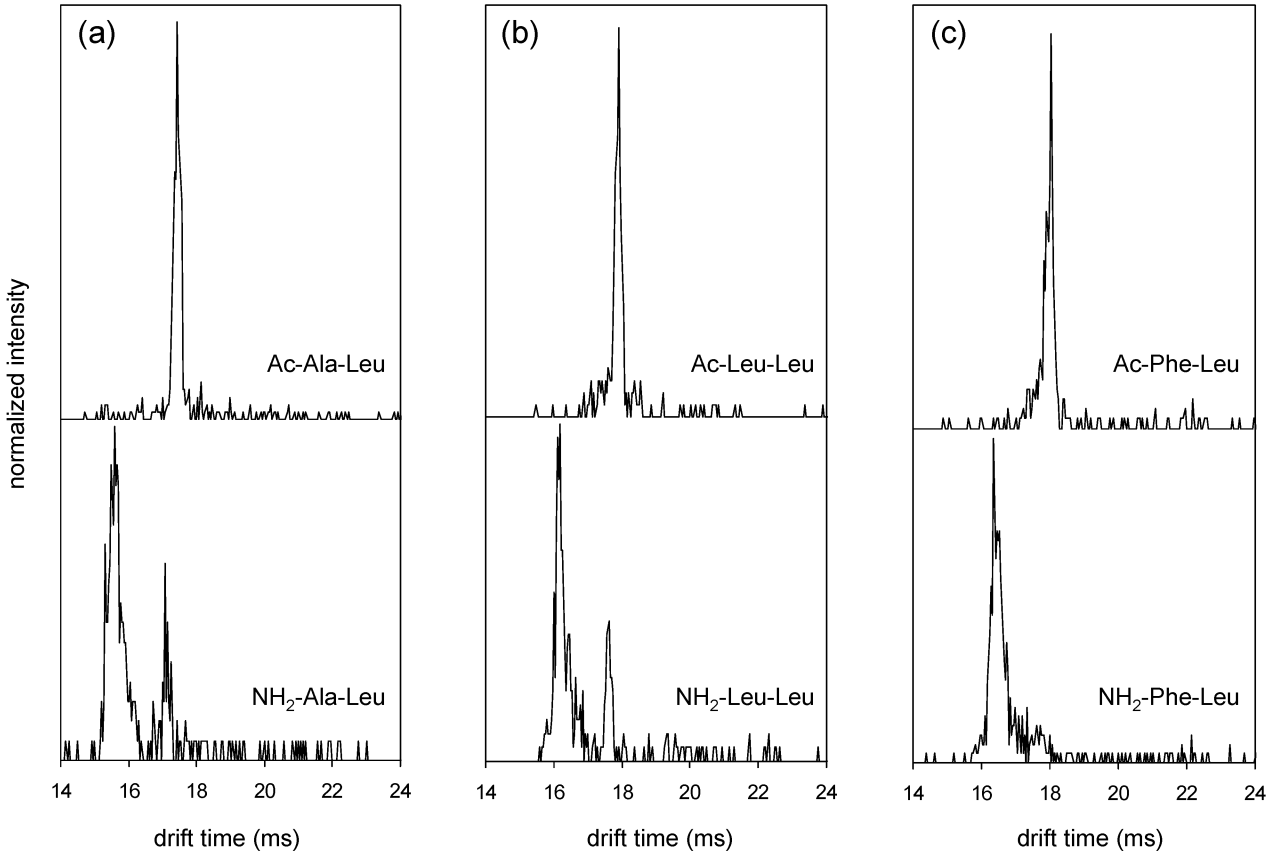

Figure 2. Ion mobility distributions obtained by integration of the ion mobility/time-of-flight data over regions of the dataset corresponding to the $[M + 2H]^{2+}$ ions of the Ala(i) → Leu($i + 4$) [part a], Leu(i) → Leu($i + 4$) [part b], and Phe(i) → Leu($i + 4$) [part c] sequences. Mobility data for the NH_2- and Ac-terminated forms of each sequence are shown in the bottom and top distributions, respectively. Drift times have been normalized to a pressure of 150.0 Torr.

show only a single conformer family. For most of these sequences, experimental drift times indicate that these ions have compact structures. The only exceptions are for Xxx = Lys, His, or Arg; these substitutions lead to peptides that have drift times that fall between the values expected for the compact and elongated forms (see Figure 5). The acetylated peptides also show well-defined peaks, consistent with elongated conformations (as observed in other libraries). However, the Xxx = Lys, His, and Arg peptides have drift times that are similar to their NH_2-terminated forms suggesting that the conformations of these sequences are not significantly influenced by the character of the N-terminal group.

Discussion

Overview of Approach Used to Group Ions into Structural Families. An understanding of how the ion mobility results presented above are related to the conformations of the peptide sequences is complicated by several factors; therefore, it is useful to provide a detailed discussion of the approach that we used to assign a given sequence to a structural type. We start by noting that the study of these sequences as $[M + 2H]^{2+}$ ions introduces a complexity associated with where protons are located along the peptide chain. A system of 15 equivalent protonation sites would have 105 {15!/[2! (15−2)!]} unique configurations for proton placement. It is not practical to perform molecular modeling simulations for all of these configurations -even for a single sequence. Therefore, we rely on previously reported thermochemistry to restrict the possible protonation sites to a reasonable number and explore these. Once reasonable configurations for different conformations have been identified, it is possible to examine some sequences in detail. In this case, we calculate cross sections for low-energy structures of specific sequences and compare the results with those of the experiment. In the end, these comparisons suggest that sequences form only a few major conformer types, compact globules and extended structures that contain at least several helical turns (in many cases it appears that these structures are almost entirely helical). Having established that most structures fall into these two categories, it is useful to present the data in a context that is relevant to the previous work involving alanine polymers. We do this by accounting for differences in the sizes of each amino acid in the peptide sequence and plotting the data on the relative cross section scale that was developed previously.[22] Below, we provide a detailed account of both the molecular modeling results and relative cross section analyses. A final caveat is that we do not intend to imply that we understand the structure at the level of atomic detail but rather that we have identified which sequences are likely to fall into a given structural family.

Influence of the Protonation Sites on the Structures and Stabilities of $[M + 2H]^{2+}$ Ions. The most dramatic effect that is observed in these studies is associated with the structural differences that are observed when the basic amino terminus is blocked by acetylation. This indicates that the locations of protons have a pronounced influence upon the conformation of ions. A complication that arises is that the location of charged sites is unknown. To gain a feeling about where charges reside, we have conducted a series of molecular modeling simulations in which protonated sites are systematically varied along the backbone of the peptide. Cassady has previously reported ab initio calculations at the MP4 level for the $[Gly_3 + H]^+$ system

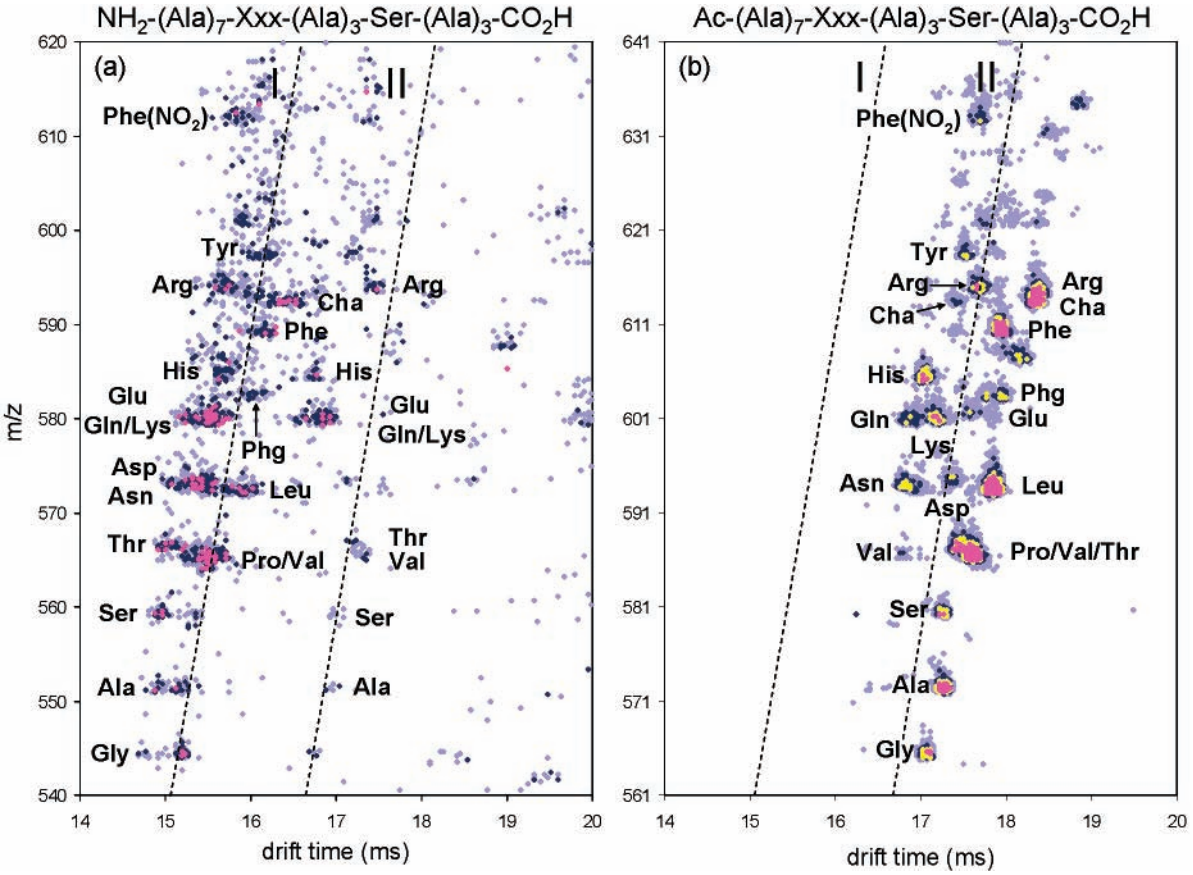

Figure 3. Ion mobility/time-of-flight data for the $NH_2-(Ala)_7-Xxx-(Ala)_3-Ser-(Ala)_3-CO_2H$ peptide library (part a) and the $Ac-(Ala)_7-Xxx-(Ala)_3-Ser-(Ala)_3-CO_2H$ (part b). These data were recorded using helium buffer gas pressures of 154.2 (a) and 155.0 Torr (b) and a drift field of 170.97 V cm^{-1}. Drift times have been normalized to 150.0 Torr.

and examined the relative proton affinities of different sites along the Gly_3 polypeptide.[40] Relative to the most basic N-terminal amino group, the next most basic position is the carbonyl group closest to the N-terminus, followed by the backbone carbonyl group closest to the C-terminus. These sites are less basic than the amino terminus by ~5.0 and 12.4 kcal mol^{-1}, respectively. Relative proton affinities should be analogous for the peptides studied here, especially if helices are formed, because addition of a proton to the C-terminal residue will stabilize the helix macrodipole.

Figure 6 shows a plot of calculated energies for a series of molecular modeling studies for the Ac−Ala(i) → Leu(i + 4) sequence with different protonation configurations. Simulations were initiated from α-helical structures with one proton fixed on the backbone carbonyl group of residue 14 and the other placed at one of the remaining carbonyl groups. The simulation was carried out at 300 K for 250 ps. A series of longer simulations performed for a few sequences suggests that results do not vary substantially from those that are presented in Figure 6. A plot of the average energies for the lowest 20 energy conformations obtained in the simulations shows that, when the first proton is positioned at residue 14 and the second is located at positions 10−13, highly helical conformations (having relatively low energies) are generated. As the position of the second charge is moved toward the N-terminal side of the peptide (from residue 9−1), helical turns are disrupted and relatively open random structures are formed. The calculated energies for these structures are substantially higher than those for helices.

An additional parameter that helps to constrain the possible proton locations is the experimental collision cross section value of 287.7 $Å^2$ for the Ac−Ala(i) → Leu(i + 4) sequence. Calculated cross sections for the low-energy structures used in Figure 6 are 289.8 ± 3.8, 288.4 ± 4.5, 306.6 ± 4.7, and 301.7 ± 1.9 $Å^2$, when the second proton is positioned at residue 13, 12, 11, and 10, respectively. Although several of these values are near the 287.7 $Å^2$ experimental value, the closest agreement comes from conformations generated upon placing the second proton at residue 12. The structures generated for this assignment favor a π-helical turn on the C-terminal end. This π character shortens the peptide helix slightly relative to an ideal α helix.

Examination of other proton configurations also is guided by Cassady's result that the second most basic site in the NH_2-Gly_3-COOH system is the backbone carbonyl group of the N-terminal residue. The results of simulations in which the first proton is fixed to this backbone group of the Ac−Ala(i) → Leu-(i + 4) sequence and the second is positioned at one of the remaining backbone carbonyl groups are shown in Figure 7. The low-energy structures that are found are globules, even though the initial structure was a helix. However, the average energies of conformations in which the second charge is located at the carbonyl groups of residues 13, 12, 11, and 10 are 33.3, 14.2, 19.8, and 9.9 kcal mol^{-1} above the values for the helices, respectively. In many cases, the globules from the simulations have cross sections that are near the experimental values for the high-mobility NH_2–terminated peptide ions.

Detailed Molecular Modeling Studies of a Few Sequences and Assignment of Structural Types. Because of the large number of different sequences that are examined here, it is not practical to carry out detailed molecular modeling simulations and cross section calculations for every peptide. Below, we use our previous estimates of the intrinsic sizes of different amino

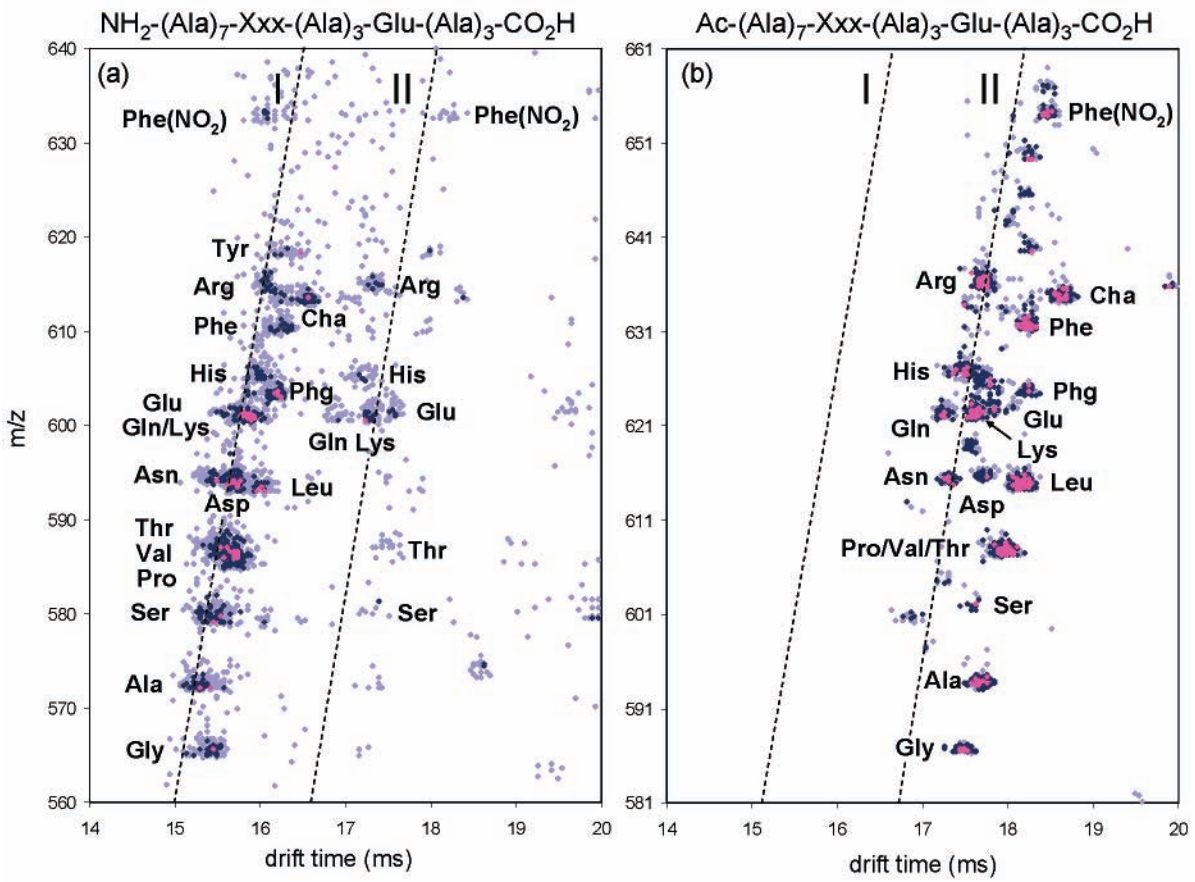

Figure 4. Ion mobility/time-of-flight data for the $NH_2-(Ala)_7-Xxx-(Ala)_3-Glu-(Ala)_3-CO_2H$ peptide library (part a) and the $Ac-(Ala)_7-Xxx-(Ala)_3-Glu-(Ala)_3-CO_2H$ (part b). These data were recorded using helium buffer gas pressures of 152.8 (a) and 152.9 Torr (b) and a drift field of 170.97 V cm^{-1}. Drift times have been normalized to 150.0 Torr.

acids[25] to remove contributions to the cross section that arise from variations in the sizes of the side chains and plot the data on Jarrold's relative cross section scale.[22] This makes it possible to make some general assignments of structural type. For these assignments, it is instructive to examine a few sequences more rigorously. In this section, we focus the discussion on the NH_2- and Ac-terminated Ala(i) → Leu(i + 4) system from above. However, we note that analogous studies of several other sequences yield similar results.

The simulations presented above regarding the energetics associated with different charge site assignments have been used to define starting points for molecular modeling simulations for different sequences. We have fixed protons at the amino terminus and closest backbone carbonyl as well as the carbonyl groups of residues 12 and 14 for the amino- and acetyl-terminated forms. The simulations indicate that extended states that display at least several helical turns are favored when charges are assigned to residues 12 and 14. Compact globular configurations are favored when charges are assigned to the amino terminus and the carbonyl group of residue 1.

Figure 8 shows the results for the lowest-energy group of conformations that were found in a detailed series of simulations for the NH_2- and Ac-terminated Ala(i) → Leu(i + 4) peptides. Protonation of the backbone carbonyl groups of residues 12 and 14 yields nearly identical π/α-helical conformations for the NH_2- and Ac-terminated peptide forms. The average calculated cross section for the 20 lowest-energy NH_2-terminated helical structures is 285.3 ± 6.6 Å^2, near the value determined from the lower intensity experimental peak in Figure 2 (281.8 Å^2). The higher intensity peak in Figure 2 for the NH_2-Ala(i) → Leu(i + 4) peptide has a drift time of 15.7 ms, which gives Ω(exp) = 256.7 Å^2. This value agrees with the calculated cross section (256.7 ± 4.2 Å^2) for the low-energy structures obtained in the simulations when charges are placed on the N-terminal amino group and the carbonyl group of residue 1. Other charge site assignments also lead to globules that are in agreement with the experiment.

Summary of Results Plotted on a Relative Cross Section Scale for Polyalanine. From the comparisons of experiments and theory for ten Xxx(i) → Yyy(i + 4) sequences as amino- and acetyl-terminated forms, we conclude that most of the peaks that we have observed in these studies fall into two basic types of structures: a structure that is compact and globular and a structure that is significantly more extended and displays several helical turns.

It is useful to plot the experimental cross sections on a scale that removes differences that arise from variations in the sizes of the different side chains. Here, we plot the results on a relative cross section scale, defined as Ω(exp) − [14.50n − p(Xxx) − p(Yyy)], where 14.50 is Jarrold's calculated average cross section per residue for an ideal polyalanine helix and n is the number of alanine residues in the sequence (n = 13 for sequences in which neither Xxx nor Yyy is Ala). The terms p(Xxx) and p(Yyy) correspond to the intrinsic contributions to the cross section from a single amino acid residue (determined from measurements of tryptic peptides, as reported previously).[25] Table 2 lists the values for p(Xxx) and p(Yyy) along with the intrinsic size parameters that are used to calculate them.

Figure 9 shows plots of relative cross sections as a function of the Xxx residue for each of the NH_2- and Ac-terminated libraries. Nearly all of these relative values fall into two general structural types. For example, relative cross sections ranging

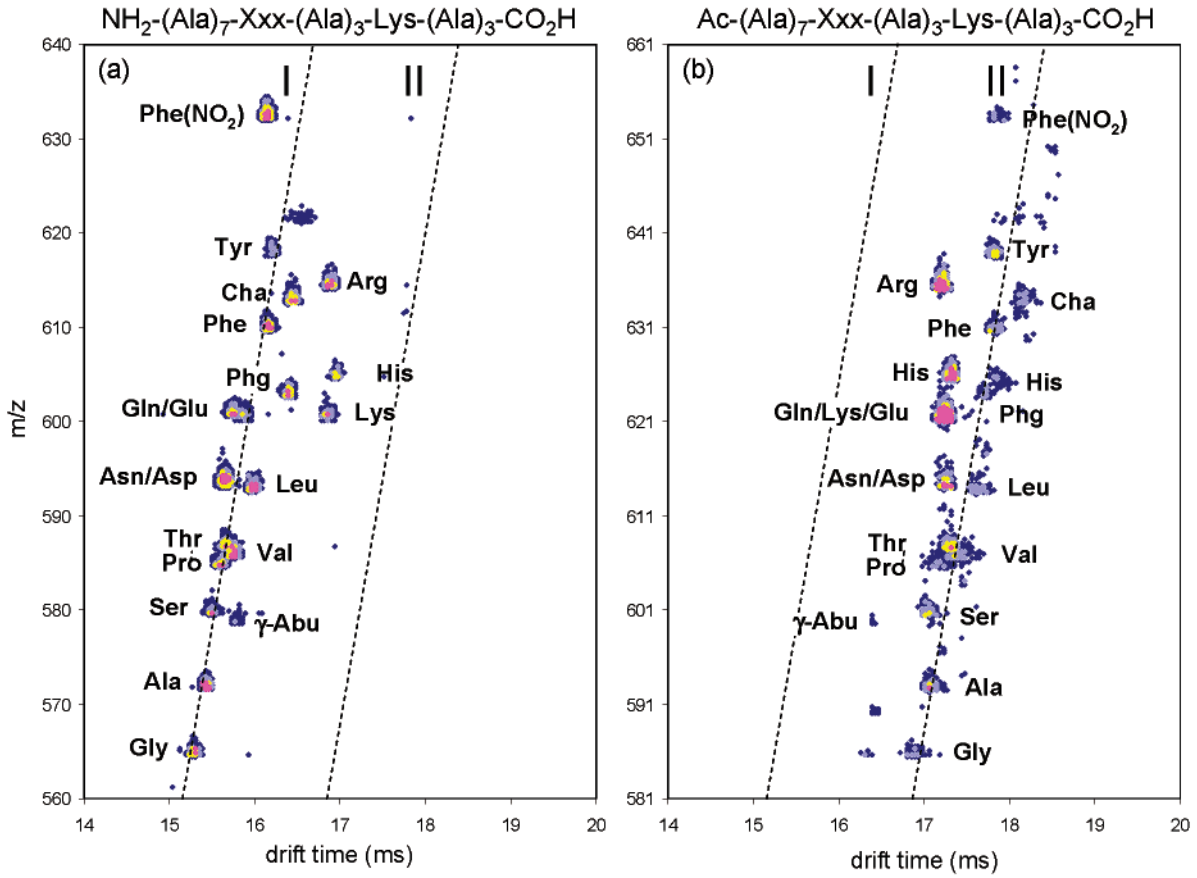

Figure 5. Ion mobility/time-of-flight data for the $NH_2-(Ala)_7-Xxx-(Ala)_3-Lys-(Ala)_3-CO_2H$ peptide library (part a) and the $Ac-(Ala)_7-Xxx-(Ala)_3-Lys-(Ala)_3-CO_2H$ (part b). These data were recorded using helium buffer gas pressures of 161.0 (a) and 161.4 Torr (b) and a drift field of 170.97 V cm^{-1}. Drift times have been normalized to 150.0 Torr.

from ~20 to 35 are observed for the majority of NH_2-terminated sequences. These values are similar to the relative cross section of ~19 reported for a singly charged $[Ala_{15} + H]^+$.[22] In other work, we have recorded the cross section for $[Ala_{15} + 2H]^{2+}$ and include the relative cross section for this ion (~35) for comparison.[23] The observation that a specific sequence has a relative cross section that falls within the ~19 to ~35 range for globules of $[Ala_{15} + H]^+$ and $[Ala_{15} + 2H]^{2+}$ suggests that it has a globular structure.

Of the 80 different NH_2-terminated sequences studied here, only one His(i) → Lys($i + 4$) does not show a peak that falls within the expected range for globules. However, the Lys(i) → Lys($i + 4$) and Arg(i) → Lys($i + 4$) sequences are also near the upper limit of our globule range. We note that, in each of these cases, it is likely that protons are associated with the highly basic side chains of these residues. Molelcular modeling results for these ions suggest that none of these sequences should be assigned as globules or helices. Instead, it appears that Coulombic interactions between the two nearby charged side chains forces the peptides to adopt extended (random-like) structures between residues 8 and 12. Overall, these three sequences appear to behave differently than the other sequences studied here.

As shown in Figure 2 many of the NH_2-terminated sequences show two peaks. The conformations that arrive at longer times have relative cross sections that are in close agreement with the values that are obtained for the Ac-terminated sequences. Nearly all of the Ac-terminated peptides have relative cross sections that fall within a range of ~48 to 67. The relative cross section for Jarrold's $[Ala_nLys + H]^+$ helix calculated using our size parameter for Lys is 59. However, as noted above in the molecular modeling studies, it appears that the best assignment of our peptides is a mixed π/α helix. The somewhat shorter structures are expected to have smaller relative cross sections.

To examine this idea in more detail, we created structures that are based on the structure in Figure 8 in which a last turn on the C-terminal side involves an π turn. We estimate the relative cross section for the largest Xxx and Yyy residues (Arg and Glu, respectively) to be 59; the relative cross section for the smallest combination Gly and Ser is 49. The majority of the Ac-terminated sequences fall within this helical region. Many of the lower mobility NH_2-terminated conformers observed for specific $i \rightarrow i + 4$ amino acid pairs also fall within this range.

The only sequences that consistently fall outside or on the edges of the range expected for compact globules and extended helices are the Lys(i) → Lys($i + 4$), His(i) → Lys($i + 4$), and Arg(i) → Lys($i + 4$) sequences. As noted above, these structures appear to be extended in the region between residues 8 and 12. Other than this, it is difficult to ascertain any other structural characteristics that are reproducible in the molecular modeling simulations. Thus, we refer to these simply as *other* conformations.

Relative Abundances of Globules, Helices and Other Conformations for the $NH_2-(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3$ Sequences. Figure 10 provides information about the relative abundances of the globular, helical, and *other* conformations for every $NH_2-(Ala)_7-Xxx-(Ala)_3-Yyy-(Ala)_3$ peptide sequence examined here. Several trends are apparent in these data. We start by noting that the $NH_2-Xxx(i) \rightarrow Lys(i + 4)$ system appears to be unique. In these peptides, no Xxx residue substitutions appear to favor formation of NH_2-terminated helical conformations, only globular structures are observed. It is likely that protonation of the basic butylamine

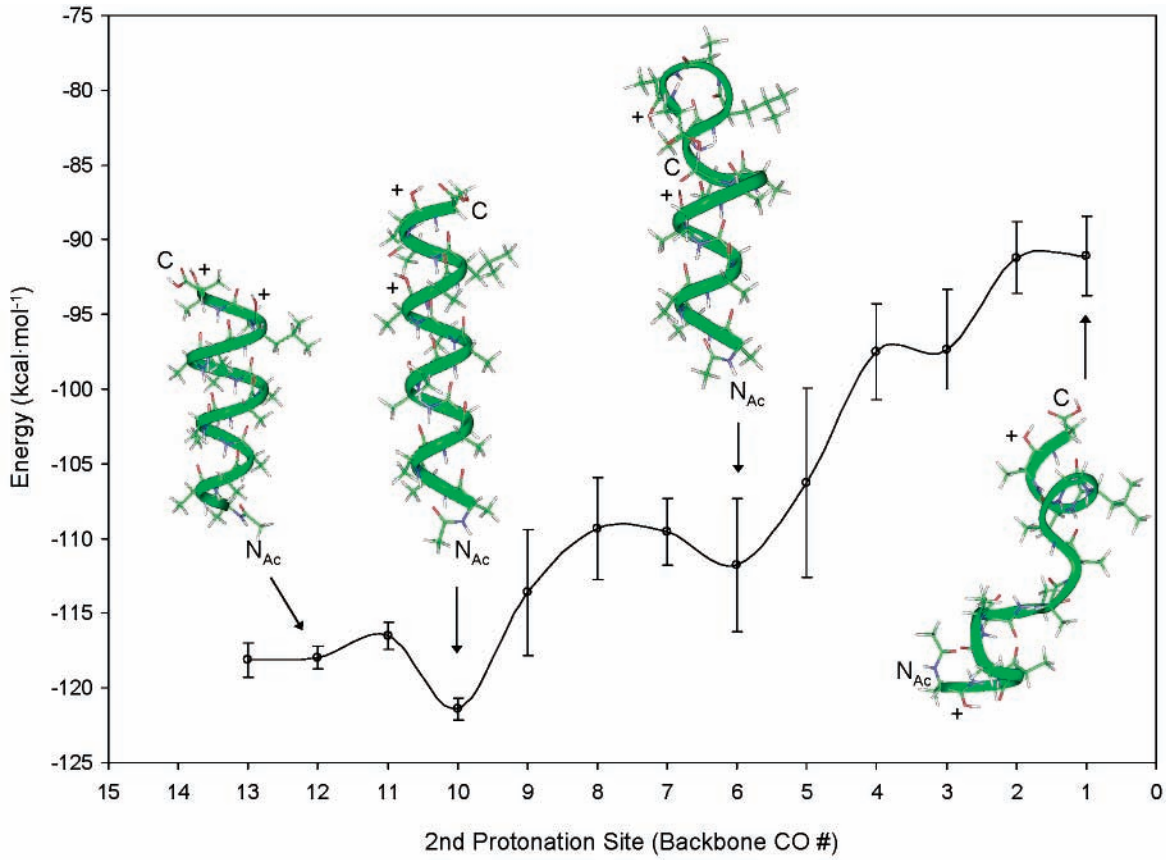

Figure 6. Average energies for the twenty lowest-energy structures obtained in a series of molecular modeling simulations for $[M + 2H]^{2+}$ ions of the Ac-terminated Ala(i) → Leu($i + 4$) sequence. The simulations were performed for 250 ps at 300 K using an α-helical starting structure. For each simulation, one proton was fixed on the backbone carbonyl group of the 14th residue. Energies are plotted with respect to the location of a second proton on each of the remaining backbone carbonyl groups. Error bars represent one standard deviation of the calculated average energies. Sample structures in the figure correspond to the lowest-energy conformers obtained in the simulations for the ($i = 14$, 13), ($i = 14$, 10), ($i = 14$, 6), and ($i = 14$, 1) proton configurations. The C-terminus and acetylated N-terminus for each structure are indicated using "C" and "N_{Ac}". In addition, protonated backbone carbonyl groups are indicated for each structure using a "+" symbol. Calculated collision cross sections for these representative structures are 287.9, 303.9, 312.1, and 306.0 Å^2, respectively. The calculated cross section for the ($i = 14$, 12) structure is in close agreement with the experimentally obtained cross section for the Ac-terminated Ala(i) → Leu($i + 4$) peptide (287.7 Å^2).

side chain prohibits the stabilization of helical structures because of unfavorable interactions with the helix dipole. Molecular modeling simulations for several representative NH_2−Xxx(i) → Lys($i + 4$) peptides show substantial disruption of the helix due to solvation of the protonated Lys side chain by electronegative backbone carbonyl groups (data not shown). As noted above, when Xxx is a basic residue, the conformer type appears to be relatively open between the 8 and 12 residues with no other discernible structure. Experimental cross sections for these conformations are consistent with conformations from modeling simulations in which the Lys and basic Xxx residues are protonated. We have classified this structural type as *other*.

NH_2-terminated helices in the Yyy = Leu, Ser, and Glu libraries appear to fall into families defined by the residues at both the Xxx and Yyy positions. For some of these sequences, helix stabilization is easily rationalized, and for others, the observation of helical structures is more difficult to understand. One trend observed in these data involves the formation of NH_2-terminated helices by peptides with polar amino acids at both the Xxx and Yyy position. The abundances of these helical conformations (relative to globular structures) are shown in the left panel of Figure 10. A partial listing of these helical sequences includes: NH_2−Ser(i) → Ser($i + 4$), NH_2−Gln(i) → Ser($i + 4$), NH_2−Lys(i) → Ser($i + 4$), NH_2−Ser(i) → Glu($i + 4$), NH_2−Gln(i) → Glu($i + 4$), NH_2−Glu(i) → Glu($i + 4$), and NH_2−His(i) → Glu($i + 4$). Polar amino acid residues placed at the Xxx(i) and Yyy($i + 4$) positions introduce the potential for favorable contacts between the side chains of residues in adjacent helical turns.

We have used molecular modeling simulations for specific combinations of Xxx(i) and Yyy($i + 4$) polar amino acids to investigate the influence of side chain hydrogen bonding interactions on helix stability. Figure 11 shows the results from a simulation for the NH_2−His(i) → Glu($i + 4$) sequence. Two different π/α-helical structures with relatively similar calculated cross sections are shown in the figure. The first structure includes a hydrogen bond between the His and Glu side chains, and the second shows the side chains rotated away from one another. The calculated cross section for the first structure is 285.6 Å^2 in good agreement with the experimental cross section of 284.9 Å^2 obtained for this peptide. The second structure in which the side chains do not interact has a calculated cross section of 289.5Å^2. The helix with the His−Glu interaction corresponds to the lowest-energy structure obtained in the simulation. The calculated energy for this structure is −157.27 kcal mol^{-1}, nearly 5 kcal mol^{-1} less than the calculated energy of the structure in which the side chains do not interact (−152.66 kcal mol^{-1}).

General Trends among Side Chains that Appear to Favor Helices. The results and molecular modeling presented above indicate that Ac-terminated peptides favor elongated states containing substantial helical character, for nearly all of the

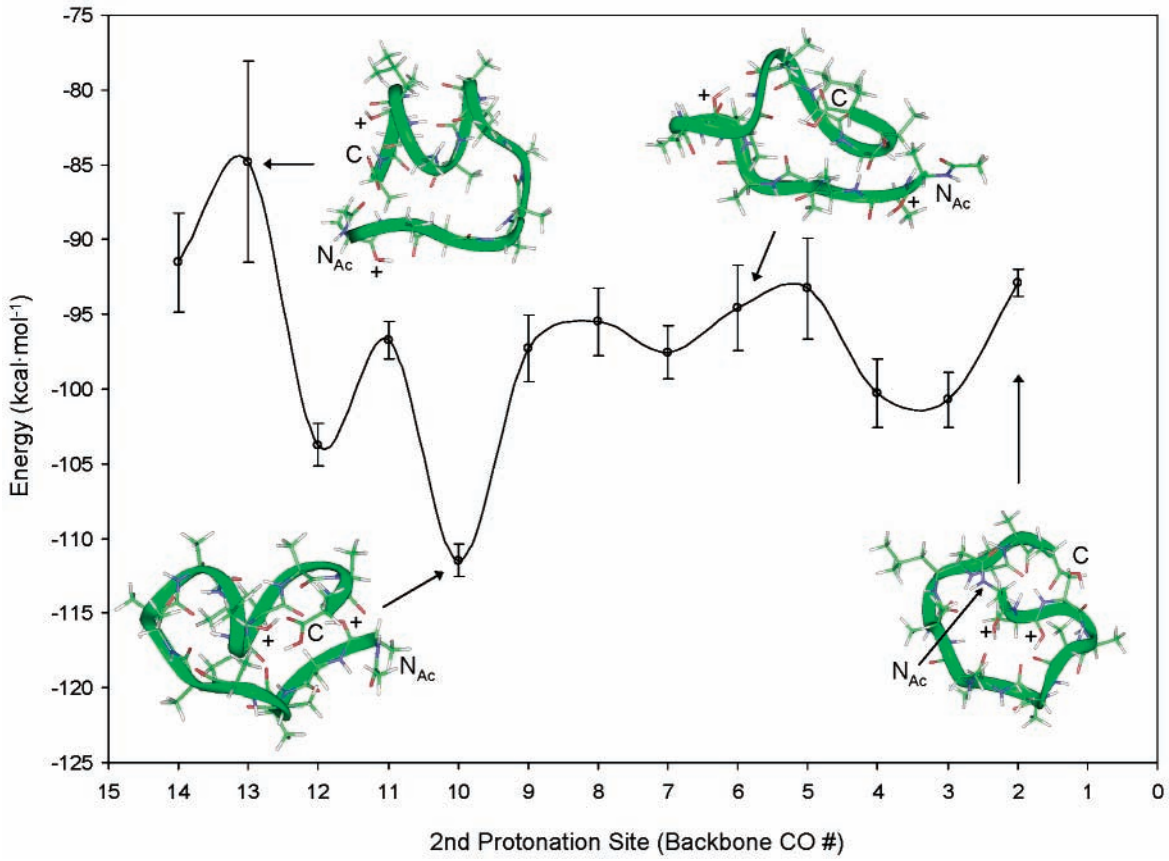

Figure 7. Average energies (and standard deviations) for the 20 lowest-energy structures obtained from molecular modeling simulations (identical to those used in Figure 6) for the Ac-terminated Ala(i) → Leu(i + 4) sequence. In these simulations, one proton was fixed on the backbone carbonyl group of the N-terminal amino acid residue. Energies are given with respect to the location of the second proton on each of the remaining backbone carbonyl groups in the peptide. Representative structures in the figure are the lowest-energy structures from the simulations for the (i = 1, 12), (i = 1, 8), (i = 1, 4), and (i = 1, 2) proton configurations. The C-terminus and acetylated N-terminus for each structure are indicated as "C" and "N_{Ac}" and the locations of protonated backbone carbonyl groups are given using a "+" symbol. Calculated collision cross sections for these representative structures are 303.9 ± 23.6, 271.0 ± 2.5, 293.5 ± 6.3, and 264.2 ± 4.3 Å^2, respectively.

sequences that we have studied. The NH_2-terminated peptides display much more interesting behavior. We start by noting that $[Ala_{15} + 2H]^{2+}$ ions favor compact globules. It appears that only specific types of substitutions are sufficient to favor extended helical states.

Figure 10 shows a plot of the relative abundances of different conformer states for the range of Xxx and Yyy substitutions that have been explored for the NH_2-terminated peptides in this study. This plot also assumes that the different states fall into three basic structural types: helix, globule, and other (as determined from the analysis given above). It is interesting to note that the Yyy = Lys system behaves differently than the Leu, Ser, and Glu systems. There is no evidence for formation of extended helical states for any of the Xxx(i) → Lys(i + 4) interactions. This is presumably because charges are localized at the amino terminus and the basic Lys residue, such that the net charge is distributed on the N-terminal side of the peptide chain in all sequences. This would be consistent with our previous work on polyalanine peptides; in this system, it appears that helices are observed only when the charge-distribution shifts to the C-terminal side of the peptide.[23]

Conversely, a number of specific Xxx(i) → Yyy(i + 4) interactions appear to favor helices in the Yyy = Leu, Ser, and Glu sequences. The interpretation of these results must be handled cautiously because the ion structure may reflect states that are trapped during the ESI process, rather than a distribution that is defined by the relative stabilities of different conformers (influenced by the side chain). However, this said, these interactions can be categorized according to the side chain functional groups of the amino acids at the Xxx and Yyy positions. For example, Xxx(i) → Leu(i + 4) interactions favor helices when Xxx corresponds to a nonpolar residue, such as Ala, Val, and Leu or the basic Lys, His, and Arg residues. The Xxx = Tyr sequence also forms a small amount of helix. It appears that interactions of the Yyy = Leu residues with other nonpolar residues as well as basic groups favor helices in the gas phase. The Yyy = Ser system favors helices when Xxx = Ala and Val but not when Xxx corresponds to Leu. Additionally, this system favors helices for several aromatic residues (Tyr and Phe(NO_2)), several polar residues (Ser, Thr, Gln, and Glu), as well as the basic Lys, His, and Arg groups. The Yyy = Glu sequences do not appear to favor helices when Xxx is a nonpolar group. Instead, the aromatic [Phe, Tyr, and Phe(NO_2)], polar (Ser, Thr, Gln, and Glu), and basic (Lys, His, and Arg) groups favor helices.

Simulations, Consistent with Favorable and Unfavorable Side Chain Interactions. Some insight regarding the influence of a given side chain combination on helix formation can be obtained by examining the structures that are observed at intermediate times during the molecular modeling simulations. If an α-helical structure is used as the initial structure, some Xxx(i) → Yyy(i + 4) combinations appear to interact favorably, effectively locking the helical structure of the backbone into place, while with other interactions the initial helix is quickly disrupted (resulting in the formation of globular conformers). An example of a favorable Xxx(i) → Yyy(i + 4) combination

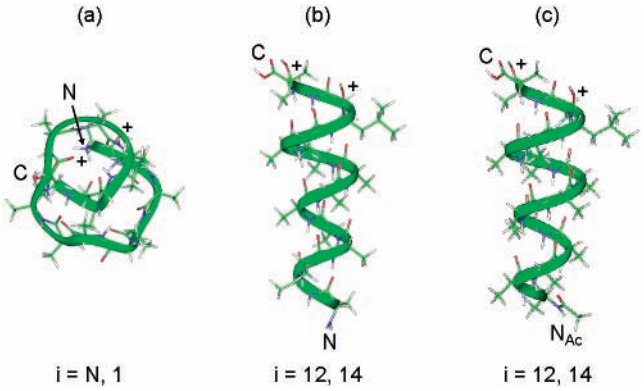

Figure 8. Lowest-energy structures obtained from molecular modeling simulations for the $[M + 2H]^{2+}$ NH_2-terminated and Ac-terminated Ala(i) → Leu($i + 4$) peptide ions. The simulations were performed for 250 ps at a temperature of 300 K. Part a shows the lowest-energy structure obtained in a simulation for the NH_2−Ala(i) → Leu($i + 4$) peptide with protons placed at the N-terminus and backbone carbonyl group of the first amino acid residue. The calculated collision cross section for this structure (251.1 Å^2) is in close agreement with the experimental cross section for the compact NH_2-terminated Ala(i) → Leu($i + 4$) peptide (256.7 Å^2). Part b shows the lowest-energy structure for the NH_2−Ala(i) → Leu($i + 4$) simulation in which protons were placed on the backbone carbonyl groups of amino acid residues 12 and 14. The calculated cross section for this conformer is 285.9 Å^2, a value close to the experimental value of 281.8 Å^2 measured for the elongated form of this sequence. Part c shows the lowest-energy structure for the Ac-terminated Ala(i) → Leu($i + 4$) peptide with protons placed on the backbone carbonyl groups of residues 12 and 14. The calculated collision cross section for this structure (287.9 Å^2) is similar to the experimentally obtained cross section for the Ac-terminated peptide (287.7 Å^2).

is shown for the His(i) → Glu($i + 4$) combination in Figure 11. The additional hydrogen bond that is formed for these side chains stabilizes the helix by ∼5 kcal·mol^{-1}. Upon disrupting the helix, we can generate higher-energy structures with globular conformations; in these structures, we find no clear interactions between side chains that would stabilize the globule. Thus, it appears that one role that side chain interactions play in favoring helices in the gas phase is to lower the energy of the helix relative to the globule.

A number of other Xxx(i) → Yyy($i + 4$) combinations show similar behavior; for example, limited modeling results (initiated from helical starting structures) show that combinations Xxx = Glu with Yyy = Tyr, Phe(NO_2), Ser, Thr, Gln, and Glu and also neutral forms of the Lys and Arg residues can be stabilized by hydrogen bonding interactions between the side chains. Additionally, interaction of the OH group of the carboxylic acid side chain of Glu with the center of the aromatic ring in Phe appears to stabilize helices; although the calculated energetics of this interaction appears to be only ∼2 kcal mol^{-1}. Similar interactions are also observed in combinations associated with the Yyy = Ser library for Xxx = Tyr, Phe(NO_2), Ser, Thr, Gln, Glu, and neutral forms of the Lys, His, and Arg residues.

Alternatively, examination of intermediate structures in simulations for some Xxx(i) → Yyy($i + 4$) combinations in the Yyy = Glu and Ser libraries suggests that side chains aid in disrupting the helix. Although we anticipated that the interactions of the polar Glu and Ser side chains with nonpolar side chains would be less stable than with polar groups, we were somewhat surprised that the presence of a single polar group in the sequence actually appears to initiate helix unfolding. In this case, in the absence of strong stabilizing interactions between the side chains, the polar residue associates with CO or NH groups along the backbone, and in doing so, the initial $i \rightarrow i + 4$ hydrogen bonding network of the helix is disrupted. This type of disruption allows the helix to unravel near the middle of the sequence, rather than from the ends, as has been reported previously. We note that the Yyy = Ser system displays less of this type of behavior in the modeling than the Glu system. Although our current molecular modeling studies have not provided a clear explanation of this, it appears that the small size of the Ser residue may prohibit it from forming hydrogen bonds with the backbone. This may be because the initial helical structure must distort (breaking bonds between backbone CO and NH groups) before the hydroxyl side chain of Ser can interact closely with polar backbone groups.

Finally, we note that the Yyy = Leu system favors helices for Xxx = Ala, Val, Leu, and Tyr. Molecular modeling simulations for these sequences do not show significant interactions between the nonpolar side chains at the Xxx and Yyy positions. Relative energies for conformers in which these side chains are in close proximity to one another are similar to energies for structures in which the side chains are rotated away from one another. These studies suggest that, although nonpolar side chain interactions do not appear to stabilize helical structures, the side chains do not destabilize the helix by interacting with backbone CO and NH groups as is observed for isolated polar side chains.

TABLE 2: Relative Collision Cross Sections Derived from Intrinisic Size Parameters

residue	intrinsic size parameter[a]	residue Ω from polyalanine fit (Å^2)[b]
Gly	0.99	13.65
Ala	1.08	17.35
γ-Abu	0.84	15.32
Ser	0.99	18.25
Pro	1.00	19.82
Val	1.08	21.69
Thr	1.00	20.35
Leu	1.19	26.02
Asn	0.94	20.66
Asp	0.89	19.66
Gln	0.98	23.10
Lys	1.23	29.01
Glu	0.91	21.55
Phg	0.95	22.90
His	0.93	22.78
Phe	1.05	26.77
Cha	1.38	35.98
Arg	1.27	33.40
Tyr	0.99	26.60
Phe(NO_2)		

[a] Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1188. [b] These data are derived by scaling the intrinsic size parameters derived previously by a relation that normalizes for differences in residue mass. In the present system, the relation used was $\Omega = 1/15(-2.724 \times 10^{-5}\,(MW_{15})^2 + 0.2141\,(MW_{15}) + 40)$. See ref 25 for a more detailed discussion of this type of treatment.

Summary and Conclusions

A combination of ion mobility measurements and molecular mechanics simulations have been used to evaluate the structures of a series of $[M + 2H]^{2+}$ ions having the general forms NH_2−$(Ala)_7$−Xxx−$(Ala)_3$−Yyy−$(Ala)_3$ and Ac−$(Ala)_7$−Xxx−$(Ala)_3$−Yyy−$(Ala)_3$, where residue 8 (Xxx) and residue 12 (Yyy) have been systematically varied in order to probe the influence of different side chain interactions upon the formation of helical structures. The results that are presented are at a very early stage of interpretation and thus the current work should be viewed

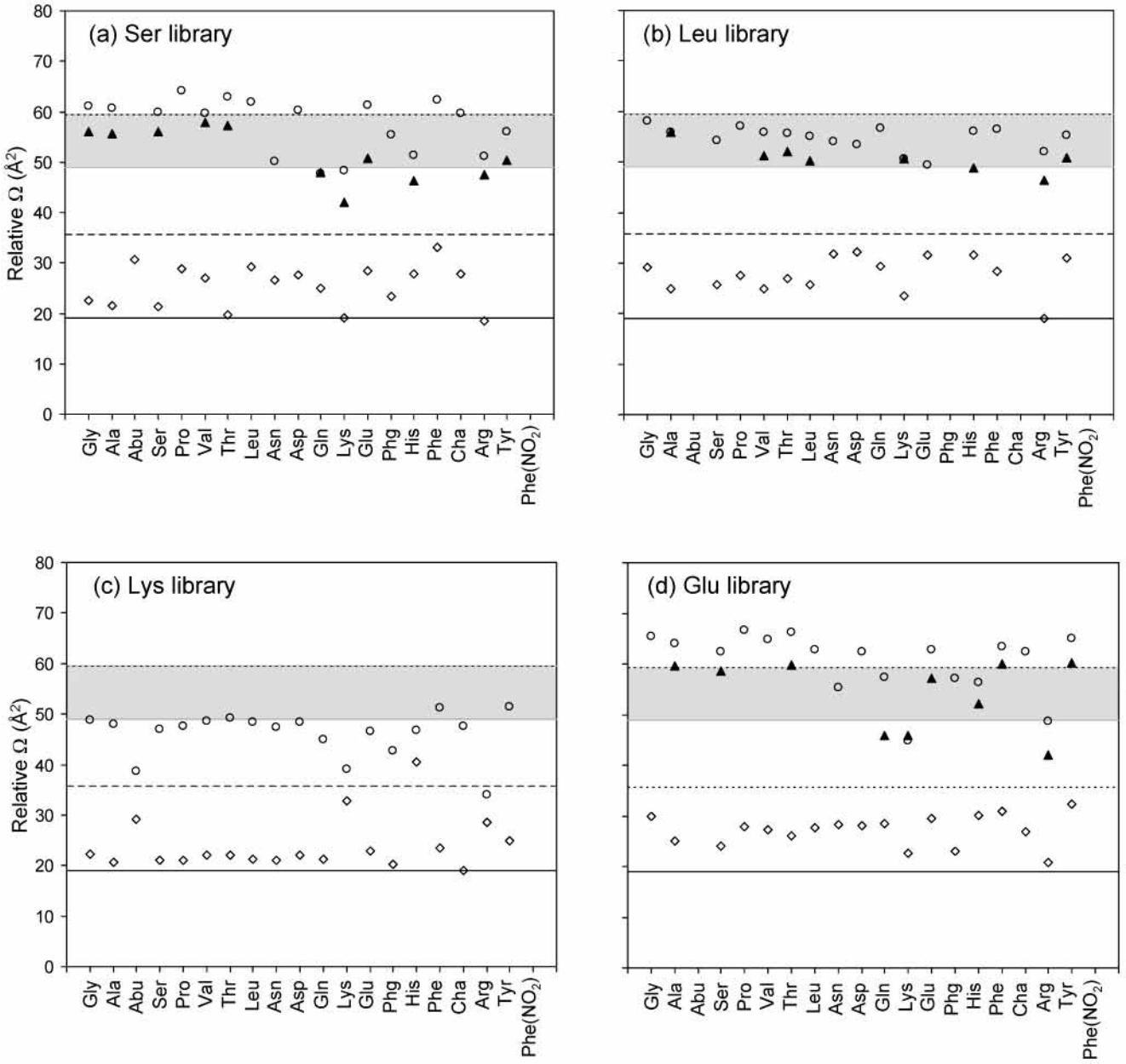

Figure 9. Plots of the relative cross section versus the Xxx(i) residue for the NH_2- and Ac-terminated Xxx(i) → Ser(i + 4) [a], Xxx(i) → Leu(i + 4) [b], Xxx(i) → Lys(i + 4) [c], and Xxx(i) → Glu(i + 4) [d] libraries. Relative cross sections were determined from experimental cross section values for each sequence according to the relation: $\Omega_{Exp} - [14.50n - p(\text{Xxx}) - p(\text{Yyy})]$, where 14.50 is Jarrold's calculated average cross section per residue for an ideal polyalanine helix and n is the number of alanine residues in the sequence ($n = 13$ for all sequences in which Xxx and Yyy is not Ala). The terms p(Xxx) and p(Yyy) correspond to the intrinsic contributions to cross section from the amino acid residues at the Xxx and Yyy positions. These values are given in Table 2 along with the intrinsic size parameters that are used to calculate them. Shaded regions of the plots correspond to relative cross section values for alanine-based π/α helices in which residues Xxx and Yyy are both small (Gly and Ser, respectively) and large (Arg and Glu, respectively). Horizontal lines represent relative cross sections for $[Ala_{15}+H]^+$ (solid line), $[Ala_{15}+2H]^{2+}$ (dashed line), and Jarrold's $[Ala_{15}Lys+H]^+$ (dotted line) determined from experimental cross sections using the relation described above.

as a survey of what types of structures are present in these sequences. Overall, NH_2-terminated peptides show evidence for helices and globules, and the abundance of these states is very sensitive to the Xxx and Yyy residues that are incorporated into the sequence. Ac-terminated peptides favor more extended helical structures. The differences between the NH_2- and Ac-terminated forms of these peptides indicate that the sites of protonation influence the gas-phase structures that are observed. Molecular modeling simulations have been carried out in order to provide information about which sites are likely to be protonated in helical and globular structures of NH_2- and Ac-terminated peptides.

A summary of the different types of structures that are observed for the range of sequences studied suggests some general trends in the data. Three types of different side chain interactions appear to govern the competition between formation of helices and globules. First, stabilizing polar−polar interactions between side chains can lock in a helical turn between the 8th and 12th residues. In this case, the stabilizing interactions between a single pair of adjacent side chains provides an energetic driving force for favoring the helix. Additionally, the formation of a single turn would provide a nucleus for propagating the helix throughout the rest of the polypeptide chain, leading to the formation of extended helical structures.

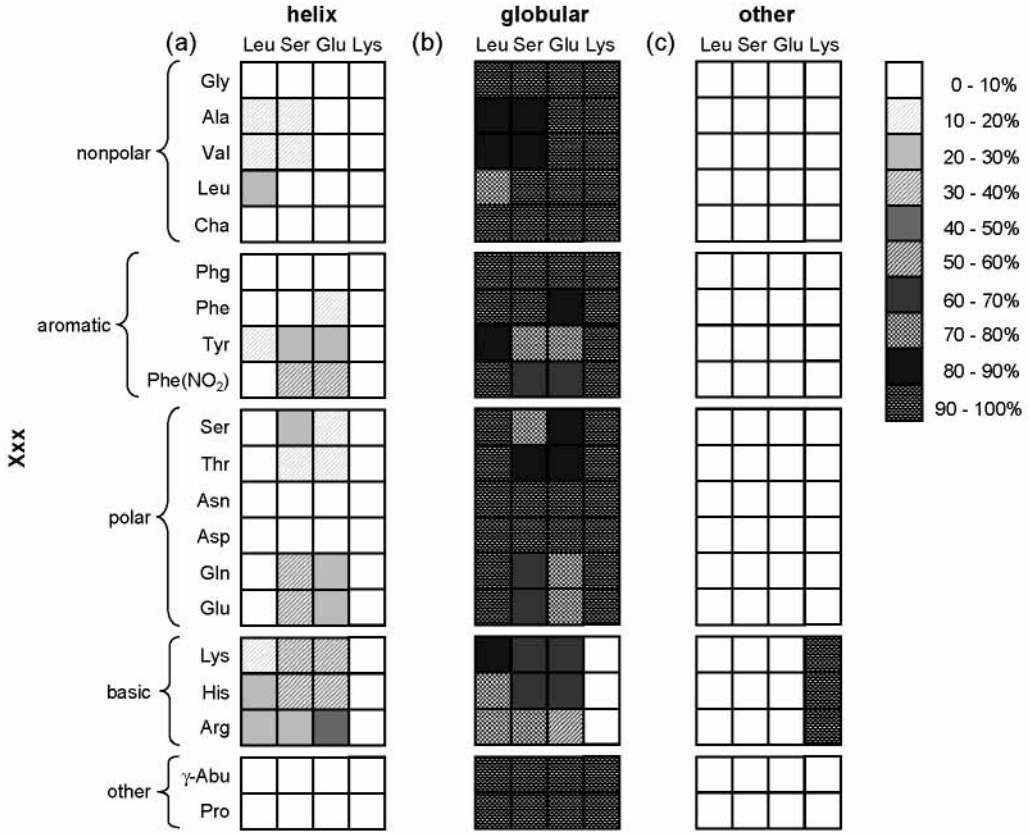

Figure 10. Plots of relative abundance of helical (a), globular (b), and other (c) conformations for all combinations of amino acid residues at the Xxx and Yyy positions in the NH_2-terminated library peptides. Relative abundances were determined from the ratios of peak areas (from experimental mobility distributions) for the different conformer types of each sequence. High mobility (type I) structures have been classified as globular and elongated (type II) as helices according to the results of the molecular modeling simulations as described in the text. The other structures represent structures with experimental cross sections intermediate between those of the compact globules and more elongated helical structures.

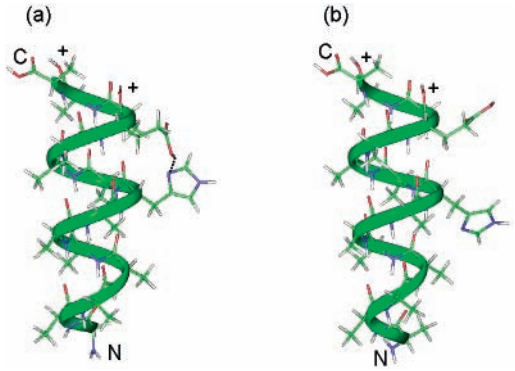

Figure 11. Structures obtained from molecular modeling simulations for the $[M + 2H]^{2+}$ NH_2−His(i) → Glu(i + 4) peptide ion. The simulations were performed for an α-helical starting structure with protons were placed on the 12th and 14th backbone carbonyl groups for 250 ps at a temperature of 300 K. Part a shows the lowest-energy structure obtained the simulation in which the His and Glu side-chains form a hydrogen bond. The calculated collision cross section for this structure (285.65 Å^2) is in close agreement with the experimental cross section for the elongated NH_2-terminated His(i) → Glu(i + 4) peptide (284.88 Å^2). The calculated energy for this structure is −157.27 kcal mol^{-1}. Part b shows a higher energy structure (E = −152.66 kcal mol^{-1}) in which the side-chains do not form a hydrogen bond that was obtained in the same simulation. The calculated collision cross section for this structure is 289.51 Å^2.

No similar interactions beween the side chains of the 8th and 12th residues are observed in globules. Second, the data presented above may suggest that helices are often disrupted when the pair of amino acids at the i and i + 4 positions includes one polar and one nonpolar residue. In such a case, a single polar−nonpolar interaction between side chains of the 8th and 12th residues is not of sufficient magnitude to lock in a helical turn. Instead molecular modeling results suggest that the polar group appears to initiate the unfolding of the helix. This may occur when the single polar side chain interacts with backbone CO and NH groups (disrupting hydrogen bonding along existing regions of helix). We anticipate that such interactions will be sensitive to the placement of polar and nonpolar residues. We are currently designing several libraries to test this idea. Finally, we note that helices are often favored upon incorporation of nonpolar residues into the Xxx and Yyy positions. In this case, it appears that higher helix propensities (observed for large non polar groups in solution) also lead to favorable helices in the gas phase. Although interactions between nonpolar side chains are weak, they do not appear to disrupt the hydrogen-bonding network associated with the helix.

In closing, throughout this work we have stressed that the present paper presents a survey of the types of structures that are favored upon substitution of the 8th and 12th residues of alanine based peptide ions. Although it is clear that many differences in structure arise upon varying sequences, these structures are established immediately after ions emerge from the electrosprayed droplet, and as such, the relative abundances of different conformations may be influenced by the final stages of the electrspray process. It should be possible to study transitions between structures in some detail by varying the temperature of drift tube buffer gas. We are also exploring the use of ion trapping techniques that allow conformers to be stored for varying time periods prior to mobility analysis in order to study transitions between structures.[41]

References and Notes

(1) Chou, P. Y.; Fasman, G. *Biochemistry* **1974**, *13*, 211.

(2) Rohl, C. A.; Fiori, W.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3682.

(3) For recent reviews, see: (a) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biomol. Struct.* **1992**, *21*, 95. (b) Baldwin, R. L. *Biophys. Chem.* **1995**, *55*, 127. (c) Chakrabartty, A.; Baldwin, R. L. *Adv. Protein Chem.* **1995**, *46*, 141. (d) Rohl, C. A.; Baldwin, R. L. *Methods Enzymol.* **1995**, *46*, 1. (e) Baldwin, R. L.; Rose, G. D. *Trends Biochem. Sci.* **1999**, *24*, 26.

(4) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

(5) Winger, B. E.; Light-Wahl, K. J.; Rockwood, A. L.; Smith, R. D. *J. Am. Chem. Soc.* **1992**, *114*, 5897; Suckau, D.; Shi, Y.; Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 790; Wood, T. D.; Chorush, R. A.; Wampler, F. M.; Little, D. P.; O'Connor, P. B.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2451.

(6) (a) von Helden, G.; Wyttenback, T.; Bowers, M. T. *Science*, **1995**, *267*, 1483. (b) Wyttenbach, T.; Bowers, M. T. *J. Am. Chem. Soc.* **1996**, *118*, 8355. (c) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995**, *117*, 10141. (d) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F.; *J. Am. Chem. Soc.* **1997**, *119*, 2240. (e) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 9586.

(7) Compagnon, I.; Hagemeister, F. C.; Antoine, R.; Rayane, D.; Broyer, M.; Dugourd, P.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **2001**, *123*, 8440.

(8) Zhang, X.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1211.

(9) (a) Schnier, P. F.; Gross, D. S.; Williams, E. R. *J. Am. Chem. Soc.* **1995**, *117*, 6747. (b) Williams, E. R. *J. Mass Spectrom.* **1996**, *31*, 831. (c) Rodriguez-Cruz, S. E.; Klassen, J. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 565.

(10) (a) Riemann, C. T.; Velazquez, I.; Tapia, O. *J. Phys. Chem. B* **1998**, *102*, 2277. (b) Riemann, C. T.; Velazquez, I.; Tapia, O. *J. Phys. Chem. B* **1998**, *102*, 9344. (c) Arteca, G. A.; Tapia, O. *J. Chem. Phys.* **2001**, *115*, 10557. (d) Arteca, G. A.; Reimann, C. T.; Tapia, O. *J. Phys. Chem. B* **2000**, *104*, 11360. (e) Arteca, G. A.; Velazquez, I.; Reimann, C. T.; Tapia, O. *J. Chem. Phys.* **1999**, *111*, 4774.

(11) (a) Sueki, M.; Lee, S.; Powers, S. P.; Denton, J. B.; Konishi, Y.; Scheraga, H. A. *Macromoleules* **1984**, *17*, 148. (b) Wojicik, J.; Altmann, K.-H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121. (c) Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 843. (d) Padmanabhan, S.; York, E. J.; Gera, L.; Stewart, J. M.; Baldwin, R. L. *Biochemistry* **1994**, *33*, 8604. (e) Luque, I.; Mayorga, O. L.; Freire, E. *Biochemistry* **1996**, *35*, 13681. (f) Padmanabhan, S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *J. Mol. Biol.* **1996**, *257*, 726.

(12) (a) Åqvist, J.; Luecke, H.; Quiocho, F. A.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2026. (b) Huyghues-Despointes, B. M. P.; Scholtz, J. M.; Baldwin, R. L. *Prot. Sci.* **1993**, *2*, 1604. (c) Armstrong, K. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11337.

(13) (a) Lyu, P. C.; Wemmer, D. E.; Zhou, H. X.; Pinker, R. J.; Kallenbach, N. R. *Biochemistry* **1993**, *32*, 421. (b) Doig, A. J.; Baldwin, R. L. *Protein Sci.* **1995**, *4*, 1325.

(14) Gans, P. J.; Lyu, P. C.; Manning, M. C.; Woody, R. W.; Kallenbach, N. R. *Biopolymers* **1991**, *31*, 1605.

(15) (a) Padmanabhan, S.; Baldwin, R. L. *J. Mol. Biol.* **1994**, *241*, 706. (b) Padmanabhan, S.; Baldwin, R. L. *Protein Sci.* **1994**, *3*, 1992. (c) Padmanabhan, S.; Jiménez, M. A.; Laurents, D. V.; Rico, M. *Biochemistry* **1998**, *37*, 17318.

(16) (a) Armstrong, K. M.; Fairman, R.; Baldwin, R. L. *J. Mol. Biol.* **1993**, *230*, 284. (b) Fernández-Recio, J.; Vázquez, A.; Civera, C.; Sevilla, P.; Sancho, J. *J. Mol. Biol.* **1997**, *267*, 184.

(17) (a) Huyghues-Despointes, B. M. P.; Scholtz, J. M.; Baldwin, R. L. *Protein Sci.* **1993**, *2*, 80. (b) Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L. *Biochemistry* **1993**, *32*, 9668. (c) Huyghues-Despointes, B. M. P.; Baldwin, R. L. *Biochemistry* **1997**, *36*, 1965. (d) Olson, C. A.; Spek, E. J.; Shi, Z.; Vologodshii, A.; Kallenbach, N. R. *Proteins Struct., Funct., Genet.* **2001**, *44*, 123.

(18) (a) Huyghues-Despointes, B. M. P.; Klingler, T. M.; Baldwin, R. L. *Biochemistry* **1995**, *34*, 13267. (b) Stapley, B. J.; Doig, A. J. *J. Mol. Biol.* **1997**, *272*, 465. (c) Smith, J. S.; Scholtz, J. M. *Biochemistry* **1998**, *37*, 33.

(19) Petukhov, M.; Kil, Y.; Kuramitsu, S.; Lanzov, V. *Proteins Struct., Funct., Genet.* **1997**, *29*, 309.

(20) Hudgins, R. R.; Mao, Y.; Ratner, M. A.; Jarrold, M. F. *Biophys. J.* **1999**, *76*, 1591.

(21) Samuelson, S.; Martyna, G. J. *J. Phys. Chem. B* **1999**, *103*, 1752.

(22) Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 12974.

(23) Counterman, A. E.; Clemmer, D. E. *J. Am. Chem. Soc.* **2001**, *123*, 1490.

(24) Kinnear, B. S.; Jarrold, M. F. *J. Am. Chem. Soc.* **2001**, *123*, 7907.

(25) (a) Valentine, S. J.; Counterman, A. E.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. *J. Phys. Chem. B* **1999**, *103*, 1203. (b) Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1188.

(26) Counterman, A. E.; Clemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 4031.

(27) The following abbreviations for synthetic amino acid residues appear throughout the text: γ-Abu (γ-aminobutyric acid), Phg (phenylglycine), Cha (cyclohexlyalanine), and Phe(NO_2) (paranitrophenylalanine).

(28) Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E. *Anal. Chem.* **1998**, *70*, 2236.

(29) The high-pressure drift tube used in these studies is identical to the configuration described previously: Counterman, A. E.; Valentine, S. J.; Srebalus, C. A.; Henderson, S. C.; Hoaglund, C. S.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 743.

(30) Lebl, M.; Krchnak, V. *Methods Enzymol.* **1997**, *289*, 336.

(31) Wellings, D. A.; Atherton, E. *Methods Enzymol.* **1997**, *289*, 44.

(32) The Xxx = γ-Abu and Phe(NO_2) peptide are expected to be present in the Xxx(i) $\rightarrow$ Leu(i + 4) libraries based on the synthetic protocol; however, there is no substantial ion intensity observed at m/z ratios corresponding to these $[M + 2H]^{2+}$ ions (m/z = 571.3 and 624.9). Our experience with γ-Abu couplings (verified by quantitative ninhydrin tests) indicates that coupling efficiencies for this amino acid are exceptionally low. In addition, the high content of alanine residues and the absence of solubilizing polar residues in the peptides required the use of high percentages of TFA as an electrospray solvent. TFA is known to significantly suppress electrospray ionization; this is consistent with our observation of an overall reduction in ion intensity for all of the libraries in TFA solvent mixtures examined here. This suppression combined with reduced abundances for the γ-Abu and Phe(NO_2) peptides due to incomplete synthetic couplings may result in their absence in the ion mobility/time-of-flight data.

(33) Our experience with these systems indicates that coupling reactions involving these residues are substantially less efficient than for other residues. In the present system, all couplings were monitored by the quantitative ninhydrin test.

(34) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley: New York, 1988.

(35) InsightII; Biosym/MSI: San Diego, CA, 1995.

(36) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86.

(37) Calculated EHSS cross sections have been calibrated according to the following relation: $\Omega_{EHSS}(cal) = 3.405 \times 10^{-5} (\Omega_{EHSS})^2 + 0.97226 (\Omega_{EHSS}) - 3.23$.

(38) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082.

(39) Counterman, A. E.; Clemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 4031.

(40) Zhang, K.; Cassady, C. J.; Chung-Phillips, A. *J. Am. Chem. Soc.* **1994**, *116*, 11512.

(41) Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. *Anal. Chem.* **2001**, *73*, 6000. Myung, S.; Badman, E.; Lee, Y. J.; Clemmer, D. E. *J. Phys. Chem. A* **2002**, *106*, 9976.