Volumes of Individual Amino Acid Residues in Gas-Phase Peptide Ions

Anne E. Counterman and David E. Clemmer*

Contribution from the Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received December 17, 1998

Abstract: Collision cross section measurements and molecular modeling calculations have been combined to extract average volumes of amino acid residues in gas-phase peptide ions. The approach uses cross sections for a series of 103 lysine-terminated peptide ions having the general form $[Xxx_nLys+H]^+$ (where Xxx is any amino acid except Lys, Arg, His, and Cys, and $n = 4$ to 8). The results show that average volumes of individual Xxx residues in the gas phase are 5 to 15% smaller than when found in protein interiors; the Lys residue is 20% smaller in the gas phase than in protein cores. The relatively small volumes of amino acids in the gas phase compared with protein core volumes are explained by considering that few gaps or crevices along interior regions exist in compact model structures that represent the experimental data. The very small volume of the Lys residue in peptide ions is consistent with the expectation that this amino acid is the charge carrier in most (if not all) of the peptides and is subject to "electroconstriction". Some implications of these results in understanding amino acid packing in gas-phase peptide ions and in protein cores are discussed.

Introduction

A common feature of native protein structures is the dense packing of amino acid residues within interior regions. Richards has described residue packing as a three-dimensional jigsaw puzzle with many intricately placed pieces.[1] A number of studies have examined the coordinates of crystal structures to investigate favorable interresidue interactions. Preferential interactions appear to be driven by electrostatic forces: ∼30% of interior charged residues exist as ion pairs;[2] and aromatic groups display preferential alignments.[3] A key parameter in understanding packing is the volume that individual amino acid residues occupy in different environments. In this paper we report average volumes for individual amino acid residues that have been derived for a series of five to nine residue peptides from structural measurements that were conducted in the gas phase. The peptides in these studies are entirely free of solvent; thus, these measurements provide a first look at individual residue volumes that are determined solely by interresidue interactions. Understanding the volumes that individual amino acid residues occupy in different environments is central to many important issues in biochemistry such as the influence of residue packing on conformer stability, function, and folding pathway.

Extensive work involving the volumes of amino acid residues in crystals and measurements of volumes of individual amino acids in solution has been carried out. Residue volumes can be derived from atomic coordinate data (either from theory or crystals) by constructing Voronoi polyhedra[4] that are modified to include atomic van der Waals radii.[5] Average volumes for residues in core regions of protein crystals are substantially smaller than values from thermodynamic measurements for individual amino acids in solution.[6−8] The volume that is associated with a specific residue often varies because gaps or crevices, due to solvent−solvent or solvent−protein interactions, are included in the space allotted to atoms in the polyhedra. Molecular modeling studies show that the average volumes occupied by residues at the surface−water interface are ∼6% larger than those for residues that are entirely buried in the core.[9] An exception is that charged residues on the surface undergo "electroconstriction" and are smaller than interior residues. Table 1 lists average volumes for amino acid residues within core regions of protein crystals[9] as well as volumes determined from thermodynamic studies of individual amino acids in solution.[6−8]

The measurements of solvent-free peptides presented below take advantage of the recently developed electrospray ionization (ESI)[10] source for mass spectrometry. Exact determination of biomolecular masses for identification, sequencing, examining modifications and mutations, and the formation of covalent and noncovalent complexes is now routine in many laboratories.[11] A few groups are attempting to extract information about the conformations of the solvent-free ions in the gas phase. Comparisons of anhydrous and solvated structures can help to delineate the roles of intramolecular interactions and solvent in establishing conformation.[12] Current methods for studying conformation in the gas phase include the following: molecular modeling,[13] hydrogen−deuterium exchange,[14] molecular ad-

(1) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
(2) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1983**, *168*, 867. Rashin, A. A.; Honig, B. *J. Mol. Biol.* **1984**, *173*, 515. Barlow, D. J.; Thornton, J. M. *Biopolymers* **1986**, *25*, 1717.
(3) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23.
(4) Voronoi, G. F. *J. Reine Angew. Math.* **1908**, *134*, 198.
(5) In this approach, the volume associated with a specified atom is deduced by construction of a series of vertices defined by nearby atoms. Richards, F. M. *J. Mol. Biol.* **1974**, *82*, 1.
(6) Rao, M. V. R.; Atreyi, M.; Rajeswari, M. R. *J. Phys. Chem.* **1984**, *88*, 3129.
(7) Mishra, A. K.; Ahluwalia, J. C. *J. Phys. Chem.* **1984**, *88*, 86.
(8) Jolicoeur, C.; Riedl, B.; Desrochers, D.; Lemelin, L. L.; Zamojska, R.; Enea, O. *J. Soln. Chem.* **1986**, *15*, 109.
(9) Harpaz, Y.; Gerstein, M.; Chothia, C. *Structure* **1994**, *2*, 641.
(10) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.
(11) For recent reviews, see: Przybylski, M.; Glocker, M. O. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 806. Smith, R. D.; Bruce, J. E.; Wu, Q.; Lei, Q. P. *Chem. Soc. Rev.* **1997**, *26*, 191. Loo, J. A. *Mass Spectrom. Rev.* **1997**, *16*, 1.
(12) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2426.

10.1021/ja984344p CCC: $18.00
Published on Web 04/02/1999

Table 1. Volumes (Å^3) of Amino Acids in Solution and Residues in Protein Crystals and Gas-Phase Peptide Ions[a]

residue	protein core[b]	solution[c]	peptide ions[d]	extended[e]
Gly	63.8(2.9)	71.7	56.5(2.3)	52.7(1.2)
Ala	90.1(4.2)	100.3	81.8(1.1)	72.1(0.6)
Val	139.1(4.7)	150.6	122.7(1.7)	106.9(0.9)
Ile	164.9(6.2)	175.4	144.1(3.0)	128.1(1.6)
Leu	164.6(5.9)	178.7	142.8(1.5)	127.5(0.7)
Met	167.7(6.7)	174.9	148.3(8.2)	134.1(3.5)
Phe	193.5(5.9)	202.3	171.7(3.1)	154.4(1.3)
Tyr	197.1(6.5)	205.3	183.5(5.1)	167.7(3.0)
Trp	231.7(5.6)	239.0	210.7(11.1)	188.2(5.3)
Ser	94.2(3.7)	100.7	89.4(3.2)	83.9(1.4)
Thr	120.0(4.8)	127.6	111.5(2.1)	99.8(0.9)
Asn	127.5(4.2)	128.4	115.9(3.3)	106.0(1.7)
Asp	117.1(4.0)	113.1	111.0(2.3)	104.3(1.2)
Gln	149.4(4.9)	156.0	134.7(8.5)	122.8(2.8)
Glu	140.8(5.3)	140.2	131.4(1.7)	124.8(0.8)
Pro	123.1(5.9)	137.2	106.4(5.3)	98.1(2.2)
Lys	170.0(5.1)	170.3	131.0(3.7)	156.5(1.8)

[a] Unless otherwise noted, values correspond to volumes for residues. [b] Values are taken from ref 9. [c] Values correspond to individual amino acids, as taken from ref 9. [d] Values correspond to the average Connolly volumes determined from trial conformers obtained from molecular modeling of 103 $[Xxx_nLys+H]^+$ peptides (listed in Table 2). See text for discussion. [e] Values were determined from Connolly volumes for fully extended forms of the $[Xxx_nLys+H]^+$ peptides listed in Table 2 by using the approach associated with eq 2. See text.

duction,[15,16] and proton-transfer reactivity studies;[14d,17] kinetic energy release measurements;[18] microscopy studies of the high-energy impacts of proteins on surfaces;[19] and measurements of collision cross sections by ion scattering[20] and ion mobility methods.[21−23] In this study, we use ion mobility techniques to determine cross sections for a series of 103 peptide ions having the general form $[Xxx_nLys+H]^+$, where Xxx is any residue except Lys, Arg, His, or Cys and $n = 4$ to 8. Molecular modeling methods are used to produce conformers of each peptide having cross sections that are in agreement with its experimental value. The volumes of these structures are then used to determine the average volumes for individual amino acid residues in the gas phase. The average volumes of residues that we obtain are given in Table 1 and show that in most cases residue volumes are ~5 to 15% smaller than when found in protein cores. We attribute this to the fact that, over the five to nine residue size range, peptides appear to have tightly packed conformations that are largely determined by the interaction of the charged site with electronegative groups; there are few gaps or crevices in such small systems. Additionally, we have derived intrinsic volumes from residues in extended linear conformations. These values provide a means of assessing the packing density of residues in protein cores and in peptide ions. From these results, it appears that polar residues pack more efficiently than nonpolar groups (in the gas phase and in protein cores). This analysis also shows that the volume of the Lys residue is substantially smaller than the intrinsic value derived from linear conformers. This result is consistent with the expectation that Lys is associated with the charge site in most (if not all) of the ions we have studied and undergoes substantial "electroconstriction" (i.e., a dense packing around the charged site due to favorable electrostatic interactions).

Experimental Approach

General. The strategy for determining average residue volumes can be summarized as follows: (1) accumulate a large database of cross sections for peptides with sequences that includes all amino acid residues of interest; (2) for each sequence, generate arrays of trial model conformers with calculated cross sections that agree with experiment; (3) compute the average volume for each peptide sequence; and (4) write and solve a system of equations that relates peptide volumes for each sequence to the frequency of occurrence and average volume of each amino acid residue. The procedures for each of these steps are described in this section.

Nested Drift (Flight) Time Measurements. A schematic diagram of the ESI/ion trap/ion mobility/time-of-flight apparatus that was used in these studies is shown in Figure 1 and has been described in detail previously.[24,25] Only a brief description is given here. The instrument is specifically configured for investigation of mixtures. Briefly, mixtures of peptides were electrosprayed into the gas phase and accumulated in an ion trap. Ions are pulsed from the trap into a drift tube and separated by differences in their mobilities. As ions exit the drift tube, they enter the source region of the time-of-flight mass spectrometer. Here, high-voltage, high-frequency pulses synchronous with the initial injection pulse are used to initiate time-of-flight measurements in the mass spectrometer. Because flight times in the mass spectrometer are much shorter than drift times in the drift tube it is possible to record hundreds of *m*/*z* distributions at specified drift times for each pulse of ions that is injected into the drift tube. This is referred to as a nested drift (flight) time measurement and, as shown below, the approach allows cross sections for numerous ions to be determined from a single data set. The peptides that are studied here were generated by tryptic digestion of common proteins. Complete descriptions of ion traps[26] and ion

(13) See, for example: Martyna, G. J. *J. Chem. Phys.* **1996**, *104*, 2018. Samuelson, S. O.; Tobias, D. J.; Martyna, G. J. *J. Phys. Chem. B* **1997**, *101*, 7592. Reimann, C. T.; Velazquez, I.; Tapia, O. *J. Phys. Chem. B* **1998**, *102*, 2277.

(14) (a) Winter, B. E.; Light-Wahl, K. J.; Rockwood, A. L.; Smith, R. D. *J. Am. Chem. Soc.* **1992**, *114*, 5897. (b) 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. (c) 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. (d) Cassady, C. J.; Carr, S. R. *J. Mass Spectrom.* **1996**, *93*, 3143.

(15) Fye, J. L.; Woenckhaus, J.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 1327.

(16) McLuckey, S. A.; Goeringer, D. E. *Anal. Chem.* **1995**, *67*, 2493. Stephenson, J. L., Jr.; McLuckey, S. A. *J. Am. Chem. Soc.* **1996**, *118*, 7390. Stephenson, J. L., Jr.; McLuckey, S. A. *Anal. Chem.* **1996**, *68*, 4026. McLuckey, S. A.; Stephenson, J. L., Jr.; Asano, K. G. *Anal. Chem.* **1998**, *70*, 0, 1198. Stephenson, J. L., Jr.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 585.

(17) Ogorzalek Loo, R. R.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 207. Ogorzalek Loo, R. R.; Winger, B. E.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 1064. Schnier, P. F.; Gross, D. S.; Williams, E. R. *J. Am. Chem. Soc.* **1995**, *117*, 6747. Williams, E. R. *J. Mass Spectrom.* **1996**, *31*, 831. Rodriguez-Cruz, S. E.; Klassen, J. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 565. Zhang, X.; Cassady, C. J. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 1211.

(18) Kaltashov, I. A.; Fenselau, C. C. *J. Am. Chem. Soc.* **1995**, *117*, 9906. Adams, J.; Strobel, F.; Reiter, A. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 30. Kaltashov, I. A.; Fenselau, C. C. *Proteins* **1997**, *27*, 165.

(19) Sullivan, P. A.; Axelsson, J.; Altmann, S.; Quist, A. P.; Sunqvist, B. U. R.; Reimann, C. T. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 329. Reimann, C. T.; Sullivan, P. A.; Axelsson, J.; Quist, A. P.; Altmann, S.; Roepstorff, P.; Velazquez, I.; Tapia, O. *J. Am. Chem. Soc.* **1998**, *120*, 7608.

(20) (a) Covey, T. R.; Douglas, D. J. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 616. (b) Cox, K. A.; Julian, R. K.; Cooks, R. G.; Kaiser, R. E. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 127. (c) Collings, B. A.; Douglas, D. J. *J. Am. Chem. Soc.* **1996**, *118*, 4488. (d) Chen, Y.-L.; Collings, B. A.; Douglas, D. J. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 681.

(21) (a) von Helden, G.; Wyttenbach, T.; Bowers, M. T. *Science* **1995**, *267*, 1483. (b) Wyttenbach, T.; von Helden, G.; Bowers, M. T. *J. Am. Chem. Soc.* **1996**, *118*, 8355.

(22) (a) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995**, *117*, 10141. (b) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2240. (c) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 9586.

(23) (a) Valentine, S. J.; Clemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 3558. (b) Valentine, S. J.; Anderson, J.; Ellington, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **1997**, *101*, 3891. (c) Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 954.

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

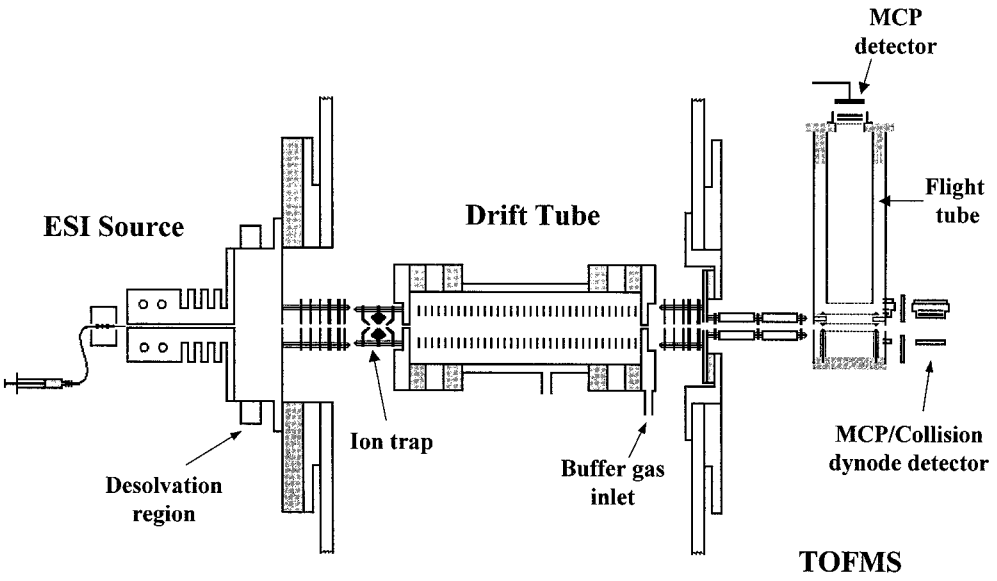

Figure 1. Schematic diagram of the experimental apparatus.

mobility/MS methods[27,28] (including our hybrid ion trap/ion mobility/time-of-flight approach)[25] have been given previously.

Formation of Mixtures of Peptide Ions. Mixtures of tryptic digest peptides for each protein[29] were generated by addition of 150 μL of a 0.2 mg/mL trypsin (Sigma, sequencing grade) solution in 0.2 M ammonium bicarbonate (EM Science) to 0.5 mL of a 20 mg/mL solution of each protein. After incubation for 20 h at 37 °C the trypsin was filtered from the digest with a microconcentrator (microcon 10, Amicon, Inc.) and the peptides that remained were lyophilized. Positively charged (protonated) ions were formed by electrospraying solutions containing 0.5 mg/mL of the tryptic digest in 49:49:2 (% v) water:acetonitrile:acetic acid. The ESI needle was biased +3200 V relative to the entrance of the desolvation region. Typical solution flow rates were 0.05 to 0.10 mL/h. Solutions were electrosprayed at atmospheric pressures into a 5 cm long differentially pumped desolvation region that operates at a pressure of 1 to 10 Torr.

Ion Trap/Ion Mobility/Time-of-Flight Considerations. Some ions exit the source through a 0.010 cm diameter orifice and enter the main chamber of the instrument where they are accumulated in an ion trap (typically for 100 to 300 ms). Concentrated packets of ions are then injected into a 40.4 cm long drift tube that was operated using ~2 to 3 Torr of He buffer gas and a uniform 8.663 V cm^{-1} drift field. Ions were injected into the drift tube using injection voltages of 70 to 110 V. As ions drift through the gas they are separated by differences in their mobilities. Flight time distributions at specified drift times (defined by a programmable delay generator, Lecroy 4222) are recorded using a time-to-digital converter (Lecroy 4208). These data are stored as a three-dimensional array of drift times (in the ion mobility instrument), flight times (in the mass spectrometer), and ion intensities for all ions in the electrospray distribution.

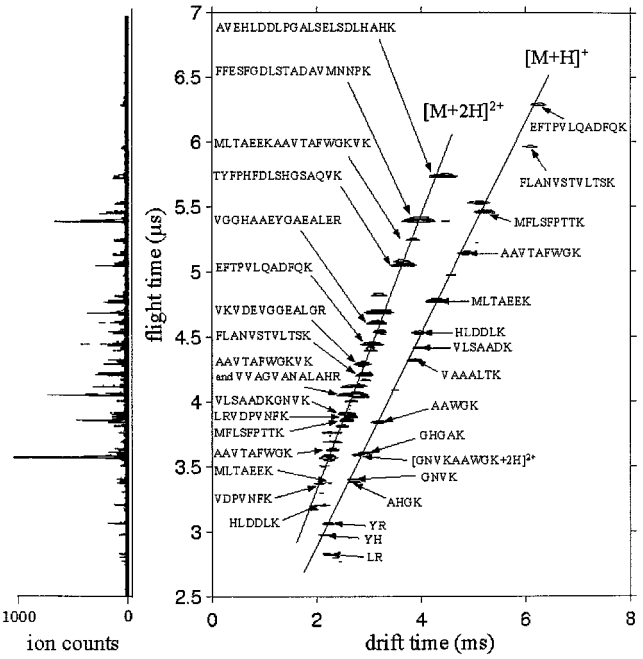

Figure 2. Contour plot of drift times and flight times obtained by electrospraying a mixture of peptides from tryptic digestion of the α- and β-chains of human hemoglobin. Drift times are normalized to a buffer gas pressure of 2.000 Torr. Peak assignments are based on the comparison of masses calculated from charge state and measured flight times with molecular weights of sequences expected from the tryptic digestion. The flight time distribution shown to the left is obtained by compressing the drift time axis.

Example of a Nested Drift(Flight) Time Data Set. Figure 2 shows a two-dimensional contour plot of one example of a data set obtained for a tryptic digest of the α- and β-chains of hemoglobin (mol wt = 15126.36 and 15867.22, respectively). A mass spectrum, obtained by compression of the drift time axis, is also shown. The appearance of these data is similar to results obtained for tryptic digests of most of the proteins that we have studied. We have recently shown that most ions observed upon ESI of tryptic digests fall into families corresponding to $[M + H]^+$ or $[M + 2H]^{2+}$ ions.[30] For a given mass-to-charge (m/z), $[M + 2H]^{2+}$ peptides have higher mobilities than $[M + H]^+$ peptides because the former ions experience twice the drift force. For a given charge state, compact conformations have higher mobilities

(25) Henderson, S. C.; Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *Anal. Chem.* **1999**, *71*, 291.

(26) March, R.; Huges, R. *Quadrupole Storage Mass Spectrometry*; Wiley: New York, 1989. Cooks, R. G.; Kaiser, R. E., Jr. *Acc. Chem. Res.* **1990**, *23*, 213. Glish, G. L.; McLuckey, S. A. *Int. J. Mass Spectrom. Ion Processes* **1991**, *106*, 1. March, R. E. *J. Mass Spectrom.* **1997**, *32*, 351.

(27) Hagen, D. F. *Anal. Chem.* **1979**, *51*, 870. Tou, J. C.; Boggs, G. U. *Anal. Chem.* **1976**, *48*, 1351. St. Louis, R. H.; Hill, H. H. *Crit. Rev. Anal. Chem.* **1990**, *21*, 321. Hill, H. H.; Siems, W. F.; St. Louis, R. H.; McMinn, D. G. *Anal. Chem.* **1990**, *62*, 1201A. von Helden, G.; Hsu, M.-T.; Kemper, P. R.; Bowers, M. T. *J. Chem. Phys.* **1991**, *95*, 3835. Karpas, Z.; Stimac; Rappaport, Z. *Int. J. Mass Spectrom. Ion Processes* **1988**, *83*, 163. Jarrold, M. F. *J. Phys. Chem.* **1995**, *99*, 11.

(28) For recent reviews of ion mobility studies of biomolecules see: Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32*, 577. Liu, Y.; Valentine, S. J.; Counterman, A. E.; Hoaglund, C. S.; Clemmer, D. E. *Anal. Chem.* **1997**, *69*, 728A.

(29) Complete descriptions of enzymatic digestion can be found in the following references: Findlay, J. B. C.; Gelsow, M. J., Eds. *Protein Sequencing: A Practical Approach*; IRL Press: Oxford, 1989; p 43. Burdon, R. H.; Knippenberg, P. H., Eds. *Laboratory Techniques in Biochemistry and Molecular Biology*; Elsevier: New York, 1989; Vol. 9, p 73.

(smaller cross sections) than elongated conformers.[21a,22a] The $[M + H]^+$ and $[M + 2H]^{2+}$ charge state families in tryptic digests arise because trypsin cleaves proteins at basic residues. Thus, upon complete digestion, each peptide has two preferred protonation sites: the basic C-terminal Arg or Lys residue; and the N-terminal amino group. Inspection of the sequences of peptides shown in Figure 2 shows that this is the case for most of the peaks that are observed. Sequences for each peak are assigned by comparing the experimentally derived *m*/*z* ratios to those that are expected for peptide fragments that should result from digestion. Additionally, the mobilities of ions (i.e., positions of peaks with respect to other peaks which make up the $[M + H]^+$ or $[M + 2H]^{2+}$ families) aid in the assignment of charge states and sequences.[25,30]

Experimental Collision Cross Sections. The measured arrival time for a packet of ions at the detector is a composite of the drift time, flight time, and the time required to travel through other portions of the instrument. Thus, it is necessary to account for the flight time and a small correction associated with transport of the ions from the exit of the drift tube to the entrance of the time-of-flight region in determining the drift times. The correction times are small (80−140 μs) compared to the 2 to 10 ms drift times in these experiments. The experimental drift time can be converted into a collision cross section (Ω) by using the relation[28]

$$\Omega = \frac{(18\pi)^{1/2}}{16}\frac{ze}{(k_b T)^{1/2}}\left[\frac{1}{m_1}+\frac{1}{m_B}\right]^{1/2}\frac{t_D E}{L}\frac{760}{P}\frac{T}{273.2}\frac{1}{N} \qquad (1)$$

where t_D is the drift time (taken as the maximum of each peak), E is the drift field, T and P are the temperature and pressure of the buffer gas, respectively, L is the drift tube length, ze is the ion's charge, N is the neutral number density, k_b is Boltzmann's constant, and m_I and m_B are the masses of the ion and buffer gas, respectively. All of the parameters E, L, P, T, and t_D can be precisely measured. The reproducibility of measured cross sections is excellent; typical relative standard deviations are less than ±1.5% (for sequences with more than three independent measurements). Ion energies in the drift tube are determined by E/N. The present experiments were carried out at low E/N where mobilities are independent of the applied drift field and drift velocities are small compared with the thermal velocity of the buffer gas. Under these conditions, ions are not expected to align in the drift tube and we assume that collision cross sections correspond to an average of all possible orientations.

A final note involves the determination of drift times. We take the drift time to be the peak maximum for all peptides in this study. In almost all cases, the peak maximum is located in the center of the peak distribution and peak shapes are accurately represented by calculated peak shapes for arrival of a single conformer from the transport equation.[28,31] Although this suggests that only a single conformer may be present, the result is also consistent with a distribution of structures that rapidly interconvert on the time scale of the experiment. Molecular modeling calculations of peptide ion dynamics at 300 K suggest that for most sequences with compact conformations, structures interconvert rapidly compared with the experimental time scales. Thus, we believe that cross sections correspond to many states that are centered about the energetically most favorable conformer. In some cases, peaks are skewed slightly, suggesting that multiple stable states that interconvert near the experimental time scale (or over longer times) are present. We do not observe multiple resolved peaks for any of the small peptides studied here.

Molecular Modeling, Cross Section, and Volume Calculations. Trial conformations of the 103 peptide sequences studied here were generated using the AMBER force field and the Insight II molecular modeling program.[32] Structures were considered after subjecting the ions to a two-stage simulated annealing cycle. For each peptide sequence, 100 trial conformers were generated in an initial simulated annealing procedure where extended forms of each protonated sequence were heated to 1000 K over 2 ps, equilibrated for 2 ps, and then cooled over 1 ps to 300 K. The five lowest energy structures were selected, reheated to 500 K over 2 ps, equilibrated over 2 ps, and then cooled for 1 ps to 300 K. The second annealing cycle was carried out 100 times for each of the five lowest energy conformers to generate 500 final conformations.

An issue that arises in these studies is the selection of the protonation site. From considerations of gas-phase basicities of different residues we expect the lysine residue to be protonated. However, recent work suggests that salt-bridged structures in which a proton from the carboxylic acid is transferred to the amino terminus may also be favorable.[21,33] Here, all peptide sequences have been modeled as Lys-protonated as well as salt-bridged charge configurations (where the lysine and amino sites were protonated and the carboxylic acid site was deprotonated). To generate salt-bridge conformers, it was necessary to modify the two-stage annealing process slightly; a distance constraint of 5 Å between protonation and deprotonation sites in the salt-bridged conformers was included in the first annealing stage. This constraint was removed for the second stage of annealing.

Three methods for calculating cross sections have been discussed in detail previously: the projection approximation,[34] exact hard sphere scattering (EHSS),[37] and trajectory calculations.[35,36] Trajectory calculations are expected to be most accurate; however, these calculations are too expensive to be used here. Here, cross sections are calculated by the EHSS method[37] and also by the projection approximation.[38] As noted before, the projection method ignores all scattering and potential interactions between the ion and the buffer gas; however, for relatively small ions (mol wt ≤ 1500) this method should be accurate to within a few percent.[35,39] The EHSS method ignores potential interactions but does include a scattering term, which has been shown to be especially important in calculating accurate cross sections for large ions as well as capturing concave shapes.[20d,37] The method has been calibrated by comparison with more than 100 cross sections obtained from trajectory calculations.[40] Over the ∼100 to 300 Å^2 range of cross sections studied, EHSS cross sections should agree with trajectory methods to within 1% (relative uncertainty). Comparisons of EHSS and projection cross sections for hundreds of conformations for each sequence show that cross sections usually agree to within 2%; in some cases the methods yield cross sections that differ by as much as ∼5%.

For each peptide ion sequence we select conformers produced by molecular modeling and calculate volumes. The criterion used for selection is that the calculated cross sections must be within the uncertainty of the experiment; additionally, only the lowest 10 kcal/mol structures are considered. The latter requirement is somewhat arbitrary; however, it is necessary to reduce the number of required volume calculations to 10 to 30 conformers for most sequences and prohibit inclusion of structures with unrealistically high energies. Volume calculations are carried out for Lys-protonated and salt-bridged conformers found from both methods of cross section calculation.

Peptide volumes are calculated from atomic coordinate data by a method developed by Connolly[41] that is available in the Insight II

(30) Valentine, S. J.; Counterman, A. E.; Hoaglund, C. S.; Reilly, J. P.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1213.
(31) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley: New York 1988.
(32) Insight II.; BIOSYM/MSI: San Diego, CA, 1995.
(33) Campbell, S.; Rodgers, M. T.; Marzluff, E. M.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1995**, *117*, 12840. Schnier, P. D.; Price, W. D.; Jockusch, R. A.; Williams, E. R. *J. Am. Chem. Soc.* **1996**, *118*, 7178.
(34) von Helden, G.; Hsu, M.-T.; Gotts, N.; Bowers, M. T. *J. Phys. Chem.* **1993**, *97*, 8182.
(35) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082.
(36) Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Chem. Phys.* **1998**, *108*, 2416.
(37) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86.
(38) The hard-sphere impact parameters used for the projection calculations were 2.38 Å for H and 3.02 Å for C, N, and O, as given by: Wyttenbach, T.; Bushnell, J. E.; Bowers, M. T. *J. Am. Chem. Soc.* **1998**, *120*, 5098.
(39) Wyttenbach, T.; von Helden, G.; Batka, J. J., Jr.; Carlat, D.; Bowers, M. T. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 275.
(40) EHSS cross sections were calculated as described previously in ref 37. These calculations use impact parameters of 2.2 for H, 3.1 for C and O, and 2.8 for N. The results have been calibrated to values obtained from the trajectory method by Jarrold and co-workers by the following relation: $\Omega_{EHSS}(cal) = -3.23 + 0.97226\Omega_{EHSS} + (3.4205 \times 10^{-5})\Omega_{EHSS}^2$. Jarrold, M. F. (personal communication).
(41) Connolly, M. L. *J. Am. Chem. Soc.* **1985**, *107*, 1118.

software.[32] The approach treats the molecule as a collection of hard spheres defined by their van der Waals radii and relative positions.[42] The calculated volumes correspond to the region of solvent (in this case, He) that is excluded from interior regions of each rigid structure. The He probe radius was taken to be 1.1 Å. Uncertainties introduced in volume calculations by varying this radius by ±10% are negligible (~0.01%). We tested the accuracy of volume calculations on several types of geometries with analytical solutions, including solid and hollow spheres of varying sizes, and found that calculated volumes were in near perfect agreement with analytical solutions. Thus, uncertainties associated with calculation of volume are insignificant compared with the other sources discussed above.

Results and Discussion

Experimental Collision Cross Sections. A consideration in the derivation of gas-phase residue values is that sequences should reflect the compositions of residues that are found in naturally occurring proteins. To accomplish this, we start with a series of 103 $[Xxx_nLys+H]^+$ peptides that are obtained by tryptic digestion of common proteins; as discussed below, the residue compositions are similar to those found in protein cores.[9] Table 2 lists 103 cross sections, sequences, and source proteins for the series of singly protonated peptides. These data are a subset of a database of cross sections for peptides from tryptic digestion. The complete database contains cross sections for 660 different ions including the following: 420 singly protonated peptides ranging in size from 2 to 15 residues and 240 doubly protonated peptides with 4 to 24 residues. The data in Table 2 include 23 five residue, 17 six residue, 29 seven residue, 20 eight residue and 14 nine residue peptides that are singly-protonated. Plots of cross sections as a function of peptide molecular weight show that cross sections for peptides with similar molecular weights vary as much as ~10 to 15% over the range of molecular weights that are studied.[43]

Cross Sections and Volumes for Trial Model Conformers. The average residue volumes that are obtained below suggest that the approach taken here is relatively insensitive to the two-stage simulated annealing method and details of the charge assignment. Figure 3a shows an example scatter plot[21b] of cross sections that were obtained after a single-stage and two-stage annealing cycle for the $[FMMFESQNK+H]^+$ ion, where the charge has been placed on the C-terminal lysine residue. Analogous plots for salt-bridged charge assignments appear similar. After a single annealing cycle, four of the conformers have cross sections and energies that meet the criteria described in the Experimental Section. After two annealing cycles there are 26 low-energy conformers with cross sections that are in agreement with experiment. Similar results are found for the other peptides in this study.

Figure 3b shows a plot of Connolly volumes as a function of cross section (projection approximation) for a range of different conformations of $[FMMFESQNK+H]^+$. Here, trial structures vary from linear to compact. As the peptide chain stretches into a more linear conformation, the average cross section increases. However, the Connolly volumes decrease as the peptide opens up. This can be understood by considering that in the limiting linear structure there are no interior gaps or crevices such as those that arise in folded conformers.[44] Volumes for folded conformers are typically 5 to 9% greater than those for linear structures over the size range of this study.

A comparison of average volumes obtained for each peptide sequence (C-terminal Lys proton assignment) by selecting conformations by using the projection and EHSS models is shown in Figure 4. Although differences between cross sections

Table 2. Sequences and Collision Cross Sections of $[Xxx_nLys+H]^+$ Peptides ($n = 4$ to 8)[a]

5 residues	6 residues	7 residues	8 residues	9 residues
AAWGK[s,u,v,x]	AAAAEK[p]	APVDAFK[z]	AADALLLK[p]	AAVTAFWGK[s]
157.36(1.02)	160.38(1.06)	189.05(0.76)	223.86(2.78)	237.76(0.53)
ADLAK[b,e,g]	ANIDVK[p]	ATDEQLK[g]	ADFAEISK[e]	AAVTGFWGK[x]
159.31(1.06)	176.78(0.76)	205.89(0.16)	218.06(0.55)	233.00(0.18)
AFDEK[b]	ASEDLK[ae]	ATEEQLK[b]	ADFAEVSK[e]	ANELLINVK[h]
168.36(0.44)	175.16(0.58)	206.40(1.81)	214.17(0.67)	249.69(0.86)
AIAEK[z]	EAMAPK[k]	DGADFAK[i]	ADFTDVTK[g]	EAVLGLWGK[v]
160.73(1.84)	176.19(1.24)	185.83(0.90)	214.56(1.13)	235.37(1.14)
APNAK[q]	EMPFPK[k]	DSAIMLK[l]	ADFTEISK[f]	FMMFESQNK[l]
147.31(0.98)	193.26(1.43)	203.77(1.32)	219.61(0.21)	259.37(4.27)
AWGGK[t]	IEEIFK[n]	ELTEFAK[g]	AEFAEVSK[e]	FQPLVDEPK[f]
152.16(2.53)	197.00(2.97)	209.40(0.69)	214.97(4.51)	255.89(0.74)
DIAAK[ae]	IVAPGK[l]	EVTEFAK[f]	ALQASALK[i]	MFLGFPTTK[v,w]
155.37(2.98)	173.33(1.17)	202.91(0.29)	218.92(0.62)	250.02(1.45)
DLLFK[j,l,z,ad]	LIFAGK[ae]	FNDLGEK[c]	DIVGAVLK[h]	MFLSFPTTK[s,u]
183.11(3.14)	186.10(1.58)	206.73(1.10)	206.07(0.54)	255.16(0.70)
FFSDK[m]	LVEDLK[r]	GDVAFVK[j,l,ad]	DLGEENFK[e]	QSALAELVK[c]
172.73(2.52)	192.17(1.05)	200.45(1.77)	223.98(1.41)	234.39(1.35)
GGNMK[n]	MQIFVK[ae]	GGVVGIK[t]	DLGEQYFK[f]	QTALVELLK[g]
147.44(1.24)	203.9(0.59)	175.87(0.62)	232.22(0.30)	245.01(1.53)
GITWK[o]	NDIAAK[ae]	IATAIEK[p]	DSADGFLK[j]	QTALVELVK[c]
169.34(0.58)	173.78(0.84)	202.92(0.37)	209.30(1.09)	242.43(0.50)
GTFAK[t,v,w]	NLDNLK[t]	ILLSSAK[c]	EYEATLEK[c]	SAVTALWGK[u,w]
153.97(0.91)	192.40(0.69)	202.97(0.75)	229.28(0.93)	231.09(1.09)
IFVQK[o]	NVPLYK[p]	IVTDLAK[f]	FGVNGSEK[l]	SLVSGLWGK[t]
181.96(0.51)	195.27(0.90)	204.75(1.40)	201.76(0.59)	236.29(0.69)
IIAEK[aa]	NYQEAK[b,e,g]	IVTDLTK[c]	GASIVEDK[r]	TFQSFPTTK[t]
172.54(0.50)	191.16(0.92)	207.02(0.28)	205.22(1.78)	245.22(0.25)
LDALK[c]	TEAEMK[ae]	LVTDLTK[b,e]	IDALNENK[aa]	
172.36(1.48)	182.54(2.25)	205.76(0.98)	225.09(1.45)	
NLNEK[ad]	TPVSEK[b,c]	MIFAGIK[o]	IGDYAGIK[h]	
167.97(0.94)	175.98(0.73)	207.13(0.39)	210.4(0.63)	
NTYEK[ad]	YLTTLK[z]	MLTAEEK[s,x]	LIVTQTMK[aa]	
175.90(2.66)	197.94(0.70)	209.77(0.59)	243.91(0.71)	
TAWEK[w]		NPDPWAK[ad]	TYETTLEK[e]	
170.03(1.25)		198.09(0.38)	239.55(1.30)	
TGQIK[p]		VAAALTK[s,x]	VLTPDLYK[n]	
157.62(0.28)		190.32(0.73)	230.73(1.13)	
TLTGK[ae]		VADALTK[v]	YLGEEYVK[ad]	
157.34(2.11)		194.69(1.73)	238.79(1.26)	
TPGSK[ab]		VDPVNFK[s,t,u,v]		
145.50(1.17)		208.73(0.91)		
TVGGK[j]		VLAAVYK[i]		
140.18(3.52)		206.91(1.08)		
YYPLK[a]		VLPVPQK[k]		
187.31(0.38)		206.94(0.57)		
		VLSAADK[s,v,x]		
		193.03(1.28)		
		VLSPADK[t,u,w]		
		196.51(1.02)		
		VLTSAAK[f]		
		194.37(0.13)		
		VSEALTK[w]		
		198.59(2.27)		
		VVTDLTK[e]		
		202.35(0.99)		
		WNMQNGK[m]		
		206.26(0.19)		

[a] Collision cross sections for each sequence correspond to those measured for singly-protonated ions. See text for details. Uncertainties correspond to one standard deviation of the mean and are given in parentheses. The superscripts given for each sequence denote the following proteins that were used in the digestion: [b]albumin (bovine), [c]albumin (dog), [d]albumin (horse),[e]albumin (human), [f]albumin (pig), [g]albumin (sheep), [h]alcohol dehydrogenase (yeast), [i]aldolase (rabbit), [j]apotransferrin (bovine), [k]β-casein, [l]κ-casein, [m]conalbumin (chicken), [n]concanavalin A (Canavalis ensiformis), [o]creatine phosphokinase (rabbit), [p]cytochrome *c* (horse), [q]enolase (yeast), [r]glucose oxidase (*Aspergillus niger*), [s]L-glutamic dehydrogenase, [t]hemoglobin (bovine), [u]hemoglobin (dog), [v]hemoglobin (human), [w]hemoglobin (pig), [x]hemoglobin (rabbit), [y]hemoglobin (sheep), [z]lactoferrin (bovine), [aa]β-lactoglobulin, [ab]lysozyme (turkey), [ac]myoglobin (horse), [ad]transferrin (human), and [ae]ubiquitin (bovine).

(42) The van der Waals radii used were the following: 1.60, carbonyl oxygen; 1.70 hydroxyl oxygen; 1.65, −NH.; 1.70, $-NH_2$; 1.75, $-NH_3^+$; 1.85, aliphatic −CH; 1.90, aliphatic $-CH_2-$; 1.95, $-CH_3$; 1.80, aromatic or carbonyl carbon; 1.90, aromatic −CH− group; and 1.90, sulfur. These parameters for extended atoms are identical with those used by: Gelin, B. R.; Karplus, M. *Biochemistry* **1979**, *18*, 1256.

(43) Valentine, S. J.; Counterman, A. E.; Hoaglund Hyzer, C. S.; Clemmer, D. E. *J. Phys. Chem. B* **1999**, *103*, 1203.

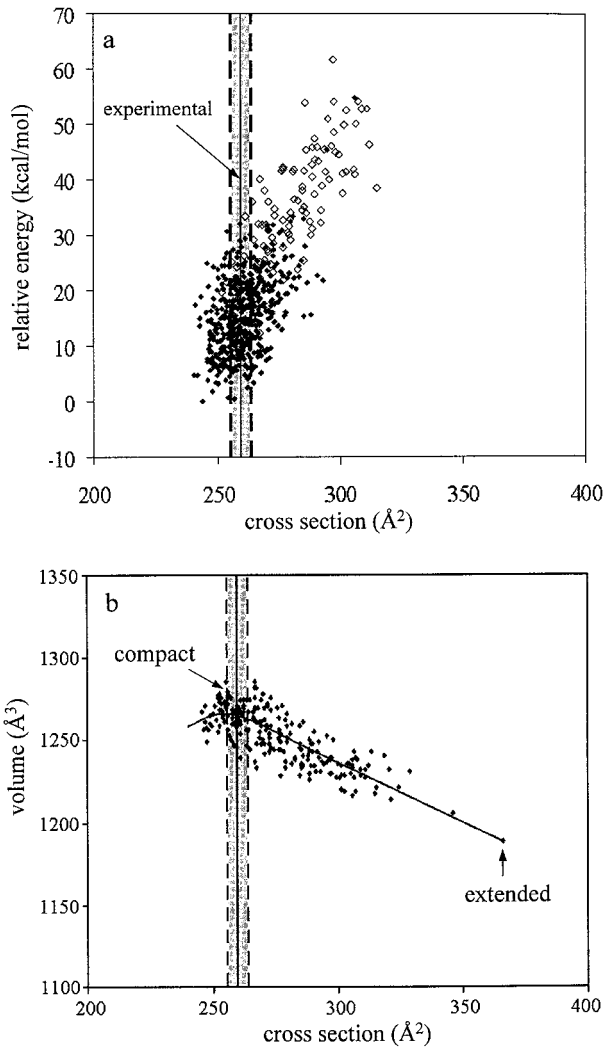

Figure 3. Part a shows calculated cross sections and energies for 100 conformers of $[FMMFESQNK+H]^+$ obtained after a single annealing cycle (open diamonds) and the 500 conformers obtained after a two-stage annealing cycle (filled diamonds). The energy axis has been defined relative to the lowest energy structure that was found in the two-stage cycle. Part b shows volumes and cross sections calculated for $[FMMFESQNK+H]^+$ conformers. The experimental cross section of 259.3 ± 4.1 Å² is shown for comparison. The arrow at 365.9 Å² shows the conformation with the smallest volume (the linear conformation).

for specific conformers can arise from the projection and EHSS methods, the volumes that are derived agree to within 1.5% for all of the peptides studied here. Volumes for individual amino acids from projection and EHSS cross section constraints (derived below) show that the two methods yield indistinguishable results. Differences found for individual sequences effectively cancel in the final calculation of volumes because many different sequences and conformers (for each sequence) are considered. A final note is that there are no significant differences between average volumes when peptides are protonated at the Lys residues compared with salt-bridge conformations.

Average Volumes of Individual Residues. The 17 unknown volumes for the individual Xxx and Lys residues are related to their occurrence frequency and peptide volumes by

(44) The Connolly volumes for linear structures reduce to near the values obtained by summing volumes of individual amino acids (for the appropriate sequence) that have been corrected for the removal of water.

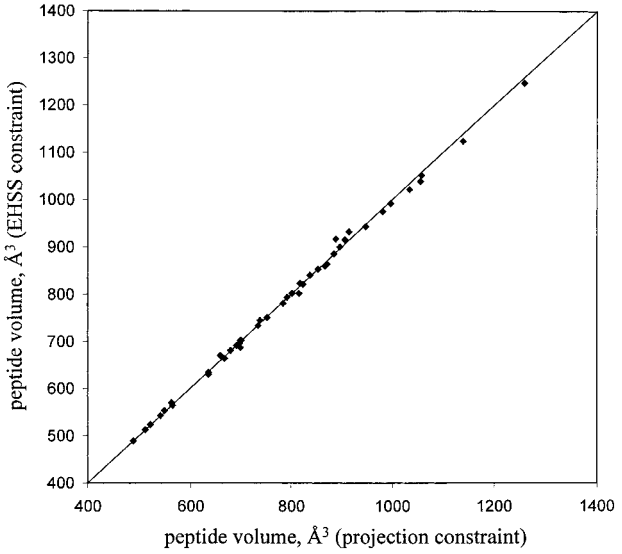

Figure 4. Average peptide volumes obtained by constraining trial conformers by comparing cross sections calculated by the projection approximation (*x* axis) and EHSS model (*y* axis) to the experimental cross section. The line shows values for which volumes constrained by the two methods are identical. See text for discussion.

$$\sum_i n_{ij}v_i = V_j(\text{calc}) \qquad (2)$$

where n_{ij} corresponds to the number of times an amino acid i occurs within a given sequence of each peptide j and the unknown amino acid residue volumes are denoted v_i with i = 1 to 17. The calculated volume of each peptide, V_j(calc) is an average of volumes calculated for all of the lowest energy conformers that meet the criteria established above. It is necessary to correct each peptide volume by subtracting 11.1 Å³ to account for an extra H atom and OH group at the N- and C-terminus of the peptide, respectively.[45] The 17 unknown volumes (v_i) are obtained by solving the system of 103 equations using the format $Ax = b$, where A is the matrix defined by n_{ij} and b is the vector comprised of the average calculated volumes, V_j. The equations are solved for v_i by computing $x = (A^TA)^{-1}A^Tb$, where x is the vector comprised of v_i.[46] This solution yields values of v_i that are effectively the linear least-squares best fit. With this approach it is possible to determine volumes for all Xxx residues. The Lys value can also be obtained and the accuracy of this value should not be impugned by the fact that this residue occurs in every sequence. Uncertainties in each value (δv_i), which correspond to one standard deviation about the mean, are determined from the residuals ($\epsilon = b - Ax$) by $(\delta v_i)^2 = (A^TA)^{-1}\epsilon^T\epsilon/(n_{\text{tot},i})$, where $n_{\text{tot},i}$ is the total number of occurrences of amino acid i in the series of peptides studied.[47] In the present studies, uncertainties are small (1.1 to 3.7 Å³ for 12 residues, ~5 Å³ for two residues, and >6 Å³ for only three residues, Table 1), an indication that volumes of individual residues are similar for the different sequences and peptide lengths in Table 2.

A comparison of average residue volumes determined from gas-phase peptide ions and those previously derived from Voronoi volumes in protein cores[9] is shown in Figure 5. Trends observed for different residue types in the gas phase and in protein cores are similar. We have grouped the residues as

(45) Chothia, C. *Nature* **1975**, *254*, 304.

(46) Leon, S. J. *Linear Algebra with Applications*, 3rd ed.; Macmillan: New York, 1990; p 208.

(47) Bethea, R. M.; Duran, B. S.; Boullion, T. L. *Statistical Methods for Engineers and Scientists*, 2nd ed.; Marcel Dekker: New York, 1985.

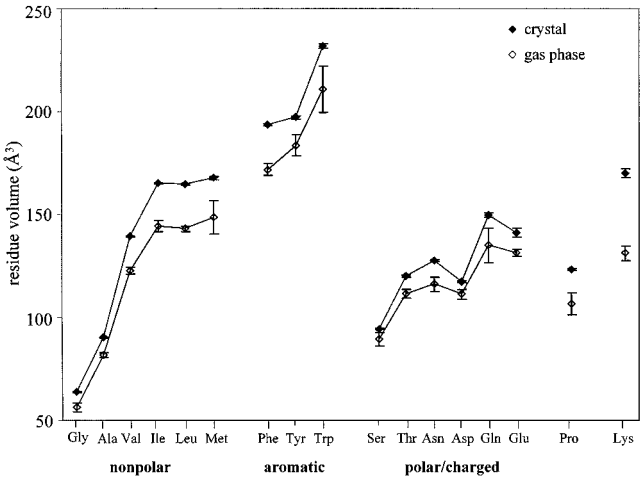

Figure 5. Average volumes and uncertainties for individual amino acid residues found in gas-phase peptide ions (open diamonds) and for residues in protein cores (solid diamonds), which are taken from ref 9.

nonpolar, aromatic, and polar/charged. The Pro and Lys residues are plotted separately. In all cases, the relative changes observed between residue types in the gas phase correlate with changes observed between residues in protein cores. The volumes determined for residues in gas-phase peptide ions are smaller than those determined for residues in protein cores. The largest differences are found for the nonpolar Val, Ile, and Leu residues, where the gas-phase values are smaller by ~15% and for Pro and Lys residues which are smaller by 14 and 23%, respectively. Residues in the polar/charged grouping differ by ~5 to 10%.

Possible Origin of Differences in Volumes. It is important to consider possible origins for differences in average volumes for peptide ions compared with protein cores. First, we have considered the possibility that large differences in the residue-He scattering dynamics associated with cross section measurements and calculations might explain some differences between residue volumes. We rule this possibility out because volumes for individual residues derived from EHSS cross sections and projection approximation are identical within the experimental uncertainty. If the dynamics of scattering alone were responsible for differences in volumes for different residues (or the smaller volumes observed for residues in the gas phase), we would have expected to observe variations related to the method used to calculate cross sections. As noted above, none are apparent.

There are several other possible origins for differences in volumes. Volumes for gas-phase peptides could differ from those in protein cores because of differences in the average residue compositions of the peptide ions studied here, compared with the protein cores. Recently, Chothia and co-workers have revised the original reports of volumes of residues in protein cores[45] by carefully excluding residues with any surface exposure.[9] This analysis led to residue volumes that were 1.1 to 8.5% smaller than previous values. As discussed above, we have studied naturally occurring residue sequences; however, it is worthwhile to consider the relative abundances of the residues in more detail. Table 3 lists the percentages of each residue in the peptide ion sequences (Table 2) as well as those used previously for core regions.[9] The sequences considered in the gas-phase peptides and protein cores show that the percentages of most types of amino acid residues are similar; however, some differences are notable. Nonpolar aliphatic residues (Gly, Ala, Val, Ile, Leu, and Met) account for 60% of the core regions of proteins, but only 42% of the tryptic digest peptides. The populations of aromatic residues (Phe, Tyr, and Trp) are 11% in crystal cores and 9% in the gas-phase peptides. Charged residues (Asp, Glu, Arg, and Lys) comprise only a small percentage (5%) of the protein cores while making up 30% of residues in gas-phase peptide ions. The smaller average residue volumes that are observed in the gas phase could be explained (at least in part) by the different compositions of amino acids. The larger fraction of polar and charged residues should lead to favorable interresidue interactions and may increase packing densities compared with the more nonpolar composition of the protein core.

Table 3. Percentages of Different Residues in the $[Xxx_nLys+H]^+$ Peptides and Protein Cores

residue type	residue	interior of protein crystals[a]	peptide ions[b]
nonpolar aliphatic	Gly	11	6
	Ala	13	12
	Val	13	8
	Ile	9	4
	Leu	12	10
	Met	2	2
		total 60	total 42
aromatic ring	Phe	6	6
	Tyr	3	2
	Trp	2	1
		total 11	total 9
polar aliphatic	Asn	1	4
	Gln	2	2
	Ser	6	4
	Thr	5	7
		total 14	total 17
	Pro	3	3
charged	Asp	3	6
	Glu	1	7
	Lys	1	17
		total 5	total 30

[a] Percentages of residues in protein cores as taken from ref 9. The total composition of nonpolar aliphatic, aromatic ring, polar aliphatic, and charged residues are given. [b] Determined from the sequences of 103 $[Xxx_nLys+H]^+$ peptides listed in Table 2.

An extreme difference in the composition of the peptide ions studied here and protein cores arises from the fact that every sequence in the gas phase contains a highly localized charge. Overall, the observation of smaller volumes for individual residues in the gas phase is consistent with the tightly packed nature of conformers that were found by molecular modeling. The shapes and general structures of model conformers are similar to those reported for the $[RPPGFSPFR+H]^+$ bradykinin sequence[21] and more recently for 8 to 16 residue polyalanine oligomers with an N-terminal proton.[48] Generally, molecular modeling of bradykinin,[21] polyalanine peptides,[48] and the tryptic fragment ions studied here shows that the charged site is solvated by interactions with electronegative oxygen atoms associated with carbonyl groups and side chains. There are few chances for gaps and crevices to arise in such small, tightly packed conformers. An example showing the tightly packed nature of the conformation of the heptapeptide, $[VDPVNFK+H]^+$, is shown in Figure 6. The experimental cross section for this ion is 208.7 ± 0.9 Å^2. The representative conformer (chosen randomly) has calculated cross sections of Ω(EHSS) = 212.4 Å^2 (within 2% of experiment)[49] and Ω(projection)= 209.2 Å^2, in good agreement with experiment. Also shown are the atomic positions of the same sequence of residues as they are found in the core of human hemoglobin, as taken from the crystal

(48) Samuelson, S. O.; Martyna, G. J. *J. Chem. Phys. B* **1999**, *103*, 1752. Hudgins, R. R.; Jarrold, M. F. *Biophys. J.* **1999**, *76*, 1591. Clemmer, D. E. Unpublished results.

(49) In this case, this conformer would not have been selected by the EHSS method.

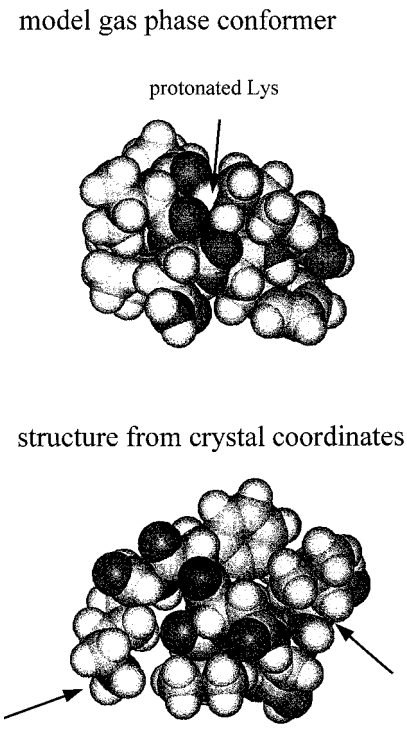

Figure 6. (a) A representative compact trial conformer of $[VDPVNFK+H]^+$ having calculated cross sections of Ω(EHSS) = 212.4 Å^2 and Ω(projection) = 209.2 Å^2 (in good agreement with the experimental value of 208.7 ± 0.9 Å^2) is shown. The arrow indicates the protonation site on the C-terminal Lys residue, which is solvated by electronegative carbonyls of the peptide backbone. (b) The conformation of the VDPVNFK sequence in the folded structure of human hemoglobin (Protein Data Bank accession number 1A3N). The arrows on the left and right indicate the terminal Lys residue and starting Val residues (respectively) by which the sequence is connected in the protein.

coordinates.[50] Visual inspection of these structures shows that the VDPVNFK sequence in the hemoglobin core is less tightly packed than the gas-phase ion. In the protein, the structure of the VDPVNFK sequence is defined by other secondary and tertiary constraints. In the gas-phase peptide ion, these residues are free to adopt a substantially more compact structure. Here, the driving force for formation of the compact gas-phase structure arises from favorable electrostatic interactions between electronegative groups and the charged Lys residue. These large differences in packing are consistent with the differences in average volumes of individual residues that are observed in gas-phase peptides compared with protein cores.

Residue Packing Efficiency. The packing efficiency can be defined as a dimensionless term called the *packing density*, which is the intrinsic volume of a moiety divided by the actual volume that it occupies in a specific environment.[51] To gain a feeling for the differences in packing efficiencies between residues in small gas-phase peptide ions and residues in protein cores, we have derived average volumes from linear conformers for each peptide sequence in Table 2. This was done by calculating Connolly volumes for conformers that have been stretched to create highly linear forms (within reason) of each sequence. Then, the method associated with eq 2 was used to determine individual residue volumes. The average volumes of individual residues in linear conformers should be near a lower limit for each residue, analogous to the lower limit shown for the volume of the $[FMMFESQNK+H]^+$ peptide in Figure 3b. We refer to these values as the intrinsic volume of each residue; each value is listed in Table 1.

(50) Tame, J.; Vallone, B., coordinates from the Protein Data Bank (accession code 1A3N).

(51) An elegant discussion of packing density is given in: Richards, F. M.; Lim, W. A. *Q. Rev. Biophys.* **1994**, *26*, 423.

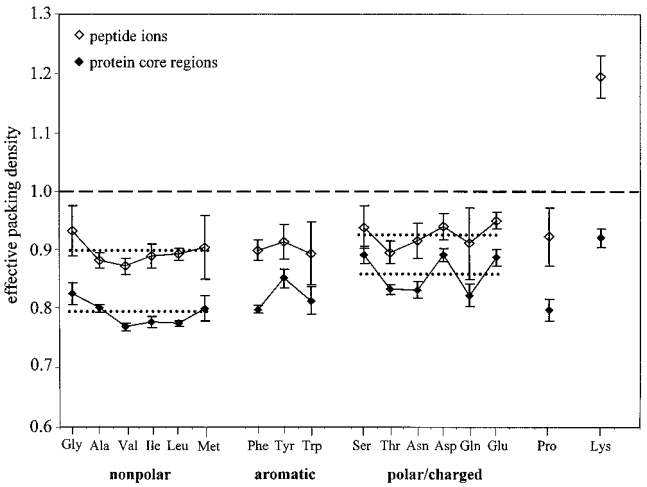

Figure 7. The effective packing densities shown are the ratio of volumes calculated for individual residues in fully extended $[Xxx_n-Lys+H]^+$ conformers to residue volumes for $[Xxx_nLys+H]^+$ gas-phase peptides (open diamonds) and those reported in ref 9 for residues in protein cores. The dashed line at 1.00 indicates packing densities that are equivalent to those of residues in fully extended sequences. The dotted lines represent the averages of values calculated for each type of residue (nonpolar, aromatic, and polar/charged) for gas-phase peptides and protein cores.

Figure 7 shows the average residue volumes in cores and in peptide ions divided by the average intrinsic residue volumes derived from trial linear conformers (denoted as effective packing density). In this plot, residues which pack together perfectly, such that there are no gaps or cavities, would have a ratio of 1.00. Comparison of the average effective packing densities of nonpolar residues in protein cores with those in gas-phase peptides (indicated by the dotted lines) shows a larger difference in packing efficiency than that between polar residues in the two systems. Thus it appears that nonpolar residues pack less efficiently in protein cores than polar residues. The nonpolar Val, Ile, and Leu residues appear to be the least efficiently packed residues; this can be understood by considering that the side chains of these residues are large, flexible, and thus may occupy a larger space than would be ascribed if their van der Waals radii alone were considered. On average, a closer agreement between effective packing densities for protein cores and gas-phase peptides is found for polar residues. In both cases, Asp, Glu, and Ser are the most efficiently packed residues. The higher packing density of these residues is consistent with the relatively long-range attractive interactions between these residues and others.

An additional outcome of this analysis is that the lysine residue is found to have a smaller volume in the gas-phase ion than is expected from average intrinisic volumes derived from the limiting linear structure. This result comes about because the charge is associated with the Lys residue in both of the models (Lys-protonated and salt-bridged) that we have considered. Examination of model structures found by simulated annealing shows that the charged Lys side chain is often buried due to charge solvation by electronegative groups. This decrease in volume is similar to the molecular modeling calculations of Chothia and co-workers for charged residues in solution, showing that electroconstriction causes volumes for charged residues to decrease.[9] The substantially smaller volume that we have found for Lys here requires that the residue behaves in a

similar fashion for most (if not all) peptide ions. It is generally accepted that, in gas-phase ions, protons should reside at the most basic groups[52] and numerous correlations between the number of basic residues and maximum observable ESI charge states[53] as well as relationships of basic residue positions and fragmentation patterns[54] have been noted. The electroconstricted volume for Lys found here is an additional corroboration that the Lys residue is associated with the charged site in tryptic digest peptides, and provides a glimpse of the general impact that this has on structure.

Summary

Measurements of collision cross sections and molecular modeling methods have been used to determine the average volumes of 17 amino acid residues in a series of $[Xxx_nLys+H]^+$ peptides (where Xxx is any naturally occurring amino acid except Cys, His, or Arg and $n = 4$ to 8). The average volumes of the individual Xxx residues in these peptide ions are 5 to 15% smaller than corresponding values determined from crystallographic data for residues in protein cores. The average volume of the Lys residue in gas-phase peptides is 20% smaller than the volume this residue occupies in protein cores. These results have been explained by considering that there are few gaps or crevices in the small peptide ion systems; the driving force for forming such compact forms is favorable electrostatic interactions upon solvation of the charge. The observation that the Lys volume in the gas phase is substantially smaller than in protein cores is consistent with the expectation that this residue is the site of protonation in most (if not all) of these peptides.

Calculations of volumes for linear conformations of each of the 103 peptide sequences studied here provide a lower limit for the typical volume that residues will occupy in typical protein sequences, or, an intrinsic volume. Comparison of these values with the volumes of residues in gas-phase peptide ions and with residues in protein cores provides insight into the effective packing densities of different residues. The highest average packing densities in protein cores and in the gas phase are found for polar and charged residues, especially Thr, Ser, Asp, and Glu. The lowest packing densities are found for the large nonpolar Val, Ile, Leu, Met, and Phe residues. These findings are consistent with the expected magnitudes of attractive interactions between these residue types that should increase with increasing polarity. The packing densities of these residues are measurably higher in the gas-phase peptide ions than in protein cores. Sequences in gas-phase peptide ions are densely packed due to favorable electrostatic interactions. Additionally, these small systems are not subject to constraints found in larger proteins, where other stuctural features can influence packing.

A primary factor in the success of these methods in determining average volumes is the large database that is available for many small ions. It seems likely that the high degree of control associated with mass spectrometry-based methods will make it possible to clarify volume changes that occur as a function of peptide size in a detailed fashion. This should allow this approach to address the nature of packing within domains of larger proteins.

Acknowledgment. This work is supported by a grant from the National Institute of Health (Grant No. 1R01GM55647-01).

JA984344P

(52) Williams, E. R. *J. Mass Spectrom.* **1996**, *31*, 831.

(53) See, for example: Covey, T. R.; Bonner, R. F.; Shushan, B. I.; Henion, J. *Rapid Comm. Mass Spectrom.* **1988**, *2*, 249. Loo, J. A.; Udseth, H. R.; Smith, R. D. *Rapid Comm. Mass Spectrom.* **1988**, *2*, 207. Schnier, P. D.; Gross, D. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1086.

(54) See, for example: Martin, S. A.; Biemann, K. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 213. Johnson, R. S.; Martin, S. A.; Biemann, K. *Int. J. Mass Spectrom. Ion Processes* **1988**, *86*, 137. Jones, J. L.; Dongre, A. R.; Somogyi, A.; Wysocki, V. H. *J. Am. Chem. Soc.* **1994**, *116*, 8368. Vachet, R. W.; Asam, M. R.; Glish, G. L. *J. Am. Chem. Soc.* **1996**, *118*, 6252.