Prediction of Peptide Ion Mobilities via a priori Calculations from Intrinsic Size Parameters of Amino Acid Residues

Alexandre A. Shvartsburg and K. W. Michael Siu
Department of Chemistry and Centre for Research in Mass Spectrometry, York University, Toronto, Ontario, Canada

David E. Clemmer
Department of Chemistry, Indiana University, Bloomington, Indiana, USA

Ion mobility spectrometry (IMS) has recently been established as a powerful tool to separate the protease digest mixtures and identify their peptide components. As accurate calculation of mobilities is critical for this technique, a new rapid method based on intrinsic size parameters (ISPs) of amino acid residues has been devised. However, those parameters had to be obtained by tedious statistical analysis of a large body of experimental data. Here we demonstrate that they can instead be derived a priori, based on the stoichiometry of a residue. Our main finding is that the ISP of a residue is essentially determined by its density, that is, the average mass/size ratio of its constituent atoms. This is in accordance with an interpretation in which peptides assume compact conformations in the gas phase dominated by the solvation of ionic charge. (J Am Soc Mass Spectrom 2001, 12, 885–888) © 2001 American Society for Mass Spectrometry

Since the development of soft-ionization methods such as electrospray and matrix-assisted laser desorption/ionization, which enable efficient production of large intact biomolecular ions in the gas phase, biological mass spectrometry (MS) has experienced an explosive growth [1]. Today, in the post-genomics era, sub-picomole sensitivity, high mass accuracy, and high resolution of modern mass spectrometry have rendered it the technique of choice for protein identification. In particular, (partial) sequencing of proteins has become routine. This is typically accomplished by first cleaving a protein into a number of peptides in solution or in the polyacrylamide gel (digestion), then lifting the peptide mixture thus produced into the gas phase as ions, and identifying its components based on their measured masses and sequences [2]. It is obviously desirable to be able to digest a protein in a controlled manner by selective scission at certain residues. One common procedure involves the enzyme trypsin that severs the peptide linkage on the C-terminal side of lysine and arginine residues, thus yielding a mixture of Lys- and Arg-terminated peptides (the tryptic digest) [3].

The power of MS is greatly enhanced by interfacing it with a suitable separation method, such as two-dimensional gel electrophoresis or liquid chromatography. Lately, coupling MS with IMS has become topical [4]. This is not only because IMS, as a fast separation method, is ideal for high-throughput analysis, but also because the measured collision integrals reveal the structure of ions investigated. A method for obtaining the mass and mobility spectra simultaneously has recently been developed, dramatically increasing the acquisition speed in IMS/MS [5]. This has allowed the mobilities for a large number of peptide ions to be tabulated. In particular, the collision cross sections for hundreds of tryptic digests with up to 15 residues in length have been reported [3]. The features observed in IMS experiments are typically assigned by calculating the mobilities for plausible candidate systems and comparing the results with the measurements. These calculations can be performed with reasonable accuracy using the methods developed over the last several years. In the order of increasing sophistication, they are the projection approximation [6], the exact hard-spheres scattering model [7], trajectory calculations [8, 9], and the scattering on electron density isosurfaces treatment, in its original form [10] or coupled with trajectory calculations [11]. All these techniques involve modeling the geometry of an ion and computing its mobility in a given gas by numerically solving the transport equations under a certain approximation for the interaction potential between the ion and buffer gas atoms. This

Address reprint requests to Dr. K. W. Michael Siu, Department of Chemistry, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada. E-mail: kwmsiu@yorku.ca

1044-0305/01/$20.00
PII S1044-0305(01)00269-0

Received January 16, 2001
Revised March 27, 2001
Accepted March 27, 2001

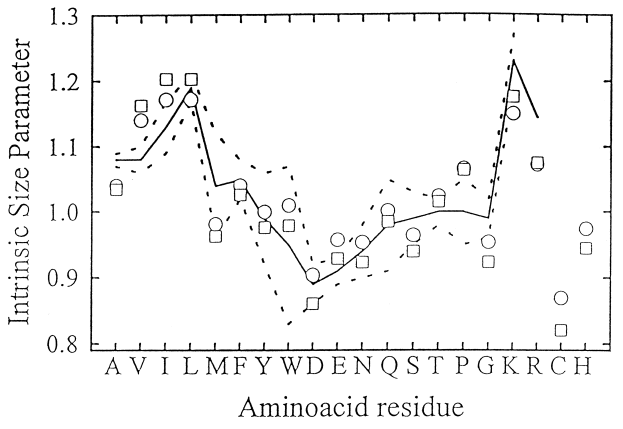

Figure 1. Fig. 1. Intrinsic size parameters of amino acid residues. Solid line represents the set produced by fitting the cross sections measured for 113 Lys-terminated peptides with 5–10 residues, the dashed lines delineating the error margin [3, 12]. The value for arginine was obtained from the set of 38 Arg-terminated peptides [12]. Symbols represent a priori values derived from the radii of atoms comprising the residue. Circles and squares represent the result from the assumptions of minimum and maximum plausible radii, respectively (see Appendix 1). Values for cysteine and histidine residues have not been determined experimentally [3, 12, 13].

first-principles approach reliably produces quality results; however, it is computationally intensive both for generating the needed trial geometries and for evaluating their mobilities. Thus, its large-scale employment in real-time analysis of IMS data for tryptic digests, which routinely contain dozens of peptides, appears at present to be impractical.

Clemmer and coworkers have proposed an alternative simple procedure to predict mobilities that is based on the ISPs of amino acid residues comprising the peptide [4, 12, 13]. In this treatment, the cross sections are calculated by multiplying a function of the molecular weight by the average of ISPs for all residues present. The ISPs have been derived empirically by a linear fit to the mobilities measured for a set of peptides containing a variety of residue types. Basically, an ISP is the indicator of how bulky a particular residue tends to be in a typical peptide environment. A limitation of this approach is that it ignores the peptide sequence: the calculated mobility depends only on the composition of residues. In reality, some sequence inversions can be separated using IMS [14, 15]. Nonetheless, the ISP-based formalism was shown to be superior to an earlier model where the cross section, regardless of the identity of the residues present, is a function of peptide weight only. Further, a substantial variation of the ISPs between residues (Figure 1) was interpreted to reflect the differences in the interaction of their side chains, specifically with regard to their solvation and packing properties. For example, low ISPs for polar residues were explained by their more pronounced solvation within a peptide, attributable to stronger long-range interactions between polar groups, polar groups and the peptide backbone, or polar groups and the ionic charge [4, 12, 13]. Low ISPs for aromatic residues were taken as an indication of efficient stacking of aromatic rings [13], and higher ISPs for residues with long side chains were regarded as a result of larger numbers of conformational degrees of freedom [4, 12]. Finally, a high ISP of lysine was presumed to be the result of a tryptic peptide located at the C-terminus [12].

Here we investigate whether the ISPs for amino acid residues can be derived a priori without using any experimental data, and how this would affect the peptide cross sections deduced from the ISPs of constituent residues. Indeed, assuming that all tryptic peptides are close-packed without major internal cavities and are near-spherical in shape, the volume (and thus the cross section) of a peptide would depend only on its mass and density. Different amino acids do not have the same proportion of light and heavy atoms, and thus have different densities. To see if this may be one cause of the variations in ISPs, a model was set up where the ISP of a residue is simply proportional to the sum of projection area contributions from all constituent atoms divided by the total mass:

$$\mathrm{ISP} \propto \sum \pi r_i^2 \Big/ \sum m_i \qquad (1)$$

Here r_i are the radii of atoms and m_i are their masses (see Appendix 1). The resulting values were then uniformly scaled to yield the best agreement with the measurements [3] for Lys-terminated peptide cations with 5–10 residues. The ISPs obtained in this way are compared with those derived experimentally in Figure 1: the two sets are virtually identical within the experimental and computational error margins combined (see Appendix 2). The latter arises from some freedom in choosing the atomic radii in the present model (Appendix 1).

We now test the capability of this model to predict the mobilities of peptides by comparing the cross sections resulting from a priori-calculated ISPs of the constituent amino acid residues with experimental data. This may be of analytical utility as a means to assign peaks in the ion mobility spectra of an unknown protein digest in proteomics, [3], (e.g., as demonstrated by Hoaglund-Hyzer et al. [16]). In Table 1, we revisit the 10 Lys-terminated tryptic peptides from sperm whale myoglobin and yeast enolase [3]. These data have been used by Valentine et al. [3] to show that their ISP-based model was superior to a weight-based fit. Our present a priori-derived ISPs match the measurements even better, despite the fact that ISPs in reference [3] have been fit to specifically reproduce the mobilities of Lys-terminated peptides with 5–10 residues. While this performance is impressive, a sample of just 10 data points does not suffice to assess the applicability of proposed model to peptides in general. So we have analyzed the data for all 163 singly-charged Lys-terminated peptide ions with 5–15 residues [3]. This set includes the 10 species listed in Table 1. However, it excludes all

Table 1. Cross sections for tryptic peptides from sperm whale myoglobin and yeast enolase: predictions by several different methods versus experiment.

	Cross section, Å^2				
		Calculated			
Peptide Sequence	Measured[3]	Weight[a]	Fitting[b]	Apriori[c]	Apriori[d]
DIAAK	155.37	152.48	156.63	153.85	154.33
TGQIK	157.92	158.17	158.60	160.49	161.27
AAAAEK	159.57	160.89	164.58	159.33	158.97
ANIDVK	178.62	179.86	177.84	180.48	180.72
ASEDLK	175.16	180.42	177.37	176.56	175.61
TEAEMK	181.27	189.04	186.57	181.86	180.08
NVPLYK	195.63	193.68	191.80	196.87	197.43
YDLDFK	198.72	205.90	201.58	200.48	198.44
IATAIEK	203.73	195.89	200.66	200.62	202.04
AADALLLK	221.35	208.44	218.01	215.55	217.58
Standard deviation δ, %		3.1	1.8	1.3	1.1
Average |dev.|,[e] %		2.5	1.5	1.1	0.9
Number of points deviating by >2% (>3%)		5 (5)	2 (1)	1 (0)	1 (0)
Maximum |dev.|,[e] %		5.8	3.1	2.6	2.1

[a]Produced by the quadratic polynomial fit to measurements based on the molecular weight only.[3]
[b]Obtained using the intrinsic size parameters derived by fitting the mobilities of 113 lysine-terminated peptide cations with 5–10 residues.[3]
[c]Calculated using the standard van der Waals atomic radii (Appendix 1).
[d]Calculated using the largest reasonable radii (Appendix 1).
[e]The abbreviation |dev.| stands for the modulus of relative deviation.

peptides containing arginine, cysteine, or histidine residues because their ISPs have not been obtained from the fit to experimental data. Thus, our a priori-derived mobilities could not have been compared to earlier work. For this set, our a priori approach yields a standard relative deviation (δ) of 2.2% (calculated using the standard Van der Waals radii in Appendix 1). This matches the value resulting from calculations using the experimentally fit ISPs [3] and is better than that of the weight-based fit (δ = 3.1%). Employing another measure of quality, 29 out of 163 points (18%) derived a priori are outside of a 3% error margin. This is virtually equal to 28 points using the experimental ISPs [3] and is only half of the 56 points (34%) in the fit by molecular weight only. For a truly unbiased evaluation, we should apply the present treatment to peptides containing a residue for which no ISP has been derived experimentally. The mobilities measured in reference [3] include the data for 46 singly charged histidine-containing peptides with 5–15 residues; however, the ISP of the His residue has not been obtained. Using the a priori value, we calculated the mobilities and found that only six (13%) deviate from the experimental points by more than 3%. By comparison, this happens for 21 points (46%) in the weight-based fit. The value of δ has decreased from 3.7% to 2.4%. It is important to determine the performance of the proposed model for non-Lys terminated peptides. The only other large class of peptides for which the mobilities have been measured are Arg-terminated species [3]. Out of 69 Arg-terminated singly charged peptides with 5–15 residues, the mobilities computed by the present method are outside of a 3% margin in 15 cases (22%), or less than half of the 35 points (51%) in the fit by weight. Here, the use of calculated ISPs reduces the value of δ from the weight-based procedure of 3.6% to 2.5%. Overall, the mobilities for 271 singly-charged peptides (either Lys- or Arg-terminated) have been measured. In this case, the a priori model described has δ = 2.3% and fails to place 49 peptides (18%) within the 3% margin. Again, this is less than half of the 106 points (39%) in the weight-based fit with δ = 3.3%. The deviations for all sets would still be ~1%, even if the calculations had been exact, because of experimental error inherent in mobility measurements. Thus a systematic reduction of δ from 3–4% in the weight-based fits to 2–2.5% obtained using our model actually corresponds to halving the standard deviation attributable to mobility computation from 2–3% down to 1–1.5%.

In summary, a priori-calculated intrinsic size parameters of the constituent amino acid residues can be used to consistently predict the mobilities of peptides more accurately than on the basis of molecular weight only, and in some cases, even more accurately than by retrofitting the parameters derived from measurements. The capability to predict the size parameter of any amino acid (or other peptide constituent) a priori may be of particular value for argentinated and other metalated peptides [17–20] used in peptide sequencing [17–19], and peptides involving non-standard residues that result from post-translational modifications [21]. As those reagents or complex atoms or additional residues are not encountered frequently, the determination of their ISPs by fitting experimental data would be challenging.

A more fundamental aspect of this work is the implication of the above findings to our understanding

of peptide folding and self-solvation, if only in the gas phase. In fact, if the variations in the measured ISPs of amino acid residues can be explained by their different densities (Figure 1), such variations cannot be due to putative differences in the interaction between residues (such as peptide contraction by long-range charge-dipole and dipole-dipole interactions of polar groups or efficient stacking of aromatic rings [3, 12, 13]. Instead, the polar and aromatic residues may simply be denser because of higher ratios of heavy atoms to hydrogen. Likewise, on the basis of its comparatively low ISP, methionine, which is normally characterized as a non-polar aliphatic residue in solution, has been assumed to display polar characteristics in the gas phase [12]. This was ascribed to the relatively high polarizability of the sulphur atom, inducing stronger long-range interactions which ultimately contract the peptide. However, the low methionine ISP is accounted for here by the high density of the S atom without relevance to the packing of the rest of the peptide. Similarly, high ISPs of lysine and other residues with long side chains now appear unrelated to the lysine position at the peptide end or the greater conformational freedom of longer side chains. Instead, they may be caused by low ratios of heavy atoms to hydrogen, which results in low density. In conclusion, our results strongly suggest that folding of tryptic peptides in the gas phase is largely independent of the chemical nature of constituent residues. These peptides apparently assume compact conformations dominated by the solvation of ionic charge. Thus, the model we describe for calculating peptide mobilities can also be used as a rapid method to screen substantial deviations from such compact conformations.

Acknowledgments

We thank Anne Counterman for useful discussion. This research was supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix 1

The radii of atoms for the purpose of eq 1 can be chosen somewhat arbitrarily within a reasonable range. For the standard model, we have adopted the Van der Waals values of 1.1 Å for hydrogen, 1.6 Å for carbon, nitrogen, and oxygen, and 2.0 Å for sulphur. To test the dependence of the findings on the radii assumed, we recalculated the ISPs with several different sets of values and found the impact to be insignificant. In particular, we tested the maximum reasonable radii obtained by adding the collision radius of He atom (assumed to equal 1.1 Å) to all the Van der Waals radii above (Figure 1).

Appendix 2

We also considered a model where the ISPs are defined a little differently, as

$$\mathrm{ISP} \propto \pi \left(\sum r_i^3 \right)^{2/3} \Big/ \sum m_i \qquad (2)$$

This more closely emulates the physical reality of a near-spherical, densely-packed ball of atoms, by first calculating the total volume and then deriving the corresponding cross section. However, the ISP values obtained using eq 1 and eq 2 are very close, as both formalisms qualitatively introduce the same factor—the unequal densities of different atoms that result in the unequal densities of amino acids with different composition. In the rest of this work, eq 1 has been adopted.

References

1. McLafferty, F. W.; Fridriksson, E. K.; Horn, D. M.; Lewis, M. A.; Zubarev, R. A. *Science* **1999,** *284,* 1289.
2. Beavis, R. C.; Chait, B. T. *Proc. Natl. Acad. Sci. USA* **1990,** *87,* 6873.
3. Valentine, S. J.; Counterman, A. E.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **1999,** *10,* 1188.
4. Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997,** *32,* 577.
5. Hoaglund, C. S.; Valentine, S. J.; Sporleder, C. R.; Reilly, J. P.; Clemmer, D. E. *Anal. Chem.* **1998,** *70,* 2236.
6. Von Helden, G.; Hsu, M. T.; Gotts, N. G.; Bowers, M. T. *J. Phys. Chem.* **1993,** *97,* 8182.
7. Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996,** *261,* 86.
8. Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996,** *100,* 16082.
9. Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Chem. Phys.* **1998,** *108,* 2416.
10. Shvartsburg, A. A.; Liu, B.; Jarrold, M. F.; Ho, K. M. *J. Chem. Phys.* **2000,** *112,* 4517.
11. Shvartsburg, A. A.; Liu, B.; Siu, K. W. M.; Ho, K. M. *J. Phys. Chem.* **2000,** *104,* 6152.
12. Valentine, S. J.; Counterman, A. E.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. *J. Phys. Chem. B* **1999,** *103,* 1203.
13. Henderson, S. C.; Li, J.; Counterman, A. E.; Clemmer, D. E. *J. Phys. Chem. B* **1999,** *103,* 8780.
14. Wu, C.; Siems, W. F.; Klasmeier, J.; Hill, H. H. *Anal. Chem.* **2000,** *72,* 391.
15. Srebalus, C. A.; Li, J.; Marshall, W. S.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2000,** *11,* 352.
16. Hoaglund-Hyzer, C. S.; Li, J.; Clemmer, D. E. *Anal. Chem.* **2000,** *72,* 2737.
17. Lee, V. W.-M.; Li, H.; Lau, T.-C.; Siu, K. W. M. *J. Am. Chem. Soc.* **1998,** *120,* 7302.
18. Chu, I. K.; Guo, X.; Lau, T.-C.; Siu, K. W. M. *Anal. Chem.* **1999,** *71,* 2364.
19. Rodriquez, C. F.; Fournier, R.; Chu, I. K.; Hopkinson, A. C.; Siu, K. W. M. *Int. J. Mass Spectrom. Ion Processes* **1999,** *192,* 303.
20. Taraszka, J. A.; Li, J.; Clemmer, D. E. *J. Phys. Chem. B* **2000,** *104,* 4545.
21. Wieland, T.; Bodanszky, M. *The world of peptides*; Springer-Verlag: Berlin, 1991.