

Optimization of Algorithms for Ion Mobility Calculations

Alexandre A. Shvartsburg,^{*,†} Stefan V. Mashkevich,[‡] Erin Shammel Baker,[†] and Richard D. Smith[†]

Biological Sciences Division, Pacific Northwest National Laboratory, P. O. Box 999, Richland, Washington 99352, and Schrödinger, 120 West 45th Street, New York, New York 10036-4041

Received: October 23, 2006; In Final Form: January 4, 2007

Ion mobility spectrometry (IMS) is increasingly employed to probe the structures of gas-phase ions, particularly those of proteins and other biological macromolecules. This process involves comparing measured mobilities to those computed for potential geometries, which requires evaluation of orientationally averaged cross sections using some approximate treatment of ion–buffer gas collisions. Two common models are the projection approximation (PA) and exact hard-spheres scattering (EHSS) that represent ions as collections of hard spheres. Though calculations for large ions and/or conformer ensembles take significant time, no algorithmic optimization had been explored. Previous EHSS programs were dominated by ion rotation operations that allow orientational averaging. We have developed two new algorithms for PA and EHSS calculations: one simplifies those operations and greatly reduces their number, and the other disposes of them altogether by propagating trajectories from a random origin. The new algorithms were tested for a representative set of seven ion geometries including diverse sizes and shapes. While the best choice depends on the geometry in a nonobvious way, the difference between the two codes is generally modest. Both are much more efficient than the existing software, for example faster than the widely used Mobcal (implementing EHSS) ~ 10 – 30 -fold.

Introduction

An increasingly common approach to structural characterization of gas-phase ions is ion mobility spectrometry (IMS), usually in conjunction with mass spectrometry (MS).^{1–22} While early work focused on atomic clusters,^{1–8} coupling IMS to electrospray (ESI) and matrix-assisted laser desorption ionization (MALDI) sources a decade ago^{9,10} enabled probing conformations of macromolecular ions such as peptides,^{11–14} proteins and their complexes,^{9,15–18} oligonucleotides,^{19,20} saccharides and their adducts,^{21,22} catenanes,²³ and organic polymers.^{24,25} Structural elucidation using IMS involves collecting plausible candidate geometries (for example, using optimization, homology, known solution or solid-state structures, or insight from other gas-phase studies), computing their mobilities (K), and comparing the results with experimental data. IMS experiments are normally performed in the low-field regime, where K is independent of the electric field E and determined by the orientationally averaged cross-section $\Omega_{av}^{(1,1)}$ (or simply Ω) between the ion (of mass m) and buffer gas molecule (of mass M):²⁶

$$K = [3ze/(16N)][2\pi(m + M)/(mMk_B T)]^{1/2}/\Omega \quad (1)$$

where z is the ion charge state, e is the elementary charge, N is the gas number density, k_B is the Boltzmann constant, and T is the gas temperature. The values of N and T can be readily measured within $<0.3\%$ and calibrated for yet higher accuracy, so the uncertainty of calculated K normally follows from the evaluation of Ω .

The specificity of structural characterization using IMS is set by combined errors of experiment and mobility calculation. A

limited IMS resolving power ($R \sim 10$ – 15) in early IMS/MS instruments of the injected drift tube design^{1,9–11,15,27} meant low measurement accuracy, which permitted deriving mobilities from a simple projection approximation (PA) that replaces the cross-section of an ion by its projection (shadow).²⁷ The increase of R to >100 in relatively high-pressure IMS systems (coupled to MS)²⁸ has improved the measurement accuracy to $\sim 1\%$ and better for relative values,^{4,29,30} stimulating the development of more sophisticated methods that compute Ω through simulation of collisional events on an approximate energy surface of ion–buffer gas interactions. The simplest such model is the exact hard-spheres scattering (EHSS) that assumes an infinite hard wall potential between each atom of the ion and gas molecule.³¹ This approach accounts for multiple scattering during an ion–molecule collision, which is ubiquitous for ions with concave or rough surfaces (such as those found for macromolecules) and may increase Ω by over^{15,31} 20%. Proper consideration of this effect has proven critical for the correct assignment of IMS data, for example, for fullerene dimers,^{4,29} semiconductor nanoclusters,³² or proteins such as BPTI or ubiquitin.¹⁵ In reality, scattering occurs on the frontier electronic orbitals of the ion, and Ω is controlled by the position of the electron cloud rather than that of atomic nuclei. This issue is much more important for anions, which have significantly more extended electronic orbitals than cations.³³ This increase of Ω due to electron spillover is quantifiable using the scattering on electron density isosurfaces (SEDI)³³—a version of EHSS where an ion is represented by a hard body of arbitrary shape, numerically defined on a 3-D grid by selecting points with a certain electron density.

While the description of repulsive ion–molecule interactions in EHSS and SEDI is reasonable for mobility calculation purposes, the attractive charge-induced dipole and dispersion force interactions are ignored. Accounting for the attractive

[†] Pacific Northwest National Laboratory.

[‡] Schrödinger.

potential (for example, assuming pairwise 6–12 interactions between all ion atoms and buffer gas molecule) improves the agreement with IMS measurements significantly, especially when comparing ions with grossly dissimilar sizes or shapes.^{34,35} This correction may be added to that due to electron spillover, creating the highest-level method known for calculating mobilities of polyatomic ions.³⁶ However, these approaches require rigorous, computationally expensive molecular dynamics (MD) simulations. The cost scales as $\sim n^{4/3}$, where n is the number of atoms in the ion, precluding routine application to large species. Further, proteins and most other macroions are flexible, featuring numerous low-energy conformers separated by modest barriers. Hence, under realistic IMS conditions, macroions exist not as a single (global minimum) geometry but as an ensemble of isomers,^{17,37–39} many not interconverting during IMS analyses as evidenced by the peak widths in IMS spectra^{17,37,38} and confirmed in IMS/IMS experiments.³⁹ Thus mobilities must be computed for a representative set of low-energy geometries obtained using molecular mechanics,^{11,40–45} often considering competing protonation (charging) schemes.^{42,44} A prohibitive cost of doing that in conjunction with simulations of ion–buffer gas scattering on realistic potentials has motivated efforts to account for attractive interactions by empiric schemes based on EHSS results and other parameters of ion geometry.^{40–45} Similar parametrizations were proposed for the dependence of ion mobilities on gas temperature.⁴⁶

A single PA or EHSS calculation is not expensive unless for macroions such as large proteins or DNA, but hundreds of calculations for a conformer ensemble can be a real endeavor. Performing SEDI is extremely costly even for a single geometry of moderate size,^{33,36} because defining an ion surface with sufficient accuracy calls for a grid not sparser than ~ 0.1 Å, so the number of anchor points exceeds that of constituent atoms by $\sim 10^3$ times. Hence a major acceleration of EHSS and (especially) SEDI would substantially help structural characterization using IMS, particularly for large biological ions. While much effort has been put into improving the accuracy of ion mobility calculations, their procedures have not been optimized for any ion–buffer gas interaction model. Here we describe new algorithms that, thanks to three fundamental changes, speed up the implementation of PA and EHSS by over an order of magnitude.

Computational Methods

Previous Approaches. At weak electric fields in IMS, ions (except some extremely large species with dipole moments over $\sim 10^4$ D)⁴⁷ experience free rotation. Hence, mobility calculations involve averaging of the ion–buffer gas cross-sections over all collision geometries. In previous methods including the PA,²⁷ EHSS,³¹ and MD calculations,³⁴ the orientational averaging was described as

$$\Omega = \frac{1}{8\pi^2} \int_0^{2\pi} d\theta \int_0^\pi d\varphi \int_0^{2\pi} d\gamma \Omega_p(\theta, \varphi, \gamma) \quad (2)$$

where θ , φ , and γ are spatial angles defining the ion orientation and Ω_p is the corresponding partial cross-section that depends on the model for ion–buffer gas interaction. In PA:²⁷

$$\Omega_p(\theta, \varphi, \gamma) = 2\pi \int_0^\infty bM(\theta, \varphi, \gamma, b) db \quad (3)$$

where b is the impact parameter and M is unity when a hard-sphere collision occurs for the configuration defined by θ , φ , γ , and b and null otherwise. In EHSS (or SEDI)³¹

$$\Omega_p(\theta, \varphi, \gamma) = 2\pi \int_0^\infty b[1 - \cos \chi(\theta, \varphi, \gamma, b)] db \quad (4)$$

where χ is the scattering angle for defined configuration.

However, in implementation, the integrand was defined not via a radial coordinate b , but via Cartesian coordinates inside a rectangle (with sides along y and z) drawn around the projection of an ion onto the yz plane.^{27,31,34} Averaging over those coordinates produces a directional cross-section Ω_{dir} , with eq 3 replaced by

$$\Omega_{\text{dir}}(\varphi, \gamma) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} M(\varphi, \gamma, y, z) dy dz \quad (5)$$

and eq 4 by

$$\Omega_{\text{dir}}(\varphi, \gamma) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [1 - \cos \chi(\varphi, \gamma, y, z)] dy dz \quad (6)$$

Equation 5 or eq 6 was convoluted with eq 2 using Monte Carlo integration, by shooting a certain number (m) of gas molecules at the ion along a selected axis (for example, x) and counting hits (in PA)²⁷ or saving χ after following trajectories through any and all collisions with the target until they leave it for good (in EHSS³¹ and SEDI³³). In the publicly released code Mobcal,⁴⁸ the ion is rotated around all three axes by random θ , φ , and γ before each shot. That software is broadly used for EHSS and PA calculations and is adopted as a benchmark in this work. In the code Sigma⁴⁹ that implements PA only, the ion is rotated after a series of shots.

Present Development. As mentioned above, all mobility calculations comprised two procedures: (I) trajectory propagation (to evaluate Ω_{dir}) and (II) target rotation (to average Ω_{dir} over ion orientations). In both PA and EHSS (SEDI), I requires checking for possible collisions with all atoms (surface points) of the ion, which scales with n linearly. In EHSS (SEDI), the expense increases for rough/concave objects because multiple reflections must be evaluated, but even in extreme cases most trajectories reflect only once. For example, for native ubiquitin ions in He where multiple scattering in EHSS increases Ω by 24% over the PA value, just 25% of trajectories undergo two reflections and 8% experience three or more. Thus the effect of ion shape on the expense of I is small, and its scaling with n remains close to linear. Step II means recalculating the Cartesian coordinates of all atoms (surface points). The cost of that scales as the number of coordinates ($3n$) and thus is also proportional to n , but the coefficient usually exceeds that in I by a large factor (R_{II}) because of laborious trigonometry involved. From the individual timing of steps I and II, typical R_{II} values are ~ 100 – 300 for PA and ~ 25 – 60 for EHSS. (Modeling multiple scattering in EHSS is obviously more expensive than counting hits in PA, while the cost of ion rotation in the two cases is equal.) Therefore, when an ion is rotated after each shot as in Mobcal, the typical fraction of calculation spent on step II is $>99\%$ for PA and ~ 96 – 98% for EHSS. Hence, the primary path to optimize those calculations is cutting the number of rotational steps needed for a given accuracy of Ω .

Expressions 5 and 6 already incorporate the rotation around x defined by θ : the integration over $dy dz$ replaces that over db as

$$\Omega_{\text{dir}}(\varphi, \gamma) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \Omega_p(\theta, \varphi, \gamma) \quad (7)$$

In simple words, an object rotated around the line of sight conserves its projection area and cross-section, regardless of

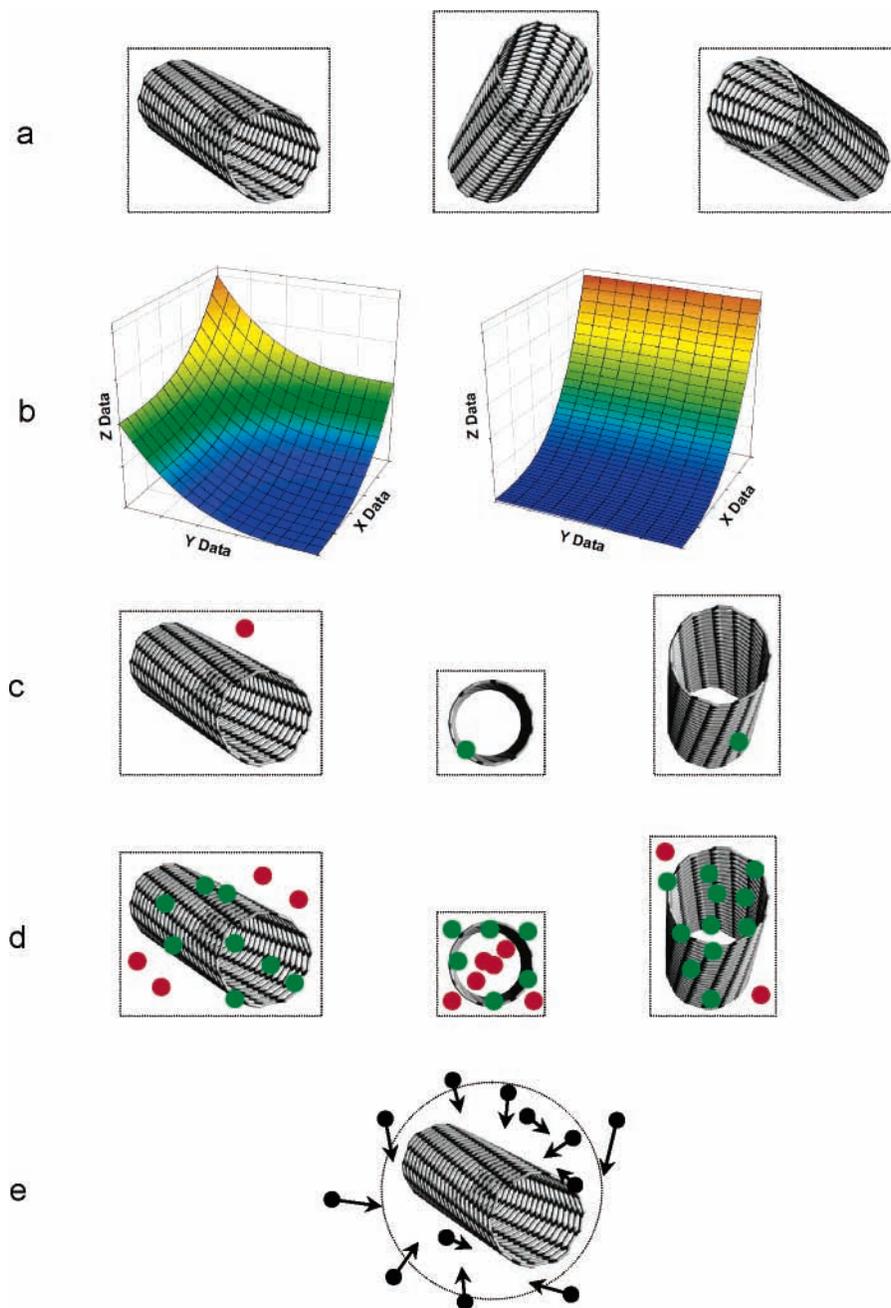


Figure 1. New algorithms for orientational averaging in ion mobility calculations. (a) Rotating an object around the line of sight does not affect its projection or cross-section. (b) Optimum numerical integration of a $z(x,y)$ function involves a uniform sampling of x and y when evaluating z for new x and y has equal costs (e.g., $z = x^3 + y^3$, left panel), but a preferential sampling of x when evaluating z for a new x is significantly cheaper (e.g., $z = x^3 + \{\ln[(\cos \sum_{k=1,12} y^k)^2]\}^2$, right panel). (c) PA2 and EHSS2: The ion is rotated around the two axes orthogonal to line of sight after a single shot (green circles for hits and red ones for misses). (d) PA2/ k and EHSS2/ k : The ion is rotated as in c after a round of 12 shots. (e) PA0 and EHSS0: shooting from a random direction obviates the need for ion rotation but requires aiming inside a circle rather than a rectangle.

the ion–buffer gas potential (Figure 1a). Then eq 2 may be condensed to

$$\Omega = \frac{1}{4\pi} \int_0^\pi d\varphi \sin \varphi \int_0^{2\pi} d\gamma \Omega_{\text{dir}}(\varphi, \gamma) \quad (8)$$

We have implemented eq 8 by rotating ions in PA or EHSS by random angles around two axes instead of three in existing codes. This should cut the expense of each step II by \sim one-third and thus accelerate the overall calculation almost in proportion. We may improve on that by precalculating certain quantities involved in rotating an ion (such as the distance of each atom from the rotational axes) once for all rotation operations. This could be done for either 3-D or 2-D rotation

but is more valuable for the latter because a greater fraction of variables may be precalculated. We termed the PA and EHSS algorithms using 2-D rotation with that shortcut PA2 and EHSS2, with the previous algorithms renamed PA3 and EHSS3 (for 3-D rotation).

Greater gains require reducing the number of steps II relative to that of I. Variables in a multidimensional Monte Carlo integration do not have to be sampled with equal density: one may compensate for a sparser sampling in some dimension(s) by denser sampling in the other(s) and keep the statistical error constant. When reevaluating the integrand takes equal time whichever variable changes, a uniform sampling density across dimensions is optimum. If fixing some variable(s) accelerates

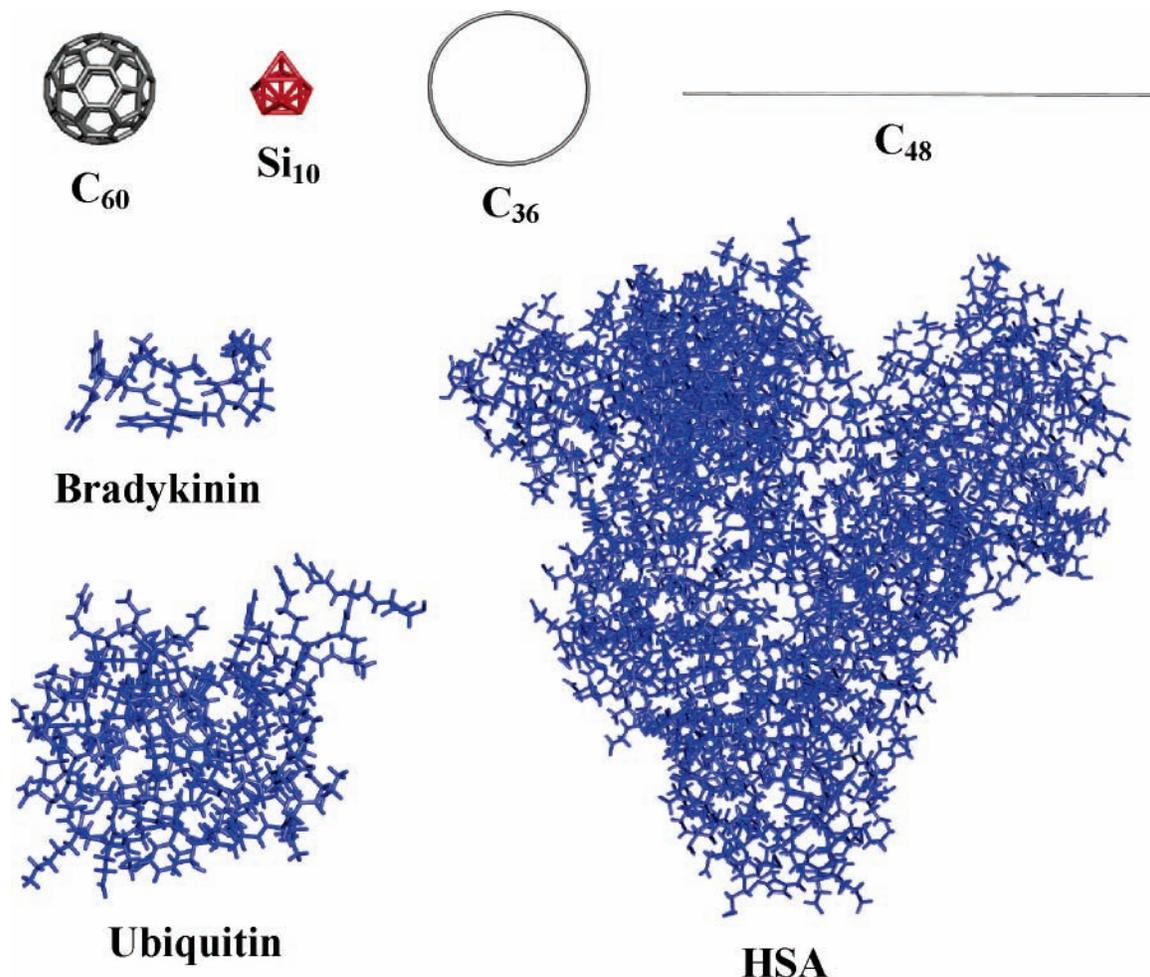


Figure 2. Ion geometries used to benchmark mobility calculation methods.

reevaluation more than fixing others, the above preferential sampling makes sense (Figure 1b). That is the case here: holding $\{\varphi; \gamma\}$ when varying $\{y; z\}$ reduces the cost of computing Ω_{dir} by a factor of $(R_{\text{II}} + 1)$, while the opposite makes no significant difference. Then we would benefit from performing multiple steps I between II—that is, shoot rounds of certain length k rather than single shots³¹ at the target ion in any particular orientation (Figure 1c,d). That is possible with either 3-D or 2-D rotation but is implemented here with the latter because accelerating steps II is always advantageous no matter the number of I in between. The resulting algorithms are termed PA2/ k and EHSS2/ k , for example, EHSS2/5 for 5 shots per orientation. As mentioned above, a similar approach for PA was employed in Sigma.⁴⁹ However, ions were rotated around three axes using eq 2, and the value of k was not optimized but in effect set to $\sim 10^4$ by requiring the statistical error of $< 1\%$ for each Ω_{dir} .²⁷

Instead of randomly turning an ion, one may aim at it from a random direction with a random impact parameter (Figure 1e), which we implement in codes called PA0 and EHSS0. This requires shooting into a sphere enclosing the ion rather than into rectangles defined by extremities of the yz projection in a given orientation (Figure 1c,d). That facet, the need to produce two additional random numbers defining the initial direction of trajectory, and computational aspects of guiding it starting from an arbitrary direction increase the cost of steps I. Also, for a significantly nonspherical ion, the projection of a surrounding sphere is larger than the areas of rectangles containing the ion projections in most or all orientations. Hence a higher fraction of shots is missed, so a greater number of shots

is needed for a certain accuracy of Ω . Therefore, while PA0 and EHSS0 algorithms avoid the expensive steps II, they may or may not be superior to PA2/ k and EHSS2/ k .

The EHSS2, EHSS2/ k , and EHSS0 algorithms and their PA analogs were coded in Fortran. The PA and EHSS programs were benchmarked versus our implementations of PA3 and EHSS3 that were identical except for the algorithmic differences described. That has allowed an accurate characterization of the merits of new algorithms apart from software-level differences between the present EHSS3 and its version in Mobcal⁴⁸ or between the present PA codes and the PA implemented in Sigma.⁴⁹ However, a comparison between EHSS3 and Mobcal shows the improvement of present codes over the previously available software. All tests were run using a PC with a 3.0 GHz Intel Xeon Processor (Dell Computer) in a Windows XP environment, trying several Fortran compiler options as described below.

Values of Ω (in He gas at 300 K) were calculated for a set of ions covering a broad range of sizes and shapes (Figure 2). Various shapes are represented by a spherical C₆₀ fullerene, a compact Si₁₀ cluster (a tricapped trigonal prism),⁶ an extremely prolate C₄₈ chain, and an oblate C₃₆ ring. (A straight chain and a circular ring were constructed assuming cumulenic C=C bonds of 1.29 Å length.)³⁵ Biological ions that span a wide range of sizes and have rough surfaces are exemplified by a low-energy zwitterion conformer¹⁴ of bradykinin 2+ (151 atoms) and native geometries of bovine ubiquitin (1232 atoms) and human serum albumin (HSA, 9136 atoms) from the RCSB data bank.⁵⁰ For

TABLE 1: Benchmarks of Codes Implementing EHSS and PA Models (Execution Times and Relative Standard Errors of Result Per 10^6 Total Shots)^a

metric	method	C ₆₀	Si ₁₀	C ₄₈ chain	C ₃₆ ring	bradykinin	ubiquitin	HSA
t, s	EHSS3 (Mobcal) ⁴⁸	129	25	102	77	318	2560	19100
	EHSS3 (present)	127	22	99	72	312	2550	19000
	EHSS3 (-O)	93	16.6	75	54	235	1910	14300
	EHSS2 (-O)	50	9.1	41	31	127	1030	7800
	PA2 (-O)	50	8.8	40	30	119	1000	7500
	EHSS0 (-O)	6.3	1.6	2.4	2.9	9.1	66	621
	PA0 (-O)	0.9	0.74	2.2	1.4	4.2	31	211
$\sigma_r \times 10^3$	EHSS3 or EHSS2	0.8	0.8	2.0	1.1	0.9	0.9	1.0
	EHSS0	0.6	0.9	3.9	1.2	1.4	1.5	1.2
	PA3 or PA2	0.5	0.6	1.6	0.8	0.7	0.7	0.8
	PA0	0.2	0.6	3.4	0.9	1.1	1.3	1.1

^a Where indicated by the -O flag, level 2 compiler optimization was used.

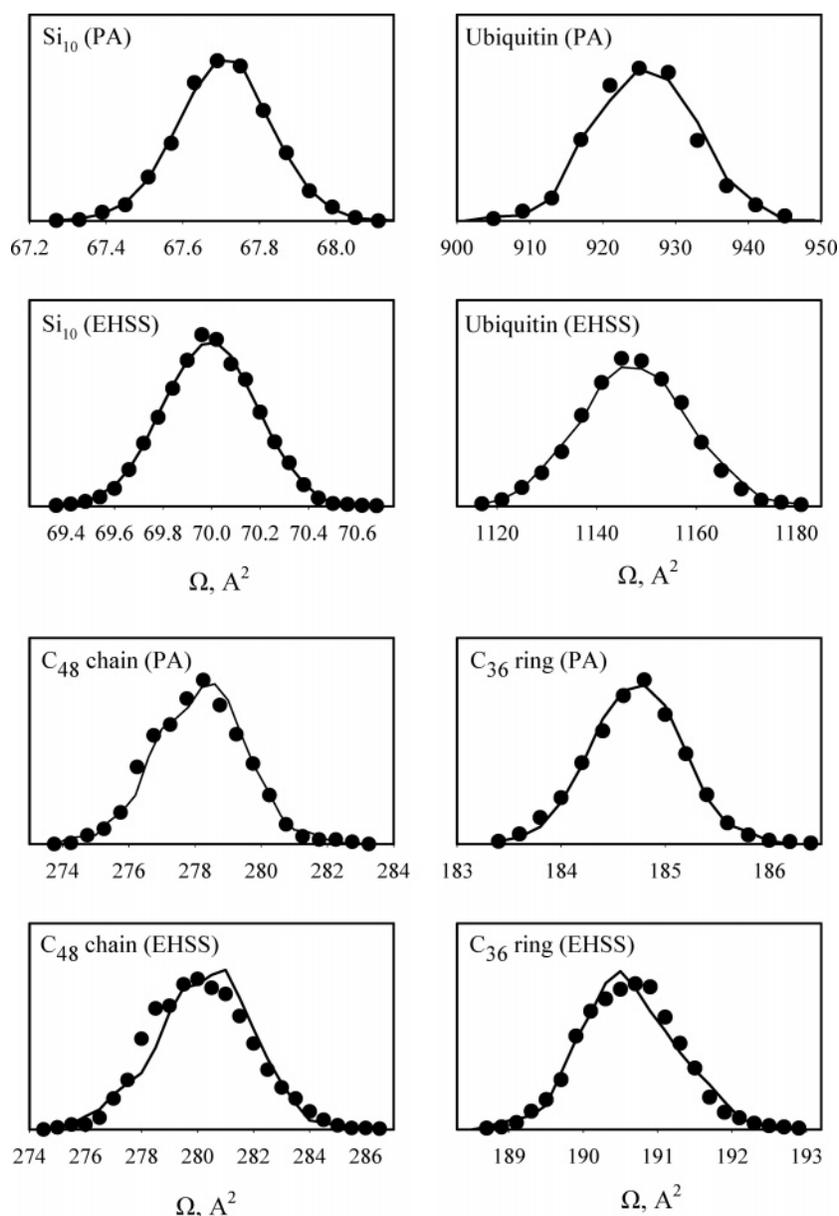


Figure 3. Distributions of Ω computed for four representative geometries using PA and EHSS (as labeled) with 3-D rotation (line) and 2-D rotation (circles).

the atomic collision radii, we used 1.0 (He), 1.2 (H), and 1.7 Å (other atoms) that produced reasonable Ω values.⁴⁸ The accuracy of ion geometries or collision radii is immaterial for this study.

For each ion, we accumulated extensive statistics of Ω by repeating the calculation (each for $m = (4-100) \times 10^3$, depending on n) 1000 times with different random number

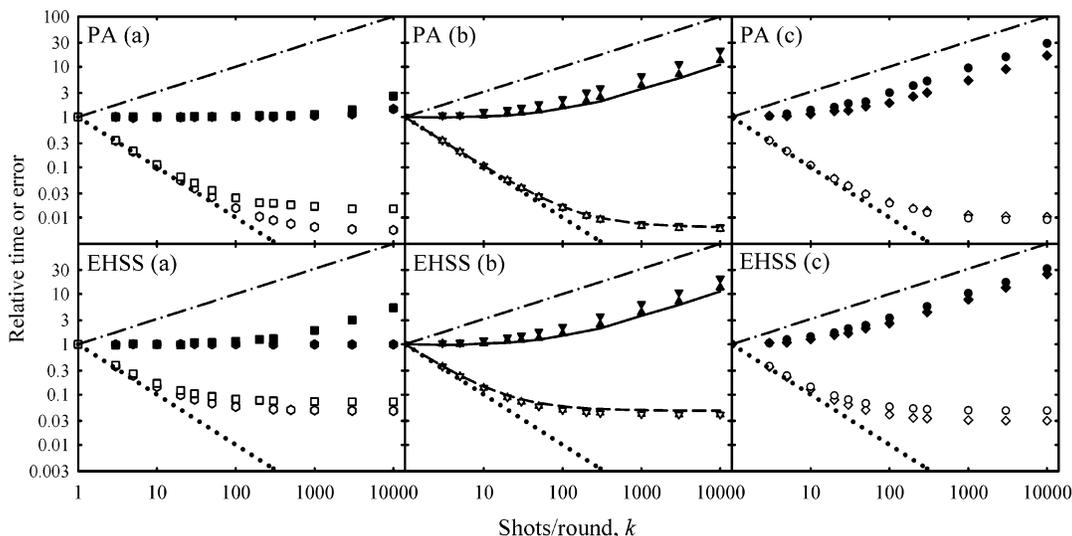


Figure 4. Characteristics of PA2/ k and EHSS2/ k calculations: execution time (empty symbols or dashed line) and statistical error of results (filled symbols and solid line) normalized such that the values for $k = 1$ equal 1. Data are for C_{60} fullerene (hexagons) and Si_{10} (squares) in a; ubiquitin (triangles up), bradykinin (triangles down), and HSA (lines) in b; and C_{48} chain (diamonds) and C_{36} ring (circles) in c. The absolute possible minimum of t , $t(k) = t(1)/k$, and maximum of σ_r , $\sigma_r(k) = \sigma_r(1)\sqrt{k}$, are marked by dotted and dashed-dotted lines, respectively.

sequences. From those data, we extracted the mean relative standard errors of Ω (termed σ_r) and average execution times t , both expressed per $m = 10^6$.

Results

Evaluation of EHSS3. The EHSS3 code is algorithmically identical to the EHSS implemented in Mobcal⁴⁸ and thus produces the same Ω , but features minor software-level shortcuts that improve compilation flexibility and accelerate execution by a few percent (Table 1). The level 2 compiler optimization (-O2 or -O option) accelerated EHSS3 for all geometries by 25–30% (Table 1). Deeper optimizations (-O3 and -O4 options) provided no further improvement, and all codes below were tested using -O. Optimization of Mobcal at any level has failed; hence, the speed gain upon optimization of EHSS3 adds to its advantage over Mobcal. Perhaps Mobcal could be optimized using other compilers, producing similar speed gains. The key point is that EHSS3 (used as a benchmark below) is at least as fast as Mobcal, with or without compiler optimization.

Evaluation of PA2 and EHSS2. Since averaging of $\Omega_{\text{dir}}(x,y)$ over θ is redundant, eqs 2 and 8 must produce the same result. That has been verified by comparing the statistics of Ω calculated using those equations with equal m for four different geometries, including the least symmetric prolate and oblate shapes where Ω is most sensitive to imperfect orientational averaging (Figure 3). In both PA and EHSS, the distributions of Ω obtained using eq 2 and eq 8 have proper Gaussian shapes and are identical in all cases: replacing 3-D rotation by 2-D has no effect on Ω or σ_r . However, the execution speed almost doubles, with a $\sim 50\%$ gain coming from the change to 2-D rotation (as anticipated) and another $\sim 25\%$ from precalculation of needed quantities (Table 1). Notably, σ_r is always greater for EHSS than for PA with the same number of shots, presumably reflecting the accumulation of statistical error during multiple scattering in EHSS.

Evaluation of PA2/ k and EHSS2/ k . From the definition of R_{II} , the cost of those algorithms depends on k as

$$t(k) \approx t(1)(1 + R_{II}/k)/(1 + R_{II}) \quad (9)$$

While the transition to 2-D rotation has reduced R_{II} by a factor

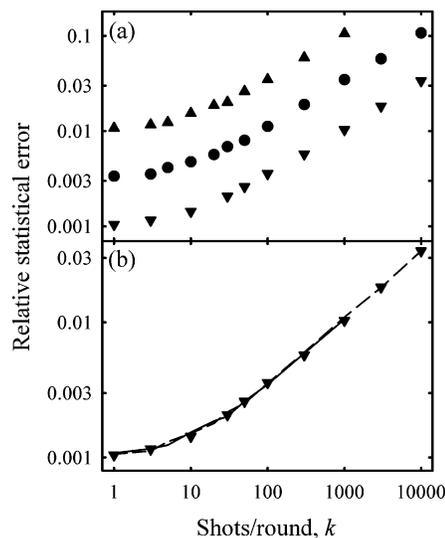


Figure 5. (a) Relative statistical error (σ_r) of EHSS2/ k calculations for C_{36} ring with $m = 10^4$ (triangles up), 10^5 (circles), and 10^6 (triangles down); (b) values of $\sigma_r\sqrt{m/10^6}$ for $m = 10^4$ (solid line), 10^5 (dashed line), and 10^6 (triangles).

of ~ 1.8 , steps II still take $>98\%$ of the total time for PA2 and $\sim 93\text{--}97\%$ for EHSS2. Hence, by eq 9, increasing k should initially accelerate the codes nearly in proportion, especially for PA. That is indeed the case; for example $k = 10$ produces a speed gain of 9–10 times for PA2/ k and 6–8 times for EHSS2/ k (Figure 4). As rotations take a smaller fraction of t , further increase of k becomes less effective and t stabilizes when $R_{II}/k \sim 0.01$, which happens at $k \sim 10^4$ for PA2/ k and $\sim 10^3$ for EHSS2/ k (Figure 4).

Of course, increasing k raises the statistical error of Ω for a given m because the number of sampled orientations (l) equals m/k . While $t(k)$ depends on the ion geometry weakly, $\sigma_r(k)$ is strongly influenced by it (Figure 4). For near-spherical ions, values of Ω_{dir} for all orientations are close, and σ_r does not increase much at higher k . For example, Ω of C_{60} calculated using either PA or EHSS by averaging Ω_{dir} over just ~ 100 orientations ($m = 10^6$, $k = 10^4$) are as accurate as those for $l = 10^6$ ($m = 10^6$, $k = 1$), Figure 4a. For rotationally asymmetric

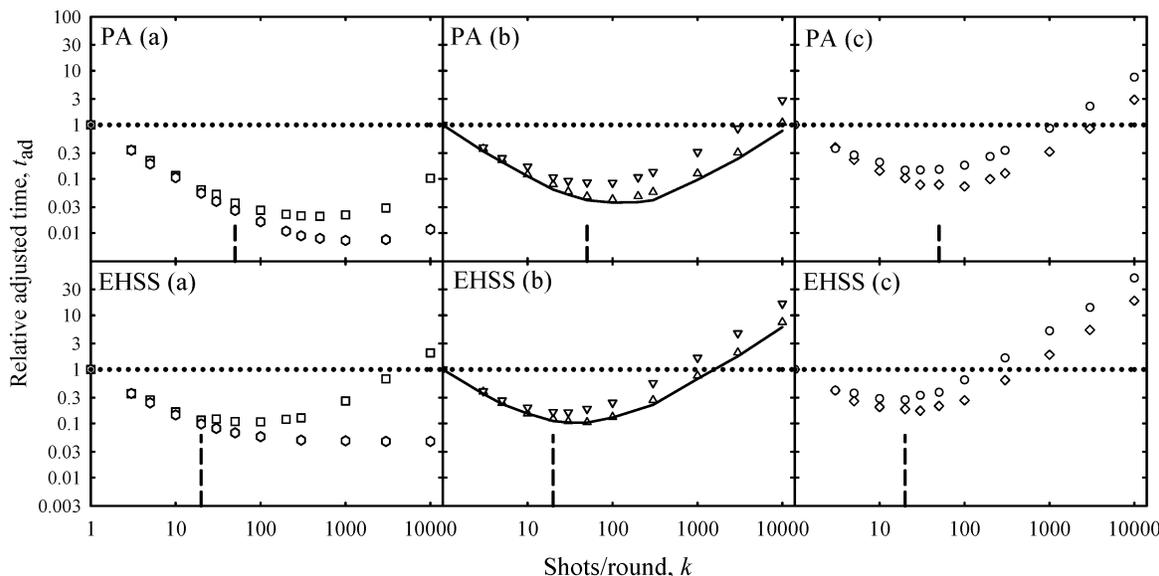


Figure 6. Cost of PA2/ k and EHSS2/ k calculations relative to PA2 and EHSS2 with equal statistical error (marked by dotted lines). Symbols are as in Figure 4; the solid line in b is for HSA. Vertical dashed bars indicate $k = 50$ for PA and $k = 20$ for EHSS.

TABLE 2: Comparison of Optimum Algorithms with and without Ion Rotation (for Both PA and EHSS Models)^a

method	metrics	C ₆₀	Si ₁₀	C ₄₈	C ₃₆	ubiquitin	bradykinin	HSA
PA2/ k	optimum k	10 ³	500	100	20	100	100	100
	minimum t_{ad}	0.007	0.020	0.073	0.14	0.042	0.085	0.036
	t_{ad} at $k = 50$	0.026	0.036	0.079	0.15	0.047	0.086	0.040
	t_{ad} at $k = 10^4$	0.012	0.10	2.9	7.6	1.1	2.9	0.78
PA0	t_{ad} (vs PA2)	0.0021	0.093	0.25	0.060	0.10	0.12	0.049
EHSS2/ k	optimum k	> 10 ³	100	30	20	50	30	50
	minimum t_{ad}	0.047	0.10	0.17	0.27	0.10	0.16	0.10
	t_{ad} at $k = 20$	0.098	0.12	0.18	0.27	0.12	0.16	0.11
EHSS0	t_{ad} (vs EHSS2)	0.066	0.19	0.22	0.11	0.18	0.18	0.13

^a The range of k for C₆₀ reflects the statistical uncertainty of calculations.

ions, σ_r starts growing rapidly when k exceeds a certain threshold that decreases for lower ion sphericities. For example, $\sigma_r(k)$ of PA2/ k exceeds 1.5 $\sigma_r(1)$ when k reaches 10⁴ for Si₁₀, ~200 for HSA, ~100 for ubiquitin, ~50 for C₄₈, ~30 for bradykinin, and ~20 for C₃₆ (Figure 4). That happens because the addition of shots per round ceases to compensate for a sparser sampling of ion orientations, though the accuracy of each Ω_{dir} still improves. Eventually, adding shots per round does not improve the accuracy of Ω_{dir} much and σ_r becomes controlled by the number of sampled orientations, scaling with $1/\sqrt{l} = \sqrt{k}$ (Figure 4). The same occurs for EHSS2/ k , and, again indicating an error accumulation in the course of multiple scattering, $\sigma_r(k)$ increase somewhat faster than those for PA2/ k (Figure 4).

The countervailing trends of $t(k)$ and $\sigma_r(k)$ suggest the existence of an optimum k , where most of the speed gain due to elimination of steps II is captured but the increase of statistical error is still small. As with any Monte Carlo integration, σ_r of PA2/ k and EHSS2/ k for a given k scale as $1/\sqrt{m}$ (Figure 5). Hence the growth of σ_r at greater k may be *exactly* offset by increasing the total number of shots at a commensurate cost (Figure 5), and the relative execution times of PA2/ k and EHSS2/ k adjusted for a *fixed* accuracy of results are

$$t_{\text{ad}}(k) = [t(k)/t(1)] [\sigma_r(k)/\sigma_r(1)]^2 \quad (10)$$

Except for C₆₀, $t_{\text{ad}}(k)$ indeed minimizes at a finite k (Figure 6), ~20–500 for PA2/ k and ~20–100 for EHSS2/ k (Table 2). For C₆₀, $\sigma_r(k)$ barely increases at the highest k (Figure 4a) and t_{ad} has virtually no minimum. The optimum k are lower for EHSS

than for PA as $\sigma_r(k)$ increase faster and $t(k)$ decrease slower, as discussed above. In either method, the optimum k are lower for less spherical geometries (Table 2) because the rapid increase of $\sigma_r(k)$ starts at lower k : adequately averaging over the orientations of less symmetric objects requires sampling a greater number of orientations.

However, the minima of $t_{\text{ad}}(k)$ are shallow (Figure 6), and in practice no geometry-specific optimization of k is needed: the cost of EHSS2/ k at $k = 20$ is within ~10% of the minimum for all ions studied except the special case of C₆₀ (Table 2). In that simple approach, the efficiency gain over EHSS2 is by a factor of ~4–6 for highly elongated C₄₈ and C₃₆, ~6–9 for mildly aspherical geometries common for biomolecules, and ~8–10 for compact structures such as Si₁₀ and C₆₀. In view of the benchmarks in Table 1, EHSS2/20 is faster than Mobcal *at equal accuracy* by a factor of ~7–19 or ~9–26 counting the compiler optimization advantage of EHSS2. A similar near-optimum value of k in PA2/ k for all but the most compact geometries is 50 (Table 2), and the gain versus PA2 is yet greater, by ~6–40 times. The algorithm adopted in Sigma⁴⁹ (in essence PA3/ k with $k \sim 10^4$) is virtually equivalent to PA2/ k with the same k : at $k \sim 10^4$ steps II take <3% of the total time and a difference by one-third between the costs of 2-D and 3-D rotation matters little. This means a computational efficiency below that of near-optimum PA2/50 by ~20–50 times, except for most compact ions (Table 2). In fact, PA2/10000 is often less efficient than even PA2 with rotation after each shot.

Evaluation of PA0 and EHSS0. A complete elimination of rotations allows PA0 and EHSS0 to run faster than PA2 and

EHSS2 by ~ 20 – 60 and ~ 6 – 17 times, respectively (Table 1). The statistical error also increases somewhat because of a greater number of missed shots as discussed above, except for a near-spherical C_{60} where σ_r decreases because the projection of the surrounding sphere is smaller than the rectangles containing the projections of the ion itself. In all cases, the speed gain far outweighs the decrease of accuracy (if any) when t_{ad} is evaluated using eq 10: the PA0 and EHSS0 algorithms outperform PA2 and EHSS2 by factors of ≥ 10 and ≥ 5 , respectively (Table 2). The comparison with optimum PA2/ k and EHSS2/ k procedures is less trivial. In PA calculations, PA2/50 is superior to PA0 for geometries other than C_{60} and C_{36} , but the difference is within a factor of 3 (except for C_{60}). The performance of EHSS2/20 is close to that of EHSS0 except for C_{36} , where the latter is ~ 2.5 times more efficient (Table 2). In the result, the gain of EHSS0 over Mobcal at equal accuracy is by a factor of ~ 8 – 30 or ~ 11 – 40 if the compiler optimization gain is counted. Averaging over all seven geometries, the nonrotating algorithm appears slightly more effective for EHSS than for PA: EHSS0 is equivalent to EHSS2/20, while PA0 lags behind PA2/50 by a factor of 1.6 (or 1.8 excluding the special case of C_{60}). This likely stems from a relatively greater efficiency of PA2/ k compared to EHSS2/ k allowed by higher R_{II} values in PA.

Conclusions

We have optimized the algorithms for computing the mobilities of polyatomic ions in gases using the widely adopted projection approximation (PA) and exact hard spheres scattering (EHSS) models. The new codes have been evaluated for a set of ion geometries that covers a broad range of sizes and shapes.

Both PA and EHSS calculations may be greatly accelerated by (i) forsaking redundant orientational averaging around one of the three axes and (ii) reducing the number of expensive ion rotation steps through evaluation of multiple (k) ion/buffer gas collisions per orientation. The best k depends on the ion shape somewhat, but in most realistic cases $k = 50$ and 20 are near-optimum for PA and EHSS, respectively. For EHSS, the efficiency gain over the previously available Mobcal software⁴⁸ is by a factor of ~ 1.8 because of i and ~ 4 – 10 depending on the ion geometry because of ii, making ~ 7 – 18 times total. The value increases for more spherical objects, and that for typical biomolecules is ~ 10 – 15 . Similar or greater gains have been achieved for PA calculations. Alternatively, all ion rotation steps in both PA and EHSS may be eliminated by aiming the gas molecule at an ion from a random direction. This approach is broadly as good as i and ii combined (slightly better or worse depending on the ion geometry) and more efficient than EHSS in Mobcal by ~ 10 – 30 times. All new codes are amenable to compiler optimization on a PC, providing a further speed gain of $\sim 30\%$.

A major acceleration of calculations enabled by new codes would help modeling mobilities of macromolecular ions and especially sets of geometries representing conformer ensembles and/or time-dependent dynamics. For example, the time needed to evaluate the EHSS mobility of serum albumin, a prototypical large protein of 66 kDa mass, with an accuracy of 0.3% at 99.7% confidence (3 standard deviations) decreases from > 5 h using Mobcal to ~ 14 min.

The algorithms developed here are transferable to calculations of high-field mobilities that require a nonuniform orientational averaging to account for molecular alignment.⁴⁷ These algorithms are also applicable to methods using more realistic ion–buffer gas potentials.^{34,35} However, the trajectory propagation in those calculations is far costlier than in PA or EHSS, and

ion rotations do not dominate the expense (i.e., the R_{II} are much lower). Therefore, the present approaches that drastically reduce or eliminate rotation steps will produce only modest efficiency gains, and efforts would have to focus on accelerating the trajectory propagation.

The mobilities of objects in media are directly related by the Einstein equation to their diffusion coefficients,²⁶ which are also determined by the orientationally averaged collision cross-sections. Hence present algorithms will also be useful to calculate the diffusion coefficients of gas-phase species (ions or neutrals). Comparison of those values with measured diffusion properties^{51–53} may provide structural information^{51,52} that, for neutrals, is not obtainable by IMS.

Acknowledgment. We thank Dr. Thomas Wyttenbach (University of California—Santa Barbara) and Dr. Nelu Marginean (PNNL) for useful discussions of mobility calculation methods, and Professor K. W. M. Siu (York University) for providing his published bradykinin geometry. This work was supported by the NIH National Center for Research Resources (Grant RR 18522). PNNL is a multiprogram national laboratory operated by the Battelle Memorial Institute for the U.S. Department of Energy through Contract DE-AC05-76RLO1830.

References and Notes

- (1) Bowers, M. T.; Kemper, P. R.; von Helden, G.; van Koppen, P. A. *M. Science* **1993**, *260*, 1446.
- (2) Bowers, M. T. *Acc. Chem. Res.* **1994**, *27*, 324.
- (3) Clemmer, D. E.; Jarrold, M. F. *J. Mass Spectrom.* **1997**, *32*, 577.
- (4) Shvartsburg, A. A.; Hudgins, R. R.; Gutierrez, R.; Jungnickel, G.; Fraunheim, T.; Jackson, K. A.; Jarrold, M. F. *J. Phys. Chem. A* **1999**, *103*, 5275.
- (5) Shvartsburg, A. A.; Hudgins, R. R.; Dugourd, P.; Gutierrez, R.; Fraunheim, T.; Jarrold, M. F. *Phys. Rev. Lett.* **2000**, *84*, 2421.
- (6) Shvartsburg, A. A.; Hudgins, R. R.; Dugourd, P.; Jarrold, M. F. *Chem. Soc. Rev.* **2001**, *30*, 26.
- (7) Gilb, S.; Weis, P.; Furche, F.; Ahlrichs, R.; Kappes, M. M. *J. Chem. Phys.* **2002**, *116*, 4094.
- (8) Jackson, K. A.; Horoi, M.; Chaudhuri, I.; Fraunheim, T.; Shvartsburg, A. A. *Phys. Rev. Lett.* **2004**, *93*, 013401.
- (9) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1995**, *117*, 10141.
- (10) von Helden, G.; Wyttenbach, T.; Bowers, M. T. *Science* **1995**, *267*, 1483.
- (11) Wyttenbach, T.; von Helden, G.; Bowers, M. T. *J. Am. Chem. Soc.* **1996**, *118*, 8355.
- (12) Hudgins, R. R.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 12974.
- (13) Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Woods, A. S.; Gillig, K. J.; Russell, D. H. *J. Proteome Res.* **2002**, *1*, 303.
- (14) Rodriguez, C. F.; Orlova, G.; Guo, Y. Z.; Li, X. M.; Siu, C. K.; Hopkinson, A. C.; Siu, K. W. M. *J. Phys. Chem. B* **2006**, *110*, 7528.
- (15) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2240.
- (16) Hoaglund Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E. *Chem. Rev.* **1999**, *99*, 3037.
- (17) Shvartsburg, A. A.; Li, F.; Tang, K.; Smith, R. D. *Anal. Chem.* **2006**, *78*, 3304.
- (18) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science* **2005**, *310*, 1658.
- (19) Hoaglund, C. S.; Liu, Y. S.; Ellington, A. D.; Pagel, M.; Clemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 9051.
- (20) Baker, E. S.; Bernstein, S. L.; Bowers, M. T. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 989.
- (21) Lee, S.; Wyttenbach, T.; Bowers, M. T. *Int. J. Mass Spectrom. Ion Processes* **1997**, *167/168*, 605.
- (22) Leavell, M. D.; Gaucher, S. P.; Leary, J. A.; Taraszka, J. A.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 284.
- (23) Schalley, C. A.; Hoernschmeyer, J.; Li, X.; Silva, G.; Weis, P. *Int. J. Mass Spectrom.* **2003**, *228*, 373.
- (24) Wyttenbach, T.; von Helden, G.; Bowers, M. T. *Int. J. Mass Spectrom. Ion Processes* **1997**, *165/166*, 377.
- (25) Baker, E. S.; Mitchell, C.; Haddad, T. S.; Bowers, M. T. *Chem. Mater.* **2005**, *17*, 2537.
- (26) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley: New York, 1988.

- (27) Von Helden, G.; Hsu, M. T.; Gotts, N.; Bowers, M. T. *J. Phys. Chem.* **1993**, *97*, 8182.
- (28) Dugourd, P.; Hudgins, R. R.; Clemmer, D. E.; Jarrold, M. F. *Rev. Sci. Instrum.* **1997**, *68*, 1122.
- (29) Shvartsburg, A. A.; Hudgins, R. R.; Dugourd, P.; Jarrold, M. F. *J. Phys. Chem. A* **1997**, *101*, 1684.
- (30) Asbury, G. R.; Hill, H. H. *J. Microcolumn Sep.* **2000**, *12*, 172.
- (31) Shvartsburg, A. A.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86.
- (32) Liu, B.; Lu, Z. Y.; Pan, B. C.; Wang, C. Z.; Ho, K. M.; Shvartsburg, A. A.; Jarrold, M. F. *J. Chem. Phys.* **1998**, *109*, 9401.
- (33) Shvartsburg, A. A.; Liu, B.; Jarrold, M. F.; Ho, K. M. *J. Chem. Phys.* **2000**, *112*, 4517.
- (34) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Phys. Chem.* **1996**, *100*, 16082.
- (35) Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *J. Chem. Phys.* **1998**, *108*, 2416.
- (36) Shvartsburg, A. A.; Liu, B.; Siu, K. W. M.; Ho, K. M. *J. Phys. Chem. A* **2000**, *104*, 6152.
- (37) Hudgins, R. R.; Woenckhaus, J.; Jarrold, M. F. *Int. J. Mass Spectrom.* **1997**, *165/166*, 497.
- (38) Li, J.; Taraszka, J. A.; Counterman, A. E.; Clemmer, D. E. *Int. J. Mass Spectrom.* **1999**, *185/186/187*, 37.
- (39) Koeniger, S. L.; Merenbloom, S. I.; Clemmer, D. E. *J. Phys. Chem. B* **2006**, *110*, 7017.
- (40) Counterman, A. E.; Clemmer, D. E. *J. Am. Chem. Soc.* **1999**, *121*, 4031.
- (41) Taraszka, J. A.; Counterman, A. E.; Clemmer, D. E. *Int. J. Mass Spectrom.* **2001**, *204*, 87.
- (42) Counterman, A. E.; Clemmer, D. E. *J. Am. Chem. Soc.* **2001**, *123*, 1490.
- (43) Srebalus Barnes, C. A.; Clemmer, D. E. *J. Phys. Chem. A* **2003**, *107*, 10566.
- (44) Mao, Y.; Woenckhaus, J.; Kolafa, J.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1999**, *121*, 2712.
- (45) Kinnear, B. S.; Kaleta, D. T.; Kohtani, M.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 9243.
- (46) Wyttenbach, T.; von Helden, G.; Batka, J. J., Jr.; Carlat, D.; Bowers, M. T. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 275.
- (47) Shvartsburg, A. A.; Bryskiewicz, T.; Purves, R. W.; Tang, K.; Guevremont, R.; Smith, R. D. *J. Phys. Chem. B* **2006**, *110*, 21966.
- (48) Mobcal downloadable from <http://nano.chem.indiana.edu>.
- (49) Sigma available from Dr. Thomas Wyttenbach (University of California, Santa Barbara), http://bowers.chem.ucsb.edu/theory_analysis/cross-sections/sigma.shtml.
- (50) RCSB data bank is accessible at www.rcsb.org/pdb.
- (51) Mack, E. *J. Am. Chem. Soc.* **1925**, *47*, 2468.
- (52) Melaven, R. M.; Mack, E. *J. Am. Chem. Soc.* **1932**, *54*, 888.
- (53) Rouholahnejad, F.; Tabrizchi, M. *J. Phys. Chem. A* **2006**, *110*, 11208.