Effective Ion Mobility Calculations for Macromolecules by Scattering on Electron Clouds

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · AUGUST 2014

Impact Factor: 2.69 · DOI: 10.1021/jp505012c · Source: PubMed

3 AUTHORS, INCLUDING:

Yuri Alexeev
Argonne National Laboratory
29 PUBLICATIONS **359** CITATIONS

SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

Effective Ion Mobility Calculations for Macromolecules by Scattering on Electron Clouds

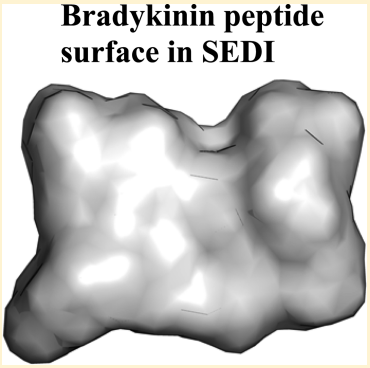
Yuri Alexeev,[†] Dmitri G. Fedorov,[‡] and Alexandre A. Shvartsburg*[,§]

[†]Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, Illinois 60439, United States

[‡]Nanosystem Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8568, Japan

[§]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99354, United States

ABSTRACT: Broad commercialization and increasing resolving power of ion mobility spectrometry/mass spectrometry (IMS/MS) platforms have engendered an explosion of IMS applications to structural characterization of gas-phase biomolecules. That has renewed interest in more accurate and rapid ion mobility calculations that are needed to elicit ion geometries from the measurements. An approach based on scattering on electron density isosurfaces (SEDI) that mirrors the physics of molecular collisions was proven superior to the common methods involving atomic coordinates a decade ago but has remained impractical for large ions because of extreme computational demands. Here, we accelerate SEDI by up to ~500 times using the fragment molecular orbital approach for surface generation and the multiplexed scattering algorithm in conjunction with the new grid extrapolation procedure for cross section evaluations. Parallelization of the code on a supercomputer has produced major further speed gains, allowing SEDI calculations for proteins (defined by over a million surface points) with a precision of <0.1% in 1 min. Initial tests reveal the anticipated dependence of mobility on the ion charge state and lower cross sections in view of reduced surface roughness. Present developments are expected to lead to broad application of SEDI in IMS studies of macromolecules, enabling more accurate and reliable structural assignments.



INTRODUCTION

Ion mobility spectrometry (IMS), and particularly IMS coupled to mass spectrometry (MS), has emerged as a powerful universal approach to resolve the structural isomers of gas-phase ions and characterize their morphologies.[1,2] The technology was initially applied in the 1990s to atomic (carbon, semiconductor, or metal) clusters comprising up to $n \approx 200$ atoms.[3−10] Since 1995, coupling of soft-ionization sources such as matrix-assisted laser desorption/ionization (MALDI)[11] and electrospray ionization (ESI)[12] to IMS/MS has expanded the method to biological macromolecules, first peptides and proteins[13−16] and then ever-larger protein and protein−ligand complexes.[17−19] Conventional IMS methods that determine absolute mobilities (K) in electric fields of moderate intensity (E) can elucidate ion structures *a priori*, by matching the observed K values to those computed for candidate geometries employing first-principles molecular dynamics (MD).[3,8,20−23] This capability is enabled by the amenability of relatively simple ion−molecule scattering to tractable simulations with sufficient accuracy, in contrast to the dynamics in liquids and on solid−liquid interfaces (underlying chromatography and electrophoresis in condensed phases) that permits only empirical modeling.[24,25] Such measurements may be performed using various IMS modes, including drift−tube IMS with constant field and static gas,[1,3−16,26] traveling-wave IMS with oscillatory field and static gas,[27] and differential mobility analysis with constant field and gas flow,[28] but not differential or field asymmetric waveform IMS that measures the difference between mobilities in strong and weak fields.[29,30]

The mobility in the low-E limit depends, in the first approximation of kinetic theory, on the orientationally averaged momentum-transfer collision integral (Ω_{Avg}), the momentum-transfer cross section convolved over the distribution of ion−molecule collision geometries and energies:[26]

$$K = \frac{3q(1+a)}{16N\Omega_{Avg}}\sqrt{\frac{2\pi}{\mu k_B T}} \quad (1)$$

where q is the ion charge, N is the gas number density, μ is the reduced mass of the ion−molecule pair, k_B is the Boltzmann constant, T is the gas temperature, and a is the correction factor. For atomic ions, a may be estimated from higher-order kinetic theory and is normally well within 10%, especially when ions are much heavier than the gas molecules (which is true for polyatomic species addressed here).[26,31,32] The kinetic theory is less advanced for molecular systems, but α should be comparable or smaller, given that the form of ion−molecule interaction potential (Φ) impacts mobilities less.

The quantity Ω_{Avg} can be found rigorously for atomic ions and gas molecules where Φ is spherical,[31,32] but practical cases of polyatomic species necessitate numerical calculations that

Received: May 21, 2014
Revised: July 28, 2014

dx.doi.org/10.1021/jp505012c | *J. Phys. Chem. A* XXXX, XXX, XXX−XXX

approximate both the collision process and the form of Φ. Whereas individual collisions with molecular ions involve redistribution of translational, rotational, and vibrational energy, long-term conservation of each in the steady-state regime implies elasticity on average and mobility calculations have historically assumed elastic collisions. Nonetheless, Ω_{Avg} integrated over all collisions may not match that in the elastic treatment, in particular as the integration is now over the microscopic collision frequencies, not cross sections. Even crude modeling demonstrates that the coupling of rotations or vibrations (in particular, manifested as diffuse instead of specular scattering) may strongly influence Ω_{Avg}, especially for smaller ions and heavier gases.[33−35] While attention to this subject grows, the focus to date has been on improving the description of Φ while assuming elastic collisions and $a = 0$ in eq 1.

The earliest approach was the projection approximation (PA) that equated the cross section to the shadow area of the ion (defined as a set of spheres centered on atomic nuclei) in a parallel beam of impinging molecules.[3,36] The corresponding potential Φ is an infinitely deep vertical well within the ion boundary and zero outside of it. The PA method neglects the momentum of molecules departing after a collision and thus ignores the successive reflections unavoidable with concave ion surfaces. Such scattering is enhanced for larger and therefore less convex ions, especially biomacromolecules that have irregular surfaces with valleys and cavities, increasing Ω_{Avg} by up to ∼1/3 over the PA results.[20] This deficiency is remedied in the exact hard-spheres scattering (EHSS) approach that defines ions as in PA, but postulates Φ with infinite vertical wall at the ion boundary.[20,23] This method accounts for multiple scattering on the repulsive wall, but not the attractive part of Φ that grows in importance for larger ions and/or gas molecules because of higher polarizability and (often) dipole moment. Incorporating that effect requires full MD simulations on the Φ surface. Computational constraints mandate crudely approximating Φ for all but the smallest ions, typically as a sum of charge-induced dipole potential (derived from calculated partial charges on ion atoms and the gas molecule polarizability, a_p) and additive pairwise Lennard-Jones interaction between all atoms of the ion and the gas molecule.[21,22] This approach works fine when the attractive potential affects Ω_{Avg} modestly and thus even a significant inaccuracy of Φ has a muted impact. At room temperature, that is generally the case with helium gas (as He atoms have the lowest $a_p = 0.2$ Å^3 for any molecule) but not N_2 ($a_p = 1.6$ Å^3). That, the absence of vibrations and rotation, and smallest size of He atoms that renders them the finest probe, have made He the choice buffer for IMS structural investigations.

Although MD simulations with He are the golden standard for mobility calculations (deployed to calibrate approximate methods such as the projected superposition approximation, PSA),[37−39] they yield deviations up to ∼5−10% for ions of known geometry.[8,22,40] That was ordinarily acceptable in 1990-s with the IMS resolving power (R) of ∼10, but not now when R > 100 and the mobility measurement accuracy of <1% mean that the utility of IMS as a structural tool is limited primarily by the error margin of mobility calculations rather than experimental factors.[37]

While those errors partly arise from the above approximations to the attractive well of Φ, the key question is how appropriate is representing ions as atomic sets? Molecules actually scatter on the outer electron orbitals, thus nuclear coordinates are relevant only as far as they control the location of those orbitals. Deviations from such "nuclear determinism" (e.g., for ions with the same geometry but different charge states) should precipitate errors in all mobility calculations inputting only atomic coordinates. In fact, the measured Ω_{Avg} for rigid carbon clusters of diverse shapes are greater for singly charged anions than cations by up to ∼15%—far more than plausible from geometric changes.[40] This difference is then due to the addition of two electrons expanding the electron cloud. That effect is not captured within any nuclear coordinate framework but is described well by the scattering on electron density isosurfaces (SEDI) model that emulates ions as surfaces of some electron density (e_{cut}), extracted from quantum-chemical (QC) calculations and defined numerically on a grid.[8]

The electron spill-out is obviously more substantial for anions. For example, the mobilities deduced by EHSS or MD deviated from the experiment by accepted <2% for Si cluster cations with $n = 3−20$, but up to 8% for anions of same sizes.[8] Utilizing SEDI compressed the error to <2% for both, allowing the elucidation of Si_n anion structures and their differences from those of cations.[8] This aspect is especially pertinent to multiply charged ions ubiquitous with ESI. In particular, proteins in solution normally expose polar residues on the surface, where their significant charging may shift the outer electron orbitals. The SEDI approach also captures the consequences of chemical bonding on electron orbitals (e.g., hybridization) to which the atomic paradigm is blind. A major appeal of SEDI is a single adjustable parameter (e_{cut}) regardless of the ion elemental composition,[8,40] unlike for the atom-based methods that attach distinct parameter(s) to each atom type.[38,41] This permits more reliable calculations with less uncertainty due to inevitably arbitrary multiparameter calibration. This point is moot for homoatomic clusters to which SEDI was applied,[8,40] but crucial for heteroatomic ions and particularly biomolecules that feature C, O, N, H, and frequently S, P, and heavier atoms. Finally, smooth surfaces in SEDI lack the artificial crevices at joints of neighboring atoms that are inherent to EHSS and ought to exaggerate the surface roughness,[42] causing overestimation of Ω_{Avg}.

While SEDI clearly represents the repulsive part of Φ more faithfully than atomic approaches, it (like EHSS) ignores the attractive part. Fortunately, the impacts of two parts on Ω_{Avg} are largely multiplicative, and the SEDI-MD hybrid that factorizes them is currently the most sophisticated tool for mobility calculations.[40] An exceptional diversity of carbon clusters in terms of size (from diatomics to fullerenes with $n > 500$), shape (from linear chains to mono- and polycyclic rings to planar graphene sheets to near-spherical fullerenes and concave fullerene oligomers), and chemical bonding (comprising sp and sp^2 hybridizations) makes these species perhaps the greatest challenge to consistent ion mobility calculations.[22,40] Within that set, the maximum discrepancy of calculated mobilities from measurements decreases from ∼20% with PA or EHSS to ∼10% with MD to ∼4% with SEDI-MD (and essentially to zero if distinct C−He interaction parameters are ascribed to the two hybridizations).[40]

The above findings for C and Si cluster cations and anions persuasively demonstrate the superior accuracy of SEDI and SEDI-MD methods. However, they were not thus far tried for species with over $n \approx 100$ or any biomolecules because of formidable computational challenges of creating the electron surfaces and evaluating Ω_{Avg} for them. Indeed, the cost of single-point density functional theory (DFT) calculations to set

dx.doi.org/10.1021/jp505012c | J. Phys. Chem. A XXXX, XXX, XXX−XXX

up a surface tend to scale as $\sim n^3$, because of the Fock matrix diagonalization. The expense of computing Ω_{Avg} is approximately proportional to n in EHSS[20,23] and the number of points delineating the surface (n_{sur}) in SEDI,[8] but, with a grid size (g) fine enough[8] for accurate findings (~0.1 Å), n_{sur} exceeds n by several orders of magnitude. For instance, a model nonapeptide bradykinin (Bk) 1+ with n = 150 has Ω_{Avg} ≈ 240 Å^2,[43,44] which would translate into $n_{sur} \approx 10^5$. These costs had made SEDI impractical for macromolecules, especially considering the need to integrate Ω_{Avg} over the conformer ensembles obtained from MD simulations.[43,45]

Here, we apply SEDI to biomolecules and drastically accelerate both the surface generation and mobility calculation steps using novel algorithms. Coupled with massive computer speed gains since the SEDI development[8,40] in 2000 and parallelization of execution on a supercomputer, the present code enables routine calculations for large biomolecules (including proteins). This capability should broadly improve the accuracy of mobility calculations and allow rigorously exploring the issues of electron spill-out, effect of the ion charge state, surface roughness, and contribution of metal and other heavy atoms (for which the interaction parameters with gas molecules are poorly known).

FULL QC CALCULATIONS OF ISOSURFACES AND APPLICATION TO BIOMOLECULES

First, we extend the SEDI approach[8,40] to biomolecules. The target ions were calculated within GAMESS[46] using the QC methods summarized below. The code outputs the electron densities on a grid inside a box enclosing the ion. Isosurfaces aggregating points with e_{cut} = 0.0025 (au)$^{3-}$ (fit to the measured[40] room-temperature mobility of C_{60}^+ fullerene in He) were extracted from those data using Matlab.[47] Projections (PR) and cross sections (CS) were computed by hard-body scattering as previously,[8,40] assuming a spherical gas molecule with the collision radius (R) of 1.1 Å (suitable for He).[20] The baseline EHSS model adopted the customary R = 2.2 Å for H and 2.7 Å for other atoms in the ion.[20] All CS calculations averaged at least 10^7 MD trajectories for the statistical convergence of under ~0.05%. Except for the work on a supercomputer, execution times are given for parallelized calculations employing the four-core Intel i7-960 processor (3.2 GHz).

Selection of Theory Level for Full QC Treatment. Here, we compare QC methods with speeds differing by over 1000-fold using the benchmark of Bk^{2+} (isomer im1).[44] First, we look at the importance of theory selection by comparing the restricted Hartree−Fock (RHF) and gradient-corrected DFT with two common functionals: Perdew, Burke, Ernzerhof (PBE)[48] and Becke three-parameter, Lee, Yang, Parr (B3LYP).[49,50] The PRs and CSs for isosurfaces derived by these methods are within a negligible 0.1% of each other (Table 1). However, the RHF calculation was ~3 times faster than DFT methods (Table 1) and thus is preferred.

With any theory, the accuracy of QC strongly depends on the basis set. The minimal set is STO-3G with three primitive Gaussian orbitals fitted to a single Slater-type orbital, and other established sets (ordered by improving quality) are the split-valence 6-31G*, 6-31G**, and the augmented correlation-consistent aug-cc-pvdz. The computational expense varies by >2000 times: from 5 min to >9 days (Table 1). The Ω_{Avg} values increase by 0.7% from STO-3G to 6-31G** and then decrease by 0.2% to aug-cc-pvdz. The 6-31G* set leads to the result within 0.1% of the largest and most accurate aug-cc-pvdz set, while the CS/PR ratio that characterizes the surface roughness converges (Table 1). The cost of computing the surface using 6-31G* is ~10 times that of STO-3G, but just ~0.5% that of aug-cc-pvdz. Hence the RHF/6-31G* level appears a fair trade-off between quality and cost.

Table 1. Projections and Cross Sections Obtained Using SEDI (g = 0.1 Å) for Bk^{2+} Isosurfaces, Created by Full QC with Different Theories and Basis Sets

Theory Comparison (6-31G* basis)				
method	SEDI-PR, Å^2	SEDI-CS, Å^2	CS−PR diff, %	time, min
RHF	246.61	276.41	12.08	64
PBE	246.57	276.39	12.09	227
B3LYP	246.70	276.63	12.13	211
Basis Set Comparison (RHF)				
basis set (no. of basis functions)	SEDI-PR, Å^2	SEDI-CS, Å^2	CS−PR diff, %	time, min
STO-3G (455)	244.35	275.51	12.75	5
6-31G* (1290)	246.61	276.41	12.08	64
6-311G** (1894)	247.14	277.36	12.23	231
aug-cc-pvdz (2575)	246.97	276.87	12.11	13,421

Application to Biomolecular Ions: Effects of Charge and Surface Roughness. To begin exploring SEDI for biological ions, we have looked at the effect of charge state for smaller species (single amino acids) and larger ones (a protein). Cations and anions of aspartic acid (Asp, 133 Da), lysine (Lys, 146 Da), and bovine ubiquitin (Ub, 8.6 kDa) in the native conformation (from the PDB database, 1UBQ)[51] were processed using RHF/6-31G* upon residue protonation in the program Avogadro[52] for several pH values in the 2−12 range. With PA and EHSS, the CSs of cations exceed that of corresponding anions, in EHSS by 2% for amino acids and 0.4% for Ub at the edges of above pH range (Table 2). This reflects the added H atoms on the surface expanding the molecular volume, and the effect is much weaker for Ub because those atoms are a smaller fraction of the total: ~1% versus ~10−20% for amino acids. In SEDI, the trend for all three systems is opposite: Ω_{Avg} values for anions exceed those for cations by 1.3−1.5% for amino acids and 0.1% for Ub. Thus, the electron spill-out for anions more than offsets a smaller atomic volume. The relative swing of Ω_{Avg} between EHSS and SEDI here is ~3.5% for amino acids and 0.5% for Ub. While the magnitude for Ub compares with the experimental error margin of top IMS methods, that for amino acids exceeds it by about 10-fold and may well affect the structural assignments. Thus, accounting for electron clouds is important when calculating the mobilities for smaller ions.

Another key issue is the surface roughness. Electron clouds generally have spherical shapes for single atoms but lose symmetry upon chemical bonding, exhibiting shared density between the atoms. Thus, discontinuities on the surfaces of atomically represented ions at sphere joints (Figure 1A) are artifacts—real surfaces are continuous and smoother (Figure 1B). Hence, EHSS arguably overestimates the surface roughness and thus Ω_{Avg}, with true CSs lying between the PA and EHSS values.[42,53] Calculations using SEDI prove that qualitatively, but the overvaluation of Ω_{Avg} (with He) by EHSS relative to PA is only ~1% consistently from single amino acids to proteins (Table 2). The difference for larger gas molecules should be less yet, as the ridges on EHSS surfaces diminish at larger molecular collision radii.[42] Hence EHSS

dx.doi.org/10.1021/jp505012c | J. Phys. Chem. A XXXX, XXX, XXX−XXX

Table 2. Cross Sections Relative to Projections for Selected Biomolecular Ions, Represented by Atomic Sets (PA and EHSS Models) and Electronic Surfaces (SEDI-PR and SEDI-CS Using RHF/6-31G* and g = 0.1 Å)

system	n	PA, Å^2	EHSS, Å^2	SEDI-PR, Å^2	SEDI-CS, Å^2	SEDI-CS−EHSS diff, %	EHSS−PA diff, %	SEDI-CS−PR diff, %	col(8) − col(9)
Asp^+ (pH 2)	17	61.42	63.47	60.58	62.18	−2.08	3.34	2.64	0.70
Asp^{2-} (pH 11)	14	60.29	62.24	61.43	63.01	1.24	3.23	2.57	0.66
Lys^{2+} (pH 2)	26	74.90	78.77	74.22	77.15	−2.10	5.17	3.95	1.22
Lys^- (pH 11)	23	73.39	77.13	75.34	78.27	1.48	5.10	3.89	1.21
Bk^{2+}	151	243.44	275.92	246.61	276.41	0.49	13.34	12.08	1.26
Ub^0 (pH 6.5)	1231	905.56	1114.25	906.65	1106.44	−0.71	23.05	22.04	1.01
Ub^{11+} (pH 2)	1242	907.38	1117.39	906.25	1106.34	−1.00	23.14	22.08	1.06
Ub^{5-} (pH 12)	1226	904.12	1112.47	907.06	1107.06	−0.49	23.05	22.05	1.00
mean									1.02

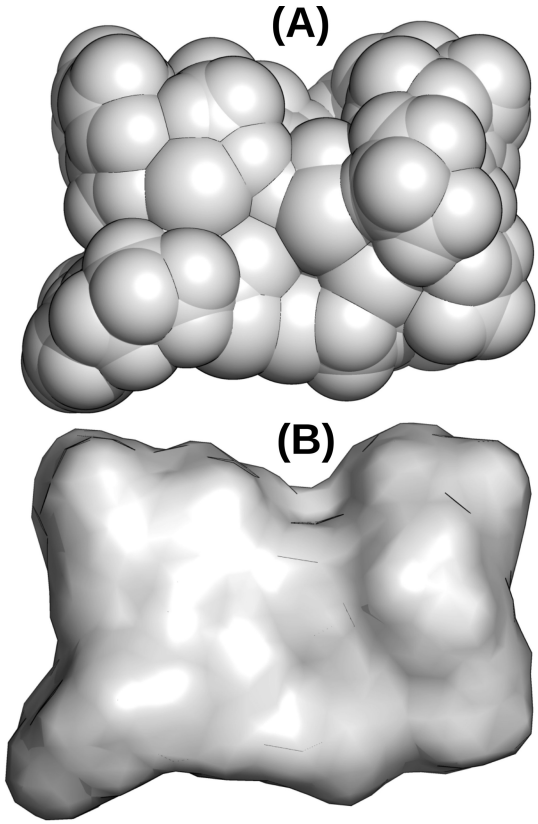


Figure 1. Exemplary Bk^{2+} peptide as seen in (A) EHSS and (B) SEDI (full QC at the RHF/6-31G* level, g = 0.1 Å).

overestimates the surface roughness only marginally; in the absence of SEDI data, one may discount the EHSS values by 1%.

The SEDI values for biomolecules across a wide size range (with surfaces mostly exposing H atoms) are close to the EHSS values and measurements (Table 2). This confirms the applicability of SEDI with same e_{cut} to ions of different elemental composition. While the original SEDI protocol[8,40] can handle biomolecules including proteins (in part, thanks to the recent breakthroughs in computing power), routine calculations for macromolecules require a huge algorithmic acceleration.

NEW FRAGMENT MOLECULAR ORBITAL APPROACH TO ISOSURFACE GENERATION

Computational Development. Our first advance is the method to construct electron density surfaces that is dramatically faster than full QC[8,40] and scales with n linearly.[54] This method leverages the fragment molecular orbital (FMO) approach developed over the past decade to model macromolecules.[55] The idea is to split large systems and calculate the fragments and their pairs from first principles in the electrostatic field of whole system.[56,57] For two-body expansion (FMO2), the energy of a system divided into K fragments is the sum of the fragment energies E_I and pair corrections evaluated using pair energies E_{IJ}:

$$E^{\mathrm{FMO2}} = \sum_{I}^{K} E_I + \sum_{I>J}^{K} (E_{IJ} - E_I - E_J) \tag{2}$$

With the electrostatic dimer approximation, the number of pairs scales linearly (with RHF or DFT).[58] The method has been successful for organic and inorganic species containing up to ∼24,000 atoms, including proteins, ligated proteins, DNA, zeolites, Si nanowires, boron nitride nanorings, and ionic liquids.[59−66] Here, we use FMO implemented in GAMESS package.[67,68]

While most FMO studies focus on the energy per eq 2, our goal is to evaluate the electron density (e) that has been computed using FMO[69,70] and other fragment-based methods.[71−78] In analogy to eq 2, the total density e at a grid point $\mathbf{r}$ is computed in FMO2 from the values for fragments e_I and their pairs e_{IJ} as[69]

$$e^{\mathrm{FMO2}}(\mathbf{r}) = \sum_{I}^{N} e_I(\mathbf{r}) + \sum_{I>J}^{N} [e_{IJ}(\mathbf{r}) - e_I(\mathbf{r}) - e_J(\mathbf{r})] \tag{3}$$

This calculation is easy at each point, but involves a very large number of points for macromolecules. Hence, we also test a much faster one-body expansion (FMO1) that truncates eq 3 at the first term:

$$e^{\mathrm{FMO1}}(\mathbf{r}) = \sum_{I}^{N} e_I(\mathbf{r}) \tag{4}$$

In this "PL" mode incorporating many-body polarization, all fragments are calculated self-consistently in the embedding electrostatic potential of overall system that depends on the electron density for all fragments. In another option "PL0", the embedding potential is computed for isolated fragments.[79] This isolated state is trivially defined for a molecular cluster, but not for covalently bound fragments.

We now introduce a new mode, "PLH", requiring neither an isolated state definition nor full self-consistency. Here, extended Huckel (H) orbitals are constructed for each fragment and projected onto the actual basis set. The resulting densities are

dx.doi.org/10.1021/jp505012c | J. Phys. Chem. A XXXX, XXX, XXX−XXX

used to build the initial embedding potential polarizing the fragment electronic states, which serves to start two self-consistent charge iterations in FMO. Physically, the many-body polarization is estimated using those states with the embedding from Huckel-derived densities, wherein the second iteration ensures that realistic SCF densities are employed when calculating the embedding potential to evaluate the total electron density. This approach resembles[80] the FMO/PCM[1(2)], where the many-body polarization of the induced solvent charges is evaluated in two iterations, starting from the one-body expansion in PCM[1]. The use of fragments in FMO is ideal for parallel computing,[70,81] and we have benchmarked the code on a massively parallel machine. Isosurfaces are extracted from FMO data as for the full QC. While all species considered below are closed-shell, the FMO approach equally applies to radicals.[82,83]

The other critical novelty is the grid space truncation. The e_I and e_{IJ} values for fragments and their pairs are intrinsically localized, while the delocalization effects are incorporated via pair corrections in eq 3. Then, as the electron density rapidly decays with distance, computing the contributions of e_I and e_{IJ} to all grid points for a macromolecule is superfluous. Instead, we evaluate density corrections to only the points within a box containing the fragment (pair) with a surrounding layer of fixed thickness (here, 2 au provided sufficient accuracy).

Dramatic Acceleration for Larger Macromolecules. Here, we benchmark the accuracy and costs of FMO treatments outlined above for the Bk^{2+} and Ub^{7+} ions, and discuss the grounds for accuracy of resulting isosurfaces. The surfaces produced by full QC and all three FMO modalities have equal number of points within 0.2% for Bk and 0.02% for Ub, and the resulting PRs and CSs lie within 0.1% for Bk and 0.01% for Ub (Table 3). These deviations are essentially within the statistical error of mobility calculations and do not go up as the FMO level is downgraded: the rudimentary FMO1/PLH procedure still substitutes for the full QC perfectly. The computational acceleration is a modest ~5 times for Bk but a staggering ~500 times for the much larger Ub: the benefit of fragmentation naturally increases for larger objects. In absolute terms, speeding up the generation of isosurface for Ub on a regular personal computer (PC) from 3 days to 8 min opens the door to routine deployment of SEDI for proteins, even when the geometry evolves in time.[43,45] A yet greater advantage is expected for larger proteins and complexes, enabling use of SEDI for those systems.

Table 3. Projections and Cross Sections Computed by SEDI Using Surfaces from Full QC and Various FMO Methods (All at the RHF/STO-3G Level), and Times for the Surface Generation Step

type of calculation (no. of surface points)	SEDI-PR, Å^2	SEDI-CS, Å^2	SEDI-CS−PR diff, %	time, min
Bradykinin^{2+}, 151 atoms				
full QC (147,806)	244.35	275.51	12.75	5
FMO2 (147,472)	244.21	275.43	12.78	8
FMO1/PL (147,627)	244.25	275.47	12.78	2
FMO1/PLH (147,638)	244.27	275.49	12.78	1
Ubiquitin^{7+}, 1238 atoms				
full QC (1,166,777)	1027.27	1270.12	23.64	3990
FMO2 (1,166,780)	1027.26	1270.10	23.64	135
FMO1/PLH (1,166,441)	1027.37	1270.18	23.63	8

A major contributor to this success is our novel FMO1/PLH simplification that takes two iterations (by definition) versus ~15−20 iterations for FMO1/PL. The final code acceleration is less than the ~7−10 times suggested by these numbers (because of the fixed time to evaluate the density on a grid), but is still material.

An extraordinary forgiveness of mobility calculations to even the crudest FMO approximations may surprise as typical molecular energy evaluations require far more elaborate treatments, where even FMO2 may fail and one must resort[67] to much more expensive three-body FMO3. To rationalize this, we note that the commonly desired accuracy of 1 kcal/mol amounts to a microscopic fraction of the total energy (for example, 2.6 ppb for a 24,060-atom system with the energy of -3.88×10^8 kcal/mol).[61] Conversely, the accuracy of calculated Ω_{Avg} compares with that of electron density itself, with nothing under ~0.1% ever sought.

Further, the major deficiency of FMO1 treatments is a distorted description of severed inter-fragment bonds, corrected in FMO2 by considering the fragment pairs with those bonds intact. That is crucial for energy evaluation, as the covalent bond energies are large (e.g., ~350 kJ/mol per C−C bond). However, those severed bonds form the molecular backbones that are generally buried inside ions, barely affecting the surfaces (made of side chains and H atoms) and thus the CSs. Another problem of FMO1 is the neglect of inter-fragment charge transfer, such as that due to hydrogen bonding or salt bridges. That did not diminish the quality of isosurfaces here (Table 3), which makes sense as the typical charge transfer per hydrogen bond[79] is ~0.02−0.03 electron, or <0.1% of ~40−120 valence electrons in the two bonded residue side chains and even smaller percentage relative to the whole ion. The effect might be larger for salt bridges not tried here, which could be remedied by a FMO scheme confining each salt bridge within one fragment.

GRID EXTRAPOLATION FOR FAST CROSS-SECTION CALCULATIONS

Ultrafast FMO methods to generate electronic surfaces move the onus of SEDI to cross-section calculations. While the surface for Ub with g = 0.1 Å (comprising 1.2 mln. points) can now be drawn on a PC in 8 min (Table 3), running 10^7 trajectories using Mobcal[20] (the only code existent in 2000 when SEDI was introduced)[8,40] would take ~27 days. Subsequent algorithmic optimization has sped up these calculations by ~20 times,[23] but ~30 h is still hardly practical. Here, we report a novel concept of rapid yet precise Ω_{Avg} calculations for fine grids via extrapolation using sparser grids.

All SEDI calculations thus far adopted g = 0.08−0.1 Å. Raising g elevates PRs and CSs for any species (Table 4) and the CS/PR ratio, as sparser grids roughen the surface.[40] The scaling of those increases with respect to g is superlinear, lying between g and g^2 for PR and near g^2 for CS (Figure 2). The best fit for Bk^{2+} (over g = 0.07−1 Å) is provided by $g^{1.48}$ for PR and $g^{2.01}$ for CS (Figure 3). The optimum for Ub, which is topical as a large species where evaluating Ω_{Avg} is costlier, is close at $g^{1.61}$ and $g^{2.06}$ (Figure 3). An excellent fit quality, especially for the more relevant CS (r^2 = 0.997 for Bk and 0.9995 for Ub with the g^2 scaling), permits robust extrapolation to g = 0 from the data for g > 0.1 Å (Figure 4). Simple two-point extrapolations from the values for g_1 and $2g_1$ provide a stable and extremely accurate result of 276.5 ± 0.1 Å (CS) up to at least g_1 = 0.5 Å (Figure 4a). As the cost of computing Ω_{Avg} is nearly proportional to n_{sur}

dx.doi.org/10.1021/jp505012c | J. Phys. Chem. A XXXX, XXX, XXX−XXX

Table 4. Projection and Cross Sections Computed for Bk^{2+} Surfaces (from B3LYP/6-31G*) and Times for Their Evaluation (with an Existing Surface) using 10^7 Trajectories, Depending on the Grid Size

grid size, Å (no. of surface points)	SEDI-PR, Å^2	SEDI-CS, Å^2	SEDI-CS−PR diff, %	time, min
0.07 (291,011)	246.73	276.66	12.13	429
0.1 (142,507)	246.84	276.87	12.16	237
0.15 (63,330)	246.91	277.12	12.23	127
0.2 (35,598)	247.11	277.60	12.34	61
0.3 (15,705)	247.64	278.87	12.61	40
0.4 (8,836)	248.30	280.72	13.05	23
0.5 (5,620)	249.10	282.95	13.59	14
0.6 (3,886)	249.99	285.69	14.28	10
0.7 (2,822)	250.51	288.42	15.13	7
0.8 (2,158)	252.16	293.68	16.47	5
0.9 (1,674)	252.10	296.05	17.43	4
1 (1,336)	253.73	302.26	19.13	3

that scales as g^{-2} (Table 4), those extrapolations take just ∼25% longer than the calculations for g_1 alone. For g_1 = 0.5 Å, that means ∼3% of the time for single calculation for g = 0.07 Å (Figure 4a), namely 17 min for Bk and ∼1.5 h. for the above Ub example. Other sensible two-point extrapolations, e.g., using g_1 and $4g_1$ grids, produce similar outcomes and speed gains (Figure 4). To gauge the adequacy of result, one may switch to three-point extrapolations that could still be much faster than the single calculation for a dense grid. For example, computing Ω_{Avg} for g = 0.3, 0.4, and 0.6 Å costs ∼1/6 of the time for 0.07 Å alone (Figure 4b). We emphasize that these gains do not degrade the accuracy, which actually improves compared to the use of g = 0.1 or 0.07 Å (Figure 4a): the errors of extrapolation are less than those due to finite grid steps even for those finest grids. Together, the optimum multiplexing of scattering and grid extrapolation[23] accelerate the CS evaluation by up to ∼500 times over the original SEDI procedure,[8,40] about matching the speed-up of surface creation delivered by FMO.

Concluding, extrapolation to zero grid size from two or three values in the g = 0.3−1 Å range permits accelerating the evaluation of CSs in SEDI by over an order of magnitude. As the result, the whole task (creating the surface and calculating Ω_{Avg}) can be completed on a PC in ∼20 min for typical peptides and ∼2 h. for a small protein. While those times are reasonable and allow averaging over conformer ensembles, a supercomputer greatly speeds up the process.

PARALLELIZATION ON A SUPERCOMPUTER FOR SEDI CALCULATIONS OF PROTEINS IN ONE MINUTE

The FMO method in GAMESS used to generate isosurfaces is parallelized.[82] The multiplexed CS calculation code[23] (EHSS 2/20 version) has been parallelized for this project using the message passing interface (MPI), with the total count of MD trajectories divided by the number of MPI ranks to obtain the share of trajectories per rank. A different random number generator seed is prepared for each rank, and the computed PR and CS values are averaged over all ranks.

The parallel calculations were performed on an IBM Blue Gene/Q supercomputer at ANL (Mira) with 655,360 cores and 10 Petaflop theoretical speed, loading up to 65,536 cores. The speed gains were tracked for ubiquitin, using the FMO1/PLH theory with RHF/6-31G* basis set and SEDI with two-point

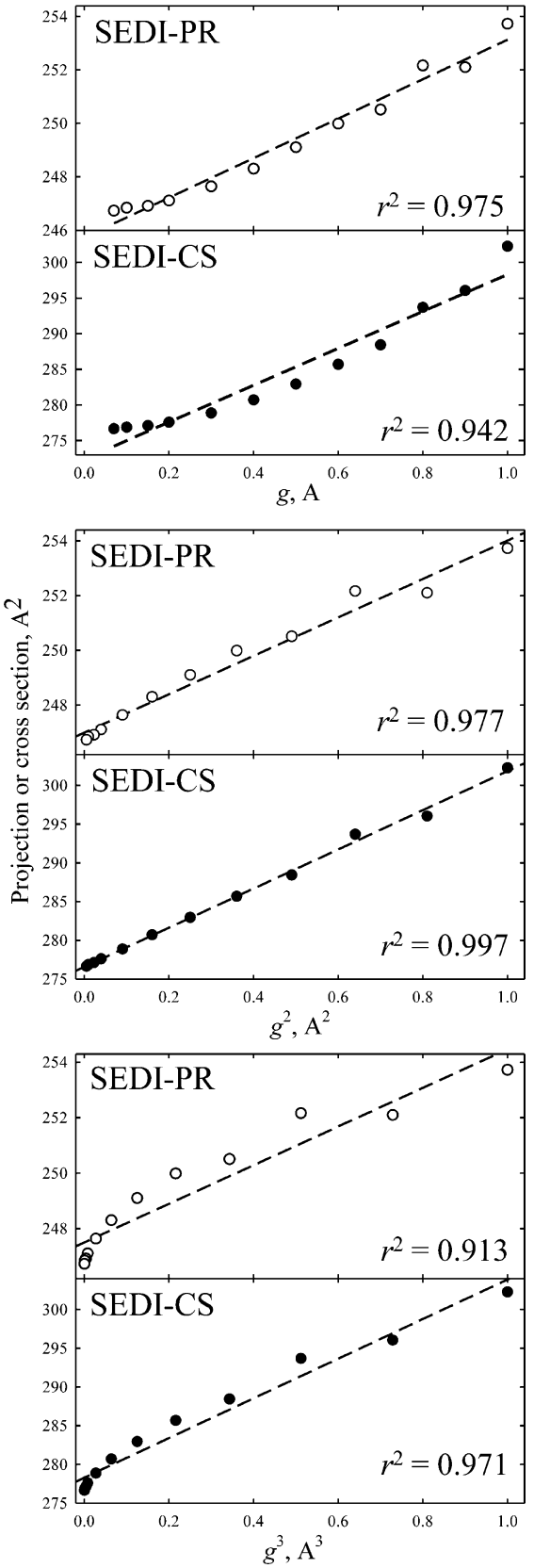


Figure 2. Projections and cross sections calculated for Bk^{2+} using SEDI with various grid sizes (circles), plotted versus powers of g. Lines are first-order regressions through the data; coefficients of determination are given for each.

dx.doi.org/10.1021/jp505012c | J. Phys. Chem. A XXXX, XXX, XXX−XXX

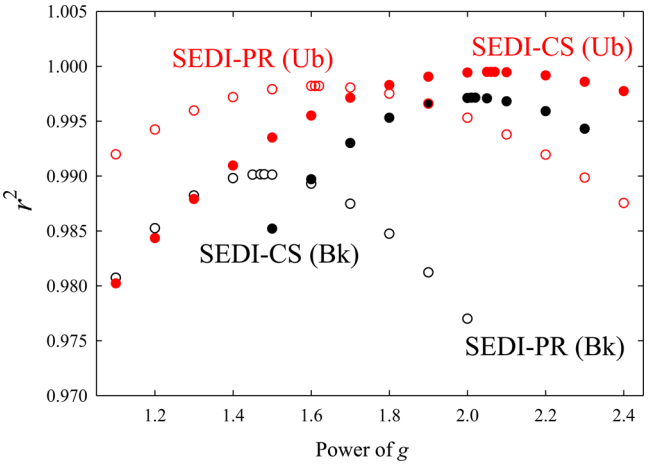


Figure 3. Coefficients of determination for plots such as those in Figure 2, versus the power of g.

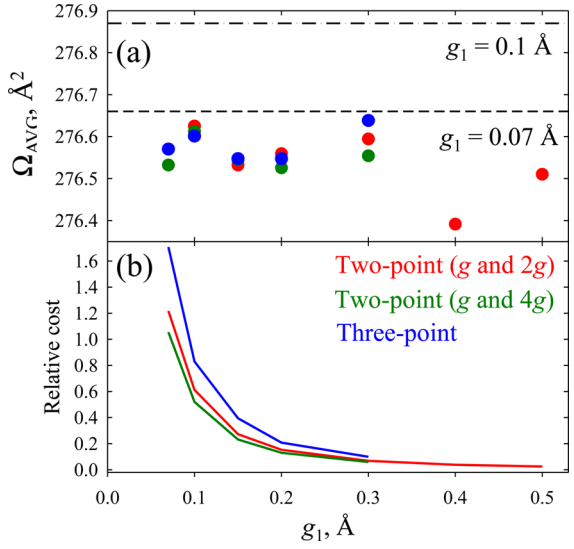


Figure 4. Extrapolation of Ω_{Avg} for Bk^{2+} from SEDI-CS data vs the smallest grid size involved: (a) resulting values; (b) total execution time for all needed CS calculations relative to single evaluation for g = 0.07 Å. The outcomes are shown for three extrapolation models (color-coded as labeled).

extrapolation from g = 0.5 and 1 Å (Figure 5). Favorable scalability extends to 32,768 cores for FMO and (at least) 65,536 cores for CS evaluations, with the execution times of 41 and 31 s, respectively. Thus, full SEDI calculation can be accomplished in ∼70 s, and similar times for the surface generation and Ω_{Avg} evaluation steps mean that speed-up of either by targeted code optimization for the Blue Gene architecture (efforts in progress) would reduce the total further.

CONCLUSIONS

The scattering on electron density isosurfaces (SEDI) paradigm for ion mobility calculations, which foundationally diverges from the alternatives by visualizing ions as electron clouds rather than atomic sets, has been expanded from small inorganic clusters[8,40] to large macromolecules with cloud surfaces encompassing millions of points. This becomes possible upon revolutionary acceleration of both the steps of surface generation (employing bare-bones fragment molecular orbital methods too inaccurate for usual energy calculations)

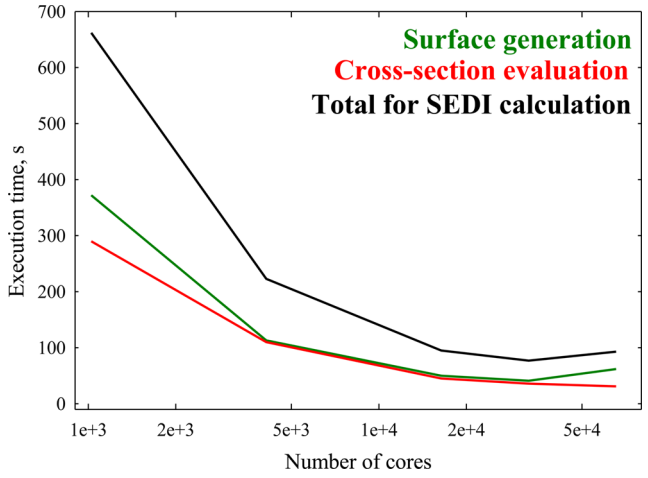


Figure 5. Scalability of SEDI calculations for Ub^0 on Mira vs the number of cores used (log scale). Execution times are plotted for the surface generation, evaluation of Ω_{Avg} via a two-point extrapolation, and the sum of two steps (color-coded as labeled).

and cross section evaluation (by the multiplexed scattering algorithm optimized for EHSS and a new approach of grid extrapolation that replaces fine grids by much sparser ones with fewer points). The method has been efficiently parallelized on a supercomputer, permitting very precise SEDI calculation for a protein in about 1 min.

Initial applications to biomolecules show the expectedly greater CSs for anions than cations with identical atomic geometries, which is significant for smaller species. The SEDI surfaces are devoid of unphysical discontinuities at atomic junctures. This reduces multiple scattering and ensuing ratios of CSs to projections by ∼1% relative to EHSS, with no apparent correlation to the ion size. While this effect was self-evident qualitatively, its presently determined magnitude is less than that commonly imputed for proteins. A hallmark of SEDI is a single adjustable parameter (the surface electron density, e_{cut}) that is transferable across ions made of different elements. We have affirmed that major advantage: the e_{cut} values fit for carbon clusters lead to accurate mobilities for peptides and proteins with mostly hydrogens on the surface. This uniformity should extend to other elements such as metals or halogens, which should improve the mobilities computed for species with those atoms on the surface.[41,84,85] Evaluation of SEDI in that context is ongoing.

Like with clusters, the SEDI treatment of macromolecules can be combined with MD tools that account for the long-range ion−molecule potential to produce highly accurate mobilities.[40] However, such calculations for large species are laborious, and their coupling would negate the tremendous speed gains achieved in this work. In this situation, one can replace MD methods by their phenomenological parametrizations such as PSA.[37−39]

Finally, the more-accurate calculations of elastic Ω_{Avg} provided for polyatomic ions by SEDI and its derivatives call for a greater focus on other previously obscure factors, such as various facets of inelastic scattering and higher-order kinetic effects.

AUTHOR INFORMATION

Corresponding Author

*E-mail: aas996@yahoo.com.

dx.doi.org/10.1021/jp505012c | J. Phys. Chem. A XXXX, XXX, XXX−XXX

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Next Generation Super Computing Project and Nanoscience Program (MEXT, Japan), Strategic Programs for Innovative Research (SPIRE, Japan) at AIST, and PNNL Technology Commercialization Office. This research used resources of the Argonne Leadership Computing Facility at ANL, supported by the U.S. DOE Office of Science under contract DE-AC02-06CH11357, and the Environmental Molecular Sciences Laboratory, a DOE national scientific user facility at PNNL.

■ REFERENCES

(1) Eiceman, G. A.; Karpas, Z.; Hill, H. H. *Ion Mobility Spectrometry*; CRC Press: Boca Raton, FL, 2013.
(2) Wilkins, C. L., Trimpin, S., Eds. *Ion Mobility Spectrometry—Mass Spectrometry: Theory and Applications*; CRC Press: Boca Raton, FL, 2010.
(3) von Helden, G.; Hsu, M. T.; Gotts, N.; Bowers, M. T. Carbon Cluster Cations with up to 84 Atoms: Structures, Formation Mechanism, and Reactivity. *J. Phys. Chem.* **1993**, *97*, 8182−8192.
(4) Bowers, M. T.; Kemper, P. R.; von Helden, G.; van Koppen, P. A. M. Gas-Phase Ion Chromatography: Transition Metal State Selection and Carbon Cluster Formation. *Science* **1993**, *260*, 1446−1451.
(5) Lee, S.; Gotts, N. G.; von Helden, G.; Bowers, M. T. Evidence from Ion Chromatography Experiments that Met-Cars are Hollow Cage Clusters. *Science* **1995**, *267*, 999−1001.
(6) Clemmer, D. E.; Hunter, J. M.; Shelimov, K. B.; Jarrold, M. F. Physical and Chemical Evidence for Metallofullerenes with Metal Atoms as Part of the Cage. *Nature* **1994**, *372*, 248−250.
(7) Shvartsburg, A. A.; Hudgins, R. R.; Dugourd, P.; Gutierrez, R.; Frauenheim, T.; Jarrold, M. F. Observation of "Stick" and "Handle" Intermediates Along the Fullerene Road. *Phys. Rev. Lett.* **2000**, *84*, 2421−2424.
(8) Shvartsburg, A. A.; Liu, B.; Jarrold, M. F.; Ho, K. M. Modeling Ionic Mobilities by Scattering on Electronic Density Isosurfaces: Application to Silicon Cluster Anions. *J. Chem. Phys.* **2000**, *112*, 4517−4526.
(9) Shvartsburg, A. A.; Hudgins, R. R.; Dugourd, P.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Applications to Semiconductor Clusters. *Chem. Soc. Rev.* **2001**, *30*, 26−35.
(10) Shvartsburg, A. A.; Jarrold, M. F. Tin Clusters Adopt Prolate Geometries. *Phys. Rev. A* **1999**, *60*, 1235−1239.
(11) von Helden, G.; Wyttenbach, T.; Bowers, M. T. Conformation of Macromolecules in the Gas Phase: Use of Matrix-Assisted Laser Desorption Methods in Ion Chromatography. *Science* **1995**, *267*, 1483−1485.
(12) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Naked Protein Conformations: Cytochrome *c* in the Gas Phase. *J. Am. Chem. Soc.* **1995**, *117*, 10141−10142.
(13) Jarrold, M. F. Peptides and Proteins in the Vapor Phase. *Annu. Rev. Phys. Chem.* **2000**, *51*, 179−207.
(14) Hill, H. H.; Hill, C. H.; Asbury, G. R.; Wu, C.; Matz, L. M.; Ichiye, T. Charge Location on Gas Phase Peptides. *Int. J. Mass Spectrom.* **2002**, *219*, 23−37.
(15) Ruotolo, B. T.; Russell, D. H. Gas-Phase Conformations of Proteolytically-Derived Protein Fragments: the Influence of Solvent on Peptide Conformation. *J. Phys. Chem. B* **2004**, *108*, 15321−15331.
(16) Shvartsburg, A. A.; Li, F.; Tang, K.; Smith, R. D. Characterizing the Structures and Folding of Free Proteins Using 2-D Gas-Phase Separations: Observation of Multiple Unfolded Conformers. *Anal. Chem.* **2006**, *78*, 3304−3315.
(17) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. *Science* **2005**, *310*, 1658−1661.
(18) Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *J. Am. Chem. Soc.* **2012**, *134*, 3429−3438.
(19) Hall, Z.; Politis, A.; Robinson, C. V. Structural Modeling of Heteromeric Protein Complexes from Disassembly Pathways and Ion Mobility−Mass Spectrometry. *Structure* **2012**, *20*, 1596−1609.
(20) Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **1996**, *261*, 86−91.
(21) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long Range Potential. *J. Phys. Chem.* **1996**, *100*, 16082−16086.
(22) Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Mobilities of Carbon Cluster Ions: Critical Importance of the Molecular Attractive Potential. *J. Chem. Phys.* **1998**, *108*, 2416−2423.
(23) Shvartsburg, A. A.; Mashkevich, S.; Baker, E. S.; Smith, R. D. Optimization of Algorithms for Ion Mobility Calculations. *J. Phys. Chem. A* **2007**, *111*, 2002−2010.
(24) Petritis, K.; Kangas, L. J.; Ferguson, P. L.; Anderson, G. A.; Pasa-Tolic, L.; Lipton, M. S.; Auberry, K. J.; Strittmatter, E. F.; Shen, Y.; Zhao, R.; Smith, R. D. Use of Artificial Neural Networks for the Accurate Prediction of Peptide Liquid Chromatography Elution Times in Proteome Analyses. *Anal. Chem.* **2003**, *75*, 1039−1048.
(25) Dwivedi, R. C.; Spicer, V.; Harder, M.; Antonovici, M.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Krokhin, O. V. Practical Implementation of 2D HPLC Scheme with Accurate Peptide Retention Prediction in Both Dimensions for High-Throughput Bottom-Up Proteomics. *Anal. Chem.* **2008**, *80*, 7036−7042.
(26) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley: New York, 1988.
(27) Santos, L. F. A.; Iglesias, A. H.; Pilau, E. J.; Gomes, A. F.; Gozzo, F. C. Traveling-Wave Ion Mobility Mass Spectrometry Analysis of Isomeric Modified Peptides Arising from Chemical Cross-Linking. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 2062−2069.
(28) Hogan, C. J.; Ogorzalek Loo, R. R.; Loo, J. A.; de la Mora, J. F. Ion Mobility−Mass Spectrometry of Phosphorylase B Ions Generated with Supercharging Reagents but in Charge-Reducing Buffer. *Phys. Chem. Chem. Phys.* **2010**, *12*, 13476−13483.
(29) Guevremont, R. High-Field Asymmetric Waveform Ion Mobility Spectrometry: a New Tool for Mass Spectrometry. *J. Chromatogr. A* **2004**, *1058*, 3−19.
(30) Shvartsburg, A. A. *Differential Ion Mobility Spectrometry: Nonlinear Ion Transport and Fundamentals of FAIMS*; CRC Press: Boca Raton, FL, 2008.
(31) Viehland, L. A.; Lozeille, J.; Soldán, P.; Lee, E. P. F.; Wright, T. G. Spectroscopy of Na^+·Rg and Transport Coefficients of Na^+ in Rg (Rg = He−Rn). *J. Chem. Phys.* **2003**, *119*, 3729−3736.
(32) Viehland, L. A.; Lozeille, J.; Soldán, P.; Lee, E. P. F.; Wright, T. G. Spectroscopy of K^+·Rg and Transport Coefficients of K^+ in Rg (Rg = He−Rn). *J. Chem. Phys.* **2004**, *121*, 341−351.
(33) Shvartsburg, A. A.; Mashkevich, S. V.; Siu, K. W. M. Incorporation of Thermal Rotation of Drifting Ions into Mobility Calculations: Drastic Effect for Heavier Buffer Gases. *J. Phys. Chem. A* **2000**, *104*, 9448−9453.
(34) Larriba, C.; Hogan, C. J. Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. *J. Comput. Phys.* **2013**, *251*, 344−363.
(35) Larriba, C.; Hogan, C. J. Ion Mobilities in Diatomic Gases: Measurement Versus Prediction with Non-Specular Scattering Models. *J. Phys. Chem. A* **2013**, *117*, 3887−3901.
(36) Mack, E. Average Cross-Sectional Areas of Molecules by Gaseous Diffusion Methods. *J. Am. Chem. Soc.* **1925**, *47*, 2468−2482.
(37) Bleiholder, C.; Wyttenbach, T.; Bowers, M. T. A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (I). Method. *Int. J. Mass Spectrom.* **2011**, *308*, 1−10.

(38) Bleiholder, C.; Contreras, S.; Do, T. D.; Bowers, M. T. A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (II). Model Parameterization and Definition of Empirical Shape Factors for Proteins. *Int. J. Mass Spectrom.* **2013**, *345 - 347*, 89−96.
(39) Anderson, S. E.; Bleiholder, C.; Brocker, E. R.; Stang, P. J.; Bowers, M. T. A Novel Projection Approximation Algorithm for the Fast and Accurate Computation of Molecular Collision Cross Sections (III). Application to Supramolecular Coordination-Driven Assemblies with Complex Shapes. *Int. J. Mass Spectrom.* **2012**, *330 - 332*, 78−84.
(40) Shvartsburg, A. A.; Liu, B.; Siu, K. W. M.; Ho, K. M. Evaluation of Ionic Mobilities by Coupling the Scattering on Atoms and on Electron Density. *J. Phys. Chem. A* **2000**, *104*, 6152−6157.
(41) Wyttenbach, T.; von Helden, G.; Batka, J. J.; Carlat, D.; Bowers, M. T. Effect of the Long-Range Potential on Ion Mobility Measurements. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 275−282.
(42) Wyttenbach, T.; Bleiholder, C.; Bowers, M. T. Factors Contributing to the Collision Cross Section of Polyatomic Ions in the Kilodalton to Gigadalton Range: Application to Ion Mobility Measurements. *Anal. Chem.* **2013**, *85*, 2191−2199.
(43) Wyttenbach, T.; von Helden, G.; Bowers, M. T. Gas-Phase Conformation of Biological Molecules: Bradykinin. *J. Am. Chem. Soc.* **1996**, *118*, 8355−8364.
(44) Rodriquez, C. E.; Orlova, G.; Guo, Y.; Li, X.; Siu, C. K.; Hopkinson, A. C.; Siu, K. W. M. Gaseous Bradykinin and Its Singly, Doubly, and Triply Protonated Forms: a First-Principles Study. *J. Phys. Chem. A* **2006**, *110*, 7528−7537.
(45) Wyttenbach, T.; Pierson, N. A.; Clemmer, D. E.; Bowers, M. T. Ion Mobility Analysis of Molecular Dynamics. *Annu. Rev. Phys. Chem.* **2014**, *65*, 175−196.
(46) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Shiro, K.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(47) http://www.mathworks.com/products/matlab/
(48) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(49) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(50) Stephens, P. J.; Devlin, F. J.; Chablowski, C. F.; Frisch, M. J. *Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(51) Available at www.rcsb.org.
(52) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: an Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4*, 1−17.
(53) Ninonuevo, M. R.; Leary, J. A. Ion Mobility Mass Spectrometry Coupled with Rapid Protein Threading Predictor Structure Prediction and Collision-Induced Dissociation for Probing Chemokine Conformation and Stability. *Anal. Chem.* **2012**, *84*, 3208−3214.
(54) Goedecker, S.; Scuseria, G. E. Linear Scaling Electronic Structure Methods in Chemistry and Physics. *Comput. Sci. Eng.* **2003**, *5*, 14−21.
(55) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: an Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.
(56) Fedorov, D. G.; Kitaura, K. The Fragment Molecular Orbital Method for Geometry Optimizations of Polypeptides and Proteins. *J. Phys. Chem. A* **2007**, *111*, 6904−6914.
(57) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring Chemistry with the Fragment Molecular Orbital Method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562−7577.
(58) Nakano, T.; Kaminuma, T.; Sato, T.; Fukuzawa, K.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment Molecular Orbital Method: Use of Approximate Electrostatic Potential. *Chem. Phys. Lett.* **2002**, *351*, 475−480.
(59) Komeiji, Y.; Ishida, T.; Fedorov, D. G.; Kitaura, K. Change in a Protein's Electronic Structure Induced by an Explicit Solvent: an *Ab Initio* Fragment Molecular Orbital Study of Ubiquitin. *J. Comput. Chem.* **2007**, *28*, 1750−1762.
(60) Watanabe, T.; Inadomi, Y.; Fukuzawa, K.; Nakano, T.; Tanaka, S.; Nilsson, L.; Nagashima, U. DNA and Estrogen Receptor Interaction Revealed by Fragment Molecular Orbital Calculations. *J. Phys. Chem. B* **2007**, *111*, 9621−9627.
(61) Sawada, T.; Fedorov, D. G.; Kitaura, K. Binding of Influenza A Virus Hemagglutinin to the Sialoside Receptor is Not Controlled by the Homotropic Allosteric Effect. *J. Phys. Chem. B* **2010**, *114*, 15700−15705.
(62) Fedorov, D. G.; Avramov, P. V.; Jensen, J. H.; Kitaura, K. Analytic Gradient for the Adaptive Frozen Orbital Bond Detachment in the Fragment Molecular Orbital Method. *Chem. Phys. Lett.* **2009**, *477*, 169−175.
(63) Fedorov, D. G.; Jensen, J. H.; Deka, R. C.; Kitaura, K. Covalent Bond Fragmentation Suitable To Describe Solids in the Fragment Molecular Orbital Method. *J. Phys. Chem. A* **2008**, *112*, 11808−11816.
(64) Izgorodina, E. I.; Rigby, J.; MacFarlane, D. R. Large-Scale *Ab Initio* Calculations of Archetypical Ionic Liquids. *Chem. Commun.* **2012**, *48*, 1493−1495.
(65) Avramov, P. V.; Fedorov, D. G.; Sorokin, P. B.; Sakai, S.; Entani, S.; Ohtomo, M.; Matsumoto, Y.; Naramoto, H. Intrinsic Edge Asymmetry in Narrow Zigzag Hexagonal Heteroatomic Nanoribbons Causes Their Subtle Uniform Curvature. *J. Phys. Chem. Lett.* **2012**, *3*, 2003−2008.
(66) Okiyama, Y.; Tsukamoto, T.; Watanabe, C.; Fukuzawa, K.; Tanaka, S.; Mochizuki, Y. Modeling of Peptide-Silica Interaction Based on Four-Body Corrected Fragment Molecular Orbital (FMO4) Calculations. *Chem. Phys. Lett.* **2013**, *566*, 25−31.
(67) Fedorov, D. G.; Kitaura, K. The Importance of Three-Body Terms in the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2004**, *120*, 6832−6840.
(68) Alexeev, Y.; Mazanetz, M. P.; Ichihara, O.; Fedorov, D. G. GAMESS as a Free Quantum-Mechanical Platform for Drug Research. *Curr. Top. Med. Chem.* **2012**, *12*, 2013−2033.
(69) Inadomi, Y.; Nakano, T.; Kitaura, K.; Nagashima, U. Definition of Molecular Orbitals in Fragment Molecular Orbital Method. *Chem. Phys. Lett.* **2002**, *364*, 139−143.
(70) Ikegami, T.; Ishida, T.; Fedorov, D. G.; Kitaura, K.; Inadomi, Y.; Umeda, H.; Yokokawa, M.; Sekiguchi, S. Full Electron Calculation Beyond 20,000 Atoms: Ground Electronic State of Photosynthetic Proteins. *Supercomputing, 2005*, Proceedings of the ACM/IEEE SC 2005 Conference, Nov 12−18, 2005; IEEE: Piscataway, NJ, 2005.
(71) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.* **2012**, *112*, 632−672.
(72) Otto, P.; Ladik, J. Investigation of the Interaction Between Molecules at Medium Distances: I. SCF LCAO MO Supermolecule, Perturbational and Mutually Consistent Calculations for Two Interacting HF and CH_2O Molecules. *J. Chem. Phys.* **1975**, *8*, 192−200.
(73) Gao, J. Toward a Molecular Orbital Derived Empirical Potential for Liquid Simulations. *J. Phys. Chem. B* **1997**, *101*, 657−663.
(74) Babu, K.; Gadre, S. R. *Ab Initio* Quality One-Electron Properties of Large Molecules: Development and Testing of Molecular Tailoring Approach. *J. Comput. Chem.* **2003**, *24*, 484−495.
(75) Exner, T. E.; Mezey, P. G. Evaluation of the Field-Adapted ADMA Approach: Absolute and Relative Energies of Crambin and Derivatives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061−4069.
(76) Chen, X.; Zhang, Y.; Zhang, J. Z. H. An Efficient Approach for *Ab Initio* Energy Calculation of Biopolymers. *J. Chem. Phys.* **2005**, *122*, 184105.
(77) He, X.; Merz, K. M. Divide and Conquer Hartree−Fock Calculations on Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405−411.
(78) Kobayashi, M.; Nakai, H. How Does It Become Possible to Treat Delocalized and/or Open-Shell Systems in Fragmentation-Based Linear-Scaling Electronic Structure Calculations? The Case of the

Divide-and-Conquer Method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7629−7639.
(79) Fedorov, D. G.; Kitaura, K. Pair Interaction Energy Decomposition Analysis. *J. Comput. Chem.* **2007**, *28*, 222−237.
(80) Fedorov, D. G.; Kitaura, K.; Li, H.; Jensen, J. H.; Gordon, M. S. The Polarizable Continuum Model (PCM) Interfaced with the Fragment Molecular Orbital Method (FMO). *J. Comput. Chem.* **2006**, *27*, 976−985.
(81) Fletcher, G. D.; Fedorov, D. G.; Pruitt, S. R.; Windus, T. L.; Gordon, M. S. Large-Scale MP2 Calculations on the Blue Gene Architecture Using the Fragment Molecular Orbital Method. *J. Chem. Theory Comput.* **2012**, *8*, 75−79.
(82) Fedorov, D. G.; Olson, R. M.; Kitaura, K.; Gordon, M. S.; Koseki, S. A New Hierarchical Parallelization Scheme: Generalized Distributed Data Interface (GDDI), and an Application to the Fragment Molecular Orbital Method (FMO). *J. Comput. Chem.* **2004**, *25*, 872−880.
(83) Nakata, H.; Fedorov, D. G.; Nagata, T.; Yokojima, S.; Ogata, K.; Kitaura, K.; Nakamura, S. Unrestricted Hartree-Fock Based on the Fragment Molecular Orbital Method: Energy and its Analytic Gradient. *J. Chem. Phys.* **2012**, *137*, 044110.
(84) von Helden, G.; Gotts, N. G.; Maître, P.; Bowers, M. T. The Structures of Small Iron-Carbon Cluster Anions. Linear to Planar to Three-Dimensional. *Chem. Phys. Lett.* **1994**, *227*, 601−608.
(85) von Helden, G.; Porter, E.; Gotts, N. G.; Bowers, M. T. Structures of Carbon Clusters from Polychlorinated Graphitic Precursors: Investigations of $C_{12}Cl_x^+$ (x = 0−10) Using the Ion Chromatography Method. *J. Phys. Chem.* **1995**, *99*, 7707−7714.