Journal Name

ARTICLE

A parallelized molecular collision cross section package with optimized accuracy and efficiency

Christian Ieritano,[a] Jeff Crouse,[a] J. Larry Campbell,[a,b] W. Scott Hopkins*[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Ion mobility-based separation prior to mass spectrometry has become an invaluable tool in the structural elucidation of gas-phase ions and in the characterization of complex mixtures. Application of ion mobility to structural studies requires an accurate methodology to bridge theoretical modelling of chemical structure with experimental determination of an ion's collision cross section (CCS). Herein, we present a refined methodology for calculating ion CCSs using parallel computing architectures that makes use of atom specific parameters, which we have called MobCal-MPI. Tuning of ion-nitrogen van der Waals potentials on a diverse calibration set of 162 molecules returned a RMSE of 2.60 % in CCS calculations of molecules containing the elements C, H, O, N, F, P, S, Cl, Br, and I. External validation of the ion-nitrogen potential was performed on an additional 50 compounds not present in the validation set, returning a RMSE of 2.31 % for the CCSs of these compounds. Owing to the use of parameters from the MMFF94 forcefield, the calibration of the van der Waals potential can be extended to additional atoms defined in the MMFF94 forcefield (*i.e.*, Li, Na, K, Si, Mg, Ca, Fe, Cu, Zn). We expect that the work presented here will serve as a foundation for facile determination of molecular CCSs, as MobCal-MPI boasts up to 64-fold speedups over traditional calculation packages.

Introduction

Ion mobility spectrometry (IMS) coupled to mass spectrometry (MS) is a powerful tool with applications in numerous fields of research. For instance, ion mobility measurements have been employed for structural elucidation in metabolomics, lipidomics, proteomics, and in the detection of illicit substances.[1–7] The success of the IMS approach relies on both rigorous experimental calibration and accurate theoretical modelling of ion collision cross sections (CCSs). With regard to modelling ion mobility, one is generally concerned with the rate of collision between an ion and a neutral buffer gas under specific electric field, pressure, and temperature conditions. Under low field conditions, ion mobility (K) can be described by the Mason-Schamp relation in the free molecular regime as shown in equation 1.[8]

$$K = \frac{\sqrt{18\pi}}{16}\sqrt{\frac{1}{m_{ion}}+\frac{1}{m_{gas}}}\frac{ze}{\sqrt{k_bT}}\frac{1}{\Omega_{avg}}\frac{1}{N} \quad (1)$$

Where m_{ion} is the ion molecular mass, m_{gas} is the molecular mass of the buffer gas, z is the charge, e is the elementary charge, k_b is the Boltzmann constant, T is the temperature, and N is the number density of the gas. The orientationally averaged CCS of an ion is approximated by modelling collisions between the ion and buffer gas. This can be accomplished using one of several existing methods, which include the projection approximation,[9] elastic hard sphere scattering,[10] and the trajectory method.[10] The MobCal code,[9,10] originally produced by Shvartsburg and Jarrold, has been developed and refined over the past twenty years to conduct calculations using all three methods, although the trajectory method is generally accepted to be the most accurate.

Within the trajectory method, ion CCSs are evaluated through momentum transfer integrals, which are averaged over all possible velocities and geometries of the ion and buffer gas as per equation 2.[9,10]

$$\Omega_{avg} = \frac{1}{8\pi^2}\int_0^{2\pi} d\theta \cdot \int_0^{\pi} \sin\varphi \, d\varphi \cdot \int_0^{2\pi}\frac{\pi}{8}\left(\frac{\mu}{k_bT}\right)^3 d\gamma \cdot \int_0^{\infty} g^5 \cdot \exp\left(-\frac{\mu g^2}{2k_bT}\right) dg \cdot \int_0^{\infty} 2b[1-\cos\chi(\theta,\varphi,\gamma,g,b)]\, db \quad (2)$$

Where θ, φ, and γ define the orientation of the ion with respect to the ion-collision gas centre of mass axis, g is the relative velocity, b is the impact parameter, μ is the reduced mass, and χ is the angle at which buffer gas is scattered upon interaction with the ion. Owing to the dependence of χ on molecular orientation and relative velocity, it can only be evaluated numerically as outlined in equation 3.[11]

$$\chi(\theta,\varphi,\gamma,g,b) = (\pi-2b)\int_{r_{min}}^{\infty}\left[r^2\sqrt{1-\frac{b^2}{r^2}-\frac{\Phi(r)}{\frac{1}{2}m_{red}g^2}}\right]^{-1} dr \quad (3)$$

Here, r_{min} is the distance of closest approach between the ion and buffer gas. Gas trajectories are ultimately determined by the intermolecular potential $\Phi(r)$, which is composed of three

[a] Department of Chemistry, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada. E-mail: shopkins@uwaterloo.ca
[b] SCIEX, 71 Four Valley Drive, Concord, Ontario, L4K 4V8, Canada

† Electronic Supplementary Information (ESI) available as described in the text free of charge. See DOI: 10.1039/x0xx00000x

Please do not adjust margins

contributions: van der Waals (V_{vdW}), ion-induced dipole (V_{IID}), and ion-quadrupole (V_{IQ}) interactions (see equation 4).

$$\Phi(r) = V_{vdW} + V_{IID} + V_{IQ} \quad (4)$$

The dominant term in modelling trajectories arises from the vdW interaction. While traditionally this takes the form of a 12-6 Lennard-Jones (LJ) potential, it was shown that the Exp-6 potential (see Equation 5) employed in the MM3 forcefield yields more accurate CCS predictions.[12,13]

$$V_{vdW}(r_i) = \sum_{i=1}^{n} \varepsilon_i \left[1.84 \times 10^5 \exp\left(\frac{12 r_i}{r_i^*}\right) - 2.25 \left(\frac{r_i^*}{r_i}\right)^6\right] \quad (5)$$

Here, r_i^* is the equilibrium distance between the buffer gas and interaction partner and ε_i is the depth of the potential well. Typically, vdW parameters used in trajectory method calculations are empirically optimized to best reproduce experimentally obtained CCSs in a specific buffer gas. Optimized parameters are available from a handful of sources and are limited to the most common elements (H, C, N, O, and F).[14–16] These parameter sets have limited support for heteroatoms (*i.e.*, those other than H, C, N, O, F); this is often circumvented by utilizing the same parameters for multiple atom types (*e.g.*, assignment of sulphur and phosphorous vdW parameters to those of parameterized atoms such as silicon), or by using vdW parameters from the UFF forcefield.[17–20] In these traditional implementations, atom types for a particular element are treated equivalently (*e.g.,* sp^3 *versus* sp^2 hybridized carbons) This lack of a generalized set of vdW parameters was addressed in 2017 by Lee and coworkers, who incorporated vdW parameters from molecular mechanics forcefields into CCS calculations.[12,21] Errors between experimental and calculated CCSs were ultimately minimized with the use of the Merck Molecular Force Field (MMFF)[22] vdW parameters in the Exp-6 potential.

Inclusion of additional potentials beyond the vdW potential enables a more accurate description of molecular interactions. These additional terms become increasingly prominent components of the ion-neutral potential as the polarizability of the buffer gas increases. The ion-induced dipole interaction is described by Equation 6.[23,24]

$$V_{IID}(r_i) = -\frac{\alpha}{2}\left(\frac{ze}{n}\right)^2 \left[\left(\sum_{i=1}^{n} \frac{x_i}{r_i^3}\right)^2 + \left(\sum_{i=1}^{n} \frac{y_i}{r_i^3}\right)^2 + \left(\sum_{i=1}^{n} \frac{z_i}{r_i^3}\right)^2\right] \quad (6)$$

The terms x_i, y_i, z_i, and r_i define the distance between each of the n atoms and centre of mass of the buffer gas atom/molecule, which has a polarizability α. When the buffer gas possesses a quadrupole moment (e.g., N_2), inclusion of an ion-quadrupole potential is also necessary for accurate calculation of gas trajectories. The quadrupole moment of molecular nitrogen ($4.65 \pm 0.08 \times 10^{-40}$ C·cm^2)[25] is effectively reproduced through separation of charges by negative z on each nitrogen (−0.4825e), which is balanced by a point charge of 2z (+0.965e) at the centre of mass of N_2.[15,23,24] This facilitates a relatively simple calculation of the ion-quadrupole potential by Equation 7, where the j index denotes each partial charge of molecular nitrogen, and i indicates atomic partial charges in the ion of interest.

$$V_{IQ}(r_{ij}) = \sum_{j=1}^{3} \sum_{i=1}^{n} \frac{z_i z_j e^2}{r_{ij}} \quad (7)$$

While advances in instrumentation have enabled high throughput IMS analysis,[26–28] development of an accurate, parallelized architecture for fast numerical calculations of CCSs have progressed at a slower pace. We attribute this to: (i) the lack of a diverse set of experimentally measured compounds for use in the derivation of parameters used for calculating CCSs, (ii) the challenge of conducting molecular dynamics simulations to accurately identify the molecular conformations that are present in probed gas-phase ensembles, (iii) the fact that the accuracy/resolution of traditional experimental techniques was sufficiently supported by older CCS calculation codes, and (iv) the confusion imparted by the multiple patches and inconsistent parameter sets[14–16,23] between the various versions of updated CCS calculation packages. While these issues have been addressed to some degree by Lee and coworkers,[12,21] their parameterization can be further extended to encompass the molecular space required to accurately describe a general set of molecules. It is also worth noting that Zanotto and coworkers recently released a High Performance Collision Cross Section (HPCCS) calculation package, written in C++ and makes use of OpenMP parallelization flags.[29] However, HPCCS employs the (less accurate) 12-6 Lennard Jones (LJ) potential,[12] which uses a {C, H, N, O, F, S} parameter set to describe vdW interactions. Here, we report on a new parallelized CCS calculation suite (MobCal-MPI), which expands on previous codes by including vdW potentials that are better suited for accurate CCS calculations and parameterization for nearly all atom types, and boasts speedups of up to 64-fold compared to the most recent serial implementation.[21] Parameterization of vdW potentials was performed on a diverse set of compounds, whose potential energy landscapes were thoroughly mapped with a custom-built basin-hopping routine prior to refinement with high-level electronic structure calculations. This approach employs an automated workflow, whereby geometries predicted by electronic structure programs (*e.g.* Gaussian) are converted to Mobcal-MPI inputs with minimal user intervention.

Computational Methods

Compound selection and geometric optimization

To ensure the general applicability of the parameterized model, the molecules present in the calibration/training set must exhibit a diverse range of functionalities and atom types. To achieve this, a collection of 162 compounds from various sources reporting on experimental CCSs acquired in N_2 were analyzed.[15–17,21,30–35] The selected compounds, which range in mass from 93 Da to 837 Da, were chosen such that the entire set would encompass most of the functional groups parameterized by the MMFF94 forcefield.[36] Chemical class composition of this calibration set, as determined by ClassyFire,[25] is shown in Figure 1. Analysis shows that the

calibration set contains 9 chemical superclasses, which can be further binned into 42 classes. A full taxonomic breakdown, along with their experimentally obtained CCSs, can be found in the Supporting Information.

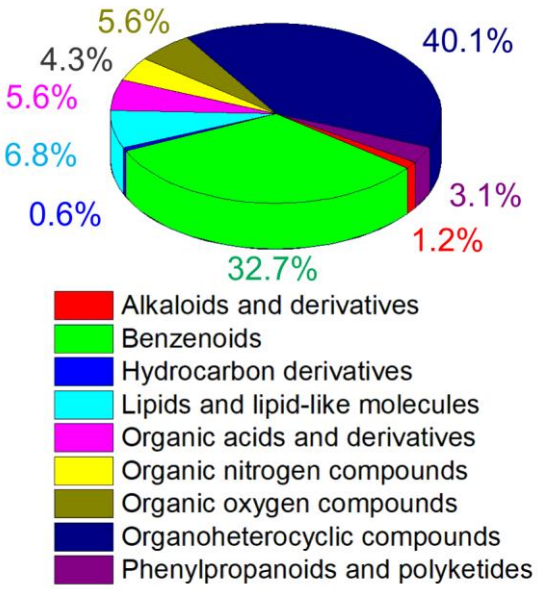

Figure 1. ClassyFire identification of molecular superclasses present in the calibration set.[25]

The ions considered in this study were all protonated molecules produced by electrospray ionization; with that in mind, all possible protonation sites and molecular conformations for each chemical species were explored. In cases where protonation sites or molecular conformations could not be easily assigned by inspection, the potential energy surface (PES) of the ionic species was mapped using a custom-written basin-hopping (BH) script interfaced with Gaussian 16.[37–44] For BH searches, ionic species were modelled using the Universal Force Field (UFF),[45] which utilized partial charges calculated for an optimized "guess structure" at the B3LYP/6-31+G(d,p) level of theory using the ChelpG partition scheme.[46–48] For each random structural perturbation in the BH search, each of the rotatable dihedral angles were assigned a random rotation of $-10° \leq \Phi \leq 10°$. In total, *ca*. 5 000 to 20 000 structures were sampled by the BH algorithm, depending on the size of the ion in question. Typically, the BH routine would identify 5 to 50 low-energy conformers for each prototropic isomer. Reasonable candidate structures were selected for higher level calculation based on energetics and chemical intuition. Candidate species were pre-optimized at the semi-empirical PM6 level of theory.[49] Protonation sites and molecular conformations within relative energies of 50 kJ mol^{-1} of the PM6 global minimum were carried forward for treatment at the B3LYP/6-31++G(d,p) level of theory. DFT calculations included the GD3 empirical correction for dispersion.[50] Iodine was treated with the Def2-TZVP basis set and effective core potential.[51,52] Normal mode analyses were conducted to verify that each isomer corresponded to a minimum on the PES. This also served to calculate the gas-phase thermochemistry for each structure, which was used in the ordering of the Gibbs-corrected electronic energies of each unique tautomer/conformer. Atomic partial charges were generated according to the Merz-Singh-Kollman (MK) partition scheme,[53,54] and constrained to reproduce the dipole moment of the isomer. All isomers found within 20 kJ mol^{-1} of the global minimum structure were carried forward to obtain theoretical CCSs via MobCal-MPI.

CCS calculations

Theoretical N_2 CCSs (Ω_{N2}) were calculated using the trajectory method.[10] All CCS calculations employed 10 complete cycles of mobility calculations that used 48 points of velocity integration and 512 points of impact parameter integration. Calculated Ω_{N2} are reported as average values with statistical errors assessed from the ten cycles of calculation. For ions exhibiting multiple low-energy conformations or prototropic isomers, a Boltzmann-weighted CCS is reported based on the standard Gibbs corrected energies (T = 298 K). Total error for a corresponding Boltzmann-weighted CCS are calculated from the standard errors for the Boltzmann-weighted CCSs of the low-energy isomers.

From the previous update of the MobCal code by Lee and coworkers,[12,21] a combination of the Exp-6-type potential (from the MM3 forcefield)[13] and MMFF94 parameters were shown to reasonably reproduce the various ion-buffer gas vdW interaction potentials. We have adapted this modified source code, generously provided by Lee et al., to read in customized atom types from MM forcefields (*e.g.,* MMFF94). This modification enables the differentiation of chemical moieties that exhibit dissimilar N_2 vdW interaction potentials based on their chemical environment (*e.g.*, amido nitrogen atoms *versus* amino nitrogen atoms). Atom type parameters can be read directly from the input file. MMFF94 interaction parameters were combined with the N_2 parameters deduced from those of monoatomic nitrogen (N) following the combination rules outlined for the forcefield.[36] To conveniently prepare input files for CCS calculations, a python module that nests the Open Babel 2.3.2[55] and sdf2xyz2sdf[56] packages was built to fully automate the MMFF94 atom type assignments and compile partial charges predicted by DFT. Although this module does streamline CCS calculations, manual verification of atom type assignments prior to submission is recommended. All up-to-date packages are freely available for download on GitHub (https://github.com/HopkinsLaboratory/MobCal-MPI) or the Hopkins group website.

Application of Forcefields and the Optimization of Scaling Factors

MMFF94 vdW parameters were obtained from the original publication.[22] In this forcefield, atom types are defined by four unique parameters (α_i, A_i, N_i, and G_i) related to physicochemical properties of the atom. Combination of these parameters with those of molecular nitrogen yields the values of r_{ij}* and ε_{ij} used in this study. Since the interaction partner in this work is always N_2, we simply refer to r_i* and ε_i. Although it is possible to fit

individual parameters, the most straight-forward procedure (owing to the rather complex combination rules associated with the MMFF94 forcefield) was to fit final r_i* and ε_i values. Since the atom types outlined in the forcefield are inherently parameterized, optimization of vdW parameters can be easily accomplished through the application of a uniform scaling factor to the distance (r_i*) and energy (ε_i) related terms as per equations 8A and 8B.

$$r^*_{i\ scaled} = r^*_i \times \rho_{dist} \quad (8A)$$

$$\varepsilon_{i\ scaled} = \varepsilon_i \times \rho_{ener} \quad (8B)$$

For *i* = C, H, O, N, F, P, S, Cl, Br, I, …

Optimization of scaling factors was completed iteratively, screening combinations of scaling factors between 0.70 and 1.00. For the final iteration, scaling factors were stepped in increments of 0.01 to ensure optimal parameterization of vdW interaction potentials. It should be noted that this procedure returned numerous scaling factors that produced similar root mean square errors (RMSEs), indicating that a comprehensive search of the scaling factor surface had to be completed.

Results and Discussion

Optimization of MobCal-MPI

Prior to this work, the updated MobCal code was only available for serial execution.[21] This severely limited its performance on High Performance Computing (HPC) architectures capable of parallelization. To remedy this, we have modified the original code to improve its computational efficiency. Briefly, modifications to the updated MobCal code now include: (i) message passing interface (MPI) to support parallel computing architectures, (ii) replacement of *goto* looping with *do* and *endo* functionality, enabling more efficient optimization, (iii) significant reworking of the *dljpot* and *mobil2* subroutines, including loop reorganization, removal of repetitive calculations of identical arrays and unused variables, (iv) consistent double precision used throughout, and (v) comments and patch notes for future use and debugging purposes. When operating on a single core (i.e., in serial), our re-worked MPI code was found to perform 2.9-fold more efficiently than the serial implementation developed by Lee and coworkers.[21] It should be noted that there are miniscule differences in CCSs calculated with the two codes owing to differences created by the pseudo-random number generator RANLUX.

Extensive benchmarking of the new parallelized code has been conducted on species ranging in size from 9 to 7,029 atoms (135 Da to 50 kDa). Benchmarking was performed on the Graham HPC cluster, which is part of the Sharcnet consortium of Compute Canada. Each CPU node is dual socket, equipped with 2 x 16 core Intel E5-2683 v4 chipsets, operating at 2.1 GHz for a total of 32 cores per node. Benchmarking was performed against serial execution of MobCal-MPI and the non-parallelized implementation developed by Lee and coworkers.[21] Speedups and overall efficiency are shown in Figure 2 and tabulated in Tables S1 and S2. Our updated MPI code substantially reduces average CCS computation time, providing 36- and 64-fold increases in efficiency when deployed on 16 and 32 cores, respectively, relative to the most recent serial version of MobCal.

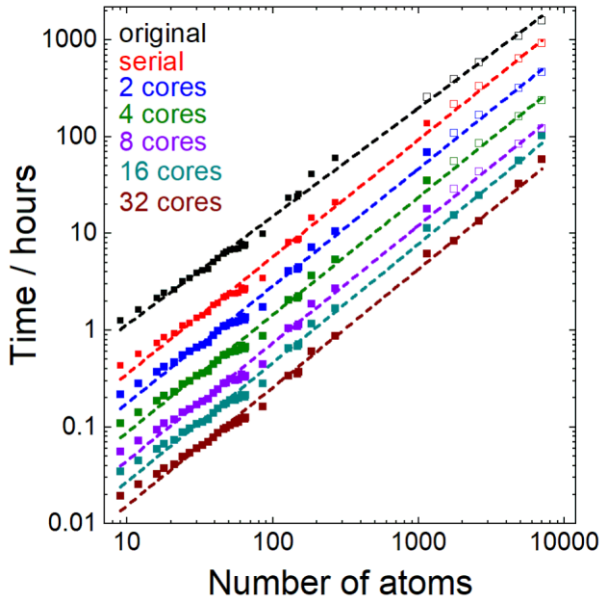

Figure 2. Total execution time in hours for numerous species ranging in size from 9 to 7,029 atoms using the most recent serial MobCal code (black),[21] and MobCal-MPI using 1 (red, serial), 2 (blue), 4 (green), 8 (purple), 16 (teal), and 32 (maroon) CPU cores for calculation. Open squares indicate extrapolated points.

Identification of Scaling Factors

To assess the CCS calculation accuracy, accurate structural models must be accompanied by reproducible CCS measurements. To accomplish this, calibrant ions with highly reproducible experimental CCS measurements were selected. The conformational space of these species was then mapped using the BH algorithm at the molecular mechanics level and refined using DFT. Note that all CCS values were obtained using N_2 as a collision gas. With refined structural models, multiple low-energy isomers for a given calibrant ion were submitted for CCS calculation with the new parallelized code. Figure 3 shows a contour plot of the RMSEs obtained through a comparison of experimental CCS with the Boltzmann-weighted CCS of each calibrant ion at specific values of the linear scaling parameters ρ_{dist} and ρ_{ener}. While numerous combinations of scaling factors return RMSEs of <3 %, the combination of ρ_{dist} = 0.78 and ρ_{ener} = 0.80 returns a minimum RMSE of 2.60 %. Note that the parallelization, benchmarking, and optimization of the vdW parameters that we report here have been conducted for CCS calculations employing the trajectory method, which is generally accepted as the 'gold standard' for computed CCS values.[2,15,57–60] Further to this, the option of utilizing helium as the collision partner is available in MobCal-MPI, but we have not

completed an extensive benchmarking or refitting of scaling parameters beyond the previous work of Lee et al.[12]

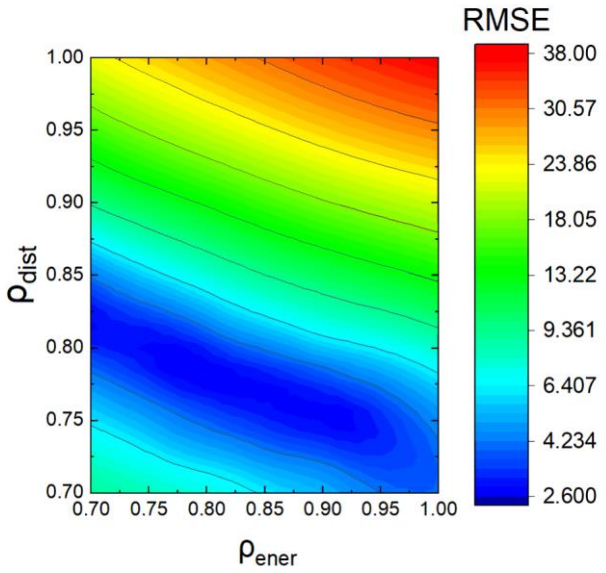

Figure 3. Contour plot indicating RMSE of the calibration set at specific combinations of the scaling parameters ρ_{dist} and ρ_{ener}.

Evaluation of Accuracy and Trajectory Method Standard Deviations

Empirical optimization methods are often accompanied by systematic bias towards the calibration set, even if the calibration set encompasses a structurally diverse set of compounds. Thus, it is necessary to test our scaled parameters on an equally diverse chemical set that does not contain compounds found in the calibration set. To this end, we constructed a validation set consisting of 50 compounds from the same compendiums, which fall within 8 molecular superclasses and 18 molecular classes based on ClassyFire assignment (Figure S1). Figures 4A and 4B show the correlations between calculated and measured CCS for both the calibration and validation sets; CCS calculations for both sets return excellent accuracy (2.60 % and 2.31 % RMSE, respectively) and no outliers can be readily identified.

An additional factor regarding the accuracy of the CCS calculations stems from the trajectory method itself. To reiterate, the CCS depends on the orientation of the molecule relative to the buffer gas, the relative impact velocity of the buffer gas, and the scattering angle. Each of these are evaluated through repeated simulations of the collision trajectories between the ion and gas molecules, where the user can specify a definitive number of velocity and impact parameter (scattering angle) integrations to be performed. The number of velocity (*inp*) and impact parameter (*imp*) integrations could potentially be optimized, but it is useful to constrain the number of integration points such that they can be efficiently distributed across parallel architecture. For example, the recommended number of points for velocity integration is 40, but this does not scale well with traditional number of cores commonly employed (*i.e.*, 8 or 16 cores). For that reason, we recommend that 48 points of velocity integration are used for even distribution across cores.

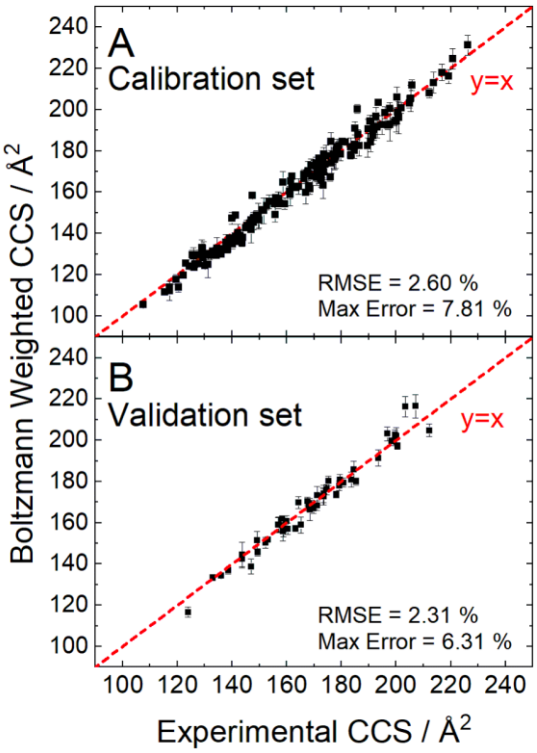

Figure 4. Comparison of Boltzmann-weighted CCSs from the calibration set (**A**) and validation set (**B**) to experiment.

To investigate the effect of changing the frequency of impact parameter integration, we calculated the CCS of the validation set using 112, 256, 512, 752, and 1008 *imp* integration points. The results of this investigation are plotted in Figure 5, which shows that CCS values are relatively unaffected by the choice of *imp* integration frequency (RMSE varies between 2.20 % and 2.36 %). However, significant variance in a single CCS calculation is observed when *imp* is set below 500 points. This is somewhat concerning for prior parameterization work, which employed 25 – 100 *imp* integration points.[12,21] When an *imp* of 112 is specified, a mean standard deviation of 3.87 Å^2 and a maximum standard deviation of 8.22 Å^2 was found for our chosen set of compounds. In contrast, when 1008 *imp* integration points were used, both the mean and maximum standard deviations decreased to 1.20 Å^2 and 2.36 Å^2, respectively. There is, of course, a trade-off in terms of computational time when increasing the size of the integration grid; using 1008 *imp* points takes 10 times longer than using 112 points. Our testing suggests that 512 *imp* integration points offers a significantly reduced standard deviation in the determination of an ion's CCS with only a moderate increase in computational cost compared to smaller grid sizes. Of course, this number must increase as ion size increases beyond those in the calibration/validation set to fully explore the surface of the ion's collision landscape.

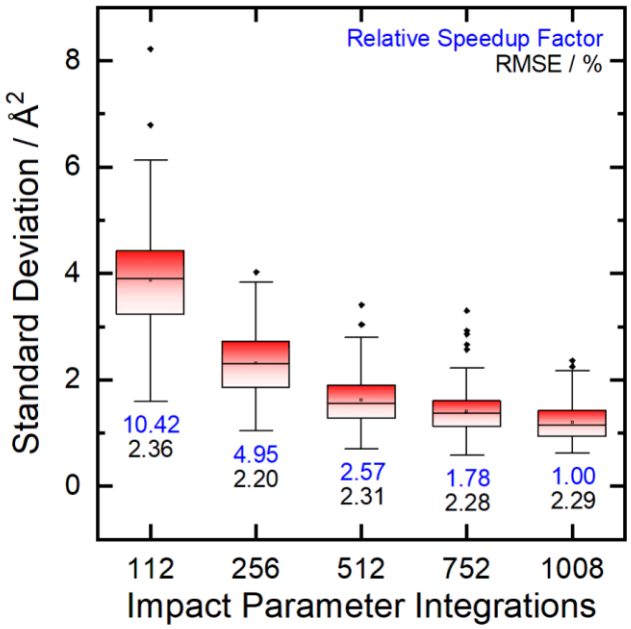

Figure 5. Box plot indicating the effect of the number of impact parameter integration points on the relative standard deviation of an individual CCS calculation for all isomers within the validation set. The horizontal line and dot within the box represent the respective median and mean relative standard deviation, the boundaries of the box indicate the 25th to 75th percentile (interquartile range; precision), and the whiskers indicate the highest and lowest observed relative standard deviations. Outliers that are outside of 1.5 times the interquartile range are shown as black dots. Speedups in calculation time are reported relative to the times for completion using 1008 *imp* points in blue, and RMSE (accuracy) for each *imp* routine is shown in black.

Influence of charge neutral interactions

Since ion-induced dipole (V_{IID}) and ion-quadrupole (V_{IQ}) interactions are significant contributors to the total ion-N_2 interaction potential, a rigorous partial charge assignment scheme must be employed. Numerous methodologies for the assignment of atomic partial charges exist; however those that determine partial charges by fitting point charges to reproduce the electrostatic potential of the molecule are generally accepted to be the most accurate (e.g., CHelpG, MK).[48,53,54] Calculated CCSs can vary dramatically when alternative atomic charge assignment schemes are used (e.g., those based on MM forcefield assignments or Mulliken charges). In principle, one could avoid this issue and decrease computational time by ignoring V_{IID} and V_{IQ} terms. However, caution must be exercised in this regard because functional groups centered around the same atom type can have similar vdW interaction descriptions but may possess dissimilar V_{IID} and V_{IQ} terms. For instance, the nitrogen atom in an amide exhibits very different electronic character (*viz*. partial charge) than does a sulfonamide nitrogen atom due to donation/resonance effects with carbonyl *versus* sulfonyl groups. However, the nitrogen atoms in the amide and sulfonamide groups have nearly identical vdW parameters and means that differentiation can only occur through atomic charge. Owing to the dependence of atomic charge on V_{IID} and V_{IQ} used to evaluate Ω_{N2}, TM calculations should employ partial charge assignments based on electrostatic potential mapping as opposed to the less reliable Mulliken population analysis or assignment of equivalent partial charges to all atoms in a particular ion (equal). Figure 6 shows impact of partial charge assignment on the calculated CCSs of the validation set with unique partial charges.

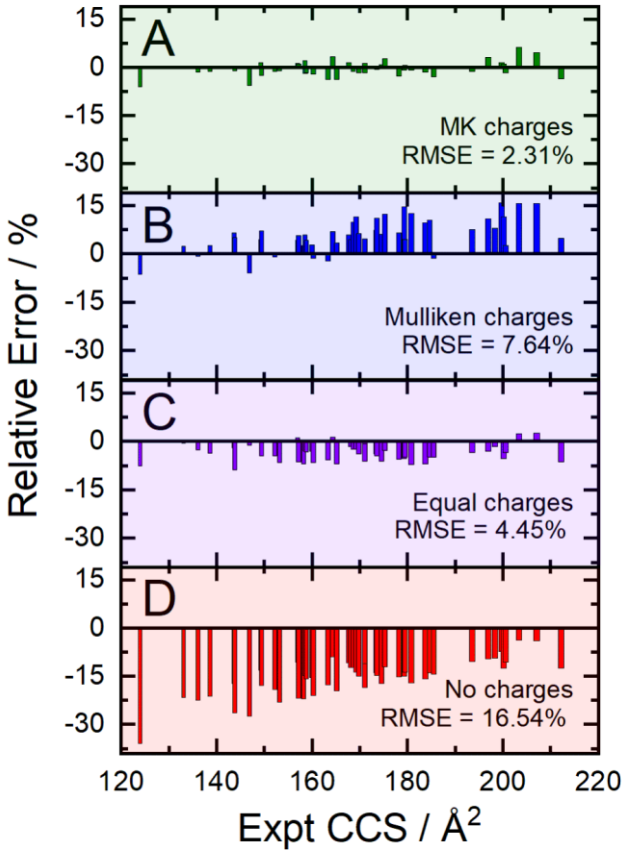

Figure 6. Relative error (%) in CCS predictions associated with assignment of partial charges by various methodologies: (**A**) MK charges constrained to reproduce dipole moment (green), (**B**) Mulliken charges (blue), (**C**) equal partial charges (purple), and (**D**) no partial charges (red).

Analysis of CCSs computed with the newly optimized scaling factors for the vdW potentials indicates that optimal accuracy is achieved when partial charges are assigned based on electrostatic potential mapping (i.e., CHelpG, MK), with the stipulation that partial charges be constrained to reproduce the molecular dipole moment. For example, the RMSE of CCSs determined using Mulliken and equal charge assignment schemes are 7.63 % and 4.45 %, respectively. Note that this effect is likely to be more important for calculations of Ω_{N2} than Ω_{He} owing to the relatively large polarizability of molecular nitrogen. We recommend that partial charge calculations should be computed at the B3LYP/6-31++G(d,p) level of theory

to maintain consistency with the calibration set. However, a change in basis set should not have a major effect unless the relative energies of local minima are significantly changed, since this would influence Boltzmann-weighted CCSs. Table S4 shows that CCSs determined using the ChelpG or MK partition schemes with different basis sets all lie within the standard deviation of the MobCal-MPI calculation itself.

It has been previously suggested that inclusion of the V_{IQ} term in the ion-N_2 potential could be circumvented by fitting appropriate scaling factors to the V_{vdW} potential.[16] While exclusion of V_{IQ} would be beneficial in terms of reducing computational cost, its exclusion will also reduce the accuracy of the calculated CCS values. To quantify the importance of the various terms in the interaction potential, we explored the effect of removing the V_{IID} and V_{IQ} terms on computational time, calculation accuracy, and trends with experimental values. Removal of both V_{IID} and V_{IQ} returned an average speedup of 2.6-fold, while removal of V_{IQ} returned a speedup of 1.2-fold (Table S3). In terms of accuracy, the red bars in Figure 6D show that removal of both V_{IID} and V_{IQ} terms returns a RMSE of 16.54 %. Figures S2 and S3 shows that individual removal of V_{IID} and V_{IQ} terms yield RMSEs of 13.38 % and 3.80 %, respectively. Based on the data, it is unlikely that removal of either potential can be compensated by modifications to the vdW potential for small ions. Moreover, the slight gain in computational efficiency results in significantly larger variance of CCS measurements, in addition to increased deviation from experiment. While there is some merit to replacement of the V_{IQ} potential with appropriate vdW parameters, linear regression indicates a significant non-unity relationship between CCSs predicted by MobCal-MPI and experiment (Figure S2). This correlates with the observation that the measurable difference in experimental Ω_{N2} of diastereomeric species,[15] tertiary/quaternary ammonium species,[23,24] and proteins[61] depends on accurate modelling of short- (V_{vdW}) and long-range (V_{IID} and V_{IQ}) interactions using appropriate potentials. If such a method were desired, individual empirical optimization of each atom type in the vdW potential would be required, which is not feasible owing to the set of 100 unique vdW parameters defined in the MMFF94 forcefield.

The exclusion of the V_{IID} and V_{IQ} terms creates a dependence of error on ion size; deviations from experimental CCSs diminish as ion size increases. In other words, V_{IID} and V_{IQ} are particularly important for accurate calculation of small molecule CCSs. This result can be rationalized in terms of surface exposure and has implications in the development of a method for macromolecular CCS calculations. In the case of small species, which effectively expose all atoms at the outer molecular surface, any changes to parameters that affect ion radii (*i.e.*, V_{vdW}, V_{IID}, and V_{IQ} terms) will have a substantial influence on buffer gas trajectories (and CCS by association). As the size of the molecular entity increases, atoms can become buried within the effective ion core, and thus will exhibit little influence on buffer gas trajectories. Continual ion growth will eventually reach a size regime where only a small fraction of the total number of atoms is exposed at the molecular surface. As this threshold is reached, changes in effective ion radii will minutely increase the size of the exposed surface and will essentially leave the volume of the ion core unchanged. Bearing this in mind, we expect the effect of corrections from V_{IID} and V_{IQ} terms to diminish to the point where macromolecular ions (e.g., proteins, DNA/RNA, etc.) can be treated relatively accurately using only a vdW interaction potential.

Comparison of MobCal-MPI to other calculation methods

Recent years have seen the emergence of machine learning (ML) methodologies being employed to predict CCSs. At the forefront of these methods is MetCCS,[62] an online tool developed by the Zhu group that makes predictions of CCS based on molecular descriptors (i.e., molecular weight, pKa, log P, polar surface area, etc.). While ML methods are far superior in terms of computational efficiency, they are challenged in cases where they must distinguish between isomers exhibiting nearly identical structural properties (e.g cis/*trans* isomers, tautomers). Moreover, ML methods are subject to bias if the training set is not of sufficient size or scope. While the training set for MetCCS contains 796 unique chemical entities, molecular properties were derived from computational predictions (e.g., ChemAxon, ALOGPS) in place of experimental values, and could be subject to error. Figure S5 shows that MetCCS was able to predict the CCS of each molecule in the validation set with a RMSE of 4.63 %. However, given the maximum deviation of 12 % for MetCCS predictions, CCS calculation approaches based on structural modelling and interaction potentials should be employed when accuracy is of the utmost importance.

With regard to structure-based approaches for calculating CCS, there are a variety of computational packages available. The most recent implementation that provides a significant speedup over the traditional MobCal package is HPCCS, which performs CCS calculations using identical numerical methodologies to both MobCal and MobCal-MPI.[29] HPCCS is a C++ analog that utilizes the traditional 12-6 LJ potential to model vdW interactions and boasts 48-fold speedups on 16 cores compared to the original MobCal implementation of Shvartsburg and Jarrold (using helium as the collision partner).[9,10] For calculations in N_2, HPCCS operates similarly to an additional parallelized package (Collidoscope),[63] in that both suites treat N_2 as a quasi-spherical entity and model CCS by defining the momentum transfer integral over various ion-buffer gas orientations and velocities. While this methodology has been well established, an additional package (IMoS) provides an alternative treatment of momentum transfer by employing a control-volume conservation of momentum approach.[16,64,65] IMoS allows for the treatment of ion-neutral collisions as either specular or diffuse. While MobCal-MPI does not currently have this ability, we note that the difference between CCS calculations from treating collisions as either diffuse or specular is minor and computational time nearly identical when the appropriate potential interactions are

included.[65] Of course, we envision there are select cases where modelling collisions as diffuse is beneficial (*e.g.,* Ω_{air} evaluations).[66–68] For the most part however, effects on collision trajectories induced by molecular rovibrational motion and/or diffuse scattering in a uniform buffer gas are effectively embedded within scaling parameters derived during the empirical optimization procedure. Upon further inspection, IMoS CCS calculations in N_2 provide speedups over the traditional MobCal code similar to that of HPCCS (*vide infra*). Since the CCS predictions of IMoS and the original MobCal agree to within < 1%,[65] and both the HPCCS and Collidoscope methods to evaluate CCSs are analogous to MobCal, a comparison of MobCal-MPI with Collidoscope and IMoS would be redundant. Thus, we sought to compare our calculation method to the most recent code (HPCCS) to assess relative accuracy and efficiency and infer comparison to the Collidoscope and IMoS CCS packages.

The LJ parameters available in HPCCS were taken from literature,[65] which were parameterized on a small set of molecules containing C, H, O, N, and F. An extension of these parameters developed by our group was included to facilitate CCS calculations of additional heteroatomic compounds (*i.e.*, P, S, Cl, Br), as these were not available in the original publication. These additional parameters were manually fit to ion mobility data previously analyzed in our group using a similar methodology outlined by Wu and coworkers.[16] A description of the methodology and optimized LJ parameters (Table S5) can be found in the supporting information. A comparison between the accuracies of the two methods on the validation set is shown in Figure 7. Utilization of the 12-6 LJ potential without atom specific vdW parameters yields a RMSE of 3.82 %, with a maximum error of 11.03 %. The increased deviation from experiment for HPCCS, especially for CCSs greater than *ca.* 175 Å^2, is consistent with the results of Zanotto and coworkers, who report a RMSE of 6.50 % and a maximum error of 14.2 % when using the trajectory method to evaluate CCSs of similar molecular entities in He. In comparison to the 12-6 LJ potential and parameters of HPCCS, the MobCal-MPI methodology is superior in accuracy for theoretical evaluation of CCSs.

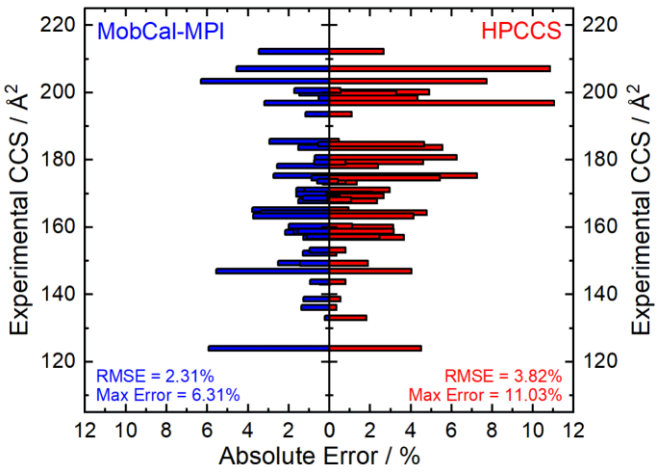

Figure 7. Comparison of the absolute error between experimental and calculated Boltzmann-weighted CCS for (**A**) MobCal-MPI (Exp-6 potential) and (**B**) HPCCS (12-6 LJ potential) for the validation set.

In terms of computational efficiency, comparison between MobCal-MPI and HPCCS was complicated owing to the use of unique vdW potentials. Evaluation of the Exp-6 type potential in programming languages requires series expansion of the exponential function (base *e*) via the infinite sum Maclaurin series. This is in direct contrast to the 12-6 LJ potential, which is evaluated through repeated multiplication to generate the appropriate powers. Thus, MobCal-MPI will inherently be a slower method owing to the use of an alternative potential that contains an exponential. Bearing this in mind, HPCCS performs 2.3-fold more efficiently on average across equivalent core allocations relative to MobCal-MPI (see Figure S6).

Without a doubt, the HPCCS framework should be utilized in the development of the next generation of trajectory method calculators owing to its optimal performance. This is reinforced by its impressive retention of efficiency upon use in parallel HPC architectures. Tables S6 and S7 indicate that HPCCS boasts 92- and 175-fold increases of efficiency when used on 16 and 32 cores compared to the MobCal implementation of Lee et al,[21] with respective performance efficiencies of 96 % and 91 %. However, for improved accuracy, use of the Exp-6 potential is recommended when calculating CCSs via the trajectory method as it has been shown to more consistently reproduce experimental results relative to other potentials.[12] Thus, future implements on the highly optimized framework of HPCCS (or IMoS/Collidoscope) would undoubtedly benefit from the implementation of both the parameterized, atom specific parameters and changes to code structure reported in this manuscript.

Conclusions

Based on a comparison between high-quality experimental data and structural models generated with high-level quantum-chemical calculations, we present a set of refined, atom specific vdW parameters for use in MobCal-MPI. MobCal-MPI is a refined suite of the commonly used MobCal code that is suitable for deployment on parallelized HPC architectures, and exhibits speedups of up to 64-fold over the most recent version when used to determine Ω_{N2}.[21] Calibration of vdW parameters minimized the RMSE between MobCal-MPI and experiment to 2.60 % for a calibration set of 162 molecules. The refined vdW parameters yield a RMSE of 2.31% when used to calculate CCSs on a structurally diverse validation set consisting of 50 molecules.

The vdW parameters are complimented by ion-induced dipole and ion-quadrupole potentials, which make use of atomic partial charges. Potential scaling factors were optimized using partial charges calculated at the B3LYP/6-31++G(d,p) level of theory using the Merz-Kollman-Singh scheme, and constrained

to reproduce the molecular dipole moment. This should be kept consistent in future uses of MobCal-MPI or analogous calculation packages to ensure retention of accuracy. This is also applicable for basis set selection, although CCS calculations should not be expected to vary appreciably if isomer relative energies are consistent between calculation methods.

The MobCal-MPI methodology performs CCS evaluations on static ion cores. Of course, full treatment of atomic/ion motion, including allowing molecules to rotate/vibrate would be the most rigorous methodology. However, this would require order(s) of magnitude more computational time to evaluate ion motion at specific time steps, especially over the course of > 10^6 trajectory calculations. We approximate this here by conducting a basin-hopping search of the potential energy surface for ions and employing a Boltzmann-weighted average CCS for compounds that exhibit numerous low energy isomers/conformers (as optimized at the DFT level of theory). Our methodology is supported by numerous studies who approximate molecular motions through targeted molecular dynamics or PES searches, where inclusion of population-averaged (*i.e.,* Boltzmann-weighted) CCSs results in improved matches with experiment.[12,21,69,70]

While no method is free of systematic bias, the compounds present in the calibration and validation set were carefully selected to exhibit a diverse set of functional groups that encompass the molecular space defined by the MMFF94 forcefield. The derivation of vdW parameters relied on rigorous structural modelling of compounds and can be regarded as a comprehensive exploration of the vdW potential surface. We expect the evaluation of vdW parameters outlined here to serve as the framework for future development of CCS calculation packages. This is particularly relevant for future IM-MS applications, which now contains a computational avenue for CCS calculation with high accuracy and efficiency and could be extended for use towards macromolecular species.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Acknowledgements

We gratefully acknowledge the high-performance computing support from the SHARCNET consortium of Compute Canada. We are also indebted to Dr. Jong Wha Lee, Dr. Hugh I. Kim, and coworkers for providing us with the serial version of their modified MobCal code. The authors would like to acknowledge the financial support provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Notes and references

1 D. C. Collins and M. L. Lee, *Anal. Bioanal. Chem.*, 2002, **372**, 66–73.
2 C. Lapthorn, F. Pullen and B. Z. Chowdhry, *Mass Spectrom. Rev.*, 2013, **32**, 43–71.
3 G. Paglia and G. Astarita, *Nat. Protoc.*, 2017, **12**, 797–813.
4 J. R. Verkouteren and J. L. Staymates, *Forensic Sci. Int.*, 2011, **206**, 190–196.
5 R. Cumeras, E. Figueras, C. E. Davis, J. I. Baumbach and I. Gràcia, *Analyst*, 2015, **140**, 1376–1390.
6 F. Lanucara, S. W. Holman, C. J. Gray and C. E. Eyers, *Nat. Chem.*, 2014, **6**, 281–294.
7 T. Cajka and O. Fiehn, *Anal. Chem.*, 2016, **88**, 524–545.
8 E. A. Mason and E. W. McDaniel, *Transport properties of ions in gases*, John Wiley and Sons, New York, 1988.
9 T. Wyttenbach, G. Von Helden, J. J. Batka, D. Carlat and M. T. Bowers, *J. Am. Soc. Mass Spectrom.*, 1997, **8**, 275–282.
10 A. A. Shvartsburg and M. F. Jarrold, *Chem. Phys. Lett.*, 1996, **261**, 86–91.
11 Z. Li and H. Wang, *Phys. Rev. E*, 2003, **68**, 61206.
12 J. W. Lee, K. L. Davidson, M. F. Bush and H. I. Kim, *Analyst*, 2017, **142**, 4289–4298.
13 J. H. Lii and N. L. Allinger, *J. Am. Chem. Soc.*, 1989, **111**, 8576–8582.
14 C. K. Siu, Y. Guo, I. S. Saminathan, A. C. Hopkinson and K. W. M. Siu, *J. Phys. Chem. B*, 2010, **114**, 1204–1212.
15 I. Campuzano, M. F. Bush, C. V. Robinson, C. Beaumont, K. Richardson, H. Kim and H. I. Kim, *Anal. Chem.*, 2012, **84**, 1026–1033.
16 T. Wu, J. Derrick, M. Nahin, X. Chen and C. Larriba-Andaluz, *J. Chem. Phys.*, 2018, **148**, 74102.
17 G. Paglia, J. P. Williams, L. Menikarachchi, J. W. Thompson, R. Tyldesley-Worster, S. Halldórsson, O. Rolfsson, A. Moseley, D. Grant, J. Langridge, B. O. Palsson and G. Astarita, *Anal. Chem.*, 2014, **86**, 3985–3993.
18 H. Lavanant, V. Tognetti and C. Afonso, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 572–580.
19 S. Poyer, C. Lopin-Bon, J. C. Jacquinet, J. Y. Salpin and R. Daniel, *Rapid Commun. Mass Spectrom.*, 2017, **31**, 2003–2010.
20 X. Zhang, J. Krechmer, M. Groessl, W. Xu, S. Graf, M. Cubison, J. T. Jayne, J. L. Jimenez, D. R. Worsnop and M. R. Canagaratna, *Atmos. Chem. Phys.*, 2016, **16**, 12945–12959.
21 J. W. Lee, H. H. L. Lee, K. L. Davidson, M. F. Bush and H. I. Kim, *Analyst*, 2018, **143**, 1786–1796.
22 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 520–552.
23 H. Kim, H. I. Kim, P. V. Johnson, L. W. Beegle, J. L. Beauchamp, W. A. Goddard and I. Kanik, *Anal. Chem.*, 2008, **80**, 1928–1936.
24 H. I. Kim, H. Kim, E. S. Pang, E. K. Ryu, L. W. Beegle, J. A. Loo, W. A. Goddard and I. Kanik, *Anal. Chem.*, 2009, **81**, 8289–8297.
25 Y. Djoumbou Feunang, R. Eisner, C. Knox, L. Chepelev, J. Hastings, G. Owen, E. Fahy, C. Steinbeck, S. Subramanian, E. Bolton, R. Greiner and D. S. Wishart, *J. Cheminform.*, 2016, **8**, 1–20.
26 I. K. Webb, S. V. B. Garimella, A. V. Tolmachev, T.-C. Chen, X. Zhang, R. V. Norheim, S. A. Prost, B. LaMarche, G. A. Anderson, Y. M. Ibrahim and R. D. Smith, *Anal. Chem.*, 2014, **86**, 9169–9176.

27 C. D. Chouinard, G. Nagy, I. K. Webb, T. Shi, E. S. Baker, S. A. Prost, T. Liu, Y. M. Ibrahim and R. D. Smith, *Anal. Chem.*, 2018, **90**, 10889–10896.

28 T. C. Chen, Y. M. Ibrahim, I. K. Webb, S. V. B. Garimella, X. Zhang, A. M. Hamid, L. Deng, W. E. Karnesky, S. A. Prost, J. A. Sandoval, R. V. Norheim, G. A. Anderson, A. V. Tolmachev, E. S. Baker and R. D. Smith, *Anal. Chem.*, 2016, **88**, 1728–1733.

29 L. Zanotto, G. Heerdt, P. C. T. Souza, G. Araujo and M. S. Skaf, *J. Comput. Chem.*, 2018, **39**, 1675–1681.

30 L. Righetti, A. Bergmann, G. Galaverna, O. Rolfsson, G. Paglia and C. Dall'Asta, *Anal. Chim. Acta*, 2018, **1014**, 50–57.

31 K. M. Hines, D. H. Ross, K. L. Davidson, M. F. Bush and L. Xu, *Anal. Chem.*, 2017, **89**, 9023–9030.

32 J. Regueiro, N. Negreira and M. H. G. Berntssen, *Anal. Chem.*, 2016, **88**, 11169–11177.

33 L. Bijlsma, R. Bade, A. Celma, L. Mullin, G. Cleland, S. Stead, F. Hernandez and J. V. Sancho, *Anal. Chem.*, 2017, **89**, 6583–6589.

34 X. Zheng, N. A. Aly, Y. Zhou, K. T. Dupuis, A. Bilbao, V. L. Paurus, D. J. Orton, R. Wilson, S. H. Payne, R. D. Smith and E. S. Baker, *Chem. Sci.*, 2017, **8**, 7724–7736.

35 A. Bauer, J. Kuballa, S. Rohn, E. Jantzen and J. Luetjohann, *J. Sep. Sci.*, 2018, **41**, 2178–2187.

36 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.

37 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.

38 C. Ieritano, P. J. J. Carr, M. Hasan, M. Burt, R. A. Marta, V. Steinmetz, E. Fillion, T. B. Mcmahon and W. S. Hopkins, *Phys. Chem. Chem. Phys.*, 2016, **18**, 4704–4710.

39 W. S. Hopkins, R. A. Marta, V. Steinmetz and T. B. McMahon, *Phys. Chem. Chem. Phys.*, 2015, **17**, 28548–28555.

40 C. Liu, J. C. Y. Le Blanc, J. Shields, J. S. Janiszewski, C. Ieritano, G. F. Ye, G. F. Hawes, W. S. Hopkins and J. L. Campbell, *Analyst*, 2015, **14**, 6897–6903.

41 M. Burt, K. Wilson, R. Marta, M. Hasan, W. Scott Hopkins and T. McMahon, *Phys. Chem. Chem. Phys.*, 2014, **16**, 24223–24234.

42 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, B. V. Barone, G. Mennucci and A. Petersson, Gaussian 16, Revision A.03, Gaussian, Inc., Wallingford CT, 2016.

43 C. Liu, J. C. Y. Le Blanc, B. B. Schneider, J. Shields, J. Federico, H. Zhang, G. W. Kauffman, D. Kung, C. Ieritano, M. Verbuyst, L. Melo, M. Hasan, D. Nasar, J. S. Janiszewski, W. S. Hopkins and J. L. Campbell, *ACS Cent. Sci.*, 2017, **3**, 101–109.

44 C. Ieritano, J. Featherstone, P. J. J. Carr, R. A. Marta, E. Loire, T. B. McMahon and W. S. Hopkins, *Phys. Chem. Chem. Phys.*, 2018, **20**, 26532–26541.

45 A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

46 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

47 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

48 C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, 1990, **11**, 361–373.

49 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–1213.

50 J. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.

51 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

52 K. A. Peterson, *J. Chem. Phys.*, 2003, **119**, 11099–11112.

53 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1984, **5**, 129–145.

54 B. H. Besler, K. M. Merz and P. A. Kollman, *J. Comput. Chem.*, 1990, **11**, 431–439.

55 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminform.*, 2011, **3**, 1–14.

56 P. Tosco, T. Balle and F. Shiri, *J. Mol. Model.*, 2011, **17**, 3021–3023.

57 S. Akashi and K. M. Downard, *Anal. Bioanal. Chem.*, 2016, **408**, 6637–6648.

58 J. Boschmans, S. Jacobs, J. P. Williams, M. Palmer, K. Richardson, K. Giles, C. Lapthorn, W. A. Herrebout, F. Lemière and F. Sobott, *Analyst*, 2016, **141**, 4044–4054.

59 G. B. Gonzales, G. Smagghe, S. Coelus, D. Adriaenssens, K. De Winter, T. Desmet, K. Raes and J. Van Camp, *Anal. Chim. Acta*, 2016, **924**, 68–76.

60 C. Lapthorn, F. S. Pullen, B. Z. Chowdhry, P. Wright, G. L. Perkins and Y. Heredia, *Analyst*, 2015, **140**, 6814–6823.

61 D. Canzani, K. J. Laszlo and M. F. Bush, *J. Phys. Chem. A*, 2018, **122**, 5625–5634.

62 Z. Zhou, X. Shen, J. Tu and Z. J. Zhu, *Anal. Chem.*, 2016, **88**, 11084–11091.

63 S. A. Ewing, M. T. Donor, J. W. Wilson and J. S. Prell, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 587–596.

64 C. Larriba and C. J. Hogan, *J. Comput. Phys.*, 2013, **251**, 344–36.

65 V. Shrivastav, M. Nahin, C. J. Hogan and C. Larriba-Andaluz, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 1540–1551.

66 H. Ouyang, C. Larriba-Andaluz, D. R. Oberreit and C. J. H. Jr, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 1833–1847.

67 C. Larriba and C. J. Hogan, *J. Phys. Chem. A*, 2013, **117**, 3887–3901.

68 A. Maißer, J. M. Thomas, C. Larriba-Andaluz, S. He and C. J. Hogan, *J. Aerosol Sci.*, 2015, **90**, 36–50.

69 F. Turecek, C. L. Moss, I. Pikalov, R. Pepin, K. Gulyuz, N. C. Polfer, M. F. Bush, J. Brown, J. Williams and K. Richardson, *Int. J. Mass Spectrom.*, 2013, **354–355**, 249–256.

70 R. Pepin, A. Petrone, K. J. Laszlo, M. F. Bush and X. Li, *J. Phys. Chem. Lett.*, 2016, **7**, 2765–2771.

