Published in final edited form as:

Analyst. 2016 June 20; 141(13): 4084–4099. doi:10.1039/c6an00453a.

Modular calibrant sets for the structural analysis of nucleic acids by ion mobility spectrometry mass spectrometry

Jennifer L. Lippens^a, Srivathsan V. Ranganathan^b, Rebecca J. D'Esposito^a, Daniele Fabris^{*,a,b}

^aUniversity at Albany, Albany, New York, USA.

^bSUNY, Albany, The RNA Institute, 1400 Washington Avenue, Albany, New York, USA

Abstract

This study explored the use of modular nucleic acid (NA) standards to generate calibration curves capable of translating primary ion mobility readouts into corresponding collision cross section (CCS) data. Putative calibrants consisted of single- (ss) and double-stranded (ds) oligodeoxynucleotides reaching up to ~40 kDa in size (i.e., 64 bp) and ~ 5,700 Å² in CCS. To ensure self-consistency among reference CCS values, computational data obtained in house were preferred to any experimental or computational data from disparate sources. Such values were obtained by molecular dynamics (MD) simulations and either the exact hard sphere scattering (EHSS) or the projection superposition approximation (PSA) methods, and then plotted against the corresponding experimental values to generate separate calibration curves. Their performance was evaluated on the basis of their correlation coefficient and ability to provide values that matched the CCS of selected test samples mimicking typical unknowns. The results indicated that the predictive power benefitted from the exclusion of higher charged species that were more susceptible to the destabilizing effects of Coulombic repulsion. The results revealed discrepancies between EHSS and PSA data that were ascribable to the different approximations used to describe the ion mobility process. Within the boundaries defined by these approximations and the challenges of modeling NA structure in solvent-free environment, the calibrant sets enabled the experimental determination of CCS with excellent reproducibility (precision) and error (accuracy), which will support the analysis of progressively larger NA samples of biological significance.

Graphical Abstract

* fabris@albany.edu.

Introduction

Ion mobility spectrometry-mass spectrometry (IMS-MS) has the power of discriminating ions on the basis of both mass-over-charge ratio (m/z) and overall size/topology.^{1–3} As typically the case with hyphenated techniques, combining orthogonal dimensions results in enhanced separation capabilities that translate into greater analyte resolution, peak capacity, and dynamic range.^{4–6} In addition, IMS analysis by itself can provide valuable structural information by virtue of the close correlation between the observed arrival time (t_D) and the collisional cross-section (CCS or Ω) displayed by analyte ions interacting with a given type of bath gas.⁷ The relationship between t_D and CCS is well understood and its physical principles have been described for both drift tube^{8–10} and traveling wave instruments.^{11–13} In addition, extensive efforts have been dedicated to understanding the effects of experimental factors, such as field strength, type of bath gas, and analyte charge state on the outcome of the analysis.^{14–18} With the purpose of bringing the intrinsic potential of this technique to full fruition, different approaches have been devised to translate CCS data directly into actual structural information.^{5,7,13}

In typical MS analyzers, ions are discriminated according to measurable changes in properties, such as momentum, angular frequency, flight time, and others, which have well-defined relationships with the corresponding m/z value. In practical terms, however, such quantities are rarely determined per se and their absolute values are not readily accessible to the user. On the contrary, primary instrumental readouts are translated directly into m/z units by utilizing standards of known mass to calibrate the instrument's scale. The same principle is applicable to IMS determinations in which proper references can be used to translate primary t_D measurements into end-point CCS data.^{5,7,13} Effective calibrants tend to consist of small molecules, salt clusters, and peptides or proteins,^{6,7,17,19} which must meet essential requirements: their CCS must be accurately known; their size must cover the range of interest; and their structure must be sufficiently “rigid” to minimize possible conformational effects of experimental conditions. Reference CCS values are typically obtained experimentally by direct determination of putative calibrants on drift tube instruments, which are considered the golden standard.^{5,17,20–22} However, any experimental strategy

must account for potential variability associated with sample preparation, source design, desolvation conditions, time scale, and any other feature that may affect the conformation of detected ions.²³ The effects of such variables are minimized when calibrants and unknown analytes are analyzed under identical conditions on the same platform. However, experimental CCS values may not be immediately reproducible across different laboratories without extensive effort, or portable to other platforms with different design. These types of challenges have been discussed in great detail in a recent review.²²

The alternative strategy involves calculating the reference values from structural coordinates obtained by established high-resolution techniques, molecular modeling approaches, or combinations of both. Crystal or solution structures can be used as starting points for molecular dynamics (MD) simulations in a solvent-free environment to generate corresponding ensembles of representative structures. Reference CCS values are then obtained from these ensembles by employing accepted algorithms, such as the projection approximation (PA), the exact hard sphere scattering (EHSS) and the trajectory method (TJM) included in the widely used MOBCAL package,^{24–26} as well as the more recent projected superposition approximation (PSA) method.^{27–30} The broad application of these algorithms to a variety of structures has revealed that their accuracy may be influenced not only by the approximations made to model the interactions with background gas and the parameters employed in the actual calculations, but also by the type of sample under consideration. Although several reports have shown excellent agreement between CCS values obtained from experimental and computational data,^{31–33} there are still unresolved questions regarding whether the initial coordinates of complex biopolymers might accurately represent their conformation in the gas phase, which could cast some doubts on the accuracy of subsequent CCS calculations. This concern represents an underlying challenge that is common to any MS-based approach aiming at the investigation of biomolecular structure.^{34,35} These considerations aside, robust computational data based on state-of-the-art understanding of gas-phase structure would be expected to be platform-neutral and, at least in principle, reproducible across different laboratories.

Recent advances in the knowledge of nucleic acid (NA) function have led to increasing demand for structural characterization of species that do not owe their activity to the information encoded in their sequence, but rather to their 3D structure.^{36–39} In turn, this demand has reignited the interest in alternative structural approaches for the characterization of samples that are not directly amenable to traditional high-resolution techniques.⁴⁰ The prominent roles played by base-pairing and stacking interactions in determining NA higher-order structure have enabled the development of very accurate structure-prediction algorithms.^{41,42} At the same time, the stability manifested by these types of interactions in the absence of solvent^{43–46} has been credited for the excellent correlation observed between corresponding gas-phase and solution structures.^{32,47–52} The ability of IMS-MS determinations to distinguish between DNA helices in the A- or B-form^{31,33} and to probe the stability of triplex and quadruplex structures^{53–55} foreshadows possible strategies in which this technique could be readily employed to validate the results of structure-prediction algorithms and to guide the selection between alternative theoretical models.

The vast majority of biologically relevant species whose function is defined by higher-order structure exceed the upper limits accessible by current IMS-MS approaches. While protein analysis can count on the availability of a broad range of viable calibrants, the choice of well-characterized references for NA analysis is much more limited. Although NA samples can be detected in positive ion mode in what is known as “wrong-way-round” ionization effect,⁵⁶ their analysis is more aptly performed in negative ion mode to take advantage of the acidic character of the phosphate groups.^{57,58} Considering the prominent effects of charge state on the conformation of protein ions, their utilization as standards for anionic determinations may pose unforeseen challenges.⁵⁹ Various sets of negatively-charged polymers have been recently investigated as possible calibrants, including α,ω -carboxy-terminated polystyrene,¹⁹ α,ω -disuccinic acid-terminated poly(ethylene glycol),¹⁹ polyalanine,⁶⁰ and polymalic acid.⁶⁰ However, the maximum size covered by these sets is equivalent to that of a 10 base-pair (bp) duplex, which would limit their applicability to larger NA structures. Preliminary work in our laboratory tested an assortment of single-stranded (ss), double-stranded (ds), hairpin (hp), and quadruplex structures of increasing size, for which CCS values had been already determined either computationally through established molecular modeling approaches, or experimentally through drift tube measurements, or both.^{31–33,61} In this case, however, the overall range did not exceed ~30 bp,^{31,33} thus making extensive extrapolation a necessity for larger samples. Owing to the heterogeneous origin and limited range of the putative references, the performance of this initial calibration curve proved to be less than satisfying.

With the goal of extending the accessible range to at least 64 bp, we explored the possibility of utilizing a series of oligo-deoxyribonucleotides (ODNs) of increasing size as possible references for anionic determinations. Taking inspiration from typical NA “ladders” employed as mass markers in gel electrophoresis, we designed ODNs with modular sequences that could be readily extended to cover any desired range (**Figure 1**). Instead of relying on homopolymers (e.g., polyA, polyT, etc.) that may introduce unforeseen bias, we employed diverse sequences with mixed base composition, which were expected to more closely mimic natural samples. The putative standards were either evaluated together as a single set, or grouped in distinct ss and ds series that combined species of like structure topology. The merits of the various calibration curves were determined by using a separate set of samples, which were purposely different from those included in the ladders. The test samples were selected for their ability to provide distinctive ss, ds and hp structures, in such a way as to enable the evaluation of possible topology effects. The outcome showed that the predictive power of the calibration curves varied not only according to the type of algorithm employed to calculate the theoretical CCSs, but also to the general topology of the species involved, to a certain extent. The study identified the possible boundaries defining the performance expected from such curves in the determination of unknowns, thus providing the means for comparing results obtained from different samples and different laboratories.

Experimental

Materials

Standard ODNs were designed to obtain series of ss and ds constructs ranging from 6 to 96 nucleotides (nt) and 6 to 64 base pairs (bp), respectively (see **Supporting Material, Table 1S**). The selected sequences possessed mixed base compositions to replicate the most general possible samples. Repeating sequences were progressively added together to increase the size in modular fashion. Heat-assisted annealing was performed to generate ds species from complementary ss components (*vide infra*). The smaller ds constructs included a sufficient G-C content to maximize duplex stability and minimize transient strand dissociation (i.e., end-fraying and breathing). The expected melting temperatures and the possibility of forming unwanted secondary structures were assessed by applying the algorithms included in the UNAFold,⁶² Sfold,⁶³ and pknotsRG⁶⁴ packages. Selected ODNs employed to obtain duplex structures were also included as individual strands in the ss series, in such a way as to extend its length to at least 96 nt (**Table 1S**). All samples employed in the study were purchased from Integrated DNA Technologies (Coralville, IO). In general, species with mass above 3 kDa were subjected to buffer exchange by ultra-filtration against 150 mM ammonium acetate (pH adjusted to 7.0) by using Microcon devices from Millipore (Billerica, MA), which enabled the substitution of non-volatile cations with MS-friendly NH_4^+ . Species smaller than 3 kDa demonstrated minimal cation adduction and were thus analyzed without buffer exchange. Quartz emitters for nanospray were produced in house by using a Sutter Instruments Co. (Novato, CA) P2000 laser pipette puller.

Arrival time determination

Immediately prior to analysis, samples were diluted in 150 mM ammonium acetate to a final concentration of 4-10 μM . Samples containing complementary strands were mixed in equimolar amounts, then subjected to refolding by heating to 90°C for 3 min., followed by slow cooling to ambient temperature. Upon cooling, each solution was added with a 10% volume of 2-propanol. A ~5 μL aliquot was finally loaded onto a quartz emitter to complete the analysis by direct infusion nanospray ionization. All experiments were performed in negative ion mode on a Synapt G2 HDMS traveling-wave ion mobility mass spectrometer (Waters, Manchester, U.K.).⁶⁵ The sampling cone of the original Z-spray source was replaced with a home-built heated metal capillary to minimize any unwanted activation during ion desolvation.²³ The incidence of ammonium adducts was monitored to optimize the tuning conditions necessary to achieve the most gentle possible desolvation.⁵⁸ Such conditions were achieved by utilizing a typical emitter voltage of 0.4-0.8 kV, a cone voltage of 40-100 V, and by keeping the source temperature within the 25-45°C range.²³ Some of these parameters were progressively increased with sample size to minimize adduct formation and allow for greater resolution. In all experiments, the traveling wave element was kept at a constant pressure of 4.2 mbar of N_2 , as indicated by the instrument's gauge (uncalibrated reading). Wave height and velocity were kept at 650 V and 40 m/s, respectively. Optimal instrumental settings were determined for individual samples spanning the entire range of size and topology covered by the ladders. Considering that such

parameters differed very little across the range, we opted to utilize the lowest energy settings for all the samples in the study.

Arrival time distributions (ATDs) were recorded for all detectable charge states afforded by the samples in the study. Each charge state was isolated in the mass-selective quadrupole of the instrument and analyzed separately in the traveling wave element.^{65,66} Any species/charge state that displayed multiple discrete signals, which are diagnostic of metastable conformations,^{23,65,66} was excluded altogether from consideration to avoid any ambiguity in their structure assignment and CCS calculations. In consistent fashion, species/charge states that displayed multiple signals were excluded as possible test samples to avoid any interpretation ambiguity. Although the incidence of cation adducts increased with the size of the construct, as typically observed for these types of analytes, proper isolation in the mass-selective quadrupole enabled the exclusive analysis of un-adducted ions. MassLynx (v4.1, SCN781) was employed to carry out the initial data processing. In order to accurately identify the apex of each time-domain signal, the ATD traces were submitted to Gaussian fitting in PeakFit (v4.2 SeaSolve Software Inc., Framingham, MA) without any prior smoothing of the original raw data. This process is also capable of revealing whether multiple unresolved populations might contribute to the observed signal. Once optimal deconvolution parameters were identified, they were kept constant for all samples to avoid bias.

The fitted t_D values were corrected for typical instrumental delays to obtain the corresponding t_D' values (Eq. 1):⁷

$$t_D' = t_D - t_T - \left| \frac{c\sqrt{m/z}}{1000} \right| \quad \text{Eq. 1}$$

The correction factors consisted of delays incurred during transit across the transfer traveling-wave (t_T), which were estimated from the length of this element divided by the wave velocity employed in the experiment,^{5,7} and during travel from the transfer traveling-wave to the time-of-flight analyzer, which were calculated from the enhanced duty cycle coefficient (c) measured for our instrument, the appropriate mass-to-charge ratio (m/z), and a scaling factor (1000) necessary to express all terms in homogenous units (s).⁷

The reference CCS (Ω) values were themselves corrected for the charge (z) and the reciprocal of the reduced mass ($1/\mu$), which accounts for the mass of both analyte and background gas involved in the interaction, thus leading to the corresponding Ω' (Eq. 2):⁷

$$\Omega' = \frac{\Omega}{\left(z \times (1/\mu)^{1/2} \right)} \quad \text{Eq. 2}$$

Finally, these values were plotted against the natural logarithm of the computed CCS value ($\ln(\Omega')$) and fit to a linear relationship to generate the exponential factor (X) necessary to calculate the final t_D'' employed for the calibration curve, as shown in Eq. 3:⁷

$$t_D'' = t_D^X \times z \times (1 / \mu)^{1/2} \quad \text{Eq. 3}$$

In the context of our report, t_D'' was referred to as the corrected t_D for consistency with the treatment described in reference 7. However, this quantity was employed here as the primary readout (i.e., “measured” quantity) that was correlated with the corresponding computational CCS value (i.e., “known” value) to build the desired curve.

Calculation of reference CCS values

All-atom MD simulations were performed by using GROMACS 4.6.3⁶⁷ to obtain an ensemble of structures for subsequent CCS calculations by accepted algorithms. The Nucleic Acid Builder (NAB) package of AMBER⁶⁸ and a modified version of the AMBER99 force-field, which was specifically optimized for NA samples,⁶⁹ were employed to generate the initial structures and perform the simulations. This modified force-field was specifically selected with the goal of minimizing typical errors associated with the simulation of NA structure, which has been traditionally accomplished under solution conditions. The structural coordinates provided by these types of simulations tend to compare very favorably with those obtained experimentally by NMR or crystallography, as demonstrated by the excellent agreement displayed by corresponding CCS values.⁷⁰ However, this may not be always the case when the simulations are completed under solvent-free conditions, which is a direct reflection of the challenges of representing NA structure in the gas phase. Following the lead of accepted protocols for gas-phase NA simulations, the DNA duplexes were modelled in either the A or B form according to the criteria explained in **Results and Discussion**. The absence of counter-ion and solvent adducts observed during analysis demanded that the actual charge state be properly accounted for. For this reason, desired charge states matching those observed experimentally were achieved by neutralizing an appropriate number of phosphate groups with hydrogen atoms. The location of charged sites was varied along the phosphodiester backbone according to the criteria discussed in the **Results and Discussion** section.

Each initial structure was subjected to equilibration at selected temperatures to ensure that the coordinates employed for CCS calculations reflected the putative average structure compatible with such conditions. **Figure 2** displays representative equilibration trajectories for two ds constructs of different sizes, in which the radius of gyration (R_g) was monitored as a function of time under both solution and solvent-free conditions (solid and dashed lines, respectively). Although this quantity cannot be directly correlated to a corresponding CCS value, its variations during the course of a simulation can provide an excellent indication of whether the model undergoes major conformational changes, or minor fluctuations around an average conformation. After brief adjustment periods, the R_g variations displayed by both solution and solvent-free trajectories remained confined to relatively narrow ranges spanning $\pm 4.4\%$ around the respective averages. The fact that this quantity had ceased to swing widely over time was a clear indication that the configurations along each trajectory had settled in proximity of an average predominant structure. Such structure was still subjected to local dynamics associated with the selected temperature, as reflected by limited R_g fluctuations,

but was not affected by major conformational changes responsible for much wider R_g variations.

When evaluating dynamics in NA systems, it is important to consider the prominence of base pairing in defining higher-order structure and its influence on conformational stability. Earlier simulation studies have shown that an open base pair possesses an average lifetime of only 10 – 20 ps in solution, and predicted that these types of local dynamics would be further reduced in the absence of solvation and counter-ion effects.⁷¹ This prediction was corroborated experimentally by the absence of detectable conformational effects when the lower charge states of various base-paired constructs were analyzed by IMS in the ms scale.⁵¹ At the other end of the spectrum, global dynamics involve major conformational rearrangements and variations of the mutual position of contiguous domains, which are typically completed in the μ s to ms scale.⁷² In our case, however, all sequences in the study were vetted for their inability to produce any stable conformation other than the duplex structures anticipated for the ds series, thus effectively ruling out the possibility of global dynamics between alternative folded states. These considerations indicated that it was safe to employ the configurations obtained after 500 ps of simulation as valid representations of the average solvent-free structure assumed by each given construct. For this reason, individual configurations taken from this region of the trajectory were stored every 25 ps and employed for subsequent CCS calculations.

Different algorithms are available for translating the coordinates produced by MD simulations into actual CCS values, which are based on different models of the interaction between ions and background gas during the IMS process. The trajectory method (TJM) included in the MOBCAL^{24–26} suite has been long considered the golden standard for these types of calculations.^{27,73} However, its application to progressively larger biopolymers has revealed numerous challenges and displayed increasing computational instability. In contrast, the exact hard sphere scattering (EHSS)²⁵ method tends to incur fewer problems with biomolecules and places more limited demands on computational power.^{15,32} To help our selection, representative constructs were employed to compare the performance afforded by these algorithms with NA structures. The original versions included in the MOBCAL suite, which are set to work exclusively with a He background, were re-parameterized for N_2 to enable direct correlations with our experimental conditions. The exercise showed modest deviations between corresponding CCS values (see **Supporting Material, Table 2S**), which contributed to the decision to rely on the less expensive EHSS method. At the same time, we elected to employ also the projected superposition approximation (PSA) method,^{27–30} which is based on a different model of the interactions between background gas and complex biomolecular structures. This algorithm was already parameterized for both He and N_2 and was thus employed without modifications. Both methods were used in parallel to calculate reference CCS values from the structures produced by MD simulations. For each construct in the study, we utilized at least 20 separate configurations taken from the equilibrated region of the MD trajectory. Therefore, each reported value represented the average and standard deviation (expressed as RSD%) afforded by these distinct but related structures.

Results and discussion

Rationale

Effective calibration curves correlate a “measured” quantity with a corresponding “known” value to obtain a relationship that will enable the determination of unknowns through actual measurements. Even before any measurement is completed, the performance of a curve is first and foremost determined by the quality of the available reference values. In general, reference values with minimum deviation from the corresponding true values lead to better overall accuracy (i.e., smaller error). At the same time, self-consistent reference series with minimum sample-to-sample uncertainty produce better precision (i.e., smaller scatter and better correlating power). The accuracy of any CCS prediction ultimately rests on the accuracy of the models used to represent NA structures in the gas phase and to describe their interactions with background gas during IMS analysis. We addressed the former by adhering to the most advanced principles of NA structure in solvent-free environment, although we are fully aware of the deficiencies of current models. We addressed the latter by employing in parallel two of the leading approaches for calculating CCS values from 3D coordinates. Therefore, the overall accuracy of CCS prediction will be expected to increase with an increased knowledge of these fundamental principles.

This study specifically confronts the current heterogeneity of putative references for NA analysis, which lack the range and self-consistency necessary to support practical applications. The selection of reference values of either experimental or computational origin is faced with different challenges. The utilization of previously reported experimental values is discouraged by the fact that the corresponding samples have been analyzed on different platforms and under widely heterogeneous conditions. At the same time, computational CCS values from 3D structures generated by either high-resolution techniques or advanced modelling protocols do not offer any simple way to evaluate sample-to-sample uncertainty. For the sake of self-consistency, we elected to utilize reference values obtained in house by a single computational workflow that, applied across the board to all reference and test samples alike, would minimize the uncertainty associated with sample-to-sample variability. This computational workflow was validated by calculating the CCS of known NA species and assessing the deviation from their reported computational values (vide infra).

At the end, the goal of this work was to identify modular series of ODN standards to support the practical application of IMS-MS to NA analysis, which possessed CCS values consistent with the accuracy allowed by the current understanding of gas-phase structure and IMS behavior of these types of samples. The selected standards will be expected to provide the robust reproducible framework necessary to compare the CCS of different samples on the same instrument and across different hardware platforms, evaluate conformational changes as a function of experimental variables, and support the formulation of valid hypothesis on structure and dynamics.

Obtaining reference CCS values from MD simulations

Owing to the modular nature of the selected calibrants, the model-building operations were carried out by conveniently grouping them into separate series consisting of either ss or ds structures (**Tables 1 and 2**, respectively). The sequences of the former were carefully crafted in such a way as to avoid the possible folding of any stable higher-order structure. This condition was verified by employing structure-prediction algorithms included in the UNAFold,⁶² Sfold⁶³ and PKnotsRG⁶⁴ servers, which did not identify any preferential fold. The ds constructs were instead obtained by combining complementary strands capable of forming stable duplexes (**Figure 1**). The initial crystal-type structures were appropriately protonated to match the charge states observed experimentally and were used to perform all-atom molecular dynamics (MD) simulations to arrive at an ensemble of representative gas-phase structures. The calculations were performed in the absence of water to simulate solvent-free conditions according to established models.^{33,74}

The MD simulations were performed at a temperature of 300 K to replicate the conditions employed in previous studies^{54,74} and to eschew the uncertainty surrounding the effective temperature of ions during the analysis. Over the years, various experimental strategies have been proposed to enable the direct determination of this parameter,^{75–77} but the lack of consensus exposes the intrinsic difficulties of extending such strategies to different types of samples and hardware designs.^{16,78,79} In the case of NA structure, comprehensive MD simulations in solvent-free environment have shown that temperature has negligible effects on the structure of triplex DNA in the 300 to 372 K range.⁵⁴ Given that these conclusions seemed to contradict the significant impact of temperature on NA structure in solution, we decided to verify whether they could be safely extended also to our ss and ds references. For this reason, representative members of the ss and ds series were submitted to simulations at both 300 and 372 K. The results were examined by comparing the CCS of conformers sampled along the respective simulation trajectories, which were calculated according to both the EHSS²⁵ and PSA^{27–30} algorithms (see *Experimental*). The relative standard deviation (RSD%) obtained from 20 of such conformers was used as a measure of the possible uncertainty (i.e., precision around a predominant average structure) associated with these structures under the selected conditions. The results showed that the uncertainty did not vary significantly at the two different temperatures for both the ss species **48b** and the **18a:b** duplex (i.e., from ± 1.3 to $\pm 1.2\%$ RSD% for the former, and from ± 0.8 to $\pm 0.6\%$ RSD% for the latter).

The constructs of the ss series were modeled without enforcing any pre-defined conformation, whereas the ds species were modeled as full-fledged duplexes. Although such species are known to assume B-form helical structures in solution,⁸⁰ earlier reports have provided evidence that constructs smaller than 8 bp exhibit nondescript globular topologies in solvent-free environment, whereas those between 8 and 18 bp adopt distinctive A-form configurations.^{31,33} Based on these findings, ds constructs between 8 and 18 bp were modelled in an A-form configuration that exhibited the more extended topology characteristic of RNA molecules (e.g., 10 bp per turn with a rise of 3.4 Å per bp).⁸⁰ The simulation trajectory obtained from the representative **14a:b** duplex (**Figure 2**) showed that the solvent-free values of R_g were slightly smaller than the solution ones, consistent with a

modest compaction of the typical A-form helix. In contrast, the larger constructs were generated in the B-form topology characteristic of DNA helices (e.g., 11 bp per turn with a rise of 2.9 Å per bp).⁸⁰ In this case, the representative **64a:b** duplex produced significantly greater R_g values in solvent-free than solution calculations (**Figure 2**). When corresponding configurations on either trajectory were compared, the solvent-free structures appeared to progressively stretch in the axial direction and shrink in the radial one as compared to their solution counterparts, in agreement with the findings of prior gas-phase simulations of helical structures.^{33,74,81} This effect was not as prominent in the smaller **14a:b** duplex, in which the axial expansion was found to be less significant than the radial shrinkage, consistent with its much smaller aspect ratio.

In order to evaluate the possible precision of the modeling operations, the equilibrated configurations from the later portion of each trajectory were employed to obtain the corresponding CCSs and evaluate their variations in terms of RSD% (see *Experimental*). Both EHSS and PSA datasets showed that the ds series afforded smaller average RSD% than their ss counterparts (**Table 1** and **2**). This apparent discrepancy is consistent with the unconstrained nature of the ss structures submitted to the equilibration process, which exhibited greater conformational diversity (i.e., flexibility) from the very beginning of the modeling process. These RSD% values provided a reliable measure of the intrinsic uncertainty (i.e., precision) associated with each reference employed in the study (**Table 1** and **2**).

The overall charge of each construct was adjusted to match the charge state observed in the respective determination to enable direct correlations between theoretical and experimental data. Given that typical model-building tools generate NA structures containing exclusively deprotonated phosphates, charge adjustment was accomplished before equilibration by neutralizing an appropriate number of phosphates through addition of hydrogen atoms. This approach was expected to provide a more realistic representation of the detected ions than the possible alternative, which involved the uniform distribution of partial charges across all phosphate groups to reproduce the overall charge state.^{54,74} This approach, however, required selecting the specific phosphates to be neutralized, which may become a possible source of uncertainty. In light of the absence of solvent molecules and counterions in the system, the charged phosphates were spaced out as much as possible to minimize electrostatic repulsions and to reduce the total energy of the system. At the same time, we also tested possible alternatives by simulating arrangements in which the charges were all clustered together. Consistent with the propensity of biopolymers to minimize the effects of Coulomb repulsion, the distributed arrangements displayed better fits with the experimental observations. Earlier studies have shown that greater charging may result in the extension (i.e., “stretching”) of single-stranded structures, which may translate into progressively larger CCS values.⁶¹ However, the “stretching” effect is typically less pronounced for ds structures, even at the higher charge states.⁶¹ The distributed, discrete charge arrangement was found to minimize this effect and increase the reproducibility of models with progressively greater charging. For this reason, this strategy was employed for all the constructs in the study. The coordinate pdb files containing the equilibrated structures are publicly available for both calibrant series upon request.

The EHSS and PSA algorithms employed in the study use different approaches for assessing the effects of the interactions between analyte ions and background gas. The former relies on a hard sphere of specific radius to represent each atom of the colliding species,²⁵ while the latter replaces this quantity with a collision probability and utilizes a shape factor to better account for the molecular geometry.^{27–30} The proposed calibrants were employed to compare the performances of these algorithm with NA structures (**Table 1** and **2**). The results obtained in N₂ background showed that their respective precisions were very similar, amounting to $\pm 1.1\%$ and $\pm 1.9\%$ average RSD% for the EHSS and PSA methods. In addition, the EHSS values were consistently lower than the PSA ones by an average of 16.5%. We could speculate that the shape factor in the PSA treatment may provide a closer approximation of the ion mobility process for biomolecules of this size, and that this may be the source of the systematic discrepancy. Corroborating this possibility will require ad hoc experiments that are beyond the scope of this report. When the calibrant sets were considered separately, the systematic difference between methods amounted to 17.7% and 15.7% average deviations for the ss and ds series, respectively, thus indicating that sample topology may have modest but not null effects on CCS calculations.

The entire computational workflow was tested on structures for which CCS values had been already calculated independently by other groups that utilized analogous MD simulations and EHSS calculations.^{31,33,51,61} The relative deviations (%) between corresponding values were employed to obtain an overall relative square mean deviation (RMSD) for the entire set. The values produced by our workflow exhibited an overall deviation of $\pm 4.5\%$ RMSD from the accepted ones, with the larger %s displayed by relatively unconstrained ss species, rather than the more constrained duplex/hairpin structures (**Table 4S** in *Supporting Material*). In spite of the relatively small RMSD, the individual %s fluctuated widely without a recognizable trend, thus ruling out the possible utilization of correction factors. The indeterminate nature of these deviations epitomizes the challenge of achieving a high degree of reproducibility from comparable but perhaps not identical computational protocols. Concerns about the possible uncertainty associated with data of different origin were among the reasons for relying on reference CCSs that were obtained from a single source and, thus, would be expected to possess greater self-consistency.

Experimental t_D determinations

The primary time-domain data necessary to obtain experimental CCS values were acquired by analyzing the selected reference samples by nanoflow electrospray ionization (nanospray)⁸² under conditions of pH and ionic strength, which ensured the detection of intact duplex structures (see *Experimental*). Their intrinsic acidic character makes ODN analytes negatively charged in neutral solution and allows for their detection as deprotonated ions. Accordingly, representative data obtained from the **64b** member of the ss series (**Table 1**) displayed signals with charges ranging from -7 to -9 (**Figure 3a**). Desolvation conditions were adjusted to minimize the incidence of ammonium adducts characteristic of these types of analytes without destabilizing strand association (see *Experimental*).^{23,58} This situation is exemplified by the data obtained from a representative member of the ds series (**Table 2**), which displayed abundant signals corresponding to the deprotonated duplex with charge states ranging from -8 to -10 (**32a:b** in **Figure 3b**). Lower signals could be detected also

for one of its ss components (i.e., **32a**), but not for the complementary strand. These data indicated that the observed strand was not produced by dissociation of the initial duplex, but was instead a remnant of the annealing procedure employed to assemble the ds complex, in which one of the components was present in slight excess over the other (see *Experimental*).

A common feature displayed by the ODN analytes in the study was their relatively narrow charge state distribution, which in the case of the ss species seemed to contradict their inability to form stable secondary structures (**Figure 3a**). In spite of the presence of a least one chargeable phosphate per nucleotide, none of these samples displayed more than three charge states under the selected experimental conditions, and none reached the maximum possible charging allowed by the total number of ionizable groups. Analyte charging is typically determined by a combination of structural features (e.g., availability of chargeable groups, folded versus unfolded state, etc.), solution conditions (e.g., pH, ionic strength, organic co-solvent, etc.), and desolvation factors (e.g., source energetics, time frame, etc.).^{83–86} Preliminary experiments aimed at reproducing the charge state distributions observed for the same constructs by other groups were met with mixed results. For example, an 18-bp duplex that had been previously observed with charges ranging from –9 to –11 produced only a –5 signal in our hands, albeit sample preparation and solution conditions matched those described in the original report.³³ Any attempt to increase the observed charge state by adjusting instrumental parameters under our control (e.g., emitter and cone voltages, source temperature, etc.) resulted in either loss of signal or duplex dissociation. The challenge of consistently reproducing the charging pattern, and by extension the IMS behavior, observed on different platforms was one of the factors that discouraged the utilization of heterogeneous experimental values as possible references.

Each charge state of interest was isolated in the mass-selective quadrupole^{75–77} and then injected into the traveling wave^{65,66} element of the instrument to obtain the primary t_D data (see *Experimental*). Representative ATDs obtained from the –8 charge states of **64b** and **32a:b**, respectively, displayed well-delineated, distinctive signals with excellent Gaussian fit, which were consistent with the presence of a single recognizable population of ions (**Figure 4a and b**). In a few cases, however, multiple signals could be recognized, which corresponded to alternative ion structures (**Figure 4c**). This type of situation arises when species under investigation may assume different folds in solution, or produce alternate conformations in the gas phase, which are relatively stable within the experimental time frame.^{7,23,87,88} As mentioned in the *Experimental* section, any charge state that produced multiple detectable conformers was excluded altogether from the study to eliminate any ambiguous CCS assignment.⁷ No attempt whatsoever was made to rescue these type of data by guessing which ATD signal may represent the main conformer corresponding to the given charge state.

The computation of each reference value utilized a predominant average structure identified by the equilibration process. To obtain a valid experimental value for such structure, we strived to minimize any unwanted perturbation of sample conformation during analysis by optimizing the source conditions to achieve the gentlest possible desolvation (see *Experimental*).²³ In addition, we evaluated the effects of essential instrumental parameters, while monitoring the detected ATDs for signs of structure perturbation. For each sample,

only a single well-defined signal was detected within the explored ranges (e.g., 30-50 °C source temperature; 50-100 V cone voltage; 4-12 V trap voltage), with no sign of additional signals attributable to alternative conformers. In some cases, modest peak broadening was detected, which did not exceed a 1% increase of the full width at half-height (FWHH). This increase was ascribed to slightly greater local dynamics analogous to the low level fluctuations observed along the simulation trajectories (**Figure 2**). The fact that the position of the peak's centroid on the time scale was not affected at all indicated that the average ion structure remained effectively unchanged within the explored range of parameters, thus justifying the utilization of the corresponding t_D for CCS determinations. This type of analysis was completed for all samples in the study to ensure that any destabilizing effects introduced by instrumental conditions were reduced to a minimum.

The primary t_D readings were subjected to an established treatment that accounts for the type of drift gas used in the experiment, the flight time between the IMS element and mass analyzer, and the different acceleration experienced in such element by ions of different charge (see *Experimental*).⁷ The correction was accomplished by employing specific factors, such as reduced mass, enhanced duty cycle, and charge, which are discussed in great detail elsewhere.^{6,7} This treatment was implemented across the board for all samples in the study to obtain corresponding t_D'' values for the desired calibration curves. Each calibrant was analyzed in triplicate to assess the uncertainty associated with the corresponding t_D'' , which was again expressed in terms of RSD%. Taken together, all members of the ss and ds series afforded an average RSD% of $\pm 0.92\%$, which provides a measure of the excellent reproducibility (i.e., precision) of these experimental determinations.

Assessing the performance of calibration curves

Reference CCSs and corresponding IMS-MS data were correlated by linear regression to generate calibration curves that provided the means for translating any observed t_D'' into a corresponding CCS value (**Figure 5**). The regressions afforded by the EHSS and PSA series possessed very different slopes, with the latter showing a much sharper increase of CCS as a function of t_D'' . This feature clearly reflected the general trend manifested by the reference series, in which EHSS values were consistently lower than the PSA counterparts (**Table 1** and **2**). Overall, the EHSS and PSA curves covered a range of CCS up to 4,900 and 5,700 Å², respectively. These boundaries more than doubled the upper limit of ~2,400 Å² previously observed for reported NA structures.³³ Overall, the data points were fairly well distributed across the range, a favorable byproduct of the modular nature of the calibrants and the diverse nature of their structures. Indeed, avoiding major gaps between data points would be expected to enable closer interpolation in the determination of unknowns. The correlation coefficients (R^2) happened to match 0.984 for both regressions. Further, a closer examination indicated that the observed coefficients were not significantly affected by the uncertainty (i.e., scatter) associated with the individual data points, which appeared to be very limited across the board (see typical error bars in **Figure 5 inset**), but rather by the actual distribution of the points in each curve.

The curves in **Figure 5** included only data points corresponding to the lowest charge state detected for each calibrant. When all observed charge states were considered, however, the

data distributions looked quite different and provided dissimilar correlations coefficients (**Figure 1S** in *Supporting Material*). The effect of Coulombic repulsion between like charges has been the subject of intense debate for its possible consequences on conformational stability and uncertainty of determination.^{33,61,89} However, such effect is known to be less prominent at the lower charge states, which are generally considered more stable. This consideration may explain the discrepancies displayed by the curves that included all detected charges states. In order to eliminate this possible source of uncertainty, only the lowest charge states were employed to generate the calibration curves.

An additional criterion for evaluating the robustness of a calibration curve is represented by its ability to produce results that match as close as possible known values. We evaluated the predictive power of the calibration curves by using separate sets of samples, which were not employed to build the curves themselves (see *Supporting Material, Table 5S*). The test samples consisted of species of varying sizes and structures designed in house, which were submitted to MD simulations according to the same workflow/parameters employed for the calibrants. The samples were grouped in specific sets according to overall structure topology (i.e., ss, ds, and hp) to explore the possible influence of topology on the choice of suitable standards (**Table 3**). The selected sizes expressed in terms of oligonucleotide length (nt) were evenly distributed over the entire range covered by the calibrant series, in such a way as to probe different regions of the curves and to enable determinations without extrapolation. The configurations produced by the MD simulations were employed to calculate computational CCS values that were compared with the experimental results obtained from the curves.

In parallel, the test samples were analyzed by IMS-MS under the same experimental conditions employed for the reference compounds. Their t_D values were processed through the calibration curves to obtain experimental CCSs, which were immediately compared with their computational counterparts by calculating the respective relative deviations (%, **Table 3**). For the sake of consistency, the treatment matched test values with curves obtained by using the same CCS algorithm. The results revealed deviations that fluctuated in both positive and negative directions with no recognizable patterns. For example, the experimental values observed for the smallest and largest species of the ss test set (i.e., **16b** and **70b**) deviated by +6.6% and -4.7% from their computational EHSS values, whereas the smallest/largest species of the ds set (i.e., **10a:b** and **56a:b**) deviated by -10.8% and +8.0%, respectively. Similar trends were observed when the PSA data were examined. Therefore, no clear correlation was apparent between deviation and sample size, which could be readily translated into convenient correction factors. The only clear pattern was observed for the hp samples, which consistently displayed the largest deviations in the study, all in the positive direction. For this reason, excluding such samples from consideration dropped the overall RMSD values provided by the EHSS and PSA curves from ± 16.6 to $\pm 5.8\%$ and from ± 16.9 to $\pm 3.4\%$, respectively (**Table 3**). These RMSDs are very remarkable when considering that they account also for the intrinsic variability associated with conformation dynamics. In a broader context, small overall deviations are always very desirable when comparing computational structures with those obtained by typical experimental approaches, such as NMR and crystallography. In the RNA-Puzzle competition, for example, in which the

performance of structure-prediction algorithms is evaluated by the deviation of a model from the corresponding high-resolution structure, RMSD values ranging between ± 2.3 and $\pm 7.2\%$ are typically considered acceptable, depending on the complexity of the structure under investigation.^{41,42}

Sample topology and predictive power

The results afforded by the hp test samples suggested the possibility that sample topology might influence the predictive power of the curves, or that a systematic error might specifically plague the modeling of hp structures in the gas phase. In order to investigate this possibility, ss and ds calibrants were employed as separate sets to generate individual calibration curves (**Figure 6**). Reference CCS values obtained from EHSS and PSA calculations were again considered separately for the sake of consistency. In both cases, the results revealed very distinctive trends, in which the ss curves displayed sharper CCS increases as a function of t_D than their ds counterparts. The former provided R^2 values of 0.983 and 0.993, whereas the latter gave 0.983 and 0.985 respectively for EHSS and PSA data. The fact that each curve contained approximately half of the available calibrants resulted in more sparse point distributions than those obtained when all calibrants were utilized together (**Figure 5**). Due to their respective compositions, the ss curves also covered smaller ranges than the ds counterparts.

The test sets were again employed to assess the predictive power of the topology-specific curves (**Table 4** and **5**). The RMSDs calculated from the various %s clearly showed that the predictions were closer (i.e., smaller RMSDs) when the topology of the test set matched that of the calibrants. This observation was more significant for the EHSS dataset, in which the ss samples displayed an RMSD of $\pm 2.2\%$ when processed through the ss curve, which increased to $\pm 10.7\%$ when using the ds one. Conversely, the ds samples manifested an RMSD of $\pm 9.4\%$ with the ss curve, which dropped to $\pm 4.8\%$ with the ds one (**Table 4**). This outcome was consistent with the different relationships between computational CCS and experimental t_D displayed by ss and ds calibrants, which were clearly substantiated in the distinctive slopes of the corresponding regressions (**Figure 6**). At the same time, limited improvements were observed also for the overall RMSDs afforded by the hp samples, which were still worse than those obtained from ss and ds samples (**Table 4** and **5**). Indeed, all calibration curves tested in the study provided very comparable RMSD values for the hp samples, thus suggesting that a possible correction factor might be applicable. In this direction, when the relative deviations between computational and experimental values were considered on an individual basis, steady improvement could be observed as a function of size, regardless of the curve applied. For example, the deviation decreased from $+41.2\%$ for **16hp** to $+12.6\%$ for **48hp** in the EHSS curve obtained from all available calibrants (**Table 3**). At the same time, it also decreased from $+35.4\%$ for the **16hp** sample to $+14.6\%$ for the **48hp** sample in the EHSS curve obtained from the ss calibrants (**Table 4**), with similar patterns in the other curves. Considering that the hp structures combine a constant single-stranded tetraloop with double-stranded stems of increasing size (see *Experimental*), the observed deviations correlated well with an increase of the double-stranded character of their structures, which increased with the size of the constructs themselves. Given that these species were modeled in rigorous accordance with accepted principles,^{33,54} these

observations are a clear reflection of an insufficient understanding of hp structure in the gas phase. The fact that this type of topology was not represented at all in the calibrant sets reduced the ability of the curves to minimize the impact of this shortcoming.

The effects of topology on the predictive power of PSA calibration curves (**Table 5**) were not as marked as those observed for the EHSS counterparts (**Table 4**). In this case, the overall RMSDs for the ss test samples were very similar between the ss and ds curves, whereas the ds test samples demonstrated a more conspicuous improvement when using the ds curve. Overall, however, the RMSDs obtained from the complete calibrant set (PSA section of **Table 3**) were still significantly better than those observed with topology matching (**Table 5**). The influence of sample topology has been investigated in the context of protein analysis, which revealed that calibration curves generated from native-like calibrants achieved better predictive power for native-like analytes than those obtained from denatured counterparts, and vice versa.^{6,21} Our results indicate that the influence of topology may be reduced by the utilization of the PSA algorithm, which appears to be better equipped to deal with the overall shape of biopolymer analytes. In general, the EHSS algorithm does not fully account for the collisions with concave regions of a folded structure and, thus, may miss the significant contribution of distinctive structural features.²⁷ In contrast, the PSA algorithm employs a shape factor to describe the degree of concaveness of the structure, which may lead to a better approximation of its actual CCS. The introduction of possible factors for capturing the different behaviors of convex versus concave surfaces will be expected to further improve the accuracy of CCS calculations.²⁷

A discussion of the possible significance of sample topology cannot be uncoupled from that of charge placement, which we briefly broached above. In fact, it has been shown that the charge on analyte ions can affect the dipole moment of rather polarizable gases, such as N₂, and promote possible dipole-dipole interactions that are not accounted for by the leading CCS algorithms.^{17,90–92} Although this effect is significant for analytes of relatively small size (e.g., <500 u),^{92,93} it appears to taper off for larger multiply charged ions in which charge distribution may not necessarily support the formation of a strong analyte dipole. This scenario is likely applicable to the multiply charged NA samples in the study, in which a distributed charge arrangement was implemented by spreading the discrete charged sites over the entire span of the model. The outcome may not comprise a well-defined dipole, but rather multiple localized ones that are likely to manifest different interaction modes with polarizable gas. Accounting for local interactions will require combining knowledge of sample topology with charge localization, which is currently not included in the available CCS algorithms. These considerations may help explain the striking deviations between computational and experimental CCSs obtained from either algorithm employed in the study, and point toward a promising direction for future investigations.

Conclusions

The goal of this study was to explore the merits of modular NA standards that, mirroring the function of MS standards employed to calibrate the m/z scale, would enable the generation of effective calibration curves for translating primary ion mobility readouts into valid CCS data. Major impetus was provided by the absence of any controlled set of NA standards

extending beyond ~30 bp in size, and by the concerns associated with the problematic alternative of utilizing protein references for which reliable data are available only in positive ion mode. The selected species consisted of ODNs of increasing size and mixed base composition, which mimicked as closely as possible natural samples. Repeating sequences were utilized to facilitate the modeling of regular, expandable ladders devoid of unexpected secondary structures. In an effort to minimize possible sources of uncertainty, reference CCS data of disparate provenance were purposely eschewed in favor of values obtained in house by applying a single computational approach to all selected standards and test samples. The utilization of self-consistent computational references allowed us to compare the results afforded by the EHSS and PSA algorithms and look for the possible influence of sample topology on calibration performance. The standards covered ranges of CCS spanning from 320 to 4,900 Å² and 641 to 5,700 Å² for EHSS and PSA datasets, respectively, which far exceeded the ~2,400 Å² afforded by the largest NA sample analyzed to date by IMS-MS.³³

The linear regressions obtained by plotting each “measured” quantity (t_D readout) against the corresponding “known” value (computational CCS) provided excellent correlation coefficients and predictive power, as observed in particular for the curve generated from PSA references (**Figure 5**). The utilization of test constructs that mimicked possible unknowns revealed that curves produced from different datasets displayed very distinctive properties. In general, the curves obtained from PSA data consistently provided CCS values that were much closer to the expected ones than those obtained from curves based on EHSS data. Whereas the PSA curve comprising both ss and ds calibrant sets provided consistently the best results regardless the topology of the test analyte, the EHSS counterpart displayed a noticeable dependence on sample topology, which led to better performance when matching calibrant and sample topologies.

It is interesting to note how both EHSS and PSA curves performed better with ss than ds analytes, and significantly better with these types of constructs than with hp structures (**Table 3**). While the EHSS/PSA discrepancies are clearly ascribable to the different abilities of these algorithms to capture conformation, the discrepancies between the results obtained from ss, ds and hp constructs point towards intrinsic challenges in representing these structures in solvent-free environment. In this respect, we must point out that all ss constructs in the study were purposely unable to fold stable higher-order structures and, thus, did not possess recognizable folded/unfolded states. At the same time, double-helical and stem-loop structures were the only stable states afforded by intact ds and hp constructs, and their experimental analysis was conducted under the most gentle possible conditions to avoid any perturbation. It appears that the models used to represent unstructured ss species fared better than those employed for describing far more well-defined structures. Previous studies of ds species have identified clear boundaries between A and B-form helices in a solvent-free environment, which were not expected from the established knowledge of solution structure.³³ Our data suggests that the future investigation of progressively larger and more complex structures may reserve additional surprises.

In conclusion, the combination of the calibrant sets described here will be expected to represent an excellent aid for the practical application of IMS-MS to NA analysis. The

computational origin of the reference CCS values makes such tool completely independent from hardware platforms and analytical conditions. In particular, this feature will enable the direct utilization of the calibrant sets on both drift tube and traveling wave analyzers. Although our determinations were performed exclusively in a N₂ background, reference values calculated in both N₂ and He were provided for any interested party (see **Supporting Material, Table 3S**). The self-consistent origin of the calibrant sets places unknown samples on the same scale to allow a fair evaluation of conformational changes as a function of experimental variables and to support the investigation of structure and dynamics. Going forward, the self-consistent character of the references will allow for efficient upgrades of the calculated values, as the knowledge of gas-phase structure and IMS behavior of NA species will increase over time. These types of biopolymers are still far less understood than peptide/protein systems, which have received most of the attention from the IMS-MS community in recent years. A better knowledge of their behavior will lead not only to better parameterization for these biopolymers in current algorithms, but also to the development of better algorithms that may be capable of accounting for possible charge localization over complex structure topologies. Any advance in these directions will be expected to further increase the accuracy of the calibration curves for unknown determinations. At the same time, the modular nature of the calibrant sets will enable the effortless extension of the range covered by the calibration curves to better serve the analysis of progressively larger NA samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Dr. Maria Basanta-Sanchez for helpful preliminary discussions and training on the utilization of the IMS-MS instrument. Funding was provided by the National Institute of Health with award GM064328.

References

1. Hill HH, Siems WF, St. Louis RH, McMinn DG. Ion mobility spectrometry. *Anal. Chem.* 1990; 62:1201A–1209A.
2. Clemmer DE, Jarrold MF. Ion mobility measurements and their applications to cluster biomolecules. *J. Mass Spectrom.* 1997; 32:577–592.
3. Eiceman, GA, Karpas, Z. *Ion Mobility Spectrometry*. Second Edition.. CRC Press; 2005.
4. Verbeck GF, Ruotolo BT, Sawyer HA, Gillig KJ, Russell DH. A fundamental introduction to ion mobility mass spectrometry applied to the analysis of biomolecules. *J Biomol Tech.* 2002; 13:56–61. [PubMed: 19498967]
5. Smith DP, et al. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. *Eur J Mass Spectrom (Chichester, Eng).* 2009; 15:113–130.
6. Salbo R, et al. Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Commun. Mass Spectrom.* 2012; 16:1181–1193.
7. Ruotolo BT, Benesch JL, Sandercock AM, Hyung S-J, Robinson CV. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat Protoc.* 2008; 3:1139–1152. [PubMed: 18600219]

8. Mason EA, McDaniel EW. Transport properties of ions in gases. NASA STI/Recon Technical Report A. 1988; 89:15174.
9. Liu Y, Valentine SJ, Counterman AE, Hoaglund CS, Clemmer DE. Injected-Ion Mobility Analysis of Biomolecules. *Anal. Chem.* 1997; 69:728A–735A.
10. McLean JA, Ruotolo BT, Gillig KJ, Russell DH. Ion mobility–mass spectrometry: a new paradigm for proteomics. *International Journal of Mass Spectrometry.* 2005; 240:301–315.
11. Thalassinos K, et al. Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *Int J Mass Spectrom.* 2004; 236:55–63.
12. Shvartsburg AA, Smith RD. Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* 2008; 80:9689–9699. [PubMed: 18986171]
13. Giles K, Wildgoose JL, Langridge DJ, Campuzano I. A method for direct measurement of ion mobilities using a travelling wave ion guide. *International Journal of Mass Spectrometry.* 2010; 298:10–16.
14. May JC, McLean JA. The Influence of Drift Gas Composition on the Separation Mechanism in Traveling Wave Ion Mobility Spectrometry: Insight from Electrodynamical Simulations. *Int J Ion Mobil Spectrom.* 2003; 16:85–94. [PubMed: 23888124]
15. Michaelievski I, Eisenstein M, Sharon, M. Gas-Phase Compaction and Unfolding of Protein Structures. *Anal. Chem.* 2010; 82:9484–9491. [PubMed: 20964410]
16. Morsa D, Gabelica V, De Pauw E. Effective Temperature of Ions in Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* 2011; 83:5775–5782. [PubMed: 21682304]
17. Bush MF, Campuzano I, Robinson CV. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Anal. Chem.* 2012; 84:7124–7130. [PubMed: 22845859]
18. Jurczek E, Kalapothakis J, Campuzano IDG, Morris M, Barran PE. Effects of Drift Gas on Collision Cross Sections of a Protein Standard in Linear Drift Tube and Traveling Wave Ion Mobility Mass Spectrometry. *Anal. Chem.* 2012; 84:8524–8531. [PubMed: 22974196]
19. Hamilton JV, Renaud JB, Mayer PM. Experiment and theory combine to produce a practical negative ion calibration set for collision cross-section determinations by travellingwave ion-mobility mass spectrometry. *Rapid Commun. Mass Spectrom.* 2012; 26:1591–1595.
20. D'Atri V, Porini M, Rosu F, Gabelica V. Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass Spectrom.* 2015; 50:711–726. [PubMed: 26259654]
21. Bush MF, Hall Z, Giles K, Hoyes J, Robinson CV. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* 2010; 82:9557. [PubMed: 20979392]
22. Sun Y, Vahidi S, Sowole MA, Konermann L. Protein Structural Studies by Traveling Wave Ion Mobility Spectrometry: A Critical Look at Electrospray Sources and Calibration Issues. *J Am Soc Mass Spectrom.* 2016; 27:21–40.
23. Lippens JL, Mangrum JB, McIntyre W, Redick B, Fabris D. A simple heated-capillary modification improves the analysis of non-covalent complexes by Z-spray electrospray ionization. *Rapid Commun. Mass Spectrom.* 2016; 30:773–783. [PubMed: 26864529]
24. Mesleh MF, Hunter JM, Shvartsburg AA, Schatz GC, Jarrold MF. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J Phys Chem.* 1996; 100:16082–19086.
25. Shvartsburg AA, Jarrold MF. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Letters.* 1996; 261:86–91.
26. Shvartsburg AA, Mashkevich S, Baker E, Smith R. Optimization of Algorithms for Ion Mobility Calculations. *J Phys Chem.* 2007; 111:2002–2010.
27. Bleiholder C, Wytenbach T, Bowers MT. 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.
28. Anderson SE, Bleiholder C, Brocker ER, Stang PJ, Bowers MT. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (II). Application to Supramolecular Coordination-Driven Assemblies. *Int J Mass Spectrom.* 2012; 330-332:78–84.

29. Bleiholder C, Contreras S, Do TD, Bowers MT. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (III). Model parameterization and definition of empirical shape factors for proteins. *Int J Mass Spectrom.* 2013; 345-347:89–96.
30. Bleiholder C, Contreras S, Bowers MT. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (IV). Application to polypeptides. *International Journal of Mass Spectrometry.* 2013; 354–355:275–280.
31. Gidden J, Ferzoco A, Baker ES, Bowers MT. Duplex formation and the onset of helicity in poly d(CG)n oligonucleotides in a solvent-free environment. *J. Am. Chem. Soc.* 2004; 126:15132–15140. [PubMed: 15548010]
32. Baker ES, Bernstein SL, Gabelica V, De Pauw E, Bowers MT. G-quadruplexes in telomeric repeats are conserved in a solvent-free environment. *Int J Mass Spectrom.* 2006; 253:225–237.
33. Baker ES, Bowers MT. B-DNA helix stability in a solvent-free environment. *Journal of the American Society for Mass Spectrometry.* 2007; 18:1188–95. [PubMed: 17434745]
34. Wyttenbach T, Pierson NA, Clemmer DE, Bowers MT. Ion Mobility Analysis of Molecular Dynamics. *Annual Review of Physical Chemistry.* 2014; 65:175–196.
35. Mehmood S, Allison TM, Robinson CV. Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annu. Rev. Phys. Chem.* 2015; 66:453–474. [PubMed: 25594852]
36. Pyle AM. Ribozymes: a distinct class of metalloenzymes. *Science (New York, N.Y.)* 1993; 261:709–14.
37. Chang KY, Varani G. Nucleic acids structure and recognition. *Nat. Struct. Biol.* 1997; 4(Suppl): 854–858. [PubMed: 9377158]
38. Grosjean, H. Fine-tuning of RNA functions by modification and editing. Springer; 2005.
39. Strobel SA, Cochrane JC. RNA catalysis: ribozymes, ribosomes, and riboswitches. *Curr Opin Chem Biol.* 2007; 11:636–643. [PubMed: 17981494]
40. Leontis, N, Westhof, E. RNA 3D Structure Analysis and Prediction (Nucleic Acids and Molecular Biology). Springer; 2012.
41. Cruz JA, et al. RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA.* 2012; 18:610–625. [PubMed: 22361291]
42. Miao Z, et al. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. *RNA.* 2015; 21:1066–1084. [PubMed: 25883046]
43. Ganem B, Li Y-T, Henion JD. Detection of oligonucleotide duplex forms by ion-spray mass spectrometry. *Tetrahedron Lett.* 1993; 34:1445–1448.
44. Schnier PD, Klassen JS, Strittmatter EF, Williams ER. Activation Energies for Dissociation of Double Strand Oligonucleotide Anions: Evidence for Watson-Crick Base Pairing in Vacuo. *J Am Chem Soc.* 1998; 120:9605–9613. [PubMed: 16498487]
45. Gabelica V, De Pauw E. Comparison between solution-phase stability and gas-phase kinetic stability of oligodeoxynucleotide duplexes. *J Mass Spectrom.* 2001; 36:397–402. [PubMed: 11333443]
46. Stephenson W, et al. The Essential Role of Stacking Adenines in a Two-Base-Pair RNA Kissing Complex. *J. Am. Chem. Soc.* 2013; 135:5602–5611. [PubMed: 23517345]
47. Mo J, Hakansson K. Characterization of nucleic acid higher order structure by high-resolution tandem mass spectrometry. *Analytical and bioanalytical chemistry.* 2006; 386:675–81. [PubMed: 16855815]
48. Turner KB, Hagan NA, Fabris D. Understanding the isomerization of the HIV-1 dimerization initiation domain by the nucleocapsid protein. *J. Mol. Biol.* 2007; 369:812–828. [PubMed: 17466332]
49. Balbeur D, Widart J, Leyh B, Cravello L, De Pauw E. Detection of oligonucleotide gas-phase conformers: H/D exchange and ion mobility as complementary techniques. *Journal of the American Society for Mass Spectrometry.* 2008; 19:938–46. [PubMed: 18467119]
50. Mo J, Todd GC, Hakansson K. Characterization of nucleic acid higher order structure by gas-phase H/D exchange in a quadrupole-FT-ICR mass spectrometer. *Biopolymers.* 2009; 91:256–64. [PubMed: 19140156]

51. Baker ES, Dupuis NF, Bowers MT. DNA hairpin, pseudoknot, and cruciform stability in a solvent-free environment. *The journal of physical chemistry*. 2009; 113:1722–7. [PubMed: 19193169]
52. Fabris D, Kellersberger KA, Wilhide JA. Higher-order structure of nucleic acids in the gas phase: Top-down analysis of base-pairing interactions. *International Journal of Mass Spectrometry*. 2012; 312:155–162. [PubMed: 24027423]
53. Gabelica V, Baker ES, Teulade-Fichou MP, De Pauw E, Bowers MT. Stabilization and structure of telomeric and c-myc region intramolecular G-quadruplexes: the role of central cations and small planar ligands. *Journal of the American Chemical Society*. 2007; 129:895–904. [PubMed: 17243826]
54. Arcella A, et al. Structure of Triplex DNA in the Gas Phase. *J. Am. Chem. Soc.* 2012; 134:6596–6606. [PubMed: 22420643]
55. Balthasart F, Plavec J, Gabelica V. Ammonium ion binding to DNA G-quadruplexes: do electrospray mass spectra faithfully reflect the solution-phase species? *J. Am. Soc. Mass Spectrom.* 2013; 24:1–8. [PubMed: 23132414]
56. Zhou S, Cook KD. Protonation in electrospray mass spectrometry: wrong-way-round or right-way-round? *J. Am. Soc. Mass Spectrom.* 2000; 11:961–966. [PubMed: 11073259]
57. Nordhoff E, Kirpekar F, Roepstorff P. Mass spectrometry of nucleic acids. *Mass Spectrom. Rev.* 1996; 15:67–138. [PubMed: 27082318]
58. Fabris, D, Turner, KB, Hagan, NA. *Mass Spectrometry of Nucleosides and Nucleic Acids*. Banoub, J, Limbach, P, editors. CRC Press, Taylor and Francis Group, LLC; London, U.K.: 2010. 303–327.
59. Kelly MA, Vestling MM, Fenselau C. Electrospray of proteins: a comparison of positive-ion and negative-ion mass spectra at high and low pH. *Org. Mass Spectrom.* 1992; 27:1143–1147.
60. Forsythe JG, et al. Collision cross section calibrants for negative ion mode traveling wave ion mobility-mass spectrometry. *Analyst*. 2015; 140:6853–6861. [PubMed: 26148962]
61. Hoaglund CS, Liu Y, Ellington AD, Pagel M, Clemmer DE. Gas-Phase DNA: Oligothymidine Ion Conformers. *J. Am. Chem. Soc.* 1997; 119:9051–9052.
62. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research*. 2003; 31:3406–15. [PubMed: 12824337]
63. Ding Y, Chan CY, Lawrence CE. Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.* 2004; 32:W135–141. [PubMed: 15215366]
64. Reeder J, Steffen P, Giegerich R. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. *Nucleic acids research*. 2007; 35:W320–4. [PubMed: 17478505]
65. Giles K, et al. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* 2004; 18:2401–2414. [PubMed: 15386629]
66. Giles K, Williams JP, Campuzano I. Enhancements in travelling wave ion mobility resolution. *Rapid Commun. Mass Spectrom.* 2011; 25:1559–1566. [PubMed: 21594930]
67. Van der Spoel D, et al. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* 2005; 16:1701–1718.
68. Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* 2013; 3:198–210.
69. Chen AA, Garcia AE. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *PNAS*. 2013; 110:16820–16825. [PubMed: 24043821]
70. Rueda M, Luque FJ, Orozco M. G-Quadruplexes Can Maintain Their Structure in the Gas Phase. *J. Am. Chem. Soc.* 2006; 128:3608–3619. [PubMed: 16536534]
71. Perez A, Luque FJ, Orozco M. Dynamics of B-DNA on the Microsecond Time Scale. *J. Am. Chem. Soc.* 2007; 129:14739–14745. [PubMed: 17985896]
72. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature*. 2007; 450:964–972. [PubMed: 18075575]
73. Knapman TW, Berryman JT, Campuzano I, Harris SA, Ashcroft AE. Considerations in experimental and theoretical collision cross-section measurements of small molecules using travelling wave ion mobility spectrometry-mass spectrometry. *Int J Mass Spectrom.* 2010; 298:17–23.

74. Rueda M, Kalko SG, Luque FJ, Orozco M. The structure and dynamics of DNA in the gas phase. *J. Am. Chem. Soc.* 2003; 125:8007–8014. [PubMed: 12823023]
75. Nourse BD, Kenttamaa HI. Effective Ion Temperatures in a Quadrupole ion Trap. *Journal of Physical Chemistry.* 1990; 94:5809–5812.
76. Gronert S. Estimation of Effective Ion Temperatures in a Quadrupole Ion Trap. *Journal of the American Society for Mass Spectrometry.* 1998; 9:845–848.
77. Gabelica V, Karas M, De Pauw E. Calibration of ion effective temperatures achieved by resonant activation in a quadrupole ion trap. *Anal. Chem.* 2003; 75:5152–5159. [PubMed: 14708790]
78. Wu J, McLuckey SA. Gas-phase fragmentation of oligonucleotide ions. *International Journal of Mass Spectrometry.* 2004; 237:197–241.
79. Burmistrova A, Gabelica V, Duwez AS, De Pauw E. Ion Mobility Spectrometry Reveals Duplex DNA Dissociation Intermediates. *J Am Soc Mass Spectrom.* 2013; 24:1777–1786. [PubMed: 24009017]
80. Bloomfield, VA, Crothers, DM, Tinoco, I, Jr.. *Nucleic Acids - Structures, Properties, and Functions.* University Science Books; 2000.
81. Rueda M, Luque, Orozco M. Nature of Minor-Groove Binders–DNA Complexes in the Gas Phase. *J. Am. Chem. Soc.* 2005; 127:11690–11698. [PubMed: 16104746]
82. Wilm MS, Mann M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Proc.* 1994; 136:167–180.
83. Cheng X, Gale DC, Udseth HR, Smith RD. Charge State Reduction of Oligonucleotide Negative Ions from Electrospray Ionization. *Anal. Chem.* 1995; 67:586–593.
84. Greig M, Griffey RH. Utility of organic bases for improved electrospray mass spectrometry of oligonucleotides. *Rapid Commun. Mass Spectrom.* 1995; 9:97–102. [PubMed: 7888709]
85. Zhao Q, et al. An Ion Trap-Ion Mobility-Time of Flight Mass Spectrometer with Three Ion Sources for Ion/Ion Reactions. *J Am Soc Mass Spectrom.* 2009; 20:1549–1561. [PubMed: 19493684]
86. Kharlamova A, Prentice BM, Huang T, McLuckey SA. Electrospray droplet exposure to gaseous acids for reduction of metal counter-ions in nucleic acid ions. *Int J Mass Spectrom.* 2011; 300:158–166.
87. Badman ER, Hoaglund CS, Clemmer DE. Monitoring Structural Changes of Proteins in an Ion Trap over ~10–200 ms: Unfolding Transitions in Cytochrome c Ions. *Anal. Chem.* 2001; 73:6000–6007. [PubMed: 11791572]
88. Myung S, Badman ER, Lee YJ, Clemmer DE. Structural Transitions of Electrosprayed Ubiquitin Ions Stored in an Ion Trap over ~10 ms to 30 s. *J. Phys. Chem. A.* 2002; 106:9976–9982.
89. Gidden J, Baker ES, Ferzoco A, Bowers MT. Structural motifs of DNA complexes in the gas phase. *International Journal of Mass Spectrometry.* 2005; 240:183–193.
90. Beegle LW, Kanik I, Matz L, Hill HH Jr. Effects of drift-gas polarizability on glycine peptides in ion mobility spectrometry. *International Journal of Mass Spectrometry.* 2002; 216:257–268.
91. Matz LM, Hill HH, Beegle LW, Kanik I. Investigation of drift gas selectivity in high resolution ion mobility spectrometry with mass spectrometry detection. *J. Am. Soc. Mass Spectrom.* 2002; 13:300–307. [PubMed: 11951967]
92. Gelb AS, Jarratt RE, Huang Y, Dodds ED. A study of calibrant selection in measurement of carbohydrate and peptide ion-neutral collision cross sections by traveling wave ion mobility spectrometry. *Anal. Chem.* 2014; 86:11396–11402. [PubMed: 25329513]
93. Ruotolo BT, McLean JA, Gillig KJ, Russell DH. Peak capacity of ion mobility mass spectrometry: the utility of varying drift gas polarizability for the separation of tryptic peptides. *J Mass Spectrom.* 2004; 39:361–367. [PubMed: 15103649]

Figure 1.
Representative members of the ds series illustrate the modular nature of the calibrant ladder.

Figure 2. Variation of radius of gyration (R_g) as a function of simulation time observed for the **14a:b** and **64a:b** duplexes. Solid lines represent solution conditions, whereas dashed lines correspond to solvent-free conditions.

Figure 3.
ESI-MS spectra obtained from the a) **64b** and b) **32a:b** samples.

Figure 4. ATD's obtained from **a)** –8 charge state of **64b**; **b)** –8 of **32a:b**; and **c)** –9 of **64b**. Red lines are Gaussian-fitted curves generated by Peakfit (see *Experimental*).

Figure 5. Calibration curves obtained by correlating experimental t_D'' data with the corresponding reference values provided by the EHSS (■) and PSA (◆) methods.

Figure 6. Calibration curves obtained from separate ss and ds calibrant sets. Panel **a)** and **b)** contain curves generated by using EHSS (■) and PSA (◆) reference values, respectively.

Table 1

Experimental drift time (t_D) and corresponding computational CCS value ($\text{\AA}^2$) obtained from ss standards. The value of t_D represents the “measured” quantity that was obtained directly from raw t_D according to previously described protocol.⁷ Each CCS value represents the average and relative standard deviation (RSD %) of data provided by 20 individual configurations in the simulation trajectory (see *Experimental*). This value represents the “known” reference quantity.

ss series	z	t_D (ms)	CCS _{EHSS} ($\text{\AA}^2$) $\pm$ RSD%	CCS _{PSA} ($\text{\AA}^2$) $\pm$ RSD%
6mer	3	1.53	322.9 $\pm$ 1.1%	N/A
14b	4	3.41	743.7 $\pm$ 5.7%	875.5 $\pm$ 2.1%
24b	4	5.75	992.1 $\pm$ 1.9%	1240.2 $\pm$ 2.0%
	5	4.97	1000.6 $\pm$ 0.9%	1272.5 $\pm$ 1.5%
32b	5	6.75	1217.8 $\pm$ 1.6%	1480.5 $\pm$ 2.8%
	6	6.13	1367.9 $\pm$ 0.5%	1677.4 $\pm$ 1.5%
48b	7	8.41	1869.7 $\pm$ 1.3%	2266.0 $\pm$ 1.8%
64b	8	10.78	2507.9 $\pm$ 0.6%	3023.8 $\pm$ 1.0%
80b	9	12.69	2868.3 $\pm$ 0.6%	3446.9 $\pm$ 1.3%
96b	10	14.73	3505.1 $\pm$ 0.6%	4137.5 $\pm$ 8.5%

Table 2

Experimental drift time (t_D) and corresponding computational CCS values ($\text{\AA}^2$) obtained from ds standards. The value of t_D represents the “measured” quantity that was obtained directly from raw t_D according to previously described protocol.⁷ Each CCS value represents the average and relative standard deviation (RSD %) of data provided by 20 individual configurations in the simulation trajectory (see *Experimental*). This value represents the “known” reference quantity.

ds series	z	t_D (ms)	CCS _{EHSS} ($\text{\AA}^2$) $\pm$ RSD%	CCS _{PSA} ($\text{\AA}^2$) $\pm$ RSD%
(6mer) ₂	3	3.63	496.9 $\pm$ 1.0%	640.8 $\pm$ 1.4%
14a:b	5	6.64	967.0 $\pm$ 1.4%	1154.9 $\pm$ 1.3%
18a:b	6	9.22	1405.1 $\pm$ 0.8%	1704.0 $\pm$ 1.3%
22a:b	7	11.17	1791.8 $\pm$ 0.6%	2129.5 $\pm$ 1.4%
	8	14.08	1746.5 $\pm$ 1.4%	2045.0 $\pm$ 1.9%
24a:b	7	11.61	1868.4 $\pm$ 0.5%	2199.5 $\pm$ 1.5%
32a:b	8	14.99	2446.2 $\pm$ 1.0%	2924.6 $\pm$ 1.4%
	9	14.43	2432.5 $\pm$ 0.4%	2877.9 $\pm$ 1.3%
48a:b	10	21.71	3623.7 $\pm$ 0.4%	4220.3 $\pm$ 1.3%
	11	23.77	3668.8 $\pm$ 0.4%	4298.7 $\pm$ 1.7%
64a:b	12	32.05	4654.8 $\pm$ 0.7%	5401.0 $\pm$ 2.0%
	13	33.89	4810.8 $\pm$ 0.7%	5563.2 $\pm$ 1.2%
	14	35.93	4908.0 $\pm$ 0.7%	5716.2 $\pm$ 1.3%

Table 3

Predictive power of calibration curves. Separate test sets consisting of ss, ds, and hp constructs were employed to evaluate the ability of EHSS- and PSA-based regressions to translate experimental t_D values into corresponding CCSs. Computational CCS values (**Comp.**) were calculated for the test samples by using the EHSS and PSA algorithms (see *Experimental*). Experimental CCS values (**Exp.**) were instead obtained from corresponding t_D determinations by using either calibration curve provided in **Fig. 5**. Experimental values were the average and RSD% obtained from three replicate analyses. Relative deviations (%) were calculated from corresponding computational/experimental couples. Root mean square deviations (RMSD) were calculated from the % values produced by each calibration curve for the members of the separate test sets. Taken together, the EHSS and PSA curves afforded overall RMSD values of 16.6 and 16.9% for all the test samples in the study. Exclusion of the hp constructs afforded overall values of 5.8 and 3.4% RMSD.

Test	z	EHSS				PSA			
		Comp. CCS ($\text{\AA}^2$)	Exp. CCS ($\text{\AA}^2$)	%	RMSD	Comp. CCS ($\text{\AA}^2$)	Exp. CCS ($\text{\AA}^2$)	%	RMSD
16b	4	688.6 $\pm$ 3.4%	734.0 $\pm$ 0.5%	+6.6%	$\pm$ 4.6%	894.9 $\pm$ 3.6%	881.7 $\pm$ 0.4%	−1.5%	$\pm$ 1.0%
30a	5	1324.4 $\pm$ 0.9%	1378.8 $\pm$ 0.4%	+4.1%		1629.6 $\pm$ 1.8%	1629.7 $\pm$ 0.3%	+0.1%	
56a	7	2128.3 $\pm$ 0.8%	2140.1 $\pm$ 0.2%	+0.6%		2501.3 $\pm$ 1.5%	2537.1 $\pm$ 0.2%	+1.4%	
70b	8	2641.6 $\pm$ 3.7%	2516.9 $\pm$ 0.3%	−4.7%		2995.7 $\pm$ 3.0%	2986.1 $\pm$ 0.3%	−0.4%	
10a:b	5	775.8 $\pm$ 1.9%	691.9 $\pm$ 0.6%	−10.8%	$\pm$ 6.8%	873.8 $\pm$ 1.3%	860.2 $\pm$ 0.5%	−1.6%	$\pm$ 4.7%
30a:b	7	2399.3 $\pm$ 0.7%	2412.9 $\pm$ 0.5%	+0.6%		2846.7 $\pm$ 1.7%	2827.6 $\pm$ 0.5%	−0.7%	
36a:b	10	2797.5 $\pm$ 0.5%	2746.5 $\pm$ 0.9%	−1.8%		3307.9 $\pm$ 1.5%	3320.2 $\pm$ 0.9%	−0.4%	
56a:b	12	4261.0 $\pm$ 0.4%	4599.9 $\pm$ 0.6%	+8.0%		4937.0 $\pm$ 1.7%	5389.0 $\pm$ 0.6%	+9.2%	
16hp	4	589.1 $\pm$ 1.0%	839.8 $\pm$ 1.7%	+41.2%	$\pm$ 27.5%	714.2 $\pm$ 2.4%	1000.5 $\pm$ 1.6%	+40.1%	$\pm$ 28.4%
28hp	5	965.7 $\pm$ 0.8%	1211.9 $\pm$ 0.1%	+25.5%		1144.2 $\pm$ 1.9%	1447.7 $\pm$ 0.1%	+26.5%	
36hp	6	1208.4 $\pm$ 0.6%	1453.8 $\pm$ 1.4%	+20.3%		1412.9 $\pm$ 3.1%	1750.0 $\pm$ 1.3%	+23.9%	
48hp	7	1589.7 $\pm$ 0.6%	1790.6 $\pm$ 0.6%	+12.6%		1821.2 $\pm$ 1.1%	2158.4 $\pm$ 0.6%	+18.5%	

Table 4

Predictive power as a function of structure topology. The ss, ds, and hp samples were employed to assess calibration curves generated from either ss or ds calibrant sets. All computational and reference values were obtained by using the EHSS algorithm. Experimental CCS values are the average and relative standard deviation (RSD%) obtained from three replicate analyses. Relative deviations (%) were calculated from corresponding computational/experimental couples. Root mean square deviations (RMSD) were calculated from the % values produced by each calibration curve for the members of the separate test sets.

Test	z	Comp. CCS ($\text{\AA}^2$)	EHSS					
			ss curve			ds curve		
			Exp. CCS ($\text{\AA}^2$)	%	RMSD	Exp. CCS ($\text{\AA}^2$)	%	RMSD
16b	4	688.6 $\pm$ 3.4%	689.5 $\pm$ 0.5%	+0.1%	$\pm$ 2.2%	817.3 $\pm$ 0.4%	+18.7%	$\pm$ 10.7%
30a	5	1324.4 $\pm$ 0.9%	1362.7 $\pm$ 0.4%	+2.9%		1431.9 $\pm$ 0.4%	+8.1%	
56a	7	2128.3 $\pm$ 0.8%	2171.1 $\pm$ 0.2%	+2.0%		2130.1 $\pm$ 0.2%	+0.1%	
70b	8	2641.6 $\pm$ 3.7%	2571.1 $\pm$ 0.3%	-2.7%		2475.8 $\pm$ 0.3%	-6.3%	
10a:b	5	775.8 $\pm$ 1.9%	660.8 $\pm$ 0.6%	-14.8%	$\pm$ 9.4%	747.5 $\pm$ 0.5%	-3.6%	$\pm$ 4.8%
30a:b	7	2399.3 $\pm$ 0.7%	2440.8 $\pm$ 0.5%	+1.7%		2428.2 $\pm$ 0.6%	+1.2%	
36a:b	10	2797.5 $\pm$ 0.5%	2849.5 $\pm$ 0.9%	+1.9%		2605.8 $\pm$ 1.1%	-6.9%	
56a:b	12	4261.0 $\pm$ 0.4%	4736.9 $\pm$ 0.6%	+11.2%		4501.4 $\pm$ 0.7%	+5.6%	
16hp	4	589.1 $\pm$ 1.0%	797.8 $\pm$ 1.8%	+35.4%	$\pm$ 24.8%	924.0 $\pm$ 1.6%	+56.9%	$\pm$ 34.2%
28hp	5	965.7 $\pm$ 0.8%	1195.1 $\pm$ 0.1%	+23.8%		1257.4 $\pm$ 0.1%	+30.2%	
36hp	6	1208.4 $\pm$ 0.6%	1460.5 $\pm$ 1.4%	+20.9%		1459.5 $\pm$ 1.4%	+20.8%	
48hp	7	1589.7 $\pm$ 0.6%	1821.6 $\pm$ 0.6%	+14.6%		1759.7 $\pm$ 0.6%	+10.7%	

Table 5

Predictive power as a function of structure topology. The ss, ds, and hp samples were employed to assess calibration curves generated from either ss or ds calibrant sets. All computational and reference values were obtained by using the PSA algorithm. Experimental CCS values are the average and relative standard deviation (RSD%) obtained from three replicate analyses. Relative deviations (%) were calculated from corresponding computational/experimental couples. Root mean square deviations (RMSD) were calculated from the % values produced by each calibration curve for the members of the separate test sets.

Test	z	PSA						
		Comp. CCS (Å ²)	ss curve			ds curve		
			Exp. CCS (Å ²)	%	RMSD	Exp. CCS (Å ²)	%	RMSD
16b	4	894.9 ± 3.6%	789.9 ± 0.5%	−11.7%	±6.4%	963.0 ± 0.4%	+9.8%	±5.4%
30a	5	1629.6 ± 1.8%	1569.7 ± 0.3%	−3.7%		1694.6 ± 0.3%	+4.0%	
56a	7	2501.3 ± 1.5%	2573.8 ± 0.2%	+2.9%		2517.4 ± 0.2%	+0.6%	
70b	8	2995.7 ± 3.0%	3071.2 ± 0.2%	+2.5%		2924.6 ± 0.5%	−2.4%	
10a:b	5	873.8 ± 1.3%	836.0 ± 0.6%	−4.3%	±7.9%	917.5 ± 0.5%	+5.0%	±4.9%
30a:b	7	2846.7 ± 1.7%	2827.5 ± 0.4%	−0.7%		2840.6 ± 0.5%	−0.2%	
36a:b	10	3307.9 ± 1.5%	3564.3 ± 0.7%	+7.8%		3124.3 ± 1.0%	−5.6%	
56a:b	12	4937.0 ± 1.7%	5580.3 ± 0.5%	+13.0%		5243.3 ± 0.7%	+6.2%	
16hp	4	714.2 ± 2.4%	903.4 ± 1.7%	+26.5%	±24.4%	1102.5 ± 1.5%	+54.4%	±35.9%
28hp	5	1144.2 ± 1.9%	1404.0 ± 0.1%	+22.7%		1501.1 ± 0.1%	+31.2%	
36hp	6	1412.9 ± 3.1%	1770.9 ± 1.3%	+25.3%		1750.6 ± 1.3%	+23.9%	
48hp	7	1821.2 ± 1.1%	2233.5 ± 0.5%	+22.6%		2291.3 ± 13.3%	+25.8%	