# Distance Geometry Protocol to Generate Conformations of Natural Products to Structurally Interpret Ion Mobility-Mass Spectrometry Collision Cross Sections

Sarah M. Stow,[†,§,‖] Cody R. Goodwin,[†,§,‖] Michal Kliman,[†,§,‖] Brian O. Bachmann,[†,§,‖] John A. McLean,*[,†,§,‖] and Terry P. Lybrand*[,†,‡,§,⊥]

[†]Department of Chemistry, [‡]Department of Pharmacology, [§]Vanderbilt Institute of Chemical Biology, [‖]Vanderbilt Institute of Integrative Biosystems Research and Education, [⊥]Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, United States

ⓢ *Supporting Information*

ABSTRACT: Ion mobility-mass spectrometry (IM-MS) allows the separation of ionized molecules based on their charge-to-surface area (IM) and mass-to-charge ratio (MS), respectively. The IM drift time data that is obtained is used to calculate the ion-neutral collision cross section (CCS) of the ionized molecule with the neutral drift gas, which is directly related to the ion conformation and hence molecular size and shape. Studying the conformational landscape of these ionized molecules computationally provides interpretation to delineate the potential structures that these CCS values could represent, or conversely, structural motifs not consistent with the IM data. A challenge in the IM-MS community is the ability to rapidly compute conformations to interpret natural product data, a class of molecules exhibiting a broad range of biological activity. The diversity of biological activity is, in part, related to the unique structural characteristics often observed for natural products. Contemporary approaches to structurally interpret IM-MS data for peptides and proteins typically utilize molecular dynamics (MD) simulations to sample conformational space. However, MD calculations are computationally expensive, they require a force field that accurately describes the molecule of interest, and there is no simple metric that indicates when sufficient conformational sampling has been achieved. Distance geometry is a computationally inexpensive approach that creates conformations based on sampling different pairwise distances between the atoms within the molecule and therefore does not require a force field. Progressively larger distance bounds can be used in distance geometry calculations, providing in principle a strategy to assess when all plausible conformations have been sampled. Our results suggest that distance geometry is a computationally efficient and potentially superior strategy for conformational analysis of natural products to interpret gas-phase CCS data.
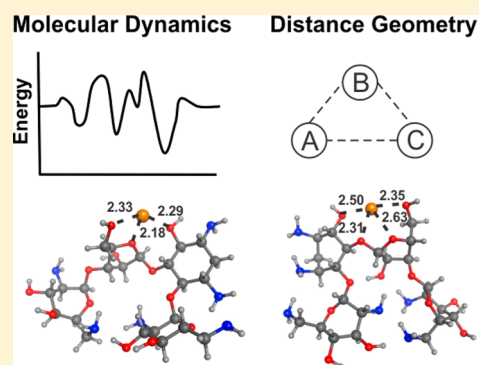
INTRODUCTION

Ion mobility-mass spectrometry (IM-MS) is an analytical technique used to separate gas-phase ions based on their structural properties such as size and shape as well as their mass, in the IM and MS dimensions, respectively. The structural properties affect the ion's collision cross section (CCS), or rotationally averaged surface area.[1−3] Ongoing efforts in our laboratories utilize IM-MS to aid in natural product discovery from bacterial colonies.[4,5] IM-MS is often able to structurally separate the low-abundance secondary metabolites from the complex biological background, a key challenge in natural product discovery by MS.

In an effort to help elucidate the structural information derived from CCS data, computational methods are often used to interpret IM-MS experiments.[6−9] A computational algorithm is used to generate conformations of the molecule, defining its conformational space. Then, a theoretical CCS is calculated for each of the conformations. Conformations that fall within the experimental CCS range can then be further interrogated to

Table 1. Representative IM-MS Structural Studies of Different Model Systems

computational methods	model system
coarse-grained modeling	protein complexes[a] virus[b]
molecular dynamics (simulated annealing, replica exchange, etc.)	peptides[c] carbohydrates[d] natural products[e]

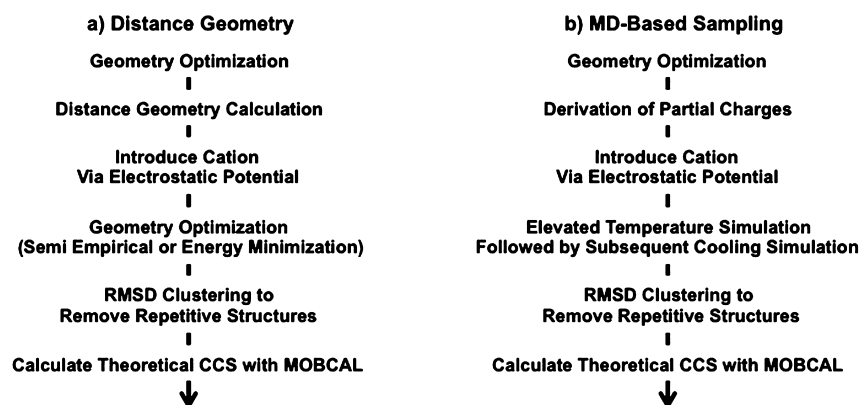[a]Refs 10−14. [b]Ref 15. [c]Refs 17 and 18. [d]Refs 19 and 20. [e]Refs 4 and 21.

provide a more detailed understanding of the molecular conformation(s) that are consistent with experiment. Table 1 lists methods commonly used for generating conformations in

Scheme 1. Schematic Workflow for Conformational Analyses Using the Distance Geometry Protocol and the MD-Based Protocol

a) Distance Geometry

Geometry Optimization
|
Distance Geometry Calculation
|
Introduce Cation
Via Electrostatic Potential
|
Geometry Optimization
(Semi Empirical or Energy Minimization)
|
RMSD Clustering to
Remove Repetitive Structures
|
Calculate Theoretical CCS with MOBCAL
↓

b) MD-Based Sampling

Geometry Optimization
|
Derivation of Partial Charges
|
Introduce Cation
Via Electrostatic Potential
|
Elevated Temperature Simulation
Followed by Subsequent Cooling Simulation
|
RMSD Clustering to
Remove Repetitive Structures
|
Calculate Theoretical CCS with MOBCAL
↓

support of IM-MS experimental data. For large systems such as protein complexes[10−14] and virus assemblies[15] coarse-grained methods are used to obtain the cross-sectional area. For smaller systems (peptides,[16−18] carbohydrates,[19,20] smaller molecules,[4,21] etc.), some form of molecular dynamics (MD) is the current method of choice.

Long MD simulations with an appropriate force field for the molecule(s) of interest are typically required to obtain a thorough and useful conformational analysis. While the IM-MS experimental data can be generated rapidly [on the order of milliseconds (IM) and microseconds (MS)], MD simulations for conformational analysis are quite time-consuming (days to weeks depending on the modeled structures). Furthermore, natural products are a very structurally diverse group of molecules, and this can make appropriate force field selection difficult. Although force fields exist that can describe many natural product molecules realistically, it is not clear that any current force field is appropriate for all members of this diverse class of molecules. If an inappropriate force field is used in the MD simulation, the resulting molecular conformations might be chemically unreasonable. One possible solution is to generate new potential function parameters for each molecule of interest, but this can be a time-consuming process if large numbers of molecules need to be studied. An alternative solution is to use computational techniques that do not rely on a force field for conformational sampling.

Distance geometry generates molecular conformations by sampling different possible interatomic distances between all pairs of atoms in the molecule.[22,23] Upper and lower distance limits, or bounds, are defined for each pair of atoms in the molecule, and then a distance within these bounds is selected randomly for each pair. Lower bounds are typically adjusted to avoid atomic overlaps (i.e., the lower bound may be set as the sum of the van der Waals radii of the atom pair), while upper bounds can be set at an arbitrarily large value to increase the number of possible conformations that are generated. Note that, as the upper bound is increased, a larger number of chemically unreasonable conformations will be generated. To limit the number of chemically unreasonable conformations generated, bounds for covalently bonded atom pairs are normally restricted to values quite close to the corresponding equilibrium bond lengths. Pair distances for atoms involved in bond angles at sp[2]- and sp[3]-hybridized atoms can also be restricted to produce chemically reasonable angle values, and additional restraints are routinely imposed to preserve stereochemistry at chiral centers. The set of selected atom-pair

distances is then converted to the corresponding set of Cartesian coordinates that define the unique molecular conformation. By selecting different random distances for each pair of atoms in subsequent iterations, a collection of unique molecular conformations is generated. Since IM-MS experiments are conducted with ionized molecules, a cation is then added to each generated conformer, based on the minimum of the calculated molecular electrostatic potential grid. The cation-associated conformations generated with this computational protocol typically require brief geometry optimization to relieve any residual, small geometrical distortions (e.g., slightly distorted bond lengths, bond angles, etc.) and to optimize the cation position. These geometry optimization calculations can be performed with either a quantum mechanical (QM) method or molecular mechanics energy minimization with an appropriate force field.

■ METHODS

The steps for the suggested distance geometry protocol as well as the MD-based method are shown in Scheme 1 with a more detailed discussion below.

Molecular Dynamics Method. An MD-based sampling protocol has been implemented where the system undergoes a single heating and cooling cycle during the calculation. One of the following two methods was used depending on which force field (GAFF[24] or MMFF94x[25]) better described the molecule of interest.

For an MD simulation performed with the GAFF force field, a geometry optimization at the Hartree−Fock level with a 6-31G* basis set was performed with Gaussian09[26] for each test molecule. Partial charges for each molecule were derived from an ab initio electrostatic potential calculation using a 6-31G* basis set and fitted using the RESP[27] program in AMBER.[28] XLEaP was then used to generate the molecule−sodium complex. Chirality constraints were applied in the form of improper torsion angles and a distance restraint was placed on the sodium ion to keep it near the molecule during the simulation. Next, 1000 steps of steepest descent/conjugate gradient energy minimization was performed with the sander module followed by a 10 ps MD simulation to heat the molecule to 1000 K. Then, a long MD simulation was run at 1000 K for 9000 ps where structural snapshots were saved every 3 ps during the simulation. Each snapshot was then cooled to 300 K during a 15 ps MD simulation followed by a short energy minimization. The MOBCAL implementation of

the projection approximation, exact hard sphere scattering, or calibrated trajectory methods (depending on the size of the molecule) were used to calculate the CCS of each sodium-coordinated complex.[29−31] Specifically, for molecules containing less than 100 atoms, the projection approximation was used. For molecules containing more than 100 atoms, the exact hard sphere scattering method was used with one exception. For erythromycin, which has several oxygen atoms and a concave region between the two sugar rings, a correlated trajectory method was used to achieve better alignment with the experimental data.

An MD-based sampling method was also performed in the Molecular Operating Environment (MOE)[32] program with the MMFF94x force field. The sodium-coordinated complexes generated for the AMBER MD simulations were used as starting structures. Chiral centers were fixed and a distance restraint was used to retain the sodium cation near the molecule during the simulation. Each sodium-coordinated complex was first energy minimized, and then heated to 800−1000 K over a 10 ps MD simulation. Next, a long MD simulation was run at that elevated temperature (800−1000 K) for 9000 ps where structural snapshots were saved every 3 ps during the simulation. These snapshots were then cooled to 300 K during a 15 ps MD simulation followed by a short energy minimization. As described above, MOBCAL was used to calculate the CCS of each sodium-coordinated complex.

Distance Geometry. All distance geometry calculations were performed with the DGEOM95 program.[33] Initial input structures for the DG calculations were obtained from structural databases or generated with a model building program (e.g., MOE), followed by geometry optimization in Gaussian09. DGEOM95 assigns connectivity and bond types based on input coordinates if connectivity information is not provided explicitly in the starting structure file. This information is used to assign distance restraints to actual bond lengths for covalently attached atoms and to detect non- and partially rotatable bonds within the molecule. Torsions, or 1−4 distance restraints, are utilized for atoms in certain structural relationships. Atoms attached to double bonds and atoms within aromatic rings are set coplanar, while amide and ester torsion angles are set to $0° \pm 15°$. For all other torsion angles, lower bounds are set to $+60°$ or $-60°$ (gauche orientations) for acyclic bonds and to eclipsed conformations for cyclic bonds or bonds adjacent to double bonds or aromatic rings by default. All other torsion angle upper bounds are set to $180°$ (trans conformations). Distance restraints for all other atom pairs are defined so that the lower distance bound is set to the sum of their van der Waals radii, while upper distance bounds are set to the length of the longest chain between the two atoms (i.e., the largest possible distance permitted by the series of bonds that connect the two atoms). In addition to these distance restraints, chiral centers are maintained by calculating the vector cross product of the tetrahedron enclosed by the four atoms attached to the chiral center and then ensuring the sign of the vector cross product remains the same in generated structures.

Once these distance restraints are generated, the triangle inequality theorem is used to verify that all distances are geometrically consistent. The triangle inequality theorem simply states that for any set of three atoms A, B, and C, the distance between atoms A−B cannot be longer than the sum of the distances between atoms A−C and B−C. Random pairwise distances that fall within the defined bounds are then selected

through a partial metrization method that ensures that the majority of the random pairwise distances satisfy the triangle inequality.[34] The set of distances are then converted to discrete Cartesian coordinates, generating a unique conformation. These steps of selecting random pairwise distances and then converting them to Cartesian coordinates are repeated until the desired number of conformations is generated. Finally, a clustering step is performed so that degenerate conformations (pairwise root mean squared difference (RMSD) < 1.0 Å) are removed. For the molecules in this test set, we generated between 2000−20 000 conformations to assess how thoroughly the distance geometry calculations sample conformational space.

Many empirically derived mobility measurements are determined with sodium coordination of the cationizing species, in particular for natural products that can often contain oxygen-rich carbohydrate moieties and the high oxyphilicity of alkali metals. The initial conformations from distance geometry require connectivity for all atoms, which does not allow for easy incorporation of a coordinated cation. We used the XLEaP module in AMBER to coordinate a sodium cation with each conformer generated by the distance geometry program, placing the ion at the minimum of the molecular electrostatic potential computed from the partial charges. Then we tested both a semiempirical QM technique and molecular mechanics energy minimization for the geometry optimization step. For the QM geometry optimization we used the PDDG/PM3 Hamiltonian[35] with Gaussian09.[26] A QM geometry optimization calculation is somewhat more CPU-intensive than molecular mechanics energy minimization, but may be the only practical option when suitable force fields are not available. We also used molecular mechanics energy minimization calculations, with either the MMFF94x force field[25] in MOE[32] or the GAFF force field[24] in AMBER.[28] For the GAFF force field calculations, we used atomic partial charge parameters developed for the MD simulations described above. Otherwise, we used default parameter values from each force field for all molecules.

Suppose[36] and OC[37] programs were used to cluster low-energy structures generated by each computational protocol into conformational families to achieve data reduction and facilitate structural analysis. The Suppose program superimposes all possible pairs of conformations and computes the RMSD in structure for each pair. The OC program then sorts the structures into clusters based on structural similarities as determined by a threshold RMSD cutoff value. For comparison purposes an RMSD cutoff value of 1.0 Å was chosen for both distance geometry and MD calculations. The OC program was used to select a molecular conformation that most closely represents the mean structure for each cluster. As described above, MOBCAL was then used to calculate the theoretical CCS for each cluster representative.

Experimental CCS Measurements. All 10 natural products were obtained from Sigma Chemical Company (St. Louis, MO). Matrix-assisted laser desorption ionization (MALDI) was performed for 200:1 molar ratios of the analytes with either 2,5-dihydroxybenzoic acid or α-cyano-4-hydroxycinnamic acid matrix. The MALDI-IM-TOFMS has a 13.9 cm IM drift cell that is maintained at a pressure of ca. 3 Torr helium and an orthogonal reflection TOFMS with a 1 m flight path maintained at a pressure of 5×10^{-8} Torr. The temperature of the drift tube was ~293 K, and the electrostatic-field strength ranged from ~90 to 120 V cm^{-1}.
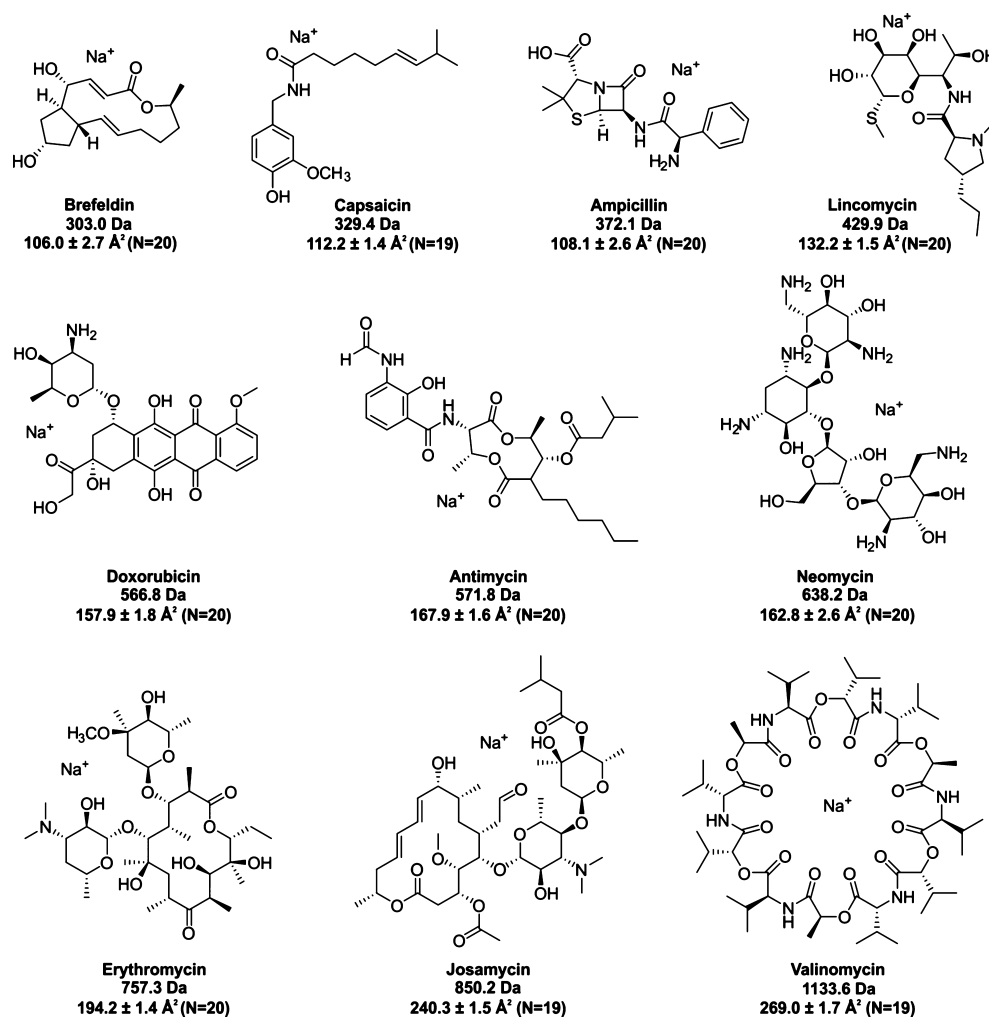
Figure 1. Two-dimensional structure representations of the 10 natural products tested in the present study. Each natural product is labeled with corresponding *m/z*, experimental CCS, and number of experimental CCS measurements.

Further experimental and instrumentation details have been presented previously in the literature.[4,38]

RESULTS AND DISCUSSION

Results. The 10 natural products studied here are shown in Figure 1. They range in size from 45 atoms ([M + Na]$^+$, 303.0 *m/z*) to 169 atoms ([M + Na]$^+$, 1133.6 *m/z*) and represent different subclasses of natural product molecules. Representatives from macrolides (brefeldin, erythromycin, josamycin), cyclic peptides (valinomycin), aromatic polyketides (doxorubicin), aminoglycoside antibiotics (neomycin), and other classes (capsaicin, lincomycin, antimycin, and ampicillin) were chosen to determine how distance geometry would perform across a wide range of natural product molecules. These different classes further emphasize the difficulty of trying to select a force field that would accurately describe such a diverse group of molecules.

The computational cost for both the MD-based method and the distance geometry calculations is presented for all 10 molecules in the histogram in Figure 2. The computational cost for the MD method is typically at least an order of magnitude greater than the distance geometry protocol. These time values reflect calculations that were run on a single processor except for calculations performed in Gaussian09, which were run on four processors. Note that the calculated times do not
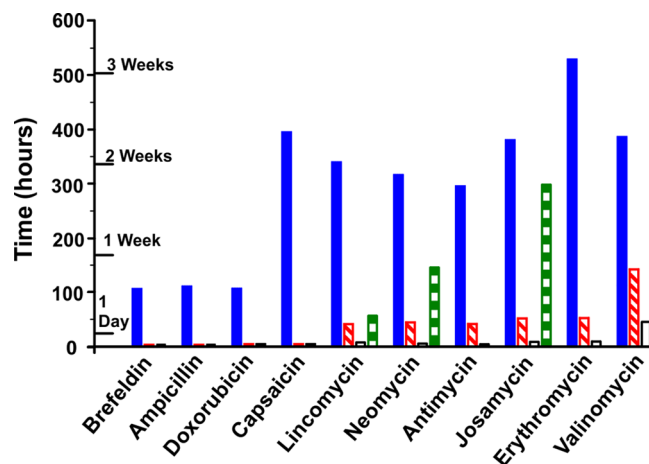

Figure 2. Histogram summarizing the computational cost of the distance geometry protocol compared to MD-based sampling. MD results are shown in solid blue, and distance geometry results are shown in dashed red and open black depending upon how many initial conformations are requested (red for 20 000 conformations and black for 8000 conformations). Results from the semiempirical geometry optimization are shown in boxed green.
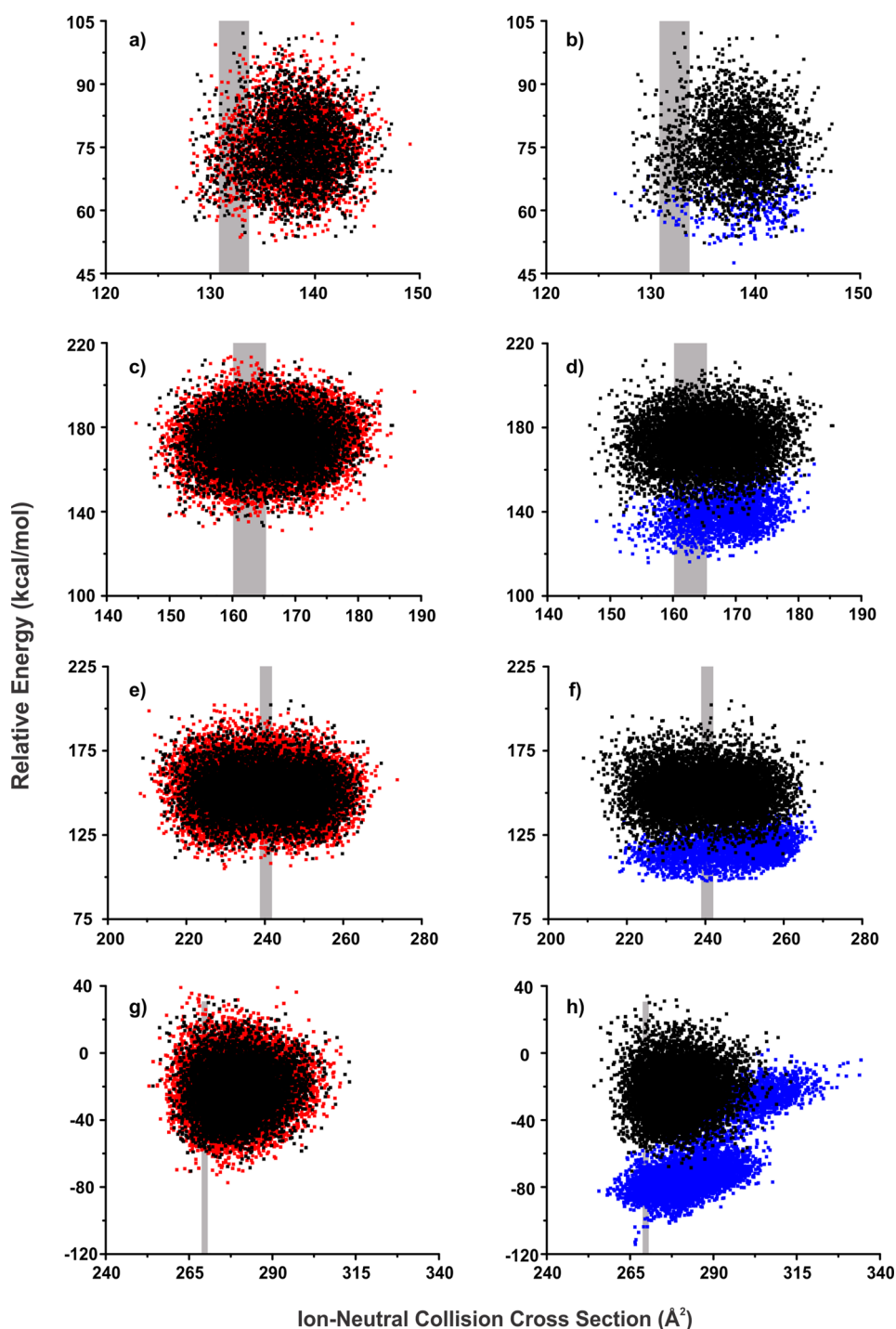
Figure 3. Scatter plots for four representative natural products to show where the generated conformations occur in relative energy vs theoretical CCS. The IM-MS measured CCS range (mean value and standard error) is indicated by the vertical gray bar. In panels a, c, e, and g, the comparison for 8000 (black) vs 20 000 (red) conformations generated by distance geometry is shown. In panels b, d, f, and h, the comparison for 8000 (black) conformers from distance geometry vs MD-based conformational sampling (blue) is displayed. Results are shown for lincomycin (a and b), neomycin (c and d), josamycin (e and f), and valinomycin (g and h).

incorporate the theoretical CCS calculation, only the time required to generate the conformations, to better compare the distance geometry and MD strategies. Although the computational cost for the QM geometry optimization with distance geometry is not much smaller than that for the MD-based sampling protocol, without an appropriate force field for the molecule, this may be the best option. Also since distance geometry calculations scale O^n and molecular dynamics

calculations scale $O(n)$, the computational advantage for the DG protocol decreases for extremely large, flexible molecules (e.g., large peptides or small proteins).

To better ascertain the robustness and reliability of the distance geometry computational protocol, we performed a detailed analysis of results as a function of total number of requested conformations. Conformational space plots are shown in Figure 3 for 4 of the 10 natural products [lincomycin
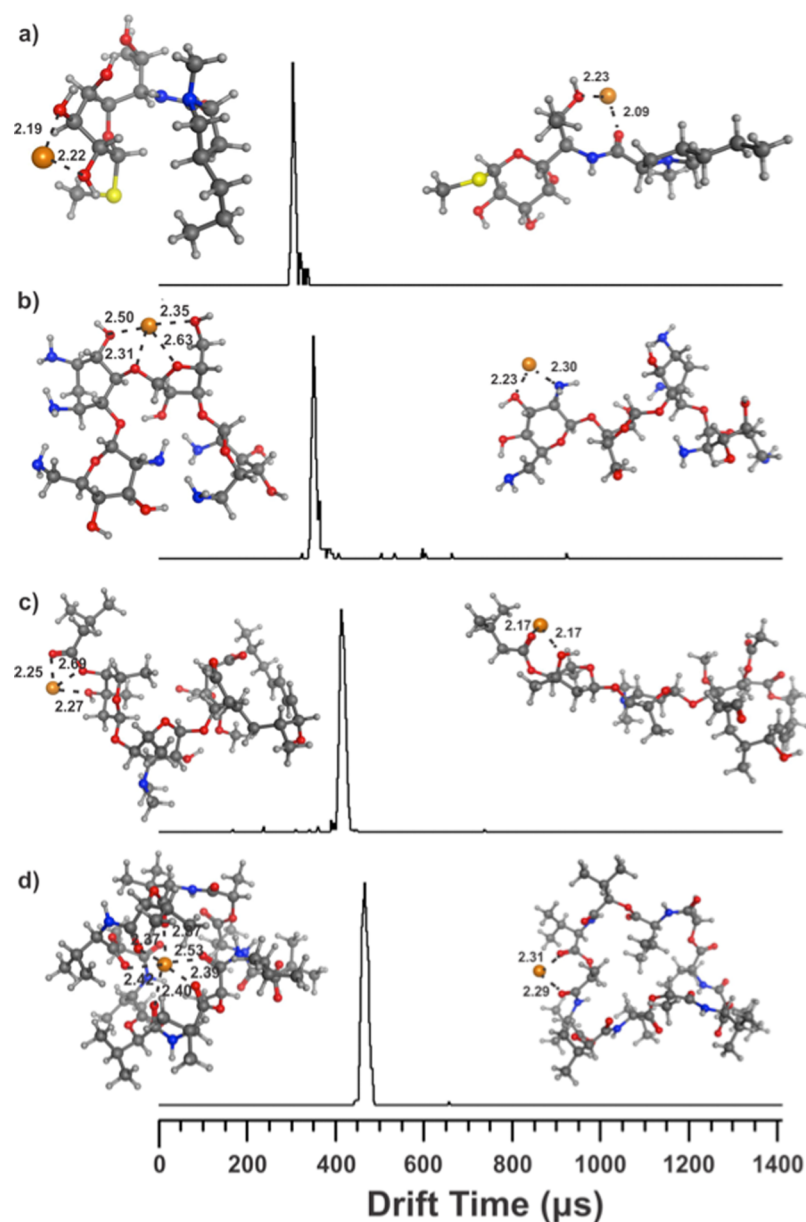
Figure 4. IM traces for the representative natural products shown in Figure 3, namely, (a) lincomycin, (b) neomycin, (c) josamycin, and (d) valinomycin. The most representative conformation generated with distance geometry from within the experimental range is shown for each natural product to the left of the mobility peak, and a conformation that does not agree with the experimental measurement is shown on the right of the mobility peak to illustrate the coordination of computation with experiment for interpretation of structure.

(Figure 3, parts a and b), neomycin (Figure 3, parts c and d), josamycin (Figure 3, parts e and f), and valinomycin (Figure 3, parts g and h)], with detailed results for the remaining six presented in the Supporting Information (Figures S-11 and S-12). In the left panel of Figure 3, conformational space plots are shown that compare the results when either 8000 or 20 000 conformations are generated by distance geometry. The theoretical CCS value is plotted against the computed energy for each conformation. The vertical gray bar indicates the experimental CCS range. When only 8000 structures are requested a similar conformational space is covered compared to the 20 000-conformation sample for all test molecules, but at a much reduced total computation time. While comparable results are obtained for some of the test molecules when only 2000 conformations are requested [data shown in the Supporting Information (Figure S-15)], this is not always the

case with some of the larger and/or more flexible molecules. Thus, we conclude that for molecules of this size and molecular complexity, 8000−10 000 generated conformations should sample conformational space adequately.

In the right panel of Figure 3, the distance geometry results and MD conformational sampling results are plotted for comparison. There is a noticeable energy difference between the two methods that is reflective of the approaches these calculations take when sampling conformational space. The MD-based conformational sampling method tends to generate predominantly low-energy conformations, since MD preferentially samples low-energy regions of the energy surface. By contrast, distance geometry is designed to explore all geometrically possible conformations (within the limits of the defined distance upper bounds), so it should always generate some slightly higher energy conformations as compared with

MD-based sampling. While MD-based conformational sampling does generate lower energy conformations (due to the extensive cooling step in the calculation), nevertheless similar conformations are generated with each method. Clustering data as well as representative structures that illustrate these similarities are presented in the Supporting Information (Figures S-16–S-44).

For lincomycin in Figure 3b, neomycin in Figure 3d, and josamycin in Figure 3f the distance geometry and MD methods perform comparably, generating a similar number of conformations within the experimental CCS range. For valinomycin in Figure 3h, the MD sampling method generates more conformations that agree with the accepted experimental gasphase conformation of valinomycin.[39] This is almost certainly because valinomycin complexes the sodium cation within the cyclic peptide structure, and its low-energy conformations are thus influenced strongly by the presence of the cation. In the absence of the sodium ion, a dramatically different conformational ensemble is sampled. By contrast, the presence or absence of sodium coordination has little or no influence on the overall ensemble of conformations obtained for the other test molecules.

To illustrate the interpretation of experimental IM conformation, IM traces for the representative natural products are shown in Figure 4 along with sample conformations generated with the distance geometry protocol for the four natural products discussed above [results for all other molecules are in the Supporting Information (Figure S-13)]. A conformation that agrees with the experimental data is shown near the mobility peak to the left, while a conformation that does not show agreement is also displayed for comparative purposes to the right. For conformations that fall within the experimental CCS range, sodium is typically coordinated with multiple atoms, generally leading to conformational contraction. By comparison, when the sodium cation is coordinated with only one or two atoms, the conformations tend to be more open and extended. While there is modest conformational contraction observed for lincomycin (Figure 4a), neomycin (Figure 4b), and josamycin (Figure 4c), the effect is far more dramatic for valinomycin (Figure 4d). The valinomycin conformation that agrees best with the experimental data has the sodium ion localized in the center of the cyclic peptide ring where it coordinates with multiple oxygen atoms. The valinomycin conformation that is inconsistent with the experimental CCS data has the sodium ion coordinated on the periphery to only two oxygen atoms, resulting in a more elongated peptide conformation. Additional representative conformations from RMSD clustering for all 10 of the natural products generated with both distance geometry and the MD sampling method can be found in the Supporting Information (Figures S-16–S-36).

Advantages and Challenges for both Distance Geometry and Molecular Dynamics Sampling. As summarized in Table 2, both distance geometry and MD sampling methods have potential advantages and challenges for the conformational sampling task. It is important to first consider how each method is sampling conformational space. An MD calculation samples conformational space by generating trajectories on an energy hypersurface. This surface is defined by a force field, which describes all covalent and noncovalent interactions for the molecule(s) of interest. A suitable force field is crucial for effective use of an MD-based method; a force field that is not properly parametrized for the molecules of

Table 2. Advantages and Challenges Associated with Distance Geometry and MD-Based Strategies

distance geometry	MD-based sampling
Advantages	
samples the entire conformational space	preferentially samples low(er) energy conformations
calculation does not depend on a force field	ion interacts with the molecule throughout the entire simulation
time efficient	
Challenges	
ions not included explicitly during DG calculations	simulation is based on force fields which are not parametrized for all molecules
	no easy way to determine when sufficient conformational sampling has been achieved
	time consuming

interest is not likely to yield relevant conformational information.

Distance geometry uses only interatomic distance data to generate a collection of conformations. The distance information is typically defined as "bounds", i.e., upper and lower distance limits, rather than a fixed value, and these bounds are defined by chemical properties such as equilibrium bond lengths and bond angle values, van der Waals radii, etc. Thus, distance geometry methods can be used for conformational sampling even when suitable force fields are not available for the molecules of interest. This situation is exacerbated when molecules contain many different and nonuniform chemical functional groups and connectivity as encountered with natural products.

Clearly, it is important to ensure that conformational space is sampled adequately. Molecular dynamics simulations preferentially sample low-energy regions of conformational space while distance geometry can generate all geometrically possible conformations allowed by the imposed bounds. Therefore, distance geometry can sample all conformational space if sufficiently loose distance bounds are set and if enough conformations are generated. In principle, MD-based methods can also sample all conformational space if a sufficiently long simulation is run. In practice, the computation time required to achieve any specified degree of "conformational space coverage" for molecules like those examined in this study will be much greater for MD-based sampling methods compared to distance geometry protocols, as our results clearly demonstrate.

Nevertheless, both of these methods can generate chemically unreasonable conformations. While distance geometry can create distorted, or strained, molecular conformations when loose bounds are used, a short energy minimization calculation can usually relax the distorted conformation. When distance bounds are defined more tightly, few, if any, distorted conformations are generated. MD-based conformational sampling strategies will typically generate relatively few strained, or high-energy, conformations. However, if elevated temperatures are used to increase conformational sampling efficiency by facilitating transitions from one low-energy region to another on the energy surface, significantly more distorted, high-energy conformations may be generated, and subsequent energy minimization performed at the end of the calculation may not always relax the distorted conformations. The unique structural motifs (e.g., heterocyclization, macrocyclization) typical of the natural products investigated in this study make them particularly susceptible to conformational distortion

during high-temperature MD simulations. Thus, while both methods can produce chemically unreasonable structures in the final solution data set, this issue impacts the MD sampling protocol efficiency more dramatically due to the greater computational cost for the MD calculations.

In order to generate structures that are relevant for interpretation of the experimental data, conformations that represent the cation–molecule complex explicitly must be generated. In an MD simulation, the ion can be present throughout the simulation to facilitate this task. In cases where the cation may exert a significant influence on molecular conformation, e.g., valinomycin, the MD sampling strategy may generate a larger number of chemically reasonable conformations that agree well with experimental CCS data. While it is possible to include the cation explicitly in the distance geometry calculation protocol, it is generally much easier to introduce the cation after the initial conformations have been generated. In cases where the cation has significant impact on the preferred conformations, it is quite possible that the distance geometry protocol will produce far fewer structures that agree well with experimental CCS data. This trade-off between computational cost or explicit inclusion of the cation during all stages of the conformational sampling procedure (e.g., MD-based methods) clearly favors distance geometry when the site of cationization is well-understood and can be fixed prior to calculation.

CONCLUSIONS

These results clearly show that the distance geometry plus geometry optimization protocol can be an effective and computationally efficient conformational sampling strategy for analysis of IM-MS data for natural products. In all but one case, the distance geometry protocol performed at least as well as the MD sampling method, but at a fraction of the computational expense. The MD method appears superior only for valinomycin, a cyclic peptide. There are two major factors that contribute to this observation for valinomycin. First, it is much larger than the other molecules in the test set, with many more degrees of freedom. As a result, many more conformations would have to be generated during a distance geometry calculation to get reasonable conformational sampling. More significantly, low-energy conformations for valinomycin are strongly influenced by the presence and position of the sodium cation. Therefore, explicit inclusion of the sodium ion during the conformational sampling process is important, but this cannot be done efficiently with our distance geometry protocol at present.

Computational analysis of the conformational space for natural products in support of structural IM-MS provides further insight into the structural motifs that cause gas-phase separation of these species from primary metabolites. Incorporating computational methods with further IM-MS studies will provide additional structural information, which could aid identification of these natural products, because of structural uniqueness compared to primary metabolites, from complex biological samples. Additionally, IM-MS is currently showing great promise as an analytical method in the fields of systems, synthetic, and chemical biology. This is due to its ability to separate and analyze complex samples containing a wide array of biological molecules such as peptides, carbohydrates, and metabolites. A computational method, such as distance geometry, that can efficiently sample the conformational space of all of these structurally different biomolecules could potentially facilitate structural interpretation of IM-MS signals on a time scale more commensurate with the experiment itself.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Protocol and performance details, RMSD clustering data as well as conformational space plots for the natural products not shown in the manuscript, sample conformations generated with the MD-based method and distance geometry for all 10 of the natural products. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Phone: 615-322-1195. E-mail: john.a.mclean@vanderbilt.edu.
*Phone: 615-343-1247. E-mail: terry.p.lybrand@vanderbilt.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) McDaniel, E. W.; Mason, E. A. *The Mobility and Diffusion of Ions in Gases*; Wiley: New York, 1973.

(2) Mason, E. A.; McDaniel, E. W. *Transport Properties of Ions in Gases*; Wiley: New York, 1988.

(3) McLean, J. A. The Mass-Mobility Correlation Redux: The Conformational Landscape of Anhydrous Biomolecules. *J. Am. Soc. Mass Spectrom.* 2009, 20, 1775−1781.

(4) Goodwin, C. R.; Fenn, L. S.; Derewacz, D. K.; Bachmann, B. O.; McLean, J. A. Structural Mass Spectrometry: Rapid Methods for Separation and Analysis of Peptide Natural Products. *J. Nat. Prod.* 2012, 75, 48−53.

(5) Derewacz, D. K.; Goodwin, C. R.; McNees, C. R.; McLean, J. A.; Bachmann, B. O. Antimicrobial Drug Resistance Affects Broad Changes in Metabolomic Phenotype in Addition to Secondary Metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 2336−2341.

(6) Clemmer, D. E.; Jarrold, M. F. Ion Mobility Measurements and Their Applications to Clusters and Biomolecules. *J. Mass Spectrom.* 1997, 32, 577−592.

(7) Wyttenbach, T.; Bowers, M. T. *Modern Mass Spectrometry, Topics in Current Chemistry*; Schalley, C. A., Ed.; Springer: Berlin Germany, 2003; Vol. 225, pp 207−232.

(8) Zakharova, N. L.; Crawford, C. L.; Hauck, B. C.; Quinton, J. K.; Seims, W. F.; Hill, H. H., Jr.; Clark, A. E. An Assessment of Computational Methods for Obtaining Structural Information of Moderately Flexible Biomolecules from Ion Mobility Spectrometry. *J. Am. Soc. Mass Spectrom.* 2012, 23, 792−805.

(9) Hall, Z.; Politis, A.; Bush, M. F.; Smith, L. J.; Robinson, C. V. Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. *J. Am. Chem. Soc.* 2012, 134, 3429−3438.

(10) Pukala, T. L.; Ruotolo, B. T.; Zhou, M.; Politis, A.; Stefanescu, R.; Leary, J. A.; Robinson, C. V. Subunit Architecture of Multiprotein Assemblies Determined Using Restraints from Gas-Phase Measurements. *Structure (Cambridge, MA, U. S.)* 2009, 17, 1235−1243.

(11) Politis, A.; Park, A. Y.; Hyung, S.-J.; Barsky, D.; Ruotolo, B. T.; Robinson, C. V. Integrating Ion Mobility Mass Spectrometry with Molecular Modelling to Determine the Architecture of Multiprotein Complexes. *PLoS One* **2010**, *5*, e12080.

(12) Wang, S. C.; Politis, A.; Di, B. N.; Bavro, V. N.; Tucker, S. J.; Booth, P. J.; Barrera, N. P.; Robinson, C. V. Ion Mobility Mass Spectrometry of Two Tetrameric Membrane Protein Complexes Reveals Compact Structures and Differences in Stability and Packing. *J. Am. Chem. Soc.* **2010**, *132*, 15468–15470.

(13) Bernstein, S. L.; Liu, D.; Wyttenbach, T.; Bowers, M. T.; Lee, J. C.; Gray, H. B.; Winkler, J. R. α-Synuclein: Stable Compact and Extended Monomeric Structures and pH Dependence of Dimer Formation. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1435–1443.

(14) Uetrecht, C.; Rose, R. J.; Van Duijn, E.; Lorenzen, K.; Heck, A. J. R. Ion Mobility Mass Spectrometry of Proteins and Protein Assemblies. *Chem. Soc. Rev.* **2010**, *39*, 1633–1655.

(15) Knapman, T. W.; Morton, V. L.; Stonehouse, N. J.; Stockley, P. G.; Ashcroft, A. E. Determining the Topology of Virus Assembly Intermediates Using Ion Mobility Spectrometry-Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3033–3042.

(16) Ruotolo, B. T.; Verbeck, G. F.; Thomson, L. M.; Gillig, K. J.; Russell, D. H. Observation of Conserved Solution-Phase Secondary Structure in Gas-Phase Tryptic Peptides. *J. Am. Chem. Soc.* **2002**, *124*, 4214–4215.

(17) Albrieux, F.; Calvo, F.; Chirot, F.; Vorobyev, A.; Tsybin, Y. O.; Lepere, V.; Antoine, R.; Lemoine, J.; Dugourd, P. Conformation of Polyalanine and Polyglycine Dications in the Gas Phase: Insight from Ion Mobility Spectrometry and Replica-Exchange Molecular Dynamics. *J. Phys. Chem. A* **2010**, *114*, 6888–6896.

(18) Dupuis, N. F.; Wu, C.; Shea, J.-E.; Bowers, M. T. Human Islet Amyloid Polypeptide Monomers Form Ordered β-hairpins: A Possible Direct Amyloidogenic Precursor. *J. Am. Chem. Soc.* **2009**, *131*, 18283–18292.

(19) Plasencia, M. D.; Isailovic, D.; Merenbloom, S. I.; Mechref, Y.; Clemmer, D. E. Resolving and Assigning N-linked Glycan Structural Isomers from Ovalbumin by IMS-MS. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1706–1715.

(20) Yamaguchi, Y.; Nishima, W.; Re, S.; Sugita, Y. Confident Identification of Isomeric N-glycan Structures by Combined Ion Mobility Mass Spectrometry and Hydrophilic Interaction Liquid Chromatography. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2877–2884; S2877/1–S2877/7.

(21) Poully, J. C.; Lecomte, F.; Nieuwjaer, N.; Manil, B.; Schermann, J. P.; Desfrancois, C.; Gregoire, G.; Ballivian, R.; Chirot, F.; Lemoine, J.; et al. Combining Ion Mobility Mass Spectrometry and Infrared Multiphoton Dissociation Spectroscopy to Probe the Structure of Gas-Phase Vancomycin-Ac2 LKDADA Non-Covalent Complex. *Int. J. Mass Spectrom.* **2010**, *297*, 28–35.

(22) Crippen, G. M. *Distance Geometry and Conformational Calculations, Chemometrics Research Studies Series 1*; Wiley: New York, 1981.

(23) Blaney, J. M.; Dixon, J. S. Distance Geometry in Molecular Modeling. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, 1994; Vol. 5, pp 299–335.

(24) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(25) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1998**, *17*, 490–519.

(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, revision A. 01; Gaussian, Inc.: Wallingford, CT, 2009.

(27) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(28) Case, D. A.; Darden, T. A.; Cheatham Iii, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M. *AMBER 11*; University of California: San Francisco, CA, 2010; p 142.

(29) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential. *J. Phys. Chem.* **1996**, *100*, 16082–16086.

(30) Shvartsburg, A. A.; Jarrold, M. F. An Exact Hard-Spheres Scattering Model for the Mobilities of Polyatomic Ions. *Chem. Phys. Lett.* **1996**, *261*, 86–91.

(31) Bush, M. F.; Campuzano, I. D. G.; Robinson, C. V. Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas & Calibration Strategies. *Anal. Chem.* **2012**, *84*, 7124–7130.

(32) *Molecular Operating Environment (MOE)*, 2012.10; Chemical Computing Group Inc.: Montreal, Canada, 2012.

(33) Blaney, J. M.; Crippen, G. M.; Dearing, A.; Dixon, S.; Spellmeyer, D. C. *DGEOM95*; Chiron Corporation: Emeryville, CA, 1984–1995.

(34) Kuszewski, J.; Nilges, M.; Brünger, A. T. Sampling and Efficiency of Metric Matrix Distance Geometry: A Novel Partial Metrization Algorithm. *J. Biomol. NMR* **1992**, *2*, 33–56.

(35) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. PDDG/PM3 and PDDG/MNDO: Improved Semiempirical Methods. *J. Comput. Chem.* **2002**, *23*, 1601–1622.

(36) Smith, J. A. *Suppose—Superposition Software*, 2006; Vanderbilt University: Nashville, TN, 2006.

(37) Barton, G. J. *OC—A Cluster Analysis Program*, 2002; University of Dundee: Scotland, U.K., 2002.

(38) Sundarapandian, S.; May, J. C.; McLean, J. A. Dual Source Ion Mobility-Mass Spectrometer for Direct Comparison of Electrospray Ionization and MALDI Collision Cross Section Measurements. *Anal. Chem.* **2010**, *82*, 3247–3254.

(39) Wyttenbach, T.; Batka, J. J., Jr.; Gidden, J.; Bowers, M. T. Host/Guest Conformations of Biological Systems: Valinomycin/Alkali Ions. *Int. J. Mass Spectrom.* **1999**, *193*, 143–152.