Published in final edited form as:

J Am Soc Mass Spectrom. 2022 February 02; 33(2): 369–381. doi:10.1021/jasms.1c00351.

Investigation of charge-state-dependent compaction of protein ions with native ion mobility-mass spectrometry and theory

Amber D. Rolland[a], **Lejla S. Biberic**[a], **James S. Prell**[a,b,*]

[a]Department of Chemistry and Biochemistry, 1253 University of Oregon, Eugene, OR, USA 97403-1253

[b]Materials Science Institute, 1252 University of Oregon, Eugene, OR, USA 97403-1252

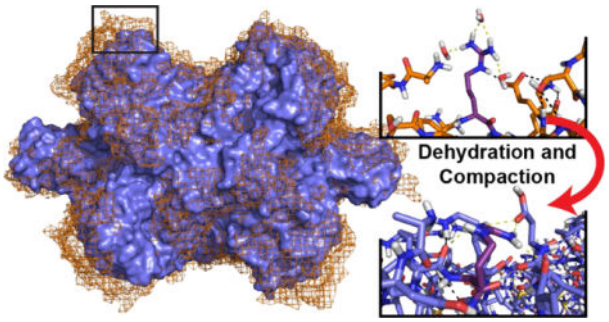
Abstract

The precise relationship between native gas-phase protein ion structure, charge, desolvation, and activation remains elusive. Much evidence supports the Charge Residue Model for native protein ions formed by electrospray ionization, but scaling laws derived from it relate only to overall ion size. Closer examination of drift tube CCSs across individual native protein ion charge state distributions (CSDs) reveals deviations from global trends. To investigate whether this is due to structure variation across CSDs or contributions of long-range charge-dipole interactions, we performed *in vacuo* force field molecular dynamics (MD) simulations of multiple charge conformers of three proteins representing a variety of physical and structural features: β-lactoglobulin, concanavalin A, and glutamate dehydrogenase. Results from these simulated ions indicate subtle structure variation across their native CSDs, although effects of these structural differences and long-range charge-dependent interactions on CCS are small. The structure and CCS of smaller proteins may be more sensitive to charge due to their low surface-to-volume ratios and reduced capacity to compact. Secondary and higher-order structure from condensed-phase structures is largely retained in these simulations, supporting the use of the term “native-like” to describe results from native ion mobility-mass spectrometry experiments, although, notably, the most compact structure can be the most different from the condensed-phase structure. Collapse of surface side chains to self-solvate through formation of new hydrogen bonds is a major feature of gas-phase compaction and likely occurs during the desolvation process. Results from these MD simulations provide new insight into the relationship of gas-phase protein ion structure, charge, and CCS.

Graphical Abstract

*Address correspondence to: jprell@uoregon.edu, Fax: +1-541-346-4643.

Supporting Information: Additional analysis details, figures depicting CCS and structural feature trends, and tables containing information for proteins studied, computed CCSs, and principal component analysis data



Introduction

Native-like protein ions formed by electrospray ionization often exhibit an approximately Gaussian distribution of charge states in the mass spectrum,[1] consistent with stochastic charging during the electrospray process.[2–4] Higher charge states beyond the native charge state distribution (CSD) typically indicate a significant degree of unfolding,[5–10] as measured with complementary ion mobility (IM) experiments. However, many researchers anticipate this relationship—that higher charge indicates partial unfolding—extends to the native charge state distribution, interpreting the lowest charge state as corresponding to the most "native-like" structure. In mass spectrometry (MS) experiments which require isolation of a single charge state, such as in collision-induced dissociation/unfolding and tandem MS, mass spectrometrists must decide which charge state(s) to select for further study. Though it is commonly accepted that protein ion structure likely varies to some degree across the CSD,[11] the magnitude of these structural differences and their influence on gas-phase behavior remain poorly understood. For example, in collision-induced unfolding (CIU) experiments,[12] CIU fingerprints, in which collision cross-section (CCS) of an ion is plotted against the collision voltage used to unfold it, have often been observed to vary among different charge states of the same protein (and transition voltages do not simply scale with charge state).[12–16] Whether this is due to differences in charge or structure or both remains to be determined. CCS measurements made in native ion mobility-mass spectrometry (IM-MS) experiments[17–19] provide information on the overall size and shape of ions and enable some insight into the charge-structure relationship.[8,11,20–24] However, investigation of protein structure using native IM-MS is complicated by the well-documented observation that protein ions typically compact to some extent in the gas phase, as compared to condensed-phase (e.g., X-ray crystal, NMR, or cryo-EM) structures.[25–28] As evidenced through comparison of experimental CCSs with those computed for condensed-phase structure coordinate files,[26] many proteins compact by more than 10% (including small proteins such as ubiquitin, an 8.6-kDa monomer) and some compact by as much as ~20–30% (such as transthyretin and bovine serum albumin, 56 and 66 kDa, respectively).[25] While the lowest charge state from a protein ion native charge state distribution does often represent the most compact structure, the term "native-like" can be somewhat context-dependent based on what aspects of native structure are under discussion.[8,20]

According to Dole's "Charge Residue Model" (CRM),[2,29,30] ion charge (z) should scale approximately with mass$^{1/2}$ for ions of fixed density and globular shape, and experimental

evidence supports this.[5,6,31,32] The general relationship between CCS and quasi-globular protein ion mass can also be derived from first principles based on geometry. For hard spheres of fixed density (and ignoring charge), CCS should increase as mass$^{2/3}$.[6,33–35] Taken together, it follows that CCS should scale roughly as $z^{4/3}$ for fixed-density hard spheres, ignoring charge entirely (i.e., charge exerts no forces and simply follows the CRM; see Supporting Information). Examination of well-established experimental drift tube CCS values, which do not require external calibration[33] and thus have relatively small calibration uncertainties compared to other methods,[28,36] for a set of 17 protein cations widely used as IM-MS calibrants[33,37] generally supports these relationships between CCS with charge state and mass (Fig. 1). As shown in these plots, the expected scaling of CCS with $z^{4/3}$ and mass$^{2/3}$ most closely applies to the most abundant native charge states and to the average CCS from individual protein ion native CSDs. (Including uncertainties of one standard deviation, the overall fits for both helium and nitrogen CCS data sets, respectively, very closely follow the expected power laws, with charge exponents of 1.23±0.07 and 1.2±0.1 and mass exponents of 0.71±0.02 and 0.66±0.03. Fits to exact 4/3- and 2/3-power laws for each of these, respectively, can be found in Fig. S1.) By contrast, CCS remains almost constant across each protein's CSD (Fig. 1), an observation reported previously by Bush and coworkers for cations, anions, and charge-reduced cations of the same protein species over a wide range of charge states.[21] Closer inspection of each individual protein native charge state distribution reveals small but measurable differences in CCS (see Fig. S2 for inset version of this plot to illustrate trends for small proteins). This observation itself indicates that protein ion structure does vary across native CSDs. However, the role of charge in affecting CCS and the details of the structural differences between ions of the same protein but different charge remain unclear.

Additionally, these differences are observed from a more coarse-grained view of global protein ion size and shape, and straightforward methods to probe finer structural details of native protein ions in the gas phase remain elusive. Molecular dynamics (MD) simulations offer a method by which to shed light on these unresolved questions regarding gas-phase protein ion structure.[25,38–57] We recently reported a simple *in vacuo* force field MD protocol[25] for simulating experimental gas-phase protein compaction to enable relatively fast calculation[58] of accurate, reliable theoretical CCS values for native protein ions (Table S1). This study represents the most wide-scale comparison of the performance of different force fields in the context of gas-phase protein ion structure to date, with one simulation performed for each protein (ranging in mass from 2.8 to 336 kDa) using each of 5 different force fields for a total of 85 simulations. Validated on a set of drift tube CCS values for 17 protein cations commonly used to calibrate IM-MS experiments,[33,37] the method accurately captures the extent of gas-phase compaction that protein ions undergo experimentally, producing structures for which the theoretical CCSs differ on average from experiment by only 0±4% (with nitrogen buffer gas) and −1±3% to −2±3% (with helium buffer gas for smaller and larger proteins, respectively).[25] The most important finding from this previous work is that all five force fields on average compacted structures both globally and at the surface of the proteins while largely retaining higher-order structure. These results are consistent with theoretical and experimental work in the literature indicating collapse of surface residue side chains in self-solvation and gas-phase protein

compaction.[38,39,42,46,50,59] Additionally, hydration has been shown to affect competition for hydrogen bond formation on many small ions, indicating the potential to alter these processes.[60–63] Experiments utilizing spectroscopy and, separately, unfolding/dissociation provide evidence of preservation of secondary and higher-order structure into the gas phase.[20,64–70] While theoretical results using force fields developed for use with water models and optimized on small molecules should not be over-interpreted,[71,72] structural features that are predicted robustly across many force fields and supported by other work can reasonably provide some insight into gas-phase protein ion structure and the role of charge.

Here, we utilized this established simulation protocol to examine in detail structural variation across native protein ion charge state distributions. We selected three proteins from the set of 17 native protein IM-MS calibrants as a case study: β-lactoglobulin (BLG; 18 kDa monomer), concanavalin A (ConA; 103 kDa tetramer), and glutamate dehydrogenase (GDH; 336 kDa hexamer).[33,37] These proteins represent a wide range of masses, charge states, CCS values (measured in both helium and nitrogen buffer gas), oligomeric states, and structural features (Fig. S3). For each, we identified up to five stable charge configurations for each native charge state using the charge placement algorithm in Collidoscope, performed *in vacuo* molecular dynamics simulations of each charge conformer, and computed the CCSs of the simulated structures using the Trajectory Method.[58] We then further analyzed structural features and changes in each simulated charge conformer, as well as in the original 17-protein data set,[25] in order to determine any structural trends with respect to charge state both globally and in the context of individual protein native CSDs. Results from this investigation of structural variation across charge states and the contributions of charge in affecting CCS provide useful insight for mass spectrometrists' interpretation of structural information from unfolding/dissociation experiments, as well as for the use and development of molecular dynamics simulations with which to investigate gas-phase protein behavior.

Experimental Section

Molecular Dynamics Simulations, Charge Placement, and Collision Cross-Section Calculations

Three proteins were selected from a set of 17 native-like protein ions with experimental drift tube ion mobility data:[33,37] β-lactoglobulin (BLG; 18 kDa monomer), concanavalin A (ConA; 103 kDa tetramer), and glutamate dehydrogenase (GDH; 336 kDa hexamer). All experimental IM-MS data used here was previously reported. Experimental conditions can be found in references 33 and 37; we note that activation conditions were minimized for all of these native IM-MS calibrant protein ions to ensure reproducibility. The following condensed-phase structure coordinate PDB files were used for simulation: 3BLG (β-lactoglobulin), 3CNA (concanavalin A), and 3JCZ (glutamate dehydrogenase). Residues missing from the full sequence were added to the GDH (3JCZ) structure in PyMOL, and this modified structure was briefly (1 ns) relaxed via molecular dynamics simulation with water to ensure more reasonable folding of these added segments. The resulting structure (with water molecules removed) was used as the starting structure for all subsequent work.

The charge placement algorithm in Collidoscope[58] was used to identify stable configurations for each charge state included in the experimental native protein ion IM-MS database.

This was repeated five times for each charge state of each protein. *In vacuo* molecular dynamics simulations were performed with the GROMACS 2016.6 molecular dynamics package as previously described.[25] Briefly, topology files were generated with charges assigned according to each unique charge configuration identified with Collidoscope.[58] All structure coordinate files were simulated with the GROMOS96 43a2 force field,[73] and the BLG charge conformers for which there was experimental data reported in helium buffer gas were simulated using the GROMOS96 54b7 force field[74] (due to GROMOS96 54b7 performing better for small proteins in helium; see Table S2). A brief energy minimization step was performed first, followed by a 5 ns NTV-ensemble production run at 300 K with a modified Berendsen thermostat. (Previous results indicate that longer production runs up to 500 ns do not typically result in additional measurable structural differences.[25])

Collision cross-sections were computed for all structures (unsimulated and simulated) using the Lennard-Jones 6–12-4 parameters for either helium or nitrogen buffer gas as implemented in the Trajectory Method in Collidoscope[58] with the original number of charges, unless otherwise noted (see below; CCS values listed in Table S3). (Throughout the text, we refer to the original condensed-phase starting structures as “unsimulated” and to the vacuum MD-compacted structures as “simulated”.) Experimental CCS values for comparison were obtained from the literature set of 17 native-like protein ion IM-MS calibrants (see Table S1 for complete list of protein identities, masses, and starting PDB IDs).[33,37]

CCS Calculations for Identical Structures Varying Only in Charge

The CCS of each smallest-CCS and largest-CCS structure resulting from the above-described simulations of BLG, ConA, and GDH charge conformers was re-calculated using the Trajectory Method in Collidoscope[58] using both the lowest and highest native charge state for each protein in both buffer gases. The resulting CCS values (reported in Table S4) were used to represent extremes of the CCS range expected if native charge state distributions corresponded to identical protein ion structures varying only in charge state.

Projection Approximation CCS Calculations

To represent the size of protein ions in the absence of charge-dipole interactions, Projection Approximation CCSs were calculated using IMoS for unsimulated and simulated structure coordinate files containing no explicit charges (Table S5).[75–77]

Analysis of Structural Features

Full details of how all structural features (listed in Table S6) were determined and analyzed, including principal component analyses, are available in the Supporting Information. All structural feature analysis presented in this work, both for the set of 3 proteins simulated for the first time here and the set of 17 native-like protein IM-MS calibrants with which the protocol was previously developed,[25] represents new results.

Results and Discussion

Experimental Trends in Collision Cross-Section with Charge for Globular Protein Ions

As shown for the set of 17 native-like protein ion IM-MS calibrants in Fig. 1,[33,37] while the $z^{4/3}$ power law derived from first principles generally holds empirically for the CCS of most-abundant charge states, the trend in CCS across individual protein ion native CSDs is nearly flat. This observation indicates that different native charge states of the same protein ion do not simply "expand" or "compact" to respect the CRM (which would indicate large changes in density), and it is known that charge state does affect long-range ion-dipole interactions that contribute to charge scattering and thus can potentially alter CCS.[10,78] The assumptions inherent to simplistic derivation of power-law relationships of CCS to charge and mass from CRM and geometry—fixed ion density, charge-independent CCS, and fixed structure across the charge state distribution—predict a perfectly flat trend in CCS across individual native CSDs (see Supporting Information). The measurable (albeit small, on the order of ~2%) differences in the CCSs across native CSDs (Fig. S2) indicate that at least one of these assumptions must be invalid. Either 1) density varies to some small extent across the native charge state distribution, 2) charge exerts sufficient forces to alter CCS, or 3) protein ion structure varies.

Computation of the volumes of simulated charge conformer structures (described in detail in a later section) for individual protein ion native charge state distributions reveals only minimal variation in density (Table S7).[79] Packing density, the ratio of the van der Waals volume to the total volume, varies less than 1% across all charge states for each of the proteins regardless of size (0.9% deviation for BLG, 18 kDa, and 0.3% deviation for GDH, 336 kDa). Examination of density determined as the ratio of protein mass over volume yields similar results. The plot of experimental drift tube CCS values[33,37] as a function of mass in Fig. 1 further shows very little change in density with mass. Thus we conclude that the first assumption—that density is essentially fixed across the native CSD—generally holds true, narrowing focus to investigation of the other two assumptions, which regard the role of charge and possible structural variation.

The "global" trend in CCS across all 17 proteins depicted in Fig. 1 has a slope of ~3.2–3.4 nm^2/charge (using local linear fits for the most-abundant native charge states) using CCSs measured in helium or nitrogen buffer gas, respectively.[33,37] As determined through linear regression, the experimental CCSs of smaller proteins across their individual CSDs tend to exhibit slightly more positive trends with respect to charge state, in contrast to larger proteins, which either exhibit a negative local slope or do not strongly correlate with charge state (Table S8). Regardless, among proteins with at least three native charge states, all local slopes were of magnitude equal to or less than 1 nm^2/charge, with the majority less than 0.5 nm^2/charge, considerably less drastic than compared to the "global" trend.[21]

Overall, the trends observed between experimental drift tube CCS and charge state suggest that ions of the same protein identity but different charge states do not adopt identical structures in the gas phase. However, despite the very slight changes in CCS with respect to charge state, we also cannot rule out the other remaining assumption—that charge does not affect CCS—from these observations alone. Both long-range and short-range interactions

can affect CCS. Long-range charge-dipole interactions should typically increase CCS with increasing charge state,[10] owing to increased long-range scattering, at least for identical ion structures.[78] Thus, as previously noted by Robinson, Bush, and others,[8,20,22,24,80,81] charge-dipole interactions should in principle have a relatively greater influence on the CCS of smaller proteins than larger proteins containing more residues and multiple subunits, as evidenced by their slight increases in CCS with increasing charge (Fig. S2). If structure does vary across native CSDs, several key questions remain: 1) How large are the effects of structural differences on CCS? 2) What structural features cause these effects? 3) Are these effects monotonic with charge? Teasing apart the reason for experimental CCS deviation from the expected relationship from first principles—whether it be due to variation of structure across native charge state distributions or due to significant contributions of charge to CCS—requires further investigation into the role of charge in gas-phase protein structure, for which we draw upon experiment and theory.

Global Trends in Protein Ion Structural Features with Charge State

To help assess the role of structure and charge state on CCS trends across native CSDs, we first investigated possible global trends by re-analyzing the original 17-protein data set in greater detail.[25] In our previous systematic comparison of the results of *in vacuo* simulation of each of the 17 proteins in the IM-MS database,[33,37] we examined structural changes predicted by each of the five force fields tested in order to identify both robust and force field-specific trends. This was based on the average change for each structural feature across all 17 proteins, for which only one charge state was investigated. Here, we re-analyze these simulated structures to identify possible trends with charge. Because this original data set encompasses a wide range of charge states from 17 different proteins, trends derived from the full data set are less likely to be heavily skewed by unique features of a single protein.

Features analyzed (see Table S6 for full list and analysis details) included those related to size (CCS, compaction, percent difference with respect to experiment, and RMSD as compared to the initial condensed-phase structure used for the simulations) and measures of how compaction varies for solvent-accessible residues at the “surface” of the protein versus the remaining set of buried residues in the “interior” (interior CCS, number of surface residues, number of polar contacts involving charged side chains at the surface, and number of hydrogen bonds involving surface residues). We also analyzed non-covalent interactions (number of hydrogen bonds and salt bridges) and secondary structure content (number of residues with α-helical or β-strand geometry as determined in PyMOL). Among the five FFs tested in our original study, most resulted in the same general changes to these structural features with differences only in magnitude or the extent of the change.[25] For example, all five FFs caused structures to compact both globally and at the surface, as determined by comparison of the CCS, number of “surface” residues, and number of polar contacts involving charged side chains before and after simulation (Fig. 2). Non-covalent interactions increased on average relative to initial structures for ions simulated with each of the five FFs, while secondary structure content decreased very slightly on average, with more loss of α-helical structure than β-strand.

Linear regression analysis performed here of each of these metrics against charge state revealed significant (Pearson $R^2 > 0.9$) correlations with CCS, number of hydrogen bonds, number of surface hydrogen bonds, number of surface residues, interior CCS, and number of polar contacts involving charged residues, and all of these exhibited positive slopes (Fig. 3 and Fig. S4–S5). CCSs and surface hydrogen bonds (which should scale roughly with surface area) for these simulated structures are fit to a $z^{4/3}$ power law as expected for global trends. The total number of hydrogen bonds should correlate to volume, which scales with z^2, and is fit accordingly. While the total number of salt bridges, RMSD, and secondary structure content also increased with increasing charge state, these trends did not have significant R^2 values (Fig. S4–S5). The global linear trend in CCS as a function of charge state identified here for simulated structures (with derivative values of ~3.1 nm^2/charge and ~3.4 nm^2/charge for helium and nitrogen buffer gas, respectively) matched well to that observed for the experimental drift tube CCS measurements described above (~3.2–3.4 nm^2/charge).

Further analysis of the above types of structural features as normalized metrics (with structural features represented as either a percent change relative to the unsimulated structure or a proportion relative to the maximum possible number of interactions[38] or number of features per residue) yielded trends with no additional significant correlation coefficients (Fig. S4–S5). However, the slopes for many of these trends in structural changes relative to their respective unsimulated starting structures as a function of charge state were near zero, indicating structural features or changes that are effectively constant (slopes of approximately ±0.1% relative change/charge) across this wide range of protein charge states (Fig. S4–S5). Several of these were related to non-covalent interactions: percent of maximum number of hydrogen bonds, percent change in number of hydrogen bonds, number of hydrogen bonds per surface residue, number of polar contacts per charged residue, and percent of maximum number of salt bridges (Fig. S4). Percent change in secondary structure also exhibited a flat trend with respect to charge state for this 17-protein data set. That the relative changes in each of these features is constant across this wide range of charge states, each representing a different protein ranging in mass from 2.8 kDa to 336 kDa, suggests that the GROMOS96 43a2 force field[73] robustly predicts these typical changes regardless of the particular protein, a result that is useful in interpreting structural information from simulation results. Because this force field also outperforms others tested for recapitulating experimentally observed gas-phase compaction of native protein ions,[25] the parameters associated with these types of interactions represent a good starting point for future optimization of existing or new force fields or more sophisticated computations for gas-phase purposes in the future. As collapse of side chains at the surface of proteins is the dominant feature of gas-phase compaction, it may be possible to capture more detailed aspects of compaction in the future by focusing more computationally-expensive methods, such as quantum mechanical simulations, on the protein ion surface.

Other trends for normalized (i.e., percent change or percent of maximum) metrics with respect to charge state which had neither a significant (> 0.9) R^2 value or effectively zero slope are also included in Fig. S4–S5. Briefly, we note that the percent change in each of the following features exhibited positive, although not strong, trends: surface hydrogen bonds, polar contacts involving charged side chains, and salt bridges. By comparison, the percent

change in surface residues and “interior CCS” decreased (i.e., became more negative) with increasing charge, indicating “smoother” surfaces for larger gas-phase protein ions due to large decreases in solvent-accessibility of surface residues. Overall, findings from these linear regression analyses are consistent with the general conclusion that larger protein ions form more new non-covalent interactions during *in vacuo* MD simulation than do smaller protein ions and thus are able to compact to a greater extent.

Trends in Structural Features Across Protein Native Charge State Distributions

Having established global trends in structure with respect to charge, we analyzed BLG, ConA, and GDH in greater detail, separately simulating compaction for each charge state in their native charge state distributions. These ions were chosen because they span a wide range of masses, charge states, and oligomeric states (BLG monomer: 18 kDa, 7–9+; ConA tetramer: 103 kDa, 19–23+; GDH hexamer: 336 kDa, 37–43+), and there is no significant homology between them (Fig. S3). BLG consists of a small β-barrel with short segments of helical structure. The condensed-phase structure of ConA contains no α-helical secondary structure and is primarily composed of β-sheets with a gap at the center of its four-subunit interface. GDH features a much larger subunit interface involving all C-termini at its center and is mostly α-helical, though each subunit also contains multiple β-sheets. Drift tube CCSs for these three protein ions are also available for both helium and nitrogen buffer gas,[33,37] and these experimental values range from ~17 nm^2 to ~135 nm^2. (Monomeric BLG is the smallest protein for which CCS measurements in both gases are reported, and GDH is the largest protein of the 17 in the IM-MS calibrant data set, which excludes GroEL.[33,37])

Analysis of individual protein CSDs removes effects of mass differences and allows for closer examination of the specific role of charge in influencing features of gas-phase structure. We identified up to five high-stability charge configurations for each experimental charge state of all three proteins using the Collidoscope charge placement algorithm.[58] We then performed an *in vacuo* MD simulation on the condensed-phase structures of these three proteins for each stable charge configuration identified. As before, CCSs were computed using the Trajectory Method in Collidoscope (Table S3),[58] and structural features of both the unsimulated and simulated structures were analyzed using PyMOL.

For the structural analysis above and in our previous simulation study, we utilized only the most abundant charge state from each protein’s distribution for simulation and CCS calculation. To ensure that the accuracy of the protocol is not limited to this charge state, we first checked its accuracy and precision for all of the native charge states for the three test proteins with available consensus CCS. As shown in Fig. 4, all 24 computed average CCSs (and all 49 total computed CCSs for individual structures) fell well within the expected range of performance with only one exception—the single CCS value for BLG 9+ in helium. We thus concluded that the simulation protocol is accurate across the native charge state distributions for these ions.

We also note that these simulation results, especially for the larger proteins ConA and GDH, do not exhibit a monotonically increasing relationship between CCS and charge state (Fig. S6), indicating that the lowest charge state from a native protein ion distribution does not necessarily correspond to the most compact structure. CCSs for individual simulated charge

conformer structures for each protein were very similar to each other, with less than 1% standard deviation for ConA and GDH in both gases and for BLG in helium (Table S3). This value was slightly higher for BLG structures in nitrogen buffer gas (~3%), though this is likely due to only one stable configuration being identified for each of the two charge states.

Linear regression analysis of structural features across individual protein ion CSDs for the three proteins investigated here, however, did not yield any significantly different findings from the global study described above. In general, the only linear trends with apparently strong correlation coefficients were for BLG (Fig. 5 and Fig. S7–S8), though these are not to be over-interpreted as they originate from fitting a set of just three data points due to identification of only one stable charge configuration for each charge state. While most trends for the two larger proteins ConA and GDH were effectively flat with respect to charge state (Fig. S9–S11), structural changes related to surface hydrogen bonds, surface residues, and salt bridges exhibited the clearest trends with charge state for BLG (Fig. 5). For all of these trends, increasing charge state correlated with fewer interactions or smaller percent increases in the number of these interactions. That is, the BLG 9+ structure formed fewer hydrogen bonds involving surface residues and salt bridges than did the 7+ structure, indicating competition between Coulomb repulsion of the charge sites and self-solvation of charges that is generally a major feature of gas-phase compaction.[38,39,42,46,59] Interestingly, this trend for BLG contrasts with the *global* decrease in surface residues retained relative to the unsimulated condensed-phase structure (i.e., increased surface “smoothing” as exemplified in Fig. 2) described above as a function of protein mass.

Structural changes that were consistent across the entire native CSD for both ConA and GDH include those related to hydrogen bonding and non-covalent interactions (percent of maximum number of hydrogen bonds, percent change in total number of hydrogen bonds, number of hydrogen bonds per surface residue, number of polar contacts per charged residue, and percent of maximum number of salt bridges), as well as RMSD (relative to condensed-phase structure coordinate file) and percent change in secondary structure (Fig. S9–S11).

Expected Collision Cross-Section Trends for Identical Structures Varying Only in Charge

To further investigate the role of both long-range and short-range interactions in determining CCS, we decoupled the influence of charge-dipole and gas-phase compaction by two different methods using the simulated protomer structures of BLG, ConA, and GDH. With the first of these methods, we tested the possibility that the ions represented by a charge state distribution have identical structure (other than that of the protonated residues) and varied only in charge. As described above, charge-dipole interactions between the ions and the buffer gas should result in larger CCSs for higher charge states.[10,78] To more fully capture the variation in CCSs expected, we identified the simulated charge conformer structure with the smallest computed CCS and that with the largest computed CCS for each protein in each buffer gas and re-computed each corresponding structure’s CCS with the lowest and highest charge states from the distribution (Table S4). As shown in Fig. 6, the variation in CCS over the native CSD assuming identical structure is minimal. In helium buffer gas, the largest variation by far in CCS expected between the lowest and highest charge state

is 0.28 nm^2 for GDH, with all other ranges ~0.1 nm^2 or less. These absolute differences correspond to differences in CCS below 0.5%. Similarly, in nitrogen buffer gas, the largest variation (again for identical structures of GDH) is ~1 nm^2, with all percent differences in CCS ~1.6% or less. (Note that the scale used for CCS values on the y-axis differs in each plot and that the difference after re-calculation of ConA CCSs in nitrogen buffer gas is much smaller than for GDH.) Thus, for these ions with essentially fixed structure across the native CSD, the difference in CCS arising from charge effects is below current experimental uncertainty and effectively unmeasurable. These results suggest that contributions of long-range charge-dipole effects alone cannot account for the variation in CCS observed across individual protein ion native CSDs, consistent with previous findings.[24] The inability of a single structure to explain the full range of experimental CCSs is also consistent with the hypothesis that protein ion CSDs arise from a distribution of multiple solution-phase folded states.[82]

To further confirm that charges do not exert significant forces to alter CCS, we performed a separate set of calculations to remove contributions of charge-dipole interactions from the computation of CCSs of the simulated structures. Charge is accounted for in Trajectory Method (TM) calculations through explicit inclusion of both long- and short-range scattering effects.[58,78,83] We thus performed an additional set of CCS computations on the simulated structures using the Projection Approximation (PA) method, which does not model charge-dependent interactions, in IMoS (Table S5).[75–77] The resulting (charge-ignorant) PA CCSs qualitatively mimic the trends observed with the TM, as shown in Fig. 7 where both are plotted as percent differences in CCS relative to the condensed-phase, unsimulated structure. Although there are systematic shifts in the PA computed CCSs relative to those from the TM due to the well-understood differences in accuracy of the two methods,[83] the PA results confirm that variation in CCS for the simulated structures reported here is not merely a result of accounting for long-range charge-dipole interactions in the CCS computation itself.

Of note, the greatest percent difference in CCS relative to the condensed-phase structure for BLG is observed for the lowest charge state (Fig. 7 and Table S3), with the smallest percent difference in CCS seen for the highest charge state. Further, as discussed above, BLG CCSs follow a positive trend with increasing charge state. For BLG and other small proteins for which the extent of possible compaction is limited and for which charge density and the proportion of surface residues is greater, experiment and these simulation results together indicate increasing charges can often lead to Coulombic repulsion that preferentially stabilizes larger structures. This finding (specific to the one small protein studied here) is consistent with a common observation in the literature whereby the lowest charge states are the most compact (i.e., have the smallest CCS). However, it is important to note that this structure is the most compacted and different from the condensed-phase structure, and labeling such structures as the most “native-like”, based solely on their CCS, can be somewhat misleading. While experiment reveals that the 7+ structure has a smaller CCS than does the 9+ structure, these results illustrate unique insight from computational chemistry that the higher-charged ion is indeed “larger” than the lower-charged ion.

Principal Component Analyses of Structural Features

Because linear regression analysis of the individual structural features as a function of charge did not yield any new significant trends, we sought to determine whether some combination of these features might be more meaningful in explaining the observed trends in CCS than each feature alone. Other IM-MS researchers have used sophisticated machine learning and PA CCS computations on condensed-phase structures to analyze ion mobility results for very large data sets.[84,85] Here, we use principal component analysis (PCA) and TM computations for MD-compacted structures, owing to the relatively small data set, for which more advanced machine learning approaches might be inappropriate or extremely time-consuming. We selected one metric from each of the 7 different categories of features (hydrogen bonds, surface hydrogen bonds, surface residues, polar contacts involving charged side chains, salt bridges, α-helical and β-strand secondary structure content), expressed as a percent change relative to the original unsimulated structure so that all features would be on the same protein-normalized scale. PCA of the charge state averages of these 7 features revealed that almost all of the variation could be explained by the first two principal components alone (66.0% and 25.0%, respectively, for a total of 91.0%). Replotting these data along the first two principal components shows clustering determined by protein identity (Fig. 8).

However, while all three proteins are well separated along the first principal component, ConA and GDH fall onto the same region along the second principal component axis (Fig. 8). The eigenvectors of the first principal component are evenly weighted with a coefficient ±0.37–0.44 for all features except for percent change in the number of surface residues (0.17, Table S9). This latter feature then dominates the second principal component (coefficient of 0.69), for which ConA and GDH have similar values and are clearly separated from BLG. These two larger proteins show similar behavior for this feature, with an average decrease in the number of surface residues of ~50–56%, but differ drastically by all other metrics studied. By comparison, the average decrease in the number of surface residues for BLG is considerably lower at ~30–37%. These differences lend further support to our general conclusion that smaller proteins' tendency toward increasing CCS with increasing charge state—indicative of long-range charge-dipole interactions predominating—can be rationalized by their minimal ability to undergo significant surface compaction, which is a robust, major contributor to global gas-phase compaction for larger protein ions.[25,38,46,59] We also note that of these three proteins, ConA is unique in that its condensed-phase, unsimulated structure contains no α-helical residues (Fig. S3). The differences in the first two principal components suggest that, in these simulations, percent change in the number of surface residues and α-helical secondary structure content are not strongly linked.

We also performed PCA on our original 17-protein data set with these same 7 features, both with the aggregated data and with each individual force field (see Supporting Information and Tables S10–S12). Importantly, the two major principal components in the aggregated data set, which together account for 56.9% of the variation, do not distinctly separate these data according to force field (Fig. S12). These axes are dominated primarily by features related to hydrogen bonding, surface residues, and polar contacts with charged side chains (Table S10), similar to PCA results for charge conformers of the 3 proteins described above.

PCA of these data for individual force fields reveal a high degree of similarity, with the two GROMOS96 force fields[73,74] and, separately, CHARMM27[86,87] and OPLS-AA/L[88] being most similar to one another and AMBER94[89] sharing commonalities with each of these pairs (Tables S11–S12). These results are consistent with the findings in our original study and further suggest that a major explanation for differences in structural features and performance for gas-phase investigations among the force fields is due to the greater extent to which GROMOS-simulated structures form new hydrogen bonds and surface interactions on the timescale used in this protocol.[25] We note that compaction at the surface is to be expected given the relatively short (5 ns) production runs for these simulations, although the interior regions of most protein structures simulated with GROMOS FFs in the previous study also decreased in size (CCS) (Fig. S5). Extending the production run timescale to 50 ns and 500 ns did not result in significant additional structural changes or performance with respect to comparison to literature CCS values.[25] We attribute this to the accelerated dynamics possible with united-atom FFs such as the GROMOS FFs used, which additionally offers advantages in reducing computational expense while enabling a high degree of accuracy and precision in comparison to experimental IM-MS data.

Conclusions

Small but measurable differences in experimental CCS across individual native protein ion CSDs[33,37] provide an initial indicator that the structures of gas-phase ion charge conformers are not identical and that charge may contribute to CCS. Deviations from the expected trend cannot be explained by varying density of ions across their native CSDs, as supported both by experimental trends in CCS with charge state and by MD simulations indicating essentially fixed density across CSDs (Table S7). The greater tendency of smaller proteins to exhibit positive slopes in CCS across their native CSDs suggests their structures and CCSs are more sensitive to charge than larger proteins, as evidenced by results described throughout. Many small monomeric proteins, such as BLG (studied here), have minimal interior regions and lack cavities and grooves, resulting in a reduced ability to compact in the gas phase. Experimental and theoretical work reported here, as well as many examples of previous work in the literature,[25,38,39,42,46,59] indicate self-solvation is a major feature of gas-phase protein ion structure and compaction (Fig. 2).

CCSs for simulated structures of BLG, ConA, and GDH were very similar to each other, with less than 1% standard deviation for all values of each protein, except for the two structures of BLG for nitrogen buffer gas (3% standard deviation; Table S3). Our results thus suggest that structural differences can impart a small but measurable effect on CCS, though this effect is not necessarily monotonic with charge (Fig. S6). Differences in compaction across charge state are supported by computation of PA CCS values (Fig. 7), and long-range charge-dipole interactions do alter CCS, but to a minimal extent that is effectively unmeasurable with most current native IM instrumentation and cannot explain experimental CCS variation (Fig. 6).[24] This result, as well as those described below, supports a model whereby compaction occurs after or simultaneously with charging of the ion, and results in structure varying across native CSDs. Compaction is affected by charge, with charged residues (and residues at the surface generally) collapsing to self-solvate by forming new hydrogen bonds and other non-covalent interactions, ultimately resulting in subtle structural

differences between charge states. By contrast, the opposite scenario (in which compaction occurs first, followed by charging) should result in essentially the same compacted structure initially, especially as protons are likely somewhat mobile during the final stages of ion desolvation and migrate to relatively stable charge sites.[11,51,61] The overall effect of compaction occurring prior to charging would be a slight monotonic increase in CCS with charge state, which is unsupported by experiment and simulations. Further, it is reasonable to expect that the energy released by formation of many (tens to hundreds of) new hydrogen bonds would exceed the small concomitant increase in Coulombic energy upon compaction. Together these findings support previous work showing that CCS need not correspond one-to-one with native protein ion structure and that peaks in IM spectra often represent multiple different conformations of protein ions with indistinguishable CCS.[80,90] The MD protocol utilized here,[25] with which we were able to identify several competing protomers for each charge state that had different structures, may also be useful when integrated into ongoing efforts to model experimental arrival time distributions, especially when combined with more extensive solution MD modeling prior to vacuum MD.[91] Because CCSs for cationic, anionic, and charge-reduced cationic protein ions of the same species have been found to vary to a small extent,[21] we also anticipate investigation of how structure may vary across native CSDs obtained with negative polarity and in comparison with their positive polarity counterparts, which are often higher in charge than for negative polarity, as a future application of this protocol.

Overall, results presented from experiment and theory provide evidence that gas-phase protein ion structure varies, albeit to a small extent, across native charge state distributions. Trends (or lack thereof) in structural features across CSDs identified here (Fig. 3, 5) can better inform researchers performing experiments which require isolation of a single charge state and rationalize differences in gas-phase behavior observed in CIU. Simulated protein charge conformers reported in the present work, as well as simulated structures in our previous work, all retain much higher-order structure and similarity to their corresponding original condensed-phase structure coordinate files. This result warrants optimism for the use of native IM-MS in studying protein structure, and other evidence also supports the finding that (globular) protein ions preserve much of their solution-phase structures in the gas phase.[38,40,43,64,92] However, we would caution the native IM-MS community against interpreting the lowest charge state from a native distribution or the native ion with the smallest CCS to be the most “native-like” simply because it is the most compact. Results from literature drift tube CCS values[33,37] and from simulations of protomers here illustrate that CCS does not vary with charge state “trivially” due to long-range ion-dipole interactions. As shown for the small protein BLG (Fig. 7), the most compact structure can also be the most different from the “native” condensed-phase structure, as shown by experimental and theoretical results together.

The primary explanation for gas-phase compaction identified here—extensive formation of new hydrogen bonds and non-covalent interactions—is not a novel finding,[38–41,50,59,64] but results presented here show that “smoothing out” of the ion surface and collapse of cavities/grooves upon compaction (as exemplified in Fig. 2) can subtly depend on charge state. This appears to be a general phenomenon applying to biomolecules beyond just proteins, as similar findings of gas-phase compaction and self-solvation by hydrogen bond formation

have been reported for nucleic acids as well.[93] The general ability of our MD simulation protocol,[25] in which charges are added to condensed-phase structures before a gas-phase production run, to recapitulate both general and detailed trends with charge state suggests that charging occurs before or along with final compaction of the protein ion, as described above. Detailed structure changes upon ion desolvation are then largely determined by charge state and compaction—through self-solvation and formation of new interactions for surface residues with other parts of the protein to replace those in many cases originally fulfilled by solvent interactions.[38,39,42,46,50,59] Results from cryogenic IM-MS studies of different hydrated peptides and small ions by Russell and coworkers include examples where solvation with increasing numbers of water molecules can lead to different effects on CCS and structural changes.[63] Separately, results from Williams and Russell provide evidence that even a small extent of hydration can screen electrostatic interactions and their effects on structure.[60,61,94] In the context of this work, we posit that formation of new hydrogen bonds after charging and concomitant with compaction and surface "smoothing"—as evidenced by our simulations—is a major part of the desolvation process. Dissociating a water molecule requires a considerable amount of thermal energy (often ~40 kJ/mol), which could be offset by the formation of new hydrogen bonds. While the smallest protein studied here formed only ~14 new hydrogen bonds on average through simulation, this number was much higher for both ConA and GDH (~270 and ~200 on average, respectively), and the energy gained would be sufficient to dislodge more than many water molecules for these latter cases. Results from this work not only provide new insight into the relationship of gas-phase protein ion structure and charge that will be useful for interpretation of structural information from native IM-MS experiments, but also highlight the important role for formation of new hydrogen bonds between protein functional groups in aiding desolvation and determining final protein ion structure and CCS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements & Funding:

A.D.R. is supported by the National Institute of General Medical Sciences (Award 2T32GM007759), and she is an ARCS Scholar supported by the ARCS Oregon Chapter. A.D.R. is also a recipient of the University of Oregon Doctoral Dissertation Fellowship. A.D.R. and L.S.B. are supported by the Peter O'Day Fellowship in Biological Sciences and Office of the Vice President for Research and Innovation at the University of Oregon. Further support is provided by the National Institute for Allergy and Infectious Diseases (Award R21AI125804 to J.S.P.). The authors also wish to thank Dr. Michael J. Harms (University of Oregon) for helpful discussions. The authors declare no competing financial interest.

References

(1). Mann M; Meng CK; Fenn JB Interpreting Mass-Spectra of Multiply Charged Ions. Anal. Chem 1989, 61, 1702–1708.

(2). Felitsyn N; Peschke M; Kebarle P Origin and Number of Charges Observed on Multiply-Protonated Native Proteins Produced by ESI. Int. J. Mass Spectrom 2002, 219, 39–62.

(3). Kebarle P; Verkerk UH Electrospray: From Ions in Solution to Ions in the Gas Phase, What We Know Now. Mass Spectrom. Rev 2009, 28, 898–917. [PubMed: 19551695]

(4). Peschke M; Verkerk UH; Kebarle P Features of the ESI Mechanism That Affect the Observation of Multiply Charged Noncovalent Protein Complexes and the Determination of the Association

Constant by the Titration Method. J. Am. Soc. Mass Spectrom 2004, 15, 1424–1434. [PubMed: 15465355]

(5). Testa L; Brocca S; Grandori R Charge-Surface Correlation in Electrospray Ionization of Folded and Unfolded Proteins. Anal. Chem 2011, 83, 6459–6463. [PubMed: 21800882]

(6). Donor MT; Ewing SA; Zenaidee MA; Donald WA; Prell JS Extended Protein Ions Are Formed by the Chain Ejection Model in Chemical Supercharging Electrospray Ionization. Anal. Chem 2017, 89, 5107–5114. [PubMed: 28368095]

(7). Natalello A; Santambrogio C; Grandori R Are Charge-State Distributions a Reliable Tool Describing Molecular Ensembles of Intrinsically Disordered Proteins by Native MS? J. Am. Soc. Mass Spectrom 2017, 28, 21–28. [PubMed: 27730522]

(8). Hall Z; Robinson CV Do Charge State Signatures Guarantee Protein Conformations? J. Am. Soc. Mass Spectrom 2012, 23, 1161–1168. [PubMed: 22562394]

(9). Grandori R Origin of the Conformation Dependence of Protein Charge-State Distributions in Electrospray Ionization Mass Spectrometry. J. Mass Spectrom 2003, 38, 11–15. [PubMed: 12526001]

(10). Shelimov KB; Clemmer DE; Hudgins RR; Jarrold MF Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome C. J. Am. Chem. Soc 1997, 119, 2240–2248.

(11). Raab SA; El-Baba TJ; Laganowsky A; Russell DH; Valentine SJ; Clemmer DE Protons Are Fast and Smart; Proteins Are Slow and Dumb: On the Relationship of Electrospray Ionization Charge States and Conformations. J. Am. Soc. Mass Spectrom 2021, 32, 1553–1561. [PubMed: 34151568]

(12). Dixit SM; Polasky DA; Ruotolo BT Collision Induced Unfolding of Isolated Proteins in the Gas Phase: Past, Present, and Future. Curr. Opin. Chem. Biol 2018, 42, 93–100. [PubMed: 29207278]

(13). Donor MT; Mroz AM; Prell JS Experimental and Theoretical Investigation of Overall Energy Deposition in Surface-Induced Unfolding of Protein Ions. Chem. Sci 2019, 10, 4097–4106. [PubMed: 31049192]

(14). Donor MT; Shepherd SO; Prell JS Rapid Determination of Activation Energies for Gas-Phase Protein Unfolding and Dissociation in a Q-IM-TOF Mass Spectrometer. J. Am. Soc. Mass Spectrom 2020, 31, 602–610. [PubMed: 32126776]

(15). Hong S; Bush MF Collision-Induced Unfolding Is Sensitive to the Polarity of Proteins and Protein Complexes. J. Am. Soc. Mass Spectrom 2019, 30, 2430–2437. [PubMed: 31502225]

(16). Zhou M; Dagan S; Wysocki VH Impact of Charge State on Gas-Phase Behaviors of Noncovalent Protein Complexes in Collision Induced Dissociation and Surface Induced Dissociation. Analyst 2013, 138, 1353–1362. [PubMed: 23324896]

(17). Gabelica V; Marklund E Fundamentals of Ion Mobility Spectrometry. Curr. Opin. Chem. Biol 2018, 42, 51–59. [PubMed: 29154177]

(18). Konijnenberg A; Butterer A; Sobott F Native Ion Mobility-Mass Spectrometry and Related Methods in Structural Biology. Biochim. Biophys. Acta Proteins Proteom 2013, 1834, 1239–1256.

(19). Gabelica V Ion Mobility–Mass Spectrometry: An Overview. In Ion Mobility-Mass Spectrometry: Fundamentals and Applications, Vol. Ashcroft AE; Sobott F, Eds.; The Royal Society of Chemistry, 2022; pp 1–25.

(20). Hall Z; Politis A; Bush MF; Smith LJ; Robinson CV Charge-State Dependent Compaction and Dissociation of Protein Complexes: Insights from Ion Mobility and Molecular Dynamics. J. Am. Chem. Soc 2012, 134, 3429–3438. [PubMed: 22280183]

(21). Allen SJ; Schwartz AM; Bush MF Effects of Polarity on the Structures and Charge States of Native-Like Proteins and Protein Complexes in the Gas Phase. Anal. Chem 2013, 85, 12055–12061. [PubMed: 24224685]

(22). Canzani D; Laszlo KJ; Bush MF Ion Mobility of Proteins in Nitrogen Gas: Effects of Charge State, Charge Distribution, and Structure. J. Phys. Chem. A 2018, 122, 5625–5634. [PubMed: 29864282]

(23). Duez Q; Metwally H; Konermann L Electrospray Ionization of Polypropylene Glycol: Rayleigh-Charged Droplets, Competing Pathways, and Charge State-Dependent Conformations. Anal. Chem 2018, 90, 9912–9920. [PubMed: 30024742]

(24). Laszlo KJ; Bush MF Effects of Charge State, Charge Distribution, and Structure on the Ion Mobility of Protein Ions in Helium Gas: Results from Trajectory Method Calculations. J. Phys. Chem. A 2017, 121, 7768–7777. [PubMed: 28910102]

(25). Rolland AD; Prell JS Computational Insights into Compaction of Gas-Phase Protein and Protein Complex Ions in Native Ion Mobility-Mass Spectrometry. Trends Anal. Chem 2019, 116, 282–291.

(26). Jurneczko E; Barran PE How Useful Is Ion Mobility Mass Spectrometry for Structural Biology? The Relationship between Protein Crystal Structures and Their Collision Cross Sections in the Gas Phase. Analyst 2011, 136, 20–28. [PubMed: 20820495]

(27). Scarff CA; Thalassinos K; Hilton GR; Scrivens JH Travelling Wave Ion Mobility Mass Spectrometry Studies of Protein Structure: Biological Significance and Comparison with X-Ray Crystallography and Nuclear Magnetic Resonance Spectroscopy Measurements. Rapid Commun. Mass Spectrom 2008, 22, 3297–3304. [PubMed: 18816489]

(28). Michaelevski I; Eisenstein M; Sharon M Gas-Phase Compaction and Unfolding of Protein Structures. Anal. Chem 2010, 82, 9484–9491. [PubMed: 20964410]

(29). Dole M; Mack LL; Hines RL; Mobley RC; Ferguson LD; Alice MB Molecular Beams of Macroions. J. Chem. Phys 1968, 49, 2240–2249.

(30). de la Mora JF Electrospray Ionization of Large Multiply Charged Species Proceeds Via Dole's Charged Residue Mechanism. Anal. Chim. Acta 2000, 406, 93–104.

(31). Kaltashov IA; Mohimen A Estimates of Protein Surface Areas in Solution by Electrospray Ionization Mass Spectrometry. Anal. Chem 2005, 77, 5370–5379. [PubMed: 16097782]

(32). Heck AJR; van den Heuvel RHH Investigation of Intact Protein Complexes by Mass Spectrometry. Mass Spectrom. Rev 2004, 23, 368–389. [PubMed: 15264235]

(33). Bush MF; Hall Z; Giles K; Hoyes J; Robinson CV; Ruotolo BT Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. Anal. Chem 2010, 82, 9557–9565. [PubMed: 20979392]

(34). Marklund EG; Degiacomi MT; Robinson CV; Baldwin AJ; Benesch JLP Collision Cross Sections for Structural Proteomics. Structure 2015, 23, 791–799. [PubMed: 25800554]

(35). Ruotolo BT; Benesch JLP; Sandercock AM; Hyung S-J; Robinson CV Ion Mobility-Mass Spectrometry Analysis of Large Protein Complexes. Nat. Protoc 2008, 3, 1139–1152. [PubMed: 18600219]

(36). Edwards AN; Tran HM; Gallagher ES Propagating Error through Traveling-Wave Ion Mobility Calibration. J. Am. Soc. Mass Spectrom 2021, 32, 2621–2630. [PubMed: 34662111]

(37). Salbo R; Bush MF; Naver H; Campuzano I; Robinson CV; Pettersson I; Jørgensen TJD; Haselmann KF Traveling-Wave Ion Mobility Mass Spectrometry of Protein Complexes: Accurate Calibrated Collision Cross-Sections of Human Insulin Oligomers. Rapid Commun. Mass Spectrom 2012, 26, 1181–1193. [PubMed: 22499193]

(38). van der Spoel D; Marklund EG; Larsson DSD; Caleman C Proteins, Lipids, and Water in the Gas Phase. Macromol. Biosci 2011, 11, 50–59. [PubMed: 21136535]

(39). Steinberg MZ; Elber R; McLafferty FW; Gerber RB; Breuker K Early Structural Evolution of Native Cytochrome C after Solvent Removal. ChemBioChem 2008, 9, 2417–2423. [PubMed: 18785672]

(40). Kaltashov IA; Fenselau C Stability of Secondary Structural Elements in a Solvent-Free Environment: The Alpha Helix. Proteins 1997, 27, 165–170. [PubMed: 9061780]

(41). Li A; Fenselau C; Kaltashov IA Stability of Secondary Structural Elements in a Solvent-Free Environment. II: The Beta-Pleated Sheets. Proteins 1998, Suppl 2, 22–27.

(42). Hudgins RR; Mao Y; Ratner MA; Jarrold MF Conformations of Gly_nH^+ and Ala_nH^+ Peptides in the Gas Phase. Biophys. J 1999, 76, 1591–1597. [PubMed: 10049339]

(43). Florance HV; Stopford AP; Kalapothakis JM; McCullough BJ; Bretherick A; Barran PE Evidence for α-Helices in the Gas Phase: A Case Study Using Melittin from Honey Bee Venom. Analyst 2011, 136, 3446–3452. [PubMed: 21701716]

(44). Kulesza A; Marklund EG; MacAleese L; Chirot F; Dugourd P Bringing Molecular Dynamics and Ion-Mobility Spectrometry Closer Together: Shape Correlations, Structure-Based Predictors, and Dissociation. J. Phys. Chem. B 2018, 122, 8317–8329. [PubMed: 30068075]

(45). Pagel K; Natan E; Hall Z; Fersht AR; Robinson CV Intrinsically Disordered P53 and Its Complexes Populate Compact Conformations in the Gas Phase. Angew. Chem. Int. Ed 2013, 52, 361–365.

(46). Warnke S; Von Helden G; Pagel K Protein Structure in the Gas Phase: The Influence of Side-Chain Microsolvation. J. Am. Chem. Soc 2013, 135, 1177–1180. [PubMed: 23320566]

(47). Chen S-H; Russell DH How Closely Related Are Conformations of Protein Ions Sampled by IM-MS to Native Solution Structures? J. Am. Soc. Mass Spectrom 2015, 26, 1433–1443. [PubMed: 26115967]

(48). Krone MG; Baumketner A; Bernstein SL; Wyttenbach T; Lazo ND; Teplow DB; Bowers MT; Shea J-E Effects of Familial Alzheimer's Disease Mutations on the Folding Nucleation of the Amyloid B-Protein. J. Mol. Biol 2008, 381, 221–228. [PubMed: 18597778]

(49). Segev E; Wyttenbach T; Bowers MT; Gerber RB Conformational Evolution of Ubiquitin Ions in Electrospray Mass Spectrometry: Molecular Dynamics Simulations at Gradually Increasing Temperatures. Phys. Chem. Chem. Phys 2008, 10, 3077–3082. [PubMed: 18688371]

(50). Bakhtiari M; Konermann L Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. J. Phys. Chem. B 2019, 123, 1784–1796. [PubMed: 30724571]

(51). Konermann L Molecular Dynamics Simulations on Gas-Phase Proteins with Mobile Protons: Inclusion of All-Atom Charge Solvation. J. Phys. Chem. B 2017, 121, 8102–8112. [PubMed: 28776996]

(52). McAllister RG; Metwally H; Sun Y; Konermann L Release of Native-Like Gaseous Proteins from Electrospray Droplets Via the Charged Residue Mechanism: Insights from Molecular Dynamics Simulations. J. Am. Chem. Soc 2015, 137, 12667–12676. [PubMed: 26325619]

(53). Metwally H; McAllister RG; Popa V; Konermann L Mechanism of Protein Supercharging by Sulfolane and *m*-Nitrobenzyl Alcohol: Molecular Dynamics Simulations of the Electrospray Process. Anal. Chem 2016, 88, 5345–5354. [PubMed: 27093467]

(54). Popa V; Trecroce DA; McAllister RG; Konermann L Collision-Induced Dissociation of Electrosprayed Protein Complexes: An All-Atom Molecular Dynamics Model with Mobile Protons. J. Phys. Chem. B 2016, 120, 5114–5124. [PubMed: 27218677]

(55). Konermann L; Metwally H; McAllister RG; Popa V How to Run Molecular Dynamics Simulations on Electrospray Droplets and Gas Phase Proteins: Basic Guidelines and Selected Applications. Methods 2018, 144, 104–112. [PubMed: 29678588]

(56). Patriksson A; Adams CM; Kjeldsen F; Zubarev RA; van der Spoel D A Direct Comparison of Protein Structure in the Gas and Solution Phase: The Trp-Cage. J. Phys. Chem. B 2007, 111, 13147–13150. [PubMed: 17973523]

(57). Saikusa K; Fuchigami S; Takahashi K; Asano Y; Nagadoi A; Tachiwana H; Kurumizaka H; Ikeguchi M; Nishimura Y; Akashi S Gas-Phase Structure of the Histone Multimers Characterized by Ion Mobility Mass Spectrometry and Molecular Dynamics Simulation. Anal. Chem 2013, 85, 4165–4171. [PubMed: 23485128]

(58). Ewing SA; Donor MT; Wilson JW; Prell JS Collidoscope: An Improved Tool for Computing Collisional Cross-Sections with the Trajectory Method. J. Am. Soc. Mass Spectrom 2017, 28, 587–596. [PubMed: 28194738]

(59). Breuker K; McLafferty FW Stepwise Evolution of Protein Native Structure with Electrospray into the Gas Phase, 10^{-12} to 10^{2} s. Proc. Natl. Acad. Sci. U. S. A 2008, 105, 18145–18152. [PubMed: 19033474]

(60). Prell JS; Chang TM; O'Brien JT; Williams ER Hydration Isomers of Protonated Phenylalanine and Derivatives: Relative Stabilities from Infrared Photodissociation. J. Am. Chem. Soc 2010, 132, 7811–7819. [PubMed: 20469865]

(61). Chang TM; Prell JS; Warrick ER; Williams ER Where's the Charge? Protonation Sites in Gaseous Ions Change with Hydration. J. Am. Chem. Soc 2012, 134, 15805–15813. [PubMed: 22954377]

(62). Bush MF; Prell JS; Saykally RJ; Williams ER One Water Molecule Stabilizes the Cationized Arginine Zwitterion. J. Am. Chem. Soc 2007, 129, 13544–13553. [PubMed: 17929811]

(63). Servage KA; Silveira JA; Fort KL; Russell DH Cryogenic Ion Mobility-Mass Spectrometry: Tracking Ion Structure from Solution to the Gas Phase. Acc. Chem. Res 2016, 49, 1421–1428. [PubMed: 27334393]

(64). Seo J; Hoffmann W; Warnke S; Bowers MT; Pagel K; von Helden G Retention of Native Protein Structures in the Absence of Solvent: A Coupled Ion Mobility and Spectroscopic Study. Angew. Chem. Int. Ed 2016, 55, 14173–14176.

(65). Zhong Y; Han L; Ruotolo BT Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. Angew. Chem. Int. Ed 2014, 53, 9209–9212.

(66). Benesch JLP; Aquilina JA; Ruotolo BT; Sobott F; Robinson CV Tandem Mass Spectrometry Reveals the Quaternary Organization of Macromolecular Assemblies. Chem. Biol 2006, 13, 597–605. [PubMed: 16793517]

(67). Blackwell AE; Dodds ED; Bandarian V; Wysocki VH Revealing the Quaternary Structure of a Heterogeneous Noncovalent Protein Complex through Surface-Induced Dissociation. Anal. Chem 2011, 83, 2862–2865. [PubMed: 21417466]

(68). Ma X; Zhou M; Wysocki VH Surface Induced Dissociation Yields Quaternary Substructure of Refractory Noncovalent Phosphorylase B and Glutamate Dehydrogenase Complexes. J. Am. Soc. Mass Spectrom 2014, 25, 368–379. [PubMed: 24452296]

(69). Sever AIM; Yin V; Konermann L Interrogating the Quaternary Structure of Noncanonical Hemoglobin Complexes by Electrospray Mass Spectrometry and Collision-Induced Dissociation. J. Am. Soc. Mass Spectrom 2021, 32, 270–280. [PubMed: 33124417]

(70). Ruotolo BT; Giles K; Campuzano I; Sandercock AM; Bateman RH; Robinson CV Evidence for Macromolecular Protein Rings in the Absence of Bulk Water. Science 2005, 310, 1658–1661. [PubMed: 16293722]

(71). Lee JH; Pollert K; Konermann L Testing the Robustness of Solution Force Fields for MD Simulations on Gaseous Protein Ions. J. Phys. Chem. B 2019, 123, 6705–6715. [PubMed: 31306020]

(72). Guvench O; Mackerell AD Comparison of Protein Force Fields for Molecular Dynamics Simulations. In Molecular Modeling of Proteins, Methods in Molecular Biology: Vol. 443; Kukol A, Ed. Humana Press, 2008; pp 63–88.

(73). Schuler LD; Van Gunsteren WF On the Choice of Dihedral Angle Potential Energy Functions for *n*-Alkanes. Mol. Simulat 2000, 25, 301–319.

(74). Schmid N; Eichenberger AP; Choutko A; Riniker S; Winger M; Mark AE; van Gunsteren WF Definition and Testing of the GROMOS Force-Field Versions 54a7 and 54b7. Eur. Biophys. J 2011, 40, 843–856. [PubMed: 21533652]

(75). Larriba C; Hogan CJ Free Molecular Collision Cross Section Calculation Methods for Nanoparticles and Complex Ions with Energy Accommodation. J. Comput. Phys 2013, 251, 344–363.

(76). Larriba C; Hogan CJ Ion Mobilities in Diatomic Gases: Measurement Versus Prediction with Non-Specular Scattering Models. J. Phys. Chem. A 2013, 117, 3887–3901. [PubMed: 23488939]

(77). Larriba-Andaluz C; Fernandez-Garcia J; Ewing MA; Hogan CJ; Clemmer DE Gas Molecule Scattering & Ion Mobility Measurements for Organic Macro-Ions in He Versus N_2 Environments. Phys. Chem. Chem. Phys 2015, 17, 15019–15029. [PubMed: 25988389]

(78). 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. A 1997, 101, 968–968.

(79). Chen CR; Makhatadze GI ProteinVolume: Calculating Molecular van der Waals and Void Volumes in Proteins. BMC Bioinform. 2015, 16, 101.

(80). Laszlo KJ; Buckner JH; Munger EB; Bush MF Native-Like and Denatured Cytochrome C Ions Yield Cation-to-Anion Proton Transfer Reaction Products with Similar Collision Cross-Sections. J. Am. Soc. Mass Spectrom 2017, 28, 1382–1391. [PubMed: 28224394]

(81). Lee JW; Davidson KL; Bush MF; Kim HI Collision Cross Sections and Ion Structures: Development of a General Calculation Method Via High-Quality Ion Mobility Measurements and Theoretical Modeling. Analyst 2017, 142, 4289–4298. [PubMed: 29034911]

(82). Chalk R; Borkowska O; Azeez KA; Oerum S; Born P; Gileadi O; Burgess-Brown N Electrospray Surface Charge Describes Protein Molecular Motion. bioRxiv 2019, DOI: 10.1101/571091.

(83). Prell JS Modelling Collisional Cross Sections. In Advances in Ion Mobility-Mass Spectrometry: Fundamentals, Instrumentation, and Applications, Comprehensive Analytical Chemistry: Vol. 83; Donald WA; Prell JS, Eds.; Elsevier, 2019; pp 1–22.

(84). Turzo SMBA; Seffernick JT; Rolland AD; Donor MT; Heinze S; Prell JS; Wysocki V; Lindert S Protein Shape Sampled by Ion Mobility Mass Spectrometry Consistently Improves Protein Structure Prediction. bioRxiv 2021, DOI: 10.1101/2021.1105.1127.445812.

(85). Biehn SE; Lindert S Protein Structure Prediction with Mass Spectrometry Data. Annu. Rev. Phys. Chem 2022, 73, DOI: 10.1146/annurev-physchem-082720-123928.

(86). Mackerell AD; Bashford D; Bellott M; Dunbrack RL; Evanseck JD; Field MJ; Fischer S; Gao J; Guo H; Ha S; Joseph-McCarthy D; Kuchnir L; Kuczera K; Lau FTK; Mattos C; Michnick S; Ngo T; Nguyen DT; Prodhom B; Reiher WE; Roux B; Schlenkrich M; Smith JC; Stote R; Straub J; Watanabe M; Wiórkiewicz-Kuczera J; Yin D; Karplus M All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J. Phys. Chem. B 1998, 102, 3586–3616. [PubMed: 24889800]

(87). Mackerell AD; Feig M; Brooks III CL Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. J. Comput. Chem 2004, 25, 1400–1415. [PubMed: 15185334]

(88). Kaminski GA; Friesner RA; Tirado-Rives J; Jorgensen WL Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins Via Comparison with Accurate Quantum Chemical Calculations on Peptides. J. Phys. Chem. B 2001, 105, 6474–6487.

(89). Cornell WD; Cieplak P; Bayly CI; Gould IR; Merz KM; Ferguson DM; Spellmeyer DC; Fox T; Caldwell JW; Kollman PA A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J. Am. Chem. Soc 1995, 117, 5179–5197.

(90). Allen SJ; Giles K; Gilbert T; Bush MF Ion Mobility Mass Spectrometry of Peptide, Protein, and Protein Complex Ions Using a Radio-Frequency Confining Drift Cell. Analyst 2016, 141, 884–891. [PubMed: 26739109]

(91). Bleiholder C; Liu FC Structure Relaxation Approximation (SRA) for Elucidation of Protein Structures from Ion Mobility Measurements. J. Phys. Chem. B 2019, 123, 2756–2769. [PubMed: 30866623]

(92). Wyttenbach T; Bowers MT Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility–Mass Spectrometry Environment. J. Phys. Chem. B 2011, 115, 12266–12275. [PubMed: 21905704]

(93). Porrini M; Rosu F; Rabin C; Darré L; Gómez H; Orozco M; Gabelica V Compaction of Duplex Nucleic Acids Upon Native Electrospray Mass Spectrometry. ACS Cent. Sci 2017, 3, 454–461. [PubMed: 28573208]

(94). Demireva M; O'Brien JT; Williams ER Water-Induced Folding of 1,7-Diammoniumheptane. J. Am. Chem. Soc 2012, 134, 11216–11224. [PubMed: 22708846]

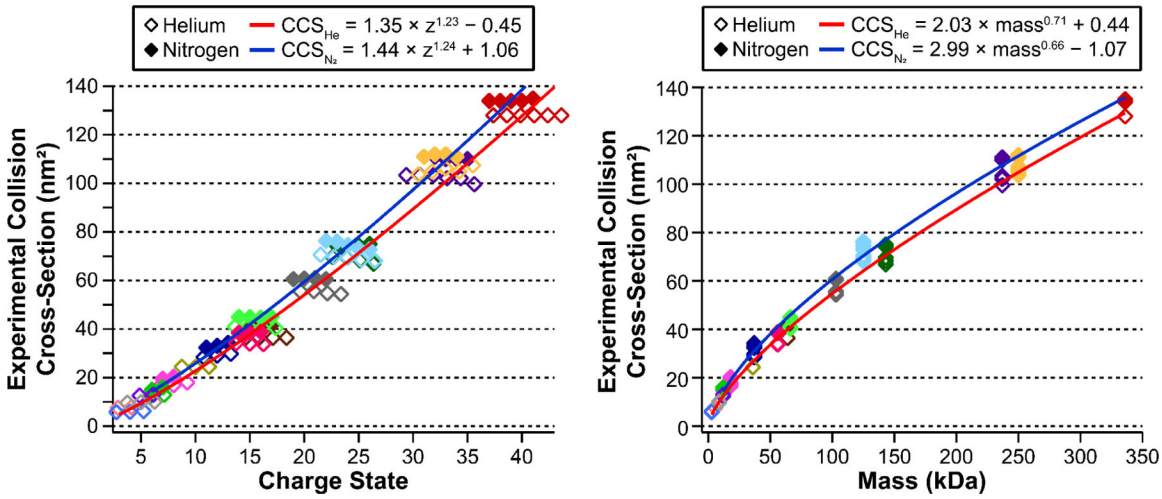


Figure 1.

Plot of experimental drift tube CCS values versus charge state (left) and mass (right) for 17-protein IM-MS calibrant data set.[33,37] Each color represents a different protein. Shaded and open markers represent CCSs measured in nitrogen or helium buffer gas, respectively. Trendlines represent CCS fits to power laws and are colored according to buffer gas identity as in legend.

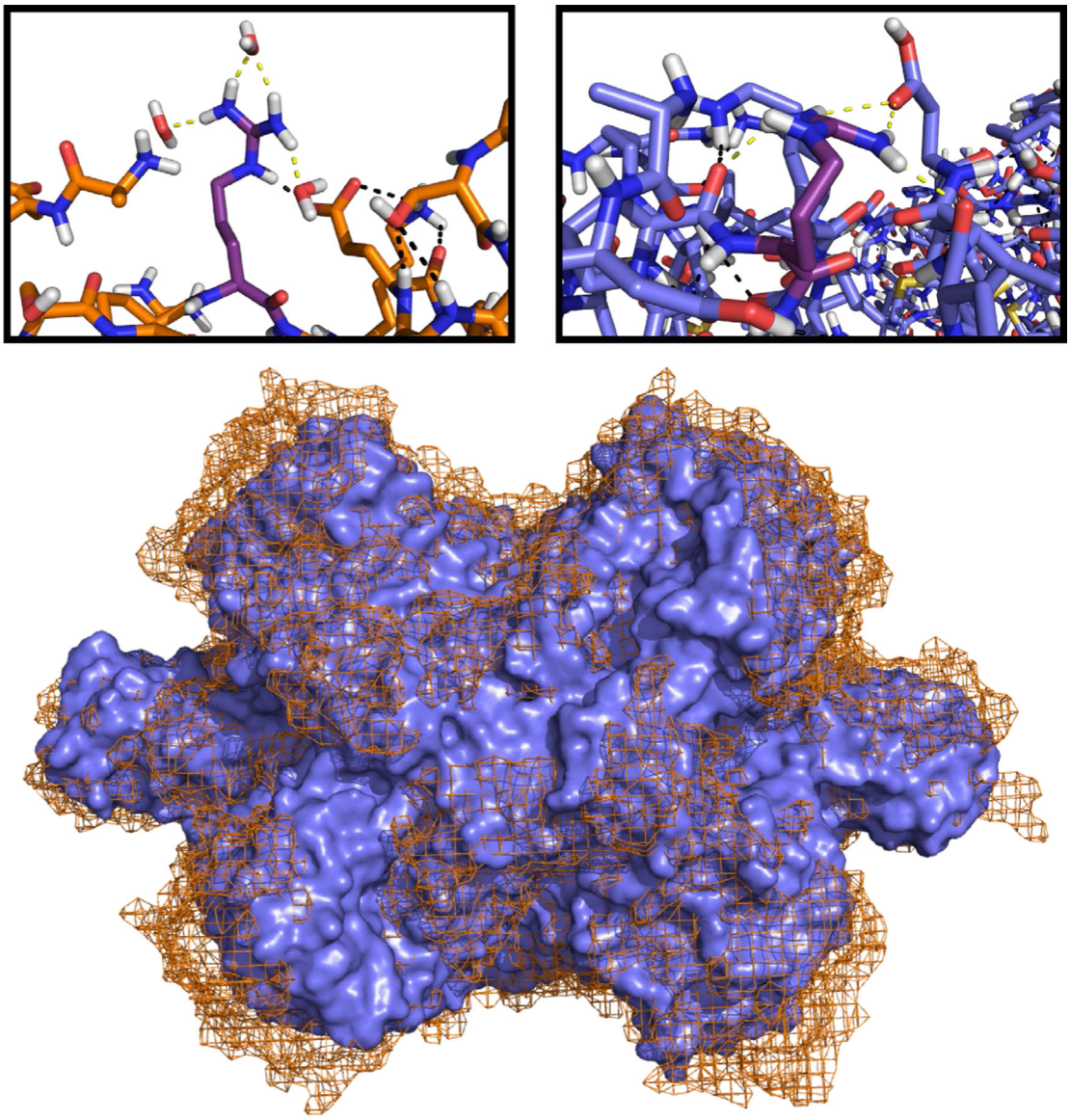


Figure 2.

Example of typical compaction with overlaid structures of a GDH 42+ charge conformer before (unsimulated cryo-EM structure, orange mesh) and after *in vacuo* molecular dynamics simulation (blue solid). Panels above show initial hydrogen bond formation for ARG338 (purple, charged site) with water in the condensed-phase structure and subsequent collapse to form interactions with other residues after simulation in absence of solvent, with yellow dotted lines representing the polar contacts involving the side chain amine groups. All other polar contacts involving side chains are indicated with black dotted lines for both structures.

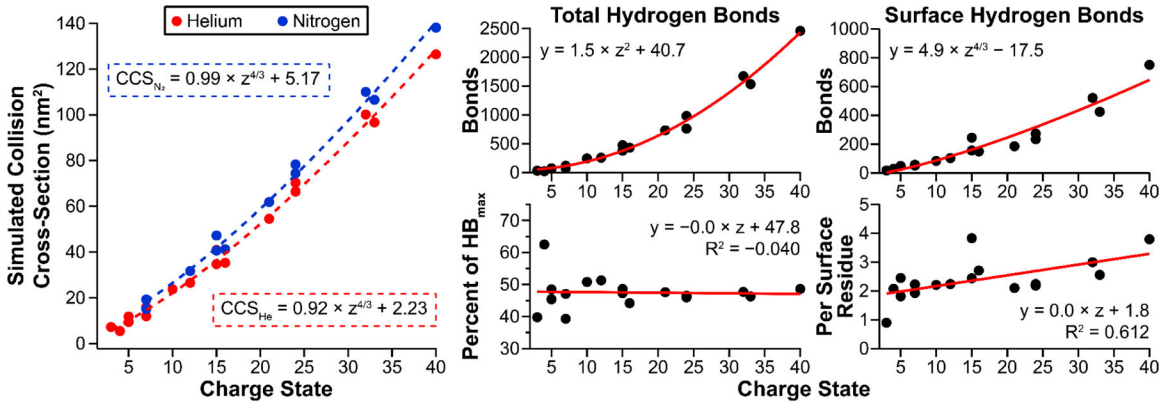


Figure 3.

Structural feature trends as a function of charge state for the original data set of simulations of 17 IM-MS calibrant proteins using one central charge state for each. CCS, total number of hydrogen bonds, and surface hydrogen bonds all exhibit power law relationships with increasing charge. Normalization of these metrics to the specific features of each protein before simulation (percent of maximum number of hydrogen bonds[38] and number of hydrogen bonds per surface residue) results in effectively flat trends.

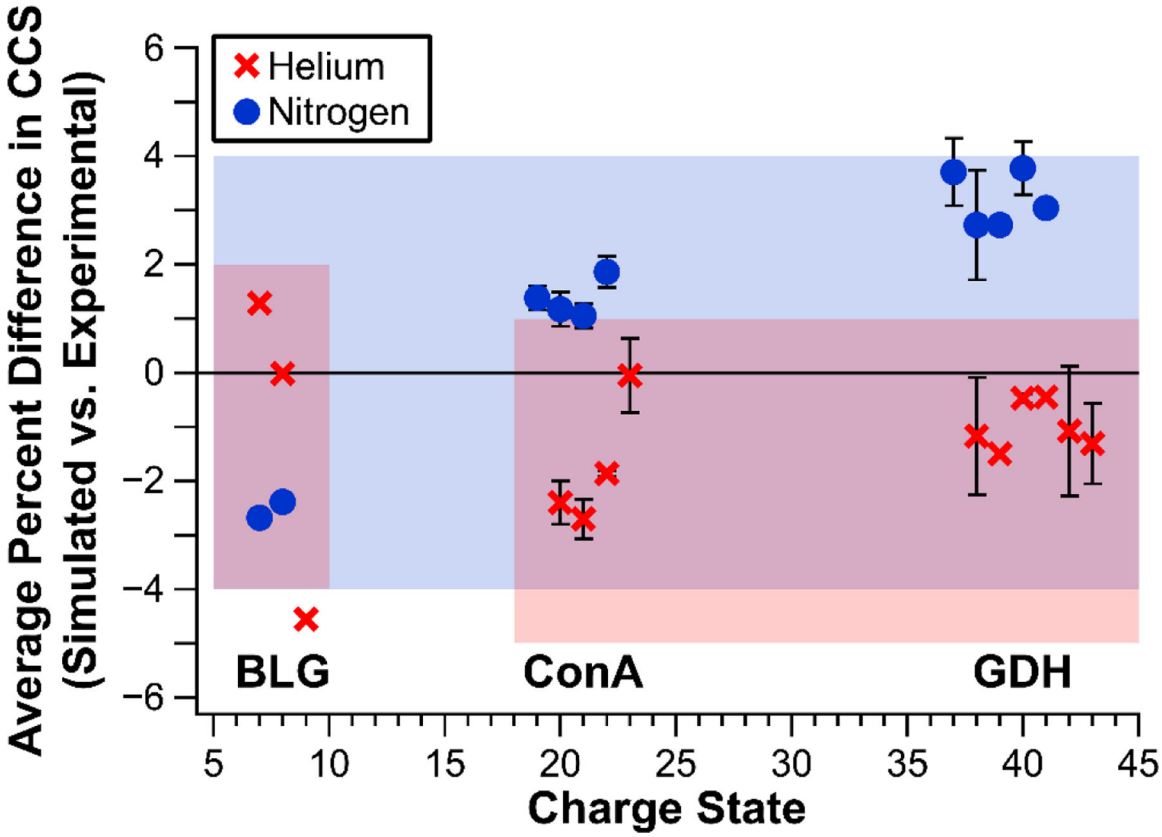


Figure 4.

Average percent difference between experimental drift tube[33,37] and simulated CCS for BLG, ConA, and GDH charge state conformations. Error bars represent 1 standard deviation. Symbols and coloring represent different buffer gases as indicated in legend. Shaded areas span the established range of accuracy and precision with the utilized MD simulation protocol for nitrogen buffer gas (0±4%) and helium buffer gas (−1±3% and −2±3% for lower charge states and higher charge states, respectively).

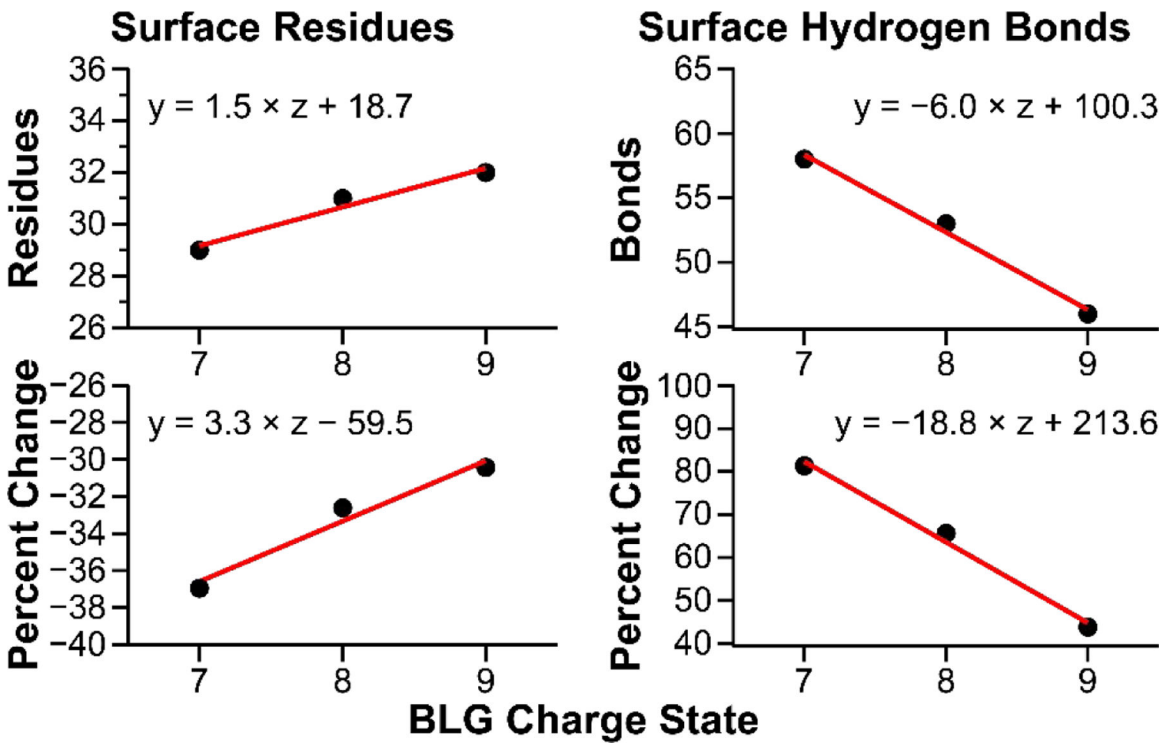


Figure 5.

Relatively strong correlations of surface residues and surface hydrogen bonds with charge state for BLG as determined with linear regression analysis. For each of these two structural features, the top plot shows the trend for the total number of surface residues and surface hydrogen bonds, while the bottom plots represent percent change relative to the unsimulated starting structure.

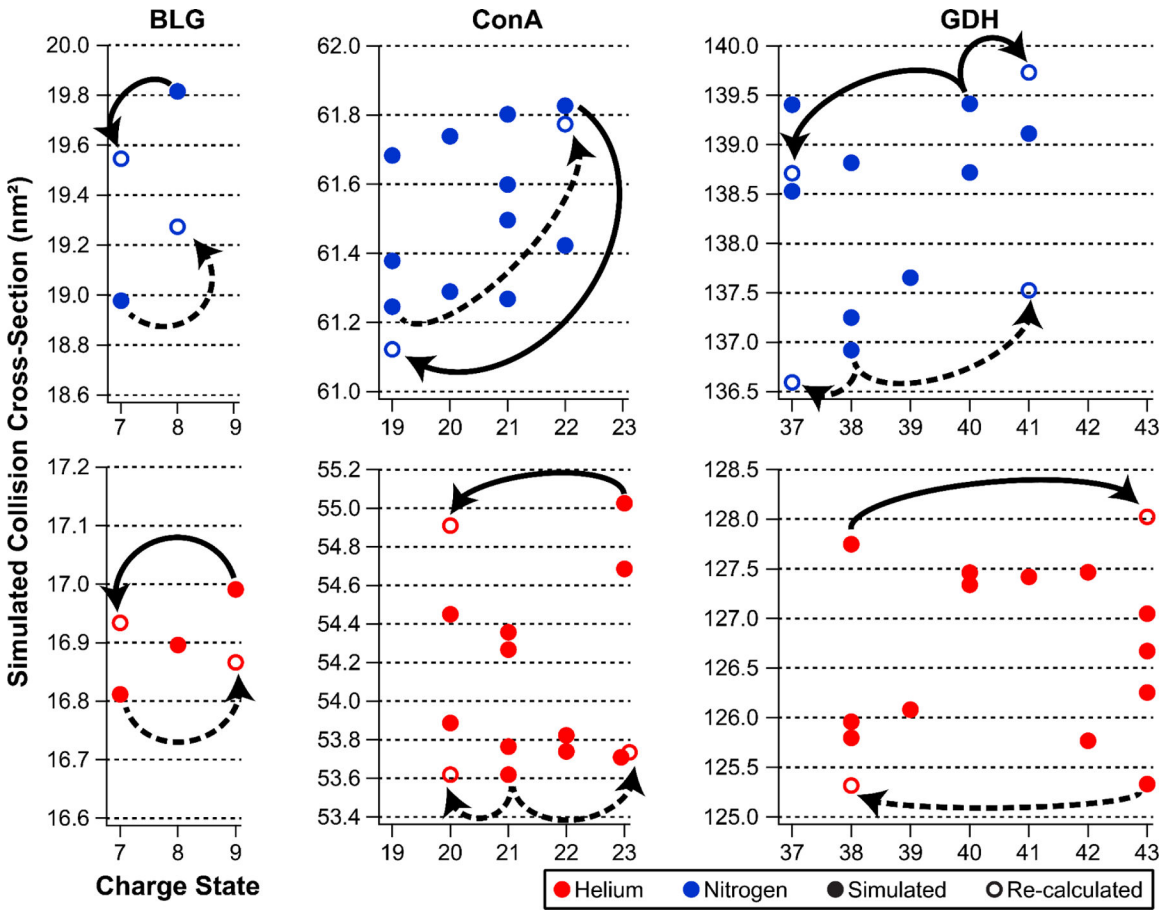


Figure 6.

Plots of computed CCSs for individual simulated charge conformers of BLG, ConA, and GDH against charge state (filled circles), with CCS values re-calculated assuming identical protein ion structure for each distribution (open circles). Arrows highlight relevant comparisons, either from the largest-CCS structure (solid arrow) or from the smallest-CCS structure (dotted arrow) to re-calculated CCSs at the lowest and highest charge states, respectively. Filled points from which two arrows originate (ConA helium and GDH nitrogen plots) indicate this structure was either the smallest or largest CCS but had not been simulated with either of the charge state extremes originally. Blue and red coloring correspond to values computed with nitrogen (top row) and helium buffer gas parameters (bottom row), respectively.

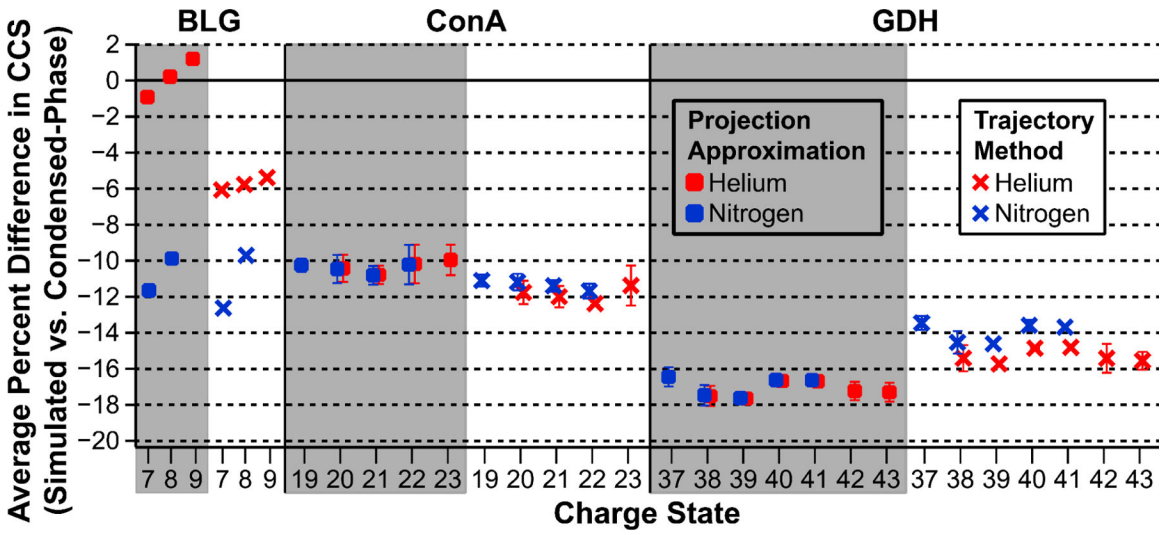


Figure 7.

Average percent difference in CCS for simulated charge conformers of BLG, ConA, and GDH in both buffer gases (color indicated in legend). Values are calculated using the difference between CCSs for the simulated and unsimulated structures, relative to the unsimulated structure. Trajectory Method calculations with Collidoscope and Projection Approximation calculations with IMoS represented by symbols indicated in legend. To provide further clarity, the full set of PA CCSs is shown first for each protein with a grey background, with the full set of TM CCSs adjacent.

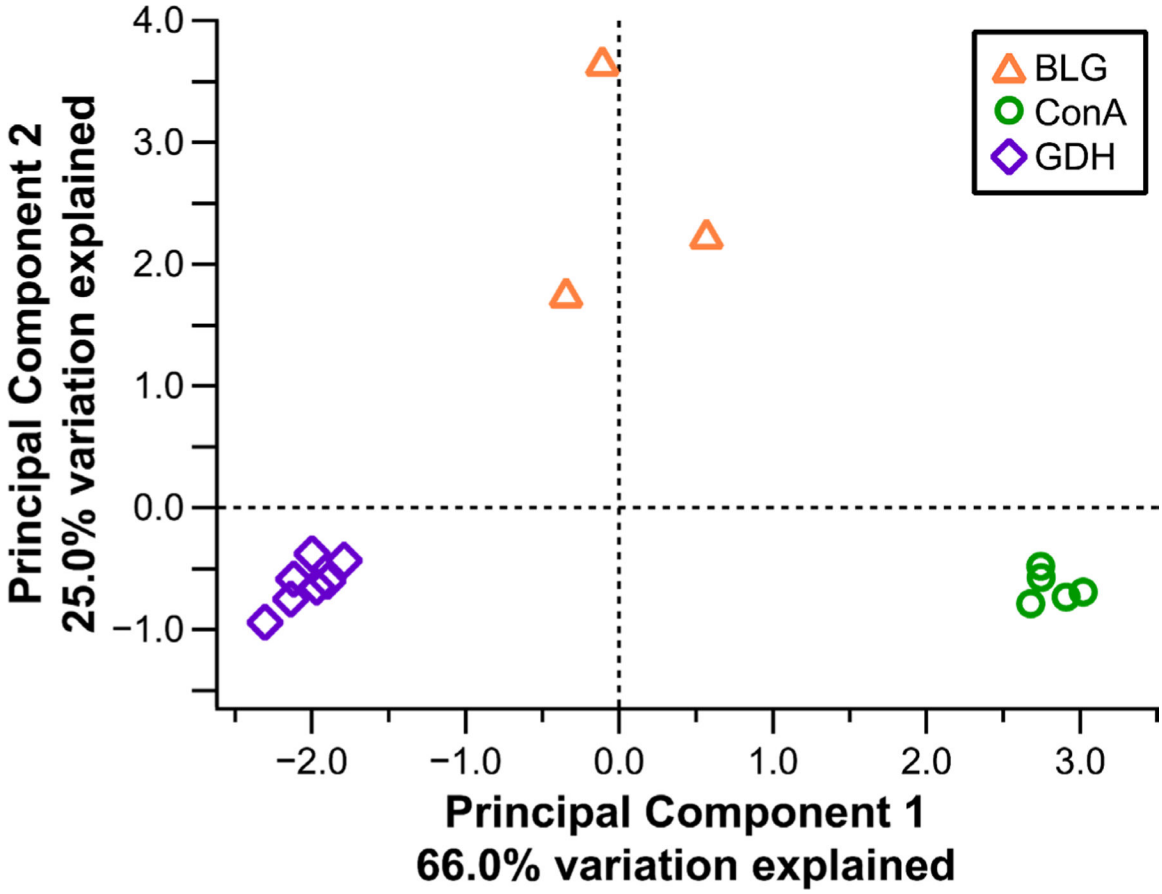


Figure 8.
Principal component analysis of BLG, ConA, and GDH structural features, with the average for each charge state plotted according to the first two principal components and proteins represented by colors/symbols as indicated in legend.