2-28-2019

# Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures.

Maryam Bakhtiari

Lars Konermann

# Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures

Maryam Bakhtiari and Lars Konermann*

*Department of Chemistry, The University of Western Ontario, London, Ontario, N6A 5B7, Canada*

* To whom correspondence should be addressed. Telephone: (519) 661-2111 ext. 86313.

Email: konerman@uwo.ca

**ABSTRACT:** Experiments and molecular dynamics (MD) simulations in the literature indicate that gaseous proteins generated by electrospray ionization (ESI) can retain native-like structures. However, the exact properties of these ions remain to be explored. Focusing on ubiquitin and lysozyme, we examined several pertinent questions. (1) We applied solvent MD runs to test whether the X-ray structures of both proteins are affected by crystal packing. Main and side chain orientations were retained in solution, providing a justification for the hitherto unscrutinized approach of relying on crystal data for "solution" vs. gas phase comparisons. (2) Most earlier gas phase protein MD investigations employed short (ns) simulation windows. By extending this time frame to 1 μs we were able to observe rare unfolding/folding transitions in ubiquitin. These predicted fluctuations were consistent with a semi-unfolded subpopulation detected by ion mobility spectrometry (IMS). (3) Most earlier modeling studies did not account for the high $H^+$ mobility in gaseous proteins. For the first time we compared static and mobile $H^+$ simulations, focusing on both positively and negatively charged ions. The MD runs revealed a strong preference for retention of a solution-like backbone fold, while titratable/polar side chains collapsed onto the protein surface. This side chain collapse was caused by a multitude of intramolecular salt bridges, H-bonds, and charge-dipole interactions. Our results generalize the findings of Steinberg *et al*. (*ChemBioChem* **2008**, *9*, 2417-2423) who had first proposed the occurrence of such side chain contacts on the basis of short-term simulations with static $H^+$. (4) Calculated collision cross sections of the MD conformers were in close agreement with IMS experiments. Overall, this study supports the view that solution-like protein structures can be retained due to kinetic trapping on the time scale of typical ESI-IMS experiments.

## Introduction

Electrospray ionization (ESI) allows the transfer of intact proteins from solution into the gas phase. The solvent-free ions generated in this way can be interrogated by mass spectrometry (MS), ion mobility spectrometry (IMS),[1-5] and various other techniques.[6-12] Protein ESI studies are usually conducted on $[M + zH]^{z+}$ ions. However, ESI can also be performed in negative ion mode, resulting in $[M - zH]^{z-}$ species.[1]

It remains controversial to what extent proteins retain solution-like structures after ESI.[4, 13-17] Some experiments suggested the occurrence of significant conformational changes.[13, 15, 18-20] On the other hand, there is now substantial evidence that many gaseous biomolecules can retain key aspects of their solution conformations and interactions, provided that the conditions are properly optimized.[2, 3, 17, 21-24] Such "native" ESI experiments employ nondenaturing aqueous solutions at near-neutral pH, with ion sampling and transmission conditions that avoid excess collisional heating.[7, 21, 25-28]

Native globular proteins in solution are tightly folded; most hydrophobic side chains are sequestered in the core, whereas titratable and polar side chains tend to be on the surface.[29] Proteins experience dramatic environmental changes during ESI. In water the hydrophobic effect represents a key stabilizing factor.[30] The absence of water after ESI suggests that hydrophobicity is not a major contributor to protein stability *in vacuo*.[31] Also, the protonation states of titratable sites can be very different in solution and after ESI.[32] The fact that ESI-generated ions carry a net charge $z \neq 0$ implies that their internal electrostatics are dominated by repulsive forces,[33-35] an aspect that is compounded by the low dielectric constant in the gas phase ($\varepsilon_{vacuum} = 1$).[36] Considering all these factors, the retention of solution-like structures after ESI is a remarkable phenomenon.[2, 3, 17, 21-24] This retention has been attributed to kinetic trapping,[14, 37, 38] i.e., the presence of activation barriers that prevent large-scale transitions on the time scale of typical ESI-IMS/MS experiments.[16]

Even under properly optimized native ESI conditions, the transition from solution into the gas phase will be accompanied by *some* structural alteration.[16] Molecular dynamics (MD) simulations[33, 39] and microsolvation studies[40] suggest that extended surface side chains such as $Lys^+$ tilt towards the protein surface where they bind to backbone carbonyls. Salt bridge formation among side chains on the protein surface may take place as well.[39] Such side chain collapse in the gas phase is consistent with the observation that IMS-derived collision cross sections (Ω) tend to be smaller than expected from crystal structures.[28, 33, 41-43] However, the exact nature of such gas phase compaction events remains incompletely understood, and predicted outcomes may depend on the modeling strategy used.[20, 34, 44-46]

Efforts to delineate protein structural differences before and after ESI are hampered by various factors. Chief among these is the lack of high-resolution structure determination methods in the gas phase. IMS-derived Ω values report on the overall compactness, but any given Ω value may be consistent with a multitude of conformations.[47] A standard approach is to compare experimental Ω values with MD-derived candidate structures.[4, 34, 47, 48] One challenge associated with gas phase MD simulations is the proper placement of charges. Proteins contain numerous sites that can be protonated (N-terminus, Arg, Lys, His) or deprotonated (Asp, Glu, C-terminus), giving rise to an astronomical number of possible charge configurations.[49] The problem becomes even more daunting when allowing for zwitterionic motifs, i.e., the presence of positively and negatively charged sites regardless of the ion polarity.[49-55] Strategies to identify the most suitable protonation pattern(s) have to consider interactions among protonated/deprotonated moieties, charge solvation effects, as well as the intrinsic proton affinities ($PA_{int}$) of individual sites.[36, 40, 56-58] Most MD studies treat protonation patterns as stationary,[34, 39, 59, 60] whereas in reality protons in gaseous proteins are mobile.[61-64] Mobile-proton MD approaches are a promising strategy for addressing these difficulties,[57, 58] but thus far such techniques are not widely used. Also, most

MD simulations focus on ns time windows that are much shorter than the tens of milliseconds required for typical IMS experiments.[4, 34, 47, 48]

Understanding ESI-associated structural changes is further complicated by uncertainties regarding the exact properties of proteins in solution. Much of the ESI literature assumes that X-ray crystallography provides an accurate representation of solution-phase conformations. Consequently, gas phase vs. "solution" comparisons usually rely on crystallographic data.[14, 39, 41, 45, 65] Although protein crystals contain some water,[66] X-ray and solution structures do not always match. Surface side chains and loops are highly dynamic in solution, while these elements tend to adopt well defined orientations after crystallization.[67-69] Such ordering arises from crystal packing, i.e., interactions with neighboring proteins in the lattice.[70, 71] Nuclear magnetic resonance (NMR) spectroscopy provides information on proteins in bulk solution. However, NMR structures are calculated largely from backbone distance and angle restraints.[72] The dynamic nature of exposed side chains implies that NMR coordinates of the corresponding atoms should be treated with caution.[69] In summary, the exact nature of protein structural differences before and after ESI remains to be uncovered.

The current work addresses the issues outlined above by taking a critical look at the crystal, solution, and gas phase structures of ubiquitin and lysozyme. We compared ESI-MS/IMS data with the results of extended (1 μs) gas phase MD data for both $[M + zH]^{z+}$ and $[M - zH]^{z-}$ ions. Mimicking MD strategies used in previous studies[34, 39, 59, 60] we initially employed static charge models. The main focus of this work, however, is on mobile-proton MD data generated with a recently developed mobile-proton algorithm[57] that accounts for the high mobility of $H^+$ in gas phase proteins.[61-64] Proof-of-principle work using this algorithm has already been presented,[57] but a critical comparison of data obtained by this approach with ESI-IMS/MS experiments is still lacking. The current work closes this gap by combining

experimental and computational data, thereby providing detailed insights into the properties of electrosprayed protein ions, and their relationship to solution and crystal structures.

## Methods

**Mass Spectrometry and Ion Mobility Spectrometry.** Bovine ubiquitin (8565 Da) and chicken egg white lysozyme (14305 Da) were purchased from Sigma (St. Louis, MO, USA). Samples were prepared in 10 mM aqueous ammonium acetate (pH 7) at a protein concentration of 10 μM. Positive ion data were acquired on a Synapt HDMS time-of-flight mass spectrometer (Waters, Milford, MA) with a standard Z-spray ESI source operated at +2.8 kV. Solutions were infused at 5 μL $min^{-1}$. Temperatures and ion transmission voltages were reduced as much as possible (source temperature 25 °C, desolvation temperature 40 °C, sampling cone 5 V, extraction cone 1 V, trap collision energy 2 V, trap DC bias 8-9 V, trap gas off). All other instrument parameters were as described.[73] These conditions were chosen to minimize collisionally or thermally induced structural changes of the protein,[28, 33] resulting in instrument operation just above the ion transmission threshold. A further drop in transmission was encountered for negative ion experiments, necessitating the use of a more sensitive instrument, a Synapt G2 (ESI voltage -2.8 kV) under otherwise identical conditions. Collision cross sections were measured using traveling wave ion mobility measurements with $N_2$ as buffer gas. The IMS data were converted to He Ω values. The calibration procedure used for this purpose was performed using a mix of collisionally activated reference proteins.[73] This procedure is well established and has been shown to yield He Ω values consistent with drift-tube IMS in positive ion mode,[73] but its use for negative ion IMS is new (see Table S1 for details). The IMS reference mix was infused in 49:49:2 (v/v/v) methanol/water/acetic acid with the cone set to 100 V, while keeping all other instrument

parameters as described above. Ω error bars represent the standard deviation of three independent measurements.

**Solution MD Simulations.** All simulations of this work were performed using Gromacs 2016.3[74] at 300 K with a 2 fs integration step. The crystal structures 1ubq (ubiquitin, 1.8 Å)[75] and 1aki (lysozyme, 1.5 Å)[76] served as starting conformations. For 1 μs solution simulations the Charmm36 force field[77] was used with TIP3P water because this combination has previously been shown to adequately model the behavior of small globular proteins.[78] The simulations employed periodic boundary conditions with a minimum 1 nm distance between protein atoms and the edge of the periodic box, and particle mesh Ewald summation. NaCl was added at a concentration of 150 mM, and additional ions were added as needed to ensure charge neutrality. All other MD parameters and procedures were as described.[79]

**Gas Phase MD Simulations.** The behavior of gaseous ubiquitin and lysozyme ions was modeled in 1 μs runs at 300 K, using the crystal data 1ubq and 1aki as starting points. The simulations employed the OPLS/AA force field[80, 81] which has been widely used for earlier gas phase applications,[15, 24, 60, 82-84] and that has been validated against *ab initio* data on model compounds in the absence of solvent.[80, 81] Our simulations were performed in a vacuum environment without cut-offs for electrostatic or Lennard-Jones interactions as described.[57] For static charge runs the protonation states of all titratable sites were set using Gromacs pdb2gmx, and these initial charge patterns were retained throughout the 1 μs runs.

Mobile-proton simulations were conducted by complementing Gromacs with in-house Fortran code and bash scripts,[57] and by breaking down 1 μs MD runs into 1 ns segments. After each of these segments the $H^+$ residing on the various protonated sites (N-terminus$^+$ [referred to as $NT^+$], $Arg^+$, $Lys^+$, $His^+$, $Asp^0$, $Glu^0$, C-terminus$^0$ [denoted as $CT^0$]) were

redistributed using a steepest descent energy minimization procedure that takes into account $PA_{int}$ of all possible acceptor sites,[85] as well as electrostatic interactions and intramolecular charge solvation. $PA_{int}$ values in kJ $mol^{-1}$ were[85] 886.6 (NT), 918.0 (Lys), 1004.0 (Arg) , 952.7 (His), 1452.7 ($Asp^-$), 1453.5 ($Glu^-$), 1423.8 ($CT^-$). This mobile-proton strategy allowed basic sites to switch between their protonated/neutral states, while acidic sites could switch between neutral/deprotonated. All proton hopping events were subject to preservation of total charge, in accordance with the net *z* value of the protein ion.[57] The presence of two titratable sites for the N-terminal Lys in lysozyme (NT and side chain) caused some difficulties with the code used here. This challenge was overcome by adding an N-terminal Gly to the sequence, such that the two charge sites were in two separate residues. X-ray data show that this N-terminus is part of a disordered tail,[76] and we conducted static charge simulations to ensure that the presence of this N-terminal Gly did not affect the structure or dynamics of lysozyme in any noticeable way. He collision cross sections were calculated using the trajectory method implemented in Collidoscope.[86]

## Results and Discussion

**Experimental Characterization of Gaseous Protein Ions.** Ubiquitin and lysozyme served as model systems for the native ESI experiments of this work. Both proteins possess a mixed α/β secondary structure. Ubiquitin does not have any disulfides, while lysozyme is stabilized by four disulfide bridges.[75, 76] The positive ESI ubiquitin mass spectrum was dominated by 6+ ions (Figure 1A). The corresponding Ω distribution had its main maximum at 960 $Å^2$. A low intensity satellite band was present around 1140 $Å^2$, revealing a small sub-population of semi-unfolded protein (Figure 1B). Negative ESI mainly produced 5- ions with a Ω maximum at ~990 $Å^2$ (Figure 1C, D). Lysozyme in positive ion mode exhibited a dominant 8+ signal, and the collision cross section of the corresponding ions was around 1390 $Å^2$

(Figure 1E, F). In negative ESI lysozyme generated 6- ions ($\Omega \approx 1470$ Å$^2$, Figure 1G, H). The standard deviation of these Ω maxima from triplicate measurements was around 1%. For both proteins the IMS peaks in positive ion mode were slightly wider, indicating a greater conformational heterogeneity. The gentle conditions chosen here broadened the lysozyme mass distribution in negative ion mode due to acetate adduction, but Ω values of adducted and non-adducted ions were almost indistinguishable (Figure S1). The ESI-IMS/MS data reported here are consistent previous literature reports [14, 28, 42, 43] Similar to other proteins, the $|z|$ values in positive ion mode were slightly higher than in negative ion mode, while the corresponding Ω values were similar for both polarities.[1, 14, 87]

**Crystal Contacts and Conformations.** As noted in the Introduction, many native ESI investigations have relied on crystal structures when comparing "solution" and gas phase proteins.[14, 39, 41, 45, 65] The underlying assumption is that crystal structures are a faithful representation of solution conformations. One objective of the current study was to test the validity of this approach, a task that requires a careful analysis of crystal structures.

The pdb coordinates of ubiquitin and lysozyme correspond to single polypeptide chains (green in Figure 2A, B). Standard visualization of these pdb files in a viewer program appears to show an isolated protein, however, such images are misleading. Figure 2A-B provides a more complete view that comprises a central protein (green) with its neighbors (gray, generated using the *symmetry mates* feature of Pymol). Analysis of the crystal packing interactions revealed a multitude of salt bridges between adjacent proteins, with O-N distances of ~4 Å or less.[88] The participating moieties include $Arg^+$, $Lys^+$, $Asp^-$, $Glu^-$ side chains, and C-terminal carboxylates (Figure 2C-D). These contacts form because the corresponding side chains point away from the protein surface towards neighboring polypeptide chains in the crystal. This trend is particularly prevalent for ubiquitin, which

exhibits tighter crystal packing than lysozyme. Both proteins have a protruding R-$COO^-$ terminus that is in contact with $Lys^+$ from a neighboring protein. In addition, lysozyme has a C-terminal $Arg^+/Asp^-$ contact (Figure 2C-D). The last four residues of ubiquitin are disordered in the crystal, and their coordinates represent a modeled average orientation.[75]

Crystal contacts such as those in Figure 2C-D can alter protein conformations,[67, 68, 70, 71] but the severity of such distortion effects has to be evaluated on a case-by-case basis. In the context of native ESI-MS it is of particular interest to explore whether the extended Lys/Arg/Asp/Glu/CT orientations seen in Figures 2C,D are caused by crystal packing. Alternatively, these moieties might already have a tendency to adopt extended orientations in bulk solution.[39] This chicken-or-egg problem is discussed in the following section.

**Crystal Structures vs. Solution MD Conformations.** MD simulations in bulk water were conducted to explore the structural preferences of ubiquitin and lysozyme in the absence of crystal packing. Three independent 1 μs solution runs for each protein yielded root mean square deviation (RMSD) values of 0.1 - 0.2 nm relative to the X-ray data, indicating that the solution conformations remained very close to the crystal structures (Figure S2). The resemblance of crystal and MD backbone coordinates is illustrated in Figure 2E-F. Noticeable differences were confined to the C-terminal tail of ubiquitin which was quite mobile for the solution structures (Figure 2E), in agreement with experimental data.[75] Of particular interest are the orientations of titratable side chains, many of which reach across to neighboring proteins in the crystal (see above, Figure 2C-D). Remarkably, these extended side chain orientations were retained in solution, albeit with a greater degree of disorder compared to the crystals (Figure 2G-H). This behavior is also evident from radius of gyration ($R_g$) analyses, which showed that the overall compactness of surface side chain atoms in solution was close to the corresponding X-ray values (Figure 2I-J).

The purpose of the current sections was to assess the appropriateness of crystal data as surrogates for solution conformations when discussing structural changes during ESI. For the two model proteins considered here the crystal conformations do indeed provide a faithful representation of the solution structure, validating the strategy used in earlier ESI studies.[14, 39, 41, 45, 65] However, this equivalence does not hold for all proteins, requiring careful case-by-case considerations.[67, 68, 70, 71] In instances where archived pdb structures provide a misleading picture of the solution state, or where condensed phase reference data are not available, the use of solution MD structures can be essential as previously noted by others.[89]

**Modeling Protons in Gas Phase MD Simulations.** Three different charge models were used when simulating the behavior of gaseous protein ions.

*(i) Static protons (positive-only or negative-only)*. This approach represents the simplest model tested here; it did not allow for proton transfer within the protein. Also, zwitterionic motifs were not considered, i.e., $[M + zH]^{z+}$ ions contained exactly $z$ positive charges at basic sites (N-terminus, Arg, Lys, His), while all acidic sites remained in their neutral R-COOH state. Conversely, $[M - zH]^{z-}$ ions were treated by having $z$ negative charges on selected $R\text{-}COO^-$ moieties (Asp, Glu, C-terminus), without any positive charges.

*(ii) Static protons (zwitterionic)*. This second model allowed the presence of zwitterionic motifs and salt bridges, reflecting the strong evidence that electrosprayed $[M + zH]^{z+}$ ions can contain $R\text{-}COO^-$ sites (and vice versa for $[M - zH]^{z-}$ ions).[49-55] Under these conditions the number of (de)protonated sites exceeds $z$ due to the presence of opposite charges. However, model 2 also did not allow for proton transfer events. Models 1[34, 48] and 2[39, 60] mimic strategies that have been widely used for earlier gas phase MD studies. Charge patterns used for all static proton runs are summarized in Tables S2 and S3.

*(iii) Mobile-protons*. The third model employed a recently developed technique[57] that accounts for the highly mobile nature of protons in gaseous proteins.[61-64] During MD runs all $H^+$ were allowed to move among basic and acidic sites, as governed by their mutual electrostatic interactions, charge solvation effects, and by the $PA_{int}$ values of protonation sites. Within this model, the formation of zwitterionic motifs and salt bridges was governed by the overall energetics of the system, without manual intervention by the user.[57]

**Overview of Gas Phase MD Results.** Multiple 1 μs simulations were generated for the dominant charge states seen in the experimental spectra (Figure 1). The overall behavior seen in these MD runs is illustrated by the $R_g$ values of Figure 3 for all three charge models.

Static proton/positive-only simulations on [ubiquitin + 6H]$^{6+}$ were conducted for various protonation patterns (Table S2). Most of the trajectories generated using this model retained tightly folded conformations, while two of the runs showed fluctuations between semi-unfolded and compact structures (Figure 3A). Subsequently we tested the behavior of [ubiquitin + 6H]$^{6+}$ in static proton/zwitterionic runs. All the trajectories produced in this way retained tightly folded structures (Figure 3B). Mobile-proton simulations on [ubiquitin + 6H]$^{6+}$ were dominated by compact conformers, with occasional unfolding/refolding for some of the trajectories (Figure 3C). Most of these reversible structural events would have gone undetected in the much shorter (~10 ns or less) time windows used for earlier gas phase simulations,[34, 39, 48] highlighting the necessity of using an extended time scale.

An analogous set of simulations was performed for [ubiquitin - 5H]$^{5-}$, resulting in the $R_g$ data of Figure 3D-E. Across the board, the negative ions displayed high stability without unfolding, regardless of the charge model used. The greater stability of negative ubiquitin ions relative to their positive counterparts is attributed to their lower internal electrostatic repulsion,[33-35] keeping in mind their $z$ values ($|5-| < |6+|$). To support this assertion, we

repeated MD runs for those positive ion patterns that consistently caused unfolding (red and blue, Figure 3A), but lowered the charge state from 6+ to 5+. As expected, the 5+ ions retained a compact conformation (Figure S4). The greater structural resilience of these charge-reduced ions is consistent with experimental observations.[90, 91]

MD simulations were also conducted on lysozyme in the gas phase, using all three protonation models in both polarities. The resulting data showed lysozyme to be highly stable, without any large-scale unfolding for any of the conditions tested (Figure S5, Table S3). The greater stability of lysozyme compared to ubiquitin is attributed to the presence of four disulfide bridges in the former.[76]

The key question explored here is to what extent ubiquitin and lysozyme retain their native solution structures in the gas phase. As discussed above, the crystal structures of both proteins provide an adequate representation of their solution conformations. Thus, Figure 4 compares mobile-proton MD conformations (gray) with the corresponding crystal data (green). Figure 4A exemplifies data from a $[\text{ubiquitin} + 6\text{H}]^{6+}$ trajectory that did not show unfolding, revealing that the backbone fold remained very close to that of the crystal structure throughout the 1 μs window. The same is true for $[\text{ubiquitin} - 5\text{H}]^{5-}$, $[\text{lysozyme} + 8\text{H}]^{8+}$, and $[\text{lysozyme} - 6\text{H}]^{6-}$ (Figure 4C-E). RMSD values relative to the corresponding crystal structures for these "folded" trajectories were on the order of 0.3 to 0.4 nm (Figures S6, S7), reinforcing the fact that the gas phase proteins remained close to their crystal structures (albeit not as close as in solution, where RMSDs were 0.1 - 0.2 nm, Figure S2).

Figure 4B illustrates one of the reversible unfolding transitions experienced by $[\text{ubiquitin} + 6\text{H}]^{6+}$. For this particular trajectory the protein opened up at $t \approx 100$ ns, (magenta), but at $t = 200$ ns it had folded back to a compact state. The fact that gaseous ubiquitin returned close to its original structure after being semi-unfolded implies that the

solution-like conformation represents a local free energy minimum, consistent with the kinetic trapping hypothesis outlined above.[14, 37, 38]

**Charge Patterns Produced in Mobile-proton Runs.** Our mobile-proton simulations started from various initial charge patterns (coinciding with those of the static charge runs, Tables S2, S3). Regardless of these starting configurations, the mobile-proton runs showed a strong propensity to generate zwitterionic patterns (Figure 5), i.e., R-$COO^-$ sites accumulated in $[M + zH]^{z+}$ ions, and $Arg^+$/$Lys^+$ (not $His^+$) sites accumulated in $[M - zH]^{z-}$ ions. $H^+$ migration had largely gone to completion after ~500 ns, producing fairly stable patterns (Figure 5).

Figure 5 also shows representative 1 μs structures, revealing that oppositely charged sites were involved in salt bridge networks at the protein surface. In other words, all R-$COO^-$ moieties in $[M + zH]^{z+}$ ions were in contact with at least one positively charged sites, and vice versa for $[M - zH]^{z-}$ ions.

Arg was the preferred basic $H^+$ residence site, in accordance with its high $PA_{int}$. Negative charges were distributed over Asp, Glu and CT sites, reflecting their similar $PA_{int}$ values.[85] The zwitterionic character after 1 μs was more pronounced for gaseous ubiquitin than for lysozyme (Figure 5). This difference is attributed to the higher percentage of basic/acidic sites in ubiquitin which promotes the formation of Coulombically favorable salt bridge networks (Tables S2, S3). A detailed summary of the average charge on each of the titratable sites in the mobile-proton simulations is provided in Figure S8.

Overall, the current work marks the first time that the structural dynamics of gaseous proteins were simulated on an extended (1 μs) time scale using a model that allowed for $H^+$ hopping. The strong preference for zwitterionic motifs seen in our simulations is consistent with experimental investigations.[49-55] Nonetheless, it is clear that the simple mobile-proton MD model used here[57] may not always capture the subtle energetic differences between

zwitterionic and neutral acidic/basic contacts. Electronic structure calculations would be required for this purpose,[56, 92] but unfortunately the size and time range explored here are out of reach for such high-level methods. At the current stage of development our mobile-proton MD model[57] represents a reasonable compromise between speed and accuracy.

**Surface Side Chain Collapse.** Gaseous ubiquitin and lysozyme retained an overall backbone fold close to the corresponding solution/X-ray structures, as discussed above. However, inspection of Figure 3 reveals the gas phase $R_g$ values for most of the simulations were 3-5% lower than for the X-ray coordinates, indicating compaction in the gas phase. To uncover the origin of this effect we calculated $R_g$ values for various sub-groups of atoms. The most pronounced $R_g$ drop was seen for surface side chains. Among these, the titratable sites Arg, Lys, His, Asp, Glu made the largest contribution. This side chain collapse occurred for both ubiquitin and lysozyme. This phenomenon was largely independent of the protonation model, and it took place for both $[M + zH]^{z+}$ and $[M - zH]^{z-}$ ions (Figures 6, S9). It has been noted previously that side chain collapse not only reduces the overall dimensions, but also lowers the net dipole moment of gaseous proteins.[39]

The origin of this side chain collapse is illustrated in Figure 7, which displays representative mobile-proton MD structures of positive and negative ubiquitin ions. Figure 7 reveals numerous intramolecular contacts that force the titratable side chains from their extended X-ray/solution structures (Figure 2) into orientations close to the protein surface. In addition to salt bridges, these contacts include charge-dipole interactions where cationic sites are solvated by backbone CO groups, while $R\text{-}COO^-$ sites interact with backbone NH sites. Many contacts also involve H-bonds among uncharged titratable sites, reflecting the capability of R-COOH, amino groups, and guanidine groups to act as both H-bond donors

(OH, NH) and acceptors (C=O, N:). Side chain contacts similar to those discussed here were also observed in static proton/zwitterionic simulations (data not shown).

Considering that salt bridges are a major contributor to side chain collapse in zwitterionic runs, it may seem surprising that a similar compaction was also seen in simulations that did not allow for zwitterions (Figure 6, S9). The reason for this effect is that even in the absence of oppositely charged moieties most of the titratable side chains formed a tightly connected network at the protein surface (Figure S10). These non-zwitterionic contacts involved numerous H-bonds among neutral moieties, exemplified by the cluster of $Lys27^0$, $Arg42^0$, $Glu51^0$, $Asp52^0$, $CT^0$ (Figure S10B). H-bonding and charge solvation involving non-titratable sites (Asn, Gln, Ser, and Thr, plus backbone NH and CO) took place under non-zwitterionic conditions as well.

The formation of extensive intramolecular contacts involving side chains, and the resulting protein compaction seen in our MD data is consistent with earlier simulations by Steinberg et al.[39] However, that earlier work[39] only tested a static proton/zwitterionic model in positive ion mode. The current study extends the work of Steinberg et al.[39] by demonstrating that similar side chain collapse takes place in both positive and negative ion mode, and under different simulation conditions (with or without zwitterions, and with or without mobile protons).

**MD and Experimental Ω Values.** Figure 8 compares MD Ω values with the corresponding experiments. Ω values were also calculated for crystal and solution MD structures (after water removal). Like Figure 4, Figure 8 contains five panels because $[ubiquitin + 6H]^{6+}$ appeared in two forms, tightly folded and semi-unfolded. The following discussion initially focuses on folded species, excluding those ubiquitin trajectories that showed conformational changes (Figure 8A, C-E). The following key points are apparent from these data:

(1) Crystal and solution MD data show very similar Ω values. This observation reflects the fact that the crystal and solution MD conformations are virtually identical, as highlighted in Figure 2. In other words, for the two model systems studied here the crystal coordinates provide an adequate representation of the protein behavior in solution.

(2) Experimental Ω values for electrosprayed ubiquitin and lysozyme are significantly smaller than expected from their crystal structures, revealing that the proteins become more compact upon transitioning from solution into the gas phase. This effect has previously been observed in experiments on various proteins in positive ESI.[28, 33, 41-43] The current study reveals that the same effect also takes place in negative ESI (Figure 8C, E). For the proteins studied here, this collapse is most prevalent for $[ubiquitin + 6H]^{6+}$ (~16%, Figure 8A), and it is least pronounced for $[lysozyme -6H]^{6-}$ (~3.5%, Figure 8E).

(3) As outlined in Figure 7, gas phase protein contraction is caused by titratable and polar side chains at the protein surface. After desolvation these side chains collapse from their initially extended crystal/solution structures onto the protein surface where they form tight networks involving salt bridges, charge-dipole contacts, and H-bonds. The occurrence of such collapse events has previously been proposed on the basis of positive ion zwitterionic MD runs.[39] The current work generalizes this finding by demonstrating that compaction takes place in both polarities, and for three different charge models.

(4) It is remarkable that the protein compaction in our simulations is virtually independent of the charge model used. All three models yielded Ω values that agreed closely with each other (Figures 8A, C-E). We initially expected that the extent of side chain collapse would depend on the charge model used. However, our data show that different types of contacts (e.g. mostly salt bridged vs. primarily H-bonded, Figures 7, S10) cause very similar side chain compaction. Gratifyingly, all the collapsed MD-generated gas phase structures had Ω values close to the experimental data for both proteins and in both polarities.

(5) In our IMS data, only [ubiquitin + 6H]$^{6+}$ showed the presence of semi-unfolded species - evident from the satellite signal at ~1140 Å$^2$ (Figure 1B). Consistent with this experimental observation, only MD runs for [ubiquitin + 6H]$^{6+}$ exhibited occasional transitions to semi-unfolded conformers (Figure 3A, C). The Ω values of these non-native MD structures (with $R_g$ > 1.3 nm) agreed closely with the experimental data (Figure 8B). In other words, non-compact conformers such as the magenta species in Figure 4B represent likely candidate structures for the ~1140 Å$^2$ signal seen in the IMS data of [ubiquitin + 6H]$^{6+}$ (Figure 1B). The unfolding/refolding transitions of [ubiquitin + 6H]$^{6+}$ would likely have gone undetected in the much shorter (ns) simulation windows used for earlier studies.

## Conclusions

This work compared MD-generated gas phase protein structures with collision cross sections from native ESI experiments. We systematically studied the behavior of three different charge models for two proteins in both polarities in an extended (1 μs) MD time window. Most of the trajectories revealed the persistence of tightly folded gas phase conformers, with backbone conformers that closely resembled the initial solution conformations. It has been suggested that the thermodynamically stable state of gaseous proteins corresponds to "inside out" conformers, where nonpolar residues are exposed to the surface, while polar and titratable side chains are buried in the interior.[93] The persistence of proper "outside out" conformers (with charged/polar sites on the surface) in our simulations is consistent with the view that solution-like structures can survive after ESI due to kinetic trapping.[14, 37, 38]

Overall, the current work highlights the potential, but also the limitations of using gas phase investigations for deducing protein structural features of in solution. We found the backbone fold to be largely retained (with the exception of flexible termini), while surface side chains underwent substantial changes in their orientations and interactions. It is hoped

that advances in hardware and software will soon extend gas phase MD time scales even further, such that it will become possible to probe kinetic trapping and structural retention in mobile-proton simulations that match the millisecond time frame of typical IMS experiments.

**Supporting Information.** Additional figures as mentioned in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

**Acknowledgments.** Funding for this work was provided by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-04243). We thank H. Metwally for experimental help, as well as B. H. Shilton, G. S. Shaw and members of the Konermann laboratory for helpful discussions.

## References


(1) Heck, A. J. R.; Van den Heuvel, R. H. H. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **2004,** *23*, 368-389.
(2) Mehmood, S.; Allison, T. M.; Robinson, C. V. Mass spectrometry of protein complexes: From origins to applications. *Annu. Rev. Phys. Chem.* **2015,** *66*, 453-474.
(3) Benesch, J. L. P.; Ruotolo, B. T. Mass spectrometry: Come of age for structural and dynamical biology. *Curr. Op. Struct. Biol.* **2011,** *21*, 641-649.
(4) Chen, S.-H.; Russell, D. H. 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.
(5) Kaddis, C. S.; Loo, J. A. Native protein ms and ion mobility: Large flying proteins with esi. *Anal. Chem.* **2007,** *79*, 1779-1784.
(6) Quintyn, R. S.; Zhou, M.; Yan, J.; Wysocki, V. H. Surface-induced dissociation mass spectra as a tool for distinguishing different structural forms of gas-phase multimeric protein complexes. *Anal. Chem.* **2015,** *87*, 11879-11886.
(7) Zhang, H.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native electrospray and electron-capture dissociation fticr mass spectrometry for top-down studies of protein assemblies. *Anal. Chem.* **2011,** *83*, 5598-5606.
(8) Morrison, L. J.; Brodbelt, J. S. Charge site assignment in native proteins by ultraviolet photodissociation (uvpd) mass spectrometry. *Analyst* **2016,** *141*, 166-176.
(9) Seo, J.; Hoffmann, W.; Warnke, S.; Huang, X.; Gewinner, S.; Schollkopf, W.; Bowers, M. T.; von Helden, G.; Pagel, K. An infrared spectroscopy approach to follow beta-sheet formation in peptide amyloid assemblies. *Nat. Chem.* **2017,** *9*, 39-44.
(10) Frankevich, V.; Barylyuk, K.; Chingin, K.; Nieckarz, R.; Zenobi, R. Native biomolecules in the gas phase? The case of green fluorescent protein. *ChemPhysChem.* **2013,** *14*, 929-935.
(11) Czar, M. F.; Zosel, F.; Koenig, I.; Nettels, D.; Wunderlich, B.; Schuler, B.; Zarrine-Afsar, A.; Jockusch, R. A. Gas-phase fret efficiency measurements to probe the conformation of mass-selected proteins. *Anal. Chem.* **2015,** *87*, 7559-7565.
(12) Mistarz, U. H.; Brown, J. M.; Haselmann, K. F.; Rand, K. D. Probing the binding interfaces of protein complexes using gas-phase h/d exchange mass spectrometry. *Structure* **2016,** *24*, 310-318.
(13) Skinner, O. S.; McLafferty, F. W.; Breuker, K. How ubiquitin unfolds after transfer into the gas phase. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1011-1014.
(14) Wyttenbach, T.; Bowers, M. T. 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.
(15) Ly, T.; Julian, R. R. Elucidating the tertiary structure of protein ions in vacuo with site specific photoinitiated radical reactions. *J. Am. Chem. Soc.* **2010,** *132*, 8602-8609.
(16) Breuker, K.; McLafferty, F. W. 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.
(17) Seo, J.; Hoffmann, W.; Warnke, S.; Bowers, M. T.; 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. Edit.* **2016,** *55*, 14173-14176.
(18) Badman, E. R.; Hoaglund-Hyzer, C. S.; Clemmer, D. E. Monitoring structural changes of protein in and ion trap over ~ 10-200 ms: Unfolding transitions in cytochrome *c* ions. *Anal. Chem.* **2001,** *73*, 6000-6007.

(19) Bich, C.; Baer, S.; Jecklin, M. C.; Zenobi, R. Probing the hydrophobic effect of noncovalent complexes by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 286-289.
(20) Devine, P. W. A.; Fisher, H. C.; Calabrese, A. N.; Whelan, F.; Higazi, D. R.; Potts, J. R.; Lowe, D. C.; Radford, S. E.; Ashcroft, A. E. Investigating the structural compaction of biomolecules upon transition to the gas-phase using esi-twims-ms. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1855-1862.
(21) Leney, A. C.; Heck, A. J. R. Native mass spectrometry: What is in the name? *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 5-13.
(22) Wang, W.; Kitova, E. N.; Klassen, J. S. Influence of solution and gas phase processes on protein-carbohydrate binding affinities determined by nanoelectrospray fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **2003,** *75*, 4945-4955.
(23) Gavriilidou, A. F. M.; Holding, F. P.; Mayer, D.; Coyle, J. E.; Veprintse, D. B.; Zenobi, R. Native mass spectrometry gives insight into the allosteric binding mechanism of m2 pyruvate kinase to fructose-1,6-bisphosphate. *Biochemistry* **2018,** *57*, 1685-1689.
(24) Konermann, L.; Metwally, H.; McAllister, R. G.; 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.
(25) Porrini, M.; Rosu, F.; Rabin, C.; Darre, L.; Gomez, H.; Orozco, M.; Gabelica, V. Compaction of duplex nucleic acids upon native electrospray mass spectrometry. *ACS Central Sci.* **2017,** *3*, 454–461.
(26) Bonvin, G.; Bobst, C. E.; Kaltashov, I. A. Interaction of transferrin with non-cognate metals studied by native electrospray ionization mass spectrometry. *Int. J. Mass Spectrom.* **2017,** *420*, 74-82.
(27) Gabelica, V.; Livet, S.; Rosu, F. Optimizing native ion mobility q-tof in helium and nitrogen for very fragile noncovalent structures. *J. Am. Soc. Mass Spectrom.* **2018,** *29*, 2189-2198.
(28) Hopper, J. T. S.; Oldham, N. J. Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: The effect of ligand binding on conformational stability. *J. Am. Soc. Mass Spectrom.* **2009,** *20*, 1851-1858.
(29) Dill, K. A.; MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **2012,** *338*, 1042-1046.
(30) Brini, E.; Fennell, C. J.; Fernandez-Serra, M.; Hribar-Lee, B.; Luksic, M.; Dill, K. A. How water's properties are encoded in its molecular structure and energies. *Chem. Rev.* **2017,** *117*, 12385-12414.
(31) Liu, L.; Bagal, D.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Hydrophobic protein-ligand interactions preserved in the gas phase. *J. Am. Chem. Soc.* **2009,** *131*, 15980-15981.
(32) Wang, G.; Cole, R. B. Disparity between solution-phase equilibria and charge state distributions in positive-ion electrospray mass spectrometry. *Org. Mass Spectrom.* **1994,** *29*, 419-427.
(33) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Protein structure in vacuo: The gas-phase conformation of bpti and cytochrome c. *J. Am. Chem. Soc.* **1997,** *119*, 2240-2248.
(34) 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.
(35) Laszlo, K. J.; Munger, E. B.; Bush, M. F. Folding of protein ions in the gas phase after cation -to-anion proton-transfer reactions. *J. Am. Chem. Soc.* **2016,** *138*, 9581-9588.

(36) Schnier, P. D.; Gross, D. S.; Williams, E. R. Electrostatic forces and dielectric polarizability of multiply protonated gas-phase cytochrome c ions probed by ion/molecule chemistry. *J. Am. Chem. Soc.* **1995,** *117*, 6747-6757.
(37) Hendricks, N. G.; Julian, R. R. Leveraging ultraviolet photodissociation and spectroscopy to investigate peptide and protein three-dimensional structure with mass spectrometry. *Analyst* **2016,** *141*, 4534-4540.
(38) Clemmer, D. E.; Russell, D. H.; Williams, E. R. Characterizing the conformationome: Toward a structural understanding of the proteome. *Accounts Chem. Res.* **2017,** *50*, 556-560.
(39) Steinberg, M. Z.; Elber, R.; McLafferty, F. W.; Gerber, R. B.; Breuker, K. Early structural evolution of native cytochrome c after solvent removal. *ChemBioChem* **2008,** *9*, 2417-2423.
(40) 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.
(41) Jurneczko, E.; Barran, P. E. 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.
(42) Shi, H.; Atlasevich , N.; Merenbloom, S. I.; Clemmer, D. E. Solution dependence of the collisional activation of ubiquitin [m + 7h]7+ ions. *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 2000-2008.
(43) Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H. 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.
(44) Pacholarz, K. J.; Barran, P. E. Distinguishing loss of structure from subunit dissociation for protein complexes with variable temperature ion mobility mass spectrometry. *Anal. Chem.* **2015,** *87*, 6271-6279.
(45) Hogan, C. J.; Ruotolo, B. T.; Robinson, C. V.; de la Mora, J. F. Tandem differential mobility analysis-mass spectrometry reveals partial gas-phase collapse of the groel complex. *J. Phys. Chem. B* **2011,** *115*, 3614-3621.
(46) Michaelevski, I.; Eisenstein, M.; Sharon, M. Gas-phase compaction and unfolding of protein structures. *Anal. Chem.* **2010,** *82*, 9484-9491.
(47) Gabelica, V.; Marklund, E. Fundamentals of ion mobility spectrometry. *Curr. Op. Chem. Biol.* **2017,** *42*, 51-59.
(48) Mao, Y.; Woenckhaus, J.; Kolafa, J.; Ratner, M. A.; Jarrold, M. F. Thermal unfolding of unsolvated cytochrome c: Experiment and molecular dynamics simulations. *J. Am. Chem. Soc.* **1999,** *121*, 2712-2721.
(49) Li, J.; Santambrogio, C.; Brocca, S.; Rossetti, G.; Carloni, P.; Grandori, R. Conformational effects in protein electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **2016,** *35*, 111-122.
(50) Yoo, H. J.; Wang, N.; Zhuang, S. Y.; Song, H. T.; Hakansson, K. Negative-ion electron capture dissociation: Radical-driven fragmentation of charge-increased gaseous peptide anions. *J. Am. Chem. Soc.* **2011,** *133*, 16790-16793.
(51) Forbes, M. W.; Bush, M. F.; Polfer, N. C.; Oomens, J.; Dunbar, R. C.; Williams, E. R.; Jockusch, R. A. Infrared spectroscopy of arginine cation complexes: Direct observation of gas-phase zwitterions. *J. Phys. Chem. A* **2007,** *111*, 11759-11770.
(52) Ogorzalek Loo, R. R.; Loo, J. A. Salt bridge rearrangement (sabre) explains the dissociation behavior of noncovalent complexes. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 975-990.

(53) Zhang, Z.; Browne, S. J.; Vachet, R. W. Exploring salt bridge structures of gas-phase protein ions using multiple stages of electron transfer and collision induced dissociation. *J. Am. Soc. Mass Spectrom.* **2014,** *25*, 604-613.
(54) Breuker, K.; Brüschweiler, S.; Tollinger, M. Electrostatic stabilization of a native protein structure in the gas phase. *Angew. Chem. Int. Ed.* **2011,** *50*, 873-877.
(55) Bonner, J. G.; Lyon, Y. A.; Nellessen, C.; Julian, R. R. Photoelectron transfer dissociation reveals surprising favorability of zwitterionic states in large gaseous peptides and proteins. *J. Am. Chem. Soc.* **2017,** *139*, 10286-10293.
(56) Marchese, R.; Grandori, R.; Carloni, P.; Raugei, S. On the zwitterionic nature of gas-phase peptides and protein ions. *PLoS Comput. Biol.* **2010,** *6*.
(57) 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.
(58) Fegan, S. K.; Thachuk, M. A charge moving algorithm for molecular dynamics simulations of gas-phase proteins. *J. Chem. Theory Comput.* **2013,** *9*, 2531-2539.
(59) Shelimov, K. B.; Jarrold, M. F. Conformations, unfolding, and refolding of apomyoglobin in vacuum: An activation barrier for gas-phase protein folding. *J. Am. Chem. Soc.* **1997,** *119*, 2987-2994.
(60) Patriksson, A.; Adams, C. M.; Kjeldsen, F.; Zubarev, R. A.; 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.
(61) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc.* **1996,** *118*, 8365-8374.
(62) Boyd, R. K.; Somogyi, Á. The mobile proton hypothesis in fragmentation of protonated peptides: A perspective. *J. Am. Soc. Mass Spectrom.* **2010,** *21*, 1275-1278.
(63) Li, J. Y.; Lyu, W. P.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.; Sobott, F.; Grandori, R.; Carloni, P. Proton dynamics in protein mass spectrometry. *J. Phys. Chem. Lett.* **2017,** *8*, 1105-1112.
(64) Cautereels, J.; Blockhuys, F. Quantum chemical mass spectrometry: Verification and extension of the mobile proton model for histidine. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 1227-1235.
(65) Hewitt, D.; Marklund, E.; Scott, D. J.; Robinson, C. V.; Borysik, A. J. A hydrodynamic comparison of solution and gas phase proteins and their complexes. *J. Phys. Chem. B* **2014,** *118*, 8489-8495.
(66) Carugo, O.; Blatova, O. A.; Medrish, E. O.; Blatov, V. A.; Proserpio, D. M. Packing topology in crystals of proteins and small molecules: A comparison. *Sci Rep* **2017,** *7*.
(67) Kowalski, J. A.; Liu, K.; Kelly, J. W. Nmr solution structure of the isolated apo pin1 ww domain: Comparison to the x-ray crystal structures of pin1. *Biopolymers* **2002,** *63*, 111-121.
(68) Vetter, S. W.; Leclerc, E. Novel aspects of calmodulin target recognition and activation. *Eur. J. Biochem.* **2003,** *270*, 404-414.
(69) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. Characterization of the fast dynamics of protein amino acid side chains using nmr relaxation in solution. *Chem. Rev.* **2006,** *106*, 1672-1699.
(70) Carugo, O.; Argos, P. Protein-protein crystal-packing contacts. *Protein Sci.* **1997,** *6*, 2261-2263.
(71) Luo, J.; Liu, Z.; Guo, Y.; Li, M. A structural dissection of large protein-protein crystal packing contacts. *Sci Rep* **2015,** *5*.
(72) Billeter, M.; Wagner, G.; Wuthrich, K. Solution nmr structure determination of proteins revisited. *J. Biomol. NMR* **2008,** *42*, 155-158.

(73) Sun, Y.; Vahidi, S.; Sowole, M. A.; Konermann, L. Protein structural studies by traveling wave ion mobility spectrometry: A critical look at electrospray sources and calibration issues. *J. Am. Soc. Mass Spectrom.* **2016,** *27*, 31-40.
(74) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1–2*, 19-25.
(75) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of ubiquitin refined at 1.8 a resolution. *J. Mol. Biol.* **1987,** *194*, 531-544.
(76) Artymiuk, P. J.; Blake, C. C. F.; Rice, D. W.; Wilson, K. S. The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 angstroms resolution. *Acta Crystallogr. Sect. B-Struct. Commun.* **1982,** *38*, 778-783.
(77) Huang, J.; MacKerell, A. D. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *J. Comput. Chem.* **2013,** *34*, 2135-2145.
(78) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110*, 5915-5920.
(79) McAllister, R. G.; Konermann, L. Challenges in the interpretation of protein h/d exchange data: A molecular dynamics simulation perspective. *Biochemistry* **2015,** *54*, 2683-2692.
(80) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. 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.
(81) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996,** *118*, 11225-11236.
(82) Fegan, S. K.; Thachuk, M. Suitability of the martini force field for use with gas-phase protein complexes. *J. Chem. Theory Comput.* **2012,** *8*, 1304-1313.
(83) Marchese, R.; Grandori, R.; Carloni, R.; Raugei, S. A computational model for protein ionization by electrospray based on gas-phase basicity. *J. Am. Soc. Mass Spectrom.* **2012,** *23*, 1903-1910.
(84) Cadenas, E.; Boveris, A.; Chance, B. Low-level chemiluminescence of hydroperoxide-supplemented cytochrome c. *Biochem. J.* **1980,** *187*, 131-140.
(85) Moser, A.; Range, K.; York, D. M. Accurate proton affinity and gas-phase basicity values for molecules important in biocatalysis. *J. Phys. Chem. B* **2010,** *114*, 13911-13921.
(86) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. Collidoscope: An improved tool for computing collisional cross-sections with the trajectory method. *J. Am. Soc. Mass Spectrom.* **2017,** *28*, 587-596.
(87) Allen, S. J.; Schwartz, A. M.; Bush, M. F. 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.
(88) Kumar, S.; Nussinov, R. Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **1999,** *293*, 1241-1255.
(89) Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Bitan, G.; Teplow, D. B.; Bowers, M. T.; Shea, J. E. Amyloid β-protein monomer structure: A computational and experimental study. *Protein Sci.* **2006,** *15*, 420-428.
(90) Pagel, K.; Hyung, S.-J.; Ruotolo, B. T.; Robinson, C. V. Alternate dissociation pathways identified in charge-reduced protein complex ions. *Anal. Chem.* **2010,** *82*, 5363-5372.

(91) Zhou, M.; Dagan, S.; Wysocki, V. H. 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.
(92) Chang, T. M.; Prell, J. S.; Warrick, E. R.; Williams, E. R. Where's the charge? Protonation sites in gaseous ions change with hydration. *J. Am. Chem. Soc.* **2012,** *134*, 15805-15813.
(93) Wolynes, P. G. Biomolecular folding in vacuo!!!(?). *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92*, 2426-2427.

## Figure Captions

**Figure 1.** Native ESI-IMS/MS data. Ubiquitin in (A, B) positive ion mode and (C, D), negative ion mode; (E, F) lysozyme in positive ion mode and (G, H) negative ion mode.

**Figure 2.** (A, B) Ubiquitin and lysozyme crystal structures, highlighting one chain (green) that is surrounded by others (gray). (C, D) Crystal salt bridges of $Lys^+/Arg^+$ (blue) and $Asp^-/Glu^-/CT^-$ (red) with side chains from adjacent proteins (Lys/Arg: cyan; Asp/Glu: magenta). (E, F) Crystal backbone structures (green) overlaid with three 1 μs solution MD structures (gray). (G, H) Titratable side chains in the crystal (green) and from three 1 μs solution MD runs (Lys/Arg: blue; Asp/Glu: red). Not all side chains are shown to reduce clutter. (I, J) $R_g$ for atoms in titratable side chains, for atoms in titratable and polar side chains, and for the whole protein during 1 μs solution MD runs. Horizontal lines represent crystal $R_g$ values.

**Figure 3.** Whole protein $R_g$ values obtained in 1 μs ubiquitin gas phase MD simulations (A-C, 6+ charge state; D-F, 5- charge state). The panels contain data from 4-14 independent runs. Data in the top row were obtained using stationary positive-only (A) and negative-only (D) charges. The center row data were generated using various stationary zwitterionic charge patterns (B, E). Panels C and F show the results of mobile-proton MD runs.

**Figure 4.** Gas phase mobile-proton MD results on ubiquitin (A-C) and lysozyme (D, E). Each panel contains the X-ray structure (green), and MD snapshots representing $t$ = 100 ns, 200 ns, … 1 μs. The magenta panel in (B) illustrates transient unfolding of $[\text{ubiquitin} + 6H]^{6+}$ at $t$ = 100 ns (blue trajectory in Figure 3C).

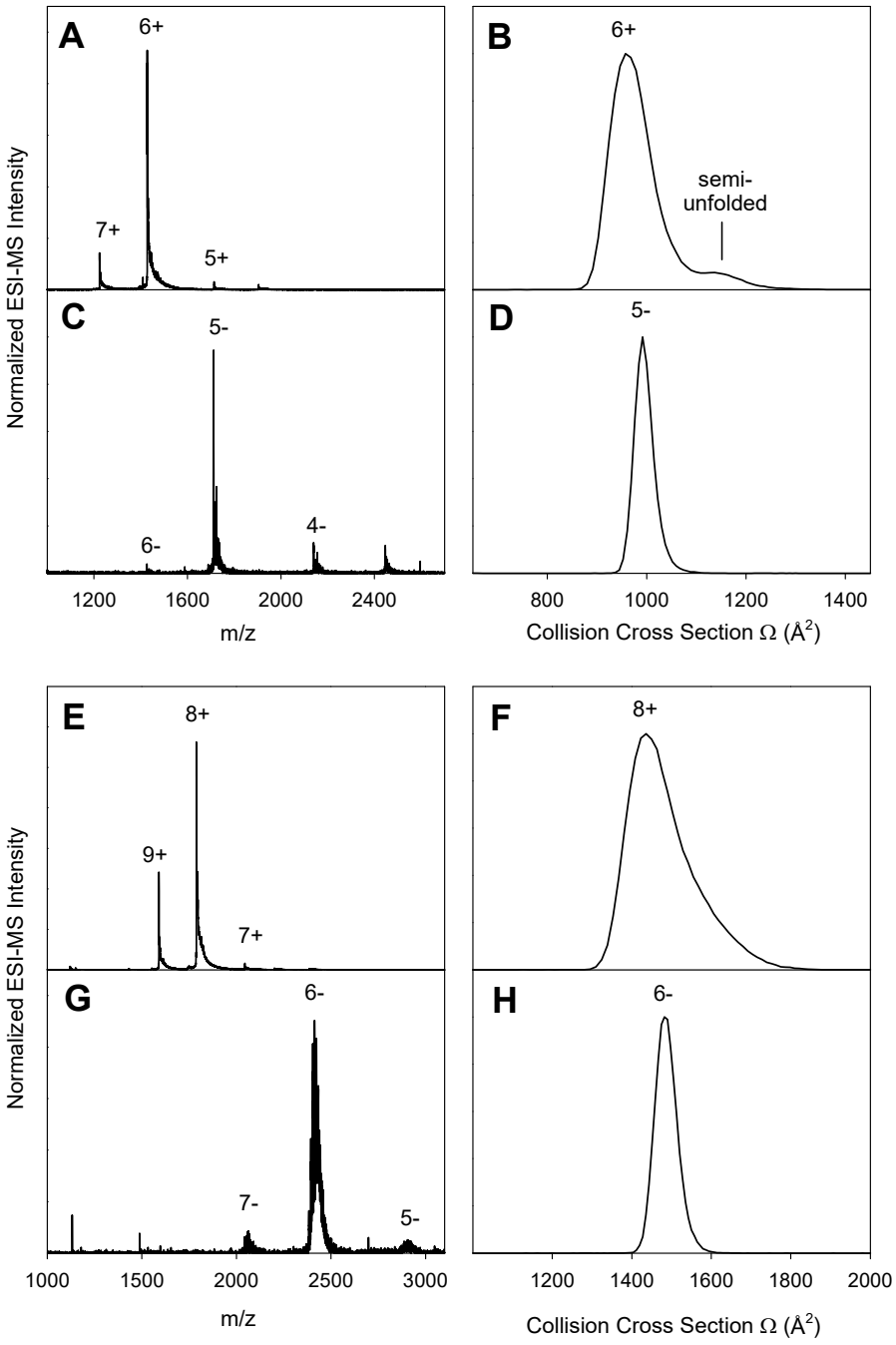
**Figure 5.** Gas phase mobile-proton MD data for ubiquitin (A, B) and lysozyme (C, D) in positive and negative ion mode. Each panel shows the accumulation of "opposite" charges, i.e., the formation of salt bridges. Zwitterionic structures are exemplified along the right hand side, using overlays of three representative 1 μs frames for each condition. Positive and negative charges are highlighted as blue and red spheres, respectively. Labels highlight sites that are consistently charged for all three 1 μs structures. For a detailed analysis of the zwitterionic patterns, see Figure S8.

**Figure 6.** Representative gas phase MD results obtained for ubiquitin in positive (A-C) and negative ion mode (D-F) with different protonation modes. Each panel shows the radius of gyration ($R_g$) for titratable side chain atoms, for titratable and polar side chain atoms, and for the whole protein. Horizontal lines represent the corresponding crystal $R_g$ values.

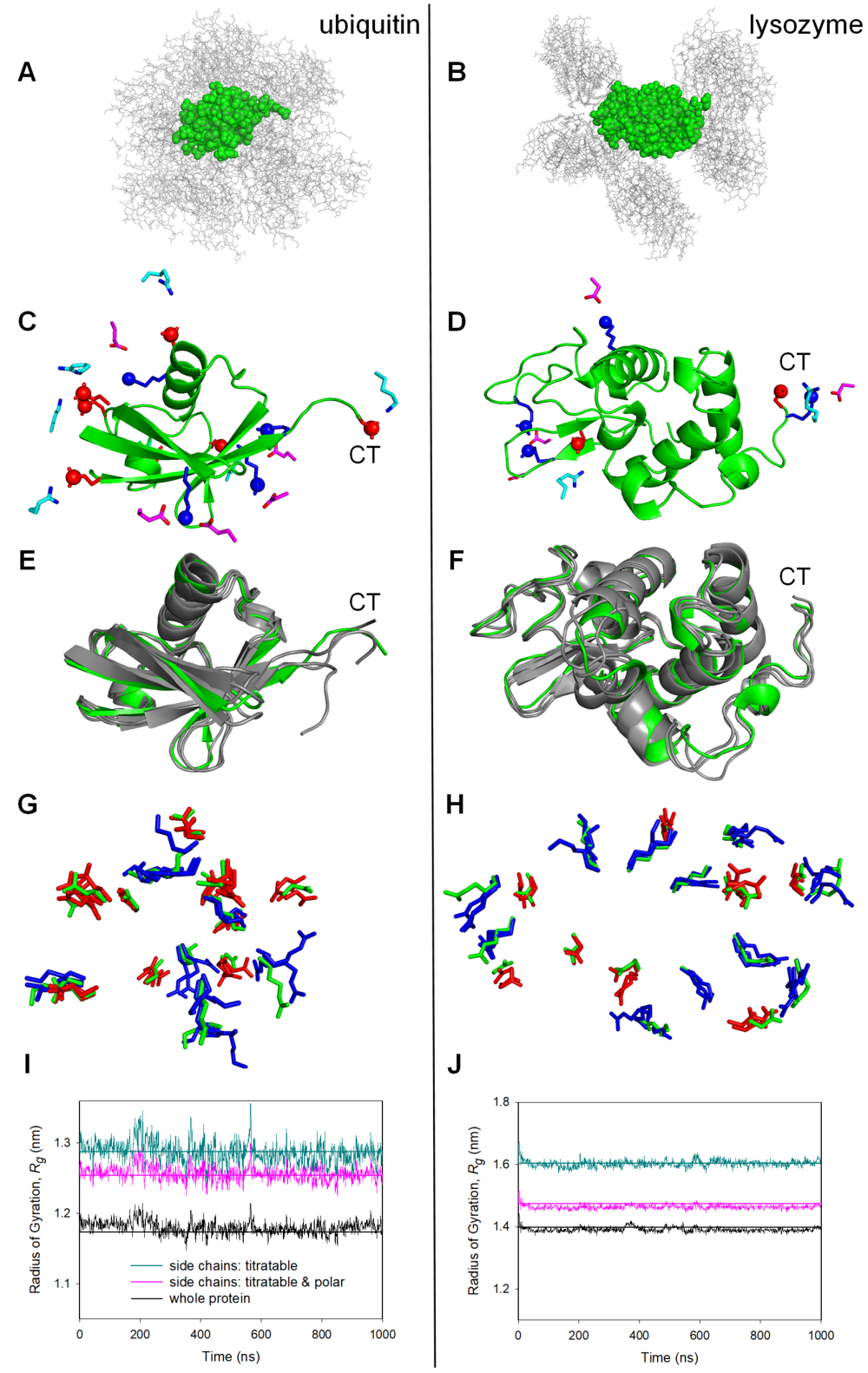
**Figure 7.** Packing of titratable/polar side chains in $t$ = 1 μs mobile-proton MD structures of ubiquitin in (A) positive and (B) negative ion mode. Cationic and anionic sites are highlighted as blue and red spheres. Selected N and O atoms are indicated as blue and red sticks, respectively. bbCO and bbNH denote backbone CO and NH groups. Not all titratable/polar moieties are shown to reduce cluttering.

**Figure 8.** Collision cross sections (Ω) from experiments and MD simulations. Vertical bars represent gas phase and solution MD Ω values. The first two panels separately display data for (A) folded and (B) semi-unfolded [ubiquitin + 6H]$^{6+}$. Colored horizontal lines represent experimental average Ω maxima, dashed lines indicate the corresponding FWHMs. Horizontal black lines represent crystal Ω values.

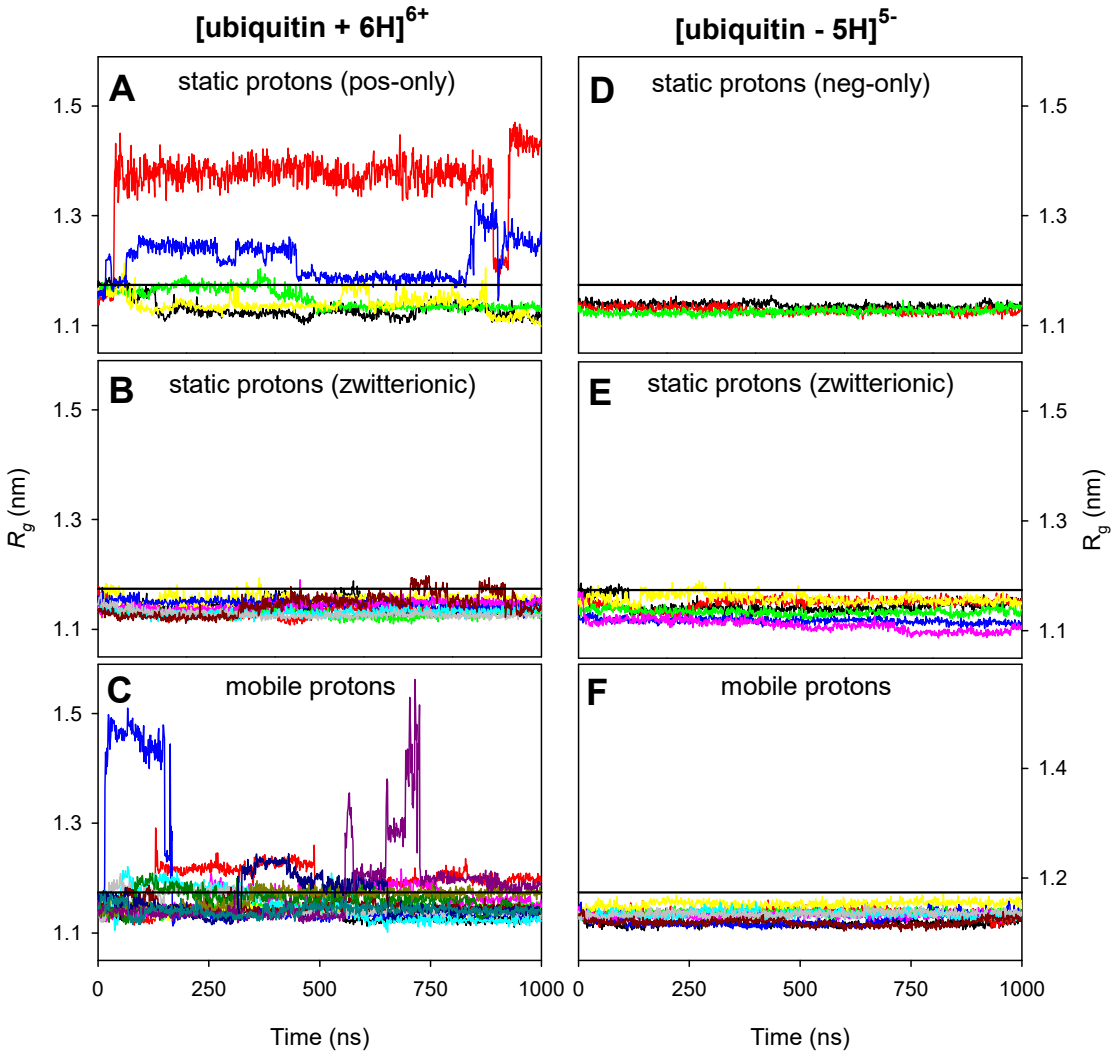
# Figure 1


A
6+
7+
5+
B
6+
semi-
unfolded
C
5-
6-
4-
D
5-
Normalized ESI-MS Intensity
1200
1600
2000
2400
m/z
800
1000
1200
1400
Collision Cross Section Ω (Å²)
E
8+
9+
7+
F
8+
G
6-
7-
5-
H
6-
Normalized ESI-MS Intensity
1000
1500
2000
2500
3000
m/z
1200
1400
1600
1800
2000
Collision Cross Section Ω (Å²)

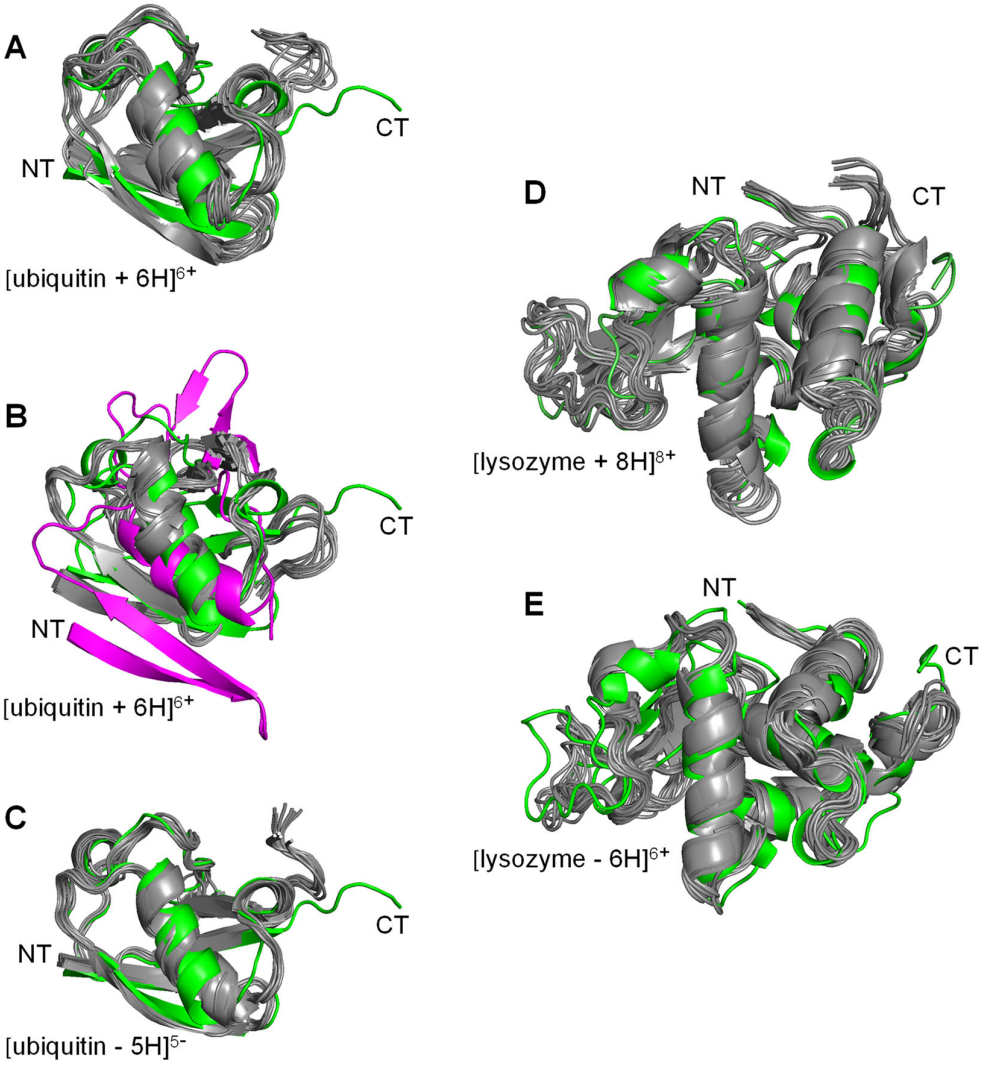
# Figure 2


ubiquitin
lysozyme
A
B
C
D
E
F
G
H
I
J
CT
CT
CT
CT
Radius of Gyration, $R_g$ (nm)
1.3
1.2
1.1
0
200
400
600
800
1000
Time (ns)
side chains: titratable
side chains: titratable & polar
whole protein
Radius of Gyration, $R_g$ (nm)
1.8
1.6
1.4
1.2
0
200
400
600
800
1000
Time (ns)

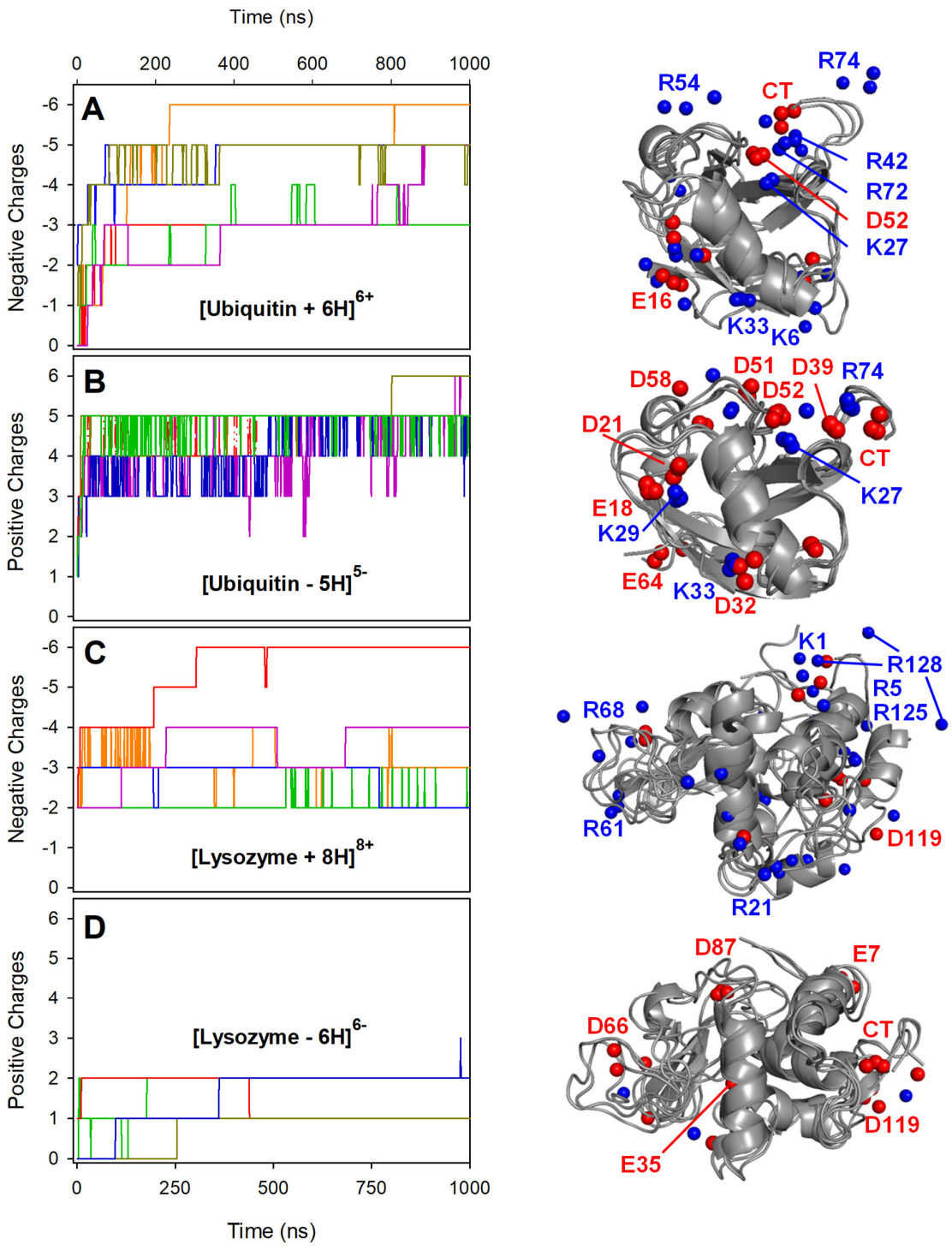
# Figure 3


[ubiquitin + 6H]6+
[ubiquitin - 5H]5-
A
static protons (pos-only)
D
static protons (neg-only)
B
static protons (zwitterionic)
E
static protons (zwitterionic)
C
mobile protons
F
mobile protons
Rg (nm)
Rg (nm)
Time (ns)
Time (ns)
1.1
1.3
1.5
1.2
1.4
0
250
500
750
1000

# Figure 4


A
NT
CT
[ubiquitin + 6H]6+
B
NT
CT
[ubiquitin + 6H]6+
C
NT
CT
[ubiquitin - 5H]5-
D
NT
CT
[lysozyme + 8H]8+
E
NT
CT
[lysozyme - 6H]6+

# Figure 5


Time (ns)
0
200
400
600
800
1000
A
Negative Charges
[Ubiquitin + 6H]6+
B
Positive Charges
[Ubiquitin - 5H]5-
C
Negative Charges
[Lysozyme + 8H]8+
D
Positive Charges
[Lysozyme - 6H]6-
0
250
500
750
1000
Time (ns)
R74
R54
CT
R42
R72
D52
K27
E16
K33
K6
D58
D51
D39
R74
D52
D21
CT
K27
E18
K29
E64
K33
D32
K1
R128
R5
R125
R68
R61
D119
R21
D87
E7
D66
CT
D119
E35

# Figure 6

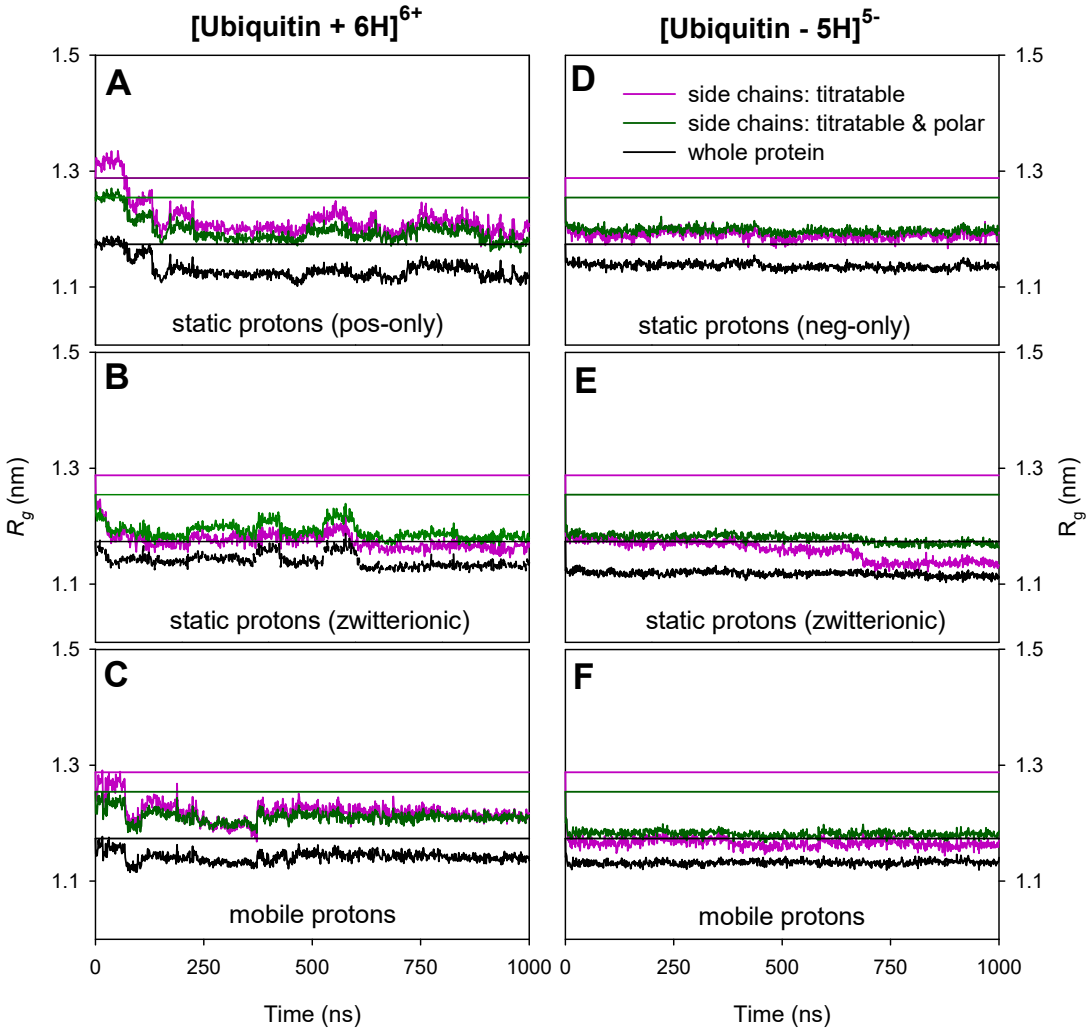
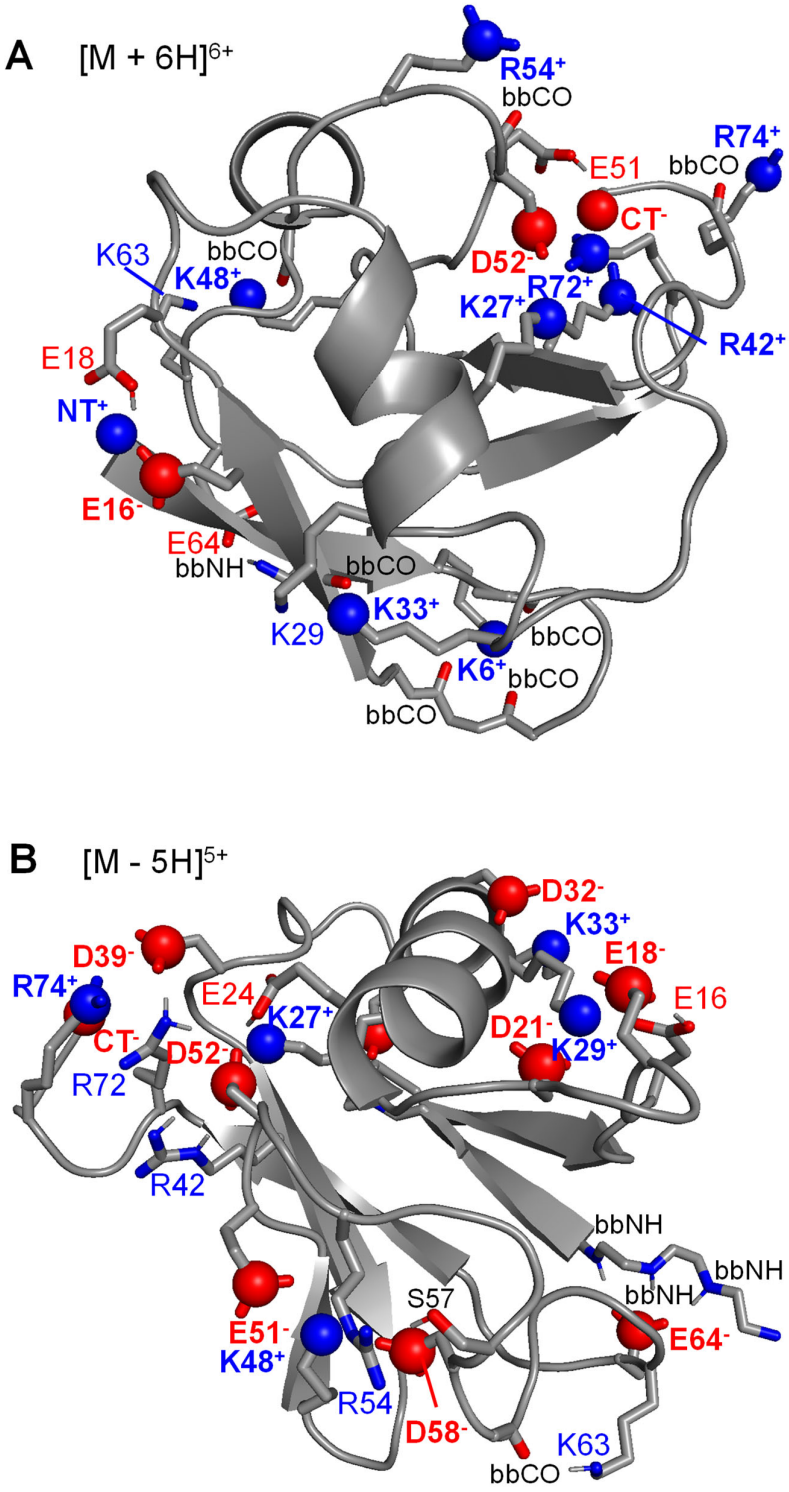
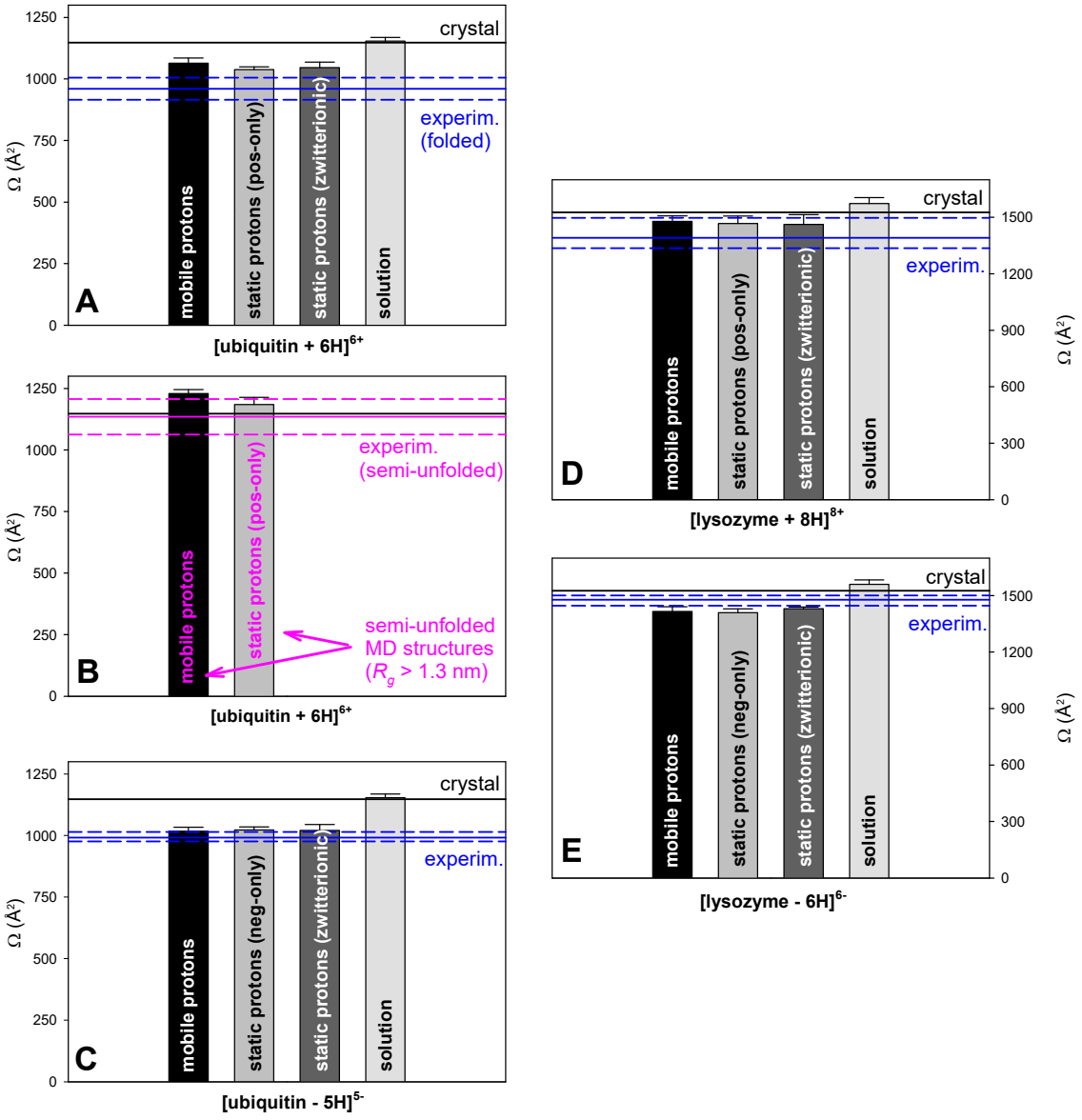

[Ubiquitin + 6H]6+
[Ubiquitin - 5H]5-
A
B
C
D
E
F
side chains: titratable
side chains: titratable & polar
whole protein
static protons (pos-only)
static protons (neg-only)
static protons (zwitterionic)
static protons (zwitterionic)
mobile protons
mobile protons
Rg (nm)
Rg (nm)
1.5
1.3
1.1
0
250
500
750
1000
Time (ns)
Time (ns)

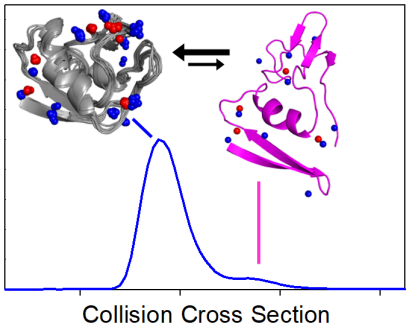
**Figure 7**


A
[M + 6H]6+
R54+
bbCO
R74+
E51
bbCO
CT-
D52-
K63
bbCO
K48+
R72+
K27+
R42+
E18
NT+
E16-
E64
bbNH
bbCO
K33+
K29
bbCO
K6+
bbCO
bbCO
B
[M - 5H]5+
D32-
K33+
D39-
E18-
R74+
E24
K27+
E16
CT-
D52-
D21-
K29+
R72
R42
bbNH
bbNH
bbNH
S57
E51-
K48+
E64-
R54
D58-
K63
bbCO

# Figure 8


A
B
C
D
E
crystal
experim.
(folded)
experim.
(semi-unfolded)
semi-unfolded
MD structures
($R_g$ > 1.3 nm)
mobile protons
static protons (pos-only)
static protons (neg-only)
static protons (zwitterionic)
solution
Ω (Å²)
[ubiquitin + 6H]$^{6+}$
[ubiquitin - 5H]$^{5-}$
[lysozyme + 8H]$^{8+}$
[lysozyme - 6H]$^{6-}$

## TOC Graphic




[TOC Graphic was pasted into the .docx document as device-independent bitmap. If resizing is required, please drag the bottom right corner.]