8-21-2018

Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry.

Haidy Metwally

Quentin Duez

Lars Konermann

Chain Ejection Model for Electrospray Ionization of Unfolded Proteins: Evidence from Atomistic Simulations and Ion Mobility Spectrometry

Haidy Metwally[1], Quentin Duez[1,2], and Lars Konermann[1,*]

[1]*Department of Chemistry, The University of Western Ontario, London, Ontario, N6A 5B7, Canada.*

[2]*Organic Synthesis and Mass Spectrometry Laboratory, University of Mons, Place du Parc, 23, Mons, 7000, Belgium*

* corresponding author: konerman@uwo.ca

Funding was provided by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2018-04243). QD was recipient of travel stipends from the Belgian Fonds National de la Recherche Scientifique, Université de Mons, and the Académie des Sciences and Faculté Wallonie-Bruxelles.

ABSTRACT: The ion evaporation model (IEM) and the charged residue model (CRM) represent cornerstones of any discussion related to the mechanism of electrospray ionization (ESI). Molecular dynamics (MD) simulations have confirmed that small ions such as Na^+ get ejected from the surface of aqueous ESI droplets (IEM), while folded proteins in native ESI are released by water evaporation to dryness (CRM). ESI of unfolded proteins yields $[M + zH]^{z+}$ ions that are much more highly charged than their folded counterparts. A chain ejection model (CEM) has been proposed to account for the protein ESI behavior under such non-native conditions (*Anal. Chem*. **2013**, *85*, 2-9). The CEM envisions that unfolded proteins are driven to the droplet surface by hydrophobic and electrostatic factors, followed by gradual ejection via intermediates where droplets carry extended protein tails. Thus far it has not been possible to support the CEM through MD simulations using realistic protein models and atomistic force fields. Such endeavors require much larger droplets than in previous MD studies. Also, the incorporation of CEM-related H^+ migration is difficult. The current work overcomes these challenges in MD simulations on unfolded apo-myoglobin (aMb) in droplets with 5.5 nm radius (~22,500 water molecules). We focused on solutions at pH ~4 where the aMb solution charge coincides with the charge on some of the electrosprayed ions (22+ to 27+), such that H^+ migration could be neglected. Na^+ ions were added to ensure a droplet charge close to the Rayleigh limit. We found that 16/17 MD runs on various protonation patterns produced $[M + zH]^{z+}$ ions via chain ejection. The predicted stretched-out aMb conformations were consistent with experimental collision cross sections. These results support the view that unfolded proteins follow the CEM. Overall, the IEM/CRM/CEM triad can account for a wide range of ESI scenarios involving various types of analytes.

Electrospray ionization (ESI) has revolutionized mass spectrometry (MS) by allowing the facile transfer of proteins and other analytes from solution into the gas phase.[1] During ESI charged droplets emanate from a high voltage capillary. Solvent evaporation and jet fission decrease the droplet size to the nanometer range, while maintaining a charge close to the Rayleigh limit.[2-7] The mechanisms whereby analyte ions emerge from ESI nanodroplets remain controversial.[4, 8-18] Recent progress in this area has been fueled by molecular dynamics (MD) simulations. MD studies revealed that small ions such as Na^+ undergo field emission from the droplet surface,[15, 19-21] consistent with the ion evaporation model (IEM).[8, 22] Simulations on peptides,[23] globular proteins,[24] and nucleic acid duplexes[25] indicated that these larger species are liberated by droplet evaporation to dryness, as envisioned by the charged residue model (CRM).[3, 4]

"Native" protein ESI experiments aim to preserve solution-like structures and interactions in the gas phase by employing non-denaturing aqueous solutions and gentle ion sampling conditions.[26-29] Structural retention under these conditions is promoted by the low CRM charge states of protein ions, which are close to the Rayleigh charge of protein-sized water droplets.[3, 4]

Electrosprayed protein ions can be further interrogated by ion mobility spectrometry (IMS)[30-35] and by various activation methods.[36-38] Collision-induced dissociation (CID) of multisubunit complexes usually causes ejection of one highly charged chain.[39] This behavior has been attributed to gradual unfolding of one subunit, H^+ migration onto the unraveling chain, and subsequent separation of this chain from the complex (Figure 1a).[38, 40, 41] This CID model is consistent with IMS data,[32, 37] the mobile nature of H^+ in gaseous proteins,[42-45] and the fact that H^+ migration onto the unraveling chain will minimize electrostatic repulsion.[46, 47]

While the CRM is widely accepted for globular proteins in native ESI,[3, 4, 11, 24, 48] the behavior of unfolded proteins is more controversial.[12, 48, 49] $[M + zH]^{z+}$ ions formed from unfolded

proteins exhibit wide charge state distributions centered at much higher *z* values than in native ESI. This effect is encountered after unfolding by acid,[50, 51] base,[52] heat,[53] disulfide cleavage,[54] mutations,[55] and cofactor removal.[51] The high charge states of unfolded proteins can boost mass analyzer performance[56, 57] and enhance top-down fragmentation.[58, 59]

Various attempts have been made to explain the dramatic shift to higher charge states seen for unfolded proteins. Early work proposed that ESI charge states reflect the titration behavior in solution,[60] but subsequent studies showed this not to be the case.[52, 61, 62] Other ideas focused on the accessibility of titratable sites,[50, 54, 63] but even in native proteins most titratable sites are accessible at the surface.[64] It has also been proposed that COO^- groups may neutralize positive sites in folded gaseous proteins.[12] Although such zwitterionic contacts are well documented,[65] it is unclear if they can account for conformation-dependent charge state changes.[48] Other studies pointed out that the gas-phase basicity of biomolecular ions could result in H^+ transfer from the solvent vapor.[12, 66] While all these ideas are interesting, they do not directly address the mechanism whereby unfolded proteins emerge from ESI nanodroplets.

Unfolded proteins in solution adopt disordered conformations, similar to certain synthetic polymers such as polyethylene glycol (PEG).[67] MD simulations of PEG-containing aqueous ESI droplets by Consta et al. revealed that PEG binds Na^+ from the solvent, followed by polymer extrusion from the droplet surface.[68] Our laboratory proposed that the ESI process for unfolded proteins follows similar avenues.[15] According to this "chain ejection model" (CEM, Figure 1b), the protein is driven to the droplet surface by electrostatic and hydrophobic factors. The protein then undergoes gradual ejection via "tadpole-like" structures where the droplet carries an extended protein tail. This CEM scenario[15] bears close parallels to the CID process of multi-subunit systems (Figure 1). Specifically, a central aspect of the protein CEM is the migration of mobile H^+ between the droplet and the protruding polypeptide tail. This H^+ transfer causes the unfolded protein to

depart as a highly charged ion, analogous to H^+ transfer during CID which causes the departing subunit to be highly charged.[15] H^+ migration is absent for sodiated PEG,[68] i.e., the CEM scenarios of unfolded proteins and PEG chains are not equivalent.

Several studies have endorsed the CEM.[15, 26, 69, 70] However, it is unsettling that protein CEM processes have never been supported by atomistic MD simulations. Instead, the idea relies on simple coarse-grained models, Monte-Carlo methods,[15] and salt adduction studies.[71] Atomistic ESI simulations on unfolded proteins face two challenges. (1) Mobile H^+ are difficult to treat computationally. MD methods are available for H^+ transfer in the gas phase,[46, 47] but H^+ migration in water (and between water and protein, Figure 1b) requires *ab initio* tools. The computational cost of these tools makes them unsuitable for ESI droplets.[72-74] (2) ESI droplets have to accommodate the analyte at the onset of the simulation. This is not a problem for native proteins, where a few thousand water molecules are sufficient for building a droplet that completely engulfs the analyte.[24, 25] In contrast, the increased dimensions of unfolded proteins[67] require larger droplets, driving up computational cost which scales as N^2 with the number of atoms.[75]

The current work scrutinizes the viability of the CEM by conducting atomistic MD simulations on ESI droplets containing unfolded proteins. Myoglobin has been used in numerous earlier ESI mechanistic investigations,[24, 30, 50, 51, 63, 69, 71] and therefore it was chosen as model protein for our work as well. Difficulties associated with the treatment of mobile H^+ were circumvented by focusing on carefully selected pH environments. Graphics processing unit (GPU)-accelerated algorithms[76] allowed us to overcome challenges related to droplet size. MD simulations were complemented by ESI-MS/IMS experiments. Our results support the view that the ESI process of unfolded proteins in aqueous solution proceeds via the CEM.

Materials and Methods

Apo-myoglobin (aMb, Sigma, St. Louis, MO) was prepared by butanone extraction. The protein was dialyzed against 10 mM neutral aqueous ammonium acetate, followed by acidification with formic acid to pH 4 or pH 2. ESI-MS and IMS data were acquired on a Synapt G2 instrument (Waters, Milford, MA). Aqueous aMb (5 μM) was infused at 5 μL min^{-1} using a standard Z-spray ion source at +2.8 kV. The source and desolvation temperatures were kept low (25 °C and 40 °C) and the cone was set to 5 V to minimize the in-source activation. IMS calibration using a set of reference proteins yielded effective He collision cross sections (Ω).[77]

MD simulations were carried out using Gromacs 2016 with GPU acceleration.[76] The Charmm36 force field[78] was chosen for the protein due to its excellent performance in previous solution-phase folding studies[79] and droplet simulations.[24] TIP4P/2005 water[80] was used because it properly reflects the water surface tension over a wide range of temperatures.[81] Unfolded protein structures were initially produced by heating aMb (1WLA without heme) from 320 K to 450 K in vacuum using canonical charge states over 20 ns. Conformers generated toward the end of these runs served as starting points for droplet simulations. Spherical water droplets with 5.5 nm radius (~22,500 water molecules) were built around the aMb chains, and the protein charge was set to 22+, 27+, or 33+ (see details below). All runs started with a droplet charge of 47+ which corresponds to the Rayleigh charge z_R of a 5.5 nm aqueous droplet,[3, 4, 24] calculated as $z_R = 8\pi/e \times (\varepsilon_0 \gamma r^3)^{1/2}$. To attain this regime, charges contributed by aMb were supplemented by Na^+ ions in random positions. The droplet simulations employed a pseudo-PBC approach[82] that is equivalent to vacuum boundary conditions[75] but provides much shorter Gromacs run times. The ESI simulation temperature was set to 370 K, reflecting the presence of heating elements in typical ion sources.[83] All runs employed trajectory stitching mode.[24] Under this scheme the system was

coupled to a Nosé-Hoover thermostat,[84] and evaporated water or Na^+ that had moved more than 100 nm from the center of mass were removed in 250 ps intervals.[24] After each of these segments new velocities were sampled from a Maxwell-Boltzmann distribution prior to starting the subsequent simulation window. Droplet evaporation in our simulations is a quasi-irreversible process that produces desolvated analyte ions, consistent with typical ESI source conditions.[83] We do not consider subsequent free jet expansion events that may cause the re-adsorption of solvent molecules for certain analytes.[85] All runs were repeated three to five times with different initial aMb structures, Na^+ positions, and starting velocities. The simulation time window was 75 ns. After release into the gas phase the desolvated protein was allowed to run for an additional 500 ns at 320 K; this lower temperature was chosen to reflect gentle ion sampling conditions.[26-28, 86] He collision cross sections were calculated using the trajectory method in Collidoscope.[87] These Ω values were determined by extracting MD structures from the 500 ns trajectories in 100 ns intervals.

Results and Discussion

ESI-MS and IMS Experiments. ESI mass spectra of aMb acquired at pH 4 showed a bimodal charge state distribution peaking at 9+ and 18+ (Figure 2a). The protein is known to be structurally heterogeneous in solution at pH 4, comprising compact conformers and disordered species.[51, 88] This heterogeneity is reflected in the spectrum of Figure 2a, where compact conformers gave rise to lower charge states (around 9+), while the more unfolded chains formed charge states around 18+.[51] Acidification to pH 2 caused further unfolding,[88] consistent with a shift to higher ESI charge states (around 20+, Figure 2b).[51] The highest detectable charge state was 27+.

IMS data for charge states 22+ to 27+ are shown in Figure 2c (the complete dataset is shown in Figure S1). 22+ ions had collision cross sections of (3870 ± 30) $Å^2$, whereas higher charge states showed larger Ω values, e.g., (4190 ± 100) $Å^2$ for 27+. This trend reflects the internal Coulomb repulsion experienced by the gaseous ions.[30, 31] For the highly charged ions considered in Figure 2c, collision cross sections measured at pH 4 and pH 2 were virtually indistinguishable.

ESI Modeling Strategy. We already noted how difficult it is to model H^+ migration in a droplet that carries a polypeptide tail (as envisioned within the CEM, Figure 1b).[15] Such H^+ transfer can cause the gas phase protein charge to be very different from that in solution.[52, 61, 62] However, the magnitude of this disparity depends on the conditions. The solution charge is dictated by the pK_a values of titratable sites and by pH.[64] Hence, a judicious choice of pH can result in a scenario where the protein solution charge resembles the charge state of the resulting gaseous ions. Figure 3 shows that at pH ~4 the aMb solution charge is around 22+ (pH 4.25) to 27+ (pH 3.75). Gas phase ions with charge states 22+ to 27+ can be produced by electrospraying aMb at pH 4 (Figure 2a). CEM-related H^+ migration between droplet and protein will be minimal under these conditions because solution charge ≈ gas phase charge. In this specific case it is reasonable to model the ESI process *without* mobile H^+, thereby greatly simplifying the computational approach.

Prior to applying the strategy outlined above, one has to consider that the droplet pH may deviate from that of the bulk solution. Solvent evaporation tends to increase the H^+ concentration. On the other hand, the analyte solutions used here contained formic acid which acts as a buffer around pH 4 (the pK_a of formic acid is 3.75). Therefore, the assumptions (1) solution pH ≈ droplet pH and (2) [protein charge in solution] ≈ [protein charge after ESI] provide a reasonable foundation for modeling the ESI behavior of unfolded aMb in the 22+ to 27+ charge states at pH 4.

Another aspect that has to be addressed is the intramolecular H^+ distribution. Most charge states can be implemented via a multitude of protonation patterns.[12] To test whether the aMb behavior is sensitive to this aspect we performed simulations on different patterns (Figure S2). Three 22+ protonation patterns were tested, referred to as 22+[A] (all sites protonated, except His), 22+[B] (N-terminus and all Arg/His/Lys protonated, all Asp and some Glu deprotonated), 22+[C] (all sites protonated, except for some Lys). 27+ simulations were conducted where all sites were protonated, except for some Lys. Simulations on fully protonated aMb (33+) were included as well, although this value is beyond the range observed in our experiments.

To ensure that unfolded aMb chains were fully contained within the initial ESI droplet we employed a droplet radius of 5.5 nm. To the best of our knowledge, these are the largest protein-containing ESI droplets modeled to date. GPU-acceleration[76] helped overcome the computational cost associated with this system size. TIP4P/2005 water[80] was chosen because it reproduces the water surface tension, thus ensuring that the simulations yielded realistic data.[24] The droplets were initially charged to the Rayleigh limit[2-4] by supplementing the aMb charge with Na^+ ions, keeping in mind that Na^+ represents a typical ESI charge carrier.[3] Other ions such as NH_4^+ or H_3O^+ would likely yield qualitatively similar results to those discussed below for Na^+.

MD Simulations Confirm CEM Behavior. Aqueous ESI droplets containing a single unfolded aMb chain were subjected to MD simulations. Typical data for protonation patterns 22+[B] and 27+ are shown in Figure 4. Within a few ns the protein migrated from the droplet interior close to the surface (Figure 4a, 5 ns; Figure 4b, 2.5 ns). This was followed by partial aMb eruption as a hydrated bulge (Figure 4a, 7 ns). At this particular point Na^+ and aMb^{22+} had partitioned into opposite regions of the droplet, highlighting the electrostatic forces within the system (see also Figure S3). Protein expulsion subsequently produced an electrostatically stretched tail that

protruded into the vapor phase (Figure 4a, 10 ns; Figure 4b, 12.5 ns). Further expulsion then caused aMb separation from the droplet. Nascent gaseous proteins retained some water which evaporated within 75 ns. Data very similar to those of Figure 4 were also seen for the other protonation patterns (22+[A], 22+[C], 33+, Figure S4), confirming that the behavior reported here is robust and reproducible. Figure 4 embodies the central result of this work: for the first time atomistic MD simulations confirm the formation of gaseous ions from unfolded proteins via the CEM.

The CEM trajectories showed slight variations. In some instances, aMb ejected with a small droplet attached to one terminus. These small droplets then evaporated without separating from the chain (Figure S5). Another variation involved ejection in a hairpin conformation. The hairpins either opened up as they departed from the droplet (22+[A], Figure S6a), or they retained looped conformations (22+[B], Figure S6b). In one instance aMb ejected without extensive stretching (22+[B], Figure S6c). All other (16 out of 17) runs showed the hallmark of the CEM, i.e., aMb ejection via droplets that carried an electrostatically stretched protein tail, consistent with the mechanism of Figure 1b.[15] None of the unfolded aMb chains showed CRM behavior, i.e., protein release via droplet evaporation to dryness.

Charge Loss at the Rayleigh Limit. Our simulations started with a droplet charge z_D close to the Rayleigh limit z_R.[2-4] Droplet shrinkage due to water evaporation (Figure 5a) had a tendency to increase the electrostatic repulsion further, forcing the droplets to shed charge. Three competing charge loss processes were encountered. The first was the IEM ejection of Na^+ (Figure 4a). The second was CEM ejection of the protein. Multiple Na^+ ejections occurred early during each run, followed by long plateaus where Na^+ loss had come to a halt (Figure 5b). These plateaus represent

the regime where formation of a protruding protein tail had decreased the electrostatic repulsion within the droplet to such an extent that Na^+ ejection became kinetically unfavorable.

The third charge loss process was the formation of Na^+-containing progeny droplets via jet fission, a phenomenon well known from imaging studies on larger ESI droplets.[5, 6] Fission events reminiscent of those experimental data,[5, 6] involving water filaments, took place in several MD runs and generated progeny droplets comprising ~50 water molecules and one or two Na^+ (Figure S7). Similar to IEM ejection of Na^+, jet fission was limited to time points prior to formation of a protein tail. Plots of z_D / z_R confirmed that the droplets stayed close to the Rayleigh limit during these early stages (Figure 5c). We limited the time frame of Figure 5c to the initial ~4 ns during which protein ejection had not started yet in most runs, keeping in mind that the z_R expression used here[2-4] applies only to spherical systems.

All three of the aforementioned charge loss processes were kinetically viable, but ultimately protein ejection was the main avenue by which the droplets relieved electrostatic stress. Specifically, IEM and progeny droplets accounted for the loss of no more than ten Na^+, while protein ejection deprived the droplet of 22 to 33 charges (Figure 5b). In addition to its electrostatic driving force, CEM behavior is favored by the fact that unfolded proteins possess many solvent-exposed hydrophobic side chains. The tendency of hydrophobic moieties to migrate to the liquid/vapor interface, rather than stay in the droplet interior, is well established.[15, 18] The case is completely different for folded proteins, where most hydrophobic sites are buried, and where water interacts favorably with solvent-exposed hydrophilic/charged residues.[64] The latter conditions cause folded proteins to remain within the ESI droplets until evaporation to dryness, resulting in CRM behavior.[15] Solvent evaporation to dryness causes CRM-produced protein ions to carry adducts arising from nonvolatile solutes, including Na^+ and other charge carriers.[3, 15] In contrast, the CEM simulations of the current work did not reveal a single instance of Na^+

adduction to aMb. This behavior is in line with the experimental finding that acid-unfolded proteins are less prone to adduction than their natively folded counterparts.[71]

Protein Conformations after CEM Ejection. For each ESI simulation the behavior of desolvated aMb after ejection from the droplet was explored in 500 ns MD runs. Representative structures from these vacuum simulations were extracted, and their Ω values were compared with experimental IMS data (Figure 6). Averaging the results of all aMb 22+ simulations yielded Ω = (3790 ± 300) Å^2, in good agreement with the experimental result of (3870 ± 30) Å^2. Simulated 27+ ions had Ω = (4190 ± 100) Å^2, close to the measured value of (4050 ± 30) Å^2. The MD structures that most closely matched the experiments are shown along the right hand side of Figure 6. These electrostatically stretched proteins had a stick-like appearance, with local α-helices, and some coil formation at the termini. Electrostatic repulsion caused 27+ aMb to be slightly longer (overall length ~ 350 Å) than the 22+ ions (~ 310 Å).

Overall, the results of MD runs on protonation patterns 22+[A], 22+[C], and 27+ agreed well with the experimental data (Figure 6). The zwitterionic pattern 22+[B] deserves a closer look. Two 22+[B] runs produced relatively compact structures that were incompatible with experiments (Figure 6b, left). Non-local salt bridges in these runs limited the extent of electrostatic stretching. A third 22+[B] run did yield a stretched structure that matched the measured Ω value (Figure 6b). Salt bridges in this third run only involved sites in close sequence proximity. Our data thus do not exclude the existence of local zwitterionic motifs in unfolded gaseous proteins, but non-local salt bridges (involving side chains separated by tens of residues) are unlikely. We note that previous evidence for zwitterionic motifs in the gas phase was obtained primarily for tightly folded proteins,[12, 65] rather than unfolded species.

The gas phase conformations generated in our 500 ns vacuum simulations were governed by the morphology of the protein during ejection. Proteins that left the droplet as stretched chains retained extended conformations (Figures 3, 4a); proteins that ejected in more compact structures (Figure S6b, c) gave rise to less extended ions (Figure 6b, left). The near-absence of conformational changes in the 500 ns runs suggests that this conformational memory persists for time periods much longer than the simulation window explored here (overlays in Figure 6). The agreement between experimental data and MD-derived Ω values in Figure 6 thus supports the fidelity of our MD data, and it bolsters the view that the observed ions are CEM products.

Conclusions

This work marks the first time that the production of gaseous ions from unfolded proteins via the CEM has been corroborated in MD simulations using an atomistic force field for protein and solvent. The MD data are supported by IMS experiments, bolstering the view that unfolded proteins are ejected from ESI droplets as electrostatically stretched chains (Figure 1b).[15] This CEM behavior can be contrasted to the CRM, which is operational in native ESI where protein are folded.[24]

Computational challenges precluded the inclusion of H^+ migration between droplet and protein (Figure 1b) in our simulations. Depending on the titration behavior of the protein, H^+ migration may increase or decrease the protein net charge during ejection. Here we focused on aMb around pH 4, where $z \approx$ *solution charge*, such that H^+ migration could be neglected. How likely is it that more advanced modeling strategies involving mobile H^+ would overthrow the viability of the CEM? Our IMS experiments demonstrated that the gas phase conformations of aMb 22+ to 27+ produced at pH 4 were indistinguishable from those generated at pH 2. This is

despite the fact CEM-related H^+ migration at pH 4 is negligible ($z \approx$ *solution charge*), whereas pH 2 will be associated with significant H^+ transfer ($z <$ *solution charge*). Hence, H^+ migration does not seem to affect the properties of these ions. In addition, our MD simulations consistently produced CEM behavior for various protonation patterns and charge states, all the way to aMb^{33+}. It thus appears that CEM behavior is a robust property of acid-unfolded proteins. Based on the arguments presented, it seems unlikely that this conclusion would change when applying computational strategies that allow for H^+ migration.

Notwithstanding the aforementioned comments, it would be desirable for future studies on the ESI behavior of unfolded proteins to employ more sophisticated models that include H^+ migration as well as H^+ transfer to background gas molecules.[89] Such strategies will be required for quantitatively explaining the wide range of ESI charge states seen for unfolded proteins (Figure 2). We hypothesize[15] that much of this charge heterogeneity can be attributed to protein ejection from differently sized droplets, in conjunction with the fact that denatured proteins comprise various conformers that may accumulate different numbers of H^+ during ejection.[12, 51] It is possible that the CEM is not the only mechanism that is operative during ESI of denatured protein. Only the highly charged ions in the spectra were attributed to the CEM, while less abundant ions in low charge states (such as 9+/10+, Figure 2b) may represent CRM products. In future work it will also be interesting to explore the behavior of ESI-induced protein aggregation, an effect that may take place under conditions where droplets contain more than one protein molecule.[90] We hope that the current work will stimulate additional studies aimed at deciphering all these intricacies associated with the ESI process, as well as the possible involvement of the CEM in "supercharging" experiments.[69]

References

(1) Fenn, J. B. *Angew. Chem. Int. Ed.* **2003**, *42*, 3871-3894.
(2) Grimm, R. L.; Beauchamp, J. L. *J. Phys. Chem. A* **2010**, *114*, 1411-1419.
(3) Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28*, 898-917.
(4) de la Mora, F. J. *Anal. Chim. Acta* **2000**, *406*, 93-104.
(5) Gomez, A.; Tang, K. *Phys. Fluids* **1994**, *6*, 404-414.
(6) Nemes, P.; Marginean, I.; Vertes, A. *Anal. Chem.* **2007**, *79*, 3105-3116.
(7) Wang, R.; Zenobi, R. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 378-385.
(8) Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287-2294.
(9) Cole, R. B. *J. Mass. Spectrom.* **2000**, *35*, 763-772.
(10) Wang, G.; Cole, R. B. *Anal. Chim. Acta* **2000**, *406*, 53-65.
(11) Iavarone, A. T.; Williams, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 2319-2327.
(12) Li, J.; Santambrogio, C.; Brocca, S.; Rossetti, G.; Carloni, P.; Grandori, R. *Mass Spectrom. Rev.* **2016**, *35*, 111-122.
(13) Hogan, C. J.; Carroll, J. A.; Rohrs, H. W.; Biswas, P.; Gross, M. L. *Anal. Chem.* **2009**, *81*, 369-377.
(14) Breuker, K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18145-18152.
(15) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. *Anal. Chem.* **2013**, *85*, 2-9.
(16) Zhao, F. F.; Matt, S. M.; Bu, J.; Rehrauer, O. G.; Ben-Amotz, D.; McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2001-2010.
(17) Nshanian, M.; Lakshmanan, R.; Chen, H.; Ogorzalek Loo, R. R.; Loo, J. A. *Int. J. Mass Spectrom.* **2018**, *427*, 157–164.
(18) Cech, N. B.; Enke, C. G. *Mass Spectrom. Rev.* **2001**, *20*, 362-387.
(19) Higashi, H.; Tokumi, T.; Hogan, C. J.; Suda, H.; Seto, T.; Otani, Y. *Phys. Chem. Chem. Phys.* **2015**, *17*, 15746-15755.
(20) Znamenskiy, V.; Marginean, I.; Vertes, A. *J. Phys. Chem. A* **2003**, *107*, 7406-7412.
(21) Consta, S.; Mainer, K. R.; Novak, W. *J. Chem. Phys.* **2003**, *119*, 10125-10132.
(22) Loscertales, I. G.; de la Mora, J. F. *J. Chem. Phys.* **1995**, *103*, 5041-5060.
(23) Kim, D.; Wagner, N.; Wooding, K.; Clemmer, D. E.; Russell, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 2981-2988.
(24) McAllister, R. G.; Metwally, H.; Sun, Y.; Konermann, L. *J. Am. Chem. Soc.* **2015**, *137*, 12667-12676.
(25) Porrini, M.; Rosu, F.; Rabin, C.; Darre, L.; Gomez, H.; Orozco, M.; Gabelica, V. *ACS Central Sci.* **2017**, *3*, 454–461.
(26) Mehmood, S.; Allison, T. M.; Robinson, C. V. *Annu. Rev. Phys. Chem.* **2015**, *66*, 453-474.
(27) Leney, A. C.; Heck, A. J. R. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 5-13.
(28) Dyachenko, A.; Gruber, R.; Shimon, L.; Horovitz, A.; Sharon, M. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 7235-7239.
(29) Gavriilidou, A. F. M.; Holding, F. P.; Mayer, D.; Coyle, J. E.; Veprintse, D. B.; Zenobi, R. *Biochemistry* **2018**, *57*, 1685-1689.
(30) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2987-2994.
(31) Wyttenbach, T.; Pierson, N. A.; Clemmer, D. E.; Bowers, M. T. *Annu. Rev. Phys. Chem.* **2014**, *65*, 175–196.
(32) Ruotolo, B. T.; Hyung, S.-J.; Robinson, P. M.; Giles, K.; Bateman, R. H.; Robinson, C. V. *Angew. Chem. Int. Ed.* **2007**, *46*, 8001-8004.
(33) Jurneczko, E.; Barran, P. E. *Analyst* **2011**, *136*, 20-28.

(34) Silveira, J. A.; Servage, K. A.; Gamage, C. M.; Russell, D. H. *J. Phys. Chem. A* **2013**, *117*, 953-961.
(35) Warnke, S.; von Helden, G.; Pagel, K. *J. Am. Chem. Soc.* **2013**, *135*, 1177-1180.
(36) Cammarata, M. B.; Thyer, R.; Rosenberg, J.; Ellington, A.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2015**, *137*, 9128-9135.
(37) Quintyn, R. S.; Zhou, M.; Yan, J.; Wysocki, V. H. *Anal. Chem.* **2015**, *87*, 11879-11886.
(38) Felitsyn, N.; Kitova, E. N.; Klassen, J. S. *Anal. Chem.* **2001**, *73*, 4647-4661.
(39) Schwartz, B. L.; Bruce, J. E.; Anderson, G. A.; Hofstadler, S. A.; Rockwood, A. L.; Smith, R. D.; Chilkoti, A.; Stayton, P. S. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 459-465.
(40) Jurchen, J. C.; Williams, E. R. *J. Am. Chem. Soc.* **2003**, *125*, 2817-2826.
(41) Benesch, J. L. P. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 341-348.
(42) Boyd, R. K.; Somogyi, Á. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1275-1278.
(43) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365-8374.
(44) Jørgensen, T. J. D.; Gårdsvoll, H.; Ploug, M.; Roepstorff, P. *J. Am. Chem. Soc.* **2005**, *127*, 2785-2793.
(45) Li, J. Y.; Lyu, W. P.; Rossetti, G.; Konijnenberg, A.; Natalello, A.; Ippoliti, E.; Orozco, M.; Sobott, F.; Grandori, R.; Carloni, P. *J. Phys. Chem. Lett.* **2017**, *8*, 1105-1112.
(46) Fegan, S. K.; Thachuk, M. *J. Chem. Theory Comput.* **2013**, *9*, 2531-2539.
(47) Popa, V.; Trecroce, D. A.; McAllister, R. G.; Konermann, L. *J. Phys. Chem. B* **2016**, *120*, 5114-5124.
(48) Nesatyy, V. J.; Suter, M. J.-F. *J. Mass Spectrom.* **2004**, *39*, 93-97.
(49) Fenn, J. B. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 524-535.
(50) Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1991**, *113*, 8534-8535.
(51) Dobo, A.; Kaltashov, I. A. *Anal. Chem.* **2001**, *73*, 4763-4773.
(52) Konermann, L.; Douglas, D. J. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1248-1254.
(53) Wang, G.; Abzalimov, R. R.; Kaltashov, I. A. *Anal. Chem.* **2011**, *83*, 2870–2876.
(54) Loo, J. A.; Edmonds, C. G.; Udseh, H. R.; Smith, R. D. *Anal. Chem.* **1990**, *62*, 693-698.
(55) Borysic, A. J. H.; Radford, S. E.; Ashcroft, A. E. *J. Biol. Chem.* **2004**, *279*, 27069-27077.
(56) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1-35.
(57) Zubarev, A. R.; Makarov, A. *Anal. Chem.* **2013**, *85*, 5288-5296.
(58) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. *Science* **2006**, *314*, 109-112.
(59) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2011**, *83*, 6868-6874.
(60) Guevremont, R.; Siu, K. W. M.; Le Blanc, J. C. Y.; Berman, S. S. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 216-224.
(61) Wang, G.; Cole, R. B. *Org. Mass Spectrom.* **1994**, *29*, 419-427.
(62) Kelly, M. A.; Vestling, M. M.; Fenselau, C. C.; Smith, P. B. *Org. Mass Spectrom.* **1992**, *27*, 1143-1147.
(63) Prakash, H.; Mazumdar, S. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1409-1421.
(64) Creighton, T. E. *Proteins*; W. H. Freeman & Co: New York, 1993.
(65) Bonner, J. G.; Lyon, Y. A.; Nellessen, C.; Julian, R. R. *J. Am. Chem. Soc.* **2017**, *139*, 10286-10293.
(66) Schnier, P. D.; Gross, D. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1086-1097.
(67) Jha, A. K.; Colubri, A.; Freed, K. F.; Sosnick, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13099-13104.

(68) Consta, S.; Malevanets, A. *Phys. Rev. Lett.* **2012**, *109*.
(69) Donor, M. T.; Ewing, S. A.; Zenaidee, M. A.; Donald, W. A.; Prell, J. S. *Anal. Chem.* **2017**, *89*, 5107-5114.
(70) Beveridge, R.; Phillips, A. S.; Denbigh, L.; Saleem, H. M.; MacPhee, C. E.; Barran, P. E. *Proteomics* **2015**, *15*, 2872-2883.
(71) Yue, X.; Vahidi, S.; Konermann, L. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1322-1331.
(72) Brini, E.; Fennell, C. J.; Fernandez-Serra, M.; Hribar-Lee, B.; Luksic, M.; Dill, K. A. *Chem. Rev.* **2017**, *117*, 12385-12414.
(73) Iyengar, S. S.; Day, T. J. F.; Voth, G. A. *Int. J. Mass Spectrom.* **2005**, *241*, 197-204.
(74) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. *Nature* **1999**, *397*, 601-604.
(75) Caleman, C.; van der Spoel, D. *J. Chem. Phys.* **2006**, *125*, 1545081-1545089.
(76) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1–2*, 19-25.
(77) Sun, Y.; Vahidi, S.; Sowole, M. A.; Konermann, L. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 31-40.
(78) Huang, J.; MacKerell, A. D. *J. Comput. Chem.* **2013**, *34*, 2135-2145.
(79) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 5915-5920.
(80) Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.
(81) Vega, C.; de Miguel, E. *J. Chem. Phys.* **2007**, *126*, 154707.
(82) Konermann, L.; Metwally, H.; McAllister, R. G.; Popa, V. *Methods* **2018**, *144*, 104-112.
(83) Covey, T. R.; Thomson, B. A.; Schneider, B. B. *Mass Spectrom. Rev.* **2009**, *28*, 870-897.
(84) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695-1697.
(85) Bush, M. F.; Saykally, R. J.; Williams, E. R. *J. Am. Chem. Soc.* **2008**, *130*, 9122-9128.
(86) Fatunmbi, O.; Abzalimov, R. R.; Savinov, S. N.; Gershenson, A.; Kaltashov, I. A. *Biochemistry* **2016**, *55*, 1918-1928.
(87) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 587-596.
(88) Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Biol.* **1998**, *5*, 148-155.
(89) Zenaidee, M. A.; Leeming, M. G.; Zhang, F. T.; Funston, T. T.; Donald, W. A. *Angew. Chem.-Int. Edit.* **2017**, *56*, 8522-8526.
(90) Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107*, 3544-3567.

Figure Captions

Figure 1. Cartoon depiction of two analogous gas phase processes. (a) CID of a noncovalent protein complex. Subunits are depicted as spheres, excess H^+ are represented by "+" signs. One subunit (red) undergoes unfolding, and the protruding tail accumulates charge due to H^+ migration from the residual complex. The subunit leaves as a highly charged unfolded ion. (b) Proposed chain ejection model (CEM) for the release of an unfolded protein from an ESI droplet. As the protein gets gradually ejected, the protruding tail undergoes charge equilibration with the droplet via H^+ migration. The protein leaves as a highly charged unfolded ion. Modified from ref.[15]

Figure 2. Experimental ESI mass spectra for aMb recorded in aqueous solution at pH 4 (a) and pH 2 (b). Selected charge states are indicated. Panel (c) shows IMS collision cross section (Ω) distributions for highly charged ions acquired at pH 4 (blue) and pH 2 (red).

Figure 3. Calculated aMb net charge in solution vs. pH (black). Also shown are the contributions of titratable sites (basic = blue; acidic = red), weighted by their abundance in the protein sequence. The number of sites and their pK_a values[64] are: 1 N-terminus$^+$ (7.4); 2 Arg$^+$ (12.0); 19 Lys$^+$ (10.8); 11 His$^+$ (6.5); 8 Asp (4.0), 13 Glu (4.4), 1 C-terminus (3.9). Hatch marks highlight the range around pH 4, where the solution charge is between 22+ (pH 4.25) and 27+ (pH 3.75). CEM production of gaseous ions in these charge states from pH ~4 droplets will involve minimal H^+ migration because *gas phase charge ≈ solution charge*.

Figure 4. Typical snapshots from CEM simulation runs on Rayleigh-charged aqueous droplets

containing unfolded aMb. The protein net charge was (a) 22+ (pattern 22+[B]) and (b) 27+. Na^+ is blue, the protein backbone is magenta, positive/negative charges on the protein are highlighted as cyan/red spheres, respectively. Water oxygen is shown in red. The zoom level decreases from top to bottom. "IEM" in (a) highlights the field emission of a Na^+ ion.

Figure 5. MD simulation data for typical CEM runs on aMb protonation patterns 22+[A] (green), 22+[B] (black), 22+[C] (orange), 27+ (blue), and 33+ (red). (a) Number of water molecules and (b) number of Na^+ in the droplet vs. time. The end of each profile marks the point where the protein chain separates from the droplet. (c) Droplet charge z_D relative to the Rayleigh charge z_R, focusing on the initial regime where protein ejection had not started yet.

Figure 6. Experimental IMS data measured at pH 4 (blue) and pH 2 (red) for charge states 22+ (a-c) and 27+ (d). Vertical black lines represent average Ω values from individual MD runs, standard deviations are indicated as horizontal bars. The four panels represent data for different protonation patterns, as indicated. Shown on the right are MD structures of those trajectories that best matched the experimental data (pink asterisks, overlays of five structures between 100 ns and 500 ns, all at the same zoom level, N-termini pointing to the left). Also included in (b) are MD structures from two runs that yielded more compact structures than those observed experimentally. Positive/negative charges on aMb are shown as cyan/red spheres, respectively.

Figure 1

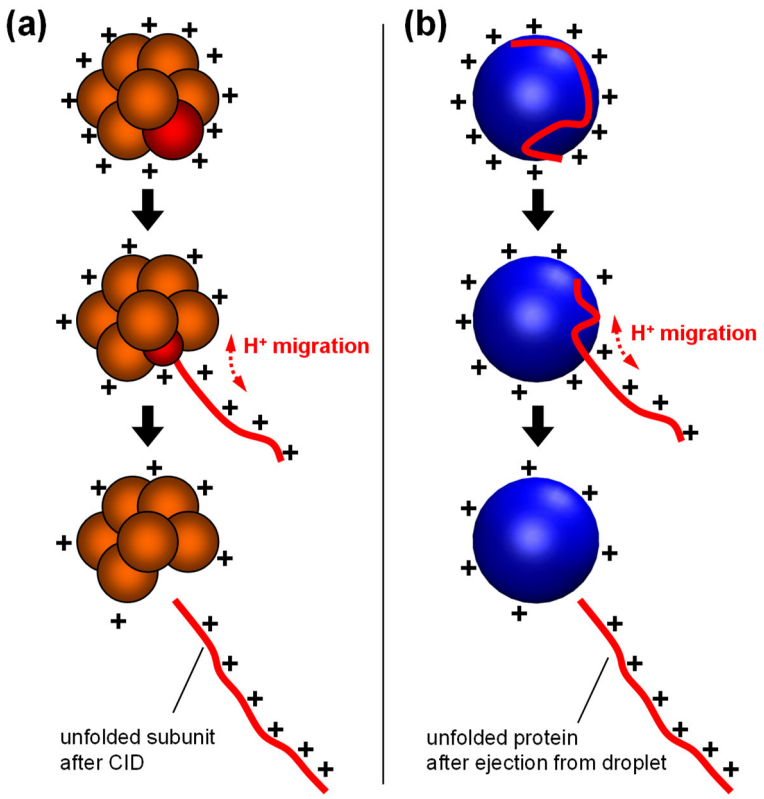

(a)
H+ migration
unfolded subunit
after CID
(b)
H+ migration
unfolded protein
after ejection from droplet

Figure 2

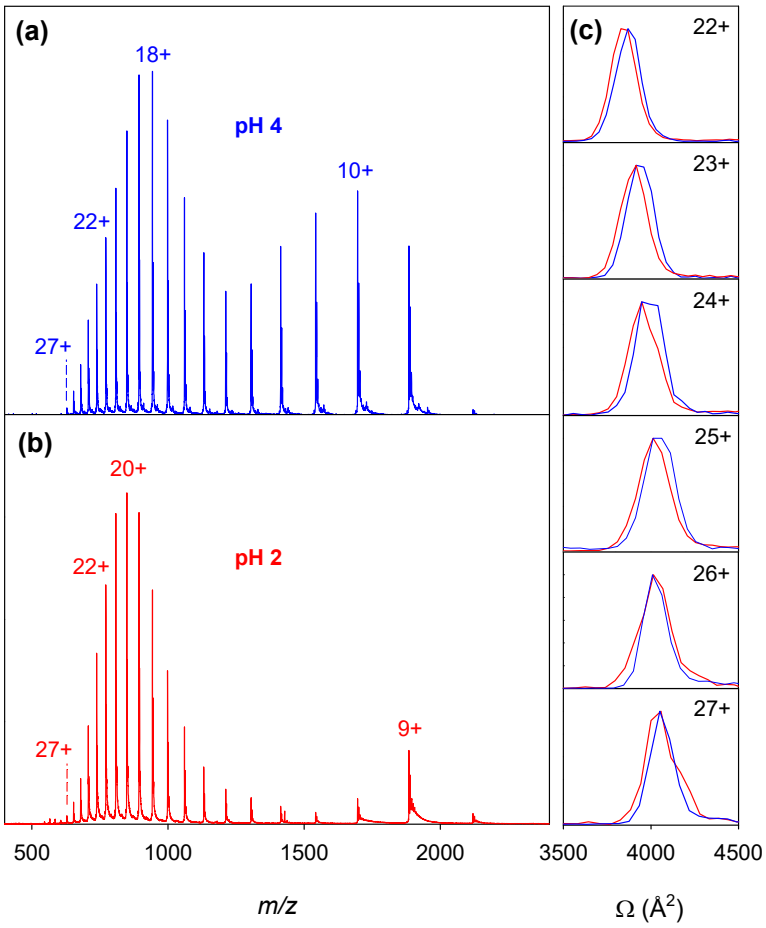

(a)
18+
pH 4
10+
22+
27+
(b)
20+
22+
pH 2
9+
27+
500
1000
1500
2000
m/z
(c)
22+
23+
24+
25+
26+
27+
3500
4000
4500
Ω (Å²)

Figure 3

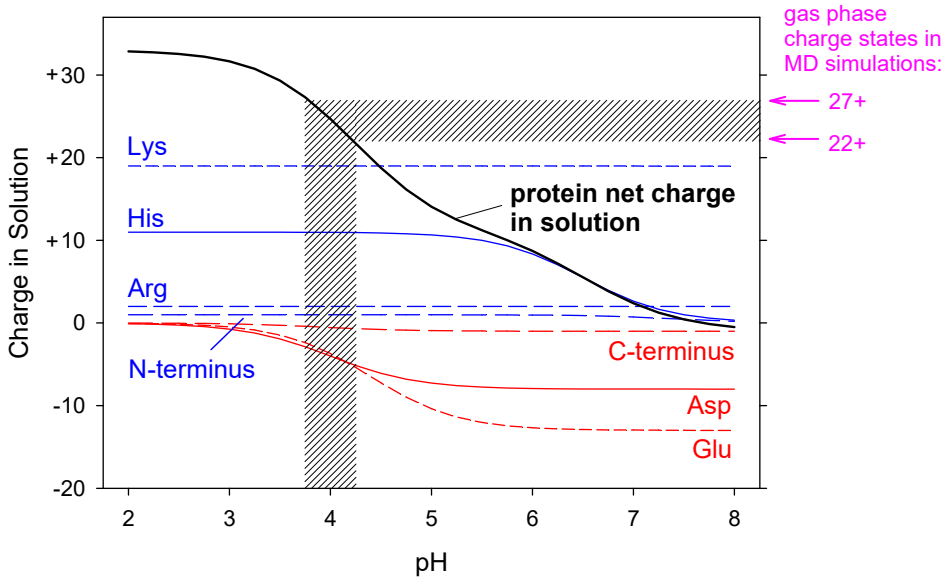

gas phase
charge states in
MD simulations:
27+
22+
+30
+20
+10
0
-10
-20
Charge in Solution
Lys
His
Arg
N-terminus
protein net charge
in solution
C-terminus
Asp
Glu
2
3
4
5
6
7
8
pH

Figure 4

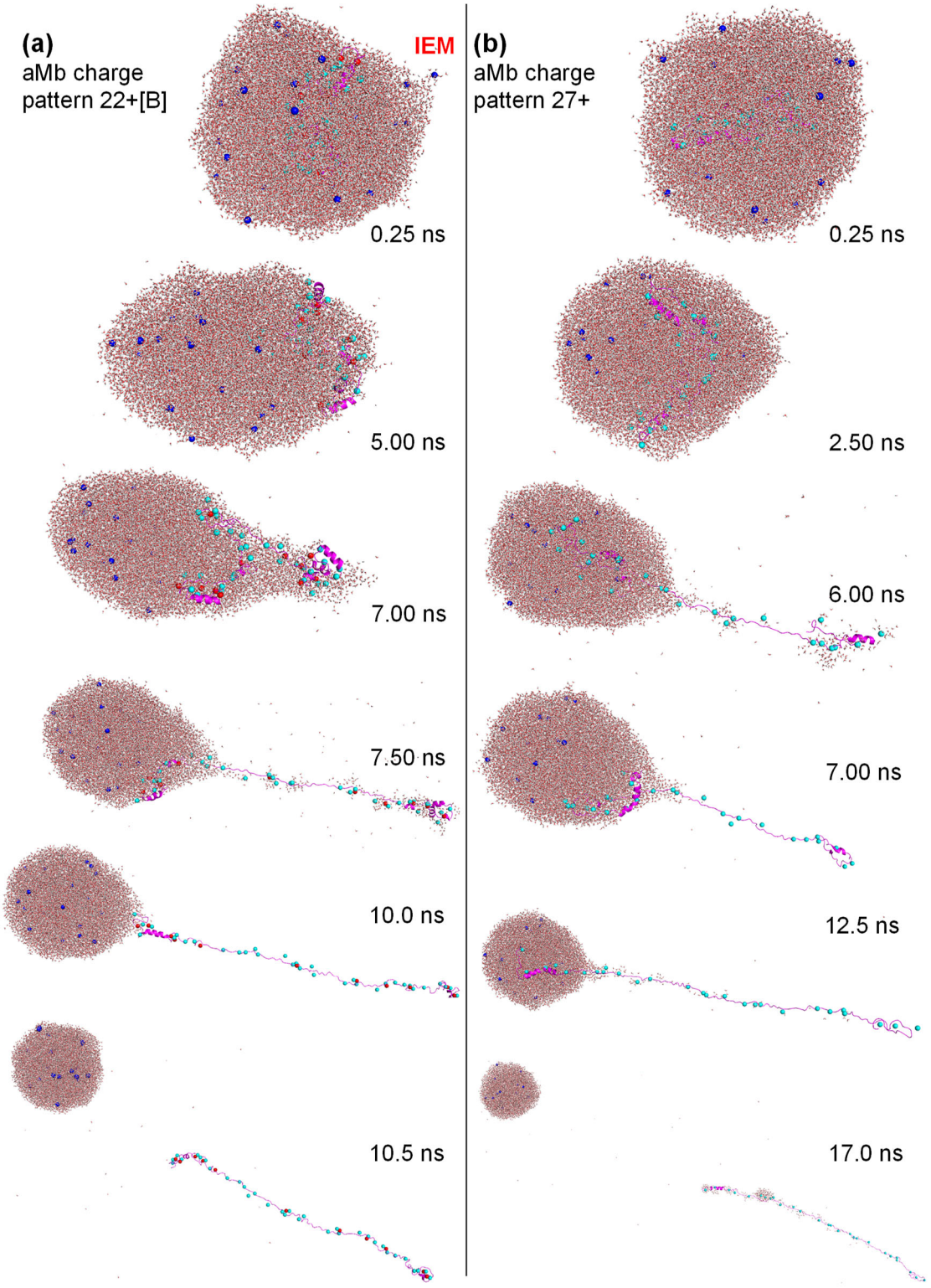

(a)
aMb charge
pattern 22+[B]
IEM
0.25 ns
5.00 ns
7.00 ns
7.50 ns
10.0 ns
10.5 ns
(b)
aMb charge
pattern 27+
0.25 ns
2.50 ns
6.00 ns
7.00 ns
12.5 ns
17.0 ns

Figure 5

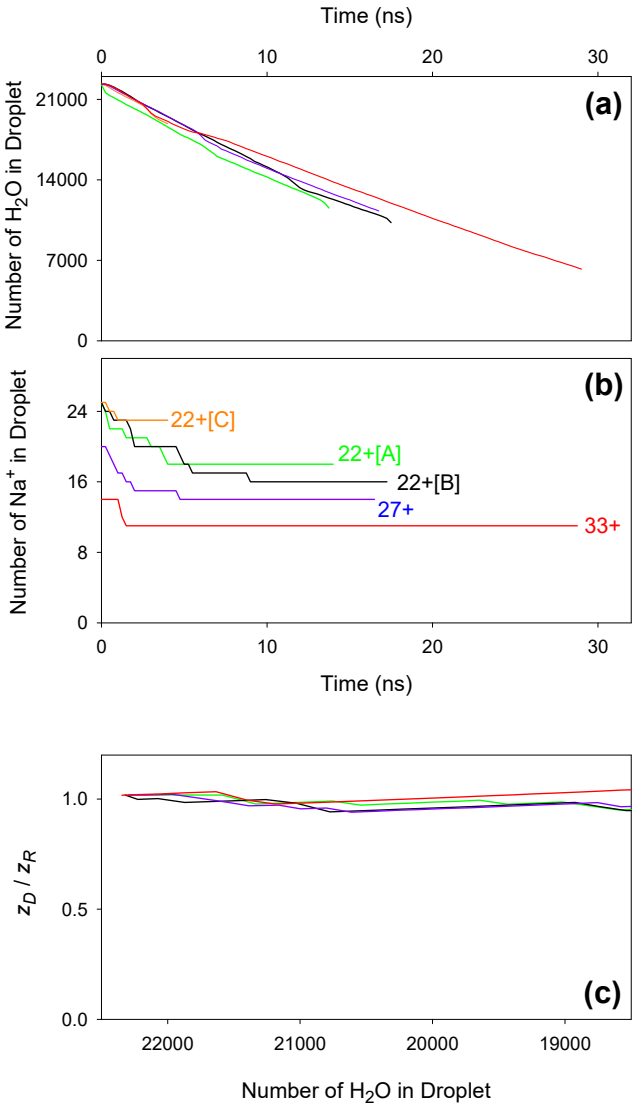
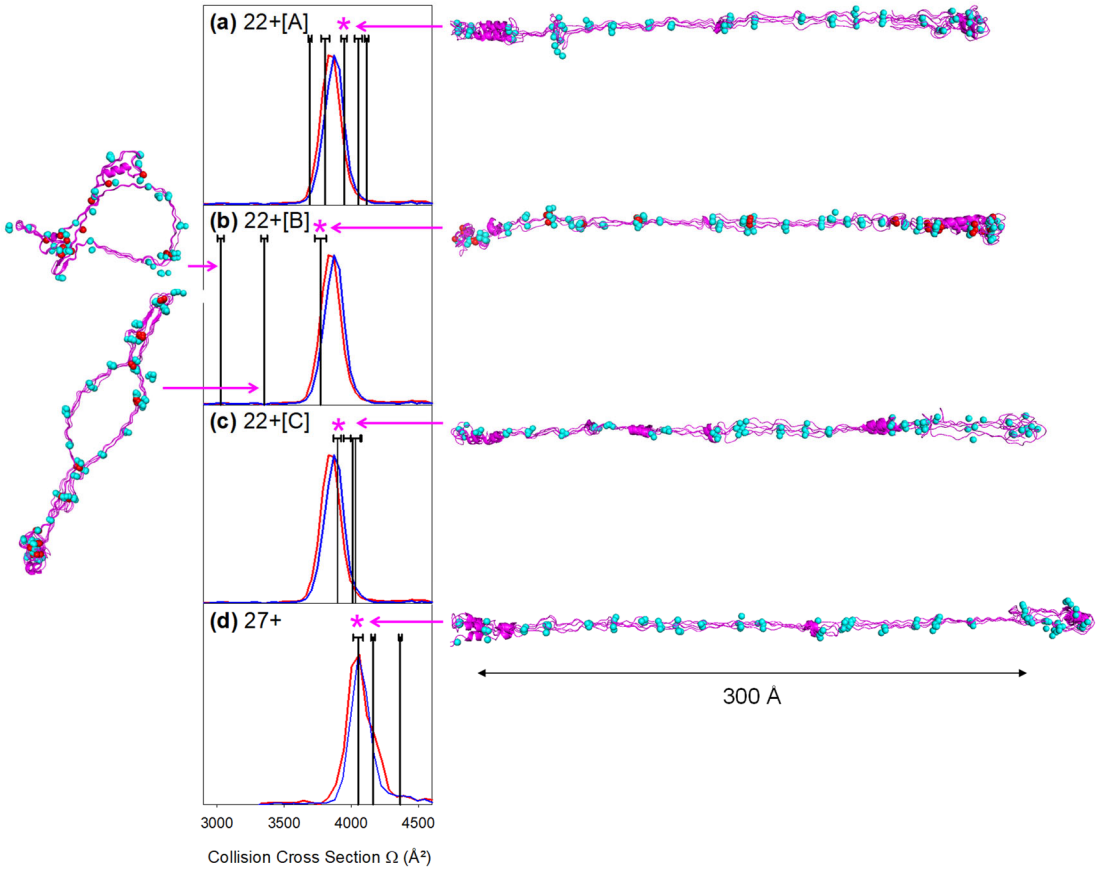
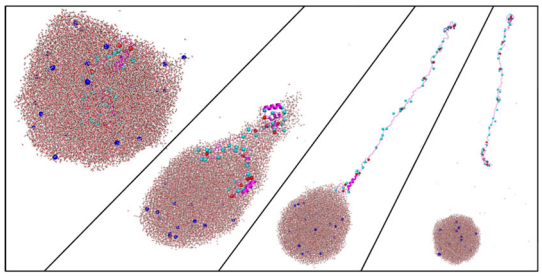

Time (ns)
0
10
20
30
Number of H2O in Droplet
21000
14000
7000
0
(a)
Number of Na+ in Droplet
24
16
8
0
(b)
22+[C]
22+[A]
22+[B]
27+
33+
Time (ns)
zD / zR
1.0
0.5
0.0
(c)
22000
21000
20000
19000
Number of H2O in Droplet

Figure 6


(a) 22+[A]
(b) 22+[B]
(c) 22+[C]
(d) 27+
3000
3500
4000
4500
Collision Cross Section Ω (Å²)
300 Å

For Table of Contents Only



Chain Ejection Model (CEM)