SUPPORTING INFORMATION

Formation of Gaseous Protein Ions from Aqueous Ammonium Acetate Droplets during Native Electrospray: Insights from Mobile-Proton Molecular Dynamics Simulations

Kasra Hanifi and Lars Konermann*

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

*E-mail address of the corresponding author: konerman@uwo.ca

Table of Contents:

SI Methods

ESI Experiments. Bovine ubiquitin, hen egg white lysozyme, porcine pepsin, human plasma transthyretin (TTR), NH_4Ac and TCEP (tris(2-carboxyethyl)phosphine·HCl) were purchased from Sigma (St. Louis, MO), see Figure S1 for sequences and other details. De-cysteinylation of TTR was performed by exposing 8 μM protein to 1 mM TCEP for 30 min at room temperature in 10 mM aqueous ammonium acetate. Excess TCEP and free cysteine were removed by 8 cycles of centrifugal filtration using 10 kDa MWCO Amicon filters, with resuspension of the protein in 10 mM aqueous NH_4Ac after each cycle (Figure S2).

Native ESI-IMS/MS was performed on a SYNAPT G2 Q-TOF (Waters, Milford, MA). Ubiquitin and lysozyme solutions were prepared at 5 μM (10 μM for pepsin, 8 μM for TTR) in 10 mM NH_4Ac at pH 7 and infused at 5 μL min^{-1}, followed by ESI at +2.8 kV under gentle ion sampling conditions (sampling cone 5 V, extraction cone 3 V, desolvation gas 40 °C, source 30 °C, trap collision voltage (trap CV) between 2 V and 4 V unless noted otherwise). Some spectra were also acquired on a SYNAPT G2-Si instrument using conditions analogous to those above. MS/MS was performed by implementing collision-induced dissociation (CID) in the Ar-filled trap cell, with trap CV between 2 V and 50 V after quadrupole selection of the precursor ion.

MD Simulation Details. GROMACS 2020.4 with leapfrog integration was used,[1] along with in-house Python and Fortran code. We employed Linux workstations with AMD Ryzen 7 9800X3D processors (8 cores, 16 threads), NVIDIA GeForce RTX 4080 SUPER GPUs, and 32 GB of DDR5 RAM. Droplets or proteins were centered in a cubic $(999.9\ nm)^3$ box under pseudo-periodic boundary conditions which are equivalent to vacuum boundary conditions.[2] The Nosé-Hoover thermostat[3] was used for temperature control, after starting each run with initial velocities that

were sampled randomly from a Boltzmann distribution at the desired temperature. For droplet simulations, protein and non-protein were modeled as separate coupling groups.[4] To eliminate LINCS instabilities, droplets were simulated using a dt = 1 fs integration step without bond constraints. Water was rigidified using SETTLE.[5] OH and NH vibrational frequencies had to be lowered slightly by adjusting the corresponding force constants to eliminate GROMACS grompp warnings regarding oscillation periods being less than 10 times dt.[6] Using previously established protocols,[7, 8] gas-phase simulations employed a 1 fs integration step and LINCS-constrained bond lengths for all atoms.

Ubiquitin (1UBQ)
M = 8565 Da, SS bonds = 0, pI = 6.6, net charge at pH 5.4 = +1

```
    1           11          21          31          41
nt  MQIFVKTLTG  KTITLEVEPS  DTIENVKAKI  QDKEGIPPDQ  QRLIFAGKQL
    51          61          71
    EDGRTLSDYN  IQKESTLHLV  LRLRGG ct
```

Lysozyme (1AKI)
M = 14305 Da, SS bonds = 4, pI = 10.5, net charge at pH 5.4 = +9

```
    1           11          21          31          41
nt  KVFGRCELAA  AMKRHGLDNY  RGYSLGNWVC  AAKFESNFNT  QATNRNTDGS
    51          61          71          81          91
    TDYGILQINS  RWWCNDGRTP  GSRNLCNIPC  SALLSSDITA  SVNCAKKIVS
    101         111         121
    DGNGMNAWVA  WRNRCKGTDV  QAWIRGCRL ct
```

Pepsin (4PEP)
M = 34504 Da, SS bonds = 3, pI = 3.2, net charge at pH 5.4 = -39

```
    1           11          21          31          41
nt  IGDEPLENYL  DTEYFGTIGI  GTPAQDFTVI  FDTGSSNLWV  PSVYCSSLAC
    51          61          71          81          91
    SDHNQFNPDD  SSTFEATSQE  LSITYGTGSM  TGILGYDTVQ  VGGISDTNQI
    101         111         121         131         141
    FGLSETEPGS  FLYYAPFDGI  LGLAYPSISA  SGATPVFDNL  WDQGLVSQDL
    151         161         171         181         191
    FSVYLSSNDD  SGSVVLLGGI  DSSYYTGSLN  WVPVSVEGYW  QITLDSITMD
    201         211         221         231         241
    GETIACSGGC  QAIVDTGTSL  LTGPTSAIAN  IQSDIGASEN  SDGEMVISCS
    251         261         271         281         291
    SIDSLPDIVF  TIDGVQYPLS  PSAYILQDDD  SCTSGFEGMD  VPTSSGELWI
    301         311         321
    LGDVFIRQYY  TVFDRANNKV  GLAPVA ct
```

Transthyretin = TTR (3GRG)
M = 55046 Da, SS bonds = 0, pI = 5.3, net charge at pH 5.4 = -4
(the parameters above refer to the tetramer)

```
    1           11          21          31          41
nt  GPTGTGESKC  PLMVKVLDAV  RGSPAINVAV  HVFRKAADDT  WEPFASGKTS
    51          61          71          81          91
    ESGELHGLTT  EEEFVEGIYK  VEIDTKSYWK  ALGISPFHEH  AEVVFTANDS
    101         111         121
    GPRRYTIAAL  LSPYSYSTTA  VVTNPKE ct
```

Figure S1 (previous page): Sequences, pdb accession codes,[9-12] and pH 5.4 charge patterns of the proteins studied in this work. Protonated basic sites are highlighted in blue, deprotonated acidic sites are shown in red; nt = N-terminus, ct = C-terminus. Also included is the MW and the isoelectric point of each protein (pI values were estimated using web.expasy.org/compute_pi/). Phosphoserine 68 in pepsin was modeled as serine, because the phosphorylated form is not parametrized in the MD force field used here. Missing N-terminal and C-terminal residues in the TTR crystal structure 3GRG were added (underlined). Two substitutions (F87M and L110M, also underlined) in the 3GRG sequence were reverted to the wild type.

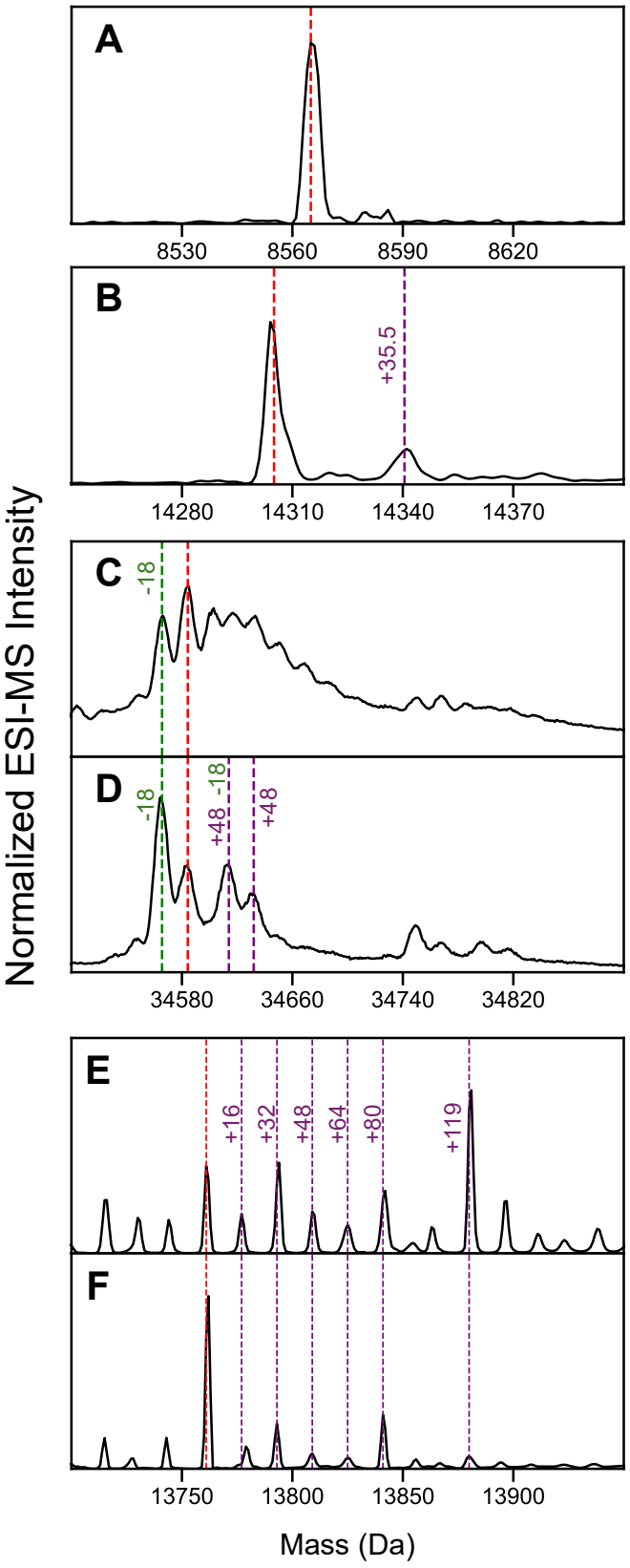
A
8530
8560
8590
8620
B
+35.5
14280
14310
14340
14370
C
-18
D
-18
+48
-18
+48
34580
34660
34740
34820
E
+16
+32
+48
+64
+80
+119
F
13750
13800
13850
13900
Mass (Da)
Normalized ESI-MS Intensity

Figure S2 (previous page): Mass distributions of the proteins examined in this work, generated by deconvoluting experimental ESI mass spectra.[13] (A) Ubiquitin and (B) lysozyme under native ESI conditions (neutral pH, 10 mM NH_4Ac). The +35.5 signal in panel B likely represents a chloride adduct; Cl^- is a ubiquitous trace contaminant in biological samples. (C, D) Pepsin. To suppress adduct formation, the pepsin spectra in this Figure were *not* acquired in neutral solution. Panel C: Direct infusion in water with 2% formic acid. Panel D: After LC/MS on a C4 column using a water/acetonitrile gradient with 0.1% formic acid under harsh ion sampling conditions (sampling cone 100 V, desolvation gas 400 °C, source 80 °C). Particularly under the conditions of panel D, pepsin undergoes water loss (-18 Da), a process that is common for S/T-rich proteins such as pepsin under activating source conditions.[14] (E, F) Spectra of monomeric TTR acquired using LC/MS under the conditions of panel D. In panel E, the protein is cysteinylated (+119 Da), which represents a common TTR covalent modification.[15] TCEP treatment removes this modification (panel F). For all panels, red lines indicate the protein mass expected from the corresponding sequences, except for pepsin where the red line represents the sequence mass with one phosphorylation (a modification that is also seen in the 4PEP crystal structure).[11, 16] Signals marked as +16, +32, etc. mark oxidation products.

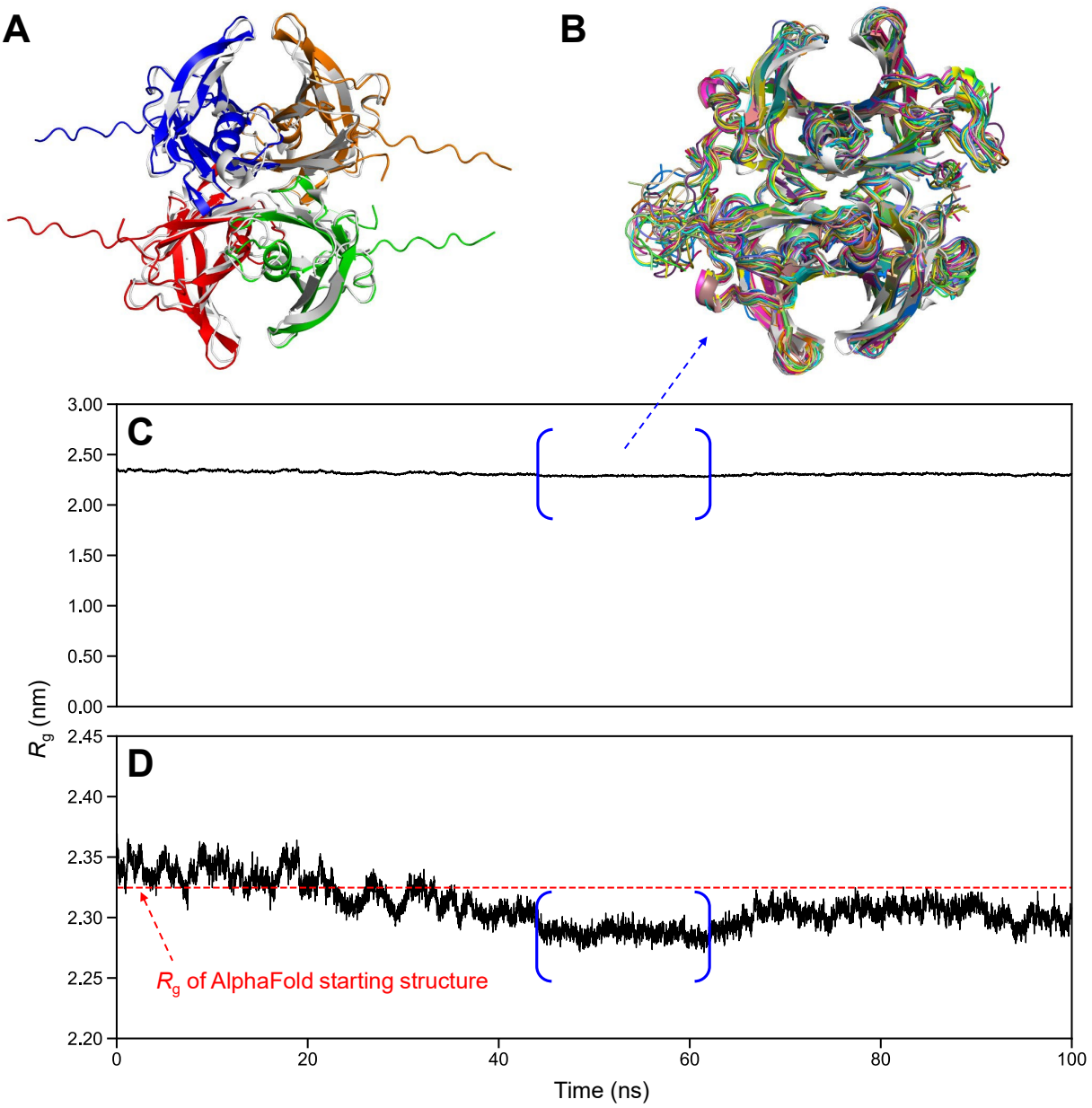

Figure S3. Generating starting structures for TTR dMPMD simulations, using Alphafold 3[17] and bulk solution MD simulations.[18] (A) Alphafold structure of the TTR tetramer, with subunits shown in different colors. Extended segments pointing to the left and right are N-termini that were missing in the X-ray structure (see Figure S1). Gray shows the X-ray structure 3GRG.[12] (B) Overlay of 20 structures generated by exposing the Alphafold coordinates of panel A to ~50 ns of MD simulations in bulk solution (TIP3P water, 150 mM NaCl, 300 K, additional Na^+ were inserted to ensure a neutral PBC box). Gray shows the X-ray structure 3GRG. (C, D) Radius of gyration (R_g) during a 100 ns MD run. The 20 MD structures in panel B were extracted in equidistant intervals from the region highlighted by the blue bracket, during which the protein attained its most compact conformation. Panels C and D show the same data with different y-axis scaling.

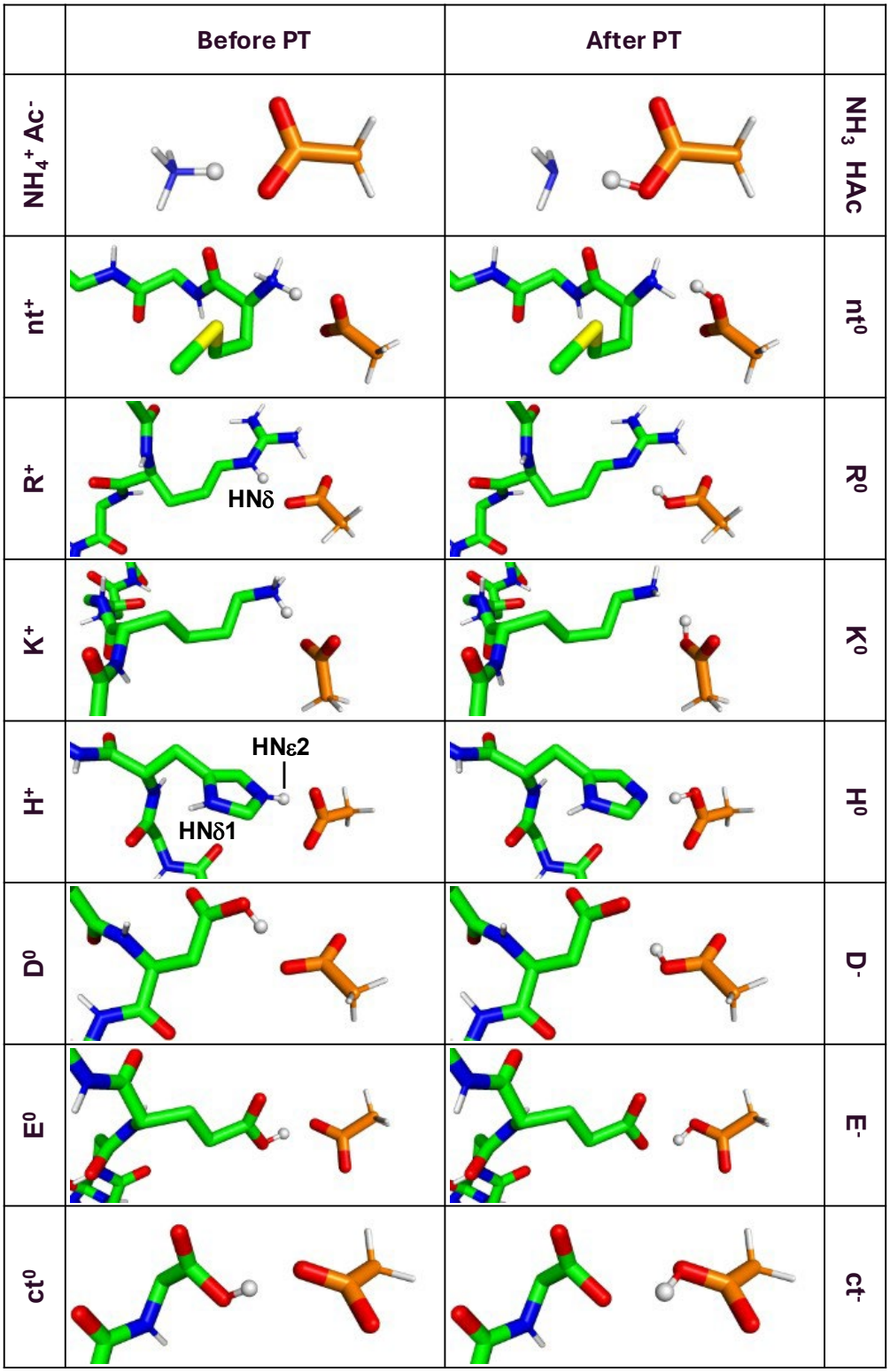

Figure S4 (A). PT examples from all possible protonated donors to Ac^- as acceptor. The transferred proton is shown as sphere. Data are from ubiquitin and lysozyme dMPMD runs in 3 nm and 6 nm droplets. Water and other molecules or ions, as well as nonpolar hydrogens were omitted.

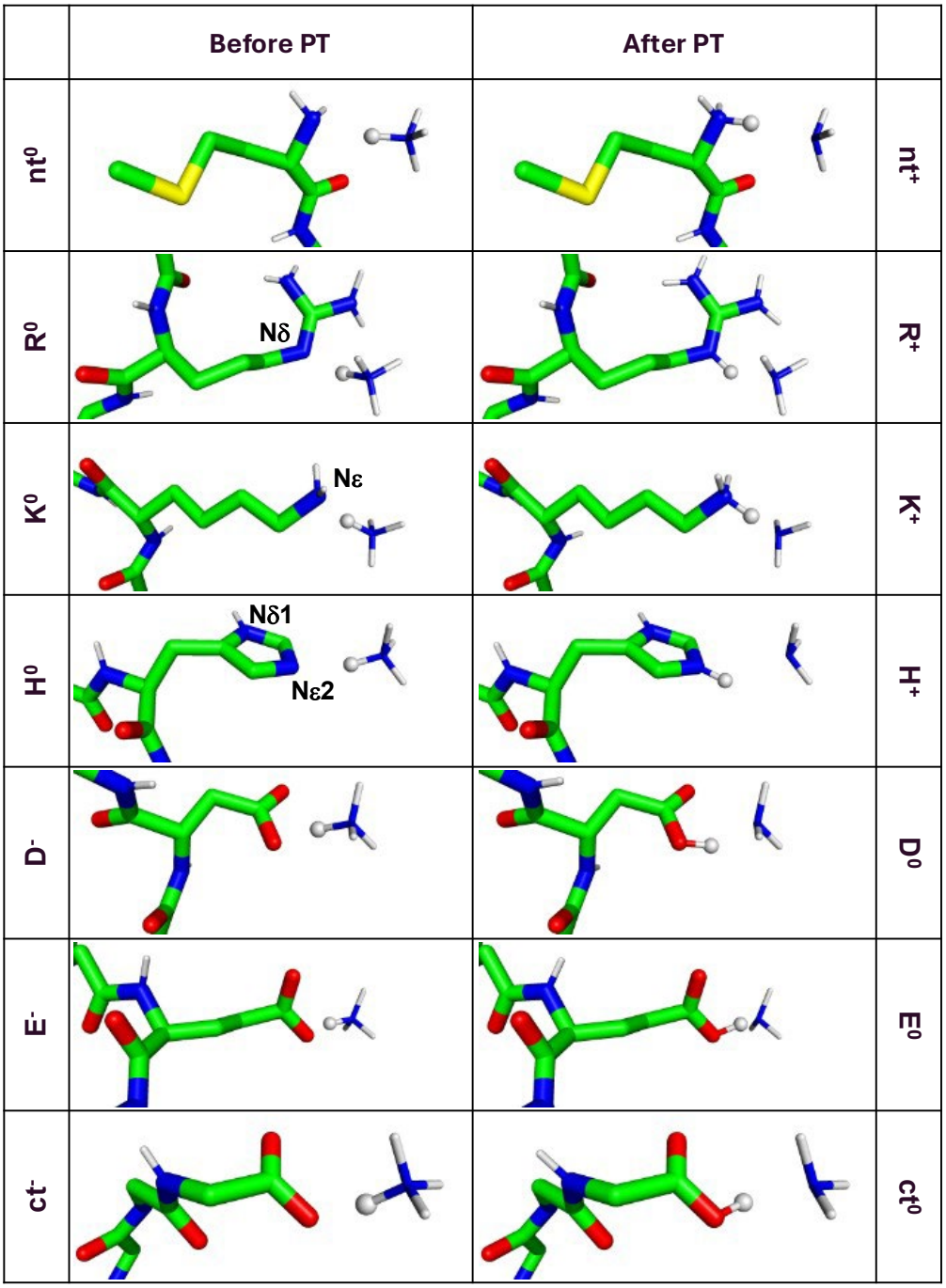

Figure S4 (B). PT examples analogous to panel A, but with NH_4^+ as proton donor and all possible deprotonated acceptors.

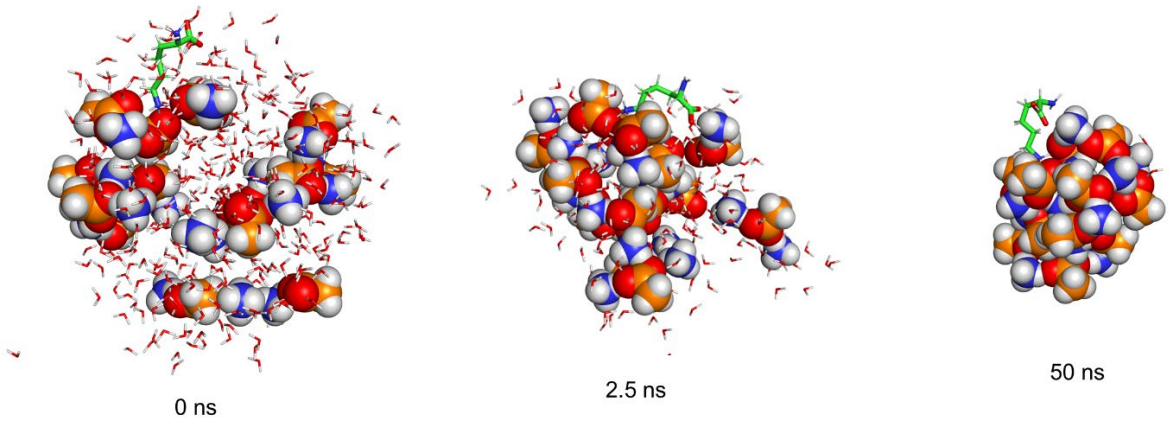

Figure S5. Trajectory snapshots from an MD run that modeled the evaporation of a r_0 = 1.5 nm droplet consisting of aqueous NH_4Ac and a single amino acid (0[K^+]0 for the example shown here, where symbols outside the brackets denote charges on nt and ct). Simulations of the type illustrated here were used for generating H-bond distance histograms for all possible donor/acceptor pairs in Figure 1 (main text). The resulting data are compiled in Figure S6. H_2O is the preferred H-bonding partner in solution, making it difficult to study donor/acceptor contacts while solvent was present. We therefore probed these contacts in dry NH_4Ac clusters, produced by evaporation of r_0 = 1.5 nm droplets with 14 NH_4^+, 14 Ac^-, and the amino acid of interest (as exemplified by the snapshots in this figure). Additional NH_4^+ were included to attain z_R = 7+. Charge patterns for tracking H-bonds with Ac^- were: +[R^+]0, 0[K^+]0, +[H^+]0, 0[D^0]0, and 0[E^0]0. For probing contacts with NH_4^+ we used 0[R^0]0, 0[K^0]0, 0[Hδ^0]0, 0[Hε^0]0, 0[D^-]0, and 0[E^-]0. Termini were probed using +[GG]0, 0[GG]-, and 0[GG]0. The r_0 = 1.5 nm droplets were evaporated at 370 K without PT for 50 ns. The resulting dry clusters were simulated for 2 ns, with H-bond tracking for histogram generation in 1 ps intervals.

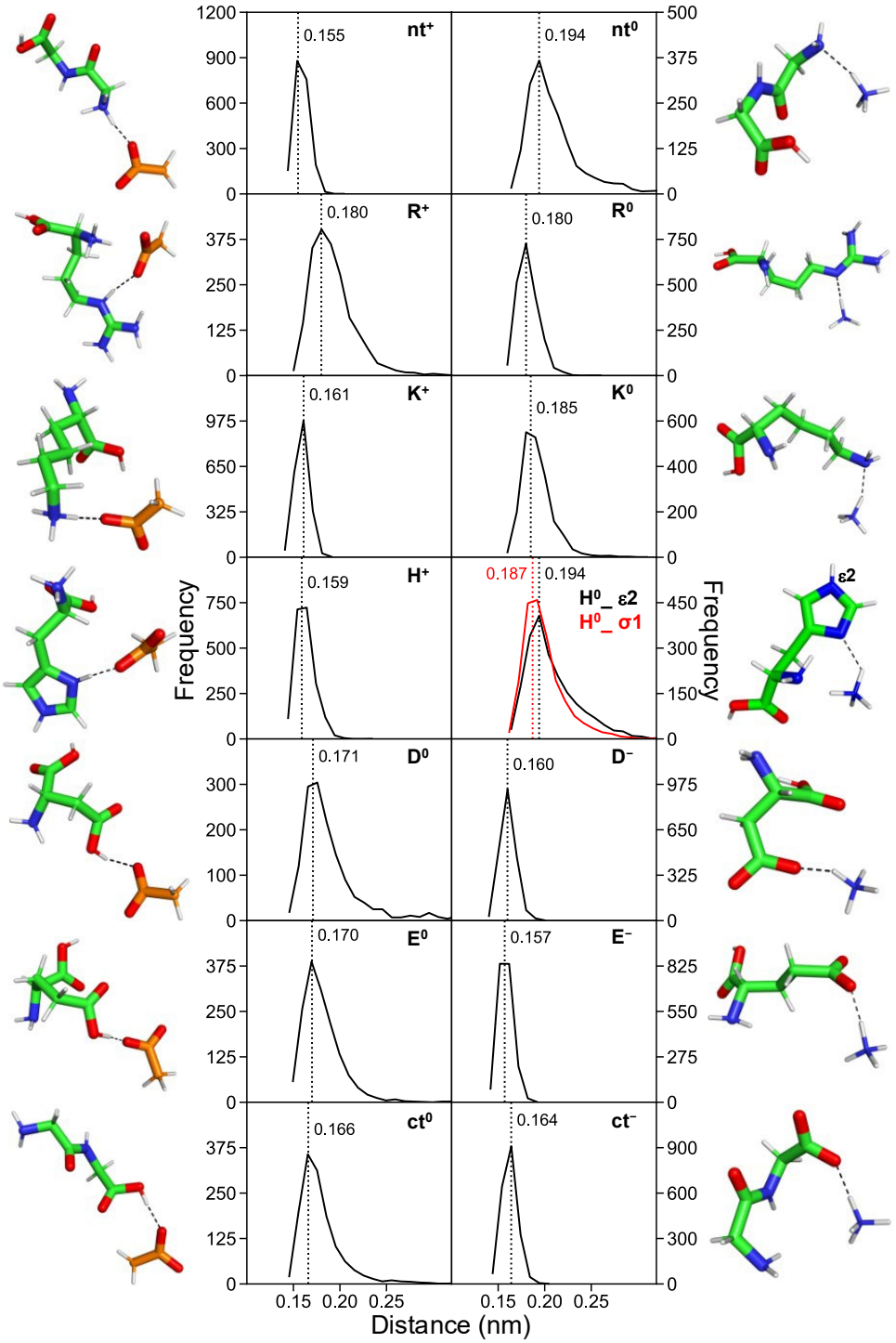

Figure S6 (A). H-bond distance distributions for all PT donor/acceptor pairs considered for dMPMD in aqueous NH_4Ac droplets. Histograms were generated using amino acid-containing NH_4Ac clusters formed via droplet evaporation (Figure S5). Maxima define PT_dist_max for each donor/acceptor pair (numerical values are indicated in each panel). In the case of H^0, PT_dist_max determined for ε2 was used for both the δ1 and the ε2 tautomer. MD snapshots corresponding to each distribution are shown along the left and right, illustrating donor/acceptor H-bonding (these snapshots omit all other species in the cluster).

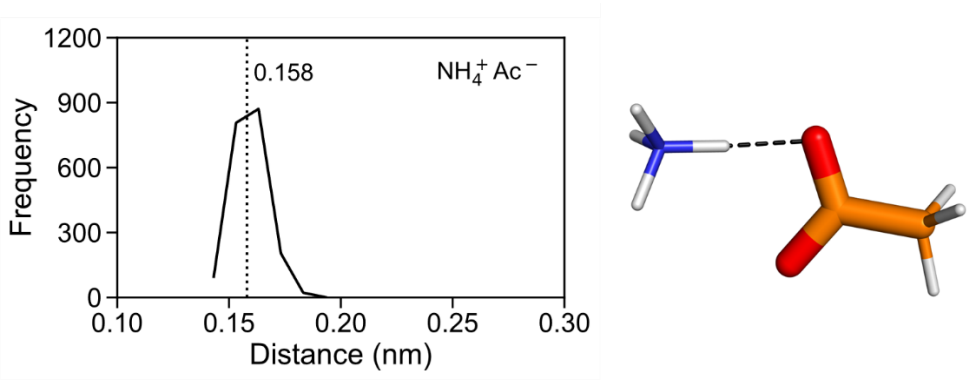

Figure S6 (B). H-bond distance distribution for determining PT_dist_max between NH_4^+ and Ac^-. For additional details, see caption of panel A. These data were generated using a method similar to that in Figure S5, except that the droplets did not include an amino acid. The resulting PT_dist_max = 0.158 nm for PT from NH_4^+ to Ac^- is close that (0.155 nm) in ref.[6]

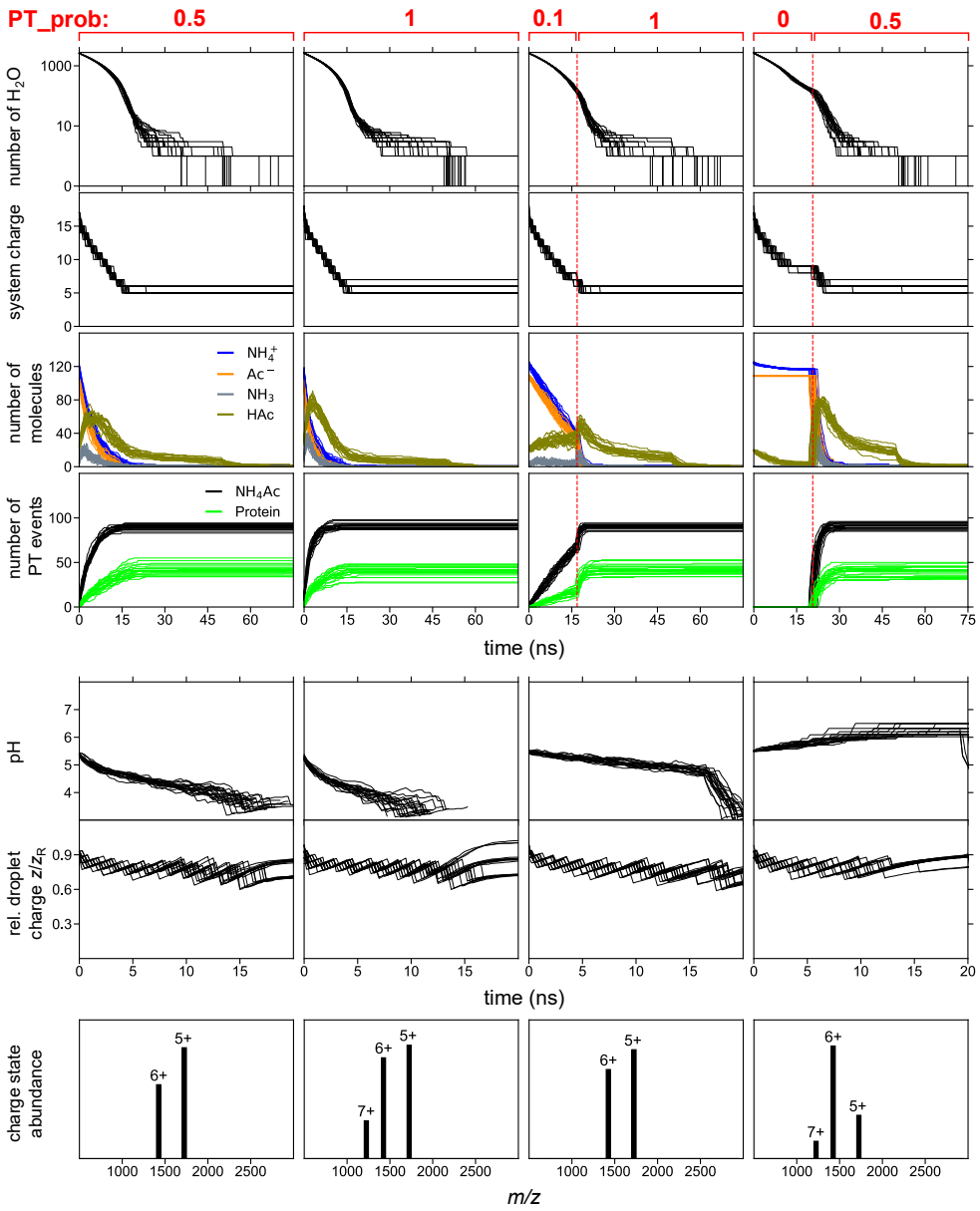

Figure S7. Ubiquitin dMPMD simulations as in Figure 2 with r_0 = 3 nm, but with different PT_prob: Column 1: PT_prob = 0.5. Column 2: PT_prob = 1. Column 3: PT_prob = 0.1 until ~17 ns, then PT_prob = 1. Column 4: PT_prob = 0 until ~21 ns, then PT_prob = 0.5. Bottom panels show simulated ESI charge state distributions. Each column summarizes data for 20 runs. The overall sequence of events, with protein ion release via droplet evaporation to dryness was, similar for all the conditions considered here and in Figure 2 (which had PT_prob = 0.1), even though time-dependent changes in droplet composition were somewhat different for the different conditions tested here. Most importantly, ESI charge states produced under all conditions covered the range of 5+ to 7+, that is consistent with experiments.[8, 19] In summary, dMPMD simulation outcomes are quite insensitive to the numerical value of PT_prob. This is true even when changing PT_prob during the simulations (analogous to the time-dependent basicity changes implemented by Cordes and Gallagher).[19] Except for the data shown here, all simulations in this work employed PT_prob = 0.1, which had already been used previously.[6]

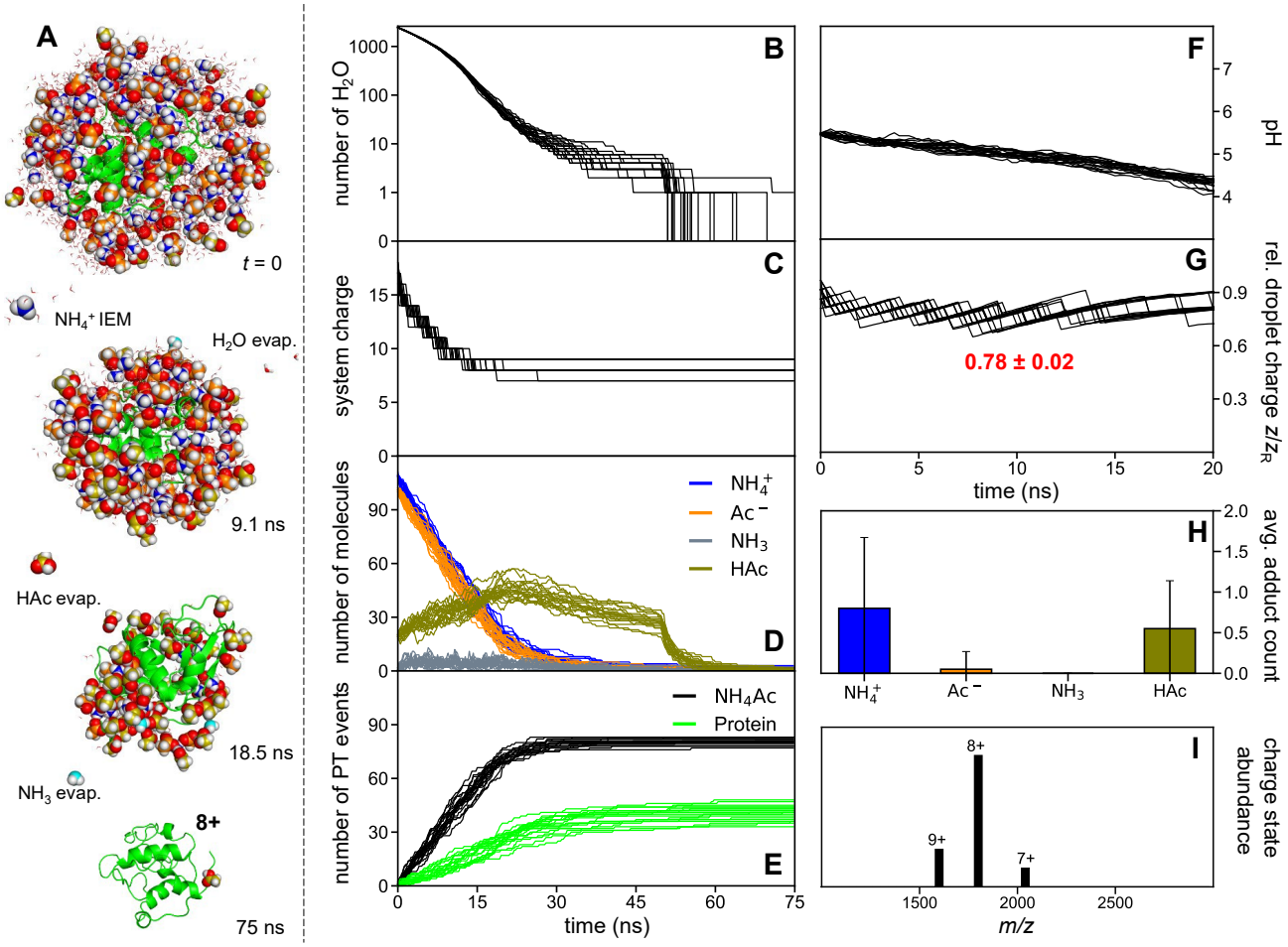

Figure S8. Lysozyme native ESI (dMPMD) simulations in r_0 = 3 nm aqueous NH_4Ac droplets, using positive ESI mode. (A) Selected trajectory snapshots, culminating in an 8+ gaseous ion that remains bound to a single HAc molecule. All other panels show data from 20 repeat runs. (B) Number of water molecules, (C) overall system charge, (D) number of solutes in the droplet, (E) cumulative number of PT events from NH_4^+ to Ac^-, and PT events involving the protein, (F) droplet pH, (G) droplet charge relative to the Rayleigh limit, (H) protein adducts after 75 ns, (I) simulated ESI charge state distribution (including adducts).

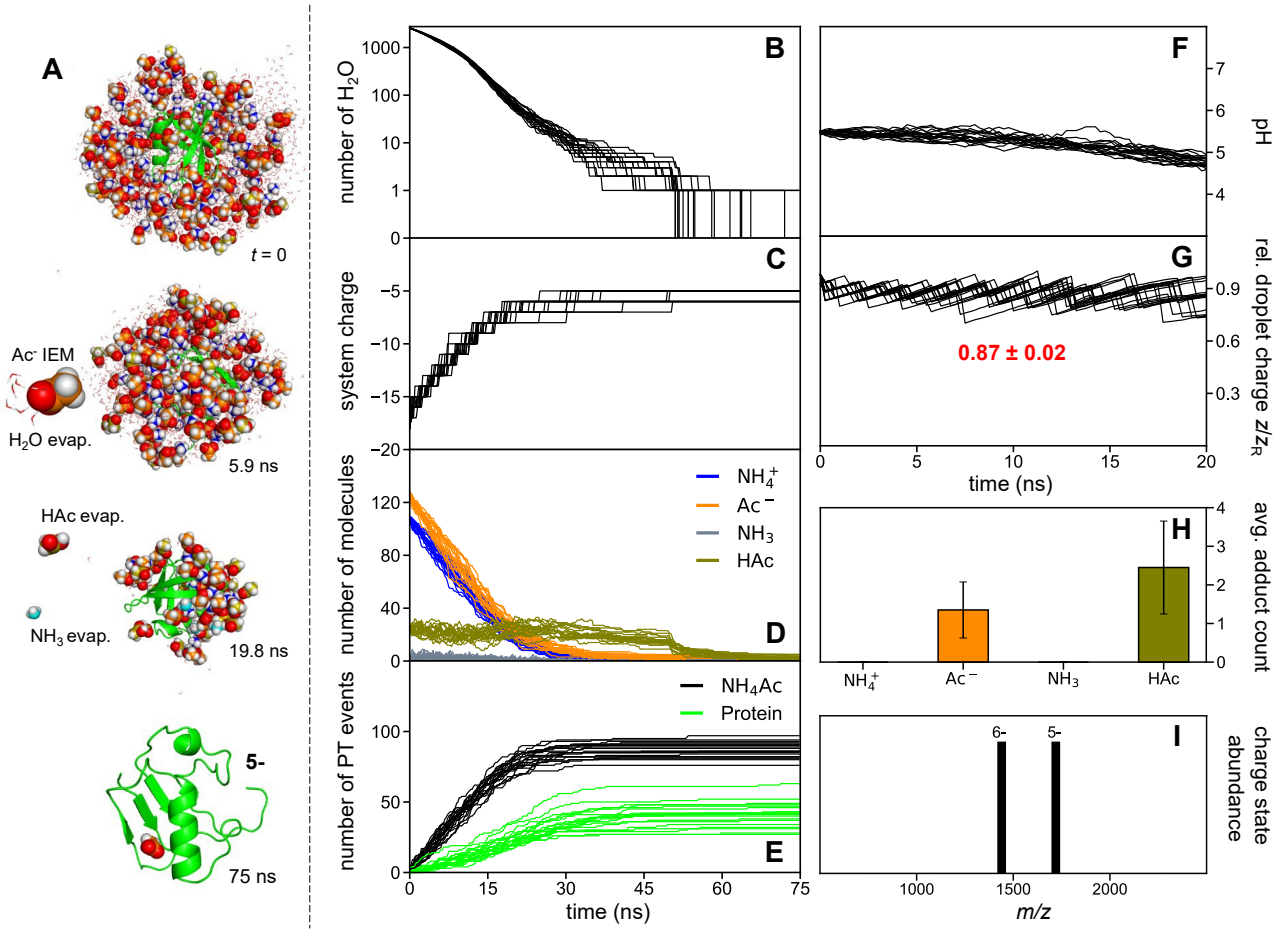

Figure S9. Ubiquitin native ESI (dMPMD) simulations in r_0 = 3 nm aqueous NH_4Ac droplets, using negative ESI mode. (A) Selected trajectory snapshots. All other panels show data from 20 repeat runs. (B) Number of water molecules, (C) net droplet charge, (D) number of solutes in the droplet, (E) cumulative number of PT events from NH_4^+ to Ac^-, and PT events involving the protein, (F) droplet pH, (G) droplet charge relative to the Rayleigh limit, (H) protein adducts after 75 ns, (I) simulated ESI charge state distribution (including adducts).

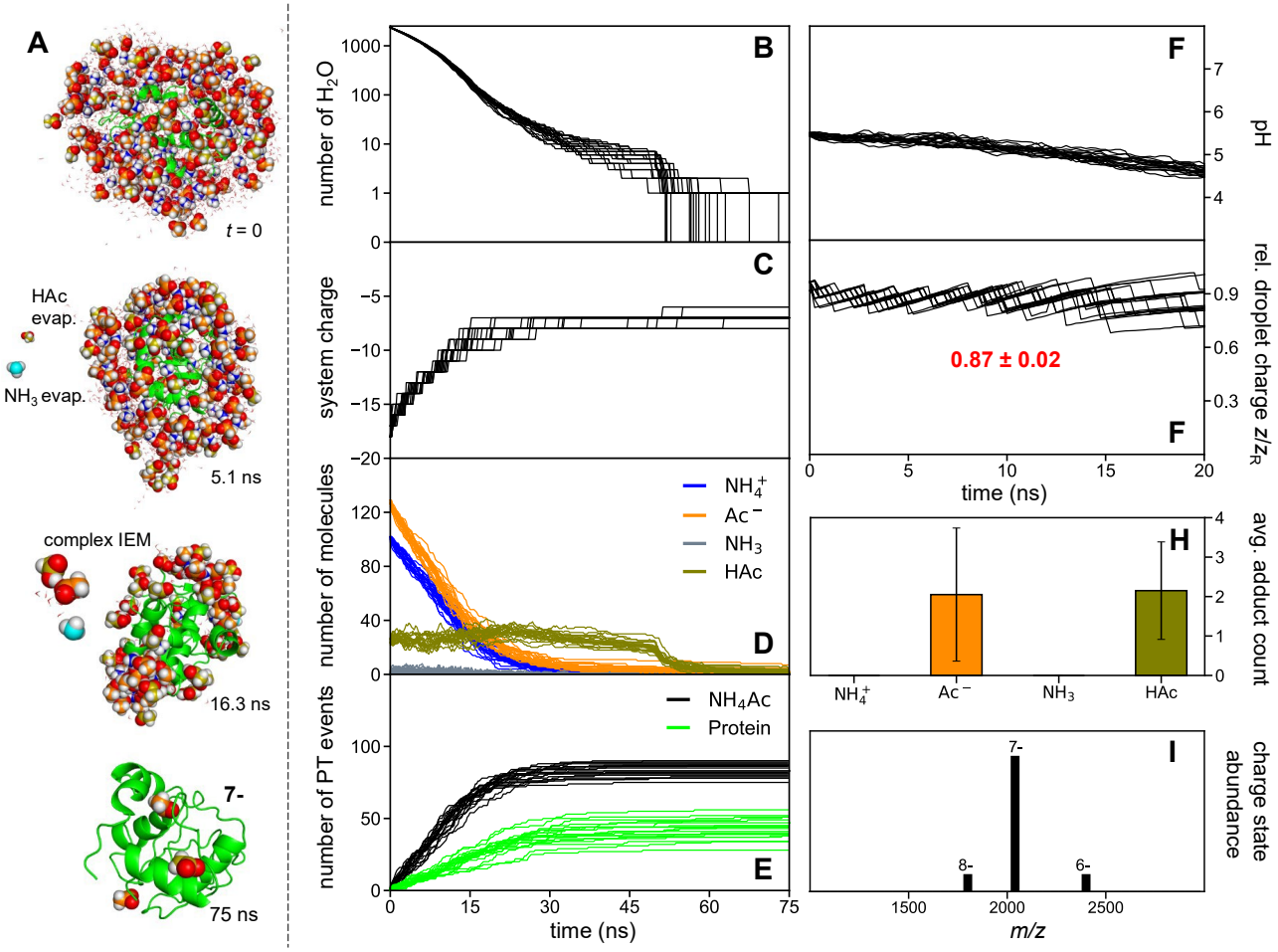

Figure S10. Lysozyme native ESI (dMPMD) simulations in r_0 = 3 nm aqueous NH_4Ac droplets, using negative ESI mode. The 16.3 ns time point in panel A illustrates a somewhat unusual IEM event that culminates in the ejection of a small cluster consisting of one Ac^-, one HAc, one NH_3, and several waters. Most other IEM events in negative and in positive ESI mode involved the ejection of an ion bound to several residual H_2O (see, e.g., Figure S9A, t = 5.9 ns). For information regarding all other panels in this figure, see the caption of Figure S8.

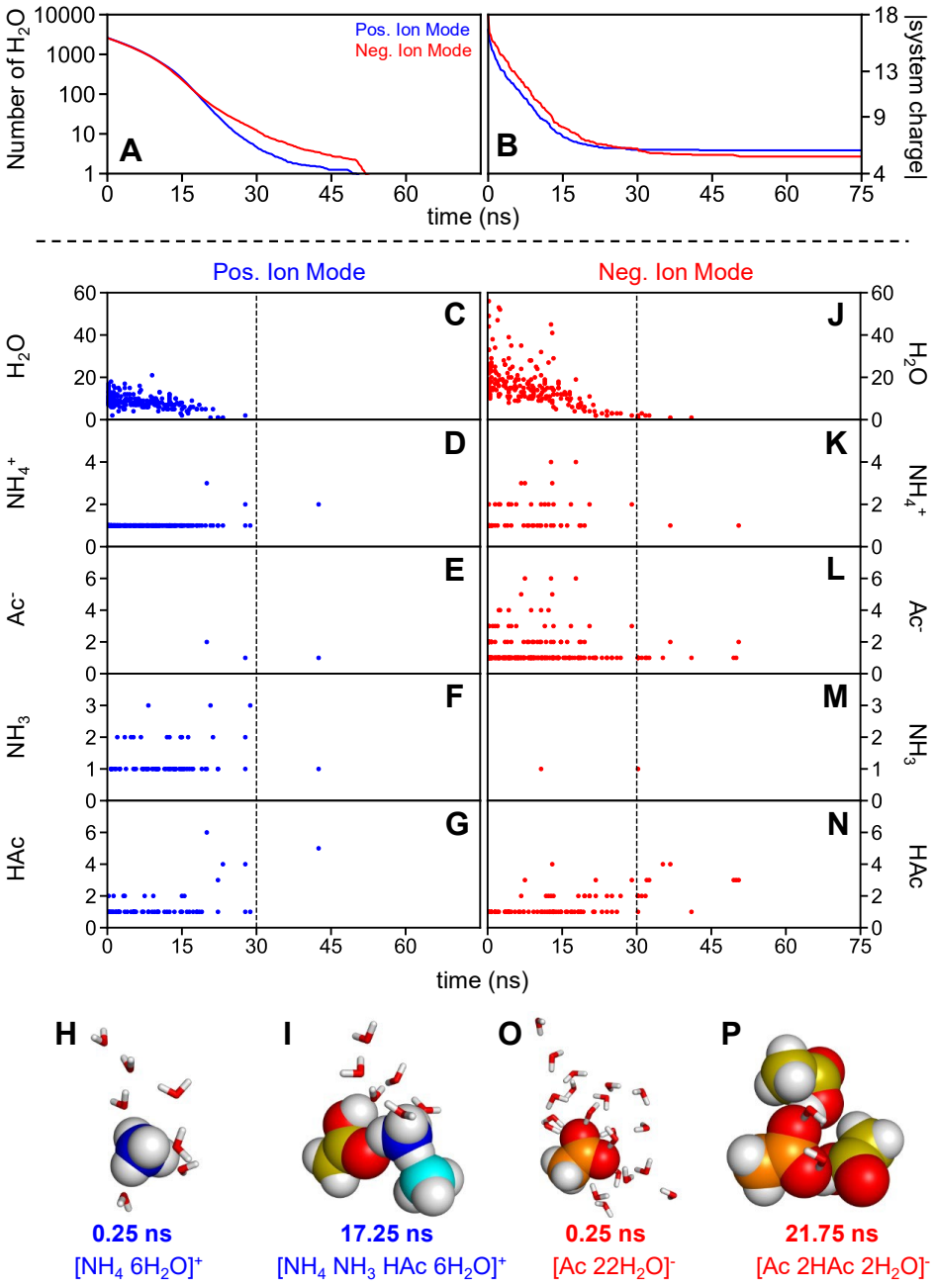

Figure S11. Charge ejection details for r_0 = 3 nm droplets containing ubiquitin, in positive (blue) and negative ion mode (red) from 20 dMPMD simulations for each condition. (A) Average number of H_2O in the droplet. (B) Average total system charge. (C-G) Vertically aligned dots represent the cluster composition for ejection events in positive ion mode, i.e., the number of H_2O, NH_4^+, Ac^-, NH_3, and HAc in each cluster. All clusters ejected in positive ion mode had a 1+ charge. (H, I) Typical cluster structures in positive ion mode. (J-P) Same as above, but for negative ion mode. Cluster charges in negative ion mode were 1- (94%), 2- (5.6%), and 3- (0.4%). Vertical dashed lines highlight differences past t = 30 ns: 1 charge ejection event in positive ion mode vs. 14 charge ejection events in negative ion mode.

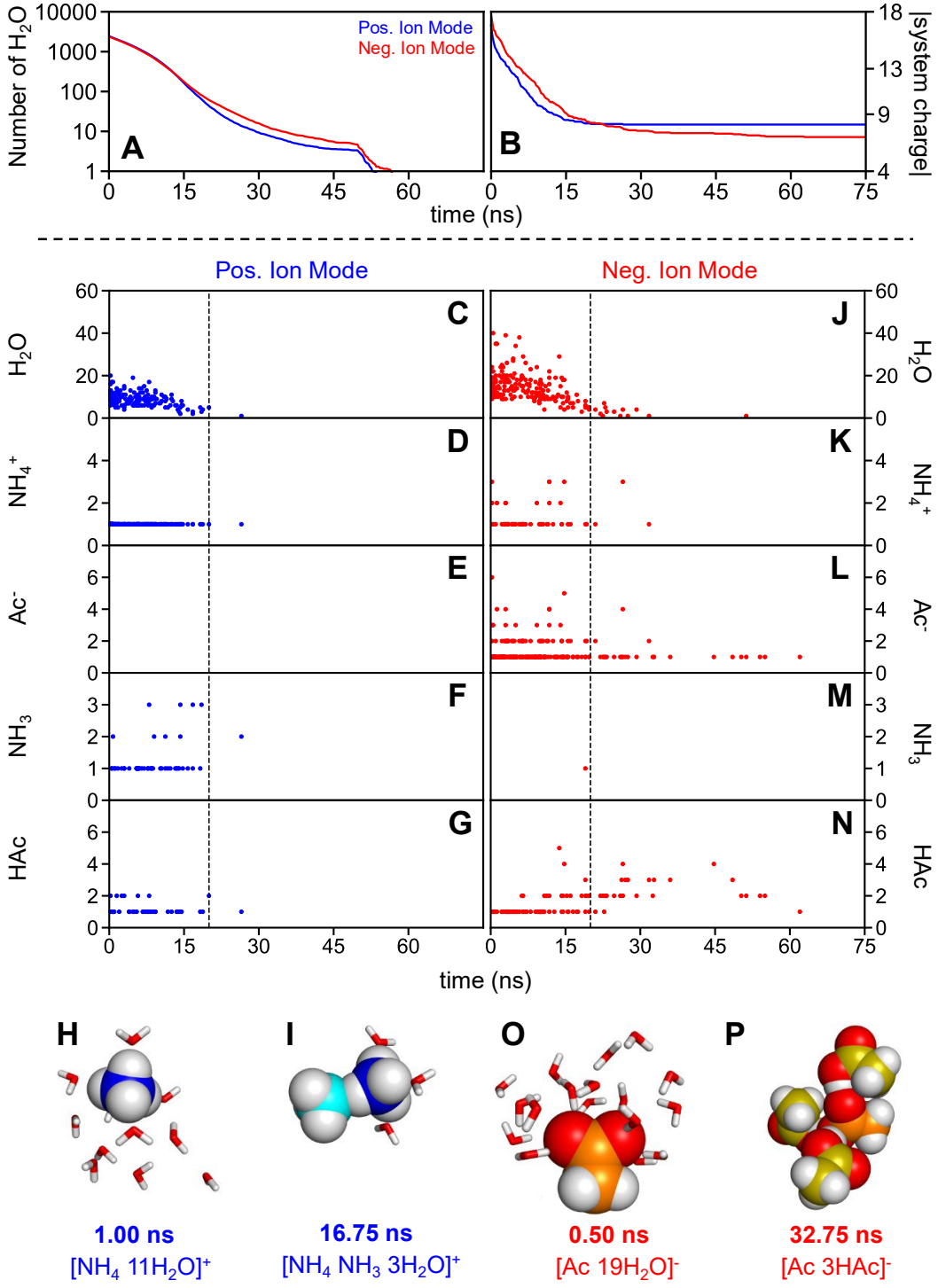

Figure S12. Same as Figure S11, but for lysozyme. All clusters ejected in positive ion mode had a 1+ charge. Cluster charges in negative ion mode were 1- (96.6%), 2- (2.9%), and 3- (0.5%). Vertical dashed lines highlight differences past t = 20 ns: 2 charge ejection events in positive ion mode vs. 25 charge ejection events in negative ion mode.

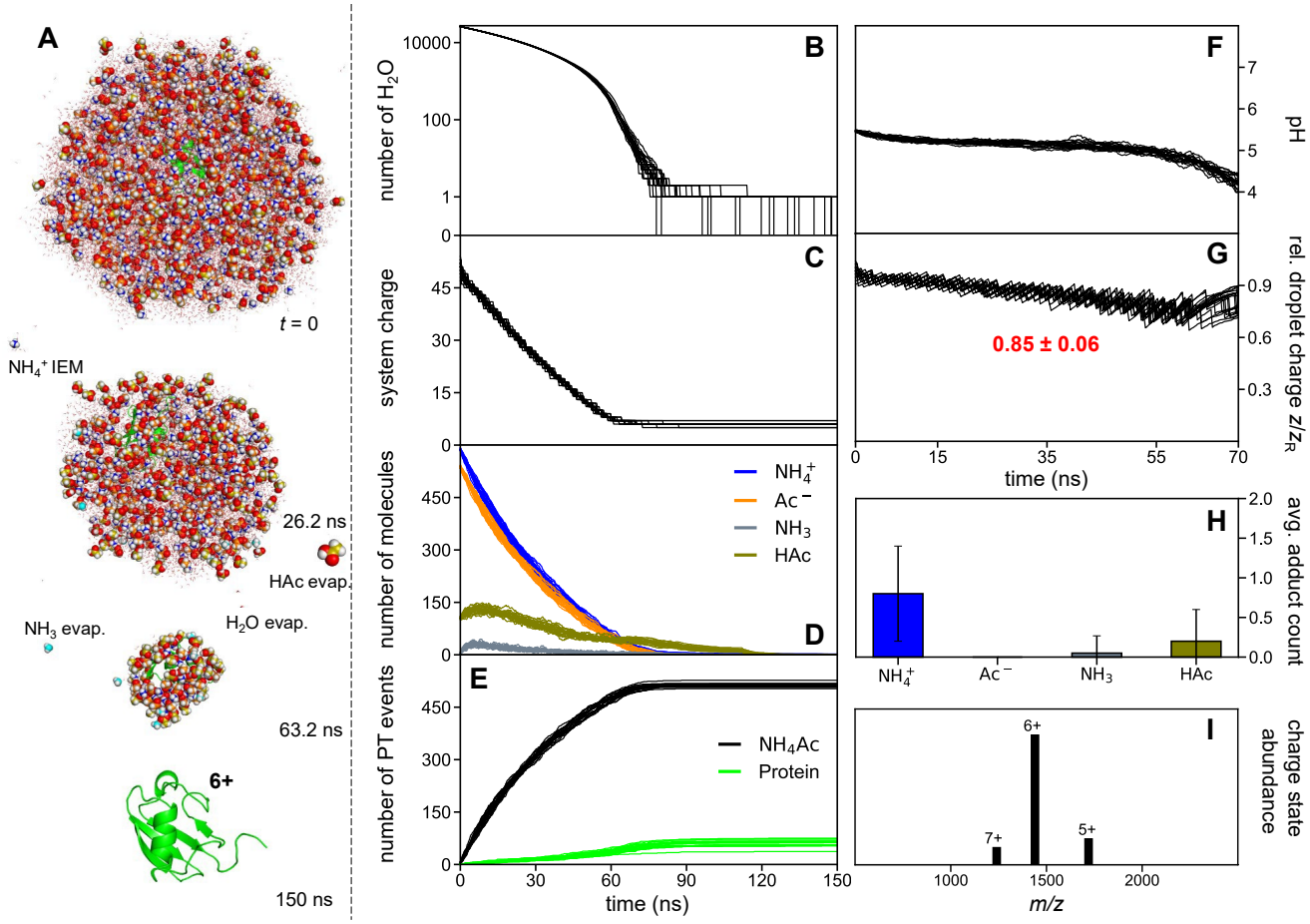

Figure S13. Ubiquitin native ESI (dMPMD) simulations in r_0 = 6 nm aqueous NH_4Ac droplets, using positive ESI mode. (A) Representative trajectory snapshots. Panels B-I summarize data from 20 replicates. For additional details, see the caption of Figure 2 (main text).

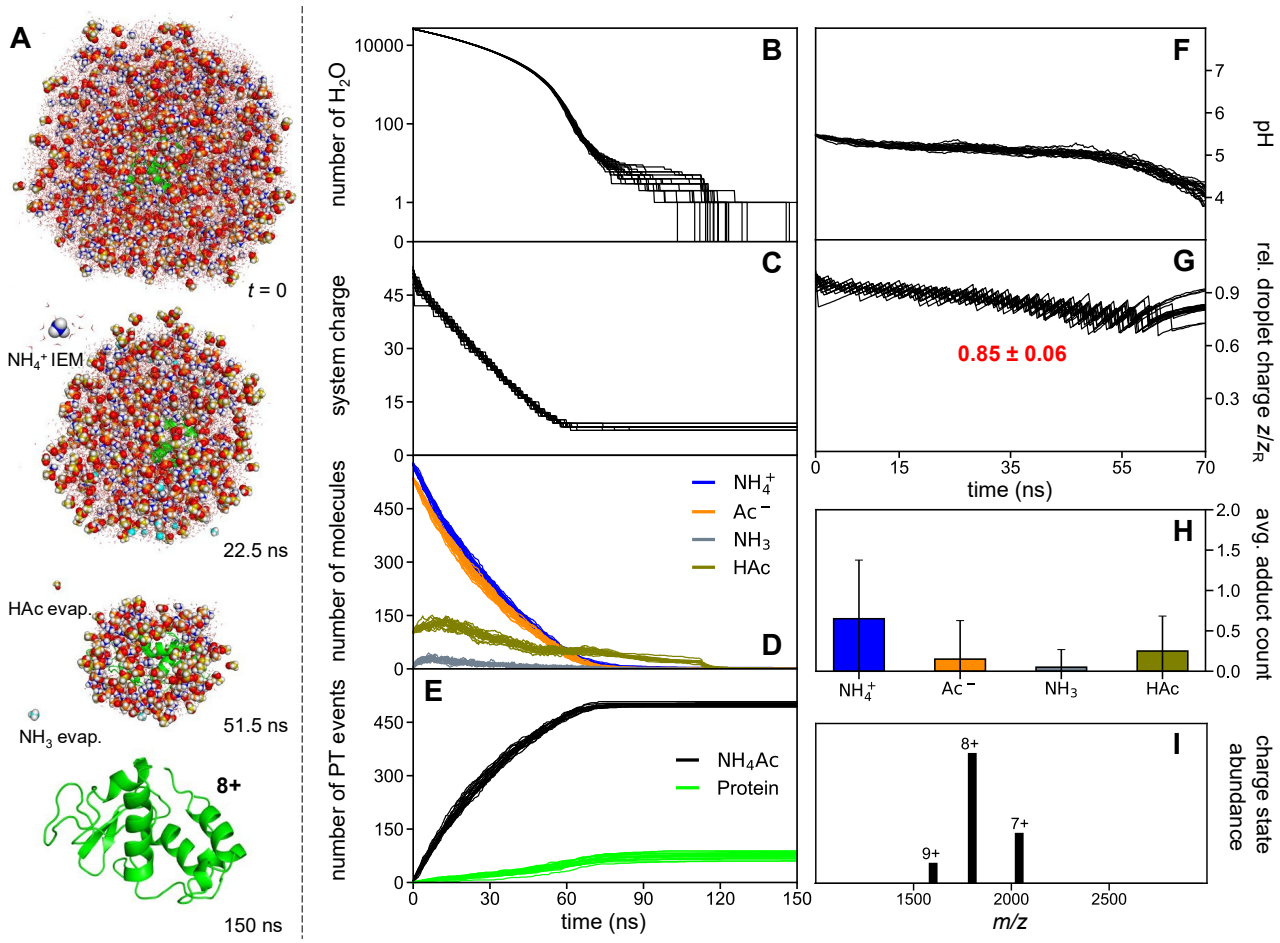

Figure S14. Lysozyme native ESI (dMPMD) simulations in r_0 = 6 nm aqueous NH_4Ac droplets, using positive ESI mode. (A) Representative trajectory snapshots. Panels B-I summarize data from 20 replicates. For additional details, see the caption of Figure 2 (main text).

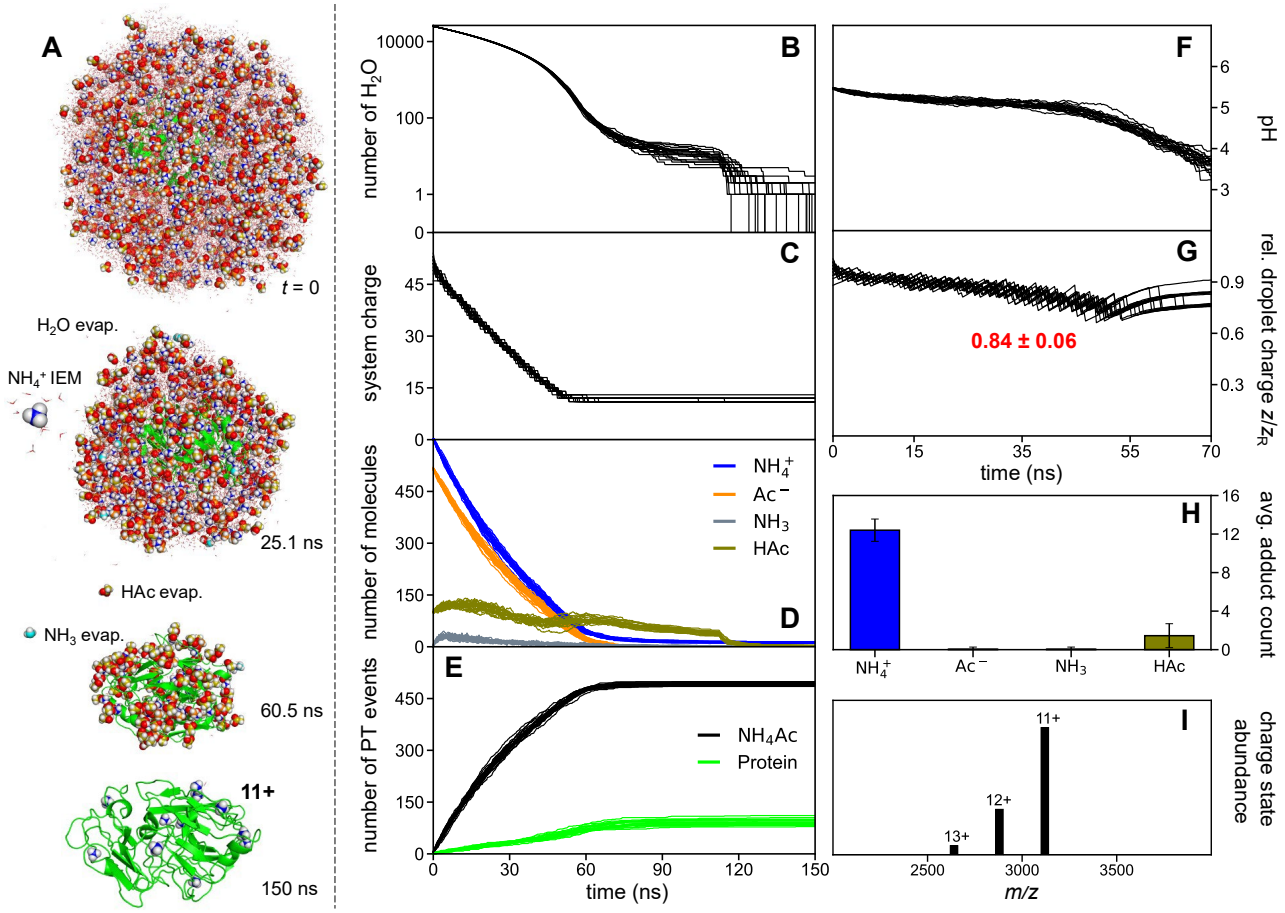

Figure S15. Pepsin native ESI (dMPMD) simulations in r_0 = 6 nm aqueous NH_4Ac droplets, using positive ESI mode. (A) Representative trajectory snapshots. Panels B-I summarize data from 20 replicates. For additional details, see the caption of Figure 2 (main text).

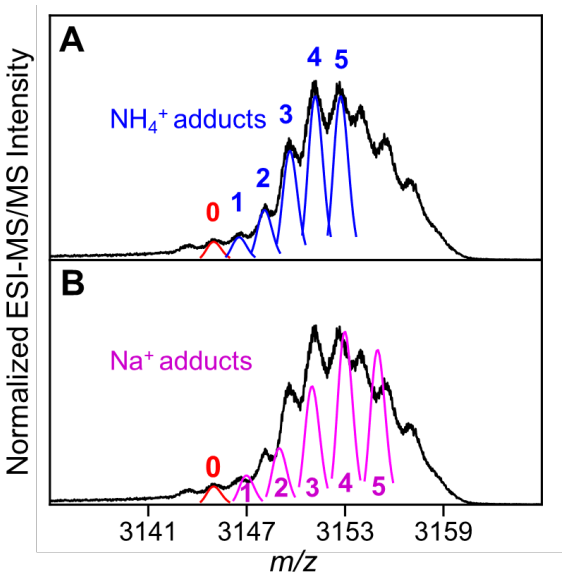

Figure S16. Both panels in this figure show a close-up of pepsin 11+ ions generated in native ESI experiments (black solid lines, identical to the data in Figure 5E of the main text). (A) The red profile represents the expected adduct-free protein mass distribution, with one phosphorylation. Blue profiles represent ammonium adducts $[M + (z\text{-}n)H + n NH_4]^{z+}$ with $z = 11$, where $n = 1, 2, \ldots$ represents the number of NH_4^+. (B) Same as in panel B, except that magenta profiles represent sodium adducts $[M + (z\text{-}n)H + n Na]^{z+}$, where $n = 1, 2, \ldots$ represents the number of Na^+. The NH_4^+ model in panel A provides an excellent fit, while the Na^+ model in panel B does not match the experimental data. This analysis confirms that pepsin adducts in these spectra are caused by NH_4^+ rather than Na^+. Isotope models were generated using Waters MassLynx.

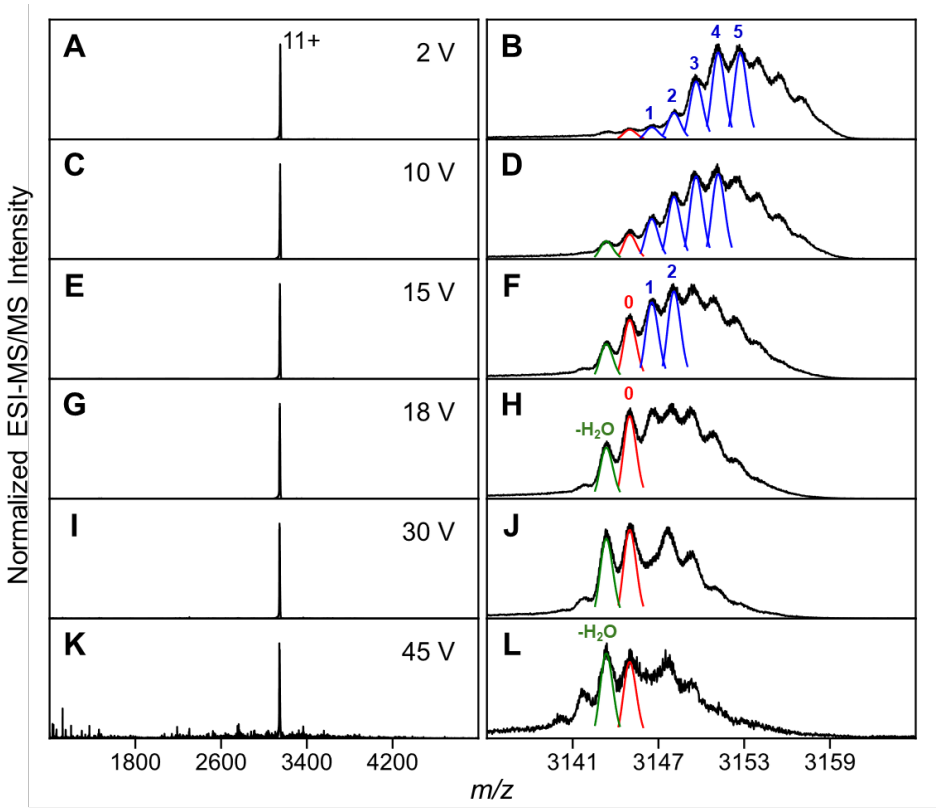

Figure S17. MS/MS of pepsin 11+ ions generated by native ESI. Data were acquired with different levels of trap cell collisional activation. Collision voltages are as indicated. Complete spectra on the left demonstrate that activation up to ~30 V only produces neutral losses, evident from the absence of charge-reduced proteins. Panels on the right show close-ups of the 11+ signal. Isotope models were generated as outlined in Figure S16. Red profiles indicate the expected adduct-free protein mass distribution, with one phosphorylation. Blue profiles represent NH_4^+ adducts. Collisional activation also promoted water loss (dark green), a common scenario for S/T-containing proteins (the S/T content of pepsin is 21%).[14]

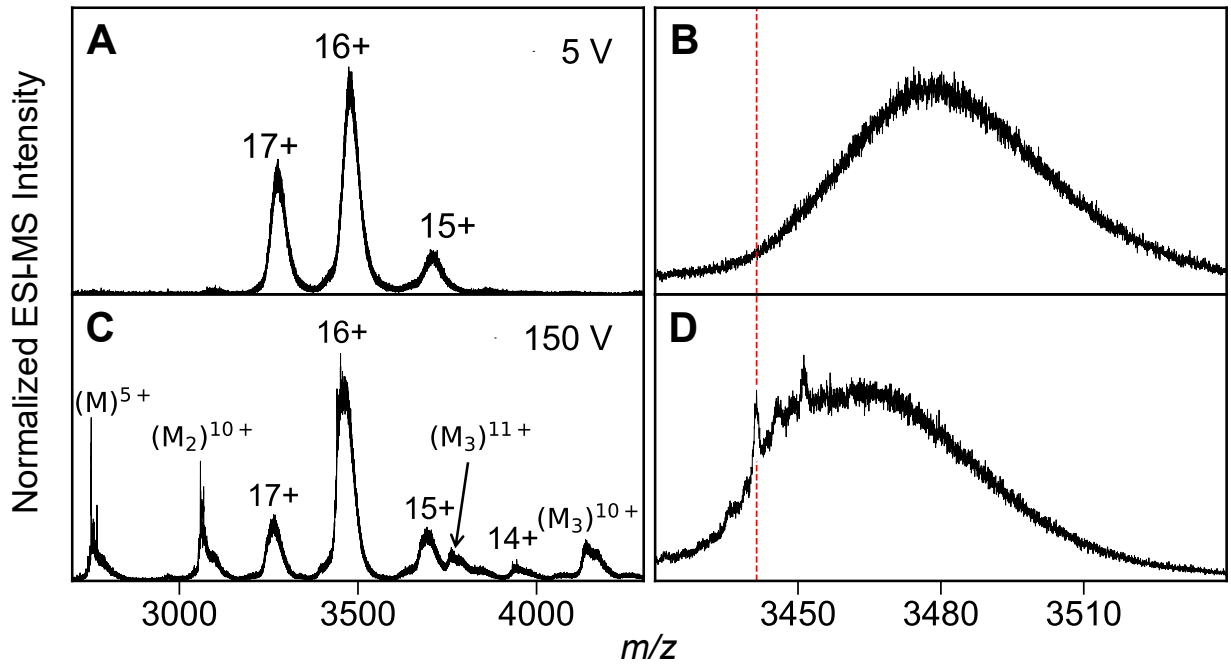

Figure S18. Native ESI mass spectra of TTR acquired (A, B) using gentle conditions with sampling cone at 5 V, and (C, D) using harsh conditions with sampling cone 150 V. Close-ups of the 16+ tetramer signal in panels B and D demonstrate how source activation promotes adduct loss, resulting in a shift to lower mass. Red dashed lines indicates the expected peak position of adduct-free tetramer ions. Charge state annotations 17+, 16+, etc. refers to TTR tetramers. Source CID products in panel C are donated as M = monomers, M_2 = dimers, and M_3 = trimers.

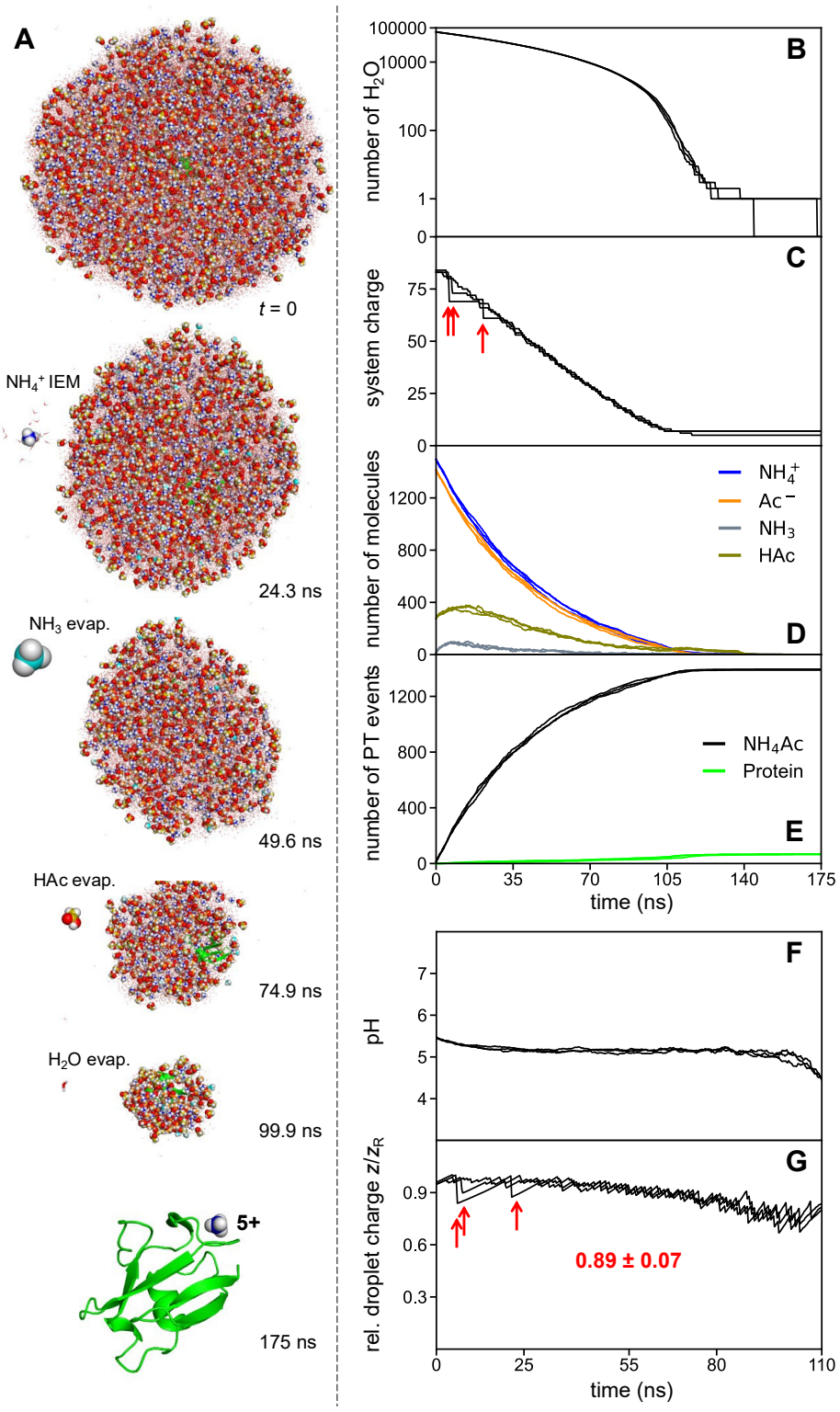

Figure S19. Ubiquitin native ESI (dMPMD) simulations in r_0 = 8.5 nm aqueous NH_4Ac droplets, using positive ESI mode. (A) Representative trajectory snapshots. (B-G): Summary of data from 3 replicates. Red arrows in panels C and G indicate jet fission events, one of which is examined in more detail in Figure 6 (main text).

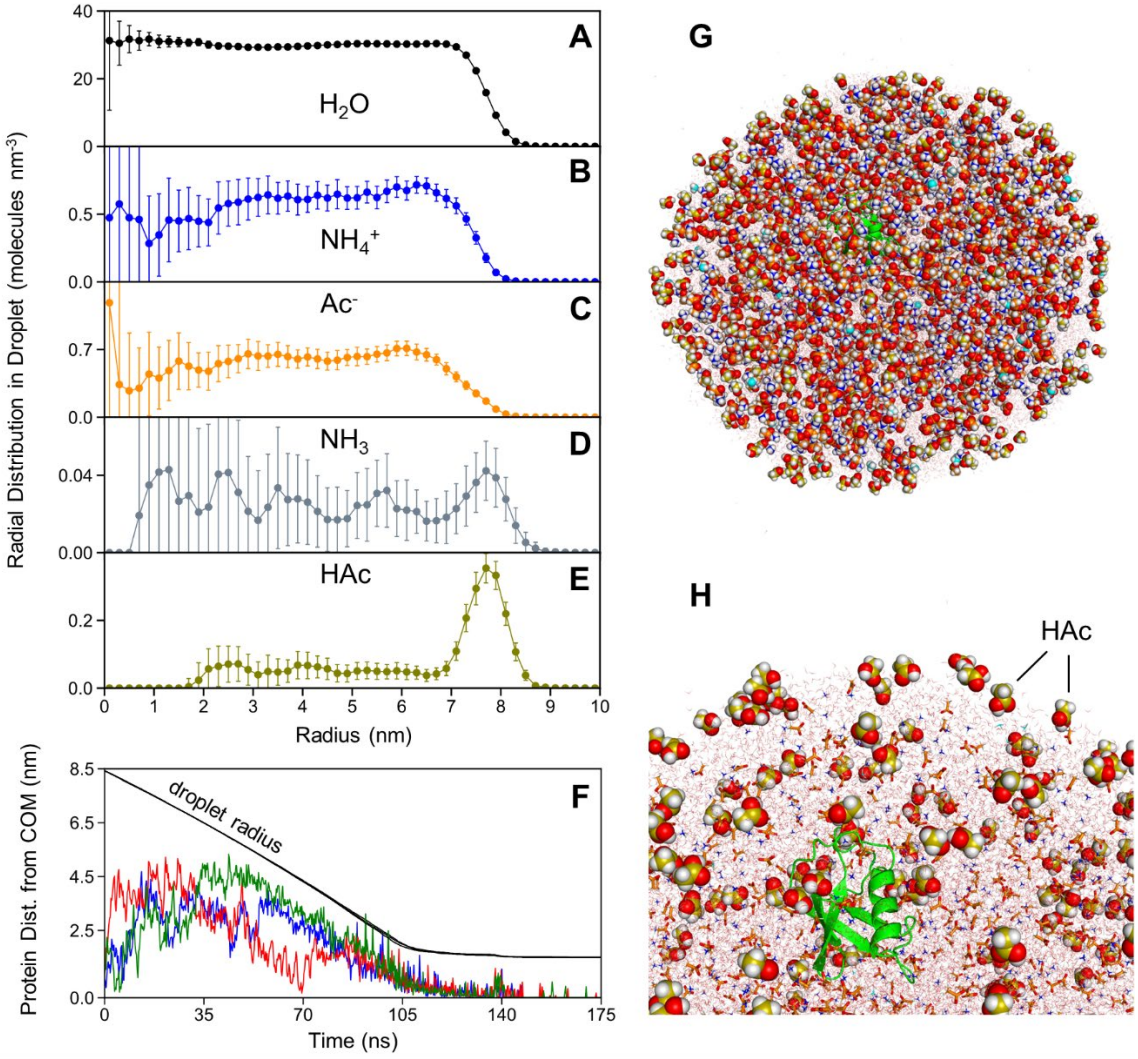

Figure S20. Details of r_0 = 8.5 nm ESI droplets containing ubiquitin. (A-E) Radial distribution functions of droplet components for a typical data set, averaged over 126 time points covering the 12.25–12.5 ns range. Error bars represent standard deviations. (F) Protein center-of-mass (COM) distance from the droplet COM for three independent runs. Also shown is the effective droplet radius that decreases over time as the solvent evaporates. (G) Trajectory snapshot at 12.5 ns. NH_4^+, Ac^-, NH_3, and HAc are shown in spacefill representation, H_2O molecules are shown as lines, the protein is colored green. (H) Close-up of a droplet section from panel G. Only HAc molecules are displayed in spacefill representation.

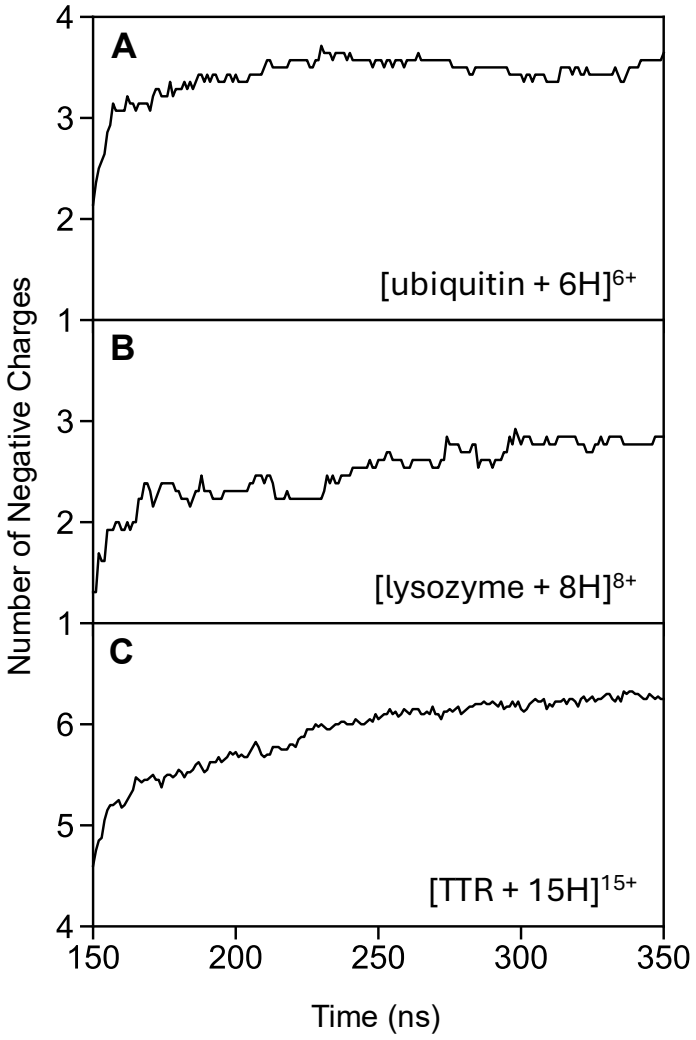

Figure S21. Increase in zwitterionic character. i.e., number of R-COO$^-$ sites, in the gas phase after protein release from r_0 = 6 nm ESI droplets. The data displayed here were averaged for multiple gMPMD runs (15 for [ubiquitin + 6H]$^{6+}$, 13 for [lysozyme + 8H]$^{8+}$, and 10 for [TTR + 15H]$^{15+}$). The remaining simulations (out of 20 for each protein) generated charge states other than those considered here. Overall, the number of negative charges during the 200 ns phase simulation window changed as follows: 2.1 → 3.6 for ubiquitin, 1.3 → 2.8 for lysozyme, and 4 × 4.6 → 4 × 6.3 for TTR tetramers.

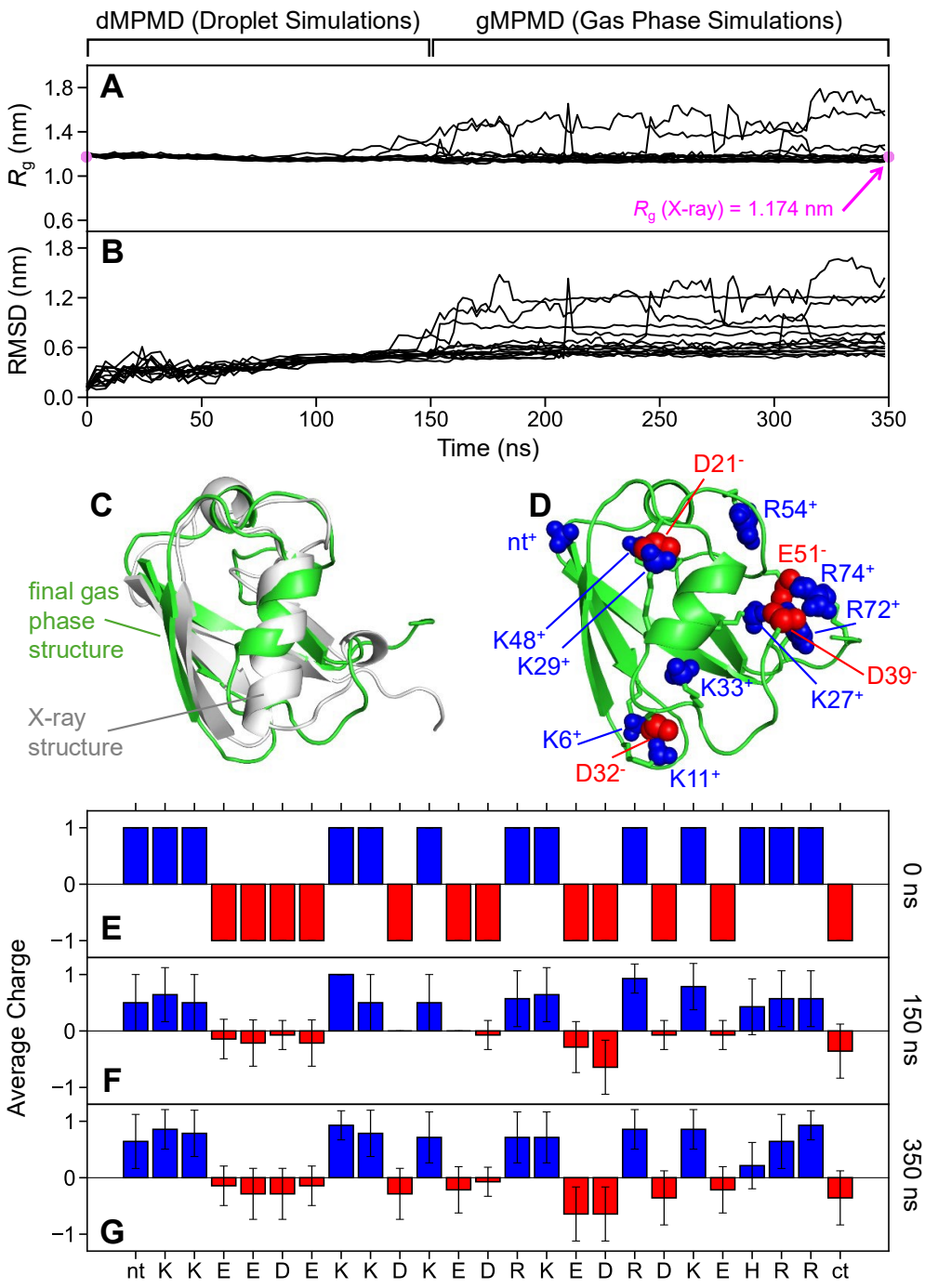

Figure S22. Structural evolution of [ubiquitin + 6H]$^{6+}$ during release from r_0 = 6 nm droplets (0-150 ns), and in the gas phase (150-350 ns). (A) Radius of gyration R_g. (B) Backbone RMSD relative to the initial X-ray structure. (C) Initial X-ray structure and representative compact final gas-phase structure. (D) Representative final gas-phase structure with positive (blue) and negative sites (red). All negative charges are involved in salt bridges. (E) Initial charge pattern in the droplet, and at the start and end (F, G) of gas-phase simulations. Error bars illustrate the variability of charge patterns for 15/20 runs; the remaining runs generated charge states other than 6+.

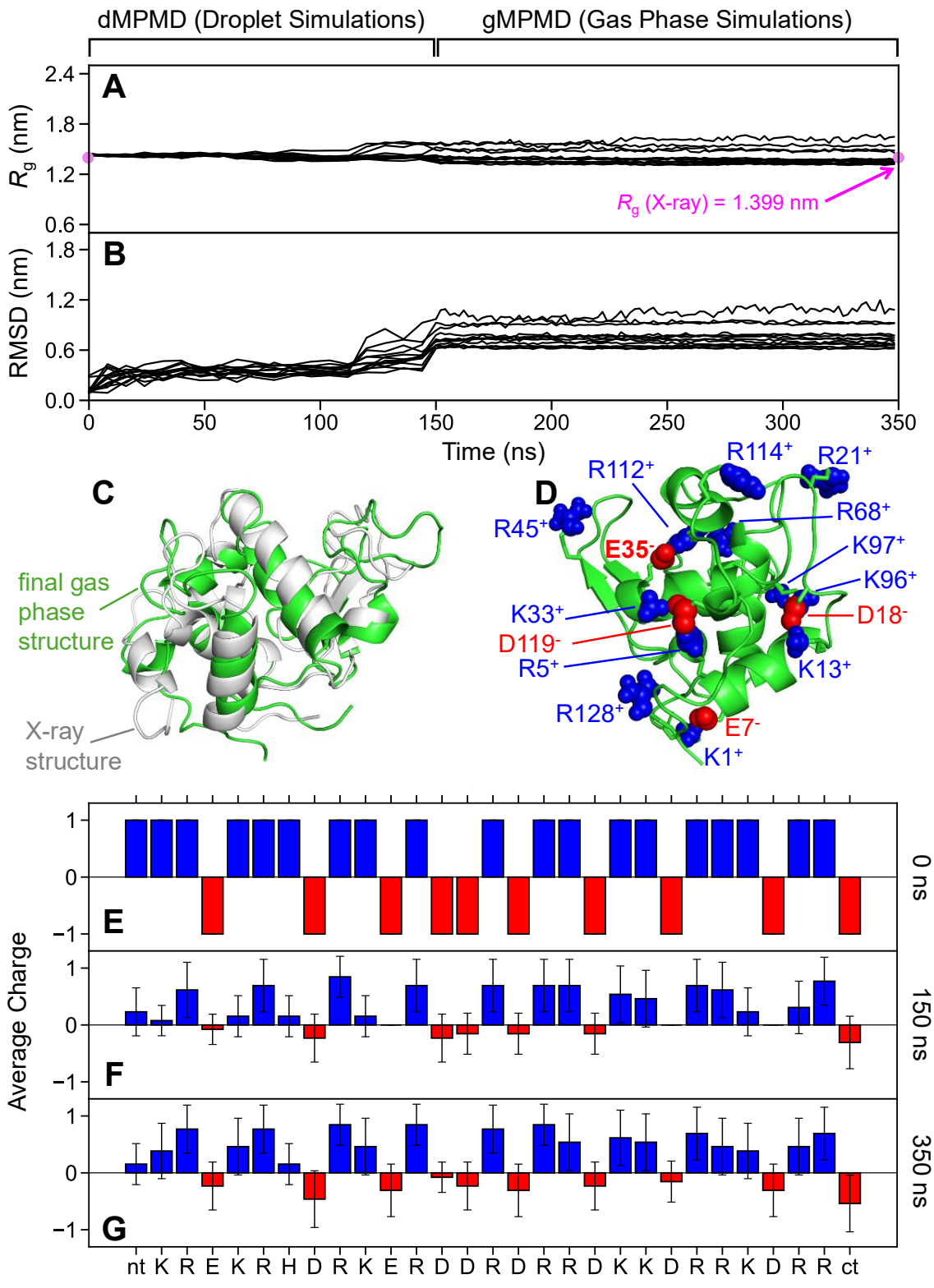

Figure S23. Same as in Figure S22, but for [lysozyme + 8H]$^{8+}$. The data shown are from 13/20 runs that generated a 8+ protein charge.

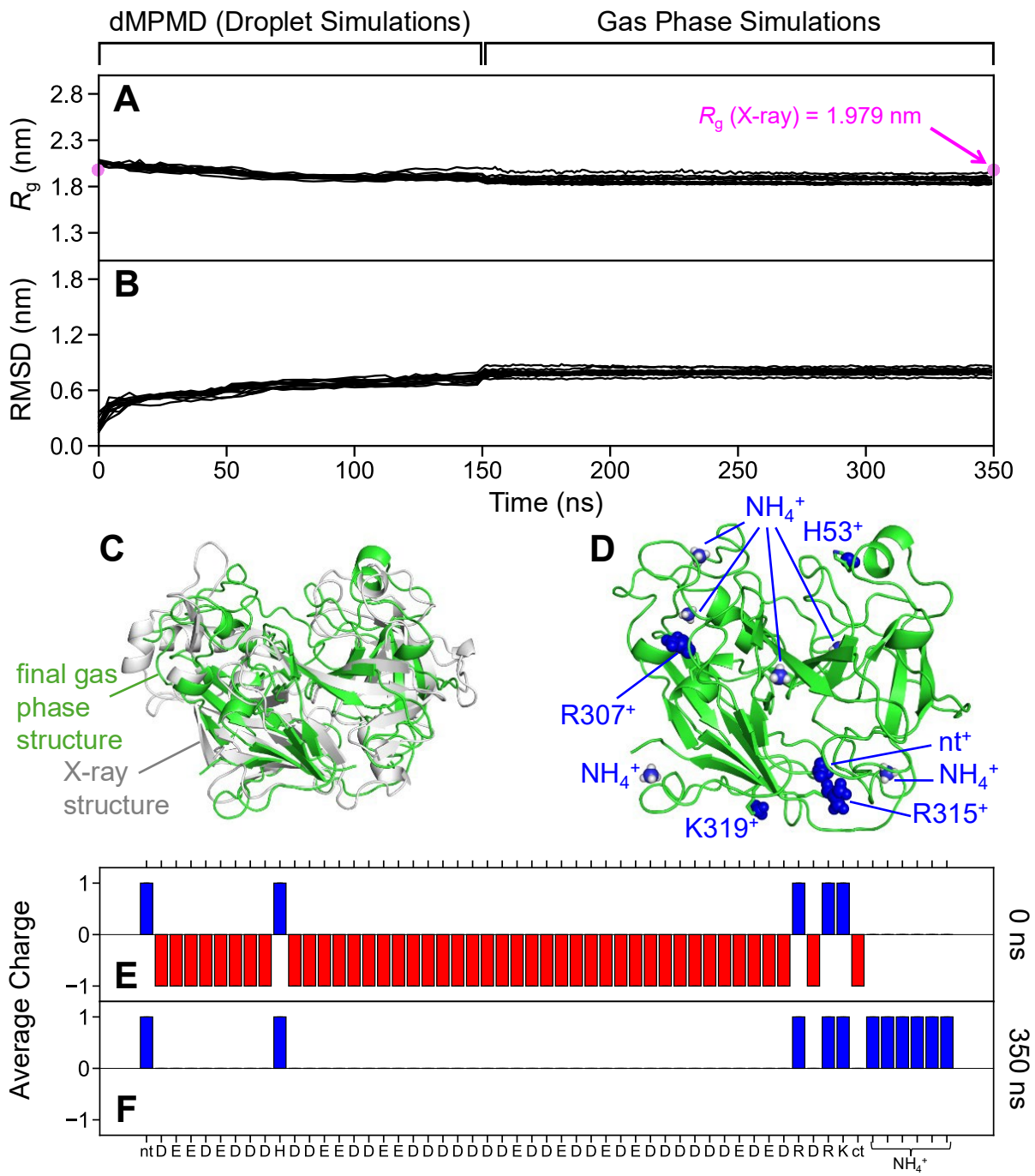

Figure S24. Structural evolution of [pepsin + 5H + 6NH4]$^{11+}$ during release from r_0 = 6 nm ESI droplets (0-150 ns), and in the gas phase (150 – 350 ns). (A) Radius of gyration R_g. (B) Backbone RMSD relative to the initial X-ray structure. (C) Initial X-ray structure and representative final gas-phase structure. (D) Representative final gas-phase structure with highlighted positively (blue) charged sites, including NH_4^+ adducts. There are no anionic sites because all carboxylates are protonated. Unlike for the other proteins examined in this work, gaseous pepsin ions were modeled with static protons. Six NH_4^+ adducts were retained to ensure a 11+ overall charge. (E) Initial charge pattern in the droplet, and (F) during the gas-phase simulations. The data shown here reflect the results of 14/20 runs; the remaining simulations generated charge states other than 11+.

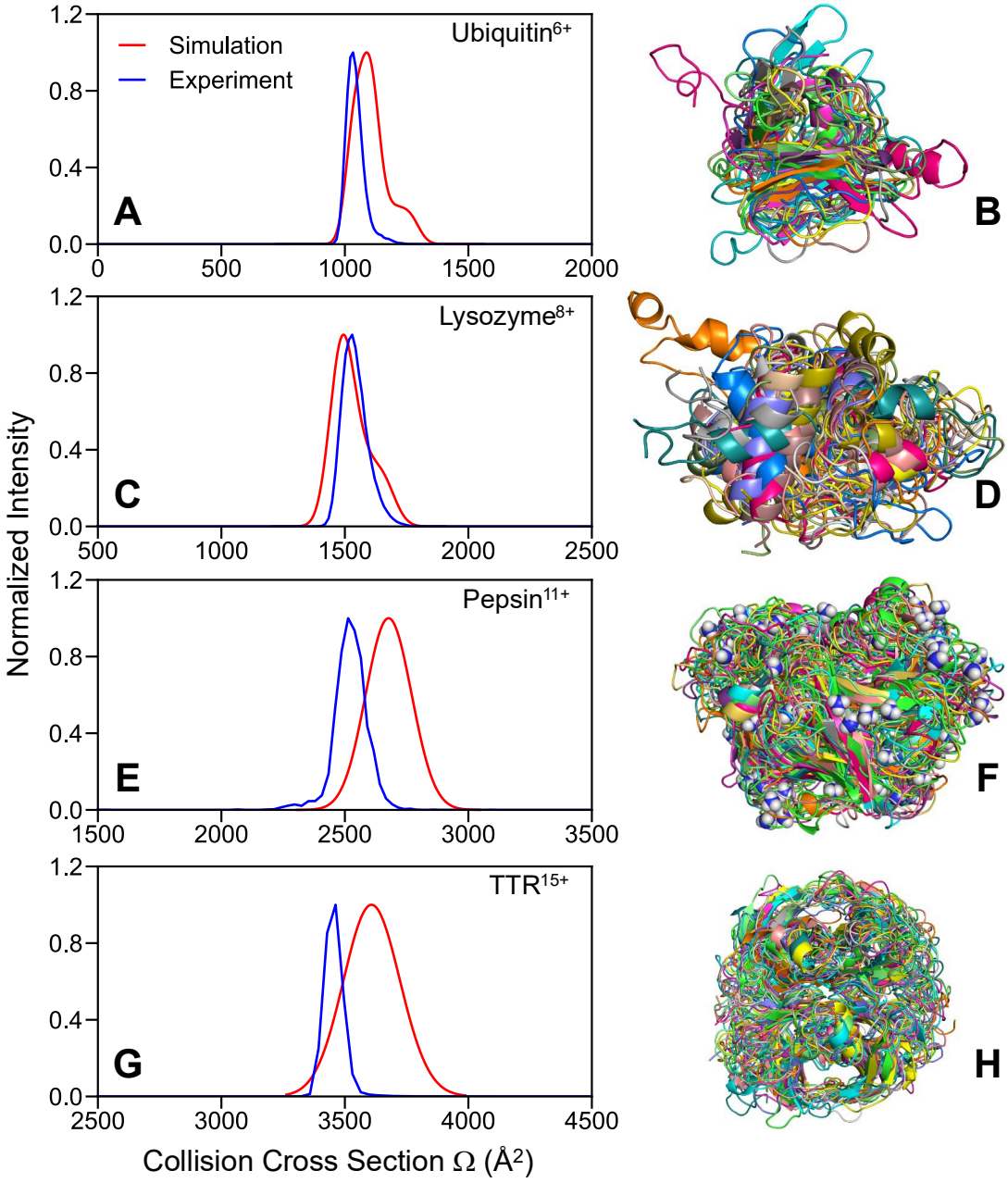

Figure S25. Experimental collision cross sections compared with the final structures produced in dMPMD/gMPMD simulations. An overlay of all simulated structures (for the charge states of interest) is shown along the right. (A, B) ubiquitin^{6+}, (C, D) lysozyme^{8+}, (E, F) pepsin^{11+}, and (G, H) TTR^{15+}. Travelling-wave IMS experiments were conducted on a SYNAPT G2 Q-TOF in He/N_2 buffer gas.[20] Drift times were converted to He collision cross sections (Ω) using a calibrant mix.[8] Ω values of simulated gas-phase structures were calculated using Collidoscope.[21] For comparison with TWIMS experiments, the contribution of each simulated protein structure was modeled as a Gaussian band centered at the corresponding Collidoscope Ω, with resolution $\Omega/\Delta\Omega = 14$. All Gaussian bands for any given protein were then added to generate the red profiles.

SI References

1. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19-25.
2. L. Konermann, H. Metwally, R. G. McAllister and V. Popa, *Methods*, 2018, **144**, 104-112.
3. W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695-1697.
4. K. Hanifi, P. M. Scrosati and L. Konermann, *J. Phys. Chem B*, 2024, **128**, 5973-5986.
5. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952-962.
6. L. Konermann, Z. Y. Liu, Y. Haidar, M. J. Willans and N. A. Bainbridge, *Anal. Chem.*, 2023, **95**, 13957-13966.
7. C. C. Moore, V. N. Staroverov and L. Konermann, *J. Phys. Chem. B*, 2023, **127**, 4061-4071.
8. M. Bakhtiari and L. Konermann, *J. Phys. Chem. B*, 2019, **123**, 1784-1796.
9. S. Vijay-Kumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531-544.
10. P. J. Artymiuk, C. C. F. Blake, D. W. Rice and K. S. Wilson, *Acta Crystallogr. Sect. B-Struct. Commun.*, 1982, **38**, 778-783.
11. A. R. Sielecki, A. A. Fedorov, A. Boodhoo, N. S. Andreeva and M. N. G. James, *J. Mol. Biol.*, 1990, **214**, 143-170.
12. L. D. Palmieri, L. Lima, J. B. B. Freire, L. Bleicher, I. Polikarpov, F. C. L. Almeida and D. Foguel, *J. Biol. Chem.*, 2010, **285**, 31731-31741.
13. A. G. Ferrige, M. J. Seddon, B. N. Green, S. A. Jarvis and J. Skilling, *Rapid Commun. Mass Spectrom.*, 1992, **6**, 707-711.
14. A. G. Harrison, *J. Am. Soc. Mass Spectrom.*, 2012, **23**, 116-123.
15. T. Nakanishi, T. Sato, S. Sakoda, M. Yoshioka and A. Shimizu, *Biochim. Biophys. Acta*, 2004, **1698**, 45-53.
16. L. Novotna, M. Hruby, M. J. Benes and Z. Kucerova, *J. Sep. Sci.*, 2008, **31**, 1662-1668.
17. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Zemgulyte, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Zídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis and J. M. Jumper, *Nature*, 2024, **630**, 24.
18. J. A. Lemkul, *Living J .Comp. Mol. Sci*, 2019, **1**, 5068.
19. M. S. Cordes and E. S. Gallagher, *J. Am. Chem. Soc.*, 2025, **147**, 15066-15076.
20. K. Giles, J. P. Williams and I. Campuzano, *Rapid Commun. Mass Spectrom.*, 2011, **25**, 1559-1566.
21. S. A. Ewing, M. T. Donor, J. W. Wilson and J. S. Prell, *J. Am. Soc. Mass Spectrom.*, 2017, **28**, 587-596.