analytical chemistry

pubs.acs.org/ac

Article

Linking Gas-Phase and Solution-Phase Protein Unfolding via Mobile Proton Simulations

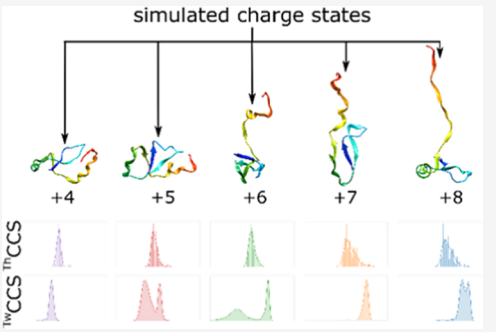
Charles Eldrid, Tristan Cragnolini, Aisha Ben-Younis, Junjie Zou, Daniel P. Raleigh, and Konstantinos Thalassinos*

Cite This: *Anal. Chem.* 2022, 94, 16113−16121

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Native mass spectrometry coupled to ion mobility (IM-MS) combined with collisional activation (CA) of ions in the gas phase (*in vacuo*) is an important method for the study of protein unfolding. It has advantages over classical biophysical and structural techniques as it can be used to analyze small volumes of low-concentration heterogeneous mixtures while maintaining solution-like behavior and does not require labeling with fluorescent or other probes. It is unclear, however, whether the unfolding observed during collision activation experiments mirrors solution-phase unfolding. To bridge the gap between *in vacuo* and *in-solution* behavior, we use unbiased molecular dynamics (MD) to create *in silico* models of *in vacuo* unfolding of a well-studied protein, the N-terminal domain of ribosomal L9 (NTL9) protein. We utilize a mobile proton algorithm (MPA) to create 100 thermally unfolded and coulombically unfolded *in silico* models for observed charge states of NTL9. The unfolding behavior *in silico* replicates the behavior in-solution and is in line with the *in vacuo* observations; however, the theoretical collision cross section (CCS) of the *in silico* models was lower compared to that of the *in vacuo* data, which may reflect reduced sampling.



The study of protein unfolding is essential for defining protein stability and provides important insight into protein aggregation in protein misfolding diseases[1] such as α1-antitrypsin deficiency,[2] transthyretin amyloidosis,[3] and β2 microglobulin amyloidosis.[4] Many techniques have been developed to study protein unfolding, including circular dichroism (CD),[5] nuclear magnetic resonance (NMR) spectroscopy,[6] electron paramagnetic resonance (EPR) spectroscopy,[7] fluorescence-based methods, and native mass spectrometry (MS) coupled to ion mobility (IM).[8−12] They all offer advantages and disadvantages, but with the exception of single-molecule methods, solution-phase methods generally have difficulties characterizing heterogeneous mixtures. While powerful, single-molecule methods require labeling with often large and bulky fluorophores, MS-based methods are particularly well suited to probing heterogeneous mixtures and have the advantage that small amounts of material are required, and modifications or labeling with probes is not required.

Native MS is widely used to probe the native-like state of proteins *via* soft-ionization techniques, such as nanoelectrospray ionization (nESI)[13−15] and enables measurement of the global protein fold,[16] ligand binding,[17] subunit composition of protein complexes[18] and proteoforms.[19] Ion mobility (IM) coupled to MS (IM-MS) adds an extra layer of information by separating isobaric protein ions *via* their 3-dimensional shape or collision cross section (CCS). IM functions by passing analyte ions through an inert buffer gas in a drift region. Ions of the same m/z but different conformation will be separated as more extended ions will experience a greater number of collisions with the buffer gas and so traverse the drift region more slowly than a more compact ion.[20,21]

IM-MS has been used to study protein dynamics and[18,22,23] domain organization[24] and to investigate the structural dynamics of disordered proteins.[25] Protein ions can be collisionally activated (CA) by increasing the energy by which they are introduced into the mobility region. They are often, but not always,[26] unfolded in a process called collision-induced unfolding (CIU). CA can give information on distinguishing features of monoclonal antibodies,[12,27,28] the number of domains within a protein,[11] and the thermal stability imparted by ligand binding.[17] Adding an extra stage of IM separation (tandem-IM), to select out particular conformers, allows even greater disambiguation of the unfolding pathways of proteins by selecting precursor ions.[21,26,29,30]

While collision activation is able to give important information about unfolding and native-like states that are retained in the gas phase[31−33] (*in vacuo*), it is not known whether gas-phase unfolding is comparable to unfolding in-

Received: August 2, 2022
Accepted: October 12, 2022
Published: November 9, 2022

solution (*in-solution*). As CCS is an inherently low-resolution structural parameter, complementary techniques are required to create a structural model of the unfolded protein *in vacuo*. Molecular dynamics (MD) simulations (*in silico*) are uniquely positioned to do so and can be coupled with IM-MS data; they provide atomistic detail, which is complementary to IM and can be benchmarked *via* comparison of experimental and theoretical CCS values.[34]

Simulations which replicate the gas-phase environment inside an IM drift cell are not as straightforward as simply simulating proteins without bulk solvent. Without the intramolecular coulombic repulsion brought about by charging, the extended states of gas-phase proteins are liable to collapse.[35,36] In positive nESI, charged sites occur on exposed ionizable sites, such as the N-terminus, lysines, arginines, and histidines and can migrate between these sites, maintaining dynamic equilibrium.[37,38] To account for these effects, frameworks which allow simulation of the dynamic protonation states of proteins in the gas-phase behavior have been developed.[39−42]

It is still unclear if unfolding *in vacuo* mimics the unfolding process *in-solution*, and it is also difficult to fully validate the assignment of *in vacuo* structures from *in silico* methods. To critically validate the methodology, a protein system for which there is detailed *in-solution* unfolding data must be used. To this end, we chose the N-terminal region of the ribosomal L9 protein from *Geobacillus stearothermophilus* as a model system.[43,44] The L9 protein comprises two distinct globular domains joined by an α-helical linker in a "dumbbell" shape. Both the N-terminal (NTL9)[45−51] and C-terminal (CTL9)[52−55] domains are stable in isolation and fold cooperatively. Each has had its *in-solution* unfolding explored in detail through fluorescence studies, CD, and NMR lineshape analysis. The structure of the N-terminal domain has been determined *via* X-ray crystallography and adopts the same fold in isolation as in the intact protein.[55] NTL9 is one of the simplest examples of the split $\beta-\alpha-\beta$ motif. The fold consists of a mixed α and β structure, with two α-helices sandwiching antiparallel 3 stranded β-sheet strands. The C-terminal helix of NTL9 forms part of the connection with the C-terminal domain, but there are no contacts between the N and C-terminal domains (Figure 1A−C).[44]

In this study, we combine native IM-MS and *in silico* unfolding of NTL9 to create an *in vacuo* model of unfolding. Upward of 100 repeats of an unbiased *in silico* method of thermal unfolding, using the approach described in Popa *et al.*,[40] were performed. The unfolded models match the *in-solution* unfolding and are in line with *in vacuo* data. While the models of unfolding are in good accordance, there are discrepancies between experimental and theoretical CCS values of the final models. Analysis of the deviations provides clues to important factors which may affect the analysis and comparison of *in-solution* and *in* vacuo unfolding.

METHODS

Sample Preparation. NTL9 was produced and purified as described previously.[52] Lyophilized NTL9 was dissolved in 100 mM ammonium acetate pH 7.5 to 50 μM and frozen at −20 °C. On the day of data collection, the sample was desalted by buffer exchange using Amicon ultra centrifugal filtration units (Merck Millipore, U.K.), 6 times centrifuged at 14.0E3 g for 15 min at 4 °C using a Heraeus Fresco 17 centrifuge. The concentration was analyzed by Qubit assay (Thermo Fisher Scientific, U.K.).

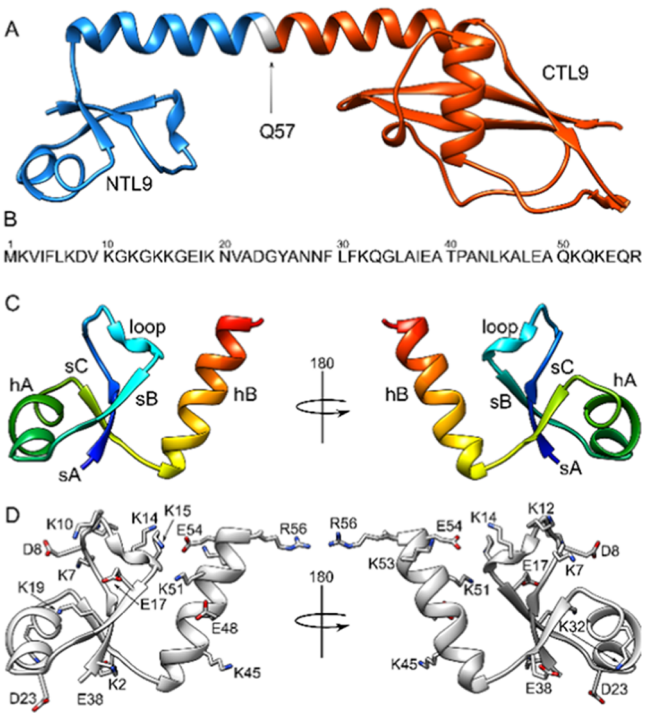


Figure 1. (A) Structure of the ribosomal L9 protein from *Geobacillus stearothermophilus* with the N-terminal construct (NTL9) colored in blue and the C-terminal construct (CTL9) colored in red. The Q57 residue, which is not included in either construct, is labeled. (B) NTL9 sequence. (C) NTL9 ribbon structure is colored in rainbow from N- to C-terminus, with the α-helices (hX) and β-strands (sY) labeled, as well as a loop that appears to show structure. (D) Location of the charged residues in NTL9. A ribbon structure is shown in gray with charged residues (aspartic acid (D), glutamic acid (E), lysine (K), and arginine (R)) in stick format, with nitrogens and oxygens in blue and red, respectively. Based on structure PDB ID: 1DIV.[44]

Data Collection. Samples were directly infused into the mass spectrometer using nESI from gold-coated capillaries prepared in-house using a Flaming Brown P97 needle puller (Sutter Instruments Co) and a Q150R S sputter coater (Quorum Tech, U.K.). Single-stage IM data was collected on a Synapt G1 (Waters Corp, U.K.) using the parameters presented in Table S1. CCS measuring by TWIMS ($^{TW}CCS_{N2\rightarrow He2}$, following notation described in Gabelica *et al.*[56]) calibration was performed using melittin (Sigma, U.K.), human insulin (Sigma, U.K.), ubiquitin (Sigma, U.K.), equine cytochrome C (Merck Millipore, U.K.), and β-Lactoglobulin (Sigma, U.K.). (Figure S2).

Simulations. The NTL9 structural model was created from residues 1−56 of the full L9 protein (PDB ID: 1DIV).[44] MD simulations were performed using Gromacs v2018.4. The simulation pipeline (see Figure S1) is as follows: the initial structure is checked for completeness, *i.e.*, all residues contain all atoms. The version of the mobile proton algorithm[40,41] (MPA) used during simulations requires nonchargeable side-chain residues at the N- and C-termini, so a C-terminal glycine residue was added *via* Modeller (v9.23); however, both the N- and C-termini remain chargeable. The Avidin model was supplied by the Konermann group and is based on PDB ID: 5IRU.[57] Initial protonated and deprotonated topology files were created using *pdb2gmx* with the OPLSAA force field;[58]

https://doi.org/10.1021/acs.analchem.2c03352
Anal. Chem. 2022, 94, 16113−16121

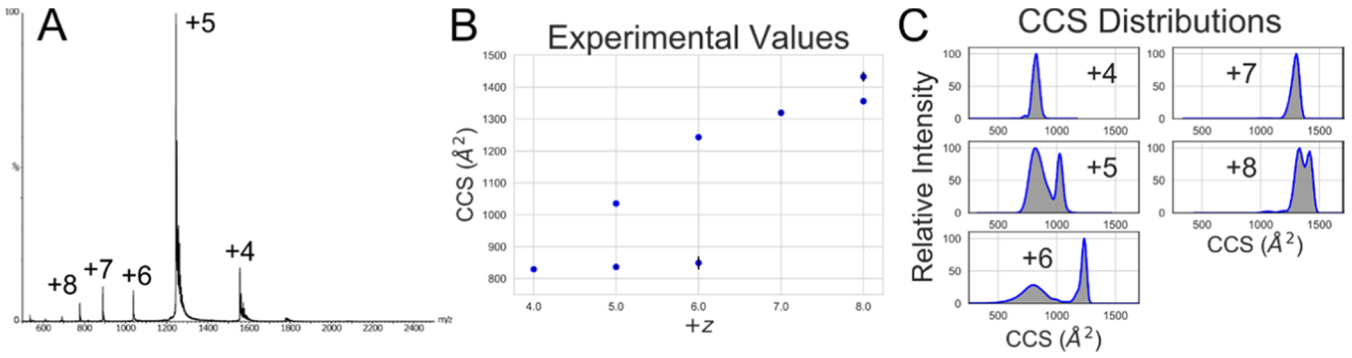


Figure 2. Representative experimental IM-MS data for NTL9. (A) Mass spectrum with charge states labeled; (B) average CCS peak top values; and (C) an example of CCS distributions.

the input values for the state of lysine, arginine, glutamate, aspartate, histidine, and termini were created using a python script (available at https://github.com/ThalassinosLab/charge_site_calculator). A charge library was created by copying the residue information from both the protonated and deprotonated topology files. A GROMACS and structure file was then created for the specific charge state. The MPA was used to distribute the protons across the structure, before equilibration. After equilibration and minimization (Tables S2−S4), a 20 ps simulation at the set temperature was run, after which the protons were rearranged. This cycle continued until the full time period had elapsed, which was either 4 ns at a set temperature (for thermal unfolding) or 100 ns (for coulombic unfolding) (Figure S1).

Data Analysis. IM-MS spectra were analyzed using MassLynx v4.1 (Waters Corp, U.K.) and DriftScope v2.1 (Waters Corp, U.K.). CCS calculations were performed using the spreadsheet available from http://www.homepages.ucl.ac.uk/~ucbtkth/resources.html, which uses the method of Thalassinos *et al.*[59,60] and $^{TW}CCS_{N2\rightarrow He2}$ values derived experimentally from first principles by Bush *et al.*[61] CCS distribution (CCSD) metrics intensity weight mean (IWM) and intensity weight standard deviation (IWSD) were calculated using the following equation

$$IWM_{CCS} = \frac{\sum_{i=1}^{n} I_i t_i}{\sum_{i=1}^{n} I_i}$$

where I is the intensity for each arrival time, t is the arrival time, and n is the number of data points in the arrival time axis.[62]

For the simulations, the theoretical CCS (^{TH}CCS) of each frame was calculated every 2 ps for the heating simulations and every 10 ps for the charging simulations using the IMPACT[63] pseudo-trajectory method, and the final structure was calculated using Collidoscope.[64] Salt-bridge analysis was performed using VMD (v1.9.2)[65] using a distance cutoff of 4 Å between the oxygen of an acidic residue and the nitrogen of a basic residue to identify salt-bridge pairings. Ensemble cluster analysis was performed using Chimera (v1.1.3).[66]

RESULTS

Summary of Structure and Solution Unfolding of NTL9. The folding and unfolding of NTL9 in solution occurs in a two-step process, progressing from a folded globular to a transition state, onward to the unfolded state.[46,47] During folding, 60−65% of the total native solvent accessible area is buried in the transition state.[45] The C-terminal helix, hB, does not form any electrostatic salt bridges to the globular structure, although there are potential intra-helical salt bridges. The last few residues of the C-terminal helix are frayed in solution, and the helix likely undergoes additional fraying during thermal denaturation prior to full unfolding of the globular structure. The C-terminal helix is also partially populated in isolation.[48] Removal of the final 5 residues, $_{51}$KQKEQR$_{56}$, destabilizes the domain.[51] Residues D8, E17, and D23 (Figure 1D) form interactions that are perturbed during unfolding.[48] D8 is in a partially ordered loop which includes 5 lysine residues ($_{7}$KDVKGKGKK$_{16}$) and may form electrostatic contacts with several different side chains, E17 contacts the amide group of K14, and D23 forms a strong salt bridge with the N-terminal amino group.[48] During thermal unfolding, the core of the structure consisting of the first helix and β-sheet likely, comprising the first 39 residues, unfolds after the unfolding/fraying of the C-terminal helix. The first 39 residues of the protein can fold in isolation but are prone to aggregate in solution.[48]

^{TW}CCS Analysis. NTL9 displays charge states ranging from +4 to +8, and there appear to be two overlapping charge state distributions (CSDs) (Figure 2A,B). IM-MS analysis shows that charge states from +4 to +6 have a compact form, and states from +5 to +8 have an extended form. This suggests that NTL9 occupies a compact state of approximately 840 Å^2, which is retained between +4 to +6, and extended states, which are maintained by intramolecular coulombic repulsion, going from 1030−1430 Å^2. The compact peaks for +5 and +6 show large FWHMs, suggesting that they contain a multiplicity of conformations.(Table 1)

Table 1. Experimental CCS Values

+z	$^{TW}CCS_{N2\rightarrow He2}$ (Å^2)
4	829 ± 3.24
5	837 ± 5.47, 1036 ± 2.15
6	850 ± 20.69, 1243 ± 2.65
7	1319 ± 6.97
8	1356 ± 5.01, 1433 ± 15.12

Activation of the NTL9 charge states leads to the +4 state coming slightly more extended; every other charge state either transitioned into a previously present extended state or did not unfold (Figure S3).

Thermal Unfolding Simulations. Initial thermal unfolding simulations were performed in triplicate on the +5 charge state, as it is experimentally the most intense and is the lowest

https://doi.org/10.1021/acs.analchem.2c03352
Anal. Chem. 2022, 94, 16113−16121

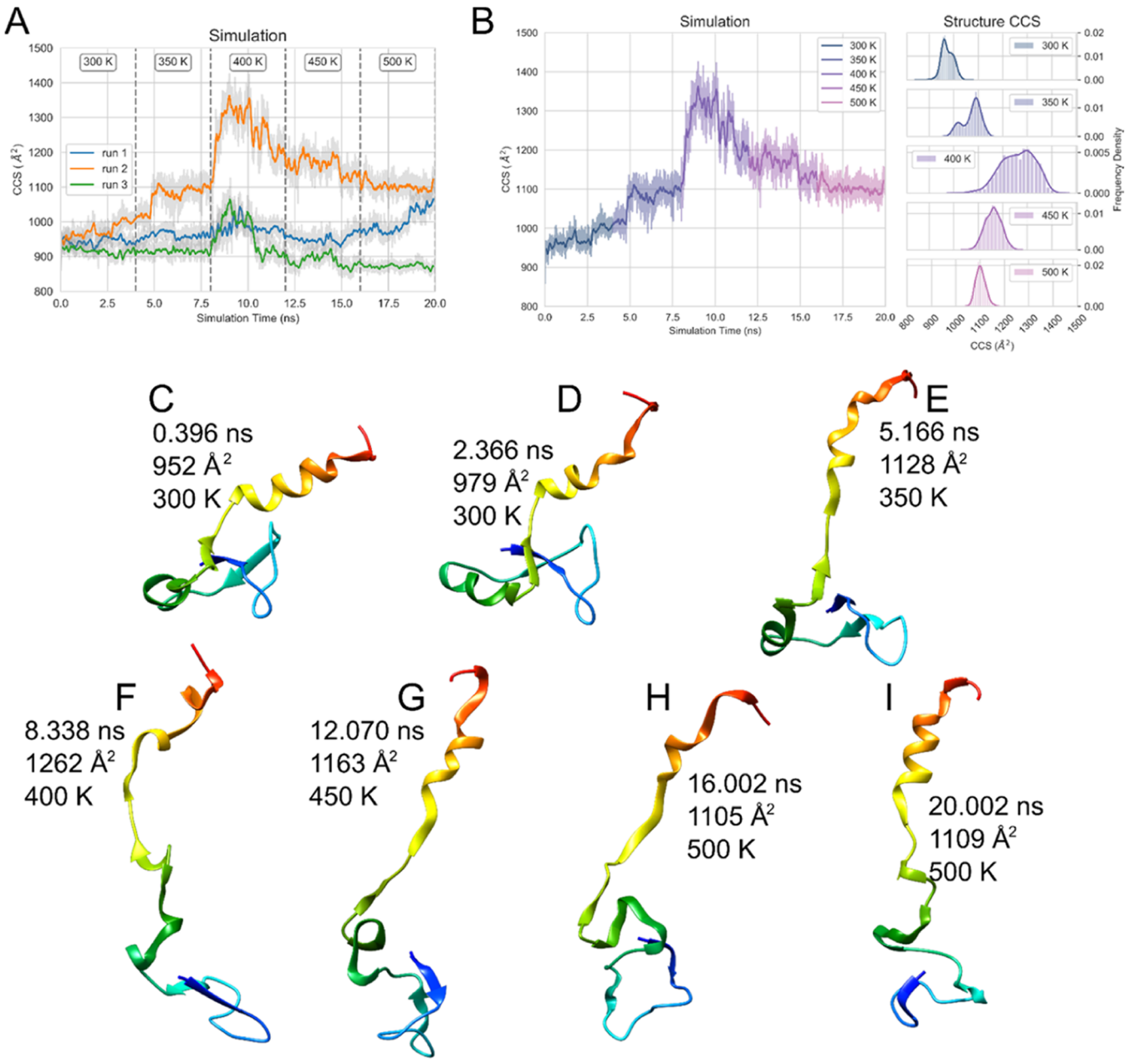


Figure 3. Simulation outcomes of the unfolding of the +5 charge state by heating. (A) Initial triplicate of simulations, with the different temperature transitions marked out with dotted lines. (B) Simulation of the +5 state, which displayed unfolding, with histograms of the states at each temperature shown to the right. The shading represents the unsmoothed data, and the solid line represents smoothed with a mean window of 50, once. (C–I) various unfolded states of the +5 model with the time elapsed, theoretical CCS, and simulation temperature.

charge state to display a clear conformational change during collision activation (Figure S3). The thermostat was increased by 50 K over 4 ns for a total simulation time of 20 ns.[40] One of the triplicates unfolded after approximately 8 ns once the temperature had increased to 400 K (Figure 3A,B); however, it recompacted under increased heating after the initial unfolding event.

The simulation suggests an unfolding pathway where there is C-terminal unfolding, which, for example, can be shown by the formation of a salt bridge between E38 and K2 at later frames in the simulation (Figure S4). Other salt bridges which are diagnostic of particular *in silico* conformations are E48/54 to K51, which are characteristic of the C-terminal helix (hB).

The *in silico* thermal unfolding simulations of the +5 charge state were repeated 100 times (Figure 4). Only 24 outcome structures had THCCS values >1100 Å^2 (Table S5), and cluster analysis suggests that 80% remain compact (Figure S6A,B Tables S5, and S7). Inspection of unique structures (structures that did not fall into a cluster ensemble) shows that only 13 structures display dissociation of hB from the protein core. This data suggests low reproducibility of *in silico* thermal unfolding. In many of the final states, structural rearrangement occurs to create a "flattened" structure, where there is a rearrangement of hA, sA, and sB, creating a set of interactions between a series of charged residues, including D8, K12, K14, K15, D48, D54, and R56.

https://doi.org/10.1021/acs.analchem.2c03352
Anal. Chem. 2022, 94, 16113−16121

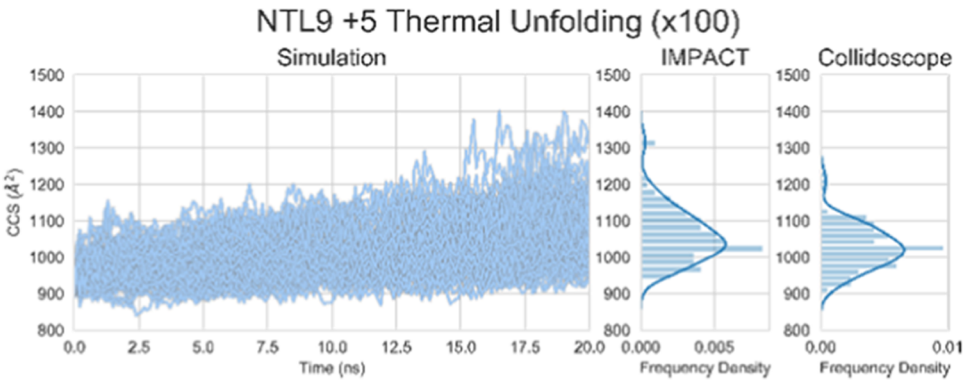


Figure 4. 100 replicates of the +5 *in silico* unfolding, showing the trace of the CCS of each frame as calculated by IMPACT, and then histograms of the final states calculated by IMPACT and Collidoscope.

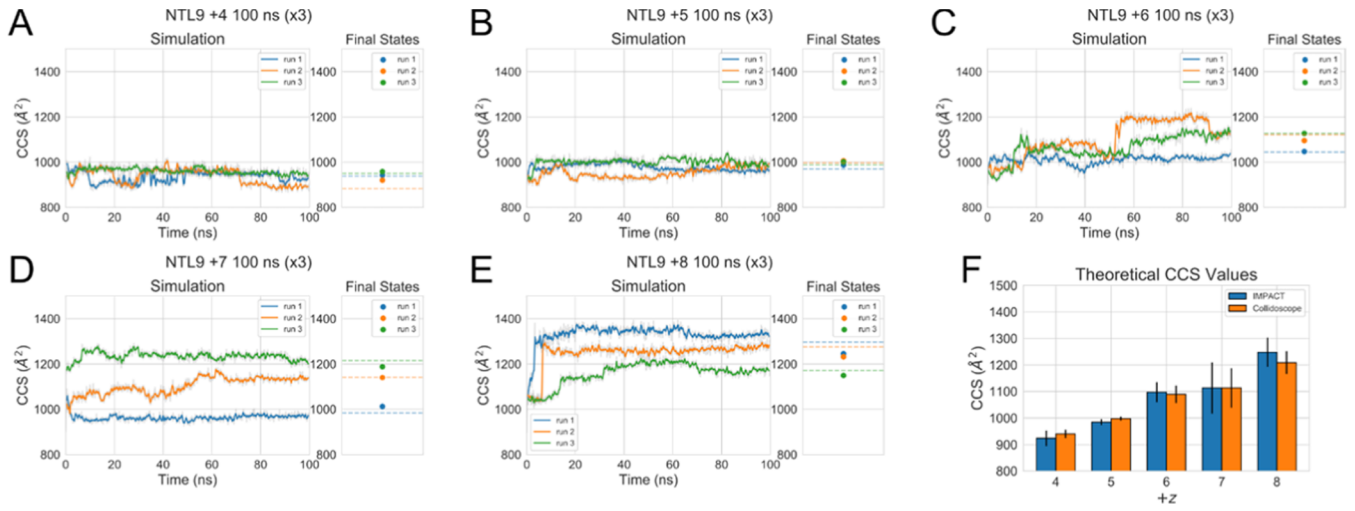


Figure 5. Theoretical CCS values of initial simulations, beginning from the same set with IMPACT analysis of full simulations in solid lines, with a separate plot showing the IMPACT value of the final structure shown as a dotted line and Collidoscope value as dots. Triplicate simulations of (A) +4, (B) +5, (C) +6, (D) +7, and (E) +8 charge state. The gray shading represents the unsmoothed data, and the solid line represents smoothed with a mean window of 5, once. (F) Mean THCCS values of the final states of the simulations calculated by IMPACT and Collidoscope.

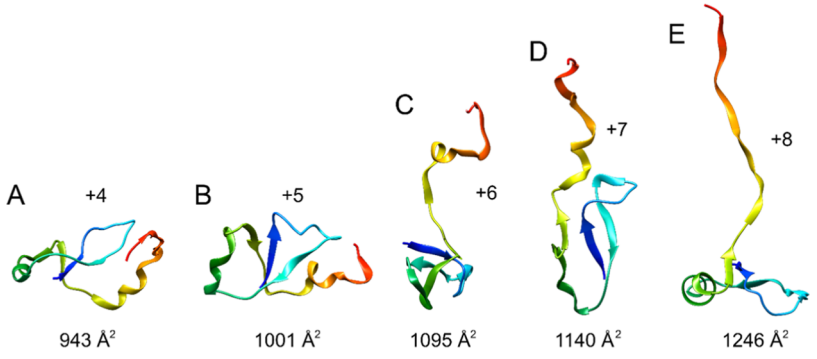


Figure 6. Selected final structures for the charging simulations after 100 ns for charges (A) +4, (B) +5, (C) +6, (D) +7, and (E) +8, with the theoretical CCS value calculated by Collidoscope.

The question of whether the low variability of unfolded outcomes was due to the system or the method was explored. NTL9 is from a thermophilic organism, and the domain is thermally stable, with a melting temperature of 78 °C at pH 5.4 in solution[44] and therefore *in silico* thermal unfolding may not be appropriate. We attempted to replicate some of the observations from the original publication, which showcased *in silico* thermal of unfolding using the MPA. In the original publication, the homotetramer transthyretin (TTR) unfolding was simulated, and the authors were able to demonstrate charge-mediated subunit ejection, which is consistent with the observation of subunit dissociation and charge stripping observed in IM-MS studies,[8,67−69] within 20 ns of simulations. Consequently, we explored the thermal *in silico* unfolding of a tetrameric protein of similar mass and structure, Avidin. None of the 100 × 20 ns simulations displayed subunit ejection; however, a partial unfolding of subunits was observed, presumably as a precursor to ejection (Figure S5 Tables S6,

https://doi.org/10.1021/acs.analchem.2c03352
Anal. Chem. 2022, 94, 16113−16121

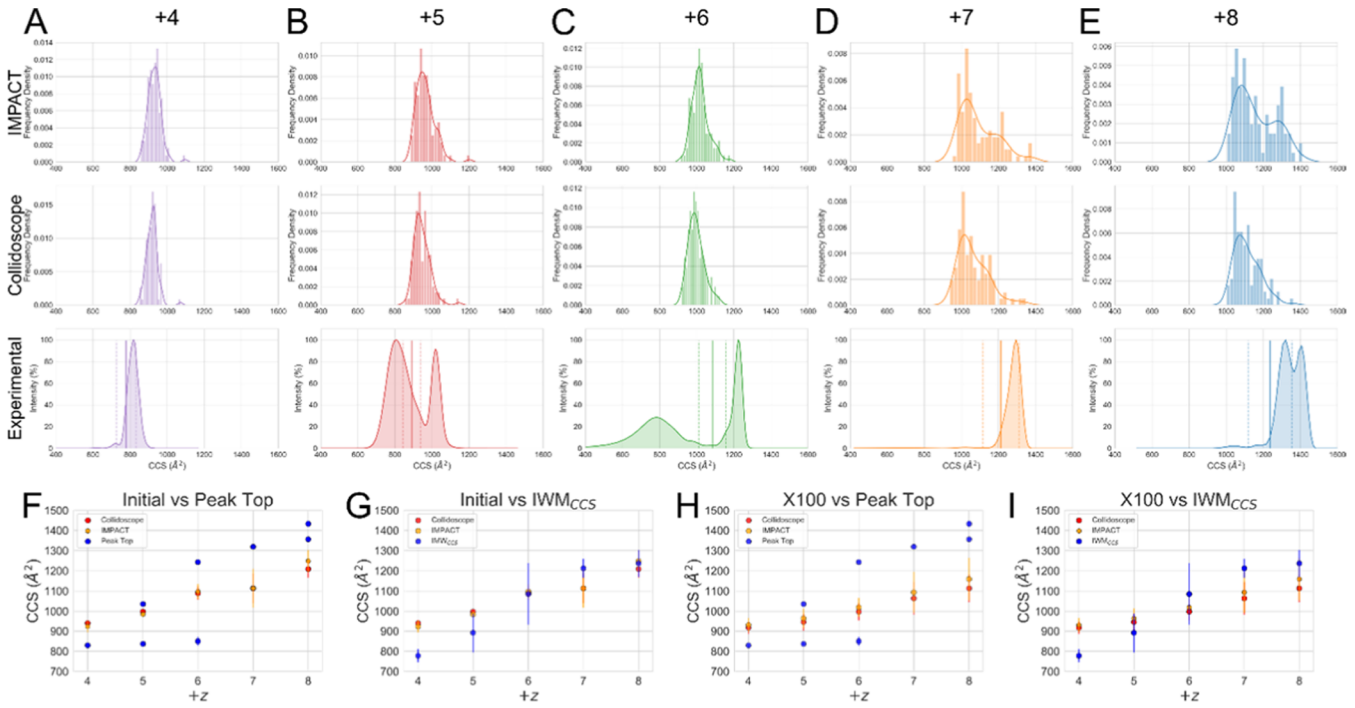


Figure 7. Comparison of simulated and experimental results. (A–E) Results of 100 replicates of the stable temperature simulations showing histograms of the final states calculated by IMPACT and Collidoscope and experimental CCS distributions, with the IWM$_{CCS}$ denoted by the dotted line, and the standard deviation shown by the faint dotted lines for (A) +4, (B) +5, (C) +6, (D) +7, and (E) +8 charge states. (F–I) Comparison of theoretical and experimental values. Mean THCCS value from initial triplicate simulation final states plotted against (A) experimental peak top CCS values and (B) IWM$_{CCS}$. Mean THCCS value from X100 simulation replicate final states plotted against (C) experimental peak top CCS values and (D) IWM$_{CCS}$.

and S7). This suggests that the original MPA workflow does not reproduce thermal unfolding *in silico* in a reliable manner.

Coulombic Unfolding Simulations. Since the *in silico* thermal unfolding showed low reproducibility, we next investigated whether increasing the charging of the protein would lead to a better match with the *in vacuo* experimental observations. Simulations were performed for the +4 to +8 charge states at a steady temperature of 300 K, for 100 ns. As the charge increases, the theoretical CCS (THCCS) of the final state increases in line with the native IM-MS data (Figures 5 and S7–S11). While the THCCS of the *in silico* model increases, not all runs lead to an extended state, for instance, for the +7 state, run 1 retains a compact structure (Figures 5 and S9A). For each run, the root-mean-square deviation of the Cα atoms between the frame and the final state decreased and became mostly stable (Figure S12).

The coulombic simulations show a consistent evolution of *in vacuo* structures (Figure 6), with +4 and +5 forming compact structures and +6 to +8 forming more unfolded structures, with the thermally unfolded +5 and coulombically unfolded +8 *in silico* models having comparable structures.

To properly compare the *in silico* and *in vacuo* data, each of the coulombic simulations was repeated 100x, to create a theoretical CCS distribution (THCCSD) from a kernel density estimation of the final *in silico* models, which could then be compared to the experimental data. Replication instead of increasing the length of the simulations was chosen, as it was clear that 100 ns was a long-enough time period for a conformational sampling of an extended state to occur from the initial triplicates.

A comparison of the THCCSDs (Figure 7) shows that increasing the replicates creates a better likelihood of a model matching the *in vacuo* data. Interestingly, all of the THCCSDs appear to show some degree of multimodal behavior. Ensemble cluster analysis of the 100 final structures of each simulated charge state support this (Figure S18). Comparison of the theoretical and experimental CCSDs shows that while a high replication number is not able to reproduce something which exactly matches the experimental distribution, an overlap exists between the THCCS of some models and experimental values, meaning a subset of models reproduce the experimental data.

DISCUSSION

Creating robust pipelines for modeling *in vacuo* unfolding *in silico* would be highly advantageous for understanding protein behavior during IM-MS analysis. It would also be advantageous to be able to objectively evaluate if the unfolding observed *in vacuo* is relevant to *in-solution* behavior. This would help CA become a more informative structural biology technique: in much the same vein as how the demonstrations that soft ionization retains the solution-like structure and behavior[70] facilitated the use of mass spectrometry a valuable tool for structural biologists.

NTL9, a well-studied system in-solution,[45−51] appears to adopt a mixture of compact and extended states during ionization, which are observable *in vacuo* (Figure 2B,C). Trying to reproduce these states in an unbiased manner *in silico*, and hence create a model of unfolding, to allow comparison of *in-solution* and *in vacuo* data has had variable success. Both thermal and coulombic unfolding *in silico* matched the expected *in-solution* model: release of the C-terminal α-helix (hB), with dissociation of the central β-sheet and the first α-helix (hA). Replications of the *in silico* thermal

https://doi.org/10.1021/acs.analchem.2c03352
Anal. Chem. 2022, 94, 16113−16121

unfolding did not show good reproducibility, and the original method for thermal unfolding[40,41] produced unfolded structures only 13/100 times. This is most likely due to the lack of conformational sampling present in unbiased simulations and was not limited to the thermostable system NTL9.[71−73] While the test system previously used in Popa *et al.*, to demonstrate the utility of the MPA workflow was transthyretin (TTR),[40] another well-studied homotetrameric system, Avidin, was used here. Avidin has a slightly higher intact mass than TTR, 64.0 kDa compared to 55.0 kDa; the structure used in our studies is truncated at the C-terminus, making it 54.8 kDa and similar in mass and size to TTR, meaning they should have a similar internal temperature when heated.[57] In the original *in silico* studies, charge-mediated subunit ejection was observed for TTR; however, Avidin displayed no charge-mediated subunit ejection on the timescale of our simulations. The TTR structure used previously (3GRG) is in fact a mutant known as M-TTR (F87M/L110M).[74] M-TTR shows a reduced self-association constant due to the substitution of F87 and L110 with the bulkier M residues at subunit interfaces,[74,75] which may have affected the outcome of the simulations. The lack of subunit ejection observed in Avidin suggests that the thermal unfolding workflow has limited conformational sampling, which may be exacerbated by the proton hopping. Energy minimization studies[64,76] show that 10^6−10^7 proton rearrangements need to be performed to reach an energy minima, which adds an extra layer of complexity outside of standard structural dynamics. While in this study, we opted for more replicates of shorter simulations, fewer numbers of longer, biased simulations may be required. Longer simulations may also allow the creation of more accurate theoretical CCSDs: while experimental CCSDs are a product of gas-phase conformations which do not interconvert on the timescales of drift separation, they are the result of dynamics of timescales exceeding the simulation time.

Comparison of ^{TW}CCS and THCCS values is an important part of bridging the gap between *in vacuo* and *in silico* studies. Here, we have compared several metrics which are used experimentally for IM-MS analysis: the peak top value and the intensity weighted mean (IWM) (Figure 7G,I). Comparison of the peak top values to the mean THCCS values is poor (Figure 7F,H) due to the fact that the average is unable to capture distinct populations. Using a method like IWM_{CCS} gives better overlap as it better captures the weighting of multi-conformer ensembles.

From the simulations, high charging, *i.e.*, +7 and +8, produce highly unfolded *in silico* structures; however, ^{TW}CCS values suggest that *in vacuo*, more compact conformers are favored. The discrepancy between the ^{TW}CCS and THCCS may be due to the difficulty of calculating THCCS of linear ions, as described by Kulesza *et al.*,[77] which is highlighted in our study by the differences between the THCCS values calculated by IMPACT and Collidoscope for the same extended structures of the +7 and+8 charge states (Figures 5E and S18G−J). It may also be a function of ^{TW}CCS calibrant class: to get an accurate TWIMS CCS calibration, molecules of the same class as the experimental molecule must be used, *i.e.*, native protein calibrants for globular proteins and denatured proteins for disordered or denatured proteins. The *in silico* structures suggest that the higher charge states of NTL9 have both ordered and disordered structural regions, meaning that neither calibrant class would be fully comparable. Other possible avenues for future exploration to close the gap between the THCCS and ^{TW}CCS include using different force fields during *in silico* model creation, as studies have shown that certain force fields, which are designed to replicate *in-solution* behavior, commonly produce models which favor either compaction or extension compared to experimentally derived CCS values.[78] Furthermore, different methods to produce simulated ions could also be employed, as the complexity of comparing the molecular dynamics to experimental data is compounded by the differing behavior of folded and unfolded protein ions during desolvation. While folded proteins are regarded as ionizing *via* the charged residue model (CRM),[79−83] unfolded proteins are believed to ionize *via* the chain ejection model (CEM).[84,85] The two models could imply differing unfolding mechanisms, which the MPA may not be able to replicate fully. Desolvation simulations function by steadily removing solvent from a charged droplet containing the protein structure to create a gas-phase ion.[31,32,86−90] While the simulation of the droplet itself is more computationally expensive, it may produce a more accurate initial structure for further simulation.

By combining *in vacuo*, *in silico* and *in-solution* data, we have shown it is possible to create coherent models of unfolding that link gas-phase and solution-phase behaviors. This approach shows promise and importantly, highlights multiple experimental and theoretical avenues to explore to further improve the methodology. We believe that the data and the analysis presented here both illustrate the power of the hybrid experimental computational approach and point the way for future developments.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c03352.

in silico pipeline, *in silico* parameter tables, experimental parameter tables, CCS calibration charts, salt-bridge analysis, triplicate final structures of coulombic simulations, RMSD charts, full tables of *in silico* CCS values for 100× coulombic and thermal simulations, and ensemble analysis for simulations (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Konstantinos Thalassinos − *Institute of Structural and Molecular Biology, Division of Bioscience, University College London, London WC1E 6BT, U.K.; Institute of Structural and Molecular Biology, Birkbeck College, University of London, London WC1E 7HX, U.K.;* orcid.org/0000-0001-5072-8428; Email: k.thalassinos@ucl.ac.uk

Authors

Charles Eldrid − *School of Biological Sciences, University of Southampton, Southampton SO16 1BJ, U.K.; Institute of Structural and Molecular Biology, Division of Bioscience, University College London, London WC1E 6BT, U.K.;* orcid.org/0000-0001-5306-3644

Tristan Cragnolini − *Institute of Structural and Molecular Biology, Birkbeck College, University of London, London WC1E 7HX, U.K.*

Aisha Ben-Younis − *Institute of Structural and Molecular Biology, Division of Bioscience, University College London, London WC1E 6BT, U.K.*

Junjie Zou − *Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States*
Daniel P. Raleigh − *Institute of Structural and Molecular Biology, Division of Bioscience, University College London, London WC1E 6BT, U.K.; Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States;* orcid.org/0000-0003-3248-7493

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c03352

Author Contributions

C.E.: experimental design, data collection, simulations, analysis, and preparation of the manuscript. T.C.: simulations, experimental design, A.B.-Y.: data collection, J.Z.: sample preparation, D.R.: experimental design, writing. K.T.: experimental design, writing.

Funding

Funding from the Wellcome Trust, 107927/Z/15/Z, is acknowledged. CE is funded by BBSRC iCASE Award with Waters BB/L015382/1; T.C. is funded by Wellcome Trust Collaborative Awards in Science 209250/Z/17/Z.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Konermann group for the algorithm files and the Avidin structure and for responding to queries regarding troubleshooting of the MPA algorithm.

■ ABBREVIATIONS USED

CA	collision activation
CCS	collision cross-section
CD	circular dichroism
CIU	collision induced unfolding
CTL9	C-terminal construct of the L9 protein
EPR	electron paramagnetic resonance
IM	ion mobility
IM-MS	ion mobility mass spectrometry
IWM	intensity weighted mean
IWSD	intensity weighted standard deviation
MD	molecular dynamics
MPA	mobility proton algorithm
MS	mass-spectrometry
NMR	nuclear magnetic resonance
NTL9	N-terminal construct of the L9 protein
THCCS	theoretical collision cross section
TTR	transthyretin
^{TW}CCS	experimental collision cross section

■ REFERENCES

(1) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2017**, *86*, 27−68.
(2) Gooptu, B.; Lomas, D. A. *J. Exp. Med.* **2008**, *205*, 1529−1534.
(3) Lai, Z.; Colón, W.; Kelly, J. W. *Biochemistry* **1996**, *35*, 6470−6482.
(4) Benseny-Cases, N.; Karamanos, T. K.; Hoop, C. L.; Baum, J.; Radford, S. E. *J. Biol. Chem.* **2019**, *294*, 9392−9401.
(5) Kelly, S. M.; Price, N. C. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1997**, *1338*, 161−185.
(6) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607−3622.
(7) Rogozea, A.; Matei, I.; Turcu, I. M.; Ionita, G.; Sahini, V. E.; Salifoglou, A. *J. Phys. Chem. B* **2012**, *116*, 14245−14253.
(8) Benesch, J. L. P. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 341−348.
(9) Eldrid, C.; Ben-Younis, A.; Ujma, J.; Britt, H.; Cragnolini, T.; Kalfas, S.; Cooper-Shepherd, D.; Tomczyk, N.; Giles, K.; Morris, M.; Akter, R.; et al. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (6), 1545−1552.
(10) Dixit, S. M.; Polasky, D. A.; Ruotolo, B. T. *Curr. Opin. Chem. Biol.* **2018**, *42*, 93−100.
(11) Zhong, Y.; Han, L.; Ruotolo, B. T. *Angew. Chem., Int. Ed.* **2014**, *53*, 9209−9212.
(12) Hernandez-Alba, O.; Wagner-Rousset, E.; Beck, A.; Cianférani, S. *Anal. Chem.* **2018**, *90*, 8865−8872.
(13) Yamashita, M.; Fenn, J. B. *J. Phys. Chem. A* **1984**, *88*, 4451−4459.
(14) Fenn, J. B.; Mann, M. M.; Meng, C. K. A. I.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.
(15) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193A−1203A.
(16) Chowdhury, S. K.; Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1990**, *112*, 9012−9013.
(17) Nyon, M. P.; Prentice, T.; Day, J.; Kirkpatrick, J.; Sivalingam, G. N.; Levy, G.; Haq, I.; Irving, J. A.; Lomas, D. A.; Christodoulou, J.; et al. *Protein Sci.* **2015**, *24*, 1301−1312.
(18) Wojnowska, M.; Yan, J.; Sivalingam, G. N.; Cryar, A.; Gor, J.; Thalassinos, K.; Djordjevic, S. *Chem. Biol.* **2013**, *20*, 1411−1420.
(19) Tamara, S.; Hoek, M.; Scheltema, R. A.; Leney, A. C.; Heck, A. J. R. *Chem* **2019**, *5*, 1302−1317.
(20) Gabelica, V.; Marklund, E. *Curr. Opin. Chem. Biol.* **2018**, *42*, 51−59.
(21) Eldrid, C.; Thalassinos, K.Developments in Tandem Ion Mobility Mass Spectrometry. In *Biochem. Soc. Trans.*; Portland Press, December 18, 2020; pp 2457−2466 DOI: 10.1042/BST20190788.
(22) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2240−2248.
(23) Dickinson, E. R.; Jurneczko, E.; Pacholarz, K. J.; Clarke, D. J.; Reeves, M.; Ball, K. L.; Hupp, T.; Campopiano, D.; Nikolova, P. V.; Barran, P. E. *Anal. Chem.* **2015**, *87*, 3231−3238.
(24) Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J. E.; Ruotolo, B. T.; Robinson, C. V.; et al. *Nat. Chem.* **2009**, *1*, 326−331.
(25) Beveridge, R.; Covill, S.; Pacholarz, K. J.; Kalapothakis, J. M. D.; Macphee, C. E.; Barran, P. E. *Anal. Chem.* **2014**, *86*, 10979−10991.
(26) Eldrid, C.; Ben-Younis, A.; Ujma, J.; Britt, H.; Cragnolini, T.; Kalfas, S.; Cooper-Shepherd, D.; Tomczyk, N.; Giles, K.; Morris, M.; et al. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1545−1552.
(27) Tian, Y.; Han, L.; Buckner, A. C.; Ruotolo, B. T. *Anal. Chem.* **2015**, *87*, 11509−11515.
(28) Deslignière, E.; Ehkirch, A.; Botzanowski, T.; Beck, A.; Hernandez-Alba, O.; Cianférani, S. *Anal. Chem.* **2020**, *92*, 12900−12908.
(29) Koeniger, S. L.; Clemmer, D. E. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 322−331.
(30) Eldrid, C.; Ujma, J.; Kalfas, S.; Tomczyk, N.; Giles, K.; Morris, M.; Thalassinos, K. *Anal. Chem.* **2019**, *91*, 7554−7561.
(31) Steinberg, M. Z.; Breuker, K.; Elber, R.; Gerber, R. B. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4690−4697.
(32) Steinberg, M. Z.; Elber, R.; McLafferty, F. W.; Gerber, R. B.; Breuker, K. *ChemBioChem* **2008**, *9*, 2417−2423.
(33) Wyttenbach, T.; Bowers, M. T. *J. Phys. Chem. B* **2011**, *115*, 12266−12275.
(34) Marklund, E. G.; Benesch, J. L. *Curr. Opin. Struct. Biol.* **2019**, *54*, 50−58.
(35) Haran, G. *Curr. Opin. Struct. Biol.* **2012**, *22*, 14−20.
(36) Holehouse, A. S.; Pappu, R. V. *Annu. Rev. Biophys.* **2018**, *47*, 19−39.
(37) Jørgensen, T. J. D.; Gårdsvoll, H.; Ploug, M.; Roepstorff, P. *J. Am. Chem. Soc.* **2005**, *127*, 2785−2793.
(38) Muller, L.; Jackson, S. N.; Woods, A. S. *Eur. J. Mass Spectrom.* **2019**, *25*, 212−218.
(39) Fegan, S. K.; Thachuk, M. *J. Chem. Theory Comput.* **2013**, *9*, 2531−2539.

(40) Popa, V.; Trecroce, D. A.; McAllister, R. G.; Konermann, L. *J. Phys. Chem. B* **2016**, *120*, 5114−5124.
(41) Konermann, L. *J. Phys. Chem. B* **2017**, *121*, 8102−8112.
(42) Konermann, L.; Metwally, H.; Mcallister, R. G.; Popa, V. *Methods* **2018**, *144*, 104−112.
(43) Donk, P. J. *J. Bacteriol.* **1920**, *5*, 373−374.
(44) Hoffman, D. W.; Davies, C.; Gerchman, S. E.; Kycia, J. H.; Porter, S. J.; White, S. W.; Ramakrishnan, V. *EMBO J.* **1994**, *13*, 205−212.
(45) Kuhlman, B.; Luisi, D. L.; Evans, P. A.; Raleigh, D. P. *J. Mol. Biol.* **1998**, *284*, 1661−1670.
(46) Kuhlman, B.; Boice, J. A.; Fairman, R.; Raleigh, D. P. *Biochemistry* **1998**, *37*, 1025−1032.
(47) Kuhlman, B.; Raleigh, D. P. *Protein Sci.* **1998**, *7*, 2405−2412.
(48) Kuhlman, B.; Luisi, D. L.; Young, P.; Raleigh, D. P. *Biochemistry* **1999**, *38*, 4896−4903.
(49) Luisi, D. L.; Wu, W. J.; Raleigh, D. P. *J. Mol. Biol.* **1999**, *287*, 395−407.
(50) Luisi, D. L.; Kuhlman, B.; Sideras, K.; Evans, P. A.; Raleigh, D. P. *J. Mol. Biol.* **1999**, *289*, 167−174.
(51) Luisi, D. L.; Raleigh, D. P. *J. Mol. Biol.* **2000**, *299*, 1091−1100.
(52) Sato, S.; Raleigh, D. P. *J. Mol. Biol.* **2002**, *318*, 571−582.
(53) Horng, J. C.; Cho, J. H.; Raleigh, D. P. *J. Mol. Biol.* **2005**, *345*, 163−173.
(54) Zhang, S.; Zhang, Y.; Stenzoski, N.; Peran, I.; McCallum, S. A.; Raleigh, D.; Royer, C. A. *Biophys. J.* **2019**, *116*, 445−453.
(55) Cho, J. H.; Meng, W.; Sato, S.; Kim, E. Y.; Schindelin, H.; Raleigh, D. P. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 12079−12084.
(56) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.et al.Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. In *Mass Spectrometry Reviews*; John Wiley & Sons, Ltd., February 1, 2019; pp 291−320 DOI: 10.1002/mas.21585.
(57) Strzelczyk, P.; Plazuk, D.; Zakrzewski, J.; Bujacz, G. *Molecules* **2016**, *21*, 1270.
(58) Robertson, M. J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2015**, *11*, 3499−3509.
(59) Thalassinos, K.; Grabenauer, M.; Slade, S. E.; Hilton, G. R.; Bowers, M. T.; Scrivens, J. H. *Anal. Chem.* **2009**, *81*, 248−254.
(60) Eldrid, C.; O'Connor, E.; Thalassinos, K. *Rapid Commun. Mass Spectrom.* **2020**, *34 Suppl 4*, No. e8613.
(61) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. *Anal. Chem.* **2010**, *82*, 9557−9565.
(62) Sivalingam, G. N.; Cryar, A.; Williams, M. A.; Gooptu, B.; Thalassinos, K. *Int. J. Mass Spectrom.* **2018**, *426*, 29−37.
(63) Marklund, E. G.; Degiacomi, M. T.; Robinson, C. V.; Baldwin, A. J.; Benesch, J. L. P. *Structure* **2015**, *23*, 791−799.
(64) Ewing, S. A.; Donor, M. T.; Wilson, J. W.; Prell, J. S. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 587−596.
(65) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33−38.
(66) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(67) Light-Wahl, K. J.; Schwartz, B. L.; Smith, R. D. *J. Am. Chem. Soc.* **1994**, *116*, 5271−5278.
(68) Benesch, J. L. P.; Aquilina, J. A.; Ruotolo, B. T.; Sobott, F.; Robinson, C. V. *Chem. Biol.* **2006**, *13*, 597−605.
(69) Beardsley, R. L.; Jones, C. M.; Galhena, A. S.; Wysocki, V. H. *Anal. Chem.* **2009**, *81*, 1347−1356.
(70) Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V. *Science* **2005**, *310*, 1658−1661.
(71) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, No. 155102.
(72) Rao, F.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 9152−9157.
(73) Harada, R.; Takano, Y.; Baba, T.; Shigeta, Y. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6155−6173.
(74) Palmieri, L. D. C.; Lima, L. M. T. R.; Freire, J. B. B.; Bleicher, L.; Polikarpov, I.; Almeida, F. C. L.; Foguel, D. *J. Biol. Chem.* **2010**, *285*, 31731−31741.
(75) Jiang, X.; Smith, C. S.; Petrassi, H. M.; Hammarström, P.; White, J. T.; Sacchettini, J. C.; Kelly, J. W. *Biochemistry* **2001**, *40*, 11442−11452.
(76) Abramsson, M. L.; Sahin, C.; Hopper, J. T. S.; Branca, R. M. M.; Danielsson, J.; Xu, M.; Chandler, S. A.; Österlund, N.; Ilag, L. L.; Leppert, A.; et al. *JACS Au* **2021**, *1*, 2385−2393.
(77) Kulesza, A.; Marklund, E. G.; MacAleese, L.; Chirot, F.; Dugourd, P. *J. Phys. Chem. B* **2018**, *122*, 8317−8329.
(78) Rolland, A. D.; Prell, J. S. *TrAC Trends Anal. Chem.* **2019**, *116*, 282−291.
(79) Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* **1968**, *49*, 2240−2249.
(80) Mack, L. L.; Kralik, P.; Rheude, A.; Dole, M. *J. Chem. Phys.* **1970**, *52*, 4977−4986.
(81) Clegg, G. A.; Dole, M. *Biopolymers* **1971**, *10*, 821−826.
(82) Tolić, L. P.; Anderson, G. A.; Smith, R. D.; Brothers, H. M.; Spindler, R.; Tomalia, D. A. *Int. J. Mass Spectrom. Ion Processes* **1997**, *165*−*166*, 405−418.
(83) Fernandez de la Mora, J. *Anal. Chim. Acta* **2000**, *406*, 93−104.
(84) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. *Anal. Chem.* **2013**, *85*, 2−9.
(85) Meyer, T.; Gabelica, V.; Grubmüller, H.; Orozco, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 408−425.
(86) Breuker, K.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18145−18152.
(87) Thirumuruganandham, S. P.; Urbassek, H. M. *Int. J. Mass Spectrom.* **2010**, *289*, 119−127.
(88) Kim, D.; Wagner, N.; Wooding, K.; Clemmer, D. E.; Russell, D. H. *J. Am. Chem. Soc.* **2017**, *139*, 2981−2988.
(89) Porrini, M.; Rosu, F.; Rabin, C.; Darré, L.; Gómez, H.; Orozco, M.; Gabelica, V. *ACS Cent. Sci.* **2017**, *3*, 454−461.
(90) 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*; Elsevier Inc., 2018; Vol. *144*, pp 104−112 DOI: 10.1016/j.ymeth.2018.04.010.

https://doi.org/10.1021/acs.analchem.2c03352
Anal. Chem. 2022, 94, 16113−16121