Gas-Phase Ion/Ion Chemistry for Structurally Sensitive Probes of Gaseous Protein Ion Structure: Electrostatic and Electrostatic to Covalent Cross-Linking

Melanie Cheung See Kit[1], Veronica V. Carvalho[1], Jonah Z. Vilseck[2,3], Ian K. Webb[1,3*]

[1] Department of Chemistry and Chemical Biology, Indiana University Purdue University Indianapolis, Indianapolis, Indiana 46202, USA

[2] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA

[3] Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA

[*] Correspondence to: Ian K. Webb; e-mail: ikwebb@iu.edu

Keywords: Native mass spectrometry, ion/ion reactions, ion mobility, electron capture dissociation

This is the author's manuscript of the article published in final edited form as:

Kit, M. C. S., Carvalho, V. V., Vilseck, J. Z., & Webb, I. K. (2021). Gas-Phase Ion/Ion Chemistry for Structurally Sensitive Probes of Gaseous Protein Ion Structure: Electrostatic and Electrostatic to Covalent Cross-Linking. International Journal of Mass Spectrometry, 463, 116549. https://doi.org/10.1016/j.ijms.2021.116549

Abstract

Intramolecular interactions within a protein are key in maintaining protein tertiary structure and understanding how proteins function. Ion mobility-mass spectrometry (IM-MS) has become a widely used approach in structural biology since it provides rapid measurements of collision cross sections (CCS), which inform on the gas-phase conformation of the biomolecule under study. Gas-phase ion/ion reactions target amino acid residues with specific chemical properties and the modified sites can be identified by MS. In this study, electrostatically reactive, gas-phase ion/ion chemistry and IM-MS are combined to characterize the structural changes between ubiquitin electrosprayed from aqueous and denaturing conditions. The electrostatic attachment of sulfo-NHS acetate to ubiquitin via ion/ion reactions and fragmentation by electron-capture dissociation (ECD) provide the identification of the most accessible protonated sites within ubiquitin as the sulfonate group forms an electrostatic complex with accessible protonated side chains. The protonated sites identified by ECD from the different solution conditions are distinct and, in some cases, reflect the disruption of interactions such as salt bridges that maintain the native protein structure. This agrees with previously published literature demonstrating that a high methanol concentration at low pH causes the structure of ubiquitin to change from a native (N) state to a more elongated A state. Results using gas-phase, electrostatic cross-linking reagents also point to similar structural changes and further confirm the role of methanol and acid in favoring a more unfolded conformation. Since cross-linking reagents have a distance constraint for the two reactive sites, the data is valuable in guiding computational structures generated by molecular dynamics. The research presented here describes a promising strategy that can detect subtle changes in the local environment of targeted amino acid residues to inform on changes in the overall protein structure.

Introduction

Over the last twenty years, mass spectrometry (MS) has emerged as a valuable approach used in structural biology.[1-2] In particular, the field of native MS routinely takes advantage of "soft" nanoflow electrospray ionization (nESI), which helps maintain non-covalent bonds as proteins transition from solution into the gas phase, so that higher order structure in the gas-phase reflect some aspects of the solution structure.[3-6] Native MS, ion mobility (IM-MS) and tandem MS (MS/MS) have been widely applied to characterize aspects of protein structure such as stoichiometry or subunit arrangement in a protein complex.[7-11]

Covalent bioconjugation reactions are routinely employed as a strategy to target amino acid residues with specific chemical characteristics and combined with MS to analyze the modified products.[12-13] These reactions provide insight about protein dynamics and the local microenvironment of the targeted residues.[14] One common type of reaction is based on N-hydroxysuccinimide (NHS) ester chemistry. NHS esters react with nucleophiles, such as the amines of a free N-terminus or the amino group in a lysine side chain to form stable amide bonds after the loss of the NHS group.[15] Sulfo-NHS ester reagents are water soluble analogs and are used in chemical cross-linking MS (XL-MS) to provide distance constraints used to map subunit topology of protein complexes. Excess cross-linking reagent is mixed with the protein complex in an aqueous environment, followed by enzymatic digestion to generate cross-linked peptides that are analyzed by liquid chromatography/ electrospray ionization-tandem mass spectrometry (LC/ESI-MS/MS) to identify cross-linked fragments.[16-17] However, one challenge with the in-solution, bottom-up XL-MS approach is the large amount of generated data, which can make it difficult to confidently assign peptide identification during cross-linked database searches.[18] There have also been some "top-down" analyses of proteins by XL-MS, where an intact protein such as ubiquitin reacts with a cross-linking reagent. Advantages of analyzing the cross-linked intact protein in the gas phase include avoiding missed coverage from undetected peptides, minimizing

the sample complexity by not forming a mixture of peptides, and a facile determination of the extent of cross-linking via an intact molecular weight measurement.[19] The reaction between excess disuccinimidyl suberate (DSS) crosslinking reagent (spacer arm = 11.4 Å) and aqueous ubiquitin resulted in cross-links identified for Lys48 – Lys63 and Lys6 – Lys11.[19] In addition, when the products of a *bis*(sulfosuccinimidyl)suberate (BS3) (spacer arm = 11.4 Å) and ubiquitin reaction were fragmented by ultraviolet photodissociation (UVPD) or higher energy collisional dissociation (HCD), the observed cross-linked residues are Lys11 – Lys33, N-term – Lys29 and N-term– Lys63.[20] These experiments have demonstrated the feasibility of top-down interrogations of cross-linking experiments. Top-down XL-MS experiments have also been used to characterize membrane-bound protein complexes and their post-translational modifications.[21]

Many of these in-solution reactions have now been conducted in the gas phase via ion/ion reactions in RF-confining ion guides and ion traps using tandem MS strategies.[22] There are many benefits to running these reactions *in vacuo*, namely greater control over reacting species and reaction conditions, *m/z* isolation to purify the reagent, millisecond reaction times, and improved reaction efficiency.[22] Gas-phase ion/ion reactions have also been used to report on three-dimensional structure of protein cations formed from native-like or denaturing solutions.[23-25] A combination of experimental[4, 7, 26-29] and computational studies[30-31] have shown that under carefully-controlled experimental conditions, the solution structures of electrosprayed peptides and proteins is conserved upon electrospray ionization, making methods for probing gas-phase three-dimensional structures of proteins an extremely valuable addition to structural biology techniques.

The covalent reaction between ubiquitin and ethylene glycol *bis*(sulfosuccinimidyl suberate) (sulfo-EGS) has been partially characterized in the gas-phase. Using a Sciex qTOF instrument modified for ion/ion chemistry[32] allowed for the isolation of the ubiquitin cations and reagent anions by Q1, which were then sequentially transferred to q2 to react and form long lived

complexes. The products were transferred back to Q1 for isolation and sent to q2 for ion-trap collision-induced dissociation (CID). The formation of a covalent cross-link was indicated by the loss of two sulfo-NHS groups, which was observed for Lys27 – Lys29 and Lys48 – Lys63 in ubiquitin 6^+ and 7^+. At higher charge states, backbone cleavage was favored over the loss of the sulfo-NHS groups, indicating that the non-covalent complex was stable. However, the sites where the sulfo groups are attached electrostatically could not be determined since CID can lead to possible scrambling of the electrostatically-bound linker during backbone fragmentation.[25]

In this study, we follow up on the previous results and implement an ion/ion, ion mobility/mass spectrometry (I^3M^2S) platform[24] to characterize the products of the gas-phase electrostatic ion/ion reaction between intact ubiquitin cation and the negatively charged sulfo-NHS based reagents (sulfo-NHS acetate and sulfo-EGS) with electron capture dissociation (ECD).[33] We observed the non-covalent complex, formed from electrostatic binding of the sulfonate group(s) of the anion to protonated residues from ubiquitin, and separate the reaction products by traveling wave ion mobility spectrometry (TWIMS)[34-35]. Fragmentation of the electrostatic product by ECD produces ions with a signature mass shift, due to the presence of the sulfo-NHS reagent, and these fragment ions are mapped to the protein sequence to identify the labeled amino acid residues. Characterizing the reactive sites of ubiquitin prepared from both aqueous and denaturing conditions provides insights into the structural variations between the two sample conditions across different protein charge states and can be used to guide molecular dynamics simulations of gas-phase structures. The results demonstrate that I^3M^2S with ECD is highly suitable for detecting and analyzing the electrostatic products formed during gas-phase ion/ion reactions, thereby providing key information to reveal changes in gas-phase protein structure from protein ions arising from different solution structural states. Furthermore, the electrostatic cross-linking and addition data should be complementary to solution and gas-phase covalent cross-linking and labeling experiments.

Materials and Methods

Materials and Sample Preparation. Ubiquitin from bovine erythrocites and ammonium acetate were purchased from Sigma-Aldrich (St. Louis, MO). Methanol, acetonitrile, acetic acid, and formic acid were obtained from Fisher Scientific (Fairmont, NJ). Water was obtained from a Milli-Q Millipore A10 (Burlington, MA) water purification system at a resistivity of 18 MΩ or greater. Ethylene glycol *bis*(sulfosuccinimidyl succinate) (Sulfo-EGS) and sulfo-n-hydroxyl succinimide acetate (Sulfo-NHS-Acetate) were purchased as sodium salts from ThermoFisher Scientific (Rockford, IL). For denaturing conditions, ubiquitin was prepared at 10 μM in a mixture of 50/50 vol/vol solution of water/methanol and 0.1% formic acid, conditions known to favor the partially-unfolded A state in solution.[36] The 6^+ charge state under aqueous native-like conditions was generated by dissolving ubiquitin in an aqueous 10 mM ammonium acetate solution for a final concentration of 10 μM. For the 7^+ and 8^+ charge states under aqueous conditions, ubiquitin was dissolved in water and the pH adjusted to 3 using a 0.1 M acetic acid solution. Sulfo-EGS and sulfo-NHS-acetate were dissolved in acetonitrile at a final concentration of 10 mM and used during the ion/ion reactions.

Ion Mobility Mass Spectrometry and Ion/Ion Reactions. All the experiments were performed on a Synapt G2-Si High Definition Mass Spectrometer (Waters Corporation, Wilmslow, U.K.), equipped with electron transfer dissociation (ETD), a NanoLockspray source, an external electrospray voltage control module (GAA Custom Electronics LLC, Kennewick, WA), and an ExD cell (e-MSion, Corvallis, OR). Instrumental details for the ion/ion reactions[24, 37] and ExD cell[33] have been described previously. The NanoLockspray source was used with proteins infused through the sample sprayer and anionic reagents through the reference sprayer.. The infusion flow rate for both the cation (protein) and anion (sulfo-NHS acetate or sulfo-EGS) reagents was 1 μL/min. The external voltage control module allows for the synchronization of the source polarity (+1.1 kV and -1.5 kV) and ion injection times (1s each) so that the anions and cations can be sequentially

introduced into the instrument. Reagent ions were mass-selected in the quadrupole and stored in the trap cell with a trap traveling wave height of zero volts to allow for the gas-phase ion/ion reaction of a specific protein charge state to occur. The ion/ion products were then separated by ion mobility (IM) using nitrogen as the mobility gas at a flow rate of 20 mL/min while the Trap, He and Transfer cell flow rates were set to 18.6, 20 and 7 mL/min respectively. This lower pressure regime was used to maintain conditions in the trap cell conducive to stable ion/ion product formation and charge state-based separation of the ion/ion products, as previously reported.[37] A ramping traveling wave velocity of 600 – 3400 m/s over a full IMS cycle and wave height of 40 V were used in the mobility cell. After IM separation, the ions were subjected to ECD fragmentation (using 3.5 eV electrons) in the ExD cell (detailed voltage settings included in Table 1 of supplementary info) before mass analysis in the time-of-flight (TOF) mass spectrometer set to 'Resolution' mode.

Molecular Dynamics Simulations. Gas-phase molecular dynamics simulations of ubiquitin 6^+ ions were performed in triplicate for 2 μs in the canonical ensemble at 298.15 K. To model structures electrosprayed from native-like conditions, the 1UBQ x-ray crystal structure[38] was used as the starting structure for native ubiquitin. To model structures electrosprayed from water/methanol/formic acid solutions, an A state[39-43] structure of ubiquitin was created manually. Following the proposed structure by Ernst and co-workers,[42] the 21 N-terminal residues of 1UBQ were combined with a modeled unfolded C-terminal end, with residues 23-34 and 39-72 forming α-helices, generated with UCSF Chimera.[44] The gas-phase protonation sites for the gas-phase simulations were determined by ECD fragmentation of the 6^+ charge state electrosprayed from both native-like and denaturing conditions (*vide infra*, **Figure 2**). Custom CHARMM patches were generated for protonated proline, glutamine, and arginine residues. Each structure was then minimized with the CHARMM molecular software package (version 44b1)[45-46] and the CHARMM36 protein force field with 250-500 steps of steepest descent and an additional 500

steps of adopted basis Newton-Raphson minimization.[44, 47-48] Prior to gas-phase sampling, the A-state model of ubiquitin was first relaxed in a MD simulation of 50:50 by volume methanol:water mixed-solvent. For this simulation, an explicit solvent box was created with the assistance of the BOSS software package and the MMTSB toolset.[49-51] Ubiquitin was placed at the center of the generated box and the entire condensed system was minimized with CHARMM. Methanol solvent molecules were parameterized with the CHARMM Generalized Force Field (CGenFF) and water was represented with the TIP3P water model.[52-55] MD relaxation was performed with OpenMM[56] in the isothermal-isobaric ensemble at 298.15 K and 1 atm for 100 ns. Periodic boundary conditions were employed with nonbonded cutoffs of 12 Å. At the conclusion of the simulation, the MD trajectory was clustered using the TTClust program[57] and the first clustered structure was used as the starting structure for gas-phase simulations. Triplicate gas-phase MD simulations of native and denatured ubiquitin were then run with OpenMM. Following the conclusion of these calculations, representative structures were generated with TTClust using the program's default clustering criteria. Protein structures were visualized in PyMOL Molecular Graphics System, v1.7.2.0 (Schrödinger, LLC, New York, NY), where the distances between amino acid residue α-carbons as well as the length of the side chain residues (Arg, Lys, Pro) were measured.

Data Analysis. Mass spectra were extracted from selected IM peaks using MassLynx v4.2 (Waters Corp.) and processed using automatic peak detection after background subtraction. The data was exported as .mgf (Mascot Generic Format) files for MS/MS annotation in LCMS spectator[58] (https://omics.pnl.gov/software/lcmsspectator), using a mass error tolerance of 20 ppm[59] and a raw intensity threshold of 1000 counts. The modified and unmodified ECD fragments were identified using the primary sequence of ubiquitin and applying custom modifications at the N and C termini. These corresponded to the mass of the electrostatic adduct formed after the ion/ion reaction, which were 235.987 Da and 616.015 Da for sulfo-NHS acetate and sulfo-EGS

respectively. The annotations were also manually confirmed and imported into Prosight Lite for visualization of the sequence ladders.[60]

Results and Discussion

Electrospray ionization of ubiquitin from aqueous and denaturing conditions produces ions with distinct charge state distributions (**Figure S1**). For the aqueous sample condition, ubiquitin in 10 mM ammonium acetate at pH 7 displayed a distribution of lower charge states ($4 \leq z \leq 6$) while protein prepared with 10 mM acetic acid at pH 3 resulted in higher charge states ($5 \leq z \leq 8$). In aqueous conditions, the structure of ubiquitin is fairly stable under a wide pH range (pH 1.2 to 8.4) and favors a native-like (N) state, which is preserved during the ESI process as the protein transitions from solution to a desolvated state.[28, 39] The lower charge state distribution under aqueous conditions suggests a compact solution structure, with fewer amino acid residues available for protonation.[4, 36] In contrast, the denaturing sample condition (in 50:50 methanol:water and 0.1 % formic acid solution), which stabilizes a partially unfolded A state, resulted in a higher charge state distribution ($5 \leq z \leq 13$). The A state is a highly dynamic conformation where the N-terminal region remains somewhat similar to the native-like β-sheet structure while the C-terminal side, initially rich in β-strand character in the native state, transitions to a more elongated structure with more α-helical character, as observed from NMR and IM-MS studies.[28, 41-42] To identify the putative protonated sites present in different charge states of ubiquitin, the protein was first electrosprayed alone and a specific charge state selected in the quadrupole for subsequent ECD fragmentation after the ion mobility cell. The weighted average charges of the *c* and *z* ions were plotted so that the trend across the sequence highlighted the amino acid position where a change in charge indicated the presence of a proton.

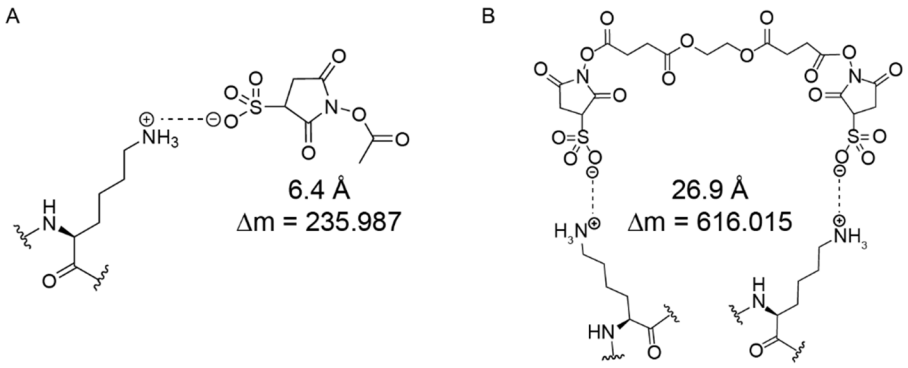
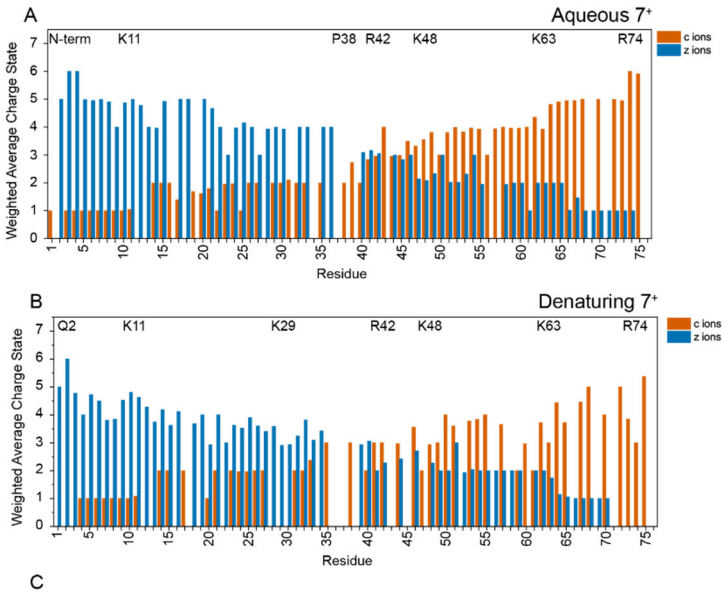


Figure 1. Electrostatic reaction between (A) sulfo-NHS-acetate or (B) sulfo-EGS with a protonated site such as a primary amine from a Lys side chain. The length of the reagent and observed mass shift due to the addition of the reagent are included.

To compare the structural differences between the aqueous, more native-like conformation, and the more extended structure favored in denaturing conditions, ubiquitin was introduced into the trap cell, where it reacted in the gas-phase with anionic reagents. The putative protonated sites, assigned from the charge distribution of the ECD fragments of ubiquitin alone, are potential reactive sites with the anionic reagents. In this study, sulfo-NHS acetate and sulfo-EGS were used since they contain one or more sulfonate groups, which facilitate the electrostatic interactions between the reagent and accessible protonated sites from ubiquitin to form a stable complex.[61] Sulfo-NHS acetate is smaller in size (6.4 Å) and can attach to accessible protonated sites while sulfo-EGS forms a crosslink between two protonated sites within a distance of 26.9 Å (**Figure 1**). After the gas-phase ion/ion reaction took place, the electrostatic product was separated by ion mobility and subjected to ECD fragmentation to identify the modified amino acid residues based on the specific mass shifts in *c*- and *z*-ions due to the presence of the sulfonate-containing reagent. Implementing an ECD cell between the IM and transfer cells allows for improved backbone fragmentation efficiency and unambiguous assignment of labeled sites after the ion/ion reaction.[62] In contrast, the use of a slow-heating activation method such as CID, which is available in the transfer cell of a standard Synapt G2-Si, involves proton migration, preferential sequence-specific backbone cleavages[63] and potential loss of labile protein modifications.[64] Hydrogen-deuterium scrambling has also been observed in CID of hydrogen-deuterium exchange

mass spectrometry, whereas using rapid electronic activation methods results in the avoidance of label scrambling.[65] Thus, ECD was used to avoid mobilizing the reagents and scrambling the electrostatic reagents' attachment sites.

Identification of Putative Protonated Sites for Ubiquitin 6^+ to 8^+ Charge States in Aqueous and Denaturing Conditions. ECD fragmentation was performed for charge states 6^+ to 8^+ electrosprayed from either aqueous or water/methanol solutions (**Figure S2**, see supplementary Table 1 for detailed ECD voltage settings). To identify the most likely protonated sites of the protein, the weighted average charge of the *c*- and *z*-ions at each residue position is calculated and mapped to the protein sequence. For ubiquitin 7^+, the putative protonated sites were similar between the aqueous and denaturing conditions, except for the N-terminus and Pro38 compared to Gln2 and Lys29 respectively (**Figure 2**). Assignment of the first protonated site to the N-terminus and Gln2 can be flexible since c_1 and z_{76} fragments were not always observed to rule out one site versus the other. The weighted average charge state plots for the 6^+ and 8^+ ions are given in **Figure S3**. Major differences between 6^+ ions include that Pro19 is protonated for the aqueous 6^+, whereas Lys11 is protonated for 6^+ electrosprayed from denaturing conditions. For the 8^+ charge state, His68 was protonated when electrosprayed from aqueous conditions, whereas for denaturing conditions, Arg54 was found to be protonated.



C

Charge State	Putative Protonated Sites
Aq. [M + 6H]$^{6+}$	Gln2, Pro19, Lys33, Arg42, Lys63, Arg74
Den. [M + 6H]$^{6+}$	Gln2, Lys11, Lys33, Arg42, Lys63, Arg72
Aq. [M + 7H]$^{7+}$	N-term, Lys11, Pro38, Arg42, Lys48, Lys63, Arg74
Den. [M + 7H]$^{7+}$	Gln2, Lys11, Lys29, Arg42, Lys48, Lys63, Arg74
Aq. [M + 8H]$^{8+}$	N-term, Lys11, Lys29, Arg42, Lys48, Lys63, His68, Arg74
Den. [M + 8H]$^{8+}$	Gln2, Lys11, Lys29, Arg42, Lys48, Arg54, Lys63, Arg74

Figure 2. Weighted average charge state plots of ECD fragments generated by [ubiquitin + 7H]$^{7+}$ in (A) aqueous and (B) denaturing conditions. Protonation sites (labeled at the top) can be identified based on the charge trends across the sequence. Fragment *c*- and *z*-ions are labeled in orange and blue respectively. (C) Summary of the putative protonated sites across ubiquitin 6+ to 8+ in aqueous and denaturing conditions.

Based on the crystal structure of ubiquitin (1UBQ), there are 5 potential ion pairs that are separated by a distance of 6 Å: N-terminus – Glu18, Lys11 – Glu34, Lys27 – Asp52, Lys29 – Asp21, Arg54 – Glu51.[66] Despite the canonical salt bridge distance being 4 Å,[67] it is reasonable to consider longer distances due to the relative strength of gas-phase ion/ion interactions and the gas-phase compaction effect, which also explains the minor difference in collision cross section values from experimental IM-MS measurements compared to those predicted from condensed phase structures obtained by NMR and X-ray crystallography studies.[68] The ion pairs Lys11 – Glu34 and Lys29 – Asp21 have been confirmed to participate in salt bridges by NMR.[43] Hence the identification of Lys11 as a protonated site for ubiquitin (6^+ to 8^+) and Lys29 for ubiquitin (7^+ and 8^+) in denaturing conditions likely indicates a disruption of the Lys11 – Glu34 and Lys29 –

Asp21 salt bridges by methanol since the proton is not shared between the two residues (i.e., the lysine and the acidic residues are both protonated with a net charge of +1). Lys11 (ubiquitin 7^+ and 8^+) and Lys29 (ubiquitin 8^+) are also protonated sites present under aqueous conditions at pH 3, unlike aqueous ubiquitin 6^+ prepared in ammonium acetate buffer, where Pro19 is the charged site, suggesting that the conformation of ubiquitin is starting to diverge from the N state. This is consistent with previous work by the Clemmer group on characterizing the collision cross section distributions of ubiquitin 7^+ and 8^+ by drift tube IM-MS.[29] In that work, ubiquitin 7^+ in 100:0 water:methanol at pH 2 comprised of multiple N states when Gaussian functions were fitted to the resulting arrival time distributions (ATDs) while a 40:60 water:methanol solution produced a mixture of a major A state peak and some N state peaks. Likewise, ubiquitin 8^+ consisted of multiple N state conformers and a minor A state conformer under aqueous conditions and the addition of 60% methanol produced a major A state conformer. Our results are also consistent with previous work from our group measuring ATDs in nitrogen with TWIMS. From ubiquitin charge states electrosprayed from various solution conditions, we observed a sharp peak at 1233 $Å^2$ for the 6^+ charge state from aqueous conditions (N state), but an intermediate state peak at 1398 $Å^2$ and a broad distribution centered around 1600 $Å^2$ from denaturing conditions, likely nitrogen-stabilized peaks resulting from various A states in solution. In addition, the putative protonated sites identified in this study are also fairly consistent with ones identified by UVPD fragmentation of ubiquitin sprayed from an ammonium acetate solution. Lys11, Lys27/29, Arg42, Arg54, His68 and Arg74 were identified as the protonated sites of a minor ubiquitin 7^+ charge site isomer and the charged sites for ubiquitin 8^+ were N-terminus, Lys11, Lys27, Pro37, Lys48, Asn60, His68 and Arg74.[66] Thus, comparison of the assigned charged sites across different charge states and sample conditions reflects the changes in solution structure caused by the addition of methanol.

Finding the Most Accessible Protonated Sites with Sulfo-NHS Acetate. The assigned protonated sites are all candidate targets for electrostatic attachment by sulfo-NHS acetate. The ATD of the IM peaks observed after the ion/ion reaction between ubiquitin 7^+ and sulfo-NHS acetate shows up to 4 reactions occurring (**Figure 3A**). The extracted mass spectrum of the IM peak (drift time = 7.3 ms) shows two major peaks (1223.81 and 1427.61 *m/z*), corresponding to the unmodified ubiquitin 7^+ and the $[M + 7H]^{6+}$ charge state resulting from a charge reduction via electrons produced for ECD fragmentation (**Figure 3B**). A zoomed-in view of the 2300 – 2800 m/z region displays some of the *c*- and *z*-ions (**Figure 3C**). The IM peak at drift time = 8.7 ms represents one electrostatic attachment to ubiquitin 7^+ ($[M+ 7H + ♦]^{6+}$ extracted peak at 1467.28 *m/z*) and ubiquitin 6^+ at 1427.61 *m/z*, produced by proton transfer from ubiquitin 7^+ to the reagent. The zoomed-in view shows the modified ECD fragments in the 2300 – 2800 *m/z* range (**Figure 3D and E**).

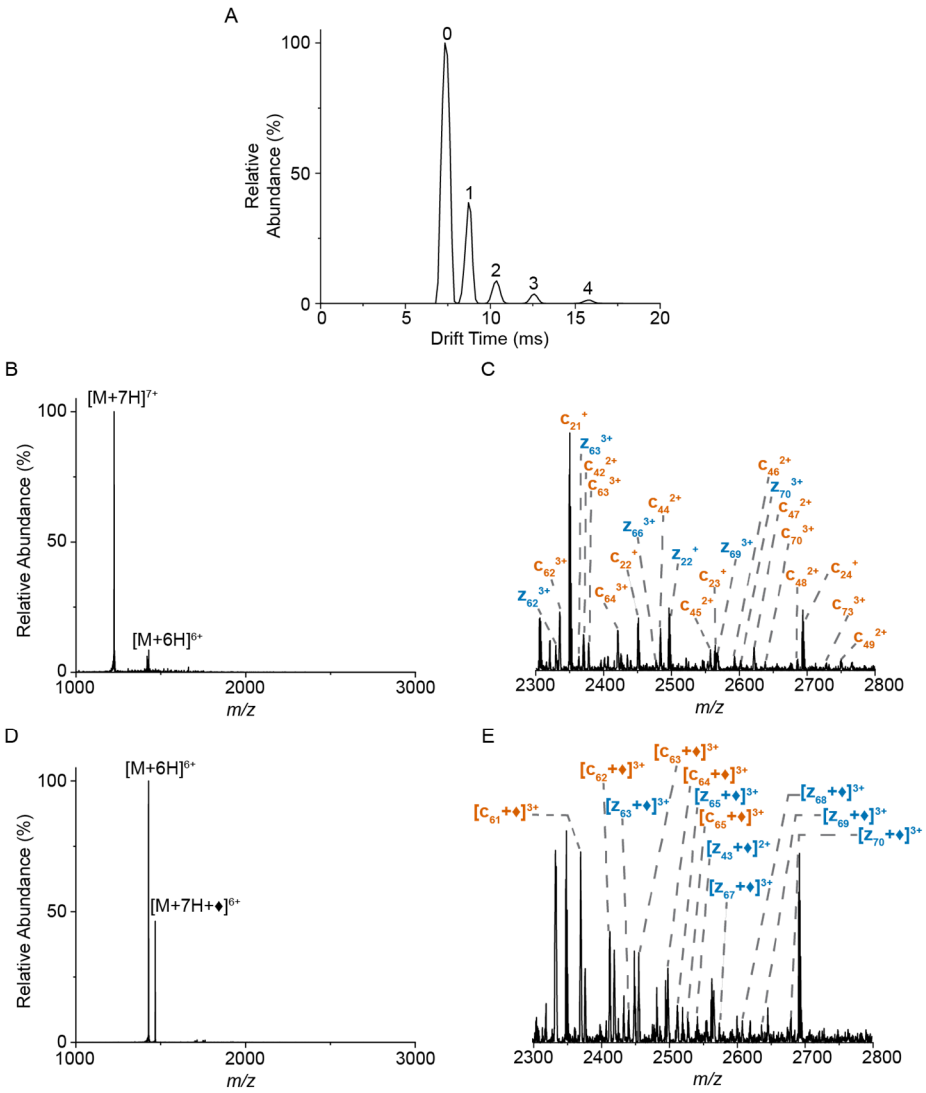


Figure 3. Electrostatic reaction between aqueous [ubiquitin + 7H]$^{7+}$ and sulfo-NHS acetate. (A) Ion mobility spectrum following ion/ion reaction, with numbered labels corresponding to the number of electrostatic attachments. (B) Mass spectrum extracted from 'zero' IM peak and (C) product ion spectrum, displaying ECD fragments from unmodified [ubiquitin + 7H]$^{7+}$. (D) Mass spectrum from one electrostatic attachment (denoted by ♦) and (E) product ion spectrum of ECD fragments generated by [ubiquitin + 7H + ♦]$^{6+}$ precursor ion.

By assigning a mass shift of 235.987 Da to both termini, all the observed modified *c*- and *z*-ions were identified, allowing the sites of electrostatic attachment of sulfo-NHS acetate to be determined. Since there are multiple protonated sites, the precursor ion corresponding to one ion/ion reaction with sulfo-NHS acetate is expected to be a mixture of proteins with different modified sites. The most probable sites for sulfo-NHS acetate attachment as determined by ECD fragmentation are highlighted in gray in **Figure 4**. Since methanol disrupts the tertiary structure of ubiquitin, it is not surprising that residues involved in salt bridges (Lys11 and Lys29) become

available for reaction with sulfo-NHS acetate, as observed across ubiquitin charge states 6^+ – 8^+ electrosprayed from denaturing conditions (**Figure 4B**). In contrast, aqueous ubiquitin is modified at different sites (**Figure 4A**), providing further evidence that aqueous ubiquitin exists as different conformers compared to the equivalent charge state from denaturing condition. Of note, Lys29 becomes one of the sites that reacts with sulfo-NHS acetate, suggesting that the higher charge state (8^+) influences the stability of the salt bridges and overall protein conformation.

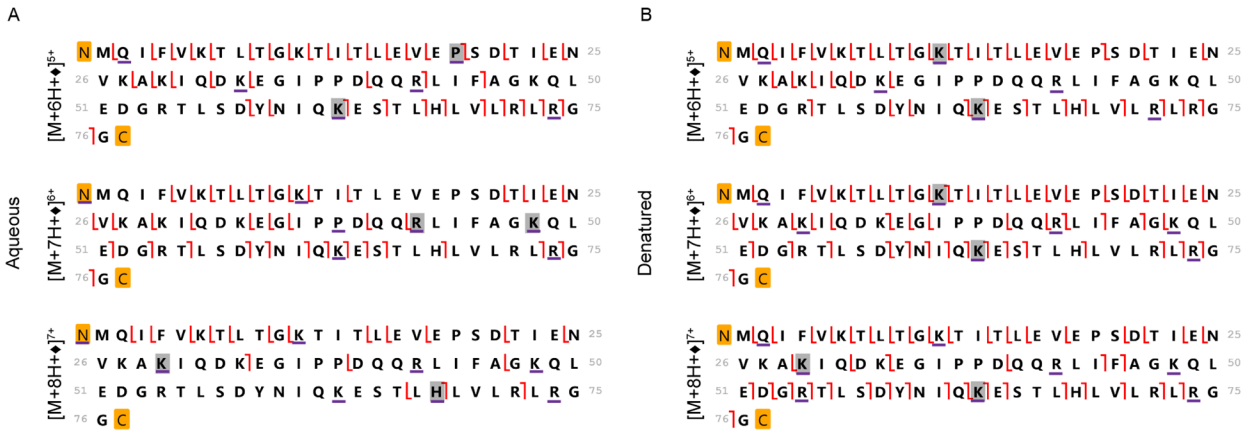


Figure 4. Sequence ladders generated from the ECD fragmentation of the ion/ion product between sulfo-NHS acetate and different charge states of ubiquitin, sprayed from (A) aqueous and (B) denatured sample conditions. Residues, where sulfo-NHS acetate likely binds, are denoted in gray. Putative protonated sites (inferred from weighted average charge state plots) are underlined in purple and the N and C termini are labeled in yellow.

Cross-linking Positively Charged Residues with Sulfo-EGS. We have exploited the strength of electrostatic bonds in the gas phase, inspired by the observation by McLuckey and coworkers that sulfonated cross-linkers bind electrostatically via ion/ion chemistry and that *b*- and *y*-ions produced by CID of the electrostatically-bound protein-cross-linker complex were bound to cross-linkers without observing covalent chemistry (through neutral loss of the sulfo-NHS leaving groups).[25] In this study, ECD was used to avoid ambiguity in assigning cross-linking sites due to the possibility of scrambling inherent in slow-heating activation methods like CID. For gas-phase structural measurements, the linker length of the anionic reagent provides an inherent distance constraint that can inform on proximities of the cross-linked sites and overall protein conformation. Sulfo-EGS has two sulfonate groups that can participate in electrostatic interactions with positively charged residues of ubiquitin. Fully extended, it spans across 26.9 Å, unlike sulfo-NHS acetate

which contains one sulfonate group and is approximately 6.4 Å (**Figure 1**). While its smaller size and single-site electrostatic reactivity allows sulfo-NHS acetate to access the most accessible protonated sites on a protein, sites that are cross-linked by sulfo-EGS will be driven by spatial constraints. The IM peaks resulting post ion/ion reaction between ubiquitin 7^+ and $[sulfo\text{-}EGS]^{2-}$ show that there are up to 2 sulfo-EGS attachments, and the small peak (drift time = 9.54 ms) contains an m/z that is assigned to the cation-to-anion proton transfer from ubiquitin to the doubly deprotonated linker reagent, followed by subsequent attachment of the nascent singly charged reagent to the protein (**Figure 5A**). The IM peak at drift time = 10.5 ms represents one electrostatic attachment to ubiquitin 7^+ ($[M+ 7H + ♦]^{5+}$ extracted peak at 1836.74 *m/z*). The zoomed-in view shows the modified ECD fragments in the 2300 – 2800 *m/z* range (**Figure 5B and C**).

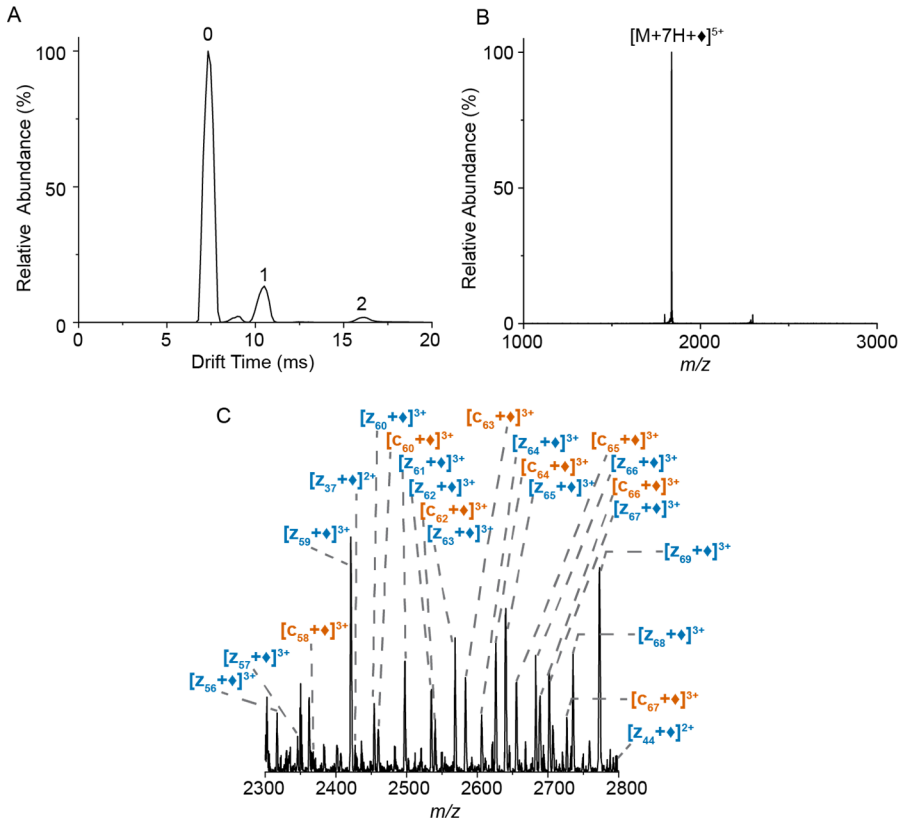


Figure 5. Gas-phase crosslinking of sulfo-EGS to aqueous $[ubiquitin + 7H]^{7+}$. (A) Ion mobility spectrum following ion/ion reaction, with numbered labels corresponding to the number of electrostatic attachments. (B) Mass spectrum extracted from IM peak with one electrostatic attachment (denoted by ♦). (C) Product ion spectrum of ECD fragments generated by $[ubiquitin + 7H + ♦]^{5+}$ precursor ion.

Following the same strategy used to identify residues modified by sulfo-NHS acetate, here a mass shift of 616.016 Da is assigned to both termini to identify cross-linked *c*- and *z*-ions and the fragment ions mapped to the protein sequence. There is a distinct lack of fragments for the region of the primary sequence spanned by the cross-linker in comparison to the results obtained with sulfo-NHS acetate where fragments were observed in the entire protein sequence, indicating that the reagent is not anchored at two residues (**Figure 6**). The cross-linker keeps the larger peptide together and effectively silences the observed ECD fragment ions for that region. The cross-linked sites for aqueous ubiquitin 6^+ and 7^+ are Pro19 – Arg42 and Arg42 – Lys48 respectively and mostly reflect the individual reactivity of the protonated sites (Pro19, Arg42 and Lys48) identified by sulfo-NHS acetate. Multiple cross-linking positions infer that there are structural isomers, which correlates with the observation of many resolvable ion mobility peaks within observed single ion mobility peaks by using high-resolution tandem ion mobility experiments.[69-71] Despite Lys63 being identified as a reactive site for aqueous ubiquitin 8^+ and denaturing ubiquitin 6^+ to 8^+, it is not a site that is targeted by sulfo-EGS. Instead, for ubiquitin 6^+ electrosprayed from denaturing conditions, the cross-linked pair is Lys33 – Arg72 while the 7^+ and 8^+ charge states have Arg42 – Lys48 as their cross-linked sites. Since Lys48 becomes protonated only once ubiquitin reaches a charge state of 7^+ or greater, it is not available in the 6^+ charge state for sulfo-EGS attachment. Moreover, Lys33 is a putative protonated site in ubiquitin 6^+ only while Lys29 is the positively charged site for the higher charge states. This likely indicates that the increase in available protons disrupts the protein structure and the structure of the 6^+ charge state is maintained by different tertiary interactions compared to the 7^+ and 8^+ states.

Figure 6. Sequence ladders generated from the ECD fragmentation of the ion/ion product between sulfo-EGS and different charge states of ubiquitin, sprayed from (A) aqueous and (B) denatured sample conditions. Crosslinked residues are shaded in gray. Putative protonated sites (inferred from weighted average charge state plots) are underlined in purple and the N and C termini are labeled in yellow.

Of note, the ion/ion reaction between aqueous ubiquitin 8^+ and sulfo-EGS does not result in the formation of cross-linked sites. Instead, Lys48 becomes modified with a one-sided, "hanging" cross-link, suggesting that sulfo-EGS first abstracts a proton through a proton transfer ion/ion reaction, thereby leaving only one effective negative charge for electrostatic interaction to Lys48 for the second ion/ion that we hypothesize gives rise to this gas-phase electrostatic hanging cross-linker phenomenon. The extracted *m/z* spectrum shows a major peak at 1530.95 *m/z*, corresponding to $[M + 7H + ♦]^{6+}$ resulting from ubiquitin 8^+, which first undergoes charge reduction to 7^+ and then reacts with singly protonated sulfo-EGS, additional evidence for hanging cross-linked peak being formed by proton transfer to the reagent and electrostatic attachment. Compared to the aqueous ubiquitin 8^+, the product formed by denaturing ubiquitin 8^+ is $[M + 8H + ♦]^{6+}$, with a larger average *m/z* of 1531.28 ($\Delta m/z = H^+$). The protonated sites are similar for aqueous and denaturing ubiquitin 8^+, except for His68 and Arg54. Hence, one possible explanation is that the proton from His68 in the aqueous state is more susceptible to a charge-reduction reaction relative to the one from Arg54 in denaturing condition, possibly due to the decreased proton affinity of histidine versus arginine.[72-73] Overall, comparing the sites cross-linked by sulfo-EGS among the different charge states and sample conditions presents some additional insight into the differences in the structure of ubiquitin.

Towards Combining the Results of Various Ion/Ion Reactions to Guide Modeling of Gas-phase Structure. While CCS measurements are useful in determining the overall protein structure, the wealth of information obtained by using different electrostatic reagents has the potential to piece together a more detailed understanding of the modified amino acid residues and their roles during subtle changes in conformations. Protonation sites obtained from the ECD fragment analysis can be used as constraints to simulate gas-phase conformations from a solution-phase structure by molecular dynamics. This approach was used to generate clusters of gas-phase structures resulting from very long (2 μs) molecular dynamics simulations modeling ubiquitin 6^+ from aqueous and denaturing conditions with structures starting from the crystal structure and A state. Three distinct clusters were produced from the modeling of the A state: namely, a more compact, an intermediate and a more elongated structure (**Figure S4**). These gas-phase structure models are useful for measuring the relative positions between modified sites, which can be used to predict why reactions with certain residues are preferred over others. By estimating the maximum distance that a cross-linking reagent such as sulfo-EGS can span between two protonated sites, this narrows down the range of interacting residues based on their relative positions in each structure model (**Fig. S5A and B**). In the case of sulfo-EGS, the distance estimated between the cross-linked sites all fall within the maximum estimate (which includes the length of the side chains, plus 4 Å as a first approximation for the length of each gas-phase electrostatic bond, and the length of the fully extended cross-linker between sulfonates). The Pro19 C_α – Arg42 C_α distance is 19.2 Å, well below the maximum estimate of 43.7 Å for the aqueous condition while the Lys C_α – Arg C_α distances for the more compact, intermediate and more elongated conformers are 41.2 Å, 36.1 Å and 42.2 Å, within the estimated limit of 48.6 Å (**Supplementary Table 2A**). This analysis is not informative enough to discern which of the denaturing conformers reacts with sulfo-EGS, owing to the linker length being too long for a relatively small protein such as ubiquitin. However, the cross-linking sites are often different than the most accessible sites identified by the sulfo-NHS acetate electrostatic addition, clear evidence

that even a linker of this length is sensitive to the three-dimensional structure of the protein. Thus, future studies involving a range of linker lengths and proteins will be employed to explore the effects of electrostatic cross-linking linker length on the identities and positions of cross-linked sites.

We have also previously published on the gas phase ion/ion reaction between ubiquitin and sulfo-benzoyl HOAt, where the sulfonate group first attaches to an accessible protonated site, followed by a covalent reaction with a nearby nucleophilic residue and the loss of the HOAt leaving group.[23] In addition, sulfo-benzoyl HOAt and sulfo-NHS acetate sulfonates likely have similar reactivity with protonated side chains. Since the through-bond distance from the reactive carbonyl carbon to the sulfonate oxygens in the reagent is approximately 6.4 Å, this provides a spatial constraint for the possible sites for covalent modification (**Figure S5**), representing electrostatic-covalent heterofunctional cross-linking. The covalent sites for aqueous ubiquitin 6^+ were Lys48 and Arg54, which agrees with our assignments of protonated sites. For Lys48 and Arg54 side chains to react as nucleophiles, they must not be protonated. Identifying the protonated sites for each charge state is useful in determining the charged residue for the first binding step, where the sulfonate group binds to a protonated site to form the electrostatic complex. Considering distances derived from molecular dynamics simulations of the aqueous 6^+ ions and the length of the sulfo-benzoyl HOAt reagent, the most accessible protonated sites (as identified by reaction with sulfo-NHS acetate) are within reasonable distance to covalently modified sites. Thus, for aqueous 6^+, sulfo-benzoyl HOAt first likely forms an electrostatic attachment to Pro19 before the nucleophilic attack by the unprotonated Lys29 (Pro C_α – Lys C_α distance of 13.6 Å, which is within the maximum estimate of 18.3 Å) as well as an electrostatic complex at Lys63 before the covalent reaction at Arg54 (Lys C_α – Arg C_α distance of 16.4 Å, which is below the maximum estimate of 24.1 Å) (**Supplementary Table 2B**). Similarly, this can be applied to the three gas-phase structures that are modeled after the in-solution A state structure. Consistent with the aqueous 6^+

structure, Lys63 is the protonated site to which the sulfonate group binds before the subsequent covalent reaction with Arg54. The C_α – C_α distance measured between Arg54 and Lys63 for all three conformers are lower than the maximum limit (**Supplementary Table 2B**). However, when protonated Lys11 reacts as the electrostatic site to position sulfo-benzoyl HOAt for a covalent reaction with Lys48, only the more compact structure provides an acceptable distance between the Lys48 and Lys11 (Lys11 C_α – Lys48 C_α equal to 16.6 Å and within the maximum estimate of 23.2 Å). In contrast, the Lys11 C_α – Lys48 C_α distance for the intermediate structure is 25.9 Å while the one for the more elongated structure is 35.5 Å. In this case, analyzing the distances between the sites modified by sulfo-benzoyl HOAt provides evidence for the presence of one conformer over other possible ones generated by molecular dynamics. The observed electrostatic-covalent and electrostatic-electrostatic cross-links are shown in the context of the MD predicted structures in **Figure 7**.

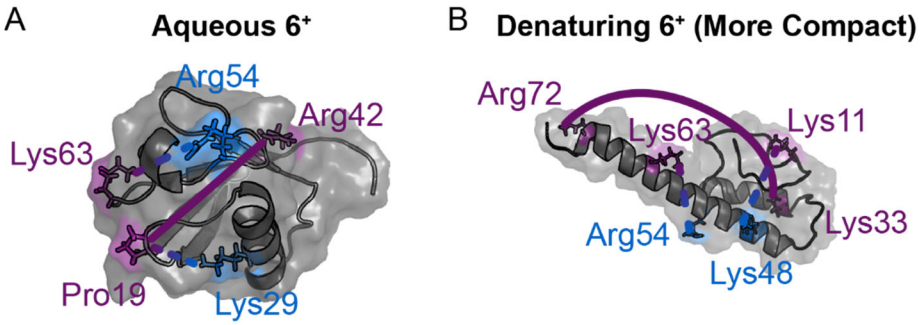


Figure 7. Modified residues represented using MD predicted structures of ubiquitin with protonated and covalent sites labeled in purple and blue respectively. (A) Aqueous 6+ is cross-linked by sulfo-EGS at Pro19 – Arg42 (solid line) and by sulfo-benzoyl HOAt at Lys29 – Pro19 and Arg54 – Lys63 (dashed lines). (B) Under denaturing condition, the pair linked by sulfo-EGS is Lys33 – Arg72 (solid line) and the pairs cross-linked by sulfo-benzoyl HOAt are Lys48 – Lys11 and Arg54 – Lys63 (dashed lines).

Conclusions. Tracking the increase in charge along the charge state distributions of ECD fragments generated from ubiquitin under different conditions allows the determination of protonated sites and provides insight on the residues that may be involved in salt bridges and other intramolecular interactions contributing to the overall protein structure. The sites identified in this study are in agreement with results published previously using 193 nm UVPD to fragment

ubiquitin 6^+ – 8^+ electrosprayed from aqueous condition.[66] The charged sites found in the structures generated from aqueous and denaturing conditions are distinct, demonstrating the combined effect of methanol and low pH on disrupting the intramolecular interactions maintaining the native structure of ubiquitin. This corroborates gas-phase structural data obtained by drift tube IM-MS demonstrating the presence of multiple N state conformers and A state intermediate structures for ubiquitin 7^+ and 8^+ under aqueous and denaturing conditions respectively.[29]

Leveraging the results obtained by specific ion/ion reaction to discern among the relevant structures generated during molecular dynamics studies is a promising approach to learn more about the subtle changes in gas-phase structures that may be overlooked when using collisional cross-section (CCS) measurements to interpret changes in protein structure. For example, work by the Clemmer group has shown that under aqueous, native-like conditions, the compact IM peak of ubiquitin 7^+ has a CCS of ~ 1010 Å, however that same peak can be fitted to multiple Gaussian functions representing substructures existing within that one peak.[29] Thus, additional work is needed in analyzing the other ubiquitin charge states to further demonstrate the benefit of this promising strategy. Also, comprehensive studies of more proteins and gas-phase reactions with electrostatic and covalent cross-linkers of different lengths are required to establish a framework for the structural accuracy and precision of this approach so that it becomes possible to tailor specific reagents based on the structure being mapped.

Acknowledgements

This work was supported by the National Institutes of Health (NIH) under Grant R21GM134408 (IKW), the Indiana University-Purdue University Indianapolis School of Science (IKW), and by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute (JZV). The authors acknowledge the Indiana University Pervasive Technology Institute for providing supercomputing and storage resources that have contributed to the research results reported within this paper. Dr. Valery Voinov and Ms. Diana Oppenheimer

of e-MSion are acknowledged for installation of the ExD cell and helpful discussions. The authors acknowledge Dr. Jared Shaw of e-MSion for helpful discussions.

References

1. van den Heuvel, R. H.; Heck, A. J., Native protein mass spectrometry: from intact oligomers to functional machineries. *Current opinion in chemical biology* **2004,** *8* (5), 519-526.
2. Benesch, J. L.; Ruotolo, B. T., Mass spectrometry: come of age for structural and dynamical biology. *Current opinion in structural biology* **2011,** *21* (5), 641-649.
3. Loo, J. A.; He, J. X.; Cody, W. L., Higher Order Structure in the Gas Phase Reflects Solution Structure. *Journal of the American Chemical Society* **1998,** *120* (18), 4542-4543.
4. Ruotolo, B. T.; Robinson, C. V., Aspects of native proteins are retained in vacuum. *Curr. Opin. Chem. Biol.* **2006,** *10* (5), 402-408.
5. Loo, J. A.; Loo, R. R. O.; Udseth, H. R.; Edmonds, C. G.; Smith, R. D., Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **1991,** *5* (3), 101-105.
6. Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass spectrometry reviews* **1997,** *16* (1), 1-23.
7. Ruotolo, B. T.; Giles, K.; Campuzano, I.; Sandercock, A. M.; Bateman, R. H.; Robinson, C. V., Evidence for macromolecular protein rings in the absence of bulk water. *Science* **2005,** *310* (5754), 1658-1661.
8. 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., Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nature chemistry* **2009,** *1* (4), 326-331.
9. Walzthoeni, T.; Leitner, A.; Stengel, F.; Aebersold, R., Mass spectrometry supported determination of protein complex structure. *Current opinion in structural biology* **2013,** *23* (2), 252-260.
10. Konijnenberg, A.; Butterer, A.; Sobott, F., Native ion mobility-mass spectrometry and related methods in structural biology. *Biochim. Biophys. Acta* **2013,** *1834* (6), 1239-1256.
11. Jurneczko, E.; Barran, P. E., How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **2011,** *136* (1), 20-28.
12. Mendoza, V. L.; Vachet, R. W., Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass spectrometry reviews* **2009,** *28* (5), 785-815.
13. Hermanson, G. T., Chapter 2 - Functional Targets for Bioconjugation. In *Bioconjugate Techniques (Third Edition)*, Hermanson, G. T., Ed. Academic Press: Boston, 2013; pp 127-228.
14. Limpikirati, P.; Pan, X.; Vachet, R. W., Covalent Labeling with Diethylpyrocarbonate: Sensitive to the Residue Microenvironment, Providing Improved Analysis of Protein Higher Order Structure by Mass Spectrometry. *Anal. Chem.* **2019,** *91* (13), 8516-8523.
15. Kalkhof, S.; Sinz, A., Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Analytical and bioanalytical chemistry* **2008,** *392* (1-2), 305-312.
16. Iacobucci, C.; Piotrowski, C.; Aebersold, R.; Amaral, B. C.; Andrews, P.; Bernfur, K.; Borchers, C.; Brodie, N. I.; Bruce, J. E.; Cao, Y.; Chaignepain, S.; Chavez, J. D.; Claverol, S.; Cox, J.; Davis, T.; Degliesposti, G.; Dong, M.-Q.; Edinger, N.; Emanuelsson, C.; Gay, M.; Götze, M.; Gomes-Neto, F.; Gozzo, F. C.; Gutierrez, C.; Haupt, C.; Heck, A. J. R.; Herzog, F.; Huang, L.; Hoopmann, M. R.; Kalisman, N.; Klykov, O.; Kukačka, Z.; Liu, F.; MacCoss, M. J.; Mechtler, K.; Mesika, R.; Moritz, R. L.; Nagaraj, N.; Nesati, V.; Neves-Ferreira, A. G. C.; Ninnis, R.; Novák, P.; O'Reilly, F. J.; Pelzing, M.; Petrotchenko, E.; Piersimoni, L.; Plasencia, M.; Pukala, T.; Rand, K. D.; Rappsilber, J.; Reichmann, D.; Sailer, C.; Sarnowski, C. P.; Scheltema, R. A.; Schmidt, C.; Schriemer, D. C.; Shi, Y.; Skehel, J. M.; Slavin, M.; Sobott, F.; Solis-Mezarino, V.; Stephanowitz, H.; Stengel, F.; Stieger, C. E.; Trabjerg, E.; Trnka, M.; Vilaseca, M.; Viner, R.; Xiang, Y.; Yilmaz, S.; Zelter, A.; Ziemianowicz, D.; Leitner, A.; Sinz, A., First Community-Wide,

Comparative Cross-Linking Mass Spectrometry Study. *Analytical Chemistry* **2019,** *91* (11), 6953-6961.
17. Liu, F.; Rijkers, D. T.; Post, H.; Heck, A. J., Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nature methods* **2015,** *12* (12), 1179-1184.
18. Sinz, A., Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom. Rev.* **2006,** *25* (4), 663-82.
19. Kruppa, G. H.; Schoeniger, J.; Young, M. M., A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003,** *17* (2), 155-162.
20. Cammarata, M. B.; Brodbelt, J. S., Characterization of Intra-and Intermolecular Protein Crosslinking by Top Down Ultraviolet Photodissociation Mass Spectrometry. *ChemistrySelect* **2016,** *1* (3), 590-593.
21. Novak, P.; Haskins, W. E.; Ayson, M. J.; Jacobsen, R. B.; Schoeniger, J. S.; Leavell, M. D.; Young, M. M.; Kruppa, G. H., Unambiguous Assignment of Intramolecular Chemical Cross-Links in Modified Mammalian Membrane Proteins by Fourier Transform-Tandem Mass Spectrometry. *Anal. Chem.* **2005,** *77* (16), 5101-5106.
22. Foreman, D. J.; McLuckey, S. A., Recent Developments in Gas-Phase Ion/Ion Reactions for Analytical Mass Spectrometry. *Anal. Chem.* **2020,** *92* (1), 252-266.
23. Carvalho, V. V.; See Kit, M. C.; Webb, I. K., Ion Mobility and Gas-Phase Covalent Labeling Study of the Structure and Reactivity of Gaseous Ubiquitin Ions Electrosprayed from Aqueous and Denaturing Solutions. *J. Am. Soc. Mass. Spectrom.* **2020,** *31* (5), 1037-1046.
24. Webb, I. K.; Morrison, L. J.; Brown, J., Dueling electrospray implemented on a traveling-wave ion mobility/time-of-flight mass spectrometer: Towards a gas-phase workbench for structural biology. *Int. J. Mass spectrom.* **2019,** *444*, 116177.
25. Webb, I. K.; Mentinova, M.; McGee, W. M.; McLuckey, S. A., Gas-phase intramolecular protein crosslinking via ion/ion reactions: ubiquitin and a homobifunctional sulfo-NHS ester. *J. Am. Soc. Mass. Spectrom.* **2013,** *24* (5), 733-43.
26. Gologan, B.; Takáts, Z.; Alvarez, J.; Wiseman, J. M.; Talaty, N.; Ouyang, Z.; Cooks, R. G., Ion soft-landing into liquids: Protein identification, separation, and purification with retention of biological activity. *J. Am. Soc. Mass. Spectrom.* **2004,** *15* (12), 1874-1884.
27. Silveira, J. A.; Fort, K. L.; Kim, D.; Servage, K. A.; Pierson, N. A.; Clemmer, D. E.; Russell, D. H., From solution to the gas phase: stepwise dehydration and kinetic trapping of substance P reveals the origin of peptide conformations. *J. Am. Chem. Soc.* **2013,** *135* (51), 19147-53.
28. Wyttenbach, T.; Bowers, M. T., Structural Stability from Solution to the Gas Phase: Native Solution Structure of Ubiquitin Survives Analysis in a Solvent-Free Ion Mobility–Mass Spectrometry Environment. *The Journal of Physical Chemistry B* **2011,** *115* (42), 12266-12275.
29. Shi, H.; Clemmer, D. E., Evidence for two new solution states of ubiquitin by IMS-MS analysis. *The journal of physical chemistry. B* **2014,** *118* (13), 3498-3506.
30. Bleiholder, C.; Liu, F. C., Structure Relaxation Approximation (SRA) for Elucidation of Protein Structures from Ion Mobility Measurements. *J. Phys. Chem. B* **2019,** *123* (13), 2756-2769.
31. Bakhtiari, M.; Konermann, L., Protein Ions Generated by Native Electrospray Ionization: Comparison of Gas Phase, Solution, and Crystal Structures. *The Journal of Physical Chemistry B* **2019,** *123* (8), 1784-1796.
32. Xia, Y.; Chrisman, P. A.; Erickson, D. E.; Liu, J.; Liang, X.; Londry, F. A.; Yang, M. J.; McLuckey, S. A., Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. *Anal. Chem.* **2006,** *78* (12), 4146-4154.
33. Williams, J. P.; Morrison, L. J.; Brown, J. M.; Beckman, J. S.; Voinov, V. G.; Lermyte, F., Top-Down Characterization of Denatured Proteins and Native Protein Complexes Using

Electron Capture Dissociation Implemented within a Modified Ion Mobility-Mass Spectrometer. *Anal. Chem.* **2020,** *92* (5), 3674-3681.
34. Giles, K.; Pringle, S. D.; Worthington, K. R.; Little, D.; Wildgoose, J. L.; Bateman, R. H., Applications of a travelling wave-based radio-frequencyonly stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **2004,** *18* (20), 2401-2414.
35. Giles, K.; Williams, J. P.; Campuzano, I., Enhancements in travelling wave ion mobility resolution. *Rapid Commun. Mass Spectrom.* **2011,** *25* (11), 1559-66.
36. Wyttenbach, T.; Bowers, M. T., Structural stability from solution to the gas phase: native solution structure of ubiquitin survives analysis in a solvent-free ion mobility-mass spectrometry environment. *J. Phys. Chem. B* **2011,** *115* (42), 12266-75.
37. Carvalho, V. V.; Cheung See Kit, M.; Webb, I., An Ion Mobility and Gas-Phase Covalent Labeling Study of the Structure and Reactivity of Gaseous Ubiquitin Ions Electrosprayed from Aqueous and Denaturing Solutions. *Journal of the American Society for Mass Spectrometry* **2020**.
38. Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J., Structure of ubiquitin refined at 1.8Åresolution. *J. Mol. Biol.* **1987,** *194* (3), 531-544.
39. Lenkinski, R. E.; Chen, D. M.; Glickson, J. D.; Goldstein, G., Nuclear magnetic resonance studies of the denaturation of ubiquitin. *Biochimica et Biophysica Acta (BBA)-Protein Structure* **1977,** *494* (1), 126-130.
40. Harding, M. M.; Williams, D. H.; Woolfson, D. N., Characterization of a Partially Denatured State of a Protein by 2-Dimensional Nmr - Reduction of the Hydrophobic Interactions in Ubiquitin. *Biochemistry* **1991,** *30* (12), 3120-3128.
41. Stockman, B. J.; Euvrard, A.; Scahill, T. A., Heteronuclear three-dimensional NMR spectroscopy of a partially denatured protein: the A-state of human ubiquitin. *Journal of biomolecular NMR* **1993,** *3* (3), 285-296.
42. Brutscher, B.; Brüschweiler, R.; Ernst, R. R., Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by 1H, 13C, and 15N nuclear magnetic resonance spectroscopy. *Biochemistry* **1997,** *36* (42), 13043-13053.
43. Sundd, M.; Iverson, N.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M.; Robertson, A. D., Electrostatic interactions in ubiquitin: stabilization of carboxylates by lysine amino groups. *Biochemistry* **2002,** *41* (24), 7586-7596.
44. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004,** *25* (13), 1605-12.
45. Brooks, B. R. B., R. E.; Olafson, B. D.; States, B. D.; Swaminathan, S.; Karplus, M. , CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983,** *4*, 187-217.
46. Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009,** *30* (10), 1545-1614.
47. Best, R. B.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr., Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of α-helix and β-hairpin formation. *Biophys. J.* **2012,** *103* (5), 1045-51.
48. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput.* **2012,** *8* (9), 3257-3273.

49. Jorgensen, W. L.; Tirado–Rives, J., Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005,** *26* (16), 1689-1700.
50. Jorgensen, W. L. *BOSS*, Yale University: New Haven, CT, 2011.
51. Feig, M.; Karanicolas, J.; Brooks, C. L., 3rd, MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **2004,** *22* (5), 377-95.
52. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010,** *31* (4), 671-90.
53. Vanommeslaeghe, K.; MacKerell, A. D., Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012,** *52* (12), 3144-3154.
54. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D., Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012,** *52* (12), 3155-3168.
55. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79* (2), 926-935.
56. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S., OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017,** *13* (7), e1005659.
57. Tubiana, T.; Carvaillo, J.-C.; Boulard, Y.; Bressanelli, S., TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries. *J. Chem. Inf. Model.* **2018,** *58* (11), 2178-2182.
58. Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; Moore, R. J.; Liu, T.; Petyuk, V. A.; Tolić, N.; Paša-Tolić, L.; Smith, R. D.; Payne, S. H.; Kim, S., Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **2017,** *14* (9), 909-914.
59. Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N., Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* **2019,** *16* (7), 587-594.
60. Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M., ProSight Lite: graphical software to analyze top-down mass spectrometry data. *Proteomics* **2015,** *15* (7), 1235-8.
61. Mentinova, M.; McLuckey, S. A., Covalent modification of gaseous peptide ions with N-hydroxysuccinimide ester reagent ions. *Journal of the American Chemical Society* **2010,** *132* (51), 18248-18257.
62. Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W., Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Analytical Chemistry* **2000,** *72* (3), 563-573.
63. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Mobile and localized protons: a framework for understanding peptide dissociation. *Journal of Mass Spectrometry* **2000,** *35* (12), 1399-1406.
64. Paizs, B.; Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005,** *24* (4), 508-548.

65. Wollenberg, D. T. W.; Pengelley, S.; Mouritsen, J. C.; Suckau, D.; Jørgensen, C. I.; Jørgensen, T. J. D., Avoiding H/D Scrambling with Minimal Ion Transmission Loss for HDX-MS/MS-ETD Analysis on a High-Resolution Q-TOF Mass Spectrometer. *Anal. Chem.* **2020,** *92* (11), 7453-7461.
66. Morrison, L. J.; Brodbelt, J. S., Charge site assignment in native proteins by ultraviolet photodissociation (UVPD) mass spectrometry. *Analyst* **2016,** *141* (1), 166-176.
67. Kumar, S.; Nussinov, R., Close-range electrostatic interactions in proteins. *ChemBioChem* **2002,** *3* (7), 604-617.
68. Rolland, A. D.; Prell, J. S., Computational insights into compaction of gas-phase protein and protein complex ions in native ion mobility-mass spectrometry. *TrAC Trends in Analytical Chemistry* **2019,** *116*, 282-291.
69. Koeniger, S. L.; Clemmer, D. E., Resolution and structural transitions of elongated states of ubiquitin. *J. Am. Soc. Mass. Spectrom.* **2007,** *18* (2), 322-331.
70. Koeniger, S. L.; Merenbloom, S. I.; Clemmer, D. E., Evidence for Many Resolvable Structures within Conformation Types of Electrosprayed Ubiquitin Ions. *The Journal of Physical Chemistry B* **2006,** *110* (13), 7017-7021.
71. Ridgeway, M. E.; Silveira, J. A.; Meier, J. E.; Park, M. A., Microheterogeneity within conformational states of ubiquitin revealed by high resolution trapped ion mobility spectrometry. *Analyst* **2015,** *140* (20), 6964-6972.
72. Schnier, P. D.; Gross, D. S.; Williams, E. R., On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization. *J. Am. Soc. Mass. Spectrom.* **1995,** *6* (11), 1086-1097.
73. Bleiholder, C.; Suhai, S.; Paizs, B., Revising the Proton Affinity Scale of the Naturally Occurring α-Amino Acids. *J. Am. Soc. Mass. Spectrom.* **2006,** *17* (9), 1275-1281.

Supplemental information

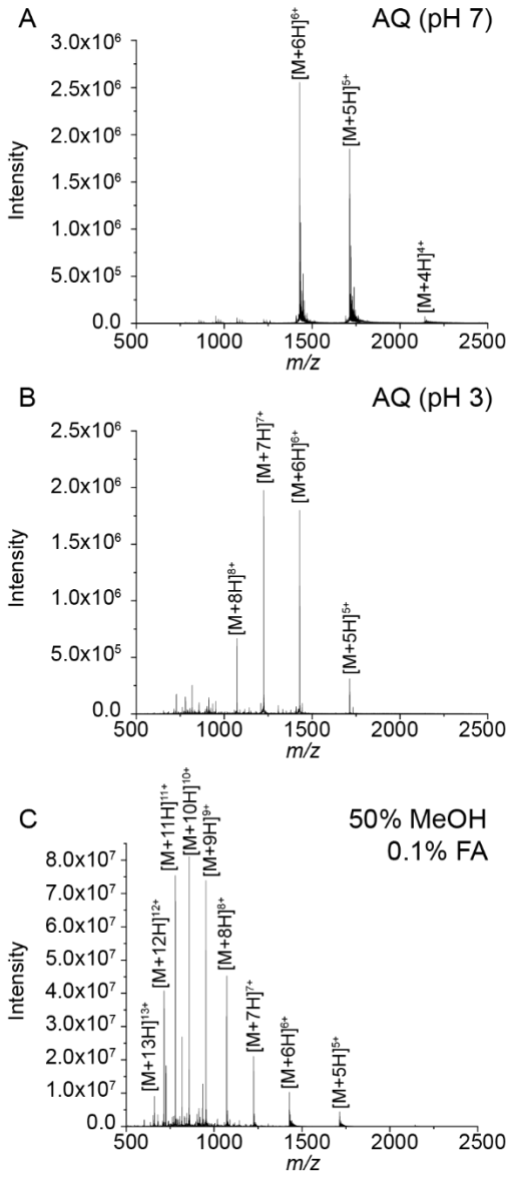


Fig. S1. Charge state distributions for aqueous (ammonium acetate pH 7 or acetic acid pH 3) compared to denaturing (50:50 methanol:water, 0.1 % formic acid) solution conditions.

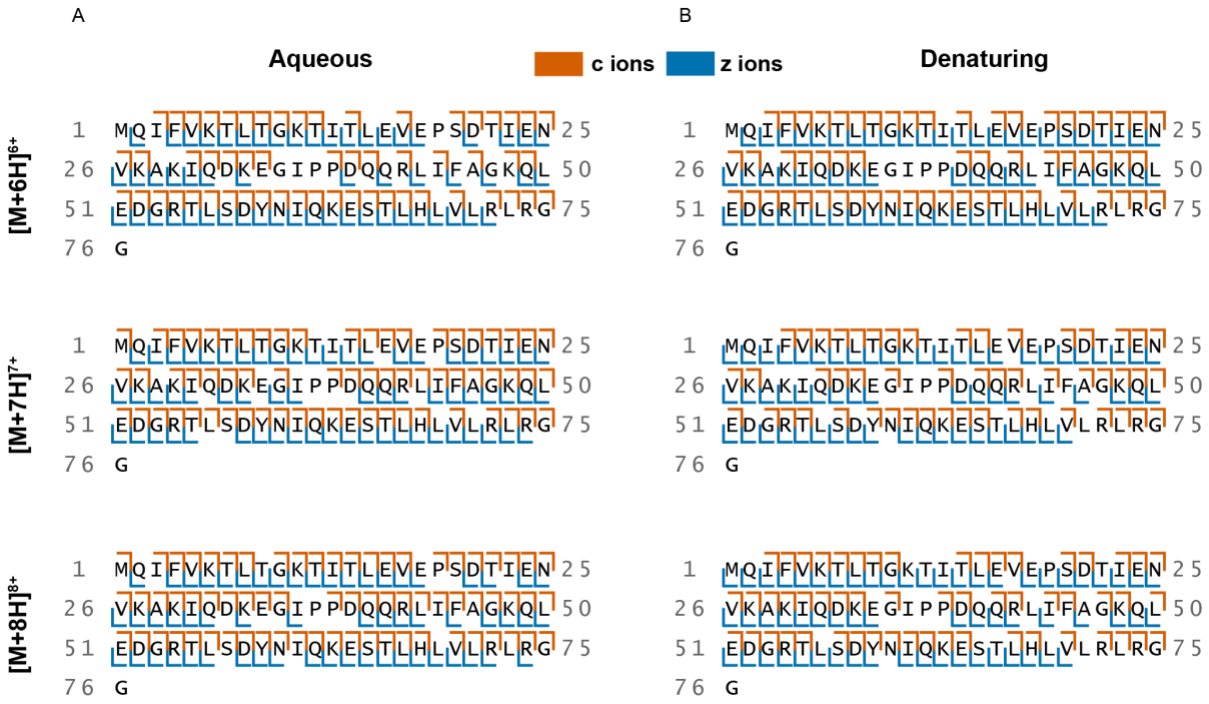


Fig. S2. ECD fragment ladder for ubiquitin (z = 6^+ to 8^+) in (A) aqueous, native-like state and (B) denaturing, more extended conformation.

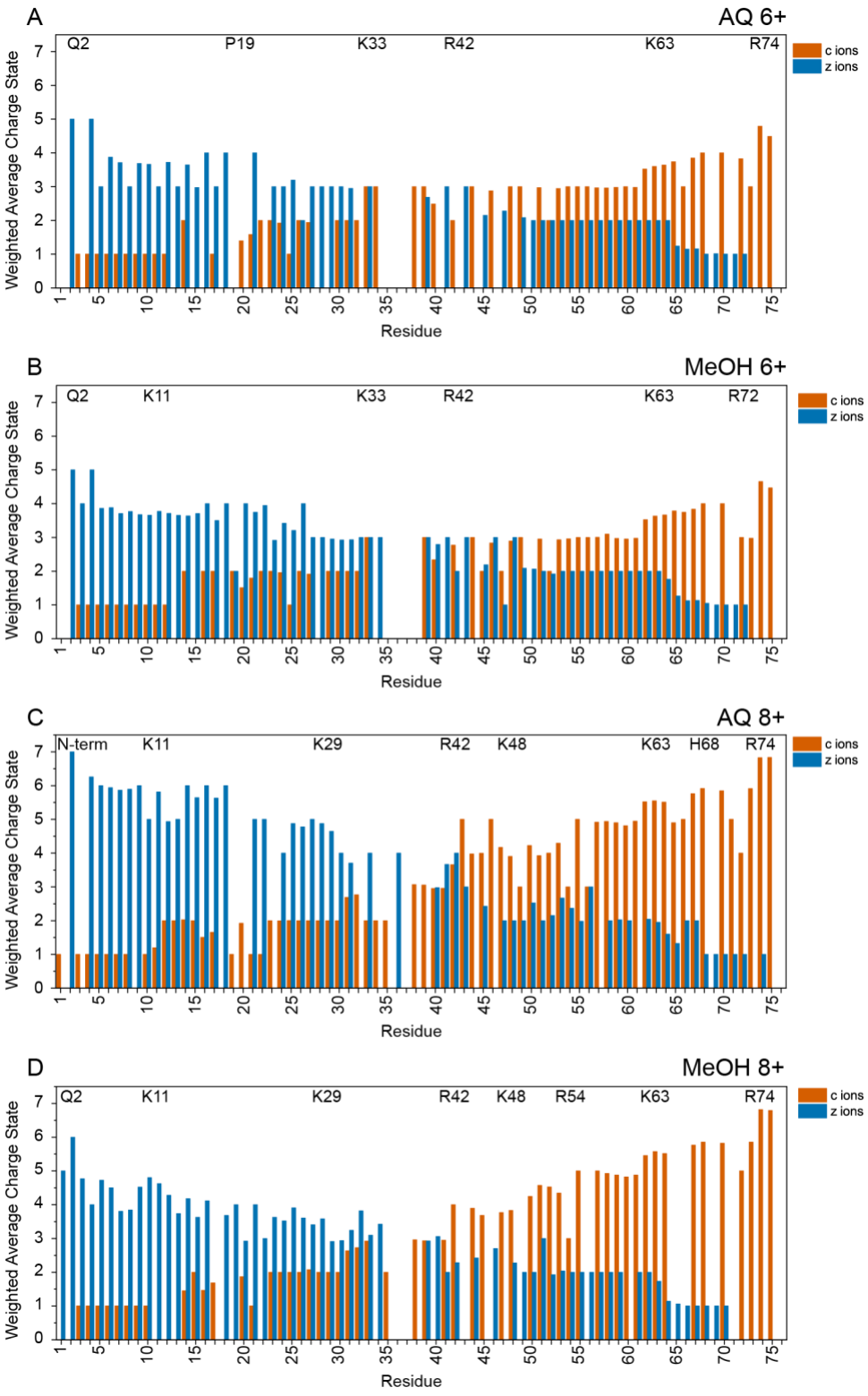


Fig. S3. Remaining weighted average charge state plots for 6+ and 8+ under aqueous and denatured conditions.

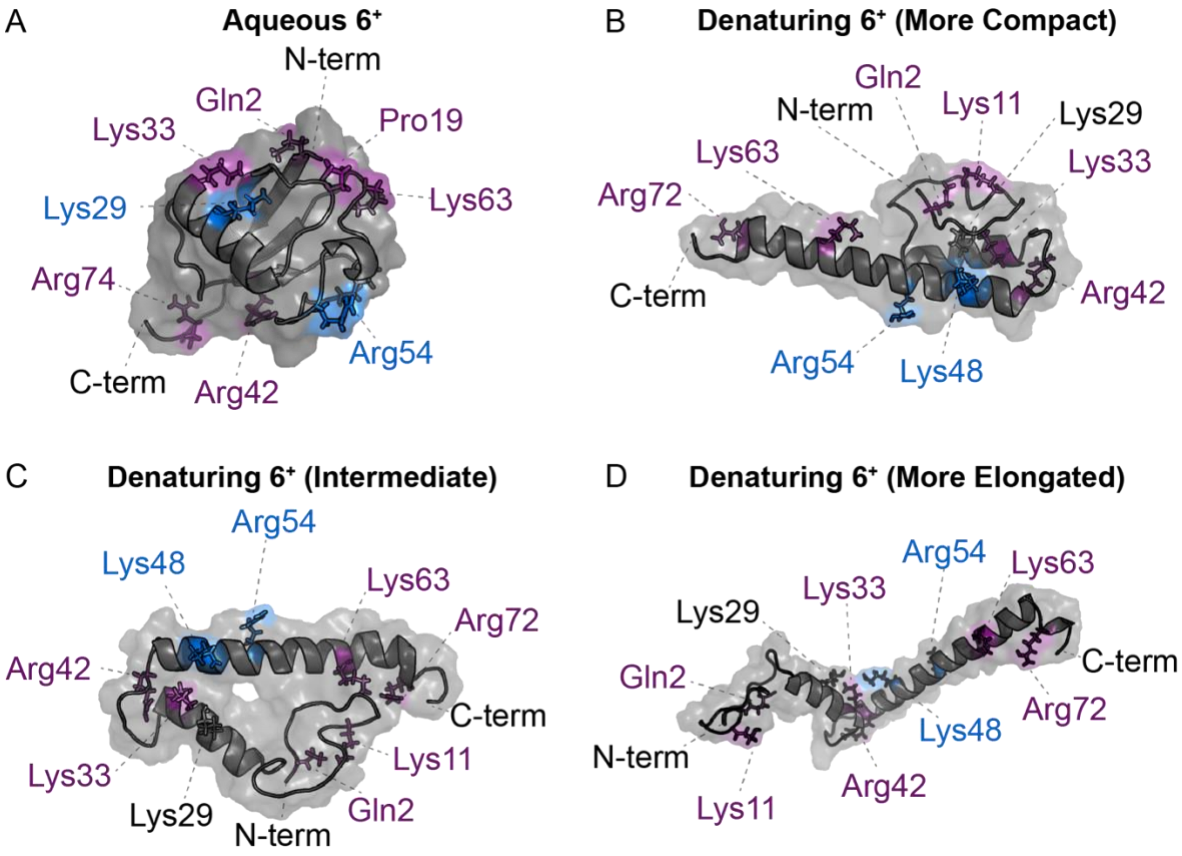
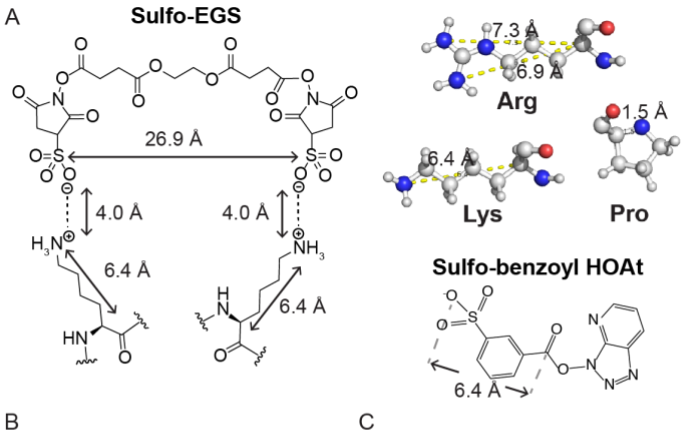


Fig. S4. Gas-phase structures of ubiquitin 6^+ modeled by molecular dynamics. (A) Aqueous ubiquitin in its N state. (B) More compact, (C) Intermediate and (D) more elongated ubiquitin conformer under denaturing condition after gas-phase modeling of the A state. Protonated sites are labeled in purple and sites modified by sulfo-benzoyl HOAt are in blue. Lys29 is labeled in gray since it remains uncharged and unmodified for ubiquitin under denaturing condition.



B

Max Estimated Distance (Å)	
Pro-Sulfo-EGS-Lys	42.8
Pro-Sulfo-EGS-Arg	43.7
Lys-Sulfo-EGS-Lys	47.7
Arg-Sulfo-EGS-Lys	48.6
Arg-Sulfo-EGS-Arg	49.5

C

Max Estimated Distance (Å)	
Pro-S-Benz-HOAt-Lys	18.3
Pro-S-Benz-HOAt-Arg	19.2
Lys-S-Benz-HOAt-Lys	23.2
Arg-S-Benz-HOAt-Lys	24.1
Arg-S-Benz-HOAt-Arg	25.0

Fig. S5. Calculation of distances spanned by reacting residue pairs. (A) Maximum distance estimates calculated based on side chain lengths generated in PyMOL (Arg, Lys and Pro: 7.3 Å, 6.4 Å and 1.5 Å), electrostatic distance estimated from canonical salt bridge (4 Å) and length of reagent (26.9 Å or 6.4 Å). Summary of the maximum distances measured between pairs of residues linked by (B) sulfo-EGS and (C) sulfo-benzoyl HOAt.

Table 1. Instrumental ExD Settings.

Setting	ECD ON	ECD OFF
Lens 1	- 27.9	- 27.9
Lens 2	- 25.3	- 25.3
Lens 3	- 5.9	- 5.9
Lens 4	6.0	6.0
Filament Bias	2.5	2.5
Lens 5	15.0	6.0
Lens 6	9.5	9.5
Lens 7	12.4	12.4
Filament Current (A)	2.35	0.1

Table 2. Modified Sites by Sulfo-EGS and Sulfo-benzoyl HOAt and Their Estimated Measured C_α - C_α Distances to Protonated Residues.

A **Sulfo-EGS**

Aqueous 6^+

Cα - Cα Distance (Å)	Q2	P19	K33	R42	K63	R74
P19	7.6	X	17.1	**19.2**	7.0	29.9
R42	19.1	**19.2**	16.8	X	19.8	12.0

Denaturing 6^+ (More Compact)

Cα - Cα Distance (Å)	Q2	K11	K33	R42	K63	R72
K33	13.8	10.2	X	6.8	28.4	**41.2**
R72	29.6	40.1	**41.2**	44.6	13.8	X

Denaturing 6^+ (Intermediate)

Cα - Cα Distance (Å)	Q2	K11	K33	R42	K63	R72
K33	21.7	27.5	X	5.2	26.8	**36.1**
R72	21.1	11.2	**36.1**	39.9	10.2	X

Denaturing 6^+ (More Elongated)

Cα - Cα Distance (Å)	Q2	K11	K33	R42	K63	R72
K33	19.2	26.9	X	5.0	31.7	**42.2**
R72	56.7	62.3	**42.2**	41.4	11.4	X

B **Sulfo-benzoyl HOAt**

Aqueous 6^+

Cα - Cα Distance (Å)	Q2	P19	K33	R42	K63	R74
K29	12.6	**13.6**	6.0	13.7	18.1	20.0
R54	19.0	13.8	21.5	13.4	**16.4**	22.4

Denaturing 6^+ (More Compact)

Cα - Cα Distance (Å)	Q2	K11	K33	R42	K63	R72
K48	14.1	**16.6**	10.2	9.8	22.3	35.5
R54	13.3	21.1	17.1	18.8	**14.0**	26.3

Denaturing 6^+ (Intermediate)

Cα - Cα Distance (Å)	Q2	K11	K33	R42	K63	R72
K48	23.3	**25.9**	9.5	9.7	22.2	32.3
R54	18.6	18.6	14.6	17.1	**13.7**	23.8

Denaturing 6^+ (More Elongated)

Cα - Cα Distance (Å)	Q2	K11	K33	R42	K63	R72
K48	27.5	**35.5**	10.4	9.9	22.2	32.9
R54	36.8	44.1	19.7	18.3	**13.9**	23.7