



Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution**

Yueyang Zhong, Linjie Han, and Brandon T. Ruotolo*

Abstract: The three-dimensional structures adopted by proteins are predicated by their many biological functions. Mass spectrometry has played a rapidly expanding role in protein structure discovery, enabling the generation of models for both proteins and their higher-order assemblies. While important coarse-grained insights have been generated, relatively few examples exist where mass spectrometry has been successfully applied to the characterization of protein tertiary structure. Here, we demonstrate that gas-phase unfolding can be used to determine the number of autonomously folded domains within monomeric proteins. Our ion mobility-mass spectrometry data highlight a strong, positive correlation between the number of protein unfolding transitions observed in the gas phase and the number of known domains within a group of sixteen proteins ranging from 8–78 kDa. This correlation and its potential uses for structural biology is discussed.

Rapidly characterizing the three-dimensional structures of proteins, and their higher-order complexes, is an unquestionably important goal in the post-genomic era.^[1] While X-ray and NMR analyses provide invaluable high-resolution structure information, ion-mobility mass spectrometry (IM-MS), which separates gas-phase ions according to both their size-to-charge and mass-to-charge ratios, can be engaged to provide structural constraints in a manner orthogonal to these more-established techniques.^[2] Many examples exist where IM-MS datasets have provided structural details for protein targets where few existed previously by offering significant advantages in terms of speed, sample consumption, and the ability to analyze complex mixtures.^[3] Despite successes in defining the overall quaternary structure of multiprotein machines, IM-MS technologies have fewer approaches available to determine the more-local elements of protein conformation within these assemblies. IM measurements have been used in conjunction with molecular

dynamics for well over a decade to generate high-fidelity structures for peptides and small proteins,^[4] but the size information contained within the IM measurement loses its ability to adequately filter models for larger protein systems, and thus their structures within larger assemblies. Therefore, there is a clear need to develop MS-based technologies capable of providing local structure information for larger, multi-domain proteins. Herein, we report the first use of gas-phase unfolding as a means of determining the number of autonomously folded domains within monomeric proteins.

Figure 1 illustrates our general experimental procedure, using a monomeric protein with two distinct domains (Figure 1a) as an example. Protein ions created using nano-electrospray ionization (nESI) are sampled by two separate analysis streams. In the first, protein ions created under native conditions (200 mM NH₄OAc) are selected for collisional activation (Figure 1b). Energy-resolved IM-MS data, acquired over a range of bias voltage values between the exit of the quadrupole mass filter and the ion trap region prior to IM separation, are used to track the unfolding of individual charge states of the protein ion population (Figure 1c). IM drift time data is converted to collision cross-section (CCS) to directly correlate the unfolding data to protein sizes and enable comparisons between different proteins and replicate runs on the same protein system (Figure 1d). These CCS data are then combined into a collision induced unfolding (CIU) fingerprint, where the energy required for protein ion unfolding can be readily identified for each conformer family detected (Figure 1e). Our detailed analysis for annotating CIU data can be found in Supporting Information, and involves the assessment of the total number of resolved features observed in excess of the initial population, and correlates this with the number of known domains within a given monomer.

The second experimental track involves proteins partially denatured in solution using a combination of organic and acidified solvents prior to nESI (Figure 1f). These data are characterized by the extensive charging of the protein ions produced, in a manner correlated with their increased surface area upon unfolding, and the drift time profiles for such ions are recorded for all of the charge states generated (Figure 1g). As observed previously, protein ions adopt a broad range of unfolded or partially unfolded structures upon charging (Supporting Information, Figure S1), ultimately achieving highly charged string-like conformations.^[4] Our analysis tracks all of this data in both IM drift time and CCS space (Figure 1h), and uses a simple derivative analysis to identify regions within IM data that exhibit the greatest rate of change in CCS (δ CCS, Figure 1i), which typically overlaps with charge states where the protein ions exist simultaneously

[*] Y. Zhong, L. Han, B. T. Ruotolo
Department of Chemistry, University of Michigan
930 North University Avenue, Ann Arbor, MI 48109-1055 (USA)
E-mail: bruotolo@umich.edu

[**] We thank Jessica Rabuck and Matthew Soellner (University of Michigan, Chemistry) for providing the Abelson kinase and for helpful discussions, and Ireena Bagai and Stephen Ragsdale (University of Michigan, Biological Chemistry) for providing BVR and CPR. This work is supported by the American Society for Mass Spectrometry (in the form of a Research Award), the Oak Ridge Associated Universities (in the form of a Ralph E. Powe Junior Faculty Enhancement Award), and the National Science Foundation (CAREER Award, 1253384).

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201403784>.

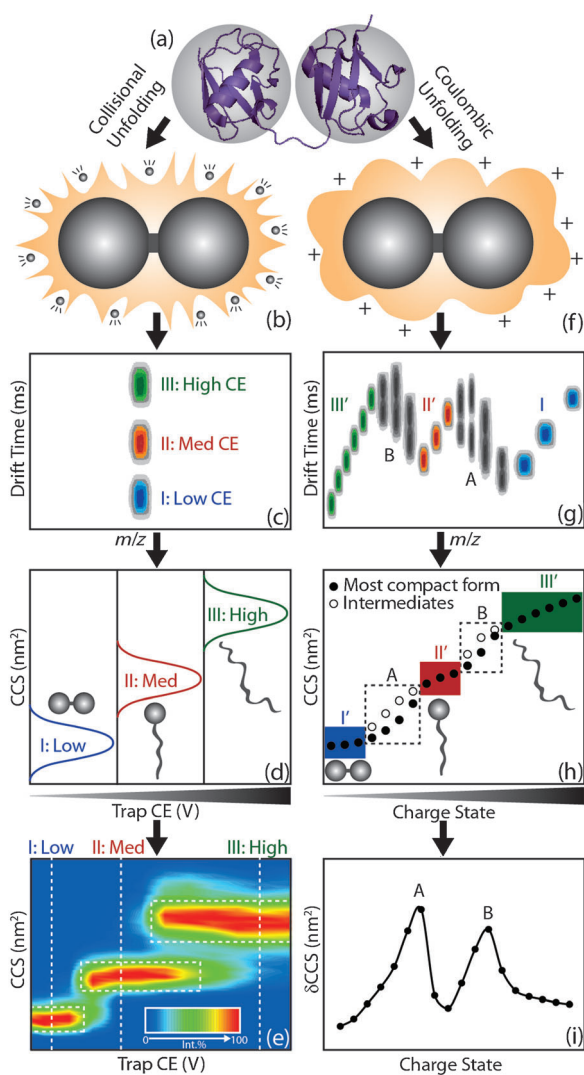


Figure 1. Diagram illustrating IM-MS data collection and analysis procedures. A native two-domain protein (a) is subjected to collisional unfolding in the gas phase (b) or Coulombic unfolding through denaturation in solution (f). c) An example overlay of collisional unfolding data collected over a range of collision voltages, resulting in three conformational families I (blue), II (red), and III (green) observed at low, medium, and high voltage values respectively. d) The drift time profile is converted into collision cross-section (CCS) data for each collision voltage. e) Combined data recorded under all collision voltages, where CCS is plotted against collision voltage. Colors denote signal intensity as indicated. g) Drift time profile of Coulombically unfolded protein ions. Along with the three general regions of protein unfolding shown I (blue), II' (red), and III' (green), two intermediate stages, marked A and B, are also observed. h) Calibrated data plotting CCS against protein charge state. i) A first derivative plot where δCCS is plotted as a function of charge state, serving to highlight stages A and B.

in multiple conformations. As above, this enables us to correlate regions of greatest δCCS directly to the number of expected protein domains.

We first examined the unfolding properties of ubiquitin, and its N-to-C-terminus, covalently linked multi-domain analogues, in the gas phase (Figure 2a–h). Single-domain ubiquitin is currently among the most-studied gas-phase

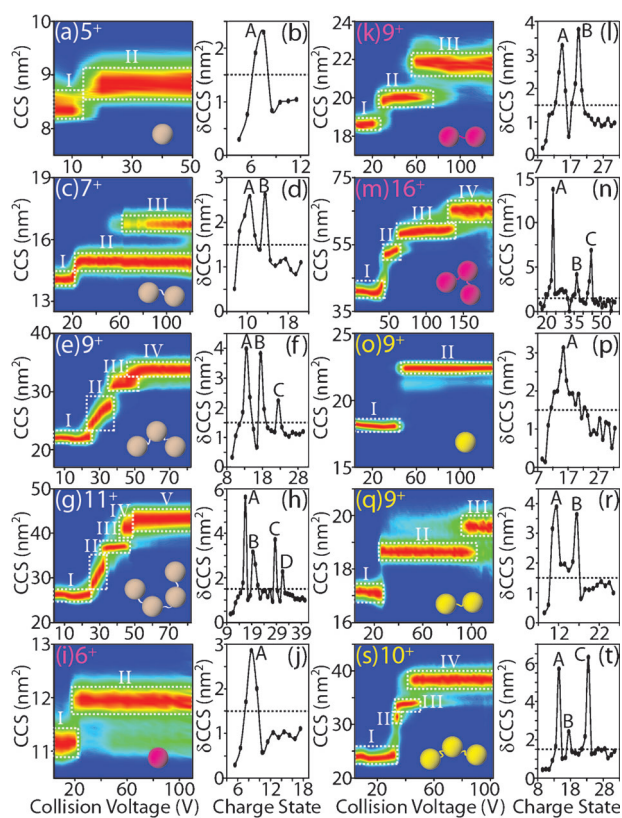


Figure 2. Collisional (a, c, e, g, i, k, m, o, q, s) and Coulombic (b, d, f, h, j, l, n, p, r, t) unfolding data for a range of proteins. Charge states used in collisional unfolding experiments are noted on the top left corner of each CIU fingerprint. Conformational families detected are noted with numerals I through V. Intermediate unfolding stages identified within Coulombic unfolding experiments are labelled A to D. The dashed line represents an empirically derived signal cut-off value at $\delta\text{CCS} = 1.5 \text{ nm}^2$. (a,b) ubiquitin, (c,d) di-ubiquitin (N-C linked), (e,f) tri-ubiquitin (N-C linked), (g,h) tetra-ubiquitin (N-C linked), (i,j) cytochrome C, (k,l) glutathione S-transferase, (m,n) serum albumin, (o,p) green fluorescence protein, (q,r) γ -D crystalline, and (s,t) fibronectin are clustered into ubiquitin proteins (with each domain bearing a mixture of α -helix and β -sheet secondary structure), primarily α -helix proteins and primarily β -sheet proteins, indicated by white, pink, and yellow, respectively. The X-ray structures and PDB codes for these proteins are provided in Supporting Information, Figure S2.

proteins.^[5] Recent work has demonstrated a clear correlation between the folded state of the biopolymer in solution and its resultant gas-phase conformer populations.^[5] The native ubiquitin fold is a mixture of α -helix and β -sheet secondary structure, making it an ideal initial target for our approach. A single ubiquitin domain exhibits one unfolding transition when low charge state protein ions are selected for CIU, and one δCCS feature in Coulombic unfolding data, both indicative of a single domain structure. As additional ubiquitin domains are added to the C-terminus of the original protein, further CIU and Coulombic unfolding signals are observed, each in precise agreement with the expected response for the known domain structure of the proteins analyzed. When our dataset is expanded to include primarily α -helical (Figure 2i–n) and β -sheet (Figure 2o–t) proteins (with their crystal structures and PDB codes shown in the

Supporting Information, Figure S2), these general correlations are retained.

Our poly-ubiquitin data reveals general trends in the domain-related unfolding of gas-phase proteins. For example, while mono- (5^+) and di- (7^+) ubiquitin CIU fingerprints contain clear, step-wise transitions between the different unfolded forms observed, tri- (9^+) and tetra- (11^+) ubiquitin CIU data reveal more-complex transitions between intermediate states. Specifically, conformational family II in both the tri- (Figure 2e) and tetra-ubiquitin (Figure 2g) datasets appears broadened on the CCS axis, and possess a diagonalized shape in CCS/voltage space. This is also observed in serum albumin (Figure 2m) and, to a lesser extent, in fibronectin (Figure 2s) datasets. It is likely that these broadened CIU features relate to significant inter-domain contacts in the gas-phase structures present. Despite these complexities, however, the number of resolved features observed is clearly correlated with the number of autonomously folded domains expected within the protein.

A careful analysis of our Coulombic unfolding data also highlights general trends. For example, mono-domain ubiquitin exhibits maximum δCCS values at about 7 charges, which compares well to previously reported IM-MS data for this protein.^[5] In comparison, di-ubiquitin exhibits maximum δCCS at about 10 charges, and this value increases to about 13 and about 15 charges for tri- and tetra-ubiquitin, respectively. This increase correlates closely with the trends in average CCS values for the intact proteins in low charge states, as well as CCS estimates based on fixed-density and protein molecular mass (linear $R^2 > 0.99$),^[6] suggesting that initial domain unfolding is strongly correlated with protein ion surface charge. This general trend continues throughout our dataset. Furthermore, while di- and tri- ubiquitin require relatively equal amounts of charge to unfold subsequent domains (4 or 5 charges on average), the Coulombic unfolding of tetra-ubiquitin appears less-evenly spaced as a function of charge, requiring between 3 and 9 charges to initiate unfolding of subsequent domains. As above, this more-complex relationship between ion charge state and unfolding is likely related to the larger number of inter-domain contacts in the gas-phase protein ion structure, and is also observed in our serum albumin (Figure 2n) and fibronectin (Figure 2t) datasets. We have assigned an empirical noise value of 1.5 nm^2 to all such datasets and, when signals are detected above this threshold, a clear correlation between unfolding data and known domain structure is observed.

One key element to successfully mapping CIU data onto protein domain structure involves the selection of protein charge state for detailed analysis. Higher protein charge states typically generate a large number of CIU transitions that may not correlate with protein domain structure. As shown on Figure 3, all of the charge states identified in this study that produce CIU data in optimal agreement with domain-resolved unfolding patterns follow a power-law correlation as a function of protein mass (Figure 3a) and solvent accessible surface area (Figure 3b). The trends in our data are mirrored by previously observed functions that are used to predict the average charging of protein ions from native solution by ESI,^[7] but are offset to charge values that are 15% and 33%

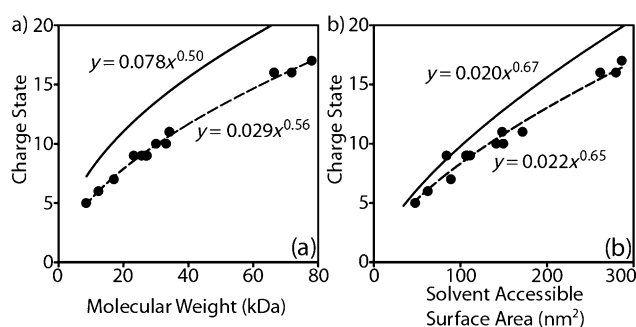


Figure 3. a) Protein charge states used in CIU experiments plotted against protein mass for thirteen of the systems studied here (●), all possessing uniform SCOP and CATH domain classifications. R^2 for the fit shown is 0.99. The solid line indicates expected average charge state from previous experiments that use fixed protein density to estimate surface charging for protein ions produced from ESI under native conditions. b) Protein charge states used in CIU experiments plotted against solvent accessible surface area for the same systems analyzed in Figure 3 a (●), as determined from available X-ray structures. R^2 for the fit shown is 0.96. The solid line indicates expected average charge states for these protein ion surface areas, generated from ESI under native conditions, derived from previous experimental data.

lower when the solvent accessible surface areas and the average masses of the proteins are considered respectively.

While clearly held within our entire dataset, we have made observations that illustrate minor caveats to the above-noted correlations. For example, green fluorescent protein CIU data shown in Figure 2o contains a small signal at 21 nm^2 along with the main unfolding transition at 22.5 nm^2 used in our domain analysis. Furthermore, the Coulombic unfolding data collected for this protein (Figure 2p) is broadened relative to others collected. We note, however, that the extra CIU signal is less than 10% of the total and owing to its low intensity can be easily ignored. Unsurprisingly, CIU data for proteins where the domains are not linked through their respective N and C termini are often more complex than those shown in Figure 2 (for example, K48 and K63 linked di-ubiquitin; Supporting Information, Figure S3), but as non-linear protein sequences comprise less than 3% of the proteome,^[8] and are easily detected by MS, spurious CIU domain assignments for such non-linear sequences are unlikely. Our data for cytochrome P450 reductase (CPR), a three-domain protein, produces CIU data in excellent agreement with known domain number, but produces Coulombic unfolding data that is more difficult to interpret, likely due to an inability to identify optimized solution conditions for unfolding (Supporting Information, Figure S4). Likewise, in the case of the Abelson protein tyrosine kinase and transferrin, the SCOP^[9] and CATH^[10] databases provide different assessments of the total number of autonomous domains contained within the proteins. In the former case, CIU and Coulombic unfolding identify a two domain structure in agreement with CATH. In the latter case, Coulombic data is in agreement with the SCOP classification, and the CIU with CATH (Supporting Information, Figure S5). Finally, in the case of billverdin reductase, BVR (Supporting Information, Figure S6), Coulombic unfolding

data is in strong agreement with the CATH domain analysis, but CIU data suggests a third domain, which is most likely due to the highly intermingled nature of the domain structure within the protein.^[11] As above, such domain architectures are expected to account for an exceedingly small portion of the proteome (less than 0.3%), thus minimally impacting the domain analysis of unknowns.^[12]

In conclusion, we have used a group of sixteen proteins to identify and validate a fundamental correlation between gas-phase protein unfolding and their domain structures in solution. While measurements of gas-phase protein unfolding are now over two decades old,^[13] and have revealed that single domain proteins and protein complex ions can adopt both compact and unfolded conformations in the absence of bulk solvent as a function of both charge state and collisional activation specific to the gas phase,^[14] neither the ability to extract physical descriptors of protein structure nor general trends within these data have been reported previously. The unfolding observed here demonstrates discrete unfolding stages for each domain, similar to atomic force microscopy (AFM) experiments where proteins tethered to a surface are mechanically unfolded.^[15] Unlike traditional solution-phase unfolding experiments, in which homologous domains often unfold cooperatively, gas-phase unfolding is likely able to make use of subtle structural differences in equivalent domains to initiate unfolding in only one region of the structure. Once triggered, asymmetric protein unfolding is undoubtedly driven by a mechanism similar to that previously described for multi-chain non-covalent protein complexes.^[16] The trends we identify are shown to be independent of protein size (8–78 kDa) and fold type. By selecting charge states that are significantly lower than the average produced by ESI from native solutions, CIU data can be generated that is highly protein domain-correlated. While having clear applications in multiprotein topology mapping,^[17] we are currently pursuing CIU and Coulombic unfolding analysis approaches aimed at the rapid assessment of intact antibodies and other biotherapeutics.

Experimental Section

A full description of the methods, detailed materials, along with supporting IM-MS experiments are given in the Supporting Information. Briefly, IM-MS data were collected using a quadrupole-ion mobility-time-of-flight mass spectrometry (Q-IM-ToF MS) instrument (Synapt G2 HDMS, Waters, Milford MA, USA) as described previously.^[18] All proteins were acquired from sources and methods described in detail in Supporting Information. Proteins were prepared separately for CIU and Coulombic unfolding experiments as described in Supporting Information, using a final protein concentration of 5–10 μM . CCS values were calibrated in a manner described previously.^[19] Solvent accessible surface areas for all proteins were

calculated with GETAREA^[20] using the appropriate Protein Data Bank entries (codes are listed in Supporting Information).

Received: March 27, 2014

Published online: July 2, 2014

Keywords: gas-phase structures · mass spectrometry · protein folding · protein structures

- [1] S. Kuhner, V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castano-Diez, W. H. Chen, D. Devos, M. Guell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Bottcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork, A. C. Gavin, *Science* **2009**, *326*, 1235.
- [2] a) A. J. R. Heck, *Nat. Methods* **2008**, *5*, 927; b) E. V. Petrotchenko, C. H. Borchers, *Mass Spectrom. Rev.* **2010**, *29*, 862; c) L. Konermann, J. Pan, Y.-H. Liu, *Chem. Soc. Rev.* **2011**, *40*, 1224; d) J. L. P. Benesch, B. T. Ruotolo, *Curr. Opin. Struct. Biol.* **2011**, *21*, 641–649.
- [3] Y. Zhong, S.-J. Hyung, B. T. Ruotolo, *Expert Rev. Proteomics* **2012**, *9*, 47.
- [4] a) D. E. Clemmer, M. F. Jarrold, *J. Mass Spectrom.* **1997**, *32*, 577; b) T. Wyttenbach, M. T. Bowers, *Modern Mass Spectrometry*, Springer, Berlin, **2003**, pp. 207–232.
- [5] a) T. Wyttenbach, M. T. Bowers, *J. Phys. Chem. B* **2011**, *115*, 12266; b) H. L. Shi, N. A. Pierson, S. J. Valentine, D. E. Clemmer, *J. Phys. Chem. B* **2012**, *116*, 3344.
- [6] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, C. V. Robinson, *Nat. Protoc.* **2008**, *3*, 1139.
- [7] a) J. Fernandez de la Mora, *Anal. Chim. Acta* **2000**, *406*, 93; b) I. Kaltashov, A. Mohimen, *Anal. Chem.* **2005**, *77*, 5370.
- [8] V. Jai Kartik, T. Lavanya, K. Guruprasad, *Int. J. Biol. Macromol.* **2006**, *38*, 174.
- [9] A. Murzin, S. Brenner, T. Hubbard, C. J. Chothia, *J. Mol. Biol.* **1995**, *247*, 536.
- [10] C. Orengo, A. Michie, S. Jones, D. Jones, M. Swindells, J. Thornton, *Structure* **1997**, *5*, 1093.
- [11] F. Rousseau, J. Schymkowitz, L. Itzhaki, *Structure* **2003**, *11*, 243.
- [12] Y. Liu, D. Eisenberg, *Protein Sci.* **2002**, *11*, 1285.
- [13] D. Suckau, Y. Shi, S. C. Beu, M. W. Senko, J. P. Quinn, F. M. Wampler, F. W. McLafferty, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 790.
- [14] M. Schennach, K. Breuker, *Angew. Chem.* **2014**, *126*, 168; *Angew. Chem. Int. Ed.* **2014**, *53*, 164.
- [15] M. Carrion-Vazquez, A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, J. M. Fernandez, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3694.
- [16] J. L. P. Benesch, *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 341.
- [17] S.-J. Hyung, B. T. Ruotolo, *Proteomics* **2012**, *12*, 1547.
- [18] a) K. Giles, J. P. Williams, I. Campuzano, *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1559–1566; b) Y. Zhong, S.-J. Hyung, B. T. Ruotolo, *Analyst* **2011**, *136*, 3534–3541.
- [19] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo, *Anal. Chem.* **2010**, *82*, 9557–9565.
- [20] R. Fraczkiewicz, W. Braun, *J. Comput. Chem.* **1998**, *19*, 319.