Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information



69451 Weinheim, Germany

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

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

ange_201403784_sm_miscellaneous_information.pdf

Supporting Information

**Sample Preparation.** Proteins were acquired from the following sources: ubiquitin (bovine, Sigma U6253), linear di-ubiquitin (human, BostonBiochem UC-700-100), linear tri-ubiquitin (human, Enzo life sciences BML-UW0780-0100), linear tetra-ubiquitin (human, Enzo life sciences BML-UW0785-0100), cytochrome C (horse, Sigma C2506), glutathione S-transferase (schistosoma japonicum, provided by the Ragsdale lab at the University of Michigan, purified and prepared using a previously-published procedure[1]), serum albumin (bovine, Sigma A7906), green fluorescent protein (GFP, Aequorea Victoria, Prospec PRO-687), γ-D crystallin (human, Creative Biomart CRYGD-3634H) and fibronectin III 8-10 N-GST (human, Kerafast EUR116, the GST tag was removed by treatment with a Thrombin CleanCleave Kit from Sigma and purified using a glutathione sepharose 4B column from GE Healthcare). Standards for collision cross-section (CCS) calibration include: a commercially-available peptide mixture (Waters 186002337), cytochrome C, and the protein complexes avidin (Sigma A9275), concanavalin A (Sigma C2010), alcohol dehydrogenase (Sigma A7011), and glutamate dehydrogenase (Sigma, G7882). The peptide mixture (~2 μM for each peptide) was prepared in a water/methanol/ acetic acid (49/49/2) solution. All other proteins were solubilized in 200 mM aqueous ammonium acetate ($NH_4Ac$) solutions, or kept in their original solution conditions, and flash frozen for storage at -80 °C. In general, stock samples (≥30 μM), were buffer-exchanged into a 200mM $NH_4Ac$ solution (described above) prior to IM-MS analysis, using a Micro Bio-Spin 6/30 column (Bio-Rad, Hercules, CA). For CIU experiments, all monomeric proteins were diluted into 200 mM $NH_4Ac$ to a final concentration of 5-10 μM before analysis, except glutathione S-transferase, which required treatment with 40% dimethyl sulfoxide and 1 M $NH_4Ac$ to convert the native dimer into its folded monomeric form. For Coulombic unfolding, proteins were denatured in solution using a combination of methanol, acetic acid and, in some cases, dimethyl sulfoxide, optimized for

each protein to achieve a maximum range of charge states. Specific denaturation conditions used for each protein are listed below in Figure S1.

**IM-MS Instrument and Methods.** IM-MS data were acquired on a Synapt G2 (Waters, Milford, MA) optimized for the transmission of intact protein and protein complexes as described previously,[2] using nano-electrospray ionization (nESI). For CIU experiments, individual protein charge states were first selected by tandem MS using a quadrupole mass filter. The selected ions were then activated in the ion trap region of the instrument using energetic collisions with Ar ($3\times10^{-2}$ mbar) prior to IM separation. IM separates proteins according to their orientationally-averaged collision cross section (CCS) by driving ions through a gas-filled chamber with an applied electric field.[2-3] Ramping the bias voltage value that dictates the static voltage offset between the exit of the quadrupole filter and the inlet of the ion trap (termed 'trap CE' in the instrument control software) allowed for the collection of CIU related unfolding in an energy-resolved manner, revealing gas-phase protein conformations stable on the millisecond timescale, as observed previously.[4] In each CIU experiment, collision voltage was ramped in 2V increments until protein backbone fragments were observed. Coulombic unfolding datasets are presented without m/z filtration or any gas phase activation. IM resolution was optimized for all datasets collected by operating the T-wave separator under conditions of high $N_2$ pressure (3-4 mbar), high wave height (40 V) and medium wave velocity (400-900 m/s).[5]

**Data Analysis.** Initial IM-MS Data processing was performed using Masslynx v4.1 and Driftscope v2.0 (both from Waters, Milford, MA). IM calibration to a He CCS axis was performed based on a previously-described method, producing highly-linear calibration plots ($R^2 > 0.96$). CIU contour plots charting protein CCS versus collision voltage were generated by Origin software package. The color scale in these plots indicates the signal intensity recorded for each conformer family at

the indicated collision energy values, normalized to the highest signal of each collision voltage. Solvent accessible surface areas for all proteins were calculated with GETAREA (www.curie.utmb.edu/getarea.html) using the appropriate Protein Data Bank entries (Table S1). We used the following rules to define different 'features' within CIU fingerprints throughout the data presented: 1) Generally, a stable conformational family survives over 5V energy increment without significant CCS variance. 2) In some cases where one single domain is not sufficiently stable over a range of collision voltages, a slight increase in CCS for an individualized conformational family might be observed. However, the rate of such CCS increase should be significantly lower than that between two neighboring conformational families which can be well-resolved, as indicated by distinct red bands on the CIU fingerprint. 3) Two conformational families may coexist under the same energy. During energy ramping, the precedent conformation is diminishing whereas the following one is developed into a dominant family. Protein domain structures in solution are mined from the SCOP (http://scop.mrc-lmb.cam.ac.uk/scop/) and CATH (http://www.cathdb.info/) databases for comparisons with IM-MS data (Table S1). We define a homologous domain is one that exhibits a likeness or similarity to another domain in terms of its structure.

**Table S1. Detailed information for proteins used in this study, including protein names, organisms, masses, PDB codes, secondary structure content (α helix and β sheet percentage) and domain numbers defined by SCOP (red) and CATH (green).** Domain numbers displayed in black indicate unambiguous domain assignments. Linear tri- and tetra- ubiquitin do not have PDB structures. Biliverdin reductase is comprised of 2 mixed domains, meaning each domain consists of different sections from the amino acid chain.

| Protein | Organism | Mass (kDa) | PDB code | Secondary Structure | Domain # (SCOP/CATH) |
|---|---|---|---|---|---|
| Ubiquitin | Bovine | 8.6 | 1UBI | 25% α+34% β | 1 |
| Linear Di-Ubiquitin | Human | 17.1 | 3AXC | 22% α+33% β | 2 |
| Linear Tri-Ubiquitin | Human | 25.7 | N/A | Similar to Ubiquitin | 3 |
| Linear Tetra-Ubiquitin | Human | 34.2 | N/A | Similar to Ubiquitin | 4 |
| Cytochrome C | Horse | 12.4 | 1HRC | 40% α+ 1% β | 1 |
| GST Monomer | Schistosoma japonicum | 27.3 | 1GTA | 50% α+ 7% β | 2 |
| Serum Albumin | Bovine | 66.4 | 4F5S | 73% α | 3 |
| Green Fluorescent Protein | Aequorea Victoria | 26.9 | 1GFL | 10% α+ 49% β | 1 |
| Gamma D Crystallin | Human | 23.2 | 2G98 | 9% α+ 42% β | 2 |
| Fibronectin III 8-10 | Human | 30.1 | 1FNF | 51% β | 3 |
| K48 linked Di-Ubiquitin | Human | 17.1 | 1AAR | 19% α+ 34% β | 2 |
| K63 linked Di-Ubiquitin | Human | 17.1 | 2JF5 | 23% α+ 32% β | 2 |
| NADPH-Cytochrome P450 Reductase | Human | 71.8 | 3QE2 | 36% α+ 20% β | 3 |
| Abelson Protein Tyrosine Kinase | Human | 33.2 | 2GQG | 39% α+ 17% β | 1/2 |
| Transferrin | Bovine | 78.0 | 1BLF | 31% α+ 17% β | 2/4 |
| Biliverdin Reductase | Human | 34.2 | 2H63 | 40% α+ 25% β | 2 (mixed) |

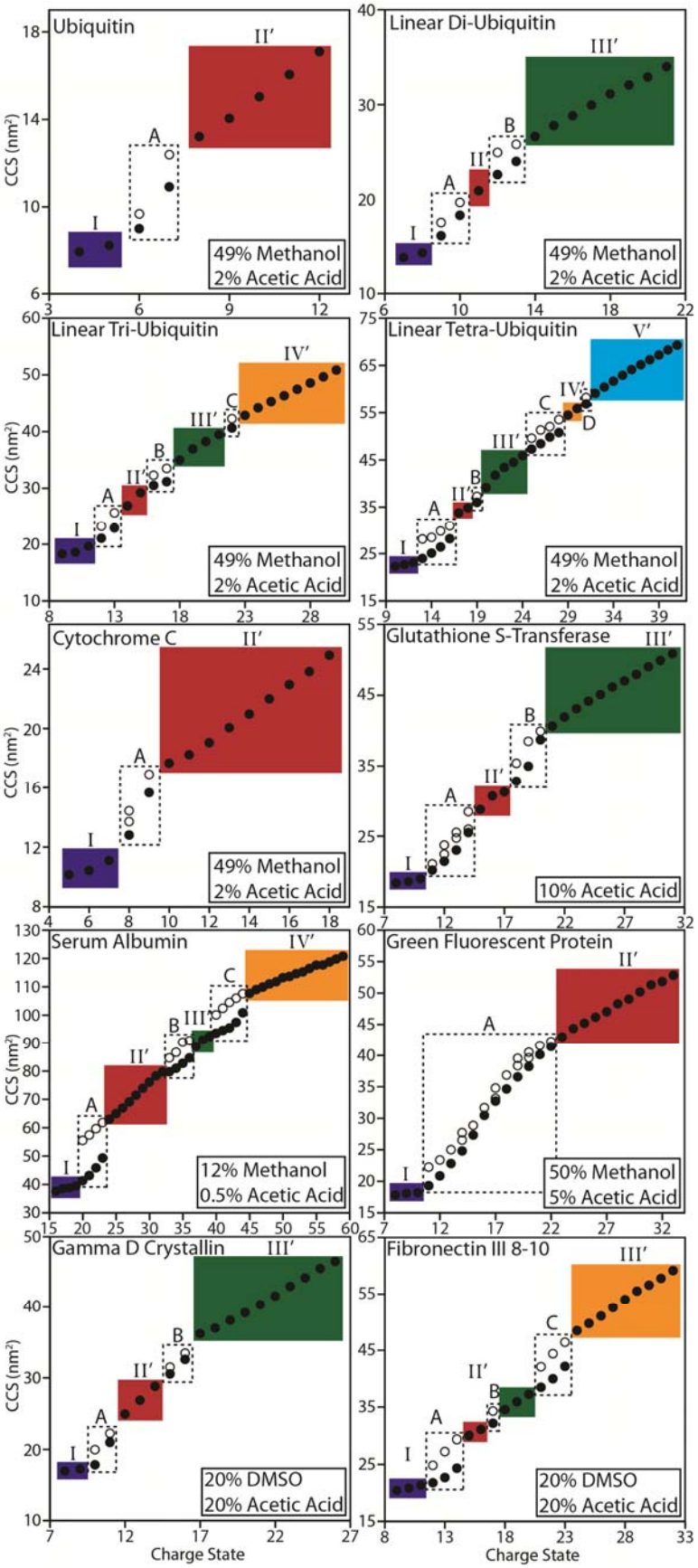


**Figure S1. Coulombic unfolding data for proteins discussed in the main text, including 1-4 ubiquitins, 3 primarily α-helix proteins and 3 primarily β-sheet proteins.** CCS is plotted against charge state adopted upon solution unfolding. The denaturing solution conditions used for each protein are listed in the lower right corner of each plot. Again, the number of known domains in each protein agrees with the number of high rate of change regions, labeled A through D, identified in our first derivative anslysis. Regions that refer to pseudo-stable conformational families as a function of charge are labeled I to V'. The most compact structure of each charge state is labeled using closed circles. The intermediate stages having large δCCS values are characterized by multiple unfolded structures that exist simultaneously. These intermediates, rather than the most compact form, are labeled as open circles.

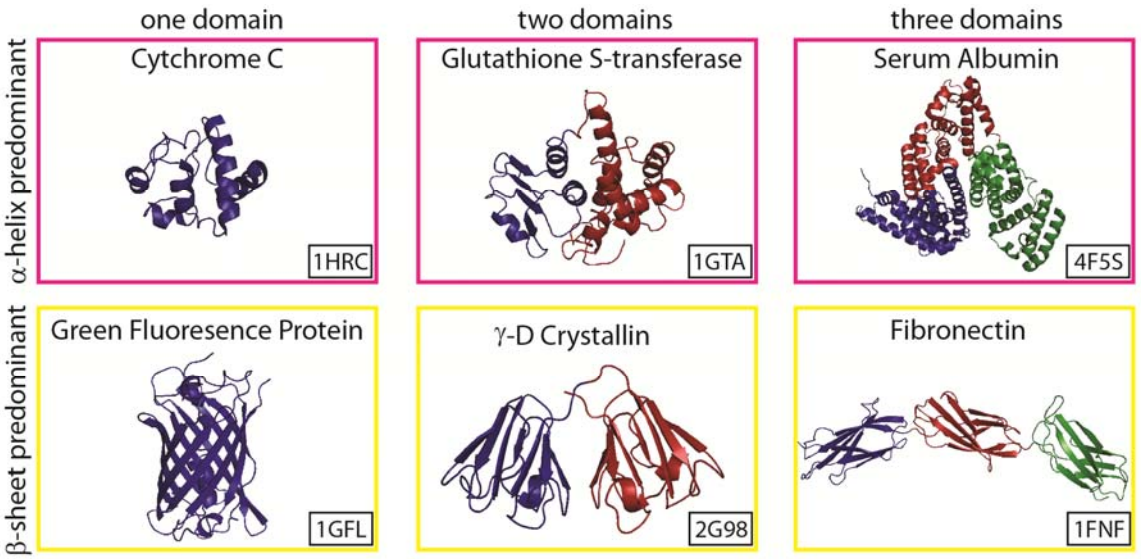


**Figure S2.** X-ray structures and PDB codes for primarily α-helix proteins (pink) and primarily β-sheet proteins (yellow), discussed in Figure 2.

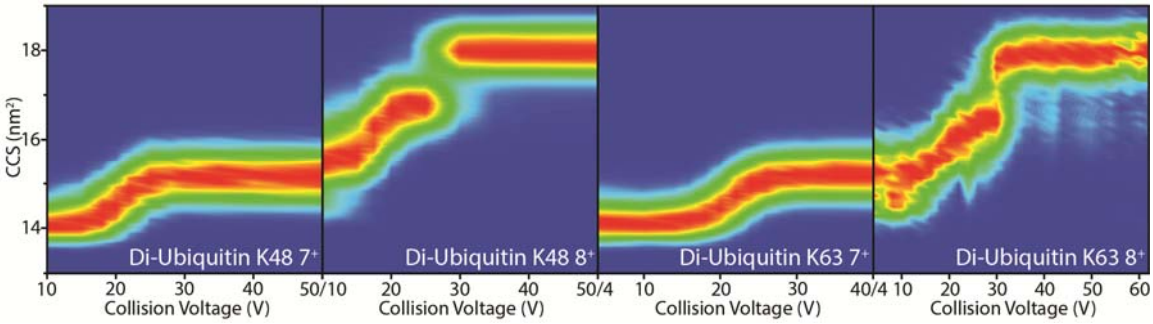


**Figure S3. Collisional unfolding data of non-linear di-ubiquitins, K48 and K63 linked at charge states $7^+$ and $8^+$.** No clear indication of a two-domain signature due to the sequence linkages that do not travel directly through the protein termini, although $8^+$ displays data with a signature more-closely associated with the expected domain structure of the protein than the $7^+$ in this case.

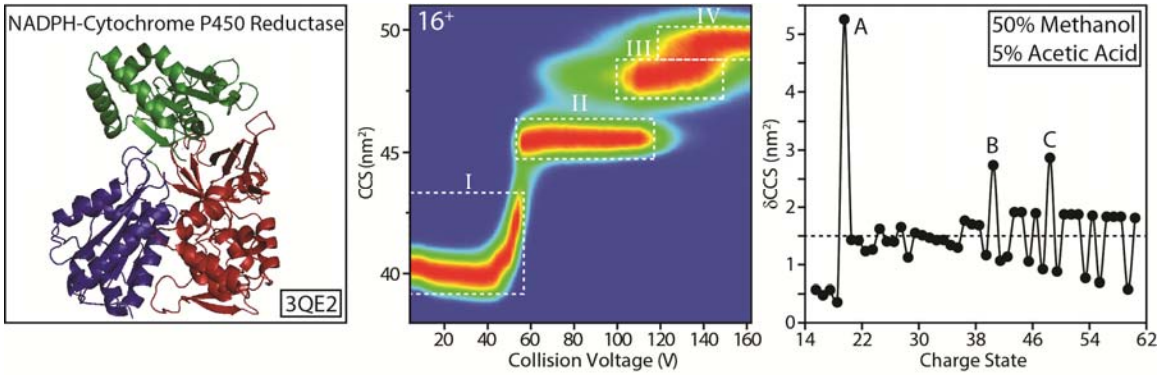


**Figure S4. Crystal structure (left), collisional unfolding (middle) and Coulombic unfolding (right) data for NADPH-Cytochrome P450 Reductase (CPR).** This three-domain protein exhibits a clear three domain CIU pattern. On the other hand, the Coulombic unfolding data is more difficult to interpret. Low intensity features for the final 2 domains, along with a higher degree of noise than expected, frustrate our analysis. It is likely that the solution condition has not been optimized for this protein.

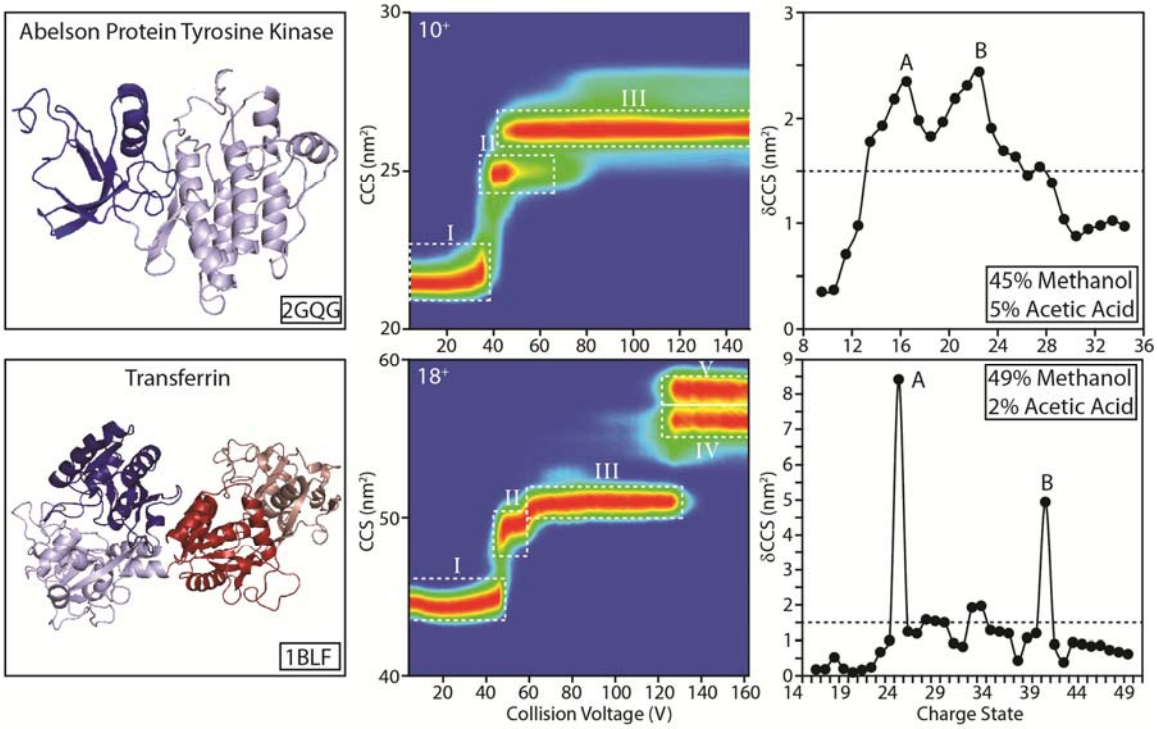


**Figure S5. Crystal structures (left), collisional unfolding (middle) and Coulombic unfolding data (right) for SCOP/CATH inconsistent proteins, Abelson protein tyrosine kinase and Transferrin.** Abelson protein tyrosine kinase is assigned as a two domain protein by CATH (dark blue and light blue), however is considered a one domain protein in SCOP. Both collisional and Coulombic unfolding data indicate a 2-domain protein. We note, however, that the δCCS peaks recorded for this protein during Coulombic unfolding are not as distinct as those found for proteins with unambiguous domain assignments. Transferrin, considered to have 2 distinct large domains with 2 subdomains in each, generates a collisional unfolding pattern that agrees well with CATH (4-domain protein), but a Coulombic unfolding pattern agrees with SCOP (2-domain protein).

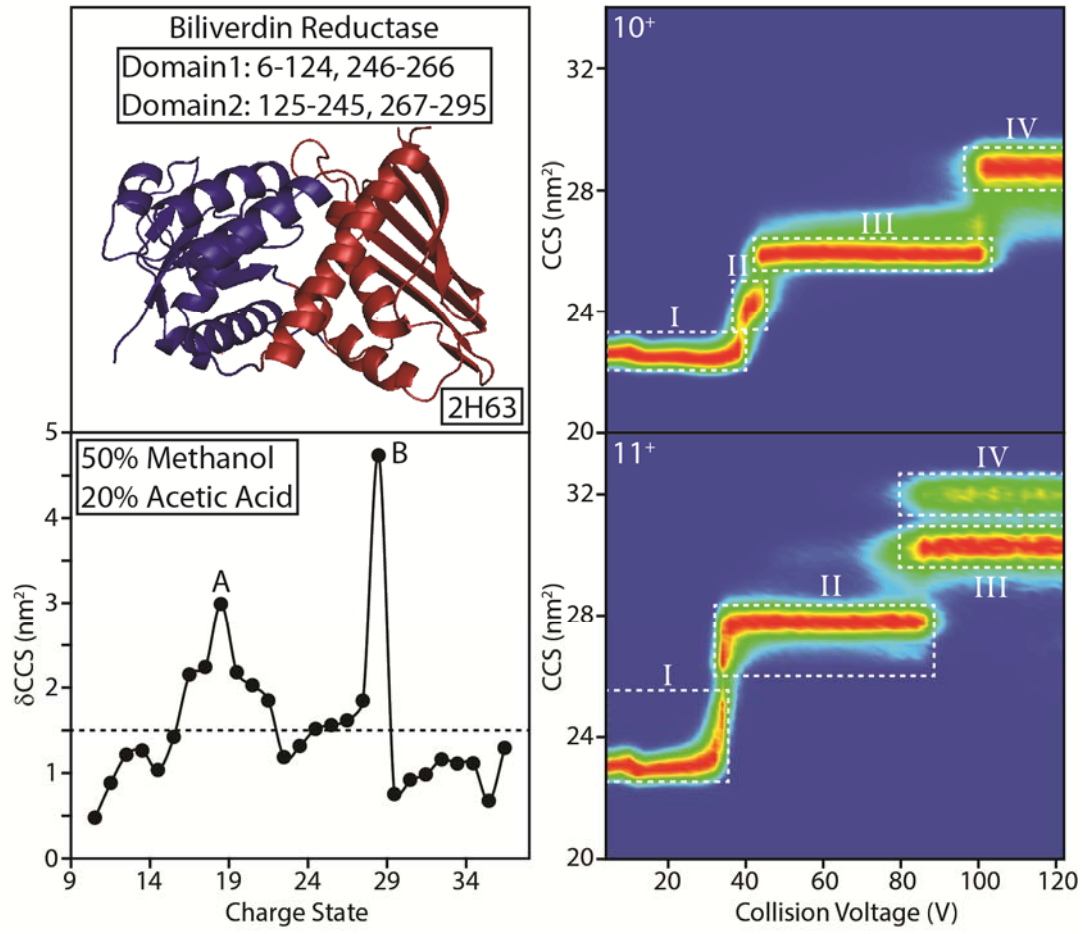


**Figure S6. Crystal structure (top, left), Coulombic unfolding (bottom left) and CIU (right) data of Biliverdin Reductase (BVR).** Each domain of BVR consists of 2 sequence segments, according to CATH. However, these two domains are mixed, meaning that the part of the sequence for domain one overlaps spatially with domain two, and vice-versa. The Coulombic unfolding data indicate a two-domain unfolding pattern. Nevertheless, both $10^+$ and $11^+$ ions subjected to CIU identify this protein as a having 3-domains, likely due to the inter-mingled nature of the domains in this protein.

**References:**


[1] D. Smith, K. Johnson, *Gene* **1988**, *67*, 31-40.

[2] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, C. V. Robinson, *Nat. Protoc.* **2008**, *3*, 1139-1152.

[3] a) K. Giles, J. P. Williams, I. Campuzano, *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1559-1566; b) A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, H. H. Hill, *J. Mass Spectrom.* **2008**, *43*, 1-22; c) J. A. McLean, B. T. Ruotolo, K. J. Gillig, D. H. Russell, *Int. J. Mass Spectrom.* **2005**, *240*, 301-315; d) C. S. Creaser, J. R. Griffiths, C. J. Bramwell, S. Noreen, C. A. Hill, C. L. P. Thomas, *Analyst* **2004**, *129*, 984-994; e) J. I. Baumbach, G. A. Eiceman, *Applied Spec.* **1999**, *53*, 338A-355A; f) D. E. Clemmer, M. F. Jarrold, *J. Mass Spectrom.* **1997**, *32*, 577-592.

[4] a) L. Han, B. T. Ruotolo, *Int. J. Ion Mobil. Spectrom.* **2013**, *16*, 41-50; b) L. Han, B. Ruotolo, *Angew. Chem. Int. Ed.* **2013**, *52*, 8329–8332; c) J. N. Rabuck, S.-J. Hyung, K. S. Ko, C. C. Fox, M. B. Soellner, B. T. Ruotolo, *Anal. Chem.* **2013**, *85*, 6995-7002; d) S. Niu, J. N. Rabuck, B. T. Ruotolo, *Current Opin. Chem. Biol.* **2013**, *17*, 809-817; e) J. Freeke, M. F. Bush, C. V. Robinson, B. T. Ruotolo, *Chem. Phys. Lett.* **2012**, *524*, 1-9; f) L. J. Han, S. J. Hyung, J. J. S. Mayers, B. T. Ruotolo, *J. Am. Chem. Soc.* **2011**, *133*, 11358-11367; g) S.-J. Hyung, C. V. Robinson, B. T. Ruotolo, *Chem. Biol.* **2009**, *16*, 382-390.

[5] Y. Zhong, S.-J. Hyung, B. T. Ruotolo, *Analyst* **2011**, *136*, 3534-3541.