Supporting Information for Publication

Towards a universal sample preparation method for denaturing top-down proteomics of complex proteomes

Zhichang Yang[1], Xiaojing Shen[1], Daoyang Chen[1], Liangliang Sun[1,*]

[1] Department of Chemistry, Michigan State University, 578 S Shaw Ln, East Lansing, MI 48824 USA

* Corresponding author. E-mail: lsun@chemistry.msu.edu

TABLE OF CONTENTS

Supplemental Experimental section

Materials and Reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise. Urea was purchased from Alfa Aesar (Haverhill, MA). LC/MS grade water, methanol, chloroform, HPLC grade acetic acid (AA), formic acid (FA), and hydrofluoric acid (HF) were purchased from Fisher Scientific (Pittsburgh, PA). Acrylamide was ordered from Acros Organics (NJ, USA). Fused silica capillaries (50 μm i.d./360 μm o.d.) were purchased from Polymicro Technologies (Phoenix, AZ). Carboxylate-modified paramagnetic beads (Sera-Mag SpeedBeads (hydrophilic) CAT # 45152105050250, and Sera-Mag SpeedBeads (Hydrophobic), CAT # 65152105050250) were purchased from GE Healthcare. Centrifugal filter units with a 30-kDa molecular weight cutoff for ultrafiltration were purchased from Millipore (MRCF0R030). CZE insert tubes (CAT # C4010-630P) were purchased from Thermo Scientific. The polyacrylamide gel for SDS-PAGE was purchased from Bio-rad (CAT # 4561094). The gel staining buffer was purchased from Bio-rad (CAT # 1610803).

SDS-PAGE analysis

The *E. coli* and HepG2 cell lysates before and after cleanup with the three methods were analyzed by SDS-PAGE according to the procedure in the literature.[1] The gel was first rinsed with D.I. water for 5 min for 3 times. Coomassie blue staining buffer was used for overnight staining with gentle swing. The destaining was processed with D.I. water rinsing.

CZE-MS/MS analysis

A 100-cm linear polyacrylamide (LPA)-coated capillary (50/360 μm i.d./o.d.) was used for CZE separation. The LPA coating was prepared following the published procedure.[2,3] One end of the capillary was etched with HF following the published procedure to reduce the outer diameter of the capillary to 70-80 μm.[4] The commercialized electrokinetically pumped sheath flow CE-MS interface (EMASS II, CMP scientific, Brooklyn, NY) was used to couple CZE to MS.[5,6] The automated CZE operations were implemented with an ECE-001 autosampler (CMP scientific). The sheath buffer contained 10% (v/v) methanol and 0.2% (v/v) FA. The sample buffer was 100 mM NH_4HCO_3 (pH 8) and the background electrolyte (BGE) was 20% or 40% (v/v) acetic acid in water. The glass emitter for the

electrospray was pulled from borosilicate glass capillary (0.75 mm i.d., 1 mm o.d.) by a Sutter P-1000 flaming/brown micropipette puller. The orifice of the emitter was controlled at 20-40 μm. The distance of the etched capillary tip to the emitter orifice was less than 300 μm and the distance of the emitter orifice to the MS entrance was around 2 mm. The sample was loaded with 5 psi for 90 s so approximately 500 nL of the sample was loaded into the capillary. The capillary sample injection end was immersed in a CE vial containing 100 mM NH_4HCO_3 (pH 8) for 10 s to neutralize the leftover acetic acid on the outer surface of the capillary injection end before it was moved into the sample vial for sample injection. After sample loading, the capillary sample injection end was moved into a BGE vial and the CZE separation was carried out by applying a +30-kV voltage for 115 min. A 15-psi pressure was applied afterwards for 5 min to flush and condition the capillary. A +2-kV voltage was applied in the sheath buffer vial for ESI.

A Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) was used for all CZE-MS/MS analyses. A data-dependent acquisition (DDA) method was employed. The full MS scan range was 600–2000 *m/z*. Mass resolution was 120,000 (at *m/z* 200) and AGC was 1E6. Maximum injection time was 50 ms. The top 3 most abundant ions in a MS spectrum were isolated in the quadrupole with a 4-m/z isolation window sequentially and fragmented with higher-energy collisional dissociation (HCD) with the normalized collision energy 20. The mass resolution for MS/MS was 60,000 (at *m/z* 200). The maximum injection time was 200 ms and the AGC was 1E5. The ion intensity threshold was 2E4 for triggering MS/MS. The dynamic exclusion was turned on and set at 30 s. Charge exclusion was enabled and ions with charges from +1 to +3 as well as ions with unassigned charge states were excluded from MS/MS.

Table S1. Intensity and migration time (MT) information of 8 high intense peaks in the MT range of 35-80 min from the CZE-MS/MS analyses of the Batch 1 and Batch 2 *E. coli* samples prepared with the MU method.

Peak #	Charge	MS (m/z)	Batch1 intensity	Batch2 intensity	RSD (%)	Batch1 MT (min)	Batch2 MT (min)	RSD (%)
1	9	1083.10	5.10E+08	6.70E+08	19.18	39.26	37.69	2.89
2	5	732.60	1.30E+08	1.47E+08	8.68	44.75	42.81	3.13
3	5	772.80	3.20E+08	2.60E+08	14.63	46.76	44.74	3.12
4	5	863.04	1.46E+08	1.12E+08	18.64	56.57	54.13	3.12
5	3	675.33	7.80E+07	7.60E+07	1.84	58.22	55.79	3.01
6	3	828.46	9.26E+07	1.09E+08	11.50	64.85	62.06	3.11
7	2	668.36	3.60E+07	4.30E+07	12.53	68.01	64.69	3.54
8	2	920.96	1.70E+07	1.85E+07	5.98	78.81	74.59	3.89

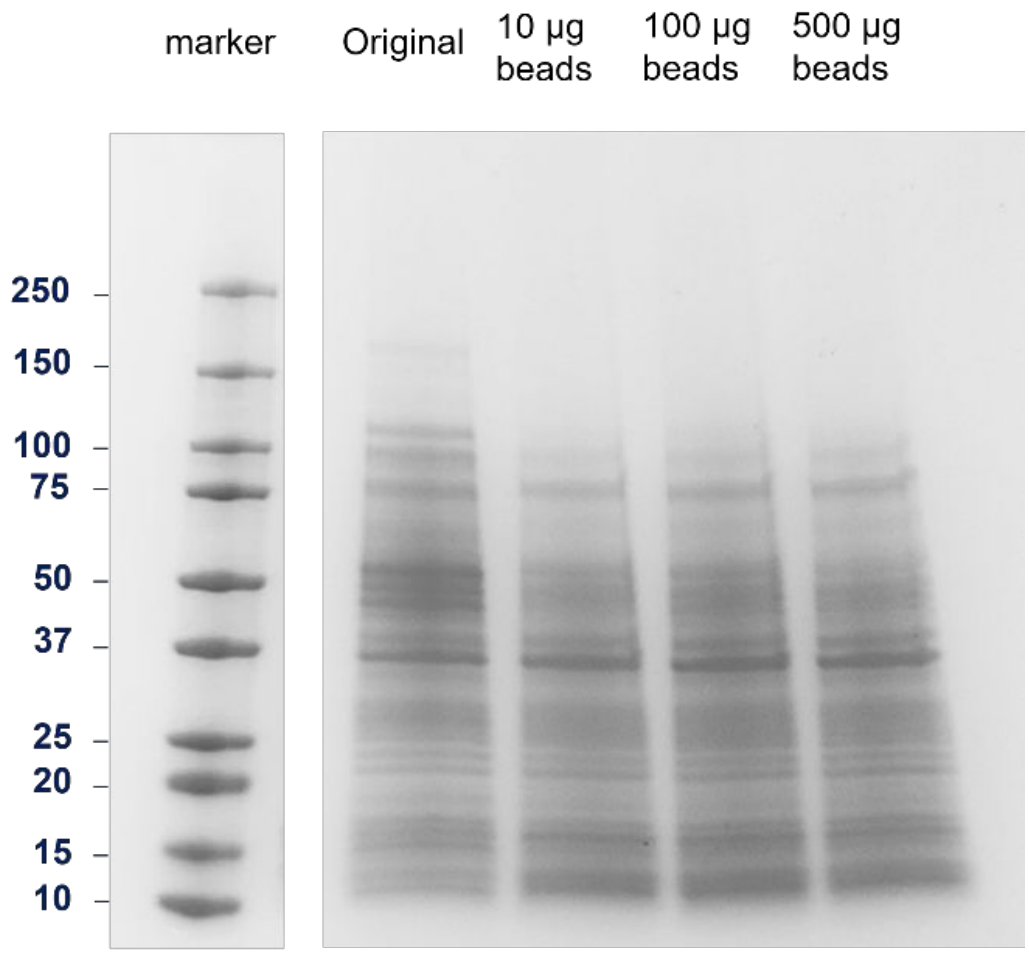

Figure S1. SDS-PAGE analysis of *E. coli* proteins processed by the SP3 method with three different amounts of beads. The original sample is the cell lysate in 1% (w/v) SDS without sample cleanup. For all the four *E. coli* samples, 6-µg *E. coli* protein was loaded for analysis.

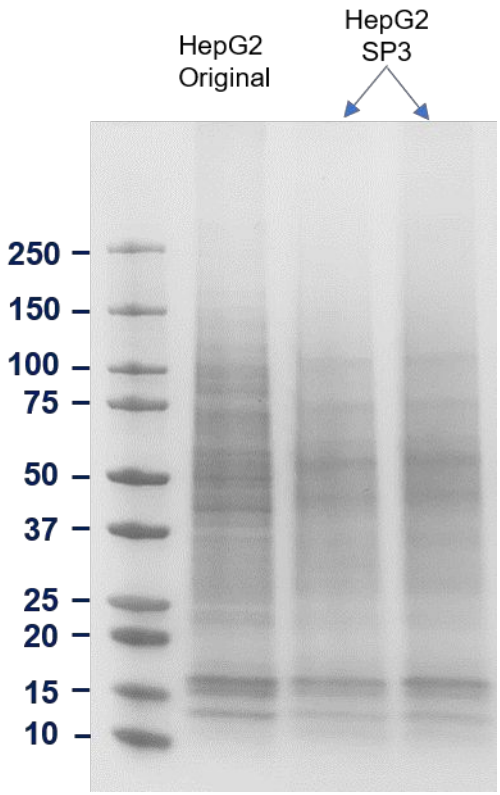

Figure S2. SDS-PAGE analysis of HepG2 proteins processed by the SP3 method (n=2). The original sample is the cell lysate in 1% (w/v) SDS without sample cleanup. 7 µg of total proteins were loaded on each lane.

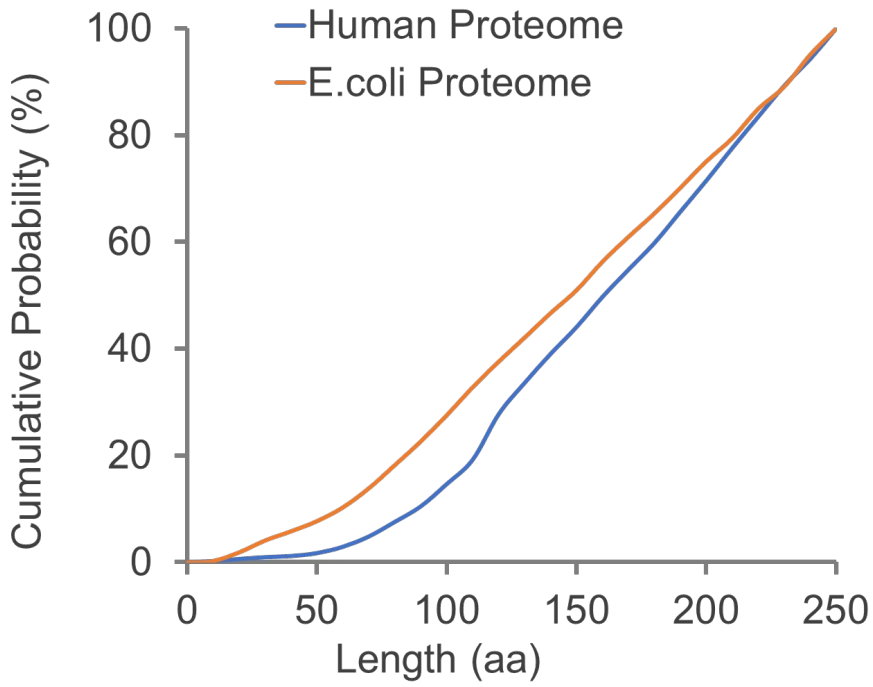

Figure S3. Cumulative distribution of the length of *E. coli* proteins and human proteins in the Swiss-Prot database in a length range of 1-250 amino acids (aa).

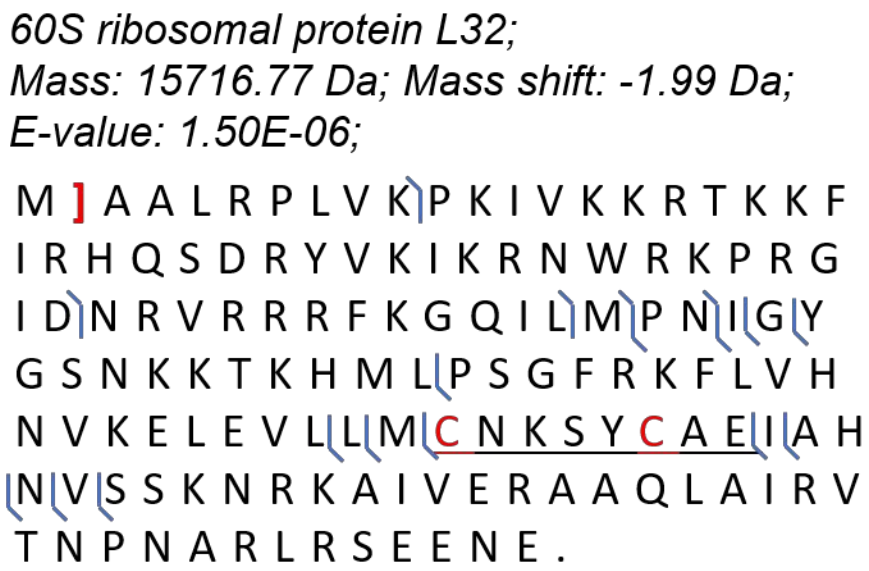

Figure S4. Sequence and fragmentation pattern of one proteoform of 60S ribosomal protein L32 with one disulfide bond between the two cystine residues marked in red.

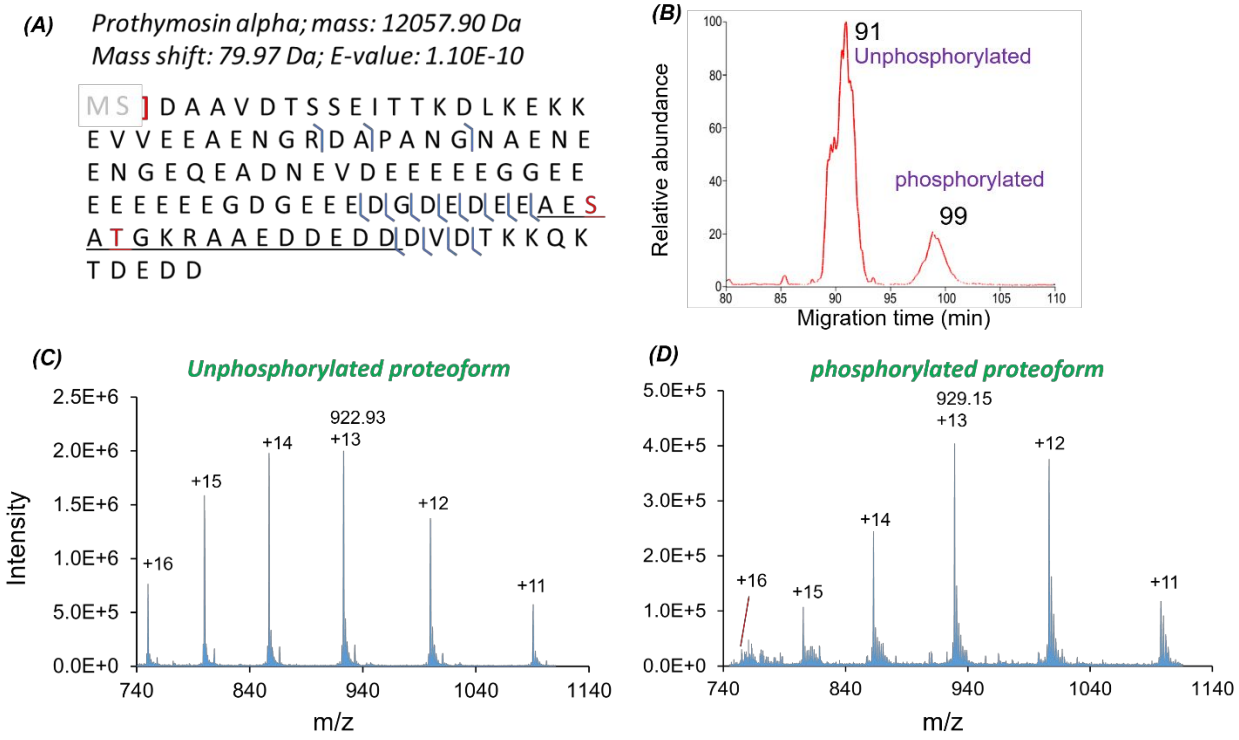

Figure S5. (A) Sequence and fragmentation pattern of the prothymosin alpha proteoform with one phosphorylation at the marked serine or threonine residue. (B) Extracted base peak electropherogram of phosphorylated and unphosphorylated proteoforms of prothymosin alpha. The highest abundant charge state (+13) was used for peak extraction with a 20-ppm mass tolerance. Mass spectra of the unphosphorylated (C) and phosphorylated (D) proteoforms of prothymosin alpha.

Supplemental References

1. Blancher, C.; Jones, A., SDS -PAGE and Western Blotting Techniques. *Methods in molecular medicine* **2001,** *57*, 145-62

2. McCool, E. N.; Lubeckyj, R.; Shen, X.; Kou, Q.; Liu, X.; Sun, L., Large-scale Top-down Proteomics Using Capillary Zone Electrophoresis Tandem Mass Spectrometry. *Journal of visualized experiments: JoVE* **2018,** *140*, e58644.

3. Zhu, G.; Sun, L.; Dovichi, N. J., Thermally-initiated free radical polymerization for reproducible production of stable linear polyacrylamide coated capillaries, and their application to proteomic analysis using capillary zone electrophoresis-mass spectrometry. *Talanta* **2016,** *146*, 839-43.

4. Sun, L.; Zhu, G.; Zhao, Y.; Yan, X.; Mou, S.; Dovichi, N. J., Ultrasensitive and fast bottom-up analysis of femtogram amounts of complex proteome digests. *Angewandte Chemie* **2013,** *52* (51), 13661-4.

5. Sun, L.; Zhu, G.; Zhang, Z.; Mou, S.; Dovichi, N. J., Third-generation electrokinetically pumped sheath-flow nanospray interface with improved stability and sensitivity for automated capillary zone electrophoresis-mass spectrometry analysis of complex proteome digests. *Journal of proteome research* **2015,** *14* (5), 2312-21.

6. Wojcik, R.; Dada, O. O.; Sadilek, M.; Dovichi, N. J., Simplified capillary electrophoresis nanospray sheath-flow interface for high efficiency and sensitive peptide analysis. *Rapid communications in mass spectrometry : RCM* **2010,** *24* (17), 2554-60.