Top-down Proteomics Analysis of Picogram-level Complex Samples using Spray-Capillary-Based Capillary Electrophoresis Mass Spectrometry

Zhitao Zhao[1], Yanting Guo[1], Trishika Chowdhury[1], Samin Anjum[1], Jiaxue Li[1], Lushuang Huang[1], Kellye A. Cupp-Sutton[1], Anthony Burgett[2], Dingjing Shi[3], Si Wu*[,4]

[1]*Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Parkway Norman, OK 73019, USA*

[2]*Dept. of Pharmaceutical Sciences, University of Oklahoma Health Sciences, 1110 N. Stonewall Ave. Oklahoma City, OK 73117, USA*

[3]*Department of Psychology, University of Oklahoma, 455 W Lindsey Street, Norman, OK 73069, USA*

[4]*Department of Chemistry and Biochemistry, University of Oklahoma, 101 Stephenson Parkway Norman, OK 73019, USA or Department of Chemistry and Biochemistry, University of Alabama, 250 Hackberry ln, Tuscaloosa, AL 35487, USA*

*Corresponding author: Si Wu e-mail: swu49@ua.edu

Keywords: Single cell proteomics, Top-down proteomics, CE-MS, Spray-capillary, Quantitative proteomics

Abstract:

Proteomics analysis, including post-translational modifications (PTMs) of mass-limited samples has become an important method for understanding biological systems in physiologically relevant contexts, such as patient samples and in multicellular organoids and spheroids. There is a growing need to develop ultrasensitive top-down proteomics techniques to provide valuable insights into PTM-regulated cellular functions in mass-limited samples, including single cells. Capillary electrophoresis–mass spectrometry (CE-MS) is a promising technique due to its high resolution and sensitivity compared to liquid chromatography (LC)-based separation techniques. We recently developed "Spray-Capillary", an electrospray ionization (ESI)-assisted device for quantitative ultralow-volume sampling and online CE-MS analysis. In this study, we present an enhanced spray-capillary-based CE-MS platform using the polyethyleneimine (PEI) coated capillary for ultrasensitive top-down proteomics analysis. Under optimized conditions, we detected >200 proteoforms from 50 pg E. coli lysate, which is approximately one-tenth of the protein mass in a single mammalian cell. Using a nanodroplet-based sample preparation method and our optimized CE-MS platform, we reproducibly detected 867 intact proteoforms in HeLa cells and 711 intact proteoforms in OVCAR-8, a type of ovarian cancer cell (~45 cells per analysis). Overall, our results demonstrate the capability of the Spray-Capillary CE-MS system to perform top-down proteomic analysis on picogram amounts of sample, and this advancement presents the possibility of meaningful top-down proteomic analysis of mass-limited samples down to the level of single mammalian cells.

Traditional proteomics studies, especially top-down studies, require analyzing bulk cell lysate produced by homogenizing a large number of cells to obtain sufficient analyte for separation and mass spectroscopy (MS) analysis[1]. However, the lack of analytical sensitivity that necessitates the bulk lysate limits the scope and biological impact of the MS analysis. Mass-limited samples such as patient clinic samples (*e.g.*, tissue biopsies or spheroids) and even single cells might not produce enough analyte for proteomics. Additionally, the requirement to blend the proteomes of thousands of different cells destroys the spatial and temporal organization of biological tissues.[2, 3] Significant progress has been made in bottom-up-based single-cell proteomics (SCP), including advancements in small-scale proteomics sample preparation, sensitive separation techniques, and high-resolution mass spectrometry instrumentation[4-15]. These developments have enabled the application of bottom-up proteomics to analyze mass-limited samples such as single cells[16]. Bottom-up proteomics, however, requires protein digestion which can result in the loss of structural information specific to the biologically active proteoforms, including post-translational modifications (PTMs)[17]. Top-down proteomics analyzes intact proteoforms directly and can observe structural variations as a result of genetic variation, alternative splicing of RNA transcripts, and PTMs[16, 18-20]. Several attempts have been made in top-down proteomics for mass-limited samples, including SCP[21-24]. However, top-down proteomics generally suffers from low sensitivity due to high sample complexity, wide dynamic range, and wide isotopic/charge state distributions of proteoforms[25-28]. This issue is particularly prominent for analysis of mass limited samples as the number of proteoform copies is limited by the total sample mass. As a result, front-end separation is critical to improving sensitivity for top-down proteomics by decreasing sample complexity and improving sensitivity[29].

Recently, capillary electrophoresis (CE) has emerged as a highly effective technique in the field of top-down proteomics, yielding remarkable results. This can be attributed to several advantages associated with CE, including its exceptional separation efficiency, ability to handle ultra-low sample volumes (*e.g.*, pL-nL), and impressive sensitivity[9, 16, 24, 30-33]. The low diffusion coefficients of large molecules result in high theoretical plate numbers for the CE separation of biological samples[34]. As sample adhesion has been demonstrated to be a primary bottleneck in mass-limited and single-cell proteomics[35], minimizing sample

injection volume requirements and sample dilution improves sensitivity. CE offers a solution by enabling the analysis of small-volume, mass-limited samples with exceptional sensitivity, thereby overcoming the limitations associated with sample adhesion. Therefore, CE is capable of analyzing small-volume, mass-limited samples with high sensitivity for top-down proteomics, as demonstrated in a few recent studies[4, 24, 27, 30, 32, 36, 37]

Despite the high separation efficiency and low sample consumption of CE, ultralow volume sample injection down to pL-nL still remains challenging. Typically, hydrodynamic[2] and electrokinetic methods[38] are used for low- volume CE sample injection; however, these approaches commonly require dedicated setups that may not be generally suitable for all labs. Additionally, these approaches can only be applied for offline sampling with relatively low throughput and may also introduce sample loss during transfer and handling. As an alternative for ultralow volume sampling, we have developed the 'Spray-Capillary'

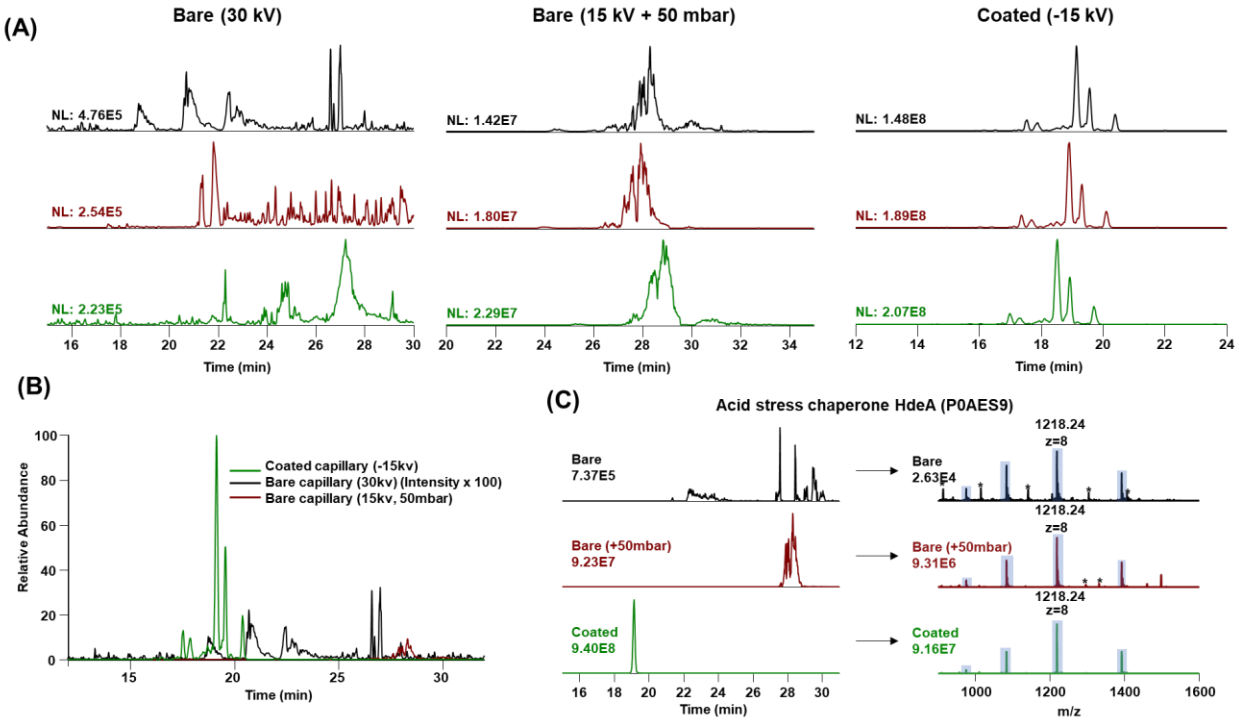

Figure 1. Evaluation of CE separation performance for bare and PEI coated capillaries. 3 ng *E. coli* cell lysate was injected with a 120 s spray-capillary injection time. (**A**) Triplicate base peak electropherograms BPEs (700-1600 *m/z*) for coated capillary (-30 kV), bare capillary (-15 kV), and bare capillary with pressure elution (-15 kV + 50 mbar). (**B**) overlayed BPEs for a coated capillary, bare capillary, and bare capillary with pressure elution run. Intensity of the black trace increased 100-fold before plotting to make the spectra visible. (**C**) Extracted ion electropherograms (EIEs) and charge state distribution of the Acid stress chaperone (HdeA) using each separation condition.

device[39, 40]. The spray-capillary utilizes the pressure difference produced by electrospray ionization (ESI) to drive ultralow volume sampling (*e.g.*, pL/s). Furthermore, this device can be directly used for online CE separation and MS detection with limited sample loss and improved reproducibility, as demonstrated in our previous application to single-cell metabolomics[31]. Here we introduced an improved spray-capillary-based CE-MS platform for ultrasensitive top-down proteomics analysis.

Intact proteins tend to adsorb onto the inner surface of the fused silica capillary during the CE-MS analysis (particularly when pH < 4), leading to decreased sensitivity[32]. To address this limitation, functionalizing the inner surface with the coating material, such as linear polyacrylamide (LPA)[41] or polyethyleneimine (PEI)[36], can significantly reduce sample loss from non-specific absorption. Moreover, it has been previously demonstrated that electroosmotic flow (EOF) can be regulated by changing the chemistry of the inner walls of the capillary[42]. In this study, we first evaluated the performance of spray-capillary devices fabricated from a bare fused silica capillary and PEI-coated capillary (**Figure 1**) to separate intact proteins.

Two spray-capillary devices (bare and PEI-coated, 85 cm length, 50 μm ID) were fabricated (details in *Supporting information*). The sampling rate was measured as previously reported[40], and the coating did not have any significant effect on the sampling rate as shown in **Supporting information Figure S1A** (150 ± 6 pL/s for PEI-coated capillary and 147 ± 6 pL/s for bare capillary). We first performed a 60 s spray-capillary sample injection with 3 kV ESI voltage to sample 3.1 ± 0.1 ng and 3.0 ± 0.1 ng *E. coli* lysate using the coated and bare spray-capillary devices, respectively. The electropherogram spectra and signal intensities were not reproducible run-to-run for the bare capillary, as demonstrated in **Figure 1A**. Reproducibility was improved for the bare capillary when applying pressure (*e.g.*, 50 mbar) to facilitate protein elution; however, despite some improvement, the separation resolution was still poor as demonstrated by overlapping, unresolved peaks. The PEI-coated spray-capillary demonstrated improved separation efficiency and reproducibility in migration times compared with the bare capillary as demonstrated by the well resolved peaks and consistent migration times. **Figure 1B** displays overlayed BPE spectra for each spray-capillary and separation condition. The average maximum BPE intensity for

the coated capillary is 1.81E8 ± 3.02E7 (RSD = 17%). This represents a more than 500-fold increase compared with the bare capillary (3.18E5 ± 1.38E5, RSD = 43%) and bare capillary with pressure elution (1.84E7 ± 4.36E7, RSD = 24%). This increase in sensitivity for the PEI-coated spray-capillary also allowed detection of a higher number of proteoforms (247 ± 58) compared with the bare capillary (42 ± 3), and the bare capillary with pressure elution (130 ± 8) (**Supporting information Figure S1B**). We observe an increase in proteoform detection for all mass ranges with the PEI-coated spray-capillary (**Supporting information Figure S1C**). In particular, the increase in sensitivity using the PEI-coated spray-capillary allowed the detection of larger molecular weight proteoforms (>15 kDa) that were not detected using the bare spray-capillary. **Supporting information Figure S2** displays the intensity correlation of the detected intact proteoforms among the triplicates under different conditions. It was observed that the R^2 values exhibited a significant increase from 0.51 ± 0.13 for the bare capillary to over 0.93 ± 0.027 for the coated spray-capillary, indicating a notable improvement in reproducibility compared to the bare capillary.

An example protein identified using the PEI-coated and bare spray-capillary devices, the acid stress chaperone HdeA (P0AES9), is shown in **Figure 1C**. The EIE intensity of the protein increased from 7.37E5 (bare capillary) to 9.40E8 (PEI-coated capillary). This represents a 1275-fold increase in maximum signal intensity for the PEI-coated spray-capillary compared with the bare capillary and demonstrates that the sensitivity was improved by PEI coating. Our study demonstrates addition of PEI coating to the spray-capillary device improves reproducibility and sensitivity for intact proteoforms. Thus, the PEI-coated spray-capillary is more suitable for intact protein CE-MS analysis of nL-level samples with limited (*e.g.*, ng-level) mass.

To evaluate the effect of capillary ID on CE separation, two PEI-coated spray-capillaries with 50 µm and 20 µm IDs were fabricated at the same length (details in *Supporting information*). The sampling rate for the two capillaries was measured as done previously[40]. The sampling rate was 24 ± 1 pL/s for the 20 µm ID PEI-coated spray-capillary and 239 ± 4 pL/s for the 50 µm ID spray-capillary (**Supporting information Figure S3A**). Appropriate injection times were selected for each spray-capillary to ensure

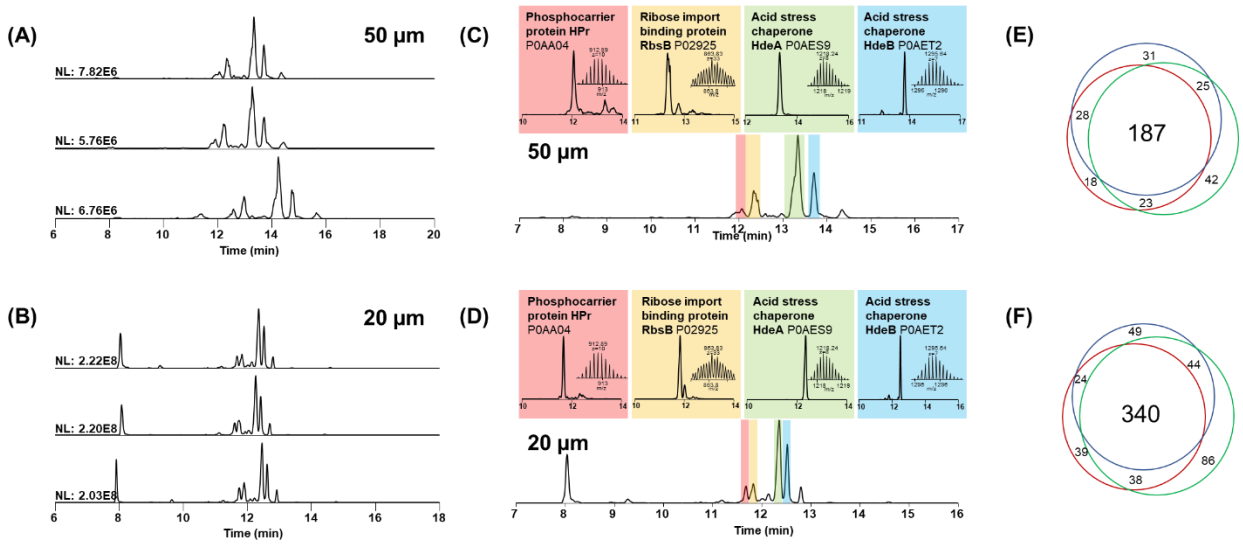

Figure 2. Evaluation of CE separation performance for PEI-coated spray-capillary with different inner diameters. Base peak electropherograms (BPEs, 700-1600 *m/z*) from triplicated analyses using (**A**) 50 µm spray-capillary and (**B**) 20 µm ID spray-capillary. Extracted ion electropherograms (EIEs) and charge state distribution of selected proteoforms using **(C)** 50 µm spray-capillary and (**D**) 20 µm ID spray-capillary. Venn diagram of detected proteoforms among triplicate runs using **(E)** 50 µm spray-capillary and (**F**) 20 µm ID spray-capillary.

the same sample amount was injected. The sample injection time was 12 s for the 50 µm ID spray-capillary and 120 s for the 20 µm ID spray-capillary for injection of 1 ± 0.02 ng and 1 ± 0.04 ng intact *E. coli* lysate, respectively. Triplicate BPEs are shown in **Figure 2A** and **2B** using the 50 and 20 µm ID PEI-coated spray-capillaries, respectively. The BPE intensity using the 20 µm ID spray-capillary was 2.2E8 ± 1E7, which was ~32-fold higher compared with the BPE intensity measured using the 50 µm ID spray-capillary (6.9E6 ± 1E6).

Figure 2C and 2D show EIEs for 4 arbitrarily selected proteoforms identified using both capillaries. A substantial enhancement in protein signal intensity across all four proteoforms was observed when the 20 µm ID PEI-coated spray-capillary was employed (**Supporting Information Table S1)**. Because the spray-capillary CE-MS uses the sheathless interface, decreasing the ID of spray-capillary leads to decreased emitter ID *(e.g.*, the emitter ID is ~20 µm for 20 µm ID spray-capillary), which can significantly increase charge density within the droplet during the ESI process and facilitate more efficient conversion of the droplet into the gas phase[43]. We also evaluated the separation efficiency for the 20 µm ID and 50 µm ID spray-capillary devices. The theoretical plate numbers of the identified proteoforms ranged from 69K to 147K when using the 20 µm ID spray-capillary and increased 22-128% compared to

the 50 μm ID spray-capillary **(Supporting Information Table S1)**. It was observed that the theoretical plate numbers of the identified proteoforms ranged from 69K to 147K when using a 20 μm ID spray-capillary, which represents an increase of 22-128% compared to the 50 μm ID spray-capillary. It has been reported that reducing the inner diameter of the separation capillary can enhance the separation efficiency in CE due to a decrease of Joule heat generation.[44] The excess heat from the current in the capillary can cause peak broadening and reduce the separation resolution. The number of detected proteoforms was 286 and 584 using the 50 μm and 20 μm ID PEI-coated spray-capillaries, respectively, as shown in **Figures 2E and 2F (**details in **Supporting information Figure S3B)**. 263 (74%) proteoforms were detected in two or more runs using the 50 μm ID capillary and 446 (72%) of the proteoforms were detected in two or more runs using the 20 μm ID spray-capillary. Overall, we found that implementation of the PEI coating and reduction of the ID of the spray-capillary improved CE separation efficiency and sensitivity for low volume samples of intact protein mixtures. The 20 μm ID, PEI-coated spray-capillary format was used for all subsequent experiments.

To evaluate the spray-capillary's ability to conduct quantitative analysis, we performed ultralow volume top-down proteomics analysis using different amounts of intact *E. coli* lysate. Using a 120 second injection time with a sampling rate of 21 ± 1 pL/s, 0.05 ± 0.002 ng, 0.1 ± 0.005 ng, 0.5 ± 0.02ng and 1 ± 0.05 ng *E. coli* lysate was injected (n = 3). The BPE intensities increased as injection mass increased, as shown in **Supporting information Figure S4A-D**. A calibration curve was plotted for average BPE intensity *vs* sample mass and a high linear correlation ($R^2 > 0.99$) was observed (**Supporting information Figure S4E)**. In addition, **Figure S4G and S4H** displayed calibration curves for two arbitrarily selected proteoforms, Acid stress chaperone HdeA and Phosphocarrier protein HPr. These examples also exhibit a strong linear correlation between the sample injection amount and the intensity of the signal ($R^2 > 0.99$). Moreover, the number of detected proteoforms increased with increased sample injection mass, from 233 for 0.05 ng to 532 for 1 ng, as expected. Overall, we have demonstrated that the spray-capillary device is a viable method for quantitative analysis of intact proteoforms in mass-limited samples.

As the total amount of protein predicted to exist within a human somatic cell is 0.5 ng[12], we further examined the 50 pg sampling amount results (*e.g.*, 1/10 of the cell content) to contextualize our results with single cell-level CE-MS analysis **(Figure 3)**. A total of 233 unique proteoforms were detected, in which 150 (67%) proteoforms were detected in two or more runs as shown in **Figure 3B**. The mass distribution of the unique proteoforms is shown in **Figure 3C,** and we find that we can detect 45 larger molecular weight proteoforms (>15 kDa). **Figure 3D** also showed a few selected proteoforms (molecular weights vary from 9k to 18k) with isotopic distributions and charge envelopes. Overall, the spray-capillary-based CE-MS platform was capable of characterizing hundreds of proteoforms from pg-level complex samples. Previous top-down CE-MS studies on complex samples such as cell lysate analyzed much larger masses of protein (*e.g.*, ~0.25 ng[36], 480 ng[45], 1 µg[46]) and LC-MS-based analysis of these

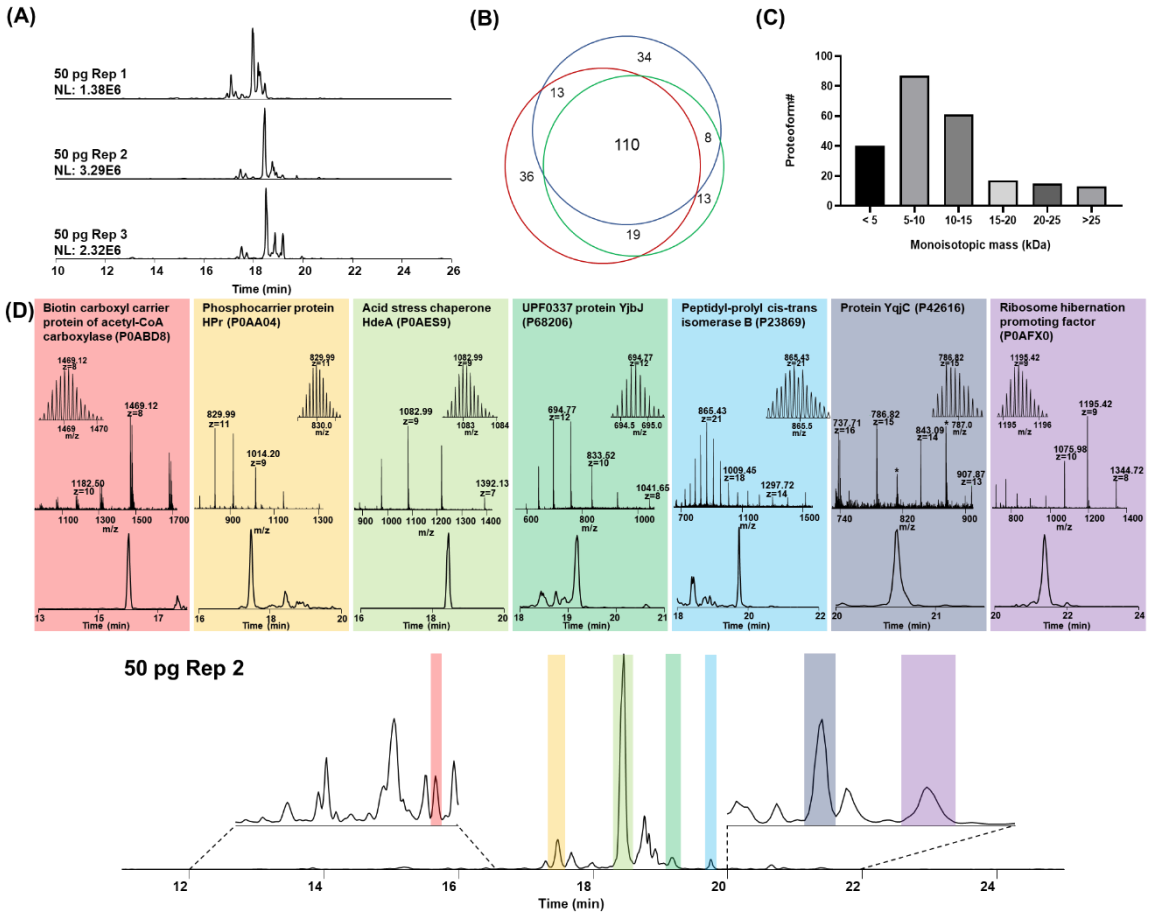

Figure 3. Ultrasensitive top-down analysis of 50pg *E. coli* cell lysate using spray-capillary CE-MS. **(A)** Base peak electropherograms BPEs (700-1600 *m/z*) from three replicated runs; **(B)** Venn diagram of detected proteoforms for three replicated runs; **(C)** Mass distribution of detected proteoforms; **(D)** Extracted ion electropherograms (EIEs) and charge state distribution of selected proteoforms.

samples also require relatively high mass (*e.g.*, ≥ 10 ng [21]). Our results demonstrate the capability of this platform to perform ultra-sensitive top-down analysis with limited mass samples (*e.g.*, 50 pg) down to the mass of protein present within a single mammalian somatic cell.

To demonstrate that the optimized spray-capillary device can be utilized to perform top-down proteomics on small numbers of intact human cells, we analyzed intact *HeLa* cells and ovarian cancer cells using a microdroplet-based sample preparation method, similar to previously reported methods used for bottom-up SCP that demonstrated minimal sample loss (platform schematic in **Supporting information Figure S5**). *HeLa* cells, OVCAR-8 cells, and bulk *HeLa* lysate were analyzed using the top-down spray-capillary CE-MS platform with this droplet-based sample preparation method (100 cells per droplet, ~45 cells per injection). To better control sample injection under the microscope, a large ID PEI-coated spray-capillary (360 µm OD, 50 µm, 80 cm length) was used for sample injection and CE-MS analysis. **Figure 4A** shows triplet runs from all three sample types (example runs shown in **Supporting information Figure S6)**. The mass distribution of the detected proteoforms from each sample type is shown in **Figure 4B**. Overall, 867 proteoforms were detected from intact *HeLa* cells, 711 proteoforms were detected from OVCAR-8 cells, and 778 were detected from *HeLa* bulk cell lysate. In total, 1397 unique proteoforms were detected from all runs of the two cell lines. The mass distribution of the unique

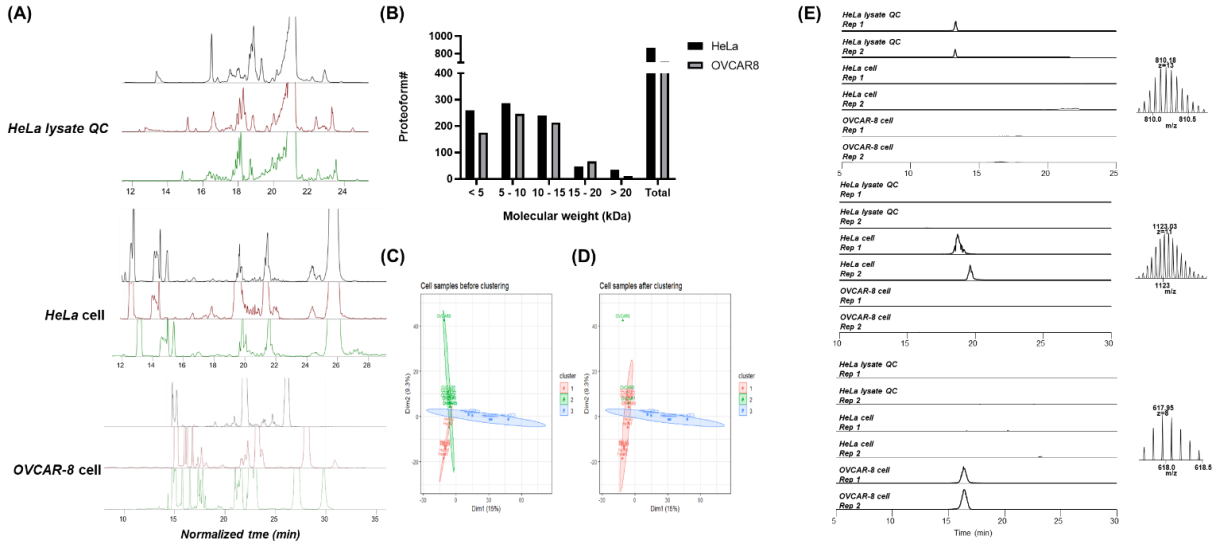

Figure 4. Spray-capillary-based CE-MS analysis of *HeLa* cells, OVACAR-8 cells, and bulk *HeLa* lysate. **(A)** Base peak electropherograms (BPEs, 700-1600 *m/z*) for three replicate runs for each sample type; **(B)** Mass distribution of detected proteoforms. PCA plots **(C)** before and **(D)** after unsupervised K-means clustering. (E) Selected proteoform examples uniquely detected in each sample type.

proteoforms is shown in **Figure 4B**. Principal component analysis was performed on the intact *HeLa* and OVCAR-8 cells as well as bulk *HeLa* lysate. The PCA plot (*e.g.*, before clustering) displays the three sample types together as expected. (**Figure 4C**). We further conducted *k*-means clustering analysis to investigate the separation of distinct group patterns among these samples (**Figure 4D**). *K*-means clustering is an unsupervised statistical learning technique to partition observations into clusters[47-49]. Using the technique, we successfully partitioned the 22 sample cells into three clusters. Figure 4D depicts clusters of cells using different symbols and colors. The larger symbols represent centroids, which are the center of the clusters. The small symbols with the same color belonged to the same cluster. The *k*-means technique well separated the *HeLa* bulk cell lysate into one distinct cluster (indicated by the red circles). The *HeLa* cells well formed a second cluster (shown as blue squares). A third cluster emerged for OVCAR-8 cells, shown in green triangles. A few OVCAR-8 cell datasets were closely situated to the *HeLa* cluster and thus were grouped with the *HeLa* cluster. The overlap in cell clustering is not unexpected cancerous cells may exemplify high levels of phenotypic similarity. Improving the clustering performance can be achieved by increasing the number of sampling points and implementing data preprocessing techniques such as handling missing values. Overall, our proof-of-principle experiments demonstrated that the top-down spray-capillary based CE-MS platform can distinguish the cell type and the lysis method for the same type of cell. This also demonstrates the potential of the top-down spray-capillary CE-MS platform for the analysis of single mammalian cells. With the implementation of well-developed cell sorting techniques (*e.g.*, FACS or CellenONE®), we expect that we can perform "true" single-cell proteomics using this platform.

Using our spray-capillary-based CE-MS platform, we demonstrated a sensitive top-down proteomics analysis of mass limited samples of intact *E. coli* lysate and small numbers of human cells. We evaluated the spray-capillary parameters including PEI coating and capillary ID for separation performance using intact protein lysate. Notably, this platform detected 233 unique mass features from only 50 pg *E. coli* cell lysate samples, indicating that this platform can reach sub-cellular sensitivity. Furthermore, we analyzed *HeLa* and OVCAR-8 cells using a droplet-based sample preparation method

and more than 700 mass features were detected for both cell lines. Our spray-capillary-based CE-MS platform performed ultrasensitive top-down analysis and established the foundation for single-cell top-down proteomics using CE-MS.

In our proof-of-principle experiments, we used a large ID PEI-coated spray capillary due to manual handling under microscope and limitations in dispensing ultralow volumes (e.g., <20 nl). New nanodroplet dispensing instruments, such as CellenONE® and Tecan UNO, can facilitate better control of much smaller nanodroplets, which can be incorporated with smaller ID (*e.g.*, 20 μm) spray-capillary CE-MS analysis for improved sensitivity. The capillary coating can be further optimized for separation by altering the chemistry of the inner capillary walls [46]. Different types of coating material such as the neutral coating LPA can be applied to the spray-capillary device to eliminate EOF for longer separation window and higher resolution so more proteoforms can be characterized in single cell samples[50]. Other potential coatings such as PVA and PS1 have also been applied for CE-MS-based proteomics and could potentially be implemented with the spray-capillary device[51, 52]. The background electrolyte can also be optimized by adding organic or changing the pH[53, 54]. Furthermore, the number of identified proteoforms was relatively low due to the limited number of ions that could be collected in the ion trap within the maximum injection time. Previously, for bottom-up proteomics on mass limited samples, TMT labeling has been implemented for multiplexed analysis and higher injection masses. Our lab has recently developed a method for intact protein TMT labeling of complex samples[55, 56]. Optimization of this intact protein TMT labeling platform for single cell analysis for multiplexed top-down proteomics would increase the injected mass and improve proteoform identification. Additionally, multidimensional separations have led to improved proteoform identification in top-down proteomics[57-59]. Coupling nanoLC with the spray-capillary CE-MS may also improve proteoform identification of limited mass samples and single cells.

Overall, we believe that our spray-capillary-based CE-MS platform is capable of ultrasensitive top-down proteomic study and has the potential for single-cell or other mass-limited sample analysis.

Supporting Information

The authors have cited additional references within the Supporting Information.

Acknowledgements

The authors declare that they have no known competing interests to declare that are relevant to the content of this article. This work was partly supported by grants from OCAST HR23-169, OCAST HR-17-116, Oklahoma Translation and Clinical Science Resource Pilot Award, Stephenson Cancer Center Gynecological Oncology Pilot Award, Presbyterian Health Foundation (PHF) Bridge Award, NIH NIAID R01AI141625, and NIH NIH/NIAID2U19AI062629, We also thank the OU Protein Production and Characterization Core (PPC) facility and Dr. Philip Bourne for supporting E. coli protein extraction. The PPC is supported by Institutional Development Awards (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health (Grants P20GM103640 and P30GM145423), the OU Vice President for Research and Partnerships, and the OU College of Arts and Sciences. SW would like to thank for the support received from the University of Alabama's startup fund.

References

1. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M.-C.; Yates, J. R., Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* **2013,** *113* (4), 2343-2394.

2. Onjiko, R. M.; Portero, E. P.; Moody, S. A.; Nemes, P., In situ microprobe single-cell capillary electrophoresis mass spectrometry: metabolic reorganization in single differentiating cells in the live vertebrate (Xenopus laevis) embryo. *Analytical chemistry* **2017,** *89* (13), 7069-7076.

3. Bach, K.; Pensa, S.; Zarocsinceva, M.; Kania, K.; Stockis, J.; Pinaud, S.; Lazarus, K. A.; Shehata, M.; Simões, B. M.; Greenhalgh, A. R.; Howell, S. J.; Clarke, R. B.; Caldas, C.; Halim, T. Y.

F.; Marioni, J. C.; Khaled, W. T., Time-resolved single-cell analysis of Brca1 associated mammary tumourigenesis reveals aberrant differentiation of luminal progenitors. *Nature Communications* **2021,** *12* (1), 1502.

4. Bagwe, K.; Gould, N.; Johnson, K. R.; Ivanov, A. R., Single-cell omic molecular profiling using capillary electrophoresis-mass spectrometry. *TrAC Trends in Analytical Chemistry* **2023**, 117117.

5. Woo, J.; Williams, S. M.; Markillie, L. M.; Feng, S.; Tsai, C.-F.; Aguilera-Vazquez, V.; Sontag, R. L.; Moore, R. J.; Hu, D.; Mehta, H. S., High-throughput and high-efficiency sample preparation for single-cell proteomics using a nested nanowell chip. *Nature communications* **2021,** *12* (1), 6246.

6. Cong, Y.; Motamedchaboki, K.; Misal, S. A.; Liang, Y.; Guise, A. J.; Truong, T.; Huguet, R.; Plowey, E. D.; Zhu, Y.; Lopez-Ferrer, D., Ultrasensitive single-cell proteomics workflow identifies> 1000 protein groups per mammalian cell. *Chemical Science* **2021,** *12* (3), 1001-1006.

7. Dou, M.; Clair, G.; Tsai, C.-F.; Xu, K.; Chrisler, W. B.; Sontag, R. L.; Zhao, R.; Moore, R. J.; Liu, T.; Pasa-Tolic, L., High-throughput single cell proteomics enabled by multiplex isobaric labeling in a nanodroplet sample preparation platform. *Analytical chemistry* **2019,** *91* (20), 13119-13127.

8. Dou, M.; Zhu, Y.; Liyu, A.; Liang, Y.; Chen, J.; Piehowski, P. D.; Xu, K.; Zhao, R.; Moore, R. J.; Atkinson, M. A., Nanowell-mediated two-dimensional liquid chromatography enables deep proteome profiling of< 1000 mammalian cells. *Chemical science* **2018,** *9* (34), 6944-6951.

9. Lombard-Banek, C.; Moody, S. A.; Nemes, P., Single-cell mass spectrometry for discovery proteomics: quantifying translational cell heterogeneity in the 16-cell frog (Xenopus) embryo. *Angewandte Chemie* **2016,** *128* (7), 2500-2504.

10. Zhu, Y.; Piehowski, P. D.; Zhao, R.; Chen, J.; Shen, Y.; Moore, R. J.; Shukla, A. K.; Petyuk, V. A.; Campbell-Thompson, M.; Mathews, C. E., Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nature communications* **2018,** *9* (1), 882.

11. Williams, S. M.; Liyu, A. V.; Tsai, C.-F.; Moore, R. J.; Orton, D. J.; Chrisler, W. B.; Gaffrey, M. J.; Liu, T.; Smith, R. D.; Kelly, R. T., Automated coupling of nanodroplet sample preparation with

liquid chromatography–mass spectrometry for high-throughput single-cell proteomics. *Analytical chemistry* **2020,** *92* (15), 10588-10596.

12. Budnik, B.; Levy, E.; Harmange, G.; Slavov, N., SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome biology* **2018,** *19*, 1-12.

13. Brunner, A. D.; Thielert, M.; Vasilopoulou, C.; Ammar, C.; Coscia, F.; Mund, A.; Hoerning, O. B.; Bache, N.; Apalategui, A.; Lubeck, M., Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Molecular systems biology* **2022,** *18* (3), e10798.

14. Rosenberger, F. A.; Thielert, M.; Mann, M., Making single-cell proteomics biologically relevant. *Nature Methods* **2023,** *20* (3), 320-323.

15. Mund, A.; Coscia, F.; Kriston, A.; Hollandi, R.; Kovács, F.; Brunner, A.-D.; Migh, E.; Schweizer, L.; Santos, A.; Bzorek, M., Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nature Biotechnology* **2022,** *40* (8), 1231-1240.

16. Cupp-Sutton, K. A.; Fang, M.; Wu, S., Separation methods in single-cell proteomics: RPLC or CE? *International Journal of Mass Spectrometry* **2022,** *481*, 116920.

17. Gregorich, Z. R.; Ge, Y., Top-down proteomics in health and disease: Challenges and opportunities. *PROTEOMICS* **2014,** *14* (10), 1195-1210.

18. Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W., Top down versus bottom up protein characterization by tandem high-resolution mass spectrometry. *Journal of the American Chemical Society* **1999,** *121* (4), 806-812.

19. Smith, L. M.; Kelleher, N. L., Proteoform: a single term describing protein complexity. *Nature methods* **2013,** *10* (3), 186-187.

20. Smith, L. M.; Agar, J. N.; Chamot-Rooke, J.; Danis, P. O.; Ge, Y.; Loo, J. A.; Paša-Tolić, L.; Tsybin, Y. O.; Kelleher, N. L.; Proteomics, C. f. T.-D., The human proteoform project: defining the human proteome. *Science advances* **2021,** *7* (46), eabk0734.

21. Zhou, M.; Uwugiaren, N.; Williams, S. M.; Moore, R. J.; Zhao, R.; Goodlett, D.; Dapic, I.; Pasa-Tolic, L.; Zhu, Y., Sensitive top-down proteomics analysis of a low number of mammalian cells using a nanodroplet sample processing platform. *Analytical chemistry* **2020,** *92* (10), 7087-7095.

22. Su, P.; Hollas, M. A.; Rubakhin, S.; Butun, F. A.; Greer, J. B.; Early, B. P.; Fellers, R. T.; Caldwell, M. A.; Sweedler, J. V.; Kafader, J. O., Top-down Proteomics of 10,000 Single Brain Cells. *bioRxiv* **2023**, 2023.05. 31.543176.

23. Toby, T. K.; Fornelli, L.; Srzentić, K.; DeHart, C. J.; Levitsky, J.; Friedewald, J.; Kelleher, N. L., A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nature protocols* **2019,** *14* (1), 119-152.

24. Johnson, K. R.; Gao, Y.; Gregus, M.; Ivanov, A. R., On-capillary cell lysis enables top-down proteomic analysis of single mammalian cells by CE-MS/MS. *Analytical chemistry* **2022,** *94* (41), 14358-14367.

25. Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M., Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **2011,** *480* (7376), 254-258.

26. Anderson, N. L.; Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Molecular & cellular proteomics* **2002,** *1* (11), 845-867.

27. Melby, J. A.; Roberts, D. S.; Larson, E. J.; Brown, K. A.; Bayne, E. F.; Jin, S.; Ge, Y., Novel strategies to address the challenges in top-down proteomics. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (6), 1278-1294.

28. Brown, K. A.; Melby, J. A.; Roberts, D. S.; Ge, Y., Top-down proteomics: challenges, innovations, and applications in basic and clinical research. *Expert review of proteomics* **2020,** *17* (10), 719-733.

29. Doucette, A. A.; Tran, J. C.; Wall, M. J.; Fitzsimmons, S., Intact proteome fractionation strategies compatible with mass spectrometry. *Expert review of proteomics* **2011,** *8* (6), 787-800.

30. Zhou, W.; Zhang, B.; Liu, Y.; Wang, C.; Sun, W.; Li, W.; Chen, Z., Advances in capillary electrophoresis-mass spectrometry for cell analysis. *TrAC Trends in Analytical Chemistry* **2019,** *117*, 316-330.

31. Huang, L.; Fang, M.; Cupp-Sutton, K. A.; Wang, Z.; Smith, K.; Wu, S., Spray-Capillary-Based Capillary Electrophoresis Mass Spectrometry for Metabolite Analysis in Single Cells. *Analytical Chemistry* **2021,** *93* (10), 4479-4487.

32. Wehr, T.; Rodriguez-Diaz, R.; Zhu, M., *Capillary electrophoresis of proteins*. CRC Press: 1998.

33. McCool, E. N.; Lubeckyj, R. A.; Shen, X.; Chen, D.; Kou, Q.; Liu, X.; Sun, L., Deep top-down proteomics using capillary zone electrophoresis-tandem mass spectrometry: identification of 5700 proteoforms from the Escherichia coli proteome. *Analytical chemistry* **2018,** *90* (9), 5529-5533.

34. Hamidli, N.; Pajaziti, B.; Andrási, M.; Nagy, C.; Gáspár, A., Determination of human insulin and its six therapeutic analogues by capillary electrophoresis–mass spectrometry. *Journal of Chromatography A* **2022,** *1678*, 463351.

35. Sun, B.; Kumar, S., Protein Adsorption Loss─The Bottleneck of Single-Cell Proteomics. *Journal of Proteome Research* **2022,** *21* (8), 1808-1815.

36. Han, X.; Wang, Y.; Aslanian, A.; Fonslow, B.; Graczyk, B.; Davis, T. N.; Yates III, J. R., In-line separation by capillary electrophoresis prior to analysis by top-down mass spectrometry enables sensitive characterization of protein complexes. *Journal of proteome research* **2014,** *13* (12), 6078-6086.

37. Xu, T.; Sun, L., A mini review on capillary isoelectric focusing-mass spectrometry for top-down proteomics. *Frontiers in Chemistry* **2021,** *9*, 651757.

38. Standke, S. J.; Colby, D. H.; Bensen, R. C.; Burgett, A. W.; Yang, Z., Mass spectrometry measurement of single suspended cells using a combined cell manipulation system and a single-probe device. *Analytical chemistry* **2019,** *91* (3), 1738-1742.

39. Li, J.; Huang, L.; Guo, Y.; Cupp-Sutton, K. A.; Wu, S., An automated spray-capillary platform for the microsampling and CE-MS analysis of picoliter-and nanoliter-volume samples. *Analytical and Bioanalytical Chemistry* **2023**, 1-13.

40. Huang, L.; Wang, Z.; Cupp-Sutton, K. A.; Smith, K.; Wu, S., Spray-Capillary: An Electrospray-Assisted Device for Quantitative Ultralow-Volume Sample Handling. *Analytical Chemistry* **2020,** *92* (1), 640-646.

41. 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*, 839-843.

42. Horvath, J.; Dolník, V., Polymer wall coatings for capillary electrophoresis. *Electrophoresis* **2001,** *22* (4), 644-655.

43. Cong, Y.; Liang, Y.; Motamedchaboki, K.; Huguet, R.; Truong, T.; Zhao, R.; Shen, Y.; Lopez-Ferrer, D.; Zhu, Y.; Kelly, R. T., Improved single-cell proteome coverage using narrow-bore packed NanoLC columns and ultrasensitive mass spectrometry. *Analytical chemistry* **2020,** *92* (3), 2665-2671.

44. Albin, M.; Grossman, P. D.; Moring, S. E., Sensitivity enhancement for capillary electrophoresis. *Analytical Chemistry* **1993,** *65* (10), 489A-497A.

45. Zhao, Y.; Sun, L.; Champion, M. M.; Knierman, M. D.; Dovichi, N. J., Capillary zone electrophoresis–electrospray ionization-tandem mass spectrometry for top-down characterization of the Mycobacterium marinum secretome. *Analytical chemistry* **2014,** *86* (10), 4873-4878.

46. Zhu, G.; Sun, L.; Yan, X.; Dovichi, N. J., Single-shot proteomics using capillary zone electrophoresis–electrospray ionization-tandem mass spectrometry with production of more than 1 250 Escherichia coli peptide identifications in a 50 min separation. *Analytical chemistry* **2013,** *85* (5), 2569-2573.

47. Bock, H.-H., Clustering methods: a history of k-means algorithms. *Selected contributions in data analysis and classification* **2007**, 161-172.

48. Lloyd, S., Least squares quantization in PCM. *IEEE transactions on information theory* **1982,** *28* (2), 129-137.

49. MacQueen, J. In *Some methods for classification and analysis of multivariate observations*, Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA: 1967; pp 281-297.

50. Hamidli, N.; Andrasi, M.; Nagy, C.; Gáspár, A., Analysis of intact proteins with capillary zone electrophoresis coupled to mass spectromery using uncoated and coated capillaries. *Journal of Chromatography A* **2021,** *1654*, 462448.

51. Luo, R. Y.; Wong, C.; Xia, J. Q.; Glader, B. E.; Shi, R.-Z.; Zehnder, J. L., Neutral-Coating Capillary Electrophoresis Coupled with High-Resolution Mass Spectrometry for Top-Down Identification of Hemoglobin Variants. *Clinical Chemistry* **2023,** *69* (1), 56-67.

52. Shen, Y.; Smith, R. D., High-resolution capillary isoelectric focusing of proteins using highly hydrophilic-substituted cellulose-coated capillaries. *Journal of Microcolumn Separations* **2000,** *12* (3), 135-141.

53. VanOrman, B. B.; Liversidge, G. G.; McIntire, G. L.; Olefirowicz, T. M.; Ewing, A. G., Effects of buffer composition on electroosmotic flow in capillary electrophoresis. *Journal of Microcolumn Separations* **1990,** *2* (4), 176-180.

54. Britz-McKibbin, P.; Bebault, G. M.; Chen, D. D., Velocity-difference induced focusing of nucleotides in capillary electrophoresis with a dynamic pH junction. *Analytical chemistry* **2000,** *72* (8), 1729-1735.

55. Guo, Y.; Yu, D.; Cupp-Sutton, K. A.; Liu, X.; Wu, S., Optimization of protein-level tandem mass tag (TMT) labeling conditions in complex samples with top-down proteomics. *Analytica chimica acta* **2022,** *1221*, 340037.

56. Yu, D.; Wang, Z.; Cupp-Sutton, K. A.; Guo, Y.; Kou, Q.; Smith, K.; Liu, X.; Wu, S., Quantitative top-down proteomics in complex samples using protein-level tandem mass tag labeling. *Journal of the American Society for Mass Spectrometry* **2021,** *32* (6), 1336-1344.

57. Wang, Z.; Yu, D.; Cupp-Sutton, K. A.; Liu, X.; Smith, K.; Wu, S., Development of an online 2D ultrahigh-pressure nano-LC system for high-pH and low-pH reversed phase separation in top-down proteomics. *Analytical chemistry* **2020,** *92* (19), 12774-12777.

58. Wang, Z.; Ma, H.; Smith, K.; Wu, S., Two-dimensional separation using high-pH and low-pH reversed phase liquid chromatography for top-down proteomics. *International journal of mass spectrometry* **2018,** *427*, 43-51.

59. Guo, Y.; Cupp-Sutton, K. A.; Zhao, Z.; Anjum, S.; Wu, S., Multidimensional separations in top–down proteomics. *Analytical Science Advances* **2023**.