An improved nanoflow RPLC-CZE-MS/MS system with high peak capacity and sensitivity for nanogram bottom-up proteomics

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

Abstract

Novel mass spectrometry (MS)-based proteomic tools with extremely high sensitivity and high peak capacity are required for comprehensive characterization of protein molecules in mass-limited samples. We reported a nanoRPLC-CZE-MS/MS system for deep bottom-up proteomics of low micrograms of human cell samples in previous work (Yang *et al.* Anal. Chem. 2018, 90, 10479-10486). In this work, we improved the sensitivity of the nanoRPLC-CZE-MS/MS system drastically via employing bovine serum albumin (BSA) treated sample vials, improving the nanoRPLC fraction collection procedure, and using a short capillary for fast CZE separation. The improved nanoRPLC-CZE produced a peak capacity of 8500 for peptide separation. The improved system identified 6500 proteins from a MCF7 proteome digest starting with only 500-ng peptides using a Q-Exactive HF mass spectrometer. The system produced a comparable number of protein identifications (IDs) to our previous system and the two-dimensional (2D) nanoRPLC-MS/MS system developed by the Mann's group with 10-fold and 4-fold less sample consumption, respectively. We coupled single spot solid phase sample preparation (SP3) method to the improved nanoRPLC-CZE-MS/MS for bottom-up proteomics of 5000 HEK293T cells, resulting in 3689 protein IDs with the consumption of a peptide amount that corresponded to only roughly 1000 cells.

Keywords: bottom-up proteomics, mass-limited sample, CZE-MS, LC-CZE, single spot solid phase sample preparation (SP3), high peak capacity, high sensitivity

Introduction

Multi-dimensional separation before electrospray ionization-mass spectrometry (ESI-MS) is indispensable for bottom-up proteomics of complex proteomes. [1-3] Reversed-phase liquid chromatography (RPLC) and capillary zone electrophoresis (CZE) are both capable for high-resolution separation of peptides and are orthogonal. [4,5] Coupling RPLC to CZE-MS for complex sample analysis has been a very active research area in the literature. [6] Both offline and online RPLC-CZE-MS systems have been developed and applied in bottom-up proteomics. In general, the online RPLC-CZE-MS should produce better sensitivity and higher throughput than the offline version due to the reduced sample loss and labor; the offline RPLC-CZE-MS should have much simpler system setup and provide better flexibility for the two-dimensional (2D) separations compared to the online version. [2] It was well demonstrated that CZE-MS outperformed RPLC-MS for analysis of mass-limited proteome samples regarding the number of protein identifications (IDs), suggesting the better sensitivity of CZE-MS compared to RPLC-MS. [7-10] Therefore, CZE is typically online coupled to MS and RPLC is employed as the first dimension for peptide fractionation.

Several research groups have developed online RPLC-CZE-MS systems for proteomics. [11-17] The Ramsey group carried out the connection of RPLC and CZE-MS using a microfluidic device. [15,16] They applied the online RPLC-CZE-MS systems in bottom-up MS characterization of standard protein digests, an antibody, and an *E. coli* proteome sample. Under the optimal condition, a peak capacity of over 1000 in one hour was produced by the online system. Recently, the Neusüß group built an online RPLC-CZE-MS system using a 4-port valve as the interface for characterization of intact proteins. [17] An RPLC eluate containing proteins was selectively transferred into a sample loop integrated into the 4-port valve, followed by CZE-MS analysis. Although some successful examples were published, some challenges remain for the online RPLC-CZE-MS. First, the separation power of RPLC and CZE cannot be fully used because we need to balance the two dimensions for online operation. Second, usually only a small fraction of the RPLC eluate can be analyzed by CZE-MS. For instance, in the design of the Ramsey group, [16] only small fractions of RPLC eluates flowed into the microfluidic device and most of them went to waste through splitting to ensure the compatibility of RPLC and CZE.

Offline RPLC-CZE-MS is a good solution for fully using the separation power of both RPLC and CZE and has been applied for large-scale bottom-up proteomics. [5,18-25] Yan *et al*. coupled offline RPLC fractionation to CZE-MS/MS for bottom-up proteomics of *Xenopus* embryos and identified 4100 proteins. [19] Chen *et al*. developed a SCX-RPLC-CZE-MS/MS platform and achieved 8200 protein IDs in 70 h from a mouse brain proteome. [24] Both studies started with hundreds of micrograms of proteome digests. Choi *et al*. coupled a microscale RPLC fractionation using a C18 zip-tip to CZE-MS/MS for analysis of 1 to 20-μg neuron proteome digests.[23] Nearly 800 proteins were identified with consumption of only 1-ng proteome digest, demonstrating the power of microscale RPLC-CZE-MS/MS for bottom-up proteomics of mass-limited samples. Although the offline RPLC-CZE-MS/MS showed its power for large-scale bottom-up proteomics, only a small fraction of peptides in each RPLC fraction (*i.e.,* less than 5%) was analyzed by CZE-MS/MS in the typical workflow because of the low sample loading capacity of CZE. The offline RPLC-CZE-MS/MS suffered from low overall sensitivity also because the serious sample loss on the analytical RPLC column (2.1-4.6 mm i.d.) used for peptide fractionation and in the Eppendorf tubes used for fraction collection.

The Mann group demonstrated that the RPLC fractionation using a capillary column (*i.e.,* 250-μm i.d.) produced drastically better sensitivity compared to that using an analytical column (*i.e.,* 2.1-mm i.d.) for bottom-up proteomic analyses of low micrograms of human cell proteins, due to the much lower sample loss from the obviously reduced surface area. [26] Recently, we coupled offline nanoRPLC fractionation using a 75-μm-i.d. column at 200 nL/min flow rate to high-capacity CZE-MS/MS for highly sensitive and deep bottom-up proteomics of MCF7 cancer cells. [27] We employed a dynamic pH junction-based sample stacking method [28] to improve the sample loading capacity of CZE and enabled a 500-nL sample injection from a 1.5-μL peptide solution, using up to 33% of the available peptide material for a CZE-MS/MS analysis. The combination of nanoRPLC and dynamic pH junction-based CZE-MS/MS identified 7500 proteins and nearly 60000 peptides starting from only 5-μg of a MCF7 proteome digest. The nanoRPLC-CZE-MS/MS can fully use the separation power of RPLC and CZE and has great overall sensitivity because the RPLC eluates can be efficiently used for CZE-MS/MS analysis.

Built upon our preliminary work, here we present an improved nanoRPLC-CZE-MS/MS with much better sensitivity than our previous system. The improved system enabled deep bottom-up proteomic analysis of nanograms of MCF7 proteome digests with the production of 6500 protein IDs starting with only 500-ng MCF7 peptides using a Q-Exactive HF mass spectrometer. Only roughly 100 ng of peptides were actually consumed. We further applied the improved nanoRPLC-CZE-MS/MS in bottom-up proteomics of 5000 HEK293T cells. The single spot solid phase sample preparation (SP3) method [29,30] was employed for preparing the 5000-cell sample. We identified about 3700 proteins with the consumption of protein content of roughly 1000 cells. The optimized nanoRPLC-CZE-MS/MS system showed high sensitivity for bottom-up proteomics because of several novelties.

First, we applied a short capillary (*i.e.,* 70-cm long) in CZE separation and the sample loading volume was 500 nL, corresponding to 36% of the total capillary volume. One CZE-MS/MS run was completed in an hour with an average peak capacity of nearly 200. The peptides migrated drastically faster compared to our previous work with a 1-meter-long capillary, [27] resulting in sharper peaks and higher peptide intensity, which are important for analysis of mass-limited samples.

Second, we pretreated the inner wall of sample vials of CZE with bovine serum albumin (BSA) to reduce the nonspecific peptide adsorption. The BSA treatment improved the peptide intensity by nearly 200%.

Third, lyophilization and redissolution of nanoRPLC eluates before CZE-MS/MS analysis were avoided to reduce sample loss, to decrease labor, and to improve throughput. The nanoRPLC eluates were directly collected into 0.6-mL low-retention Eppendorf tubes containing 1.4-2.4 µL of an ammonium bicarbonate (NH_4HCO_3) buffer (pH 8.5). Then the samples were transferred into the BSA-treated sample vials of CZE for the dynamic pH junction-based CZE-MS/MS without any sample preparation steps.

Fourth, we coupled the SP3-based sample preparation method with nanoRPLC-CZE-MS/MS for bottom-up proteomics of 5000 HEK293T cells. The SP3 method has high sample recovery for preparation of mass-limited proteome samples based on the literature. [29,30] To our best knowledge, this is the first report of bottom-up proteomics of a small number of human cells using CZE-MS/MS.

Experimental Section

"Materials and reagents", "MCF7 cell culture and proteome digestion", "high-pH nanoRPLC fractionation", and "nanoRPLC-MS/MS" are described in **Supporting Information I**.

HEK293T cell culture and **SP3-based sample preparation**

HEK293T cells were harvested, washed with PBS buffer and were immediately subjected to flow cytometry for cell sorting (BD Influx, BD Bioscience). An aliquot of 5000 cells was collected in a 1.7-mL Eppendorf tube, followed by sample preparation with the SP3 method based on references [29] and [30]. The detailed procedure is described in **Supporting Information I**.

NanoRPLC fractionation

EASY-nLC 1200 (Thermo Fisher Scientific) equipped with a capillary column (75 µm i.d. x 50 cm Length, C18, 2 µm bead, 100 Å pore, Thermo Fisher Scientific) was used for fractionation. Mobile phase A contained 2% (v/v) acetonitrile (ACN), 98% (v/v) H_2O and 0.1% (v/v) formic acid (FA). Mobile phase B contained 80% (v/v) ACN, 20% (v/v) H_2O and 0.1 % (v/v) FA. Flow rate was 200 nL/min. One-column set up was selected on the nanoRPLC system. 0.6-mL low-retention Eppendorf tube (Denville, Canada) was used for fraction collection.

50-fraction procedure. 5 µg and 500 ng of MCF7 proteome digests were fractionated into 50 fractions with a 3-h gradient using nanoRPLC, respectively. The gradient was set up as follows: from 8% to 30% (v/v) B in 100 min, from 30% to 50% (v/v) B in 50 min, from 50% to 80% (v/v) B in 15 min and stay at 80% (v/v) B for 15 min. The first fraction collection started at the sample loading and traversed the first 20 min of gradient. Then each fraction was collected every 3 min. The last fraction was collected from the 164 min to the end of the gradient. For 5-µg sample fractionation, 2.4 µL of 50 mM NH_4HCO_3 (pH 8.5) was deposited to the bottom of the 0.6-mL Eppendorf tube. For all fractions except the first and the last ones, when 0.6 µL of eluate (200 nL/min for 3 min) flowed into the buffer, the final pH was around 8.0. For 500-ng sample fractionation, the same procedure was used but only 1.4 µL of the NH_4HCO_3 buffer (pH 8.5) was deposited into the tube, resulting in a total sample volume of 2 µL for each fraction. For the first and last nanoRPLC fractions, the eluates were lyophilized and redissolved in 3 µL (for the 5-µg sample) and 2 µL (for the 500-ng sample) of a 50 mM NH_4HCO_3 buffer (pH 8.0) for CZE-MS/MS.

20-fraction procedure. A 500-ng MCF7 cell proteome digest was fractionated into 20 fractions with a 100-min gradient using the similar procedure to the 50-fraction collection described previously. For each nanoRPLC fraction, the final sample volume was 2.2 µL and the pH of the sample solution was about 8.0. The detailed description is shown in **Supporting Information I.**

10-fraction procedure. The proteome digest from 5000 HEK293T cells was fractionated by the nanoRPLC system into 10 fractions using a 60-min gradient with a similar procedure to the 50- and 20-fraction collections. For each nanoRPLC fraction, the final sample volume was 2.2 µL and the pH of the sample solution was about 8.0. The detailed description is shown in **Supporting Information I**.

Pretreatment of sample vials of CZE-MS and nanoRPLC-MS with BSA

The sample injection vials of CZE-MS and nanoRPLC-MS were treated with a BSA solution to reduce non-specific adsorption of peptides on the inner wall of vials. 10 µL of 2 mg/mL BSA solution was added into each sample injection vial and was incubated at room temperature for 10 min. After the BSA solution was removed, each vial was rinsed with 500 µL of a NH_4HCO_3 buffer (10 mM, pH 8) twice, followed by air dry in the chemical hood. The treated sample vials were ready for use. We need to note that the BSA treated sample vials were only used for the 500-ng MCF7 proteome samples and the 5000 HEK293T cell sample.

CZE-MS/MS

A detailed description of the CZE-MS/MS is given in the **supporting information I**. Briefly, a 70-cm or 100-cm linear polyacrylamide (LPA) coated capillary (50/360 µm i.d./o.d.) with an etched end by hydrofluoric acid was used for CZE separation. [31,32] The commercialized electrokinetically pumped sheath flow CE-MS interface (CMP scientific, Brooklyn, NY) was used to couple CZE to MS. [33, 34] For all CZE-MS/MS analyses, a 500-nL aliquot of each sample was loaded.

Data Analysis

The detailed information about the data analysis is described in **supporting information I**. Briefly, database searching was performed on both Proteome discoverer 2.2 (Thermo Fisher Scientific) with SEQUEST HT search engine and MaxQuant 1.5.5.1. [35,36] All the numbers of protein IDs reported in this study were from the Proteome Discoverer. The identified proteins were listed in the **Supporting information II**. The MaxQuant data was used for comparing the results from different experiments with label free quantification (LFQ).

Results and discussion

Comparing 100-cm-long and 70-cm-long capillaries for CZE-MS/MS

Longitudinal diffusion is usually the only factor causes band broadening in CZE. The longer time analytes spend in the capillary, the more band broadening would be. In our previous work, we used a 100-cm-long LPA-coated capillary for CZE separation. [27] The LPA coating significantly reduced the electroosmotic flow in the capillary, which is crucial for effective dynamic pH junction-based sample stacking and a wide separation window. In a typical run of dynamic pH junction-based CZE-MS/MS using a 100-cm-long LPA-coated capillary, we identified 8172 peptides and 1896 proteins from a MCF7 proteome digest with a Q-Exactive HF mass spectrometer in 95 min. 100-ng peptides in 500 nL of 50 mM NH_4HCO_3 (pH 8.0) were injected for analysis. The effective separation window was about 1 h, **Figure 1A**. The peak capacity was roughly 300 based on the full width at half maximum (FWHM) of peptides. We performed peak extraction at m/z 593.33 with a mass tolerance of 10 ppm and observed three peaks corresponding to three different peptides at varied migration time, **Figure 1B**. The peptide peak became obviously wider with the increase of migration time, most likely due to more significant diffusion in the capillary. Interestingly, the basic peptides clearly tended to migrate faster than the acidic peptides in the dynamic pH junction-based CZE, **Figure 1C**.

We believe that by reducing the migration time of peptides, sharper peptide peaks and higher peptide intensity could be achieved, which is important for CZE-MS/MS analysis of mass-limited samples. To reduce peptide migration time, one option is to use a shorter capillary. However, a shorter capillary can narrow the separation window, reduce the peak capacity, and decrease the number of MS/MS spectra from one CZE run. To ensure the overall peak capacity and number of MS/MS spectra of the nanoRPLC-CZE-MS/MS system, we need to collect a higher number of fractions for CZE-MS/MS analysis.

We first tested our idea by analyzing a 5-μg MCF7 proteome digest with nanoRPLC-CZE-MS/MS, in which a 70-cm-long LPA-coated capillary was employed for CZE-MS/MS analyses of 50 nanoRPLC fractions (70-cm-50-fractions). We also compared the data with our previous work that utilized a 100-

cm-long LPA-coated capillary for CZE-MS/MS analyses of 20 nanoRPLC fractions (100-cm-20-fractions). [27]

For the 70-cm-50-fractions study, we loaded 5 μg of a MCF7 proteome digest onto a 50-cm-long nanoRPLC column for fractionation over a 3-h gradient as we did in our 100-cm-20-fractions study in reference [27]. To reduce additional sample handling steps, here we put 2.4 μL of 50 mM NH_4HCO_3 buffer (pH 8.5) in each sample collection tube. When 0.6 μL of eluate (0.2 μL/min flow rate for 3 min) was combined with the NH_4HCO_3 buffer, the final pH was around 8 and final volume was 3 μL. 50 fractions were collected. Each fraction was then directly analyzed by the dynamic pH junction-based CZE-MS/MS with a 70-cm-long LPA-coated capillary. For each fraction, 500 nL out of 3 μL sample was injected. The MS analysis time per fraction was 55 min. In total, we identified 7546 proteins and 66990 peptides from the 5-μg MCF7 proteome digest. The 70-cm-50-fractions system produced similar numbers of protein IDs and peptide IDs to the 100-cm-20-fractions system in reference [27] (7546 *vs.* 7512 proteins, 66990 *vs.* 59403 peptides) but with 50% less peptide material consumption (0.8 μg *vs.* 1.6 μg). The result indicates that the 70-cm-50-fractions system has better sensitivity than the 100-cm-20-fractions platform. Less peptide material consumption also saves more material for additional analysis.

We investigated the protein-level and peptide-level overlaps between the data in the 70-cm-50-fractions and 100-cm-20-fractions studies, **Figure 2A**. The two studies shared over 70% of protein IDs and only 26% of the peptide IDs. Considering the low peptide-level overlap, we further investigated the physicochemical properties of peptides identified in the two studies, **Figures 2B-2D**. The 70-cm-50-fractions system clearly tended to identify basic peptides compared to the 100-cm-20-fractions system, **Figure 2B**. The two systems showed no drastic differences in peptide molecular weight (MW) and GRAVY value, **Figures 2C** and **2D**.

NanoRPLC-CZE-MS/MS for bottom-up proteomic analysis of 500-ng MCF7 proteome digests

The 70-cm-50-fractions system has better sensitivity than our previous 100-cm-20-fractions platform based on our data. Therefore, we further applied the 70-cm-50-fractions method in the analysis of a 500-ng MCF7 proteome digest.

We made some modifications on the nanoRPLC fraction collection and CZE-MS sample vial compared to the 5-μg-sample study to reduce the sample loss further. First, we only put 1.4-μL NH_4HCO_3 buffer in the Eppendorf tube for nanoRPLC fraction collection and we got 2-μL sample with pH about 8 in each Eppendorf tube after collection of 0.6-μL eluate from nanoRPLC. Second, we pretreated the sample vial used for CZE-MS/MS with a BSA solution to block nonspecific adsorption of peptides on the inner wall of the sample vial.

25% of peptides in each nanoRPLC fraction (500 nL out of the 2 μL sample) was loaded for CZE-MS/MS analysis, corresponding to consumption of only roughly 100-ng MCF7 proteome digest in total. We identified 6492 proteins and 47342 peptides using the 70-cm-50-fractions system. The total MS time was 45 h. The MS raw data have been deposited to the ProteomeXchange Consortium via the PRIDE [37] partner repository with the dataset identifier PXD014392.

Recently, the Mann group developed a highly sensitive 2D nanoRPLC-MS/MS system that employed a microscale capillary column for high-pH RPLC fractionation followed by low-pH nanoRPLC-MS/MS analysis. [26] The system identified 5724 proteins and 23765 peptides by MS/MS in 16 h using a Q-Exactive HF mass spectrometer with the consumption of 500-ng HeLa proteome digest. [26] Our improved nanoRPLC-CZE-MS/MS system identified 13% and nearly 100% more protein and peptide IDs than the 2D nanoRPLC-MS/MS system with four-times lower sample consumption (125 ng *vs.* 500 ng). We noted that our system consumed much longer MS time than the 2D nanoRPLC-MS/MS system (45 h vs. 16 h). We also tried to collect 20 nanoRPLC fractions for CZE-MS/MS to reduce the total MS analysis time. 4551 proteins and 24559 peptides were identified by nanoRPLC-CZE-MS/MS in 18 h with the consumption of about 100-ng peptides in total. The number of protein IDs is about 20% lower than that in reference [26] with comparable instrument time but 4-fold lower sample consumption.

To make a fair comparison between our nanoRPLC-CZE-MS/MS and the 2D-nanoRPLC-MS/MS for analysis of mass-limited proteome samples, we also analyzed a 500-ng MCF7 proteome digest using 2D-nanoRPLC-MS/MS with the same mass spectrometer. We fractionated the 500-ng digest into 36 fractions using high-pH nanoRPLC and combined them into 18 fractions. Each fraction was analyzed by low-pH nanoRPLC-MS/MS in a 2-h gradient. Considering the sample loading time of nanoRPLC-MS/MS, the total MS time for the 2D-nanoRPLC-MS/MS was about 45 h. The 2D-nanoRPLC-MS/MS identified 4758 proteins and 23589 peptides. Using the same mass spectrometer and instrument time, our nanoRPLC-CZE-MS/MS system identified 36% more proteins (6492 *vs.* 4758) and 100% more peptides (47342 *vs.* 23589) than the 2D-nanoRPLC-MS/MS from 500-ng MCF7 proteome digests. We noted that our nanoRPLC-CZE-MS/MS consumed 3-times lower (125 ng *vs.* 375 ng) peptides compared to the 2D-nanoRPLC-MS/MS. About 65% and 88% of the identified peptides and proteins from 2D-nanoRPLC-MS/MS were covered by the data from nanoRPLC-CZE-MS/MS, **Figure S1**.

The nanoRPLC-CZE-MS/MS system in this work showed high sensitivity due to several reasons. First, the system produced extremely high peak capacity for peptide separation. On average, each CZE run had a peak capacity of 170 in 55 min. The nanoRPLC-CZE with 50 fractions produced a peak capacity of 8500 for peptide separation. The extremely high peak capacity was most likely because nanoRPLC and CZE were well orthogonal for peptide separation. **Figure 3** shows the base peak chromatogram of nanoRPLC-MS/MS analysis of a 500-ng MCF7 proteome digest as well as base peak electropherograms of 20 out of 50 nanoRPLC fractions of the 500-ng digest after CZE-MS/MS analyses. It is clear that one 3-min nanoRPLC fraction can be further separated into an up to 40-min window by CZE.

Second, the pretreatment of CZE-MS sample vial with BSA drastically reduced sample loss. We compared the number of protein IDs, the number of peptide IDs, protein intensity, and peptide intensity between nanoRPLC-CZE-MS/MS studies with and without sample vial pretreatment. Twenty nanoRPLC fractions were collected for CZE-MS/MS analysis in both cases. The non-BSA treated study resulted in 39% and 58% lower number of protein (2758 vs. 4551) and peptide (10230 vs. 24559) IDs compared to the BSA-treated study. We also compared the LFQ intensity of identified proteins in both studies, **Figure 4A**. The protein LFQ intensities had good correlation between the two conditions (R^2 = 0.93) with a slope of 4.5, indicating drastically higher protein intensity with BSA-treated sample vials compared to non-treated vials. The median of peptide intensities from the BSA-treated sample vial was 2.6-fold higher than that from the non-treated sample vial, **Figure 4B**.

Third, the 70-cm-50fractions system improved the sensitivity of nanoRPLC-CZE-MS/MS obviously compared to the 100-cm-20fractions system for analysis of 500-ng MCF7 proteome digest. We fractionated 500 ng of a MCF7 proteome digest into 20 fractions with nanoRPLC and analyzed each fraction with CZE-MS/MS using a 100-cm-long capillary (100-cm-20fractions). The CZE sample vials were treated with BSA. Each CZE-MS/MS run took 120 min, and the total MS time was 40 h. The 100-cm-20fractions system identified 4723 proteins and 24975 peptides from the 500-ng MCF7 proteome digest. The number of protein and peptide IDs was 27% and 47% lower compared to that from the 70-cm-50fractions system with comparable MS time (40 vs. 45 h). The protein LFQ intensity showed good correlation (R^2 = 0.93) between the two systems with a slope of 2.3, indicating that higher protein LFQ intensities were obtained with the 70-cm-50fractions system, **Figure 4C**.

We need to note that the number of protein and peptide IDs using our nanoRPLC-CZE-MS/MS system can be boosted drastically with a more advanced mass spectrometer. In this work, a Q-Exactive HF mass spectrometer was employed and we have shown that single-shot CZE-MS/MS analysis of a

MCF7 proteome digest identified nearly 2000 proteins and 8000 peptides. Over 4000 protein IDs and 27000 peptide IDs have been achieved by single-shot CZE-MS/MS using an Orbitrap Fusion Lumos mass spectrometer with an advanced-peak-determination algorithm. [38]

We also evaluated the reproducibility of our improved nanoRPLC-CZE-MS/MS system. We fractionated a 500-ng MCF7 proteome digest into 20 fractions followed by CZE-MS/MS analysis with a 70-cm-long capillary. We performed the experiment in duplicate. The system identified 4508±60 proteins and 25034±671 peptides. The duplicate analyses shared about 70% of the identified proteins. The system was highly reproducible regarding the LFQ protein intensity, **Figure 4D**. The medians of peptide intensities from the duplicate analyses were highly consistent (6.8E6 *vs.* 6.5E6). The data here clearly indicate that the nanoRPLC-CZE-MS/MS is qualitatively and quantitatively reproducible.

Bottom-up proteomics of 5000 HEK293T cells

To further investigate the capability of our nanoRPLC-CZE-MS/MS system for bottom-up proteomic analysis of mass-limited samples, we collected 5000 HEK293T cells into an Eppendorf tube using flow cytometry and prepared the sample with the SP3 method [29] followed by nanoRPLC-CZE-MS/MS analysis. The SP3 method has been well characterized for preparation of low-µg and even sub-µg of complex proteome samples. [29,39] If we assume each HEK293T cell contains 0.1 ng of proteins, [40] 5000 cells contain 500-ng proteins in total. 16% of the peptides (500 nL out of 3 µL) corresponding to 800 HEK293T cells were first analyzed by single-shot CZE-MS/MS (Q-Exactive HF mass spectrometer). The single-shot analysis identified 1263 proteins and 5090 peptides in 2-h MS time. We then fractionated the rest of the sample using nanoRPLC into 10 fractions. Each fraction was analyzed by CZE-MS/MS using the 70-cm-long capillary in 65 min. We identified 3689 proteins in 11-h MS time using this approach with the consumption of a peptide amount corresponding to only roughly 1000 cells. We plotted the cumulative number of protein IDs versus the number of nanoRPLC fractions, **Figure S2**. We identified 2900 proteins from the first five nanoRPLC fractions using about 5-h MS time with the consumption of 12% of the total peptides corresponding to only 500 cells.

We compared our data with that in the literature about bottom-up proteomic analyses of hundreds to thousands of human cells. Wang *et al.* identified 600 proteins from 5000 MCF-7 breast cancer cells using a highly efficient NP-40-based sample preparation method and nanoRPLC-MS/MS (Q-TOF mass spectrometer) in about 4-h MS time. [41] Wiśniewski *et al.* identified 1536 and 2055 proteins from 1000 and 3000 HeLa cells using the improved filter aided sample preparation method and nanoRPLC-MS/MS (Orbitrap Velos mass spectrometer) in 4-h MS time. [42] Li *et al.* developed a highly sensitive bottom-up proteomic procedure for analysis of small numbers of human cancer cells by employing a single-tube AFA (adaptive focused acoustics)-based sample preparation method and PLOT (porous layer open tube)-nanoRPLC-MS/MS (Q-Exactive mass spectrometer). [43] A sample containing 2000 MCF7 cells was processed and a peptide aliquot corresponding to 500 cells was analyzed by nanoRPLC-MS/MS in 4-h MS time, leading to the identification of 3370 proteins. Chen *et al.* developed an integrated spintip-based sample preparation method for processing mass-limited proteome samples with high recovery, and 1270 proteins were identified from 2000 HEK 293T cells using nanoRPLC-MS/MS (Orbitrap Fusion mass spectrometer) in 1.4-h MS time. [44] Our SP3-nanoRPLC-CZE-MS/MS system achieved comparable performance compared to that developed by Li *et al.* and Chen *et al.* [43,44] regarding the number of protein IDs, MS time, and the number of human cells. The results clearly indicate that the SP3-nanoRPLC-CZE-MS/MS is an alternative bottom-up proteomic system for deep bottom-up proteomic analysis of mass-limited samples. Because CZE-MS/MS and nanoRPLC-MS/MS are complementary for protein and peptide IDs, [5-9,24,25,27] and because we only used 25% of the peptide sample in each nanoRPLC fraction for CZE-MS/MS analysis, the leftover peptide material in each nanoRPLC fraction can be further analyzed by nanoRPLC-MS/MS to boost the total number of protein and peptide IDs.

We further compared our 5000-cell data with a comprehensive bottom-up proteomic study of the same cell line published by Bekker-Jensen *et al.* in 2017. [45] In that study, 1 mg of HEK 293T cell proteome digest was analyzed by high-pH RPLC fractionation followed by nanoRPLC-MS/MS with the identification of 9246 unique protein-coding genes. Our 5000-cell study identified 3629 protein-coding genes and only 77% of them were also identified in the Bekker-Jensen's work, suggesting good complementarity between these two platforms for protein IDs. We then compared proteins identified in the 5000-cell study to a human transcription factor list that contains over 1600 transcription factors. [46] Our result covered 121 human transcription factors. It has been estimated that in mammalian cells, transcription factors have copy numbers per cell in the range of 10,000-300,000. [47] In the 5000-cell study, the peptides from the 5000 cells were eventually dissolved in 3 μL of solution before nanoRPLC-CZE-MS/MS analysis. The transcription factor concentration in the 3-μL solution should be in the range of 1-50 pM, indicating high sensitivity of our system.

The proteomic sample preparation of small numbers of human cells has been drastically improved recently. Zhu *et al.* developed nanoPOTS (nanodroplet processing in one pot for trace samples) sample preparation method that enabled efficient preparation of 1-100 human cells for bottom-up proteomic analysis. [48,49] They achieved the identification of 3000 proteins from 140 HeLa cells by coupling the nanoPOTS method with nanoRPLC-MS/MS (Orbitrap Fusion Lumos mass spectrometer). [48] The Zhang group developed novel integrated devices for proteomic preparation of 100 cells and even single cells with high sample recovery. [50,51] Over 300 proteins were identified from single HeLa cells via coupling the sample preparation method to nanoRPLC-MS/MS (Orbitrap Fusion mass spectrometer). [50] We expect that coupling these advanced sample preparation methods with CZE-MS/MS will further improve the proteomic characterization of single human cells because it has been well demonstrated that the advanced CZE-MS/MS outperformed nanoRPLC-MS/MS for mass-limited proteomic sample analysis regarding the number of protein IDs. [7-10]

Conclusions

In this work, we presented an improved nanoRPLC-CZE-MS/MS system with high peak capacity and sensitivity for bottom-up proteomics of nanograms of human cell proteome samples. The system produced 6500 protein IDs from only 100-ng MCF7 proteome digest and identified 3700 proteins from 1000 HEK 293T cells. We expect that coupling more advanced sample preparation methods with our improved nanoRPLC-CZE-MS/MS will be a powerful tool for comprehensive bottom-up proteomics of small numbers of human cells.

CZE-MS/MS has been recognized as an alternative approach to LC-MS/MS for top-down proteomics, [52,53] metabolomics, [54] and global characterization of glycans [55]. We believe that our improved nanoRPLC-CZE-MS/MS system will also benefit the scientific communities for global delineation of proteoforms, metabolites, and glycans in mass-limited biological samples.

Acknowledgments

We thank Prof. Xuefei Huang's group and Prof. Jian Hu's group at Michigan State University for kindly providing the MCF7 cells and HEK293T cells for our research. The research was funded by the National Science Foundation (CAREER Award, DBI-1846913). L. Sun was supported by the National Institutes of Health (R01GM125991) and the National Science Foundation (CAREER Award, DBI-1846913).

SUPPLEMENTARY INFORMATION

The following supporting information is available free of charge at ACS website https://urldefense.proofpoint.com/v2/url?u=http-3A__pubs.acs.org&d=DwIFaQ&c=nE__W8dFE-shTxStwXtp0A&r=v2gHFYQw0SHB4b7u7jb0oG39nInD4cKcyJGOwWHiQ8I&m=kWIYjxR504fhqOYSOOgjcFepjEBU5qrCqyxJZm1ynvM&s=QH8EOiRJS91QzFt8Nm9jGPIiYChqxOEQV4wWb5Tvrkk&e=

Supporting Information I.

-Materials and reagents

-MCF7 cell culture and proteome digestion

-HEK293T cell culture and SP3-based sample preparation

-NanoRPLC fractionation

-CZE-MS/MS

-nanoRPLC-MS/MS

-Data Analysis

-Figure S1- The protein-level and peptide-level overlaps between the 2D-nanoRPLC-MS/MS and nanoRPLC-CZE-MS/MS analysis of 500-ng MCF-7 proteome digests.

-Figure S2- Cumulative protein IDs *vs.* number of nanoRPLC fractions from the data of 5000 HEK293T cells. (PDF)

Supporting Information II.

-The identified proteins (XLSX)

References

1. Yuan, H.; Jiang, B.; Zhao, B.; Zhang, L.; Zhang, Y., Recent Advances in Multidimensional Separation for Proteome Analysis. *Analytical chemistry* **2019,** *91* (1), 264-276.
2. Motoyama, A.; Yates, J. R., 3rd, Multidimensional LC separations in shotgun proteomics. *Analytical chemistry* **2008,** *80* (19), 7187-93.
3. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews* **2013,** *113* (4), 2343-94.
4. Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J., The one hour yeast proteome. *Molecular & cellular proteomics : MCP* **2014,** *13* (1), 339-47.
5. Sun, L.; Hebert, A. S.; Yan, X.; Zhao, Y.; Westphall, M. S.; Rush, M. J.; Zhu, G.; Champion, M. M.; Coon, J. J.; Dovichi, N. J., Over 10,000 peptide identifications from the HeLa proteome by using single-shot capillary zone electrophoresis combined with tandem mass spectrometry. *Angewandte Chemie* **2014,** *53* (50), 13931-3.
6. Zhang, Z.; Qu, Y.; Dovichi, N. J., Capillary zone electrophoresis-mass spectrometry for bottom-up proteomics. *TrAC Trends in Analytical Chemistry* **2018,** *108*, 23-37.
7. Wang, Y.; Fonslow, B. R.; Wong, C. C.; Nakorchevsky, A.; Yates, J. R., 3rd, Improving the comprehensiveness and sensitivity of sheathless capillary electrophoresis-tandem mass spectrometry for proteomic analysis. *Analytical chemistry* **2012,** *84* (20), 8505-13.
8. 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 1250 Escherichia coli peptide identifications in a 50 min separation. *Analytical chemistry* **2013,** *85* (5), 2569-73.
9. Ludwig, K. R.; Sun, L.; Zhu, G.; Dovichi, N. J.; Hummon, A. B., Over 2300 phosphorylated peptide identifications with single-shot capillary zone electrophoresis-tandem mass spectrometry in a 100 min separation. *Analytical chemistry* **2015,** *87* (19), 9532-7.
10. 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,** *55* (7), 2454-8.
11. Lewis, K. C.; Opiteck, G. J.; Jorgenson, J. W.; Sheeley, D. M., Comprehensive on-line RPLC-CZE-MS of peptides. *Journal of the American Society for Mass Spectrometry* **1997,** *8* (5), 495-500.
12. Bergstrom, S. K.; Samskog, J.; Markides, K. E., Development of a poly(dimethylsiloxane) interface for on-line capillary column liquid chromatography-capillary electrophoresis coupled to sheathless electrospray ionization time-of-flight mass spectrometry. *Analytical chemistry* **2003,** *75* (20), 5461-7.
13. Bergstrom, S. K.; Dahlin, A. P.; Ramstrom, M.; Andersson, M.; Markides, K. E.; Bergquist, J., A simplified multidimensional approach for analysis of

complex biological samples: on-line LC-CE-MS. *The Analyst* **2006,** *131* (7), 791-8.
14. Zhang, J.; Hu, H.; Gao, M.; Yang, P.; Zhang, X., Comprehensive two-dimensional chromatography and capillary electrophoresis coupled with tandem time-of-flight mass spectrometry for high-speed proteome analysis. *Electrophoresis* **2004,** *25* (14), 2374-83.
15. Chambers, A. G.; Mellors, J. S.; Henley, W. H.; Ramsey, J. M., Monolithic integration of two-dimensional liquid chromatography-capillary electrophoresis and electrospray ionization on a microfluidic device. *Analytical chemistry* **2011,** *83* (3), 842-9.
16. Mellors, J. S.; Black, W. A.; Chambers, A. G.; Starkey, J. A.; Lacher, N. A.; Ramsey, J. M., Hybrid capillary/microfluidic system for comprehensive online liquid chromatography-capillary electrophoresis-electrospray ionization-mass spectrometry. *Analytical chemistry* **2013,** *85* (8), 4100-6.
17. Jooß, K.; Scholz, N.; Meixner, J.; Neusüß, C., Heart-cut nano-LC-CZE-MS for the characterization of proteins on the intact level. *Electrophoresis* **2019,** *40* (7), 1061-1065.
18. Li, Y.; Champion, M. M.; Sun, L.; Champion, P. A.; Wojcik, R.; Dovichi, N. J., Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry as an alternative proteomics platform to ultraperformance liquid chromatography-electrospray ionization-tandem mass spectrometry for samples of intermediate complexity. *Analytical chemistry* **2012,** *84* (3), 1617-22.
19. Yan, X.; Sun, L.; Zhu, G.; Cox, O. F.; Dovichi, N. J., Over 4100 protein identifications from a Xenopus laevis fertilized egg digest using reversed-phase chromatographic prefractionation followed by capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry analysis. *Proteomics* **2016,** *16* (23), 2945-2952.
20. Faserl, K.; Kremser, L.; Muller, M.; Teis, D.; Lindner, H. H., Quantitative proteomics using ultralow flow capillary electrophoresis-mass spectrometry. *Analytical chemistry* **2015,** *87* (9), 4633-40.
21. Faserl, K.; Sarg, B.; Sola, L.; Lindner, H. H., Enhancing Proteomic Throughput in Capillary Electrophoresis-Mass Spectrometry by Sequential Sample Injection. *Proteomics* **2017,** *17* (22), doi: 10.1002/pmic.201700310.
22. Faserl, K.; Sarg, B.; Gruber, P.; Lindner, H. H., Investigating capillary electrophoresis-mass spectrometry for the analysis of common post-translational modifications. *Electrophoresis* **2018,** *39* (9-10), 1208-1215.
23. Choi, S. B.; Lombard-Banek, C.; Munoz, L. P.; Manzini, M. C.; Nemes, P., Enhanced Peptide Detection Toward Single-Neuron Proteomics by Reversed-Phase Fractionation Capillary Electrophoresis Mass Spectrometry. *J Am Soc Mass Spectrom* **2018,** *29* (5), 913-922.
24. Chen, D.; Shen, X.; Sun, L., Strong cation exchange-reversed phase liquid chromatography-capillary zone electrophoresis-tandem mass spectrometry platform with high peak capacity for deep bottom-up proteomics. *Analytica chimica acta* **2018,** *1012*, 1-9.
25. Chen, D.; Ludwig, K. R.; Krokhin, O. V.; Spicer, V.; Yang, Z.; Shen, X.; Hummon, A. B.; Sun, L., Capillary Zone Electrophoresis-Tandem Mass

Spectrometry for Large-Scale Phosphoproteomics with the Production of over 11,000 Phosphopeptides from the Colon Carcinoma HCT116 Cell Line. *Analytical chemistry* **2019,** *91* (3), 2201-2208.
26. Kulak, N. A.; Geyer, P. E.; Mann, M., Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. *Molecular & cellular proteomics : MCP* **2017,** *16* (4), 694-705.
27. Yang, Z.; Shen, X.; Chen, D.; Sun, L., Microscale Reversed-Phase Liquid Chromatography/Capillary Zone Electrophoresis-Tandem Mass Spectrometry for Deep and Highly Sensitive Bottom-Up Proteomics: Identification of 7500 Proteins with Five Micrograms of an MCF7 Proteome Digest. *Analytical chemistry* **2018,** *90* (17), 10479-10486.
28. Britz-McKibbin, P.; Chen, D. D., Selective focusing of catecholamines and weakly acidic compounds by capillary electrophoresis using a dynamic pH junction. *Analytical chemistry* **2000,** *72* (6), 1242-52.
29. Hughes, C. S.; Foehr, S.; Garfield, D. A.; Furlong, E. E.; Steinmetz, L. M.; Krijgsveld, J., Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular systems biology* **2014,** *10*, 757.
30. Hughes, C. S.; Moggridge, S.; Muller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J., Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature protocols* **2019,** *14* (1), 68-85.
31. 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, doi:10.3791/58644.
32. 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.
33. 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.
34. 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.
35. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **2008**, 26(12), 1367-72.
36. Eng, J. K., McCormack, A. L., Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of The American Society for Mass Spectrometry* **1994**, 5(11), 976-89.
37. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D. J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; Pérez, E.; Uszkoreit, J.; Pfeuffer, J.; Sachsenberg, T.; Yilmaz, S.; Tiwary, S.;

Cox, J.; Audain, E.; Walzer, M.; Jarnuczak, A. F.; Ternent, T.; Brazma, A.; Vizcaíno, J. A., The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research* **2019**, 47(D1), D442-D450.
38. Zhang, Z.; Hebert, A. S.; Westphall, M. S.; Qu, Y.; Coon, J. J.; Dovichi, N. J., Production of Over 27 000 Peptide and Nearly 4400 Protein Identifications by Single-Shot Capillary-Zone Electrophoresis-Mass Spectrometry via Combination of a Very-Low-Electroosmosis Coated Capillary, a Third-Generation Electrokinetically-Pumped Sheath-Flow Nanospray Interface, an Orbitrap Fusion Lumos Tribrid Mass Spectrometer, and an Advanced-Peak-Determination Algorithm. *Analytical chemistry* **2018,** *90* (20), 12090-12093.
39. Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U., Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *Journal of proteome research* **2017,** *16* (11), 4060-4072.
40. Cohen, D.; Dickerson, J. A.; Whitmore, C. D.; Turner, E. H.; Palcic, M. M.; Hindsgaul, O.; Dovichi, N. J., Chemical cytometry: fluorescence-based single-cell analysis. *Annual review of analytical chemistry* **2008,** *1*, 165-90.
41. Wang, N.; Xu, M.; Wang, P.; Li, L., Development of mass spectrometry-based shotgun method for proteome analysis of 500 to 5000 cancer cells. *Analytical chemistry* **2010,** *82* (6), 2262-71.
42. Wisniewski, J. R.; Ostasiewicz, P.; Mann, M., High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *Journal of proteome research* **2011,** *10* (7), 3040-9.
43. Li, S.; Plouffe, B. D.; Belov, A. M.; Ray, S.; Wang, X.; Murthy, S. K.; Karger, B. L.; Ivanov, A. R., An Integrated Platform for Isolation, Processing, and Mass Spectrometry-based Proteomic Profiling of Rare Cells in Whole Blood. *Molecular & cellular proteomics : MCP* **2015,** *14* (6), 1672-83.
44. Chen, W.; Wang, S.; Adhikari, S.; Deng, Z.; Wang, L.; Chen, L.; Ke, M.; Yang, P.; Tian, R., Simple and Integrated Spintip-Based Technology Applied for Deep Proteome Profiling. *Analytical chemistry* **2016,** *88* (9), 4864-71.
45. Bekker-Jensen, D. B.; Kelstrup, C. D.; Batth, T. S.; Larsen, S. C.; Haldrup, C.; Bramsen, J. B.; Sorensen, K. D.; Hoyer, S.; Orntoft, T. F.; Andersen, C. L.; Nielsen, M. L.; Olsen, J. V., An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell systems* **2017,** *4* (6), 587-599 e4.
46. Lambert, S. A.; Jolma, A.; Campitelli, L. F.; Das, P. K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T. R.; Weirauch, M. T., The Human Transcription Factors. *Cell* **2018,** *172* (4), 650-665.
47. Biggin, M. D., Animal transcription networks as highly connected, quantitative continua. *Developmental cell* **2011,** *21* (4), 611-26.
48. 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.; Smith, R. D.; Qian, W. J.; Kelly, R. T., Nanodroplet processing platform for deep and

quantitative proteome profiling of 10-100 mammalian cells. *Nature communications* **2018,** *9* (1), 882.
49. Zhu, Y.; Clair, G.; Chrisler, W. B.; Shen, Y.; Zhao, R.; Shukla, A. K.; Moore, R. J.; Misra, R. S.; Pryhuber, G. S.; Smith, R. D.; Ansong, C.; Kelly, R. T., Proteomic Analysis of Single Mammalian Cells Enabled by Microfluidic Nanodroplet Sample Preparation and Ultrasensitive NanoLC-MS. *Angewandte Chemie* **2018,** *57* (38), 12370-12374.
50. Shao, X.; Wang, X.; Guan, S.; Lin, H.; Yan, G.; Gao, M.; Deng, C.; Zhang, X., Integrated Proteome Analysis Device for Fast Single-Cell Protein Profiling. *Analytical chemistry* **2018,** *90* (23), 14003-14010.
51. Chen, Q.; Yan, G.; Gao, M.; Zhang, X., Ultrasensitive Proteome Profiling for 100 Living Cells by Direct Cell Injection, Online Digestion and Nano-LC-MS/MS Analysis. *Analytical chemistry* **2015,** *87* (13), 6674-80.

52. Lubeckyj, R. A.; McCool, E. N.; Shen, X.; Kou, Q.; Liu, X.; Sun, L., Single-Shot Top-Down Proteomics with Capillary Zone Electrophoresis-Electrospray Ionization-Tandem Mass Spectrometry for Identification of Nearly 600 Escherichia coli Proteoforms. *Analytical Chemistry* **2017**, *89*(22), 12059-12067.

53. 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.

54. Onjiko, R. M.; Moody, S. A.; Nemes, P., Single-cell mass spectrometry reveals small molecules that affect cell fates in the 16-cell embryo. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*(21), 6545-50.

55. Qu, Y.; Sun, L.; Zhang, Z.; Dovichi, N. J., Site-Specific Glycan Heterogeneity Characterization by Hydrophilic Interaction Liquid Chromatography Solid-Phase Extraction, Reversed-Phase Liquid Chromatography Fractionation, and Capillary Zone Electrophoresis-Electrospray Ionization-Tandem Mass Spectrometry. *Analytical Chemistry* **2018**, *90*(2), 1223-1233.

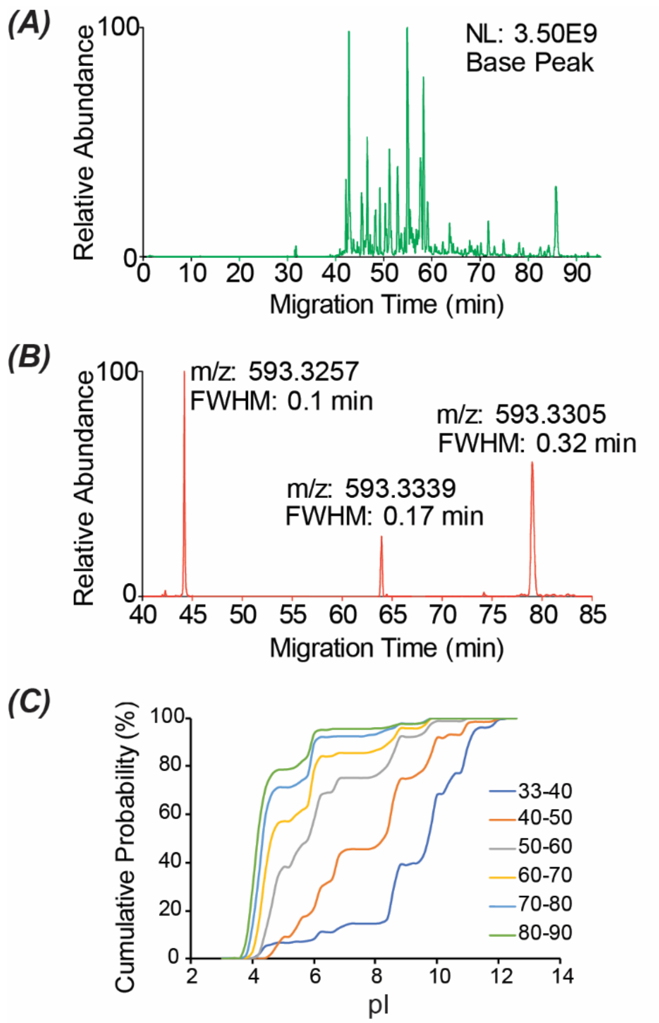


Figure 1. CZE-MS/MS analysis of a 100-ng MCF7 proteome digest. (A) Base peak electropherogram of the CZE-MS/MS analysis. (B) Extracted ion electropherogram of three peptides with similar m/z. The m/z for peak extraction was 593.33 with a mass tolerance of 10 ppm. The m/z and full width at half maximum (FWHM) were labelled. (C) Cumulative distribution of isoelectric point (pI) of peptides identified by CZE-MS/MS in different periods during a run. The identified peptides from 33 to 90 min were separated into six groups based on their migration time.

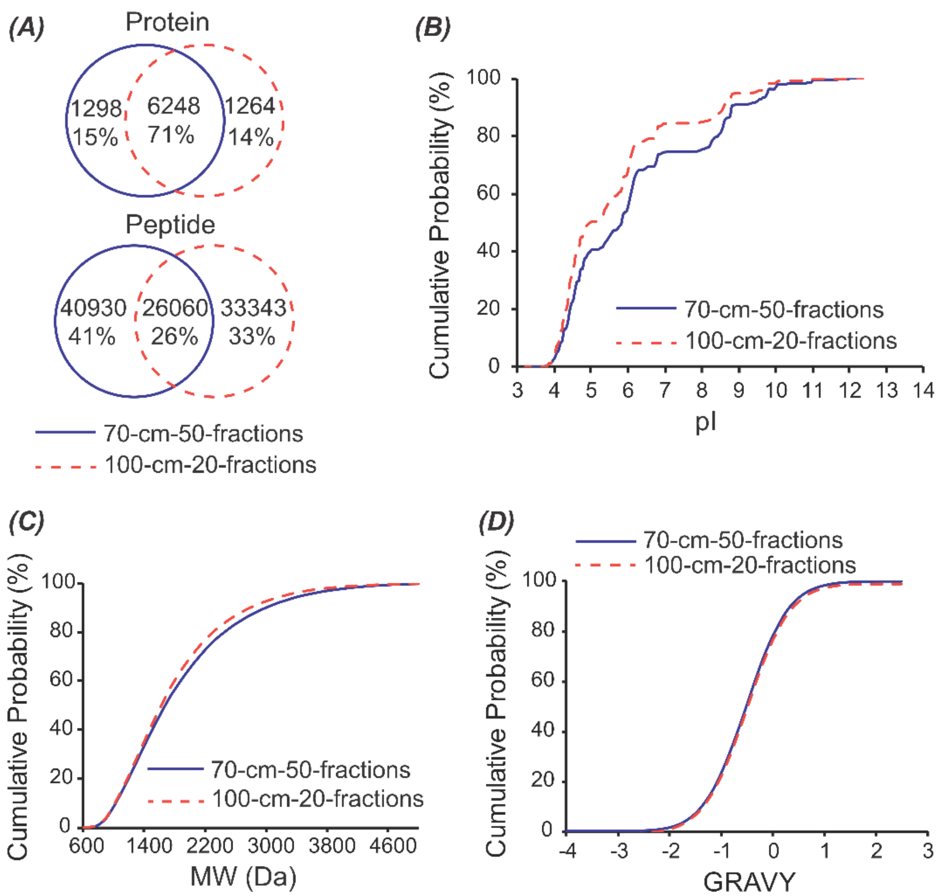


Figure 2. Summary of the 5-μg MCF7 proteome digest data from 70-cm-50-fractions and 100-cm-20 fractions experiments. (A) Overlaps of identified proteins and peptides. (B) Cumulative distribution of isoelectric point (pI) of peptides. (C) Cumulative distribution of molecular weight (MW) of peptides. (D) Cumulative distribution of GRAVY (grand average of hydropathy) value of peptides. Positive GRAVY values indicate hydrophobic and negative GRAVY values suggest hydrophilic.

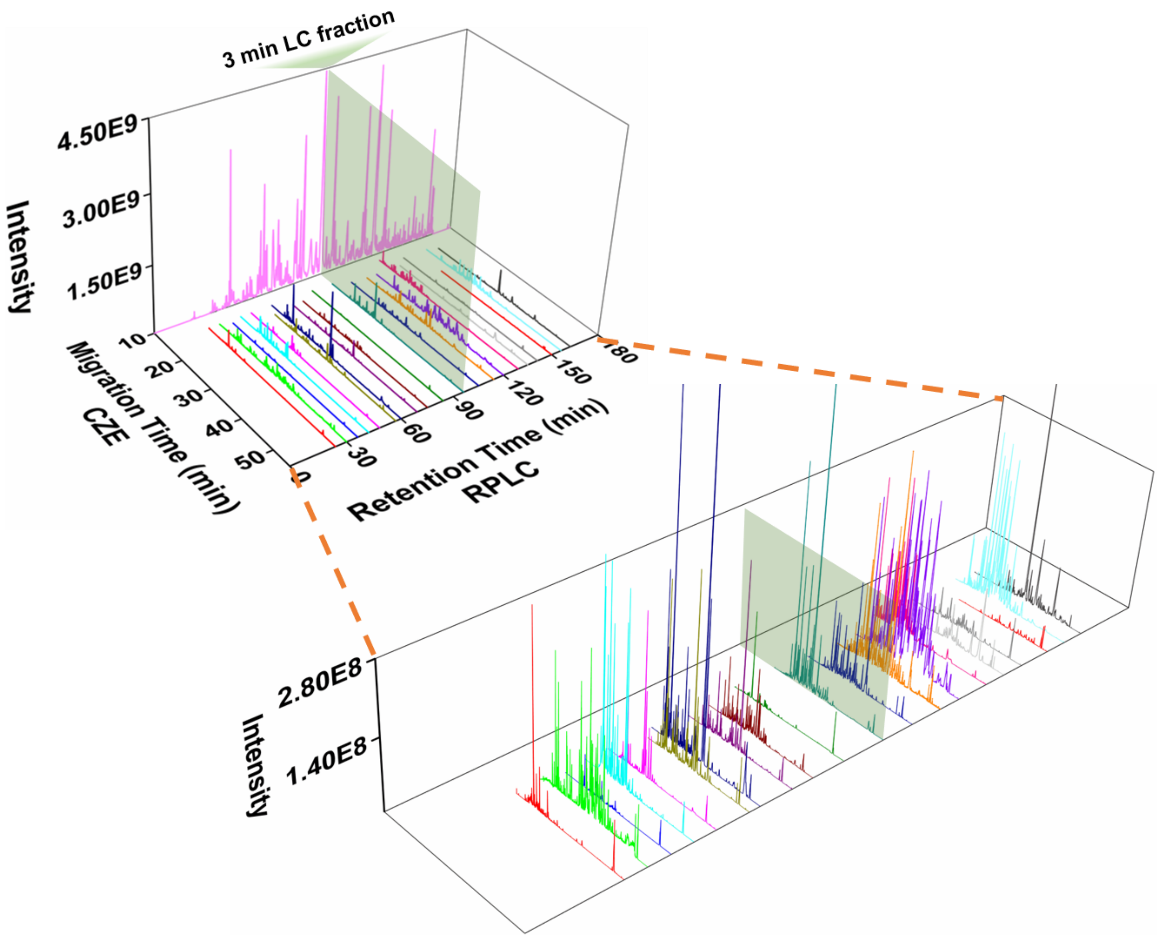


Figure 3. 3D plot of a base peak chromatogram of a 500-ng MCF7 proteome digest and base peak electropherograms of 20 nanoRPLC fractions from the 70-cm-50-fractions study of the 500-ng MCF7 proteome digest. A zoom-in plot is also shown in the figure.

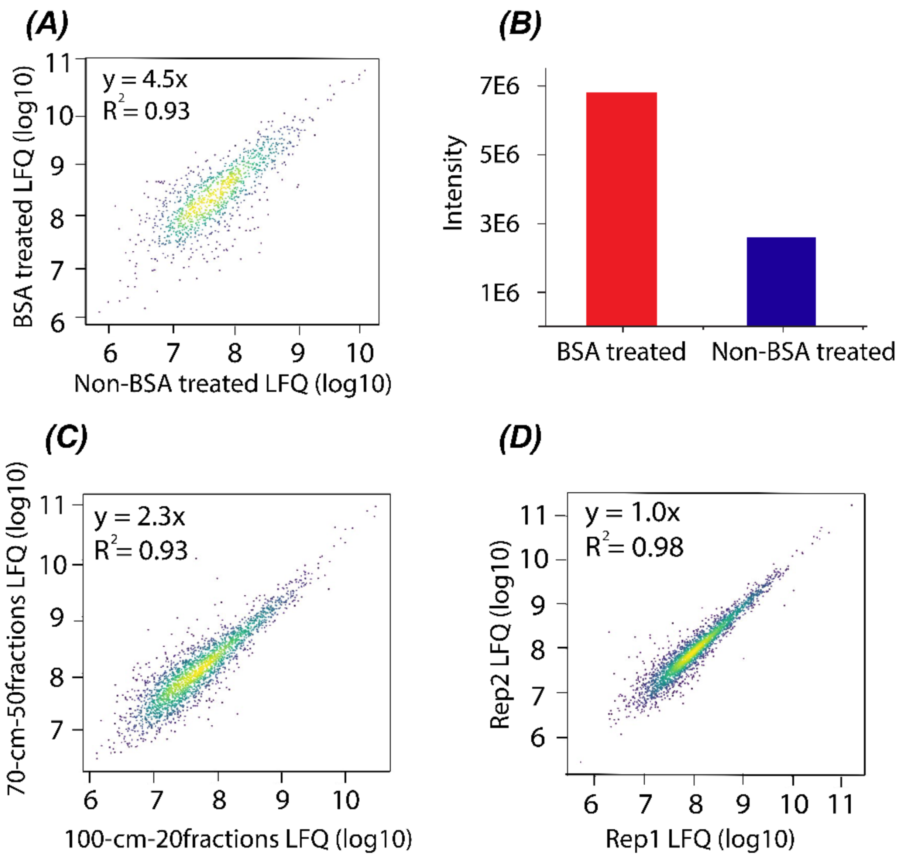
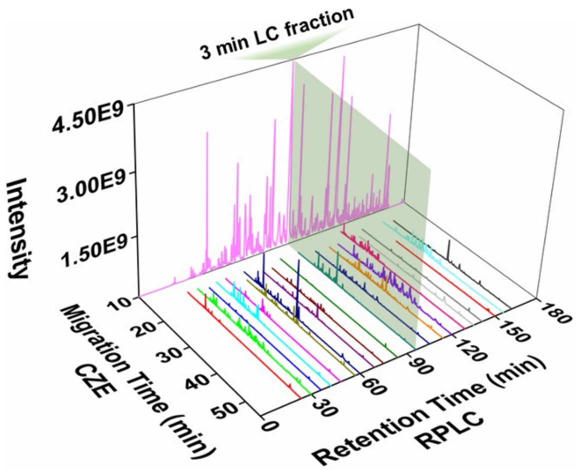


Figure 4. Summary of the 500-ng MCF7 proteome digest data from different experiments regarding peptide intensity and protein LFQ intensity. The data were from MaxQuant database search. (A) Correlation of protein LFQ intensity (log-log) between the studies using BSA treated and non-BSA treated sample vials. (B) The medians of peptide intensities from the studies using BSA treated and non-BSA treated sample vials. (C) Correlation of protein LFQ intensity (log-log) between the 100-cm-20fractions and 70-cm-50fractions studies. (D) Correlation of protein LFQ intensity (log-log) between two replicates of nanoRPLC-CZE-MS/MS analysis (Rep 1 and Rep 2).

For TOC only:



3 min LC fraction
Intensity
4.50E9
3.00E9
1.50E9
10
20
30
40
50
Migration Time (min)
CZE
0
30
60
90
120
150
180
Retention Time (min)
RPLC