Published in final edited form as:

Anal Chem. 2018 September 04; 90(17): 10479–10486. doi:10.1021/acs.analchem.8b02466.

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

Zhichang Yang, Xiaojing Shen, Daoyang Chen, and Liangliang Sun*

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

Abstract

Capillary zone electrophoresis-tandem mass spectrometry (CZE-MS/MS) has been well recognized for bottom-up proteomics. It has approached 4000-8000 protein identifications (IDs) from a human cell line, mouse brains or *Xenopus* embryos via coupling with liquid chromatography (LC) prefractionation. However, at least five hundred micrograms of complex proteome digests were required for the LC-CZE-MS/MS studies. This requirement of a large amount of initial peptide material impedes the application of CZE-MS/MS for deep bottom-up proteomics of mass-limited samples. In this work, we coupled microscale reversed-phase LC (μ RPLC) based peptide prefractionation to dynamic pH junction based CZE-MS/MS for deep bottom-up proteomics of the MCF7 breast cancer cell proteome starting with only 5- μ g peptides. The dynamic pH junction based CZE enabled a 500-nL sample injection from as low as a 1.5- μ L peptide sample, using up to 33% of the available peptide material for an analysis. Two kinds of μ RPLC prefractionation were investigated, C18 ZipTip and nanoflow RPLC. C18 ZipTip-CZE-MS/MS identified 4453 proteins from 5 μ g of the MCF7 proteome digest and showed good qualitative and quantitative reproducibility. Nanoflow RPLC-CZE-MS/MS produced over 7500 protein IDs and nearly 60000 peptide IDs from the 5- μ g MCF7 proteome digest. The nanoflow RPLC-CZE-MS/MS platform reduced the required amount of complex proteome digests for LC-CZE-MS/MS-based deep bottom-up proteomics by two orders of magnitude. Our work provides the proteomics community with a powerful tool for deep and highly sensitive proteomics.

For TOC only:

*Corresponding author. lsun@chemistry.msu.edu.

Supporting Information

Materials and Reagents; Preparation of the MCF7 breast cancer cell proteome digest; nano 2D-RPLC-MS/MS analysis; Cumulative protein IDs vs. number of RPLC fractions; Comparisons between 2D-RPLC-MS/MS and RPLC-CZE-MS/MS in terms of the protein-level and peptide-level overlaps (PDF)

The identified proteins (XLSX)

Introduction

Capillary zone electrophoresis-tandem mass spectrometry (CZE-MS/MS) has a long history in bottom-up proteomics. Yates group reported bottom-up proteomics of yeast ribosome using CZE-MS/MS in 1999.^[1] Since then, there have been many reports on exploring CZE-MS/MS for bottom-up proteomics of samples with varying complexity.^[2–14] CZE-MS/MS has some unique features compared to widely used reversed-phase liquid chromatography (RPLC)-MS/MS for bottom-up proteomics. First, it outperforms RPLC-MS/MS for analyses of low nanograms of complex proteome digests in terms of the number of protein identifications (IDs).^[3,6,9,11] This feature makes CZE-MS/MS very useful for mass-limited sample analyses, *e.g.*, characterization of single cells.^[11] Second, migration time of peptides in CZE can be predicted easily with high accuracy while prediction of peptides' retention time in RPLC is much more difficult due to various interactions between peptides and beads.^[15] This valuable feature of CZE provides the proteomics community with a new approach to validate the confidence of peptide ID from a database search. Third, CZE can reach much better separation of phosphorylated and unphosphorylated forms of peptides than RPLC,^[9] which makes CZE-MS/MS very useful for phosphoproteomics. Fourth, CZE-MS/MS has no bias against the ID of very hydrophilic peptides that usually have no or very weak retention on RPLC columns.^[3,8]

Recently, CZE-MS/MS has achieved great progress in deep bottom-up proteomics due to improvement in CE-MS interface,^[16–19] mass spectrometer,^[20,21] and online sample stacking method.^[6, 22–26] Yan *et al.* identified over 4000 proteins from *Xenopus* embryos via coupling RPLC prefractionation to CZE-MS/MS.^[25] Very recently, our group reported over 8000 protein IDs from a mouse brain proteome digest using two-dimensional (2D) LC-CZE-MS/MS.^[27] LC-CZE-MS/MS is becoming a useful tool for deep bottom-up proteomics. LC-CZE-MS/MS and the state-of-the-art 2D-LC-MS/MS are complementary in peptide and protein IDs from complex proteomes, and combining those two techniques can boost the proteome coverage from bottom-up proteomics.^[27]

Hundreds of micrograms of proteome digests were typically required for the LC-CZE-MS/MS-based deep bottom-up proteomics studies.^[5,25–27] There are at least two reasons for the requirement of large amounts of proteome digests. First, typically LC fractionation

employs a LC column with a large inner diameter (2.1 or 4.6 mm). The total surface area of beads packed in the column and the optimum flow rate of mobile phase for separation are high, leading to unavoidable sample loss on the beads and on the inner wall of Eppendorf tubes used for fraction collection.^[28] Second, the sample loading volume of CZE is low, typically less than 100 nL.^[5,25,26] A reasonable amount of peptide material is required for identification of low abundant proteins in a complex proteome sample with MS/MS. Therefore, a relatively high peptide concentration in the sample is essential for CZE-MS/MS in order to reach large-scale proteomics. The significant sample loss during LC fractionation and the requirement of high peptide concentration in the sample for CZE-MS/MS lead to the need of a large amount of starting peptide material for deep bottom-up proteomics using the LC-CZE-MS/MS. The requirement of large amount of peptides impedes the application of CZE-MS/MS for deep bottom-up proteomics of mass-limited samples.

Microscale RPLC (μ RPLC)-based peptide fractionation and CZE-MS/MS with a much higher sample loading volume are vital for dramatically reducing the required amount of initial peptide material. Several groups have tested different μ RPLC methods for fractionation of complex proteome digests prior to CZE-MS/MS or nanoRPLC-MS/MS.^[6,28–31] RP-based solid phase microextraction (RP-SPME) and strong cation exchange monolith based SPME have been online coupled with CZE-MS/MS for highly sensitive bottom-up proteomics using nanograms of proteome digests.^[6,29,30] The SPME-CZE-MS/MS platforms will be very useful for large-scale proteomics of mass-limited proteome samples, *e.g.*, single cells. C18 ZipTip has low- μ g loading capacity and it is a simple and efficient μ RPLC method for peptide fractionation. Choi *et al.* have employed the C18 ZipTip-CZE-MS/MS recently to enhance the protein ID from mammalian neuron proteome digests that correspond to the protein content of a few mammalian neurons in mass.^[31] SPME and C18 ZipTip can be considered as low-resolution μ RPLC. Kulak *et al.* recently showed that coupling high-resolution and high-pH μ RPLC fractionation to low-pH nanoflow RPLC-MS/MS enabled deep bottom-up proteomics of human cell lines using low- μ g of proteome digests.^[28] The data in that work clearly demonstrated that high-resolution μ RPLC could reach highly efficient peptide fractionation with high peptide recovery. In order to reach a high sample loading volume in CZE, efficient online sample concentration is essential. The dynamic pH junction method^[32,33] has been evaluated for the online concentration of peptides and intact proteins in CZE.^[14,24,34,35] Our group demonstrated that the dynamic pH junction based CZE-MS/MS could approach a microliter scale sample loading volume and a 2-hour separation window simultaneously.^[14,34] We speculate that coupling the low-resolution C18 ZipTip or the high-resolution μ RPLC prefractionation to the dynamic pH junction based CZE-MS/MS will enable deep bottom-up proteomics starting with low- μ g complex proteome digests.

In this work, we first performed a calibration curve experiment using the dynamic pH junction based CZE-MS/MS for bottom-up proteomics of the MCF7 proteome with 0.1–100 ng of protein digests as the starting material. Then we investigated C18 ZipTip-dynamic pH junction based CZE-MS/MS for deep bottom-up proteomics of the MCF7 proteome starting with 5- μ g protein digest. Finally, we coupled high-resolution nanoflow RPLC (nanoRPLC) to the dynamic pH junction based CZE-MS/MS to deepen the coverage of the MCF7 proteome with 5- μ g protein digest.

Experimental Section

“Materials and reagents”, “Preparation of the MCF7 breast cancer cell proteome digest” and “Nano 2D-RPLC-MS/MS analysis” are described in Supporting Information I.

Calibration curve experiment

We prepared a series of MCF7 peptide samples in 20 mM ammonium bicarbonate (NH_4HCO_3 , pH 8.0) with various concentrations: 50 ng/ μL , 5 ng/ μL , 0.5 ng/ μL , and 0.05 ng/ μL . 2 μL of each sample was put into a sample vial and a 500-nL aliquot of each sample was injected for CZE-MS/MS. The mass of peptides in the sample vial was 100 ng, 10 ng, 1 ng, and 0.1 ng corresponding to different peptide concentrations from 50 ng/ μL to 0.05 ng/ μL . Each sample was analyzed by one CZE-MS/MS run. We repeated the experiment three times to obtain triplicate analyses of each sample.

C18 ZipTip fractionation

A C18 ZipTip (Millipore, ZTC18S096) was first conditioned with 100 μL of 80% (v/v) acetonitrile (ACN) containing 20 mM NH_4HCO_3 (pH 8.0) and equilibrated with 100 μL of 3% (v/v) ACN containing 20 mM NH_4HCO_3 (pH 8.0). 5- μg MCF7 proteome digest dissolved in 20 mM NH_4HCO_3 (pH 8.0) was loaded onto the ZipTip by pipetting up and down for 20 times. The flow-through from the loading was kept for further analysis. The peptides retained on the ZipTip were separated by step-wise elution using 6%, 15%, 20%, 25%, 40%, and 80% (v/v) ACN containing 20 mM NH_4HCO_3 (pH 8.0). Each elution was processed by centrifugation at 1000 rpm for 2 min and 4000 rpm for 30 seconds using 20- μL elution buffer. The flow-through from loading and the eluate from 80% (v/v) ACN were combined. Each eluate was lyophilized and redissolved in 5 μL of 20 mM NH_4HCO_3 (pH 8.0) for CZE-MS/MS analysis.

NanoRPLC fractionation

An EASY nanoLC-1200 (Thermo Fisher Scientific) was used for the fractionation. An analytical column (75 μm i.d. $\times$ 50 cm, C18, 2 μm , 100 $\text{\AA}$, Thermo Fisher Scientific) was employed for RPLC. Buffer A containing 0.1% (v/v) formic acid (FA) and buffer B containing 80% (v/v) ACN and 0.1% (v/v) FA were used to generate gradient separation. 5- μg MCF7 proteome digest was loaded onto the RPLC column with buffer A at 800-bar pressure. Then the peptides retained on the column were separated by a linear gradient. The flow rate was 200 nL/min.

For the first RPLC fraction experiment, the gradient used for peptide separation was as follows: from 8 to 30% (v/v) B in 50 min, from 30% to 50% (v/v) B in 25 min, from 50% to 100% (v/v) B in 5 min and maintain at 100% (v/v) B for 10 min. The fraction collection started from the sample loading. The flow-through during sample loading and the peptides eluted during the first 14 min of the gradient were collected in one Eppendorf tube as the first fraction. After that, we collected one fraction every 7 min. The last fraction was collected from 70 min to 90 min. In total, 10 fractions were collected. The eluate for each fraction was directly transferred to the bottom of one 0.6-mL Eppendorf tube during fraction

collection. After lyophilization, each fraction was redissolved in 3 μL of 20 mM NH_4HCO_3 (pH 8.0) by pipetting up and down for CZE-MS/MS.

For the second RPLC fractionation experiment, the gradient for RPLC separation was as follows: from 5 to 20% (v/v) B in 100 min, from 20% to 40% (v/v) B in 50 min, from 40% to 100% (v/v) B in 15 min and maintain at 100% (v/v) B for 15 min. The fraction collection started from the sample loading. The flow-through during sample loading and the peptides eluted during the first 15 min of the gradient were collected as the first fraction. Then, each fraction was collected every 7 min. The last fraction was collected starting from 141 min until the end of the gradient. We collected 20 fractions in total into 0.6-mL Eppendorf tubes. All fractions were lyophilized and each fraction was redissolved in 1.5 μL of 20 mM NH_4HCO_3 (pH 8.0) for CZE-MS/MS.

CZE-MS/MS analysis

A one-meter linear polyacrylamide (LPA) coated fused silica capillary (50 μm i.d./360 μm o.d.) was used for CZE. The LPA coating was prepared based on references [14] and [36]. One end of the capillary was etched with hydrofluoric acid (HF) to reduce its outer diameter to ~ 70 μm . The length of the etched part was about 5 mm. The LPA coated capillary was stored at room temperature before use.

The commercialized electro-kinetically pumped sheath flow CE-MS interface (CMP scientific, Brooklyn, NY) was used to couple CZE to MS. [17–19] The automated CZE operations were carried out with an ECE-001 autosampler (CMP scientific) or a CESI 8000 plus CE system (Sciex). The background electrolyte (BGE) for CZE was 5% (v/v) acetic acid (pH 2.4). The sheath buffer was 10% (v/v) methanol and 0.2% (v/v) FA. The glass emitter for electrospray was pulled from a borosilicate glass capillary (0.75 mm i.d., 1 mm o.d.) by a Sutter P-1000 flaming/brown micropipette puller. The orifice of the electrospray emitter was 20–40 μm . The distance between the etched end of the separation capillary and the emitter orifice was less than 300 μm . The distance between the emitter orifice and the MS entrance was 2 mm.

Sample injection was performed by applying 5-psi pressure for 90 seconds. About 500 nL of the sample was loaded into the capillary for CZE separation. After sample loading, the injection end of the capillary was moved into the BGE vial. 30 kV voltage was applied at the injection end for CZE separation and 2 kV was applied in the sheath buffer vial for electrospray. The CZE separation was finished in 90 min or 100 min. After the CZE separation, 15-psi pressure and 30 kV were applied at the injection end to flush the capillary for 5 min. In order to inject a 500-nL sample plug for CZE-MS/MS from only 5 μL or lower volume of sample in a sample vial, we pushed the injection end of the capillary to the bottom of the sample vial for injection. For the experiments on the CESI 8000 plus CE system, we used the nanoVials (Part number 5043467, SCIEX) for sample injection.

A Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) was used in all the experiments. For all experiments, the resolution for full MS was 60000, the AGC was 1E6, the maximum injection time was 50 ms and the scan range was 300–1800 m/z . A Top10 data-dependent acquisition (DDA) method was applied. Only ions with a charge of 2 or

higher were isolated in the quadrupole and fragmented in the high energy collision dissociation (HCD) cell with normalized collision energy (NCE) as 28%. Other parameters were varied for different experiments.

For the calibration-curve experiment, we changed some parameters for different amounts of MCF7 peptides. For CZE-MS/MS analysis of 0.1-ng to 10-ng MCF7 peptides, the resolution for MS/MS was 60000, the AGC target for MS/MS was 1E5, and the maximum injection time for MS/MS was 250 ms. The isolation window was 4 m/z and the intensity threshold for fragmentation was 2.2E4. Dynamic exclusion was 10 s. For CZE-MS/MS analysis of 100-ng sample, all the parameters were the same as that used for 0.1-ng to 10-ng samples except the MS/MS maximum injection time as 110 ms, the isolation window as 2 m/z , the intensity threshold for fragmentation as 5E4, and the dynamic exclusion as 30 s. The parameter settings were referred to reference [37]. It has been demonstrated that a high injection time (250 ms) and a wide isolation window (4 m/z) can benefit the analysis of mass-limited samples. [37] For the 100-ng sample, a 110-ms maximum injection time and a 2- m/z isolation window were used to acquire more MS/MS spectra and reduce the co-isolation of peptides for fragmentation.

For CZE-MS/MS analysis of the fractions from C18 ZipTip and nanoflow RPLC fractionation, the resolution for MS/MS was 30000, the maximum injection time for MS/MS was 50 ms, the AGC target for MS/MS was 1E5, the isolation window was 2.0 m/z , the intensity threshold for fragmentation was 5E4, and the dynamic exclusion was 30 s.

Data Analysis

Proteome Discoverer 2.2 (Thermo Fisher Scientific) was used for the database search. SEQUEST HT searching engine was used. Precursor mass tolerance was 20 ppm while fragment mass tolerance was 0.05 Da. UniProt human proteome database (UP000005640) was used for the database searching. Trypsin was set as the enzyme with two maximum missed cleavages. The false discovery rate (FDR) was evaluated by the target and decoy approach. [38,39] Methionine oxidation, Protein N-terminal acetylation, and Asparagine or Glutamine deamination were set as variable modifications. Carbamidomethylation of cysteine was set as the fixed modification. The peptides were filtered with confidence as high, corresponding to a 1% peptide-level FDR. Protein grouping was enabled, and the strict parsimony principle was applied. All the numbers of protein and peptide IDs reported in this work were from Proteome Discoverer 2.2.

We also performed database search using MaxQuant software [40] (version 1.5.5.1) to obtain label-free quantification (LFQ) information. All the parameters were kept as default settings. The LFQ and match-between-runs function were enabled. [41] The UniProt human proteome database (UP000005640) was used for the database searching. Perseus software (version 1.6.0.7) was used for further analysis of the MaxQuant results. [42] The MaxQuant data was only used for evaluation of the quantitative reproducibility of the C18-ZipTip-CZE-MS/MS based on the LFQ information.

Grand average of hydropathy (GRAVY) values of peptides were calculated through GRAVY Calculator (<http://www.gravy-calculator.de/>). Positive GRAVY values indicate hydrophobic and negative values show hydrophilic.

Results and Discussion

Calibration curve data

CZE-MS typically can approach excellent mass detection limit and relatively unsatisfied concentration detection limit due to the low sample loading volume of CZE. Hundreds to thousands of protein IDs have been reported using CZE-MS/MS for analysis of low ng even pg of complex proteome digests.^[11,18,31,43] However, due to the low nL loading volume of CZE, the concentration of proteome digests in sample vials for injection need to be high for identification of low abundant proteins in complex proteomes, leading to the requirement of at least 1- μ g protein digest for the CZE-MS/MS studies. Recently, we boosted the sample loading volume of CZE to microliter scale based on the dynamic pH junction online sample stacking method.^[14,34] The large sample loading volume offers us the opportunity to fully use the available peptide material in the sample vial, dramatically reducing the required mass of peptides for CZE-MS/MS. Here we investigated the dynamic pH junction based CZE-MS/MS for analysis of a MCF7 proteome digest using 0.1–100 ng of peptides as the starting material. Each CZE-MS/MS run took 90 min.

Two- μ L aliquots of MCF7 peptide samples with a various concentration in 20 mM NH_4HCO_3 (pH 8.0) were put into sample vials for CZE-MS/MS. The mass of peptides in the sample vials ranged from 0.1 ng to 100 ng. Our CZE-MS/MS system enabled a 500-nL sample injection from the 2- μ L available sample in the vial, using 25% of the available sample for analysis. A diagram of the dynamic pH junction based CZE-MS/MS is shown in Figure 1A. Figure 1B shows the number of peptide and protein IDs as a function of the amount of available peptide material and injected peptide material. On average 1200 proteins and 5300 peptides were identified by the CZE-MS/MS when only 100-ng peptide was available in the vial. The data clearly indicate that our dynamic pH junction based CZE-MS/MS is able to reduce the required mass of complex proteome digests by one order of magnitude compared with the literature data (100 ng vs. 1 μ g).^[11] The CZE-MS/MS was able to identify on average 100 proteins and 15 proteins from the MCF7 proteome starting with only 1-ng and 0.1-ng peptide, which correspond to the protein content of only ten and one MCF7 cells in mass.^[44] The data highlight the potential of the CZE-MS/MS system for single cell proteomics.

As shown in Figure 1B, the number of protein and peptide IDs steadily increase as the amount of available peptide material grows. The CZE-MS/MS produced reproducible protein and peptide IDs from triplicate analyses with the relative standard deviations (RSDs) lower than 10% when 10 ng and 100 ng of peptides were available. We extracted the peaks of 12 peptides across all the raw files of 0.1–100 ng samples. The 12 peptides were randomly chosen from that identified from the 0.1-ng sample. We plotted the peptide intensity as a function of the available peptide amount, Figure 1C. The peptide intensity grows linearly as the peptide amount increases. The data suggest high confidence of those identified peptides from the 0.1-ng sample.

The calibration curve data indicate high sensitivity of our dynamic pH junction based CZE-MS/MS system and inspired us to think about deep proteomics of the MCF7 cells with only low- μ g peptides using μ RPLC fractionation-dynamic pH junction CZE-MS/MS.

C18 ZipTip fractionation-CZE-MS/MS

We investigated the C18 ZipTip fractionation-dynamic pH junction CZE-MS/MS for deep bottom-up proteomics of the MCF7 cells starting with only 5- μ g proteome digest. The C18 ZipTip fractionated the 5- μ g peptides into 6 fractions based on their hydrophobicity. Each eluate was lyophilized and redissolved in 5 μ L of 20 mM NH_4HCO_3 (pH 8.0) for CZE-MS/MS analysis. Each fraction was analyzed by one CZE-MS/MS run in 105 min with a 500-nL sample loading volume. One C18 ZipTip-CZE-MS/MS experiment took 630 min of instrument time. The whole experiment was performed three times.

The platform produced 4453 ± 304 (n=3) protein IDs, 26928 ± 2489 (n=3) peptide IDs, 93354 ± 11209 (n=3) peptide-spectrum matches (PSMs) and 229148 ± 10524 (n=3) MS/MS spectra, Figure 2A. The identified proteins are listed in Supporting Information II. The data indicate the good reproducibility of the platform. More importantly, for the first time, CZE-MS/MS was able to approach deep bottom-up proteomics of a human cell line proteome reproducibly using only low- μ g of starting peptide material. The C18 ZipTip fractionation is simple and straightforward. It can be completed within 20 min. We believe the C18-ZipTip fractionation-dynamic pH junction CZE-MS/MS can be a useful tool for deep bottom-up proteomics.

The C18 ZipTip-CZE-MS/MS was performed in triplicate and the protein-level overlap between replicates is reasonable (>70%), Figure 2B. The platform also shows great quantitative reproducibility based on the LFQ intensity of quantified proteins ($R^2 > 0.97$), Figure 2C. We further investigated the performance of C18 ZipTip for peptide fractionation, Figure 2D. We used the first replicate data and obtained the peptide-level overlap between adjacent C18 ZipTip fractions. CZE-MS/MS generated 4000–9000 peptide IDs from one C18 ZipTip fraction. The adjacent fractions share a significant portion of the identified peptides, and the number of shared peptides ranges from 2000 to 4000. The obvious peptide-level overlap between adjacent C18 ZipTip fractions is most likely due to the low resolution of C18 ZipTip for peptide separation. We speculated that coupling high-resolution μ RPLC fractionation to the dynamic pH junction CZE-MS/MS could further deepen the coverage of the MCF7 proteome.

NanoRPLC fractionation-CZE-MS/MS

We further coupled high-resolution nanoRPLC fractionation to the dynamic pH junction based CZE-MS/MS for deep bottom-up proteomics of the MCF7 cell proteome starting with only 5- μ g proteome digest. One RPLC column (75- μ m i.d., 50-cm length) packed with 2- μ m C18 beads was used for the high-resolution fractionation. The flow rate was 200 nL/min.

Two experiments were performed. In the first experiment, we fractionated the 5- μ g proteome digest with the RPLC column into 10 fractions, lyophilized the fractions, and redissolved each fraction in 3 μ L of 20 mM NH_4HCO_3 (pH 8.0) for CZE-MS/MS with a 500-nL sample loading volume. The CZE-MS/MS identified 5769 proteins and 35575 peptides from the 10

RPLC fractions. The experiment took 1100-min instrument time. In the second experiment, we fractionated another 5- μ g MCF7 proteome digest sample with the same RPLC column and obtained 20 fractions. Because the average amount of peptides in each fraction was lower than that in the first experiment due to more fractions (20 vs. 10), we dissolved the peptides in each fraction in only 1.5 μ L of 20 mM NH_4HCO_3 (pH 8.0) for CZE-MS/MS. Over 30% of the peptide material in each fraction was used for the analysis (500 nL vs. 1.5 μ L) in order to identify lower abundant proteins. CZE-MS/MS analyses of the 20 RPLC fractions in 34 hours generated 7512 protein IDs, 59403 peptide IDs, 152086 PSMs, and 543622 MS/MS spectra. The number of peptide and protein IDs was improved by 67% and 30%, respectively, compared to that from the first experiment. Compared with recent deep bottom-up proteomics work using LC-CZE-MS/MS, [5,25–27] this work reduced the required amount of complex proteome digests by two orders of magnitude (5 μ g vs. 500 μ g). We noted that different mass spectrometers used in those studies could also affect the overall sensitivity of LC-CZE-MS/MS. We also noted that CZE-MS/MS analyses of only 8 RPLC fractions in less than 14 hours could yield 6000 protein IDs, Figure S1. The nanoRPLC-CZE-MS/MS platform will become a powerful tool for highly sensitive and deep proteomics. The identified proteins are listed in Supporting Information II. The mass spectrometry proteomics raw files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD009991. [45,46]

We attribute the drastically high sensitivity of our nanoflow RPLC-CZE-MS/MS system to three major reasons. First, the nanoRPLC system significantly reduced the sample loss due to the much lower total surface area of beads packed in the column and much lower mobile phase flow rate compared with the RPLC systems used in the literature (75- μ m-i.d. column vs. 2.1-mm-i.d. column; 200 nL/min vs. 200 μ L/min). [5,25–27] The data reported recently by the Mann's group provided strong evidence for this point. [28] Second, the nanoRPLC produced high-resolution separation of peptides, leading to negligible peptide-level overlap between adjacent RPLC fractions, Figure 3A. The high-resolution RPLC fractionation reduced the sample complexity and ion suppression during the electrospray ionization. Third, the dynamic pH junction based CZE-MS/MS was able to use over 30% of the peptide material in each fraction for analysis. In previous LC-CZE-MS/MS studies, [5,25,26] the CZE-MS/MS system was only able to inject less than 1 % of the peptide material in each LC fraction for analysis.

We noted that the nanoRPLC and dynamic pH junction based CZE yielded highly orthogonal separation of peptides. RPLC separates peptides based on their hydrophobicity. As shown in Figure 3B, the peptides in later RPLC fractions tend to be more hydrophobic, which agrees well with the separation mechanism of RPLC. CZE separates peptides based on their size-to-charge ratios. [15] The retention time of peptides in RPLC and their migration time in CZE have no significant correlation, demonstrating that RPLC and CZE are highly orthogonal for peptide separation, Figure 3C. The peptides in every 7-min eluate from the nanoRPLC were further separated by CZE in a 60-min window.

Recently, the Mann's group developed a highly sensitive 2D-RPLC-MS/MS system via coupling high-pH nanoRPLC fractionation to low-pH nanoRPLC-MS/MS for deep bottom-up proteomics. [28] In this work, we also applied the nano-2D-RPLC-MS/MS for deep

bottom-up proteomics of the MCF7 proteome using 5- μ g of the proteome digest as the starting material. We fractionated the digest into 30 fractions with high-pH nanoRPLC and combined the 30 fractions into 15 fractions. Then each fraction was analyzed by low-pH nanoRPLC-MS/MS with a 2-hour gradient. The total MS analyses including the sample loading and column equilibrium took roughly 40 hours. We identified 8605 proteins and 91546 peptides after the database search. Our nanoRPLC-CZE-MS/MS generated 13% lower number of protein IDs (7512 vs. 8605 proteins) and 35% lower number of peptide IDs (59403 vs. 91546 peptides) compared to nano-2D-RPLC-MS/MS with 15% shorter MS time (34 vs. 40 hours). Those two platforms have good complementarity in peptide ID and moderate complementarity in protein ID, Figure S2. Only 70% of the peptides identified using nanoRPLC-CZE-MS/MS were covered by those identified using nano-2D-RPLC-MS/MS. 677 proteins were only identified by the nanoRPLC-CZE-MS/MS.

We need to point out that we did not fully optimize the nanoRPLC-CZE-MS/MS in this work. We speculate that performance of the nanoRPLC-CZE-MS/MS can be further improved by boosting the nanoRPLC separation of peptides, via optimizing the strategy used for fraction collection from the nanoRPLC, and through increasing the loading volume and separation window of CZE.

Conclusion

We developed two μ RPLC-CZE-MS/MS platforms for highly sensitive and deep bottom-up proteomics. The C18 ZipTip-CZE-MS/MS identified 4453 proteins from 5- μ g MCF7 proteome digest with good qualitative and quantitative reproducibility. The nanoRPLC-CZE-MS/MS platform reached over 7500 protein IDs and nearly 60000 peptide IDs with only 5- μ g MCF7 proteome digest and improved the overall sensitivity of LC-CZE-MS/MS for deep bottom-up proteomics drastically. The nanoRPLC-CZE-MS/MS is complementary with the nano-2D-RPLC-MS/MS in terms of the protein and peptide IDs from the MCF7 proteome. This work provides the proteomics community with a powerful tool for deep proteomics of mass-limited samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Prof. Xuefei Huang's group at Department of Chemistry, Michigan State University for kindly providing the MCF7 cells for our research. We thank the support from the Michigan State University and the National Institute of General Medical Sciences, National Institutes of Health (NIH), through Grant R01GM125991.

References

- [1]. Tong W; Link A; Eng JK; Yates JR, 3rd. *Anal. Chem.* 1999, 71, 2270–2278. [PubMed: 10405597]
- [2]. Garza S; Moini M *Anal. Chem.* 2006, 78, 7309–16. [PubMed: 17037937]
- [3]. Faserl K; Sarg B; Kremser L; Lindner H *Anal. Chem.* 2011, 83, 7297–305. [PubMed: 21848273]
- [4]. Sarg B; Faserl K; Kremser L; Halfinger B; Sebastiano R; Lindner HH *Mol. Cell. Proteomics* 2013, 12, 2640–56. [PubMed: 23720761]

- [5]. Faserl K; Kremser L; Müller M; Teis D; Lindner HH *Anal. Chem.* 2015, 87, 4633–40. [PubMed: 25839223]
- [6]. Wang Y; Fonslow BR; Wong CC; Nakorchevsky A; Yates JR, 3rd. *Anal. Chem.* 2012, 84, 8505–13. [PubMed: 23004022]
- [7]. Li Y; Champion MM; Sun L; Champion PA; Wojcik R; Dovichi NJ *Anal. Chem.* 2012, 84, 1617–22. [PubMed: 22182061]
- [8]. Zhu G; Sun L; Yan X; Dovichi NJ *Anal. Chem.* 2013, 85, 2569–73. [PubMed: 23394296]
- [9]. Ludwig KR; Sun L; Zhu G; Dovichi NJ; Hummon AB *Anal. Chem.* 2015, 87, 9532–7. [PubMed: 26399161]
- [10]. Sun L; Hebert AS; Yan X; Zhao Y; Westphall MS; Rush MJ; Zhu G; Champion MM; Coon JJ; Dovichi NJ *Angew. Chem. Int. Ed.* 2014, 53, 13931–3.
- [11]. Lombard-Banek C; Moody SA; Nemes P *Angew. Chem. Int. Ed.* 2016, 55, 2454–8.
- [12]. Guo X; Fillmore TL; Gao Y; Tang K *Anal. Chem.* 2016, 88, 4418–25. [PubMed: 27028594]
- [13]. Shen X; Sun L *Proteomics* 2018, 18, e1700432. [PubMed: 29577644]
- [14]. Chen D; Shen X; Sun L *Analyst* 2017, 142, 2118–27. [PubMed: 28513658]
- [15]. Krokhn OV; Anderson G; Spicer V; Sun L; Dovichi NJ *Anal. Chem.* 2017, 89, 2000–2008. [PubMed: 28208305]
- [16]. Moini M *Anal. Chem.* 2007, 79, 4241–6. [PubMed: 17447730]
- [17]. Wojcik R; Dada OO; Sadilek M; Dovichi NJ *Rapid Commun. Mass Spectrom.* 2010, 24, 2554–60. [PubMed: 20740530]
- [18]. Sun L; Zhu G; Zhao Y; Yan X; Mou S; Dovichi NJ *Angew. Chem. Int. Ed.* 2013, 52, 13661–4.
- [19]. Sun L; Zhu G; Zhang Z; Mou S; Dovichi NJ *J. Proteome. Res.* 2015, 14, 2312–21. [PubMed: 25786131]
- [20]. Scheltema RA; Hauschild JP; Lange O; Hornburg D; Denisov E; Damoc E; Kuehn A; Makarov A; Mann M *Mol. Cell. Proteomics* 2014, 13, 3698–708. [PubMed: 25360005]
- [21]. Hebert AS; Richards AL; Bailey DJ; Ulbrich A; Coughlin EE; Westphall MS; Coon JJ *Mol. Cell. Proteomics* 2014, 13, 339–47. [PubMed: 24143002]
- [22]. Busnel JM; Schoenmaker B; Ramautar R; Carrasco-Pancorbo A; Ratnayake C; Feitelson JS; Chapman JD; Deelder AM; Mayboroda OA *Anal. Chem.* 2010, 82, 9476–83. [PubMed: 21028888]
- [23]. Zhang Z; Sun L; Zhu G; Yan X; Dovichi NJ *Talanta* 2015, 138, 117–122. [PubMed: 25863379]
- [24]. Zhu G; Sun L; Yan X; Dovichi NJ *Anal. Chem.* 2014, 86, 6331–6. [PubMed: 24852005]
- [25]. Yan X; Sun L; Zhu G; Cox OF; Dovichi NJ *Proteomics* 2016, 16, 2945–2952. [PubMed: 27723263]
- [26]. Faserl K; Sarg B; Sola L; Lindner HH *Proteomics* 2017, 17, doi: 10.1002/pmic.201700310. [PubMed: 29275045]
- [27]. Chen D; Shen X; Sun L *Anal. Chim. Acta* 2018, 1012, 1–9. [PubMed: 29475469]
- [28]. Kulak NA; Geyer PE; Mann M *Mol. Cell. Proteomics* 2017, 16, 694–705. [PubMed: 28126900]
- [29]. Zhang Z; Yan X; Sun L; Zhu G; Dovichi NJ *Anal. Chem.* 2015, 87, 4572–7. [PubMed: 25822566]
- [30]. Zhang Z; Sun L; Zhu G; Cox OF; Huber PW; Dovichi NJ *Anal. Chem.* 2016, 88, 877–82. [PubMed: 26670623]
- [31]. Choi SB; Lombard-Banek C; Muñoz-LLancao P; Manzini MC; Nemes PJ *Am. Soc. Mass. Spectrom.* 2018, 29, 913–922.
- [32]. Aebersold R; Morrison HD *J. Chromatogr.* 1990, 516, 79–88. [PubMed: 2286630]
- [33]. Britz-McKibbin P; Chen DD *Anal. Chem.* 2000, 72, 1242–52. [PubMed: 10740866]
- [34]. Lubeckjy RA; McCool EN; Shen X; Kou Q; Liu X; Sun L *Anal. Chem.* 2017, 89, 12059–12067. [PubMed: 29064224]
- [35]. McCool EN; Lubeckjy RA; Shen X; Chen D; Kou Q; Liu X; Sun L *Anal. Chem.* 2018, 90, 5529–5533. [PubMed: 29620868]
- [36]. Zhu G; Sun L; Dovichi NJ *Talanta* 2016, 146, 839–43. [PubMed: 26695337]
- [37]. Sun B; Kovatch JR.; Badioung A; Merboudh NJ *Proteome. Res.* 2017, 16, 3711–3721.

- [38]. Keller A; Nesvizhskii AI; Kolker E; Aebersold R *Anal. Chem.* 2002, 74, 5383–92. [PubMed: 12403597]
- [39]. Elias JE; Gygi SP *Nat. Methods* 2007, 4, 207–14. [PubMed: 17327847]
- [40]. Cox J; Mann M *Nat. Biotechnol.* 2008, 26, 1367–1372. [PubMed: 19029910]
- [41]. Cox J; Hein MY; Luber CA; Paron I; Nagaraj N; Mann M *Mol. Cell Proteomics* 2014, 13, 2513–26. [PubMed: 24942700]
- [42]. Tyanova S; Temu T; Sinitcyn P; Carlson A; Hein MY; Geiger T; Mann M; Cox J *Nat. Methods* 2016, 13, 731–740. [PubMed: 27348712]
- [43]. Zhang Z; Peuchen EH; Dovichi NJ *Anal. Chem.* 2017, 89, 6774–6780. [PubMed: 28540730]
- [44]. Cohen D; Dickerson JA; Whitmore CD; Turner EH; Palcic MM; Hindsgaul O; Dovichi NJ *Annu. Rev. Anal. Chem.* 2008, 1, 165–90.
- [45]. Vizcaíno JA; Csordas A; del-Toro N; Dianes JA; Griss J; Lavidas I; Mayer G; Perez-Riverol Y; Reisinger F; Ternent T; Xu QW; Wang R; Hermjakob H *Nucleic. Acids. Res.* 2016, 44, 447–456.
- [46]. Deutsch EW; Csordas A; Sun Z; Jarnuczak A; Perez-Riverol Y; Ternent T; Campbell DS; Bernal-Llinares M; Okuda S; Kawano S; Moritz RL; Carver JJ; Wang M; Ishihama Y; Bandeira N; Hermjakob H; Vizcaíno JA *Nucleic. Acids. Res.* 2017, 45, 1100–1106.

Figure 1.

The calibration curve data. (A) Diagram of dynamic pH junction based CZE-MS/MS. The sample was dissolved in 20 mM NH_4HCO_3 (pH 8.0) and the BGE was 5% (v/v) acetic acid (pH 2.4). Roughly 25% of the separation capillary was filled with the sample for CZE-MS/MS. (B) The number of protein and peptide identifications as a function of the peptide amount in the sample vial (log-log plot). The peptide amount injected for CZE-MS/MS analysis was labeled on the top. The error bars represent the standard deviations of the number of identifications from CZE-MS/MS in triplicate. (C) The peptide intensity as a function of the peptide amount in the sample vial (log-log plot). Twelve peptides were randomly chosen from that identified with the 0.1-ng sample for this figure. The peptide intensity was obtained based on the extracted peptide peaks with 20-ppm mass tolerance and Gaussian smoothing (5 points). The error bars represent the standard deviations of the peptide intensity from CZE-MS/MS in triplicate.

Figure 2.

The C18 ZipTip-CZE-MS/MS data. (A) Protein IDs, Peptide IDs, peptide-spectrum matches (PSMs), and the acquired MS/MS spectra. The error bars represent the standard deviations of the IDs from the platform in triplicate. (B) The protein-level overlap between the C18 ZipTip-CZE-MS/MS replicates. (C) The correlations of protein LFQ intensity between the C18 ZipTip-CZE-MS/MS replicates. (D) The number of peptide IDs in each C18 ZipTip fraction and the number of common peptides between adjacent fractions. “FT” represents the flow through during the sample loading. The 6%, 15%, 20%, 25%, 40%, and 80% represent the concentration of ACN for peptide elution from the ZipTip.

Figure 3. The nanoRPLC fractionation-CZE-MS/MS data. The data was from the 20 RPLC fractions. The bigger RPLC fraction number indicates longer retention time in RPLC. (A) The number of peptide IDs in each nanoRPLC fraction and the number of common peptides between adjacent fractions. (B) The cumulative distribution of the GRAVY of the identified peptides in 10 out of the 20 nanoRPLC fractions. The fraction number was labeled in the figure. Positive GRAVY values indicate hydrophobic and negative values show hydrophilic. (C) The correlation between the RPLC fraction number and the migration time of peptides in CZE. All the identified peptide-spectrum matches (PSMs) were used to generate this figure. The migration time corresponding to each PSM was obtained directly from the database search result.