Supporting Information I

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, Liangliang Sun*

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

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

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 (μRPLC) 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-μg peptides. The dynamic pH junction based CZE enabled a 500-nL sample injection from as low as a 1.5-μL peptide sample, using up to 33% of the available peptide material for an analysis. Two kinds of μRPLC prefractionation were investigated, C18 ZipTip and nanoflow RPLC. C18 ZipTip-CZE-MS/MS identified 4453 proteins from 5 μg 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-μg 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.

Table of content

Experimental Section

Materials and Reagents

All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless stated otherwise. LC/MS grade water, methanol, acetonitrile (ACN), HPLC grade acetic acid (AA), formic acid (FA), and hydrofluoric acid (HF) were purchased from Fisher Scientific (Pittsburgh, PA). Urea was purchased from Alfa Aesar (Haverhill, MA). Acrylamide was ordered from Acros Organics (NJ, USA). Fused silica capillaries (50 μm i.d./360 μm o.d.) were purchased from Polymicro Technologies (Phoenix, AZ).

Preparation of the MCF7 breast cancer cell proteome digest

The human breast cancer cells (MCF-7) were kindly provided by Dr. Xuefei Huang's group at the Department of Chemistry, Michigan State University. MCF7 cells were cultured at 37°C under a 5% CO_2 in ATCC-formulated Eagle's Minimum Essential Medium, supplemented with 0.01 mg/ml human recombinant insulin and 10% fetal bovine serum. The cells were washed with PBS and lysed in a lysis buffer containing 8 M urea, 100 mM ammonia bicarbonate (NH_4HCO_3, pH 8.0) and protease inhibitors, with the assistance of ultrasonication (Branson Sonifier 250, VWR Scientific, Batavia, IL) on ice for 10 minutes. After centrifugation at 10000 g for 10 min, the supernatant was collected and the protein concentration was measured by the BCA assay.

250 μg of MCF-7 proteins were precipitated with cold acetone and stored at -20 ^{o}C overnight. After centrifugation at 14000 g for 10 min, the protein pellet was washed with cold acetone once and was air-dried for a couple of minutes at room temperature. The protein pellet was then redissolved in 100 μL of 8 M urea containing 100 mM NH_4HCO_3 (pH 8.0). The proteins were reduced by dithiothreitol (DTT) at 37^{o}C for 30 min and alkylated by iodoacetamide (IAA) at room temperature for 20 min in the dark. DTT was added to quench extra IAA. The protein sample was then diluted to 500 μL with 100 mM NH_4HCO_3 (pH 8.0), followed by tryptic digestion at 37°C for overnight with 8 μg of trypsin (Bovine pancreas TPCK-treated). The digestion process was terminated by adding 2 μL of formic acid. Peptides were then desalted using Sep-Pak C18 Cartridge

(Waters) and lyophilized with a vacuum concentrator (Thermo Fisher Scientific). The peptide sample was stored at -80°C before use.

Nano 2D-RPLC-MS/MS analysis

A C18 RPLC column (75 μm i.d. x 50 cm, C18, 2 μm, 100 Å, Thermo Fisher Scientific) connected to an EASY nanoLC-1200 system (Thermo Fisher Scientific) was used for the first dimensional high-pH RPLC separation. Buffer A containing 5 mM ammonium bicarbonate (pH 9.0) and buffer B containing 80% (v/v) ACN and 5 mM ammonium bicarbonate (pH 9.0) were used to generate gradient separation. 5-μg of the MCF7 protein 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. 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 5 min. The last fraction was collected starting from 155 min until the end of the gradient. We collected 30 fractions in total and combined fractions into 15 fractions. All fractions were lyophilized and each fraction was redissolved in 5 μL of 0.1 % (v/v) FA for low pH nanoRPLC-MS/MS.

80% of the peptides in each fraction were loaded on the analytical column (75 μm i.d. x 50 cm, C18, 2 μm, 100 Å, Thermo Fisher Scientific) for low pH RPLC separation. An EASY nanoLC-1200 (Thermo Fisher Scientific) was used. Buffer A containing 0.1% (v/v) FA and buffer B containing 80% (v/v) ACN and 0.1% (v/v) FA were used to generate gradient separation. Sample 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. The gradient for RPLC separation was as follows: from 5 to 20% (v/v) B in 60 min, from 20% to 40% (v/v) B in 40 min, from 40% to 100% (v/v) B in 10 min and maintain at 100% (v/v) B for 10 min.

The parameter settings of each nanoLC-MS/MS analysis were all consistent with CZE-MS/MS analysis except 120 min was set for data acquisition.

Proteome Discoverer 2.2 (Thermo Fisher Scientific) was used for the database search. SEQUEST HT searching engine was used. All database searching parameters were kept same with CZE-MS/MS database searching.

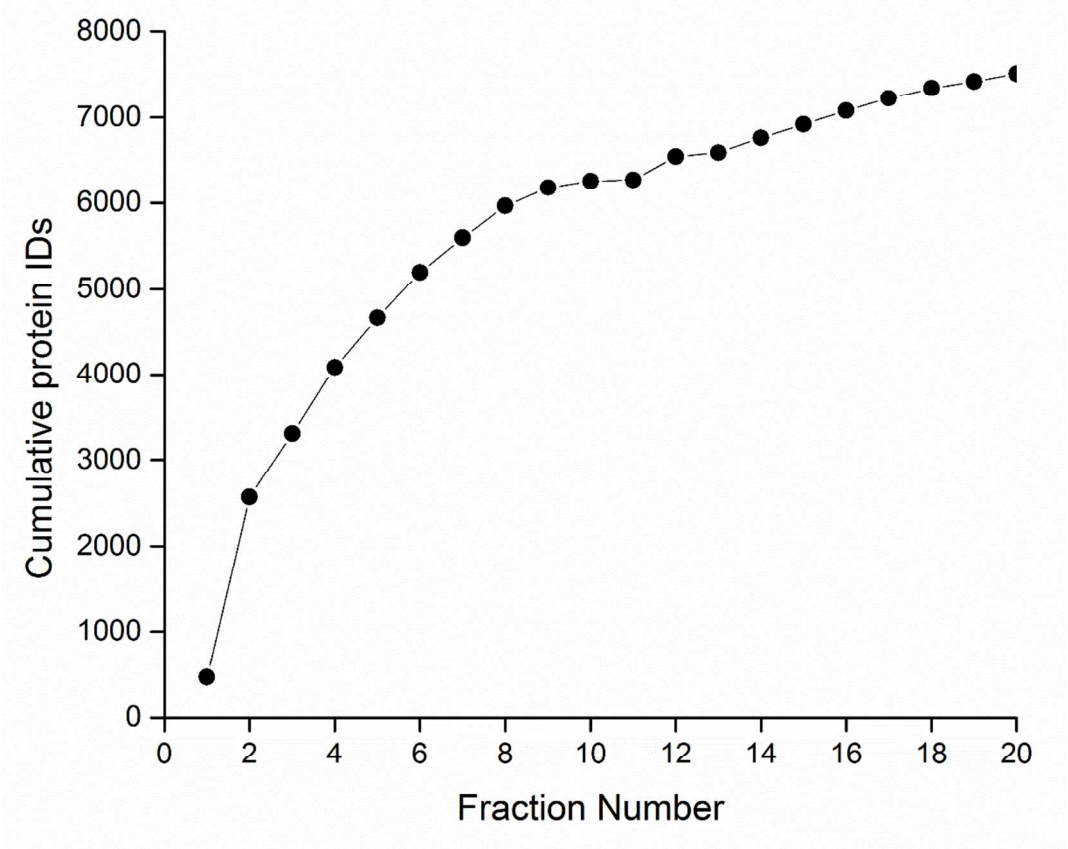

Figure S1. Cumulative protein IDs *vs.* number of RPLC fractions. The data from the nanoRPLC-CZE-MS/MS experiment (20 RPLC fractions) were used for the figure.

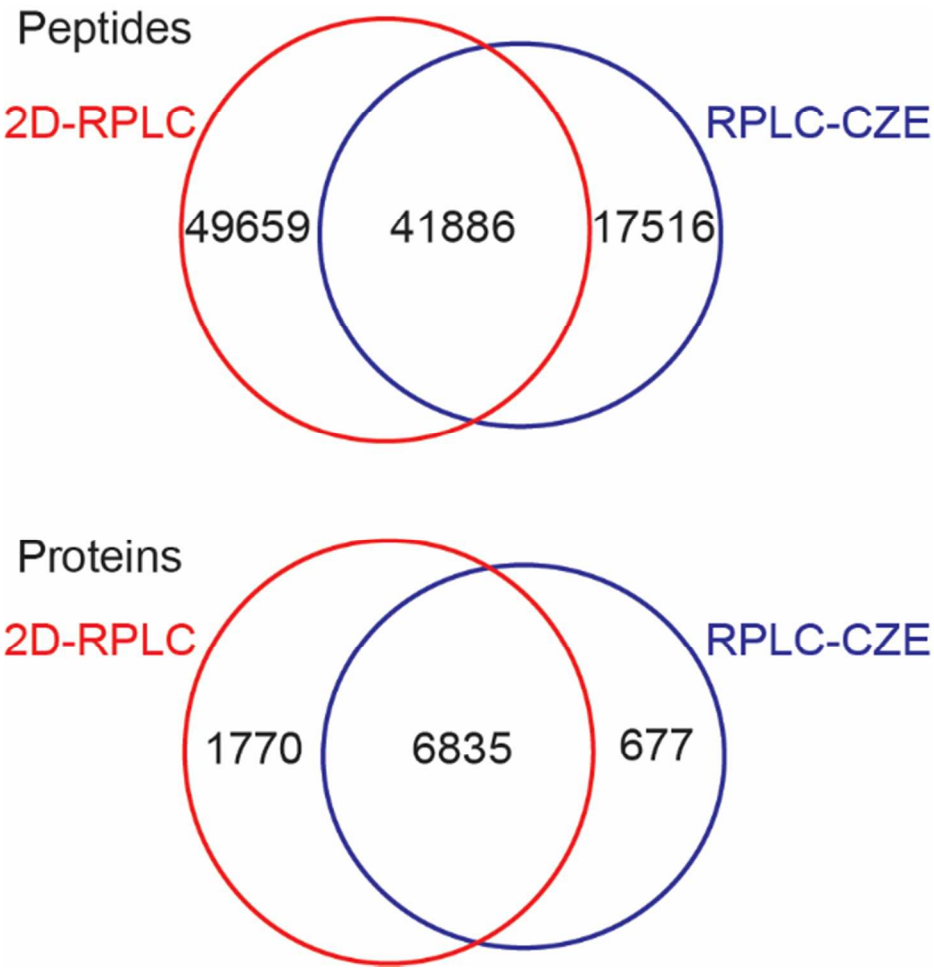

Figure S2. Comparisons between nano-2D-RPLC-MS/MS and nanoRPLC-CZE-MS/MS in terms of the protein-level and peptide-level overlaps.