Analytica Chimica Acta 1012 (2018) 1–9

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Strong cation exchange-reversed phase liquid chromatography-capillary zone electrophoresis-tandem mass spectrometry platform with high peak capacity for deep bottom-up proteomics

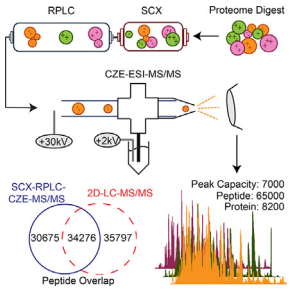
Daoyang Chen, Xiaojing Shen, Liangliang Sun*

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

HIGHLIGHTS

- CZE-MS/MS can identify over 8000 protein groups from a complex proteome.
- CZE-MS/MS and LC-MS/MS are comparable for deep proteomic sequencing.
- Combining CZE-MS/MS and LC-MS/MS can significantly boost the proteome coverage.
- The peak capacity of SCX-RPLC-CZE platform can reach ~7000.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 17 December 2017
Received in revised form
19 January 2018
Accepted 22 January 2018
Available online 5 February 2018

Keywords:
Strong cation exchange
Reversed-phase liquid chromatography
Capillary zone electrophoresis
Tandem mass spectrometry
Bottom-up proteomics

ABSTRACT

Two-dimensional (2D) liquid chromatography (LC)-tandem mass spectrometry (MS/MS) are typically employed for deep bottom-up proteomics, and the state-of-the-art 2D-LC-MS/MS has approached over 8000 protein identifications (IDs) from mammalian cell lines or tissues in 1–3 days of mass spectrometer time. Capillary zone electrophoresis (CZE)-MS/MS has been suggested as an alternative to LC-MS/MS for bottom-up proteomics. CZE-MS/MS and LC-MS/MS are complementary in protein/peptide ID from complex proteome digests because CZE and LC are orthogonal for peptide separation. In addition, the migration time of peptides from CZE-MS can be predicted accurately, which is invaluable for evaluating the confidence of peptide ID from the database search and even guiding the database search. However, the number of protein IDs from complex proteomes using CZE-MS/MS is still much lower than the state of the art using 2D-LC-MS/MS. In this work, for the first time, we established a strong cation exchange (SCX)-reversed phase LC (RPLC)-CZE-MS/MS platform for deep bottom-up proteomics. The platform identified around 8200 protein groups and 65,000 unique peptides from a mouse brain proteome digest in 70 h. The data represents the largest bottom-up proteomics dataset using CZE-MS/MS and provides a valuable resource for further improving the tool for prediction of peptide migration time in CZE. The peak capacity of the orthogonal SCX-RPLC-CZE platform was estimated to be around 7000. SCX-RPLC-CZE-MS/MS produced comparable numbers of protein and peptide IDs with 2D-LC-MS/MS (8200 *vs.* 8900 protein groups, 65,000 *vs.* 70,000 unique peptides) from the mouse brain proteome digest using comparable instrument time. This is the first time that CZE-MS/MS showed its capability to approach comparable performance to the state-of-the-art 2D-LC-MS/MS for deep proteomic sequencing. SCX-RPLC-CZE-MS/MS and 2D-LC-MS/MS showed good complementarity in protein and peptide IDs and combining those two

Abbreviations: ID, identification; LPA, linear polyacrylamide; FASP, filter-aided sample preparation.

* Corresponding author.
E-mail address: lsun@chemistry.msu.edu (L. Sun).

https://doi.org/10.1016/j.aca.2018.01.037

methods improved the number of protein group and unique peptide IDs by nearly 10% and over 40%, respectively, compared with 2D-LC-MS/MS alone.

1. Introduction

The state-of-the-art two-dimensional (2D) liquid chromatography (LC)-tandem mass spectrometry (MS/MS) has approached over 8000 protein identifications (IDs) from mammalian cell lines or tissues in 1–3 days of mass spectrometer time [1–5]. The draft human proteome containing 84% of the total annotated protein-coding genes in humans have also been generated using 2D-LC-MS/MS [6]. Over 2000 LC-MS runs were performed for the draft human proteome, but the median protein sequence coverage was still only 28% [6]. The typical median protein sequence coverage of deep bottom-up proteomics datasets are around 25% or lower. The low sequence coverage impedes the confident identification of protein isoforms.

Alternative separation techniques that are orthogonal to LC for peptide separation will be very useful to further improve the number of peptide IDs from complex proteomes in bottom-up proteomics experiments, boosting the protein sequence coverage. Capillary zone electrophoresis (CZE)-MS/MS has been suggested as an alternative to LC-MS/MS for bottom-up proteomics [7–18]. CZE separates peptides based on their size-to-charge ratios and it is orthogonal to LC for peptide separation. CZE-MS/MS and LC-MS/MS are complementary in protein/peptide ID from complex proteome digests [7–11]. CZE tends to identify small, basic and hydrophilic peptides compared with RPLC-MS. In addition, the migration time of peptides from CZE-MS can be predicted more accurately and easily than their retention time from commonly used reversed-phase LC (RPLC)-MS [19]. The electrophoretic mobility of peptides in CZE mainly relate to their size (molecular mass) and charge, which are relatively easy to be determined accurately. The retention of peptides in RPLC can be affected by various factors, *e.g.*, hydrophobic, hydrogen-bond, and ion-pairing interactions. Modeling those factors are very difficult. Recently, Krokhin et al. developed a simple model for CZE and approached very good correlation (R^2 ~0.995) between the experimental and predicted migration time of peptides in CZE based on a large-scale peptide dataset [19]. The capability for accurate prediction of peptide migration time in CZE makes CZE-MS become a powerful tool for bottom-up proteomics because it can help us further evaluate the confidence of peptide ID from the database search and even guide the database search.

Although CZE-MS has many valuable features for bottom-up proteomics, the number of protein IDs from complex proteomes using CZE-MS/MS is still much lower than the state of the art using 2D-LC-MS/MS. Much effort has been made to improve the CZE-MS for large-scale proteomics [7,15,20,21]. Sun et al. approached 2000 protein and 10,000 peptide IDs from a human cell line digest using single shot CZE-MS/MS with a neutrally coated separation capillary and an Orbitrap Fusion mass spectrometer [7]. Field-enhanced sample stacking was used to improve the sample loading volume to 100 nL and a 1-m long neutrally coated separation capillary was employed to improve the peak capacity to about 300 [7]. Yan et al. coupled RPLC prefractionation to CZE-MS/MS for bottom-up proteomics of *Xenopus* embryos, resulting in the identification of over 4000 proteins [20]. For each CZE-MS/MS run, about 50 nL of the sample was injected for analysis. Faserl et al. coupled RPLC prefractionation to CZE-MS/MS for quantitative proteomics of yeast, leading to the identification of over 3000 proteins [15]. A 1.5-mg yeast digest was used as the starting material and the sample loading volume for CZE-MS/MS was 40 nL. Very recently, Faserl et al. approached 6000 protein IDs from a human cell line proteome digest by RPLC prefractionation and sequential sample injection based CZE-MS/MS with 2 mg of peptides as the starting material [21]. The sample loading volume of CZE-MS/MS was 25 nL.

In order to further improve the CZE-MS/MS for significantly deeper proteome coverage with a reasonable mass of initial protein material, we need to improve the sample loading volume of CZE-MS/MS and meantime boost the overall peak capacity of the system. The improvement in both sample loading volume and peak capacity can evidently benefit the identification of low abundant proteins. Recently, we showed that dynamic pH junction based CZE-MS/MS could approach both micro-liter scale sample loading volume and high peak capacity (up to 380) for analysis of complex peptide or protein mixtures [22,23]. In this work, we coupled online strong cation exchange (SCX)-RPLC prefractionation to the dynamic pH junction based CZE-MS/MS for deep bottom-up proteomics. The orthogonal SCX-RPLC-CZE platform approached very high peak capacity (~7000). Because of the high peak capacity and the large sample loading volume of CZE (~0.5 μL per run), the SCX-RPLC-CZE-MS/MS system identified 8200 protein groups and 65,000 unique peptides from a mouse brain proteome digest.

2. Experimental section

2.1. Reagents and chemicals

See supporting material I for details.

2.2. Preparation of the linear polyacrylamide-coated capillary for CZE

The inner wall of the separation capillaries for CZE was coated with linear polyacrylamide (LPA) based on references [22] and [24] in order to reduce the electroosmotic flow (EOF). The detailed protocol was described in supporting material I. After that, one end of the LPA-coated capillary was etched with hydrofluoric acid (HF) to reduce its outer diameter to ~70 μm based on the protocol in Ref. [14]. The LPA-coated capillary was stored at room temperature before use.

2.3. Sample preparation

See supporting material I for details.

2.4. Online SCX-RPLC fractionation of a mouse brain proteome digest

An Agilent Infinity II HPLC system with a quaternary pump was used for the experiment. A SCX trap column (Zorbax 300SCX, 4.6 mm i.d. × 12.5 mm length, 5 μm particles, Agilent Technologies) and a C18 RP column (Zorbax 300Extend-C18, 2.1 mm i.d. × 150 mm length, 3.5 μm particles, Agilent Technologies) were directly connected with a PEEK tubing and two fittings for online 2D-LC separation. 0.1% formic acid (FA) in water (mobile phase A),