Published in final edited form as:

Angew Chem Int Ed Engl. 2013 December 16; 52(51): 13661–13664. doi:10.1002/anie.201308139.

Ultrasensitive and fast bottom-up analysis of femtogram amounts of complex proteome digests**

Dr. Liangliang Sun, Dr. Guijie Zhu, Yimeng Zhao, Xiaojing Yan, Si Mou, and Prof. Dr. Norman J. Dovichi

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556(USA)

Norman J. Dovichi: ndovichi@nd.edu

Abstract

Femtogram proteomics: We report an ultrasensitive capillary zone electrophoresis-mass spectrometry system based on an improved nanospray interface. This system is used for analysis of picogram to femtogram amounts of *E. coli* digests. Over 100 proteins were identified based on tandem mass spectra from 16 pg digests; over 60 proteins were identified from 400 fg digests based on accurate mass and time tags in 10 min.

Keywords

electrophoresis; electrospray; proteomics; ultrasensitive analysis

Bottom-up proteomics is widely used for qualitative and quantitative characterization of complex biological samples.^[1,2] Given micrograms of material, it is possible to identify more than 10,000 proteins from mammalian cell lysates and over 2,500 proteins from prokaryote lysates.^[3,4] The performance of bottom-up proteomics degrades rapidly for mass-limited samples, such as laser capture microdissected tissues, circulating tumor cells, single embryos, and single somatic cells. There have been a handful of reports of bottom-up proteomics of nanogram samples using capillary liquid chromatography (LC)-electrospray ionization (ESI)-tandem mass spectrometry (MS/MS). Mann's group identified 2,000 proteins from single pancreatic islets with protein content of several hundred ng.^[5] Karger's group identified 566 proteins from 50 ng of digest of *Methanosarcina acetivorans*^[6] and 163 proteins from ~2.5 ng of the tryptic digest of a cervical cancer cell line.^[7] Smith's group detected 870 proteins with an accurate mass and time tags (AMTs)^[8] strategy from low nanogram amounts of the digest of *Deinococcus radiodurans*. Smith's group also reported the detection of the three most abundant proteins in a 0.5 pg sample with the AMTs method, and reported a ~10 z mole detection limit for one peptide in a bovine serum albumin digest.^[9] Our group used a Q-Exactive mass spectrometer with higher energy collisional dissociation^[10] to identify ~100 protein groups from 1 ng of a digest of the RAW264.7 macrophage cell line.^[11] All of these analyses required at least one hour of instrument time. In this paper, we report an ultrasensitive and fast capillary zone electrophoresis (CZE)-ESI-MS/MS system that is based on an improved electrokinetically-pumped sheath-flow

**We thank Dr. William Boggess in the Notre Dame Mass Spectrometry and Proteomics Facility for his help with this project. This work was funded by the National Institutes of Health (R01GM096767).

Correspondence to: Norman J. Dovichi, ndovichi@nd.edu.

Supporting information for this article is available on the WWW under <http://www.angewandte.org>

interface. We demonstrate the system for the rapid bottom-up analysis of femtogram amounts of the *E. coli* protein digest.

CZE-ESI-MS/MS has attracted attention for bottom-up proteomics,^[12] and this approach consistently outperforms LC-MS/MS for low nanogram samples.^[13–16] The improved performance of CZE for small sample amounts presumably is due to its very simple design, eliminating sample loss on injectors and fittings. Beginning with the pioneering work of Smith's group,^[17] electrospray interfaces have been developed for capillary electrophoresis.^[18] Two recently developed interfaces are of note. One is a sheathless interface based on a very thin porous capillary tip developed by Moini.^[19] We have developed another interface based on an electrokinetically pumped sheath-flow interface, Figure 1A.^[20] Our interface has several advantages, including reduced sample dilution due to a very low sheath flow rate, elimination of mechanical pumps, use of a wide range of separation buffers, and stable operation in the nanospray regime. We recently coupled CZE to a triple-quadrupole mass spectrometer with this interface for quantification of Leu-enkephalin in a complex mixture using multiple-reaction monitoring, and we obtained a 335 z mole peptide detection limit,^[21] suggesting the system's potential for high sensitivity analysis.

A COMSOL model of the electrokinetically pumped sheath-flow interface predicted and experiments verified that sensitivity increases as the distal end of the capillary is brought closer to the emitter orifice.^[20] Typical distances between the capillary tip and orifice are about 1 mm, which is limited by the outer diameter of the separation capillary that butts against the conical emitter wall. In this work, we etched a few millimeters of the outside of the separation capillary tip with hydrofluoric acid to reduce its outer diameter from ~150 μm to ~60 μm . This simple step allows us to place the capillary end much closer to the emitter orifice (~200 μm), Figure 1B and C, which results in a dramatic improvement in the system's sensitivity. We used uncoated fused silica capillaries (32 cm and 40 cm, 10 μm i.d./150 μm o.d.) for electrophoresis, and a Q-Exactive mass spectrometer for peptide identification. Experimental details are provided in the Supporting information.

We first evaluated the effect of separation voltage for the analysis of 28 pg amounts of *E. coli* digests. Separations were performed at 15 kV (500 V/cm) and 10 kV (300 V/cm) in a 32-cm long capillary, Figure S1. The 10 kV potential produced a wider separation window, which resulted in more protein (129 ± 18 vs. 88 ± 14) and peptide (375 ± 27 vs. 246 ± 19) identifications compared with 15 kV. The following work used an electric field of 300 V/cm.

We then evaluated the reproducibility of our CZE-ESI-MS/MS system for analysis of 16 pg of the *E. coli* protein digests with a 40 cm capillary. We identified 105 ± 17 proteins and 256 ± 9 peptides based on triplicate bottom-up analysis of tandem mass spectra. The state of the art for tandem mass spectra analysis of complex protein digests is ~100 protein identifications at the 1 ng level.^[7,9,11,13,16] Our system produces similar number of protein identifications from two-orders of magnitude less sample.

The separations were reproducible and efficient. The signals from 50 peptides were summed to produce extracted ion electropherograms, Figure 2. The average relative standard deviation of the migration time of 154 peptides was 0.7%, Figure S2. The electrophoretic peaks were quite sharp, with an average width, defined as the standard deviation of the Gaussian function used to fit the peaks, of 0.7 s (1.6 s full width at half height), Figure S3. We consistently obtained an average of over 300,000 theoretical plates for the peptide separations, Figure S4. Peak intensity was also consistent between runs, Figure S5. Separations were complete in less than 10-min, which is an order of magnitude

improvement in analysis time compared to the state-of-the-art for high sensitivity bottom-up proteomics of complex proteomes.

We next determined the relationship between the number of identifications based on tandem mass spectra and the loaded amounts of *E. coli* digests, Figure 3. In duplicate 400 fg loadings, nine peptides corresponding to 4 ± 1 proteins were confidently identified after manual evaluation of tandem mass spectra, Figure S6. The most abundant protein in *E. coli*, elongation factor Tu, makes up ~1% of the total protein mass.^[22] An analysis of 400 fg of an *E. coli* digest will contain 4 fg of this protein, and the minimum protein amount for identification by tandem mass spectrometry is less than 4 fg, representing an improvement of two orders of magnitude in the state of art.^[9] When the sample loading amount was increased to 84 pg, the number of protein and peptide identifications increased to 162 ± 8 and 570 ± 11 , respectively, Figure 3.

We also applied Smith's accurate mass and time tag approach using the 16 pg *E. coli* data as the database. Over 60 proteins and 150 peptides were identified from the 400 fg *E. coli* digests with mass tolerance as 3 ppm, migration time tolerance as 0.3 min (without alignment), and at least two detected isotopic peaks for each peptide, Figure 3. The extracted ion electropherograms are presented in Figure S7. This result is a 20-times improvement in the number of protein identifications in the state of art for AMTs based subpicogram proteome analysis.^[9]

We finally estimated the peptide detection limit from the 400 fg *E. coli* data. We manually extracted electropherograms for three peptides from elongation factor Tu, which were identified based on MS/MS spectra with mass tolerance of 1 ppm. The signal-to-noise ratios were obtained with Xcalibur software (Thermo Fisher Scientific), using the noise region from 0.3 min to 2.3 min after the peak, Figure S8. Based on the amount of elongation factor Tu present in the sample (4 fg) and its molecular weight (~43 kDa), ~100 z mole of these peptides were taken for analysis. These peptides generated signal-to-noise ratios (S/N) of 270~290; the mass detection limit (S/N = 3) is ~1 z mole (~600 molecules), which is a one order of magnitude improvement in the state of art for MS based peptide detection.^[9]

There are several possible reasons for the high sensitivity obtained from the CZE-MS system. First, the peptides only take around 0.2 s or less to migrate from the capillary end to the spray emitter end (approximately calculated based on our previous work^[20]), which dramatically reduces the sample diffusion in the spray emitter and generates higher peptide signal, resulting in better peptide detection limits. Second, we can approximately assume that the electro osmotic flow rate in the separation capillary and in the spray emitter is same due to similar buffer and applied voltage (~300 V/cm), and the total flow rate for spray is around 20 nL/min, which should generate high ionization efficiency, resulting in high sensitivity. Third, we employ quite narrow inner diameter capillaries, which reduces sample flow rate and generates very efficient separations.

We also note that the number of protein identifications obtained in this work is ~200 proteins due to the relatively short peptide separation window, limiting the number of acquired tandem spectra, which limits the detection of relatively low abundance proteins in biological samples. One way to improve the protein identifications based on the CZE-MS system is to perform online/offline peptide pre-fractionation before the CZE-MS analysis or to employ coated capillaries to reduce electro-osmosis and increase the separation window.^[16, 21, 23]

In summary, this work demonstrates an ultrasensitive and high throughput CZE-ESI-MS/MS system for femtogram proteomics analysis. The results are a one- to two-orders of

magnitude improvement in the amount of material required for protein identification by tandem mass spectra, in the number of proteins identified by accurate mass and time tags from sub-picogram amounts of a complex protein digest, in peptide mass detection limit, and in analysis time. Faster and more sensitive mass spectrometers will enable further improvements. One obvious application of this technology will be single cell analysis. To date, the highest sensitivity tools available for single cell protein analysis have employed laser-induced fluorescence detection;^[24] while sensitive, fluorescence inherently generates a low information content signal that provides only rudimentary information on protein identity. The development of CZE-mass spectrometry systems with 1-zmol detection limits opens the door to single-cell protein analysis with confident identification of relatively high abundance proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Gstaiger M, Aebersold R. *Nat Rev Genet.* 2009; 10:617–627. [PubMed: 19687803]
2. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR III. *Chem Rev.* 2013; 113:2343–2394. [PubMed: 23438204]
3. Geiger T, Wehner A, Schaab C, Cox J, Mann M. *Mol Cell Proteomics.* 2012; 11:10.1074/mcp.M111.014050
4. Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y. *Anal Chem.* 2010; 82:2616–2620. [PubMed: 20222674]
5. Waanders LF, Chwalek K, Monetti M, Kumar C, Lammert E, Mann M. *Proc Natl Acad Sci USA.* 2009; 106:18902–18907. [PubMed: 19846766]
6. Yue G, Luo Q, Zhang J, Wu S-L, Karger BL. *Anal Chem.* 2007; 79:938–946. [PubMed: 17263319]
7. Luo Q, Yue G, Valaskovic GA, Gu Y, Wu S-L, Karger BL. *Anal Chem.* 2007; 79:6174–6181. [PubMed: 17625912]
8. Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen Y, Conrads TP, Veenstra TD, Udseth HR. *Proteomics.* 2002; 2:513–523. [PubMed: 11987125]
9. Shen Y, Toli N, Masselon C, Paša-Toli L, Camp DG II, Hixson KK, Zhao R, Anderson GA, Smith RD. *Anal Chem.* 2004; 76:144–154. [PubMed: 14697044]
10. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. *Nat Methods.* 2007; 4:709–712. [PubMed: 17721543]
11. Sun L, Zhu G, Dovichi NJ. *Rapid Commun Mass Spectrom.* 2013; 27:157–162. [PubMed: 23239329]
12. Sun L, Zhu G, Yan X, Dovichi NJ. *Curr Opin Chem Biol.* 2013; 23:1016/j.cbpa.2013.07.018
13. Wang Y, Fonslow BR, Wong CCL, Nakorchevsky A, Yates JR III. *Anal Chem.* 2012; 84:8505–8513. [PubMed: 23004022]
14. Sun L, Zhu G, Li Y, Wojcik R, Yang P, Dovichi NJ. *Proteomics.* 2012; 12:3013–3019. [PubMed: 22888077]
15. Faserl K, Sarg B, Kremser L, Lindner H. *Anal Chem.* 2011; 83:7297–7305. [PubMed: 21848273]
16. Zhu G, Sun L, Yan X, Dovichi NJ. *Anal Chem.* 2013; 85:2569–2573. [PubMed: 23394296]
17. Smith RD, Barinaga CJ, Udseth HR. *Anal Chem.* 1988; 60:1948–1952.
18. Maxwell EJ, Chen DDY. *Anal Chim Acta.* 2008; 627:25–33. [PubMed: 18790125]
19. Moini M. *Anal Chem.* 2007; 79:4241–4246. [PubMed: 17447730]
20. Wojcik R, Dada OO, Sadilek M, Dovichi NJ. *Rapid Commun Mass Spectrom.* 2010; 24:2554–2560. [PubMed: 20740530]
21. Li Y, Wojcik R, Dovichi NJ, Champion MM. *Anal Chem.* 2012; 84:6116–6121. [PubMed: 22690842]

22. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, von Mering C. *Mol Cell Proteomics*. 2012; 11:492–500. [PubMed: 22535208]
23. Yan X, Essaka DC, Sun L, Zhu G, Dovichi NJ. *Proteomics*. 2013; 13:2546–2551. [PubMed: 23798545]
24. Cohen D, Dickerson JA, Whitmore CD, Turner EH, Palcic MM, Hindsgaul O, Dovichi NJ. *Annu Rev Anal Chem*. 2008; 1:165–190.

Figure 1. CZE-ESI-MS/MS system. Sketch of the system (A), sketch of the etched capillary in the electrospray emitter (B), and micrograph of the etched capillary in the emitter (C).

Figure 2.

Extracted ion electropherograms of 50 high intensity peptides identified based on tandem spectra from 16 pg amounts of *E. coli* digests analyzed by CZE-ESI-MS/MS in triplicates. The mass tolerance for extraction was 2 ppm.

Figure 3. Relationship between loaded amounts of *E. coli* digests and identifications based on tandem spectra (points connected by lines) and accurate mass and time tags (AMTs). Protein identifications (A); peptide identifications (B). Each sample was analyzed in duplicate or triplicate. The identifications based on AMTs from 400 fg amounts of *E. coli* digests were labeled with star. The error bars are standard deviations of the mean.