Proteomics

DOI: 10.1002/ange.201409075

Over 10 000 Peptide Identifications from the HeLa Proteome by Using Single-Shot Capillary Zone Electrophoresis Combined with Tandem Mass Spectrometry**

*Liangliang Sun, Alexander S. Hebert, Xiaojing Yan, Yimeng Zhao, Michael S. Westphall, Matthew J. P. Rush, Guijie Zhu, Matthew M. Champion, Joshua J. Coon, and Norman J. Dovichi**

Abstract: *Capillary zone electrophoresis (CZE)–tandem mass spectrometry (MS/MS) has recently attracted attention as a tool for shotgun proteomics. However, its performance for this analysis has so far fallen far below that of reversed-phase liquid chromatography (RPLC)–MS/MS. The use of a CZE method with a wide separation window (up to 90 min) and high peak capacity (ca. 300) is reported. This method was coupled to an Orbitrap Fusion mass spectrometer through an electrokinetically pumped sheath-flow interface for the analysis of complex proteome digests. Single-shot CZE–MS/MS lead to the identification of over 10 000 peptides and 2100 proteins from a HeLa cell proteome digest in approximately 100 min. This performance is nearly an order of magnitude better than earlier CZE studies and is within a factor of two to four of the state-of-the-art nano ultrahigh-pressure LC system.*

In shotgun proteomics, samples are first digested with enzymes to generate peptides, followed by reversed-phase liquid chromatography (RPLC) separation and tandem mass spectrometry (MS/MS) detection. The resulting MS/MS spectra are then analyzed through correlation against a database.[1–3]

Recently, capillary zone electrophoresis (CZE)–electrospray ionization (ESI)-MS/MS has been suggested as an alternative to RPLC in shotgun proteomics.[4–6] CZE–MS and RPLC–MS generate complementary peptide identifications, and the combination of these two techniques can improve protein sequence coverage.[7–10] CZE–MS/MS can generate superior performance to RPLC–MS/MS for trace peptide analysis.[9,10] However, the performance of CZE–MS/MS for shotgun proteomics has been disappointing. To date, single-shot CZE–MS/MS proteomic analyses have enabled the identification approximately 1000 peptides and 300 proteins,[10,11] which is at least ten times less than the state-of-art RPLC–MS/MS system.[12–15]

Two obstacles limit CZE–MS performance. First, the sample-loading capacity of CZE is typically ten times lower than RPLC; this issue hinders the detection of low-abundance proteins. Second, the separation window in CZE–MS is usually less than 20 min. By contrast, RPLC routinely generates separation windows in excess of 300 min, which provides ample opportunity for peptide identification by MS/MS.

To improve the number of identifications when using single-shot CZE–MS/MS for complex proteome analysis, we report two innovations. First, we employed electrophoretic conditions that generate separation windows of up to 90 min and stacking injection conditions that increase the sample loading amount. Second, we coupled a state-of-the-art mass spectrometer with an efficient and robust CZE electrospray interface. The MS system combines a mass filter, collision cell, high-field Orbitrap analyzer, and a dual cell linear ion trap analyzer geometry (Q-OT-qIT, Orbitrap Fusion) to provide up to 20 Hz MS/MS acquisition speed.[15] The combination of a long separation window with high MS/MS sampling rates generates approximately 50 000 tandem mass spectra in a single CZE separation, thereby producing remarkable improvements in proteomic coverage over earlier reports.

Figure 1 (top trace) presents the CZE–MS/MS analysis of a HeLa cell proteome digest, with the generation of an approximately 90 min wide separation window. To our knowledge, this separation window is the widest reported for the CZE–MS/MS analysis of peptides. We used an unsupervised nonlinear least squares routine to fit a Gaussian function to the extracted ion electropherograms of the 100 most intense peptide peaks in Figure 1 (top trace). The median peak width, measured as the standard deviation of the Gaussian function, was 8.5 s; the median number of theoretical plates was 100 000; and the peak capacity was 300.[16]

The wide separation window results from the use of a linear polyacrylamide-coated capillary and a high concentration of acetic acid [5% (v/v)] in the separation buffer to produce very low electroosmotic flow. Separation efficiency

[*] Dr. L. Sun, X. Yan, Y. Zhao, Dr. G. Zhu, Dr. M. M. Champion, Prof. Dr. N. J. Dovichi
Department of Chemistry and Biochemistry
University of Notre Dame
Notre Dame, IN 46556 (USA)
E-mail: ndovichi@nd.edu
Dr. A. S. Hebert, Dr. M. S. Westphall, M. J. P. Rush, Prof. Dr. J. J. Coon
Department of Biomolecular Chemistry,
Genome Center of Wisconsin, and Department of Chemistry
University of Wisconsin, Madison, WI 53706 (USA)

[**] 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—N.J.D.) and (R01GM080148—J.J.C.) and the National Science Foundation (0701846—J.J.C.).

Supporting information for this article (including experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201409075.

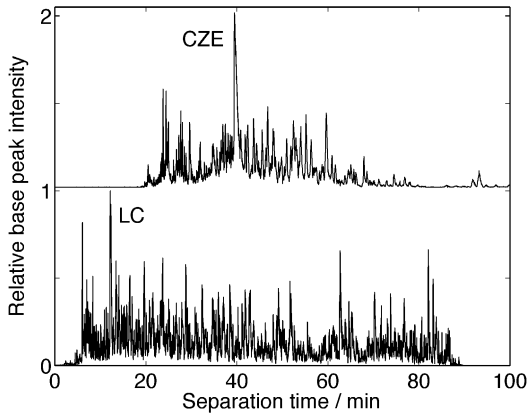

Figure 1. HeLa cell lysate digest data after single-shot CZE–MS/MS (top trace) and RPLC–MS/MS (bottom trace) analysis. Approximately 400 ng and 300 ng of sample were loaded for CZE and RPLC analysis, respectively.

was improved by dissolving the samples in 0.03–0.04 % (v/v) formic acid containing 30–40 % (v/v) acetonitrile. The conductivity of the sample buffer is much lower than that of the separation buffer, thus producing sample stacking and peak sharpening. We used a second-generation electrokinetically pumped sheath-flow interface[17] to couple CZE to the mass spectrometer, which reduced sample diffusion in the spray emitter compared with our first-generation interface.[18]

We used this system to analyze *E. coli*, yeast, and HeLa cell proteomes (Table 1). The single-shot CZE–MS/MS system identified around 1000 proteins and 4700 peptides from an *E. coli* tryptic digest in 1.5 h. This number of peptide and protein identifications (IDs) is three times larger than an earlier report of the use of CZE–MS/MS for studying an unfractionated *E. coli* proteome.[10] The improved performance is a result of the generation of six times more MS/MS spectra over a separation period that is roughly twice as long.

Table 1: A summary of the peptide and protein group identifications (IDs) from single-shot CZE–MS/MS for different samples.

Sample	Peptide IDs[a]	Protein group IDs	Analysis duration [min]
Escherichia coli	4741	956	90
Yeast	5961	1529	105
HeLa	10274	2149	105

[a] Proteome Discoverer 1.4 software with Mascot (version 2.2.4) search engine was employed for data analysis. False discovery rates of peptide identifications are $<1\,\%$.

We also performed the first application of CZE–MS/MS for analysis of the yeast proteome. The system identified more than 1500 proteins and approximately 6000 peptides in around 100 min (Table 1). Roughly 4500 proteins are expressed during log-phase yeast growth;[19] our single-shot CZE–MS/MS data thus covered 1/3 of that proteome. We further separated the yeast proteome digest into three fractions by using RPLC. CZE–MS/MS analysis of the fractions identified 2512 yeast proteins in 5 h, which covers more than 50 % of the yeast proteome. This proteome dataset is the largest generated by using CZE–MS/MS to date.

Application of CZE–MS/MS to the HeLa cell proteome produced more than 2100 proteins and 10000 peptides in 105 min (Table 1). These results represent a roughly ten-fold improvement in the number of peptide and protein IDs compared to the previous state-of-the-art single-shot CZE–MS/MS analysis. The results significantly reduce the gap between single-shot CZE–MS/MS and the state-of-art single-shot RPLC–MS/MS for the analysis of mammalian proteomes.

Next, we evaluated the dynamic range of the identified yeast proteome with the CZE–MS/MS system. We performed a database search using MaxQuant (version 1.3);[20] 1529 and 2213 protein group IDs were identified from unfractionated and fractionated yeast samples, respectively, with both peptide- and protein-level false discovery rates of less than 1 %, and these IDs are consistent with the data from Proteome Discoverer shown in Table 1. Based on the protein intensity information from MaxQuant,[14] we estimated the dynamic range for the identified yeast proteome to be roughly five orders of magnitude for the unfractionated sample, which is similar to a published comprehensive yeast proteome study based on state-of-art single-shot RPLC–MS/MS.[14] Protein identifications for these experiments are provided in the Supporting Information, and replicate CZE runs for *E. coli* and yeast are shown in Figures S1 and S2 in the Supporting Information.

A recent multiple reaction monitoring study examined the detectability of 127 proteins that represent the full range of yeast protein expression.[21] Our single-shot CZE–MS/MS data identified 11 out of 12 proteins classified as "less than 50 copies/cell" and "western-blot band not quantifiable" in that work, although this classification may need to be revisited.[22]

We generated high-quality RPLC–MS/MS data for both the HeLa and yeast samples (see Figure 1 and Figure S3). RPLC generated more than 5000 protein and 40000 peptide IDs in 90 min from the HeLa sample; these values are approximately 2.5 and 4 times higher than those for CZE. While 70 % of the CZE peptides are detected by RPLC, CZE tends to identify larger peptides than RPLC (Figure S4), a result consistent with our previous work.[10] The poorer performance of CZE is due to an average peak width roughly 2.5 times larger than that for RPLC, which is due to the large injection volume ($>5\,\%$ of the capillary volume) used in the CZE experiment. Duplicate runs for the yeast sample generated reproducible separation profiles and base peak intensities (Figures S2 and S3).

There is some literature on the use of CZE for bottom-up analysis of complex proteomes; we summarize the data in Figure 2 and Table S1 in the Supporting Information. This learning curve demonstrates improved performance that results from a number of advancements. There is a steady increase in the complexity of the proteome, which increases the number of protein targets. There is also an improvement in the speed and mass accuracy of mass spectrometers, which leads to a concomitant increase in the number of IDs. Finally,

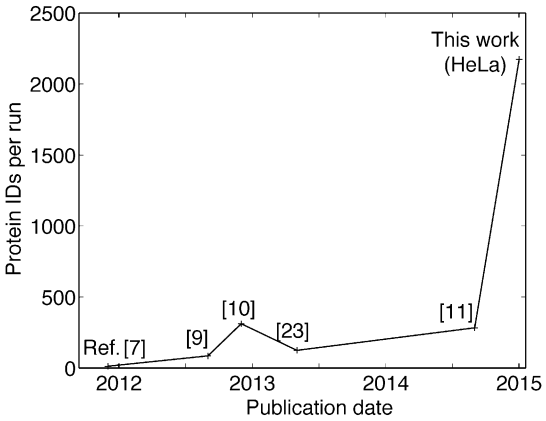

Figure 2. The numbers of protein IDs from single CZE–MS/MS analyses of complex proteomes in published studies over the years. A reference number for each point is listed in the figure. The average number of protein IDs per run from the analysis of multiple fractions was used for Refs. [7, 9, 23].

improved electrophoresis conditions yield longer runs and increase the number of IDs.

It is interesting to speculate on further improvements. Improved mass spectrometer speed and resolution will enable the identification of more peptides per run. Furthermore, the peaks generated in this separation are around 10 times broader than expected for a diffusion-limited separation. Narrower peaks will come from the optimization of injection conditions and should produce a two- to four-fold improvement in peptide and protein IDs. In this case, the performance of CZE–MS/MS should equal that of the very best RPLC–MS/MS system for bottom-up proteomics. Such technology will have at least three important advantages. First, CZE instrumentation is inherently simpler and less expensive than ultrahigh-pressure RPLC systems. Second, CZE samples a different peptide pool than RPLC; combining the results from the two technologies will allow deeper sequencing and higher protein coverage. Third, as we will show elsewhere, CZE migration time can be easily predicted from the sequence of a peptide, which improves confidence in peptide IDs.

Received: September 13, 2014
Revised: September 26, 2014
Published online: October 24, 2014

Keywords: analytical methods · CZE–MS/MS · electrophoresis · mass spectrometry · proteomics

[1] C. D. Wenger, J. J. Coon, *J. Proteome Res.* **2013**, *12*, 1377 – 1386.
[2] J. E. Elias, S. P. Gygi, *Nat. Methods* **2007**, *4*, 207 – 214.
[3] J. R. Yates, S. Speicher, P. R. Griffin, T. Hunkapiller, *Anal. Biochem.* **1993**, *214*, 397 – 408.
[4] A. A. Heemskerk, A. M. Deelder, O. A. Mayboroda, *Mass Spectrom. Rev.* **2014**, DOI: 10.1002/mas.21432.
[5] L. Sun, G. Zhu, X. Yan, N. J. Dovichi, *Curr. Opin. Chem. Biol.* **2013**, *17*, 795 – 800.
[6] L. Sun, G. Zhu, X. Yan, M. M. Champion, N. J. Dovichi, *Proteomics* **2014**, *14*, 622 – 628.
[7] Y. Li, M. M. Champion, L. Sun, P. A. Champion, R. Wojcik, N. J. Dovichi, *Anal. Chem.* **2012**, *84*, 1617 – 1622.
[8] K. Faserl, B. Sarg, L. Kremser, H. Lindner, *Anal. Chem.* **2011**, *83*, 7297 – 7305.
[9] Y. Wang, B. R. Fonslow, C. C. Wong, A. Nakorchevsky, J. R. Yates III, *Anal. Chem.* **2012**, *84*, 8505 – 8513.
[10] G. Zhu, L. Sun, X. Yan, N. J. Dovichi, *Anal. Chem.* **2013**, *85*, 2569 – 2573.
[11] L. Sun, G. Zhu, S. Mou, Y. Zhao, M. M. Champion, N. J. Dovichi, *J. Chromatogr. A* **2014**, *1359*, 303 – 308.
[12] C. D. Kelstrup, C. Young, R. Lavallee, M. L. Nielsen, J. V. Olsen, *J. Proteome Res.* **2012**, *11*, 3487 – 3497.
[13] M. Pirmoradian, H. Budamgunta, K. Chingin, B. Zhang, J. Astorga-Wells, R. A. Zubarev, *Mol. Cell. Proteomics* **2013**, *12*, 3330 – 3338.
[14] N. Nagaraj, N. A. Kulak, J. Cox, N. Neuhauser, K. Mayr, O. Hoerning, O. Vorm, M. Mann, *Mol. Cell. Proteomics* **2012**, *11*, M111.013722.
[15] A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall, J. J. Coon, *Mol. Cell. Proteomics* **2014**, *13*, 339 – 347.
[16] J. M. Busnel, B. Schoenmaker, R. Ramautar, A. Carrasco-Pancorbo, C. Ratnayake, J. S. Feitelson, J. D. Chapman, A. M. Deelder, O. A. Mayboroda, *Anal. Chem.* **2010**, *82*, 9476 – 9483.
[17] L. Sun, G. Zhu, Y. Zhao, X. Yan, S. Mou, N. J. Dovichi, *Angew. Chem. Int. Ed.* **2013**, *52*, 13661 – 13664; *Angew. Chem.* **2013**, *125*, 13906 – 13909.
[18] R. Wojcik, O. O. Dada, M. Sadilek, N. J. Dovichi, *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2554 – 2560.
[19] S. Ghaemmaghami, W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, J. S. Weissman, *Nature* **2003**, *425*, 737 – 741.
[20] J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367 – 1372.
[21] P. Picotti, B. Bodenmiller, L. N. Mueller, B. Domon, R. Aebersold, *Cell* **2009**, *138*, 795 – 806.
[22] S. S. Thakur, T. Geiger, B. Chatterjee, P. Bandilla, F. Fröhlich, J. Cox, M. Mann, *Mol. Cell. Proteomics* **2011**, *10*, M110.003699.
[23] X. Yan, D. C. Essaka, L. Sun, G. Zhu, N. J. Dovichi, *Proteomics* **2013**, *13*, 2546 – 2551.

