Bottom-up proteome analysis of *E. coli* using capillary zone electrophoresis-tandem mass spectrometry with an electrokinetic sheath-flow electrospray interface

Xiaojing Yan[1], **David C. Essaka**[1,2], **Liangliang Sun**[1], **Guijie Zhu**[1], and **Norman J. Dovichi**[1]

[1]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556 USA

Abstract

The *E. coli* proteome was digested with trypsin and fractionated using solid phase extraction on a C18 SPE column. Seven fractions were collected and analyzed by capillary zone electrophoresis (CZE)-electrospray ionization tandem mass spectrometry (ESI-MS/MS). The separation was performed in a 60 cm long linear polyacrylamide-coated capillary with a 0.1% (v/v) formic acid separation buffer. An electrokinetic sheath-flow electrospray interface was used to couple the separation capillary with an Orbitrap Velos operating in higher-energy collisional dissociation mode. Each CZE-ESI-MS/MS run lasted 50 minutes and total MS time was 350 minutes. A total of 23,706 peptide spectra matches, 4,902 peptide IDs, and 871 protein group IDs were generated using MASCOT with false discovery rate less than 1% on the peptide level. The total mass spectrometer analysis time was less than six hours, the sample identification rate (145 proteins/hour) was more than two times higher than previous studies of the *E. coli* proteome, and the amount of sample consumed (<1 μg) was roughly four-fold less than previous studies. These results demonstrate that CZE is a useful tool for the bottom-up analysis of prokaryote proteomes.

Capillary zone electrophoresis (CZE) employs very simple instrumentation, wherein a sample is separated in a buffer-filled fused silica capillary under the influence of an electric field. The simplicity of capillary electrophoresis proved invaluable in the sequencing of the human genome, where essentially all data were generated using multiple capillary electrophoresis instrumentation [1].

Despite its success in DNA sequencing, capillary electrophoresis has had negligible impact on proteomic research. Smith and colleagues reported the coupling of capillary zone electrophoresis with mass spectrometry in 1987 and presented the analysis of a set of small ions [2]. Another paper followed two years later from that group that reported the use of capillary electrophoresis for analysis of intact proteins [3]. Most subsequent examples consider the analysis of a few standard peptides or the tryptic digest of a few standard proteins.

Very few manuscripts describe the use of capillary zone electrophoresis for the bottom-up analysis of complex proteomic samples. Lindner *et al.* reported a sheathless CZE-ESI-

[2]Present address: St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105 USA

MS/MS system and compared it to RPLC-ESI-MS/MS by analyzing a rat testis linker histone protein sample digested by endoproteinase Arg-C. The total analysis time of CZE-ESI-MS/MS was shorter than nano-RPLC-ESI-MS/MS and identified more low molecular mass peptides. Eight non-histone H1 proteins were identified from the sample by capillary electrophoresis, whereas 23 proteins were identified by LC using a 10X larger sample loading [4].

Yates employed a solid-phase microextraction (SPME) technique to prefractionate the yeast ribosome digest followed by CE-MS analysis [5]. Eleven fractions were analyzed with 30-minute long CE separations. A total of 66 proteins were identified in the 5.5 hour long mass-spectrometry analysis time. Recently, the Yates group further applied an improved on-line SPME fractionation, transient isotachophoresis capillary electrophoresis–tandem mass spectrometry technique with an etched porous capillary as ESI sprayer for the proteomic analysis of a moderately complex protein mixture [6]. In total, 2,341 peptide IDs and 548 protein IDs were generated by the SPME-CE-MS/MS system in duplicate runs from *Pyrococcus furiosus* tryptic digest, and the total mass spectrometry time was about 350 min.

This group recently published a description of the use of CZE for the analysis of a proteome of intermediate complexity [7]. In that study, the secretome of *M. marinum* was analyzed by both CZE and by UPLC. The systems were constrained to the same analysis time (3 hours), and sample loadings were optimized for each separation method. The CE analysis employed reversed-phase liquid chromatography to generate 11 fractions, each of which was analyzed in a short CZE separation. The UPLC analysis employed triplicate analysis in a set of one-hour separations. The two separation methods produced similar number of protein and peptide identifications, but with modest concordance between the methods. The CZE separation identified 140 protein groups and 334 peptides. Recently, we further optimized the CZE separation of complex protein digests, and more than 1,250 *E. coli* peptide IDs could be generated by single-shot CZE-ESI-MS/MS analysis with 50 min mass spectrometry time [8], which opens the door of CZE-ESI-MS/MS for complex protein digests analysis.

Escherichia coli is an important model system for proteome analysis, and its proteome has been frequently analyzed by LC-MS/MS. Cargile *et al.* used gel based isoelectric focusing (IEF) to prefractionate tryptic *E. coli* proteome digests, and each fraction was further analyzed by RPLC-ESI-MS/MS (LCQ) [9]. The approach yielded 417 proteins and 1022 peptides. Iwasaki *et al.* directly coupled a 350 cm long monolithic silica-C18 capillary column to an LTQ-Orbitrap mass spectrometer for *E. coli* proteome analysis [10]. 2,602 proteins and 22,196 peptides were identified by this system with a 41 h LC gradient. Xia *et al.* developed a multidimensional LC platform for online protein fractionation by weak anion and cation exchange (WAX/WCX) mixed-bed microcolumn, protein digestion by immobilized trypsin microreactor (IMER), and peptide separation and identification by nano-RPLC-ESI-MS/MS (LTQ) [11]. The integrated platform identified 689 *E. coli* proteins and 2200 peptides from 8 μg protein in a 14 h analysis time; the results were comparable with traditional MudPIT method, which also consumed 8 μg peptides.

In this manuscript, we report the first example of CZE for the large-scale analysis of a complex proteome. The *E. coli* proteome was digested with trypsin and separated into seven

fractions with a SPE C18 column using a series of acetonitrile solutions. Each fraction was analyzed with our improved CZE-ESI-MS/MS procedure. Peptides were separated in a 60 cm long linear polyacrylamide (LPA)-coated capillary (50 μm i.d.×150 μm o.d.), and electrosprayed using an electrokinetically-driven sheath-flow interface into an Orbitrap-Velos mass spectrometer [12]. Higher-energy collisional dissociation (HCD) fragmentation was employed [13]. Over 850 protein groups and close to 5,000 peptides were idenfied in the 350 min analysis using MASCOT with false discovery rate (FDR) less than 1% on peptide level.

Materials and Methods

Materials

All reagents were purchased from Sigma Aldrich (St. Louis, MO, USA), unless stated. Formic acid (FA) and acetonitrile (ACN) were purchased from Fisher Scientific (Pittsburgh, PA, USA). Methanol was purchased from Honeywell Burdick & Jackson (Wicklow, IE). Water was deionized by a Nano Pure system from Thermo Scientific (Marietta, OH, USA). Linear polyacrylamide (LPA)-coated fused capillary (50 μm i.d./150 μm o.d.) was purchased from Polymicro Technologies (Phoenix, AZ, USA). Complete, mini protease inhibitor cocktail (provided in EASYpacks) was purchased from Roche (Indianapolis, IN, USA).

E. coli Sample preparation

Solid lysogeny broth (LB) was prepared by dissolving 3 g each of NaCl and tryptone, 1.5 g of yeast extract, and 6 g of agar in 300 mL of deionized water; this material was used to prepare agar plates. Another liter of the LB medium (without agar) was prepared by mixing 10 g each of NaCl and tryptone, and 5 g of yeast extract in deionized water; this material was used to grow and maintain the cultures. All media, plates, and other utensils and flasks were appropriately autoclaved before use. The medium was dispensed into the culture dishes and kept at room temperature overnight. The next day, these agar plates were stored upside down at 4°C. Before use, one plate was left at room temperature for about half an hour. Frozen cultures of E-coli (Dh5-Alpha) were thawed and plated on agar plates. A colony was inoculated on the plate and incubated at 37°C. After a day, each single colony was grown in a flask with 150 mL LB medium, and shaken overnight at 37°C.

After a few washes by PBS buffer, the cell pellet was lysed with 4% sodium dodecyl sulfate (SDS) and 100 mM Tris-HCl buffer (pH 8.0) supplemented by protease inhibitor (1 tablet of Complete Mini in 10 mL), and sonicated for 15 minutes. The lysate was centrifuged at $18,000 \times$ g for 15 minutes, and the supernatant was kept for further analysis. Protein concentration was measured by BCA method, and the protein concentration was 1.5 mg/mL. About 350 μg *E. coli* proteins were reduced in 15 mM dithiothreitol (DTT) at 95 °C for 5 min, followed by protein alkylation with iodoacetamide (IAA) and tryptic protein digestion according to filter aided sample preparation (FASP) protocol [14] developed by Mann group. The obtained protein digests were loaded on a C18 SPE column (Waters, Milford, MA, USA), and the captured peptides were eluted by a series of solutions with different ACN concentration (8%, 12%,16%, 20%, 24%, 28% and 80% (v/v)) in sequence. The obtained seven fractions were lyophilized with a vacuum concentrator (Thermo Fisher

Scientific, Marietta, OH, USA), and the dried digests were redissovled in 50% (v/v) ACN and 0.05% (v/v) FA for CZE-ESI-MS/MS analysis.

CZE-ESI-MS/MS

The electrophoresis system was assembled from components [7, 8, 12, 15–18]. High voltages were provided by two Spellman CZE 1000R high-voltage power supplies. Electrospray was generated using an electrokinetically pumped sheath flow interface [9]. The electrospray emitter was borosilicate glass capillary (1.0 mm o.d., 0.75 mm i.d., 10 cm) pulled with a Sutter instrument P-1000 flaming/brown micropipette puller. The size of the emitter opening was 5–10 μm. Voltage programming was controlled by LabView software.

Each of the dried C18-SPE fractions was separately reconstituted in 12.5 μL solution containing 0.05% (v/v) FA and 50% (v/v) ACN. The separation buffer was 0.1% FA in water and the electrospray sheath flow buffer was 10% (v/v) methanol and 0.1% (v/v) FA.

The separation capillary (50 μm i.d., 150 μm o.d., 60 cm) was LPA coated. The sample was loaded onto the separation capillary by pressure injection, and the injection volume was about 25 nL. For separation, 16.5 kV was applied on the injection end of the capillary for separation and 1.5 kV on the sheath flow reservoir for electrospray. Each fraction was analyzed once.

The LTQ-Orbitrap Velos (Thermo Fisher Scientific) was used in this experiment and HCD mode was applied. Full MS scans were acquired in the Orbitrap mass analyzer over the m/z 395–1800 range with resolution 30,000 (at m/z 400). Ten most intense peaks with charge state 2 were fragmented in the HCD collision cell, and analyzed by the Orbitrap mass analyzer with resolution 7,500. One microscan was used. Normalized collision energy was set as 40%. For MS and MS/MS spectra acquisition, maximum ion inject time was set as 500 ms and 250 ms, respectively. The precursor isolation width was 2 Da. The target value for MS and MS/MS was set as 1.00E+06 and 5.00E+04, respectively. Dynamic exclusion was enabled, and peaks selected for fragmentation more than once within 20 s were excluded from selection for 20 s.

Protein identification and data analysis

Database searching of the raw files was performed in Proteome Discoverer 1.3 with MASCOT 2.2.4 against the SwissProt database with taxonomy as *Escherichia coli* (13,477 sequences). Database searching of the reversed database was also performed in order to evaluate the false discovery rate (FDR) [19, 20]. The database searching parameters included semi tryptic digestion and allowed up to two missed cleavages, precursor mass tolerance 10 ppm, and fragment mass tolerance 0.05 Da. Carbamidomethylation (C) was set as fixed modifications. Oxidation (M), deamidated (NQ), and acetyl (protein N-term) were set as variable modifications.

Percolator software integrated in the Proteome Discoverer 1.3 was used to evaluate the identification, and the validation was based on the q-value. On the peptide level, peptide confidence value as high or medium was used to filter the peptide identification, and the corresponding FDR on peptide level was less than 1% or 5%. On the protein level, protein

grouping was enabled. Leucine and isoleucine were considered as equal, and the strict maximum parsimony principle was applied.

Results and discussion

The data generated in this study were subjected to MASCOT database searching, Table 1. In total, 871 protein groups and 4,902 peptides were generated from fractionated *E. coli* digests with about 350 min mass spectrometry time, and the FDR on peptide level was less than 1%. The detected protein groups and peptides are listed in **supporting material**. The average sequence coverage of the identified protein groups was about 22% and the average unique peptide number was about 5, which suggested high confident protein identification. This is the largest dataset of a complex proteome with CZE-MS/MS system at present. In addition, 42,164 tandem spectra were acquired and 56% of the tandem spectra had confident peptide matches (23,706).

The large number of peptides and high identification rate generated by CZE-ESI-MS/MS in this study were most likely due to three reasons. First, HCD fragmentation generated high quality tandem spectra, and also high accurate mass of fragment ions, which reduced the number of peptide hits to the reversed database, and reasonably increased the confident peptide IDs when consistent peptide level FDR was applied for filtration. Second, the efficient peptide fractionation by the C18-SPE column reduced the complexity of the *E. coli* digests, Fig. 1. The number of peptides identified in each fraction ranged from ~800 to ~1,400, and the overlap between the neighbored fractions was on the 300 peptide level, which indicated that the C18-SPE fractionation worked reasonably well. In addition, the high capacity CZE separation reduced the ion suppression of relatively low abundant peptides from high abundant ones, and the wide separation window of the CZE (from ~10 min to ~40 min) dramatically increased the number of acquired tandem spectra, Fig. 2. Third, the relatively large amount of loaded peptide (on 100 ng level) for CZE-ESI-MS/MS analysis led to high peptide intensity and reasonably high quality of tandem spectra. In addition, because the sample was dissolved in 50% (v/v) ACN and 0.05% (v/v) FA for injection and its conductivity was lower than the separation buffer (0.1% (v/v) FA), the peptides were stacked at the beginning of CZE separation, which further improved the CZE separation and also the peptide intensity.

We further determined the molecular weight (MW) and pI distributions of identified proteins, Fig. 3. The MW of identified proteins ranged from less than 10 kDa to higher than 100 kDa, and most proteins were in the 10–50 kDa range, Fig. 3A. The pI of identified proteins ranged from less than 4 to higher than 11, and most proteins were in pI 5–7 range, Fig. 3B. These results demonstrate that the CZE-ESI-MS/MS system is efficient for identifying both low and high MW proteins, and also acidic and basic proteins. We also analyzed the transmembrane domains (TMDs) of the identified proteins, and the number of TMDs was predicted by the TMHMM v. 2.0 (www.cbs.dtu.dk/services/TMHMM/). About 10% of the totally identified proteins were of at least 1 TMDs, and were considered as integral membrane proteins.

We also compared our results with three other published works for *E. coli* proteome analysis, Table 2. The IEF-RPLC- MS/MS approach was used for *E. coli* proteome analysis by Stephenson *et al.* [9]. The tryptic digests were firstly offline fractionated by gel based isoelectric focusing (IEF). The gel was cut into ~ 2.5 mm slices in order to obtain 27 fractions, and each fraction was further analyzed by RPLC-MS/MS (LCQ) with 85 min LC gradient per run. The approach yielded 417 proteins and 1022 peptides with 20 µg sample amounts and close to 40 h mass spectrometry time. Compared with our results from CZE-MS/MS (871 proteins and 4902 peptides), the relatively lower identifications from this approach might be due to the mass spectrometer (LCQ) they used. In addition, the peptide fractionation and further peptide extraction from the gel in the procedure is labor and time consuming. The RPLC-MS/MS with 3.5 m long separation column approach coupled 3.5 m monolithic silica-C18 capillary column with 41 h LC gradient to an LTQ-Orbitrap mass spectrometer for *E. coli* proteome analysis, and 2 602 proteins were identified [10]. Although the protein IDs from this approach is dramatically higher than that from our CZE-MS/MS method, the protein identification efficiency from CZE-MS/MS is two times higher than that from the RPLC-MS/MS with 3.5 m long separation column approach (145 *vs.* 63 protein IDs/h). The WAX/WCX-IMER-RPLC- MS/MS approach developed by Xia *et al.* integrated online protein fractionation, protein digestion and peptide separation and identification for *E. coli* proteome analysis [11]. The protein and peptide IDs from this approach are lower than that from our CZE-MS/MS method (689 *vs.* 871; 2 200 *vs.* 4 902), and one possible reason for this is the different mass spectrometer used (LTQ XL vs. LTQ-Orbitrap Velos). In addition, the loaded sample amounts, named consumed sample amounts, are less than 1 µg for CZE-MS/MS, which is lower than other three approaches. Results indicated that the CZE-MS/MS approach is of great potential for large-scale proteomics analysis.

One practical way for further improvement of proteome profiling depth with CZE-MS/MS method is to apply the protein level fractionation techniques (i.e. ion exchange chromatography, size exclusion chromatography and isoelectric focusing) to efficiently reduce the proteome complexity before CZE-MS/MS analysis. The total mass spectrometer analysis time was less than six hours, the sample identification rate (proteins/hour) was more than two times higher than previous studies of the *E. coli* proteome, and the amount of sample consumed was roughly four-fold less than previous studies.

Conclusions

CZE-ESI-MS/MS was applied for large-scale *E. coli* proteome analysis. 4,902 peptide IDs and 871 protein group IDs were generated with 350 min mass spectrometry time, which is the largest proteome dataset based on CE-MS/MS reported to date. The results indicate that CZE-ESI-MS/MS will be a useful tool for large scale proteomics analysis.

Acknowledgments

References

1. Dovichi NJ, Zhang J. How capillary electrophoresis sequenced the human genome. Angew Chem Int Ed Engl. 2000; 39:4463–4468. [PubMed: 11169637]

2. Olivares JA, Nguyen NT, Yonker CR, Smith RD. On-line mass spectrometric detection for capillary zone electrophoresis. Anal Chem. 1987; 59:1230–1232.

3. Smith RD, Loo JA, Barinaga CJ, Edmonds CG, Udseth HR. Capillary zone electrophoresis and isotachophoresis-mass spectrometry of polypeptides and proteins based upon an electrospray ionization interface. J Chromatogr. 1989; 480:211–232. [PubMed: 2592481]

4. Faserl K, Sarg B, Kremser L, Lindner H. Optimization and evaluation of a sheathless capillary electrophoresis–electrospray ionization mass spectrometry platform for peptide analysis: comparison to liquid chromatography–electrospray ionization mass spectrometry. Anal Chem. 2011; 83:7297–7305. [PubMed: 21848273]

5. Tong W, Link A, Eng JK, Yates JR III. Identification of proteins in complexes by solid-phase microextraction/multistep elution/capillary electrophoresis/tandem mass spectrometry. Anal Chem. 1999; 71:2270–2278. [PubMed: 10405597]

6. Wang YJ, Fonslow BR, Wong CCL, Nakorchevsky A, Yates JR III. Improving the comprehensiveness and sensitivity of sheathless capillary electrophoresis–tandem mass spectrometry for proteomic analysis. Anal Chem. 2012; 84:8505–8513. [PubMed: 23004022]

7. Li Y, Champion MM, Sun L, Champion PA, Wojcik R, Dovichi NJ. Capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry as an alternative proteomics platform to ultraperformance liquid chromatography-electrospray ionization-tandem mass spectrometry for samples of intermediate complexity. Anal Chem. 2012; 84:1617–1622. [PubMed: 22182061]

8. Zhu G, Sun L, Yan X, Dovichi NJ. Single-shot proteomics using capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry produces more than 1,250 E. coli peptide identifications in a 50-minute separation. Anal Chem. 2013; 85:2569–2573. [PubMed: 23394296]

9. Cargile BJ, Bundy JL, Freeman TW, Stephenson JL Jr. Gel based isoelectric focusing of peptides and the utility of isoelectric point in protein identification. J Proteome Res. 2004; 3:112–119. [PubMed: 14998171]

10. Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y. One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. Anal Chem. 2010; 82:2616–2620. [PubMed: 20222674]

11. Xia S, Tao D, Yuan H, Zhou Y, Liang Z, Zhang L, Zhang Y. Nano-flow multidimensional liquid chromatography platform integrated with combination of protein and peptide separation for proteome analysis. J Sep Sci. 2012; 35:1764–1770. [PubMed: 22623513]

12. Wojcik R, Dada OO, Sadilek M, Dovichi NJ. Simplified capillary electrophoresis nanospray sheath- flow interface for high efficiency and sensitive peptide analysis. Rapid Commun Mass Spectrom. 2010; 24:2554–2560. [PubMed: 20740530]

13. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods. 2007; 4:709–712. [PubMed: 17721543]

14. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009; 6:359–362. [PubMed: 19377485]

15. Li Y, Wojcik R, Dovichi NJ. A replaceable microreactor for on-line protein digestion in a two-dimensional capillary electrophoresis system with tandem mass spectrometry detection. J Chromatogr A. 2011; 1218:2007–2011. [PubMed: 21030030]

16. Wojcik R, Li Y, MacCoss MJ, Dovichi NJ. Capillary electrophoresis with Orbitrap-Velos mass spectrometry detection. Talanta. 2012; 88:324–329. [PubMed: 22265506]

17. Sun L, Zhu G, Li Y, Wojcik R, Yang P, Dovichi NJ. CZE-ESI-MS/MS system for analysis of subnanogram amounts of tryptic digests of a cellular homogenate. Proteomics. 2012; 12:3013–3019. [PubMed: 22888077]

18. Zhu G, Sun L, Yang P, Dovichi NJ. On-line amino acid-based capillary isoelectric focusing-ESI-MS/MS for protein digests analysis. Anal Chim Acta. 2012; 750:207– 211. [PubMed: 23062442]

19. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem. 2002; 74:5383–5392. [PubMed: 12403597]

20. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods. 2007; 4:207–214. [PubMed: 17327847]
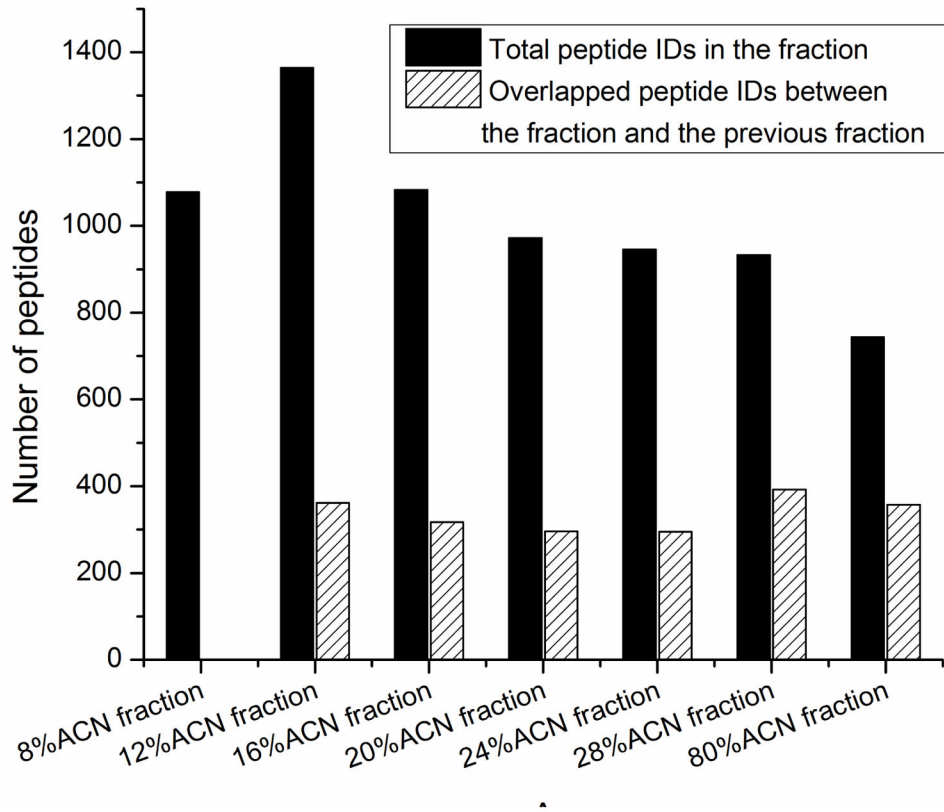
Figure 1.
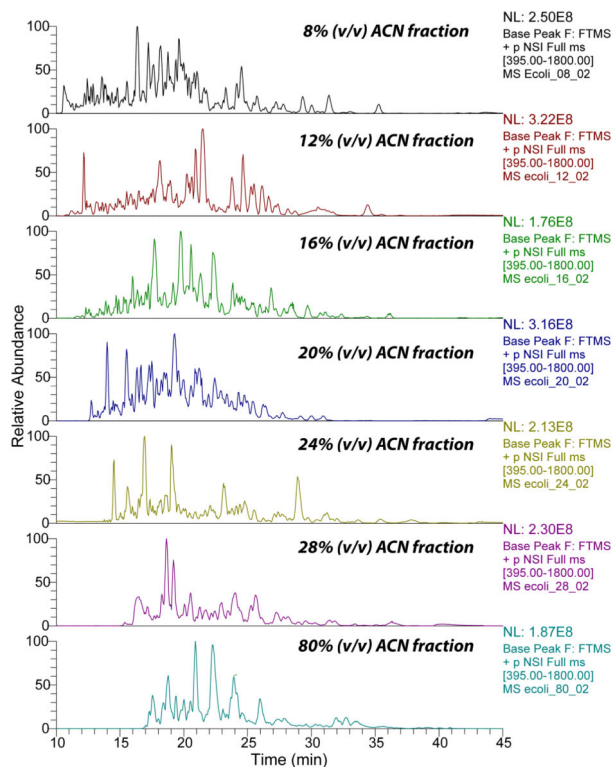Distribution of peptide IDs from different fractions identified by CZE-ESI-MS/MS.

Figure 2.
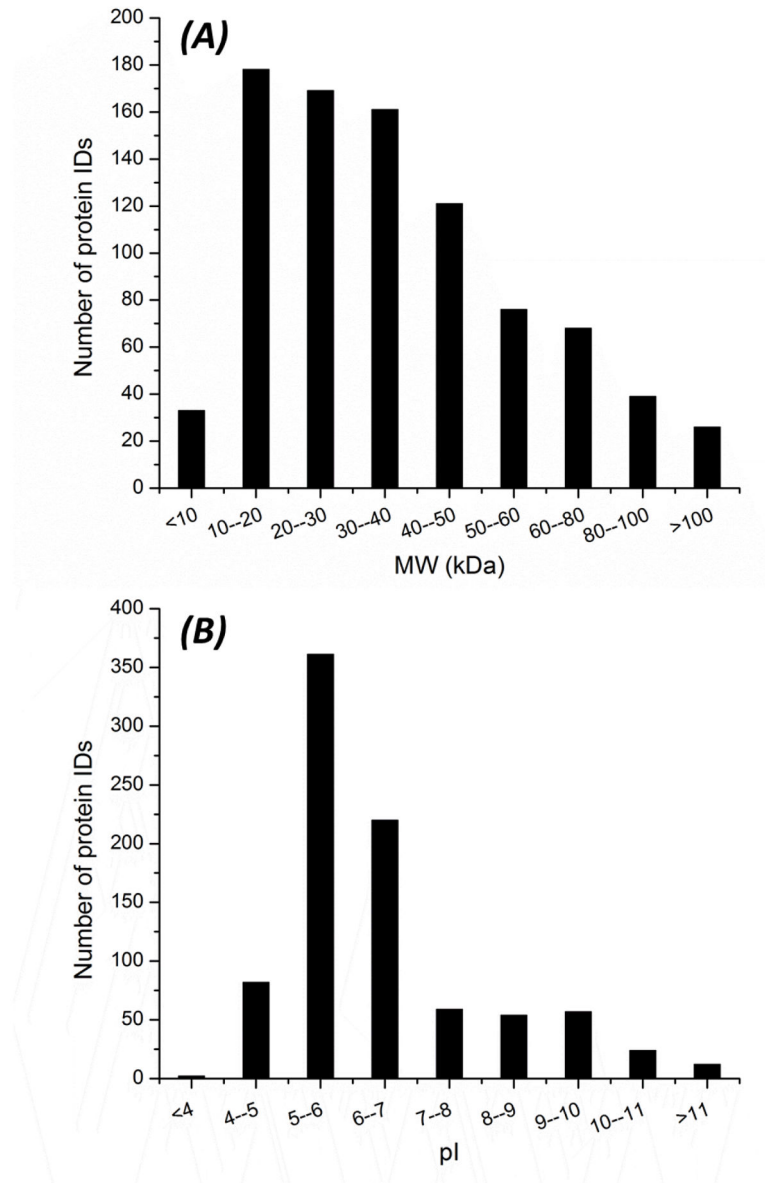Base peak electropherograms of seven fractions of *E. coli* digests analyzed by CZE-ESI-MS/MS.

Figure 3.
Distributions of molecular weight (MW) and pI of proteins identified by CZE-ESI-MS/MS.
MW (A); pI (B).

Table 1

database searching results

	Protein groups	Peptides	Peptide spectra matches (PSMs)	Acquired tandem spectra	Identification rate (%)
MASCOT (FDR<1%)	871	4,902	23,706	42,164	56
MASCOT (FDR<5%)	884	5,154	24,441	42,164	58

Table 2

Comparisons of our results with three examples of *E. coli* proteome analysis.

| | CZE-MS/MS | IEF-RPLC-MS/MS [9] | LC-MS/MS | | |
			RPLC-MS/MS with 3.5 m long separation column [10]	WAX/WCX-IMER-RPLC-MS/MS [11]	
Protein IDs	871	417	2,602	689	
Peptide IDs	4,902	1022	22,196	2,200	
Mass spectrometry time (h)	6	38	41	14	
Protein identification efficiency (number of protein IDs/h)	145	11	63	49	
Consumed sample amounts (μg)	< 1	20	4	8	
Mass spectrometer	LTQ-Orbitrap Velos	LCQ	LTQ-Orbitrap XL	LTQ XL	