Anal. Chem. **2003,** *75,* 3502−3505

Identification of Proteins in Single-Cell Capillary Electrophoresis Fingerprints Based on Comigration with Standard Proteins

Shen Hu,[†] Le Zhang,[†] Richard Newitt,[‡] Ruedi Aebersold,[‡] James R. Kraly,[†] Megan Jones,[†] and Norman J. Dovichi*,[†]

Department of Chemistry, University of Washington, Seattle, Washington 98195, and The Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103

In the previous paper in this Journal, we reported the use of capillary sieving electrophoresis to characterize proteins expressed by single cancer cells at specific phases in the cell cycle. Analysis of the data revealed one component with cell cycle-dependent changes in expression at the 99% confidence limit. However, the amount of protein present in a single cell is far too small to allow its direct identification by mass spectrometry. In this paper, we report a method by which such proteins can be tentatively identified. We perform standard SDS−PAGE electrophoresis of the proteins contained within a homogenate prepared from an HT29 cell culture. Proteins extracted from bands in the gel are identified by mass spectrometry. The proteins also provide a set of standards that can be used to spike the sample before capillary sieving electrophoresis (CSE) separation; comigration is taken as evidence for the identity of the target protein. In a proof-of-principle experiment, a single band migrating at ∼47 kDa was isolated from the SDS−PAGE gel generated from the HT29 cell line. Proteins extracted from this band were used to spike a CSE separation of the same extract. This band comigrated with a cell cycle-dependent component identified from single-cell analysis. In-gel digestion and LC/MS/MS were used to identify five proteins, including cytokeratin 18, which is the product of the most highly expressed gene in this cell line.

A single mammalian cell is typically 10 μm in diameter, has a volume of 500 fL, and contains roughly 50 pg or 2 fmol of protein.[1] There are roughly 30 000 genes within the genome, which can be translated to perhaps 100 000 different proteins through alternative splicing. The average protein is present in the zeptomole range. Unfortunately, the amount of protein contained within a cell is far too small for identification by classic methods.

We have reported a method to characterize proteins in a single cell on the basis of capillary electrophoresis and ultrasensitive laser-induced fluorescence detection.[1−5] The preceding paper in this Journal reports the separation of proteins from single cancer cells in the G1 and G2/M phases in the cell cycle.[5] One component had a cell cycle-dependent protein expression at the 99% confidence level.

In this paper, we describe a method by which proteins may be tentatively identified on the basis of comigration with standards prepared from the sample of interest. Classic SDS−PAGE is used to separate micrograms of proteins from a cellular homogenate. A band is excised from the gel and used for two purposes. Proteins are first identified on the basis of in-gel digestion with trypsin, separation of the resulting peptides by liquid chromatography, and analysis by MS/MS, followed by identification by comparison with a database. The purified proteins are then used as standards to spike the CSE sample. Comigration is taken as evidence for the presence of the standard proteins in the sample.

EXPERIMENTAL SECTION

Capillary Electrophoresis Instrument. All separations were performed using a locally constructed capillary electrophoresis instrument equipped with a sheath-flow laser-induced fluorescence detector.[6]

Reagents. Unless otherwise stated, all reagents were obtained from Sigma (St. Louis, MO). 3-(2-Furoyl)-quinoline-2-carboxaldehyde (FQ), and KCN were obtained from Molecular Probes (Eugene, OR). Pullulan was used as the sieving matrix for CSE. The sieving buffer contained 0.1 M Tris, 0.1 M 2-(cyclohexylamino)-ethanesulfonic acid (CHES), 8% pullulan, and 0.1% SDS (pH 8.6).

Cell Culture. HT29 human colon adenocarcinoma cells were grown to 80% confluence in Dulbecco's modified Eagle's medium, supplemented with 10% fetal bovine serum at 37 °C in 5% CO_2 atmosphere. The cells were washed four times with phosphate-buffered saline (PBS) to remove the residual substrate from the medium immediately before single cell analysis.

Protein Identification. A protein extract of HT29 cancer cells was denatured with BioRad Laemmli sample buffer at 90 °C for 5

[†] University of Washington.

[‡] The Institute for Systems Biology.

(1) Zhang, Z.; Krylov, S.; Arriaga, E. A.; Polakowski, R.; Dovichi, N. J. *Anal. Chem.* **2000**, *72*, 318−322.

(2) Hu, S.; Zhang, L.; Cook, L. M.; Dovichi, N. J. *Electrophoresis* **2001**, 22, 3677−3682.

(3) Hu, S.; Lee, R.; Zhang, Z.; Krylov, S. N.; Dovichi, N. J. *J. Chromatogr., B* **2001**, *752*, 307−310.

(4) Hu, S.; Jiang, J.; Cook, L.; Richards, D. P.; Horlick, L.; Wong, B.; Dovichi, N. J. *Electrophoresis* **2002**, *23*, 3136−3142.

(5) Hu, S.; Zhang, L.; Krylov, S. N.; Dovichi, N. J. *Anal. Chem.* **2003**, *75*, 3495−3501.

(6) Wu, S.; Dovichi, N. J. *J. Chromatogr., A* **1989**, *480*, 141−155.

10.1021/ac034154j CCC: $25.00 © 2003 American Chemical Society
Published on Web 05/31/2003

min. SDS−PAGE of the protein extract was performed on 7 cm × 10 cm × 1 mm minigel using the BioRad Mini-Protean III system. Two minigels were used to separate the same protein extract simultaneously. After SDS−PAGE, one minigel was visualized with BioRad Coomassie stain; the other was visualized with BioRad zinc stain.

The ~47-kDa protein band was excised from the zinc-stained gel and transferred to a 1.5-mL microcentrifuge tube. Destaining of the gel piece was performed in BioRad zinc destain solution using a three-step procedure.[7] The gel slice was first soaked in 100 μL of destain solution and 900 μL of water and vortexed for 5 min. The second soaking was performed in a fresh solution of 100 μL of destain and 900 μL of water with 10 min of vortexing. The final soaking was performed in a fresh solution of 50 μL of destain and 950 μL of water with 5 min of vortexing. The gel slice was finally crushed at the bottom of the microcentrifuge tube, and 50 μL of extraction solution was added. The protein extraction solution contained formic acid, water, and 2-propanol at a ratio of 1:3:2 (v/v/v). The microcentrifuge tube containing the gel slurry was then vigorously vortexed for 4 h. After vortexing, the standard protein mixture was centrifuged for 5 min, and the supernatant was collected for the spiking experiment. This standard protein mixture was used to spike a HT-29 protein sample, labeled with FQ, and analyzed using CSE.

The same protein band, stained with BioRad Coomassie Blue, was also excised and transferred to a 0.5-mL microcentrifuge tube. The gel band was finely minced, dehydrated in CH_3CN, and dried under vacuum. For reduction and alkylation, the gel pieces were treated with 10 mM dithiothrietol at 56 °C for 45 min and then with 55 mM iodoacetamide in darkness for 30 min. Gel pieces were dehydrated in CH_3CN and dried under vacuum again. Proteins were then in-gel digested with 10 ng/μL modified trypsin (Promega) in 50 mM NH_4HCO_3 at 37 °C overnight. Peptides were extracted with 5% HCOOH and 50% CH_3CN, and the extract was dried under vacuum and finally reconstituted in 0.005% heptafluorobutyric acid and 0.4% acetic acid.

Analysis of peptides by microelectrospray LC/MS/MS was performed essentially as described.[8] Microelectrospray columns were constructed from 360-μm o.d. × 75-μm i.d. fused-silica capillary with the column tip tapered to a 5−10-μm opening. The columns were packed with 200-Å 5-μm C18 beads to a length of 10−12 cm. The flow through the column was split precolumn to achieve a flow rate of 300 nL/min. The mobile phase used for gradient elution consisted of (A) 0.4% acetic acid, 0.005% heptafluorobutyric acid, 5% acetonitril and (B) 0.4% acetic acid and 0.005% heptafluorobutyric acid in acetonitrile. The gradient was linear from 0.5 to 45% B in 35 min followed by 45−65% B in 5 min. Tandem mass spectra were recorded on a LCQ ion trap mass spectrometer (ThermoFinnigan Corp., San Jose, CA) equipped with an in-house microelectrospray ionization source. Needle voltage was set at 1.8 kV. Ion signals above a predetermined threshold automatically triggered the instrument to switch from MS to MS/MS mode for generating collision-induced dissociated spectra (data-dependent MS/MS). The generated collision-induced dissociated spectra were searched against the National Cancer Institute human protein database using the computer algorithm, SEQUEST.[9]

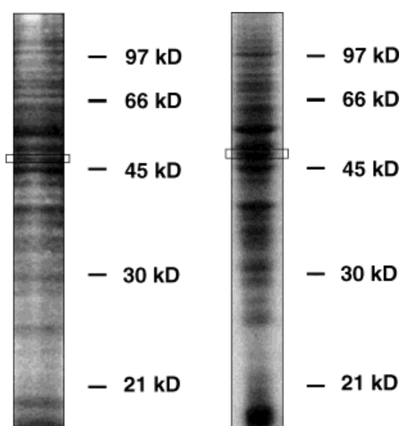

Figure 1. SDS−PAGE gel electropherogram generated from protein extract prepared from the HT-29 cancer cell line. The band used for subsequent analysis is denoted by the rectangle. The left panel is the zinc-stained gel and the right panel is the Coomassie Blue-stained gel.

RESULTS AND DISCUSSION

Capillary Sieving Electrophoresis. Capillary sieving electrophoresis is the capillary version of SDS−PAGE. Proteins are separated by molecular weight in a sieving matrix made from a soluble polymer. Polyacrylamide is not ideal for these separations because of difficulties in controlling the free-radical polymerization, and cross-linked gels are not desirable because they block fluid flow in the capillary, preventing cell injection. We instead used solutions of the polysaccharide pullulan, which provides sufficiently low viscosity to allow injection of a single cell.[2,5] Other soluble polymers, such as poly(ethylene oxide), have also been used for single cell analysis.[4]

Protein Identification by LC/MS/MS. The preceding paper in this Journal reports the cell cycle-dependent protein expression in single cells from the HT-29 colon cancer cell line.[4] One component had a cell cycle-dependent protein expression, which was significant at the 99% confidence limit. That protein had a nominal molar mass of 47 kDa, as estimated from its migration time during capillary sieving electrophoresis. The protein's mobility appears to increase as cells pass from the G1 to the G2/M phase of the cell cycle. The amount of protein within this band from a single cell is far too small to identify by classic methods. In this paper, we present a simple method to tentatively identify such proteins of interest.

Fortunately, the electropherograms produced by CSE and SDS−PAGE are similar, even though capillary electrophoresis is performed on a sample that is 5 orders of magnitude smaller than the SDS−PAGE experiment. As a result, it is possible to identify the same band in both forms of electrophoresis. We prepared two SDS−PAGE gels from the HT-29 cell line, and we tentatively identified the band the 47 kDa component in the gels, Figure 1. This band was excised from both gels. One band was extracted, and the liberated proteins were used to spike the CSE separation of the HT-29 proteins. The run-to-run reproducibility of migration

(7) Cohen, S. L.; Chait, B. T. *Anal. Biochem.* **1997**, *247*, 257−267.
(8) Gygi, S. Y.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994−999.
(9) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.

Table 1. Identification of Proteins That Comigrate with the 47-kDa Fraction Using LC/MS/MS and SEQUEST Database Searching

entry name	protein name	peptide sequence	$(M + H)^+$
ROD_human (hnRNP D0)	heterogeneous nuclear ribonucleoprotein D0	HSEAATAQR	971.0
K1CR_human	cytokeratin 18	KVIDDTNITR	1175.3
SAHH_human	adenosylhomocysteinase	VPAINVNDSVTK	1257.4
TBB1_human	tubulin β-1 chain	LHFFMPGFAPLTSR	1621.9
A3M1_human	adapter-related protein complex 3 mu 1 subunit	VVLNMNLTPTQGSYTFDPVTK	2342.7

Table 2. Protein Identified in Other Bands

18 kDa gel band	PIR2: JE0350, H2BS, UBCC, RS19, RS20, MLEN, RLA2, RL23, RL31, RLA2, RS14, RS16, SYUG, H2B0, CYPH, H4, AOPP, GP:J04801.
38 kDa gel band	ANX1, ANX2, EF1D, G3P1, G3P2, LAS1, ROA2, RLAO, RL6. ROA1, AIP, MDHM, P2AA, NPM, GP:AJ278883, GPN:BC008837, PIR:S49326.
55 kDa gel band	EF11, ENOA, ECHB, ERP5, EF1G, OST4.
60 kDa gel band	ATPA, ATPB, DHAC, ROH1, TBB1, TCPB, TBA8, TBBQ, BC001365, BC001896, DESM, DHAM, ROH2, HS72.

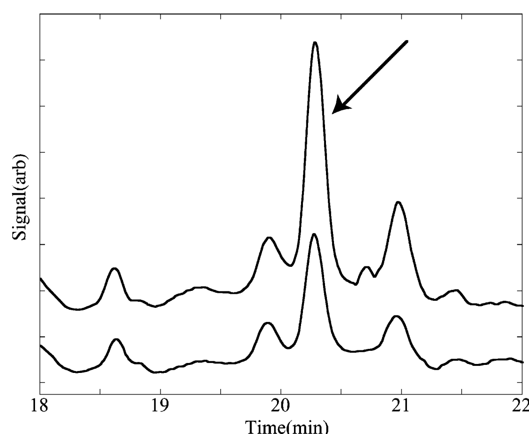

Figure 2. Spiking of capillary sieving electrophoresis using the proteins extracted from a gel band. The entire HT29 cell extract was separated by SDS−PAGE. The gel was treated with a zinc stain, and the component of interest was excised from the gel, purified, and used to spike the cell extract. The bottom electropherogram was obtained from the cell extract, and the top electropherogram was obtained from the spiked cell extract. The spiked component is marked with the arrow.

times is quite high in CSE, and comigration is easy to verify. This comigration confirmed that we had isolated the band of interest, Figure 2.

The band from the second gel was subjected to in-gel digestion and subsequent extraction of peptides for LC/MS/MS identification by peptide mapping and database searching using SEQUEST. A set of five proteins was identified and is listed in Table 1. These five proteins include two cytoskeletal proteins (K1CR and TBB1), an enzyme (SAHH), a protein associated with the ribonucleosome (ROD), and a protein associated with the Golgi apparatus (A3M1). This result points out a limitation of one-dimensional separation of complex protein samples: many components comigrate in a single band. Unfortunately, the ion intensity produced for any particular tryptic digest fragment will depend on the protein abundance, the digestion efficiency, the losses associated with handling the fragments due to adsorption, and the ionization efficiency of each fragment. As a result, we do not have information on the relative amounts of each protein in the band, and it is likely that the proteins are present at quite different levels.

Two ISO-DALT two-dimensional electropherograms have been published for this cell line.[10,11] In one case, 1500 spots were resolved and over 100 components were identified by mass spectrometry.[11] Most of the components identified in our study were also present in the two-dimensional gel. Unfortunately, the gel was not well enough annotated to guide identification of the most highly expressed proteins.

Of greater interest, SAGE (serial analysis by gene expression) data have been generated for this cell line.[12] Cytokeratin 18 is one of the most highly expressed mRNA in this cell line and constitutes 2.4% of the message in the HT-29 EST library. TBB1 was also present in this EST library for the HT-29 cell line, but at an order of magnitude lower level than cytokeratin 18. On the the basis of SAGE data, we tentatively conclude that cytokeratin 18 is the dominant protein present in the 47-kDa peak. As noted in the preceding paper, this protein undergoes a significant increase in phosphorylation upon entering the G2/M phase of the cell cycle,[13] which would account for the apparent increase in mobility for the target protein observed in single cell electropherograms.

We similarly analyzed a set of four other bands. The identified proteins are listed in Table 2. Several proteins were identified from each band. A separation with higher peak capacity will be required to resolve the proteins in those bands. We have developed an automated two-dimensional capillary electrophoresis system for protein analysis,[14] and we are working to apply that technology to the separation of proteins from single cancer cells. Identification of proteins observed in the two-dimensional capillary electrophoresis system will require the isolation of proteins from a classic two-dimensional ISO-DALT electrophoresis gel. The proteins will be identified by in-gel digestion and LC/MS/MS analysis and then

(10) Poland, J.; Sinha, P.; Siegert, A.; Schnölzer, M.; Korf, U.; Hauptmann, S. *Electrophoresis* **2002**, *23*, 1174−1184.
(11) Tan, S.; Seow, T. K.; Liang, R. C. M. Y.; Koh, S.; Lee, C. P. C.; Chung, M. C. M.; Hooi, S. C. *Int. J. Cancer* **2002**, *98*, 523−531.
(12) http://www.ncbi.nlm.nih.gov/UniGene/lib.cgi?ORG=Hs&LID=267&ALL=1
(13) Liao, J.; Lowthert, L. A.; Ku, N. O.; Fernandez, R.; Omary, M. B. *J. Cell Biol.* **1995**, *131*, 1291−1301.
(14) Michels, D.; Hu, S.; Schoenherr, R.; Dovichi, N. J. *Mol. Cell. Proteomics* **2002**, *1*, 69−74.

used as standards to spike the sample analyzed by two-dimensional capillary electrophoresis.

CONCLUSION

We demonstrate a straightforward method for the tentative identification of components in a CSE through comigration with a set of standard proteins. However, comigration is not sufficient for positive identification. Ultimately, it will be necessary to develop single-cell Western blot assays or to employ genetic engineering techniques to verify the identity of the minute amount of proteins contained within the capillary electrophoresis peak.

The main limitation to the technology is the extraction of the intact protein from the SDS−PAGE gel for use in spiking the sample used in CSE. Our current method, based on crushing of the gel and extracting the protein, results in poor recovery of the protein, which makes isolation of poorly expressed proteins difficult. Electroelution will likely be a more efficient method of protein isolation.

ACKNOWLEDGMENT

This work was supported by funding from the National Institutes of Health Grant DA016286 (N.J.D.), the Department of Defense Congressionally Directed Medical Research Program (N.J.D.), and National Institutes of Health Research Resource Grant RR11823 (R.A.).

Received for review February 17, 2003. Accepted March 29, 2003.

AC034154J