Anal. Chem. **2003,** *75,* 3495−3501

Cell Cycle-Dependent Protein Fingerprint from a Single Cancer Cell: Image Cytometry Coupled with Single-Cell Capillary Sieving Electrophoresis

Shen Hu,[†] Le Zhang,[†] Sergey Krylov,[‡] and Norman J. Dovichi*[,†]

Department of Chemistry, University of Washington, Seattle, Washington 98195, and Department of Chemistry, York University, Toronto, Ontario, Canada M3J 1P3

Study of cell cycle-dependent protein expression is important in oncology, stem cell research, and developmental biology. In this paper, we report the first protein fingerprint from a single cell with known phase in the cell cycle. To determine that phase, we treated HT-29 colon cancer cells with Hoescht 33342, a vital nuclear stain. A microscope was used to measure the fluorescence intensity from one treated cell; in this form of image cytometry, the fluorescence intensity is proportional to the cell's DNA content, which varies in a predictable fashion during the cell cycle. To generate the protein fingerprint, the cell was aspirated into the separation capillary and lysed. Proteins were fluorescently labeled with 3-(2-furoylquinoline-2-carboxaldehyde, separated by capillary sieving electrophoresis, and detected by laser-induced fluorescence. This form of electrophoresis is the capillary version of SDS−PAGE. The single-cell electropherogram partially resolved ∼25 components in a 30-min separation, and the dynamic range of the detector exceeded 5000. There was a large cell-to-cell variation in protein expression, averaging 40% relative standard deviation across the electropherogram. The dominant source of variation was the phase of the cell in the cell cycle; on average, ∼60% of the cell-to-cell variance in protein expression was associated with the cell cycle. Cells in the G1 and G2/M phases of the cell cycle had 27 and 21% relative standard deviations in protein expression, respectively. Cells in the G2/M phase generated signals that were twice the amplitude of the signals generated by G1 phase cells, as expected for cells that are soon to divide into two daughter cells. When electropherograms were normalized to total protein content, the expression of only one component was dependent on cell cycle at the 99% confidence limit. That protein is tentatively identified as cytokeratin 18 in a companion paper.

The cell is the organizing unit of life. Most cells do not proliferate (divide) in an adult organism, with the notable exception of pluripotent stem cells, which can produce different types of progenitor cells that proliferate to produce differentiated daughter cells, which do not divide. In healthy adults, proliferation is balanced by apoptosis and necrosis, so that the total cellular content is constant and homeostasis is preserved. In cancer, cells proliferate without control. Proliferating cells are also found throughout an embryo during development.

Nonproliferating cells are classified as G0. Proliferating cells pass through four phases of the cell cycle before division, and checkpoints associated with transition between those phases are vital in ensuring high fidelity genomic replication.[1] Recently divided cells are in the G1 phase of the cell cycle. Those cells progress to the S phase, in which they synthesize DNA to replicate their genetic material. Once they have duplicated their DNA, cells progress to the G2 phase, in which they rearrange their cytoskeleton and genetic material in preparation for mitosis in the M phase, in which cells divide to form two daughter cells, each in the G1 phase.

The phase of a cell in the cell cycle can be estimated from the cell's DNA content. Normal G1 cells are diploid; they contain two sets of chromosomes. Cells in the G2 and M phase are tetraploid; they contain four sets of chromosomes. S-phase cells are in the process of replicating their chromosomes and these cells contain intermediate amounts of DNA. Cancer cells contain genetic abnormalities, which often include missing, fragmented, or duplicated chromosomes. The HT-29 colon adenocarcinoma cell line used in this paper is classified as hypertriploid, in which most chromosomes are found as three copies per G1-phase cell; some cells contain additional copies of the chromosomes.[2] Despite the abnormal genomic makeup, each cell's DNA content doubles as the cell pass through the S phase of the cell cycle.

In practice, the phase of the cell in the cell cycle is determined by staining the cell with an intercalating DNA dye. These dyes are nonfluorescent until they bind to DNA. Upon binding, the dye becomes fluorescent with an intensity that is proportional to the amount of DNA; cells in the G2 and M phases generate a fluorescence signal that is twice that of G1 phase cells. Cells in the S phase generate an intermediate amplitude fluorescence signal. Cells in the G2 and M phases are not distinguished by this method and are lumped together as G2/M-phase cells.

Hoechst 33342 is a commonly used intercalating dye for cell cycle measurements.[3] This dye is a vital stain and is taken up by

[†] University of Washington.
[‡] York University.

(1) Hartwell, L. H.; Weinert, T. A. *Science* **1989**, *246*, 629−634.
(2) Kawai, K.; Viars, C.; Arden, K.; Tarin, D.; Urquidi, V.; Goodison, S. *Genes, Chromosomes Cancer* **2002**, *34*, 1−8.
(3) Arndtjovin D. J.; Jovin T. M. *J. Histochem. Cytochem.* **1977**, *25*, 585−589.

10.1021/ac034153r CCC: $25.00
Published on Web 05/31/2003

living cells. Most other dyes require that the cell be fixed before staining. We have reported the use of Hoechst 33342 to characterize the phase of a single living cell in the cell cycle before performing analysis of carbohydrate metabolism.[4]

Most studies of cell cycle-dependent protein expression focus on a specific protein, which is detected by immunostaining and cytometry of single cells. More global studies employ Western blotting of proteins extracted from a few hundred thousand cells that are arrested at a cell-cycle checkpoint.[5] There are three problems with these analyses. First, the cell populations are contaminated by as much as 30% of cells at other phases in the cell cycle, which makes difficult the detection of small changes in protein expression.[5] Second, it is not always clear if the change in protein expression is due to changes in the cell cycle or is in response to treatment to arrest the cells at the checkpoint. Third, while these studies provide information on the changes in protein expression *between* phases of the cell cycle, they provide no information on the cell-to-cell distribution of protein expression *within* a phase of the cell cycle.

In this paper, we report the analysis of protein expression from a single cell at a known phase of the cell cycle. This study provides unambiguous information on the cell-to-cell variation in protein expression at each stage in the cell cycle without treating the cells in a manner that dramatically perturbs protein expression. Because cells are not arrested at the checkpoint, we determine the distribution of protein expression within a phase of the cell cycle.

The use of separation methods to characterize the composition of a single cell has a 50-year history. The first study in 1953 considered rRNA analysis in single cells on the basis of electrophoresis on a silk fiber.[6] The earliest single-cell protein analysis was a study of hemoglobin in single erythrocytes by electrophoresis through an acrylamide fiber, published in 1965.[7] Repin reported the characterization of lactate dehydrogenase isoenzymes in single mammalian oocytes by electrophoresis in 1975.[8] Ruchel reported the first protein analysis from a single giant neuron from a sea snail the following year.[9]

Kennedy and Jorgenson inaugurated the modern era of single cell analysis by using open tubular capillary chromatography for the amino acid analysis of a single giant neuron from a snail.[10−11] At the same time, Wallingford and Ewing reported the use of a capillary to sample the internal contents of a single giant neuron; they used the same capillary for electrophoresis of biogenic amines.[12] In 1992, Hogan and Yeung reported the use of a specific label to derivatize thiols in individual erythrocytes; once the derivatization reaction was completed, a single cell was injected into a capillary, lysed, and the contents separated by capillary electrophoresis.[13] In 1995, Gilman and Ewing reported the on-column labeling of amines from a cell that had been injected into and lysed within a capillary; the capillary was used for electrophoretic separation of amines.[14] Allbritton reported in 2000 the use of a pulsed laser to lyse a cell before analysis of kinase activities by capillary electrophoresis.[15] Last year, Lillard reported the use of cell synchronization based on the shake-off method for the characterization of RNA synthesis in single cells as a function of cell cycle.[16]

This group's activity in single cell analysis is relatively recent. We reported studies of the metabolism of carbohydrates in single yeast and cancer cells in 1999.[4,17] In ref 4, the phase of the cell in the cell cycle was determined before analysis. We observed biodegradation of a disaccharide substrate in all phases of the cell cycle, but only observed biosynthesis in the G2/M phase of the cell cycle. More recently, we have reported on single cell protein analysis.[18−21] Our protein experiments are similar to those of Gilman and Ewing: a single cell is injected into a capillary and lysed, and the contents are fluorescently labeled. The products are separated by capillary electrophoresis and detected by laser-induced fluorescence.

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.[22] High voltage was provided by a 0−30 kV dc power supply (CZE 1000, Spellman, Plainview, NY). A 12-mW argon ion laser (model 2211-15SL, Uniphase, San Jose, CA) beam at a wavelength of 488 nm was used to excite fluorescence, which was collected with a 60×, 0.7 NA microscope objective, filtered with a 630DF30 band-pass filter (Omega Optical, Brattleboro, VT), and then detected with a Hamamatsu R1477 photomultiplier tube. Data were recorded at 10 Hz through a 10-Hz band-pass RC filter with a PC equipped with a 16-bit data acquisition board. Data collection software was written in LabView.

Fused-silica capillaries with 140-μm o.d., 50-μm i.d., and 40-cm length (Polymicro Technologies, Phoenix, AZ) were used in this work. The capillaries were coated with linear polyacrylamide, as described previously.[19]

Reagents. Unless otherwise stated, all reagents were obtained from Sigma (St. Louis, MO). Hoechst 33342, 3-(2-furoyl-quinoline-2-carboxaldehyde (FQ), and KCN were obtained from Molecular Probes (Eugene, OR). Pullulan was used as the sieving matrix for capillary sieving electrophoresis. 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).

(4) Krylov, S. N.; Zhang, Z.; Chan, N. W. C.; Arriaga, E.; Palcic, M. M.; Dovichi, N. J. *Cytometry* **1999**, *37*, 15−20.
(5) Ku, N. O.; Liao, J.; Omary, M. B. *EMBO J.* **1998**, *17*, 1892−1906.
(6) Edstrom, J. E. *Nature* **1953**, *172*, 908.
(7) Maholi, G. T.; Niewisch, H. B. *Science* **1965**, *150*, 1824−1826.
(8) Repin, V. S.; Akimova, I. M.; Terovskii, V. B. *Bull. Exp. Biol. Med.* **1975**, *77*, 767−769.
(9) Ruchel, R. *J. Histochem. Cytochem.* **1976**, *24*, 773−791.
(10) Kennedy, R. T.; St. Claire, R. L.; White, J. G.; Jorgenson, J. W. *Mikrochim. Acta* **1987**, *II*, 37−45.
(11) Kennedy, R. T.; Oates, M. D.; Cooper, B. R.; Nickerson, B.; Jorgenson, J. W. *Science* **1989**, *246*, 57−63.
(12) Wallingford, R. A.; Ewing, A. G. *Anal. Chem.* **1988**, *60*, 1972−1975.
(13) Hogan, B. L.; Yeung E. S. *Anal. Chem.* **1992**, *64*, 2841−2845.
(14) Gilman, S. D.; Ewing, A. G. *Anal. Chem.* **1995**, *67*, 58−64.
(15) Meredith, G. D.; Sims, C. E.; Soughayer, J. S.; Allbritton, N. L. *Nat. Biotechnol.* **2000**, *18*, 309−312.
(16) Han, F. T.; Lillard, S. J. *Anal. Biochem* **2002**, *302*, 136−143.
(17) Le, X. C.; Tan, W.; Scaman, C.; Szpacenko, A.; Arriaga, E.; Zhang, Y.; Dovichi, N. J.; Hindsgaul, O.; Palcic, M. M. *Glycobiology* **1999**, *9*, 219−225.
(18) Zhang, Z.; Krylov, S.; Arriaga, E. A.; Polakowski, R.; Dovichi, N. J. *Anal. Chem.* **2000**, *72*, 318−322.
(19) Hu, S.; Zhang, L.; Cook, L. M.; Dovichi, N. J. *Electrophoresis* **2001**, *22*, 3677−3682.
(20) Hu, S.; Lee, R.; Zhang, Z.; Krylov, S. N.; Dovichi, N. J. *J. Chromatogr., B* **2001**, *752*, 307−310.
(21) Hu, S.; Jiang, J.; Cook, L.; Richards, D. P.; Horlick, L.; Wong, B.; Dovichi, N. J. *Electrophoresis* **2002**, *23*, 3136−3142.
(22) Wu, S.; Dovichi, N. J. *J. Chromatogr., A* **1989**, *480*, 141−155.

Cell Culture. HT-29 human colon adenocarcinoma cells (American Type Culture Collection, ATCC no. HTB-38) 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: 8 g NaCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4, pH adjusted to 7.4 with HCl, diluted to 1.00 L with distilled water) to remove the residual substrate from the medium immediately before single cell analysis.

Cell Cycle Analysis. Image cytometry was used to determine the DNA content in individual HT-29 cancer cells. Hoechst 33342 was added to the cell suspension (10^6 cells/mL) to give a final concentration of 10 μg/mL. Cells were incubated at 37 °C for 1.0 h to stain the cellular DNA. A 1-μL aliquot of the cell suspension was diluted with 50 μL of PBS buffer, and this diluted suspension was placed on a glass microscope slide. The fluorescence signal was detected by using an inverted fluorescence microscope equipped with an electronically controlled PMT shutter and a Hamamatsu R1635-02 photomultiplier tube.[4] Once a cell was centered in the field of view under the microscope, the shutter was opened and DNA fluorescence signal was monitored by the photomultiplier tube. Automatic focusing of the microscope was performed to adjust and maximize the fluorescence signal when the shutter was open.

Cell Cycle-Resolved Single-Cell Protein Expression. Single cell injection was performed using a multipurpose injection apparatus.[23] The injection process was observed through the microscope. A hydraulic micromanipulator (model MX630R Newport, Nepean, ON, Canada) was used to control the capillary's position. Immediately before single cell electrophoresis, KCN was added to the suspension to give a final concentration of 2 mM. A drop of the cell suspension and a drop of 0.01 M FQ with 1.0% SDS solution were placed at different locations on a microscope slide.

The generation of the single cell protein electropherogram required several steps. First, a cell was centered in the field of view of the microscope, and the cell's DNA content was determined. Second, the capillary tip was lowered into the drop of FQ and SDS, and an 11-kPa negative pressure was applied to the distal end of the capillary for 2 s to inject a plug of SDS and FQ into the capillary. Third, the capillary tip was centered over the cell and that cell was injected into the capillary by aspiration at the 11-kPa negative pressure for 1 s. Finally, the capillary tip was moved back to the drop of FQ and SDS, which was injected again for 2 s. The capillary tip was then placed in a vial containing the pullulan sieving buffer and heated at 90 °C for 5 min to denature and label the proteins. After reaction, the separation was performed at −300 V/cm.

Cellular Homogenate. Proteins were prepared from a cellular homogenate. Roughly 10^6 HT-29 human colon adenocarcinoma cells were washed five times with PBS and resuspended in 100 μL of water. The cells were sonicated for 20 min at 4 °C. The suspension was centrifuged at 1000 rpm (82*g*) for 10 min to remove debris.

Molecular Weight Standards. A set of eight proteins (lactoalbumin, lactoglogulin, trypsin inhibitor, carbonic anhydrase, ovalbumin, BSA, conalbumin, β-galactosidase) was analyzed using identical conditions as in the single-cell analysis. The migration times of the peaks generated by these proteins were used to calibrate the molecular weight axis of the capillary sieving electropherograms.

Data Processing. Electropherograms were imported to Matlab, where they were treated with a five-point median filter to remove noise spikes and convoluted with a Gaussian function that had a four-point standard deviation to reduce noise. The time axis was normalized to correct for drift in the migration time.[24] Matlab was also used for subsequent statistical analysis of the data. Three peaks saturated the detector for at least one cell and were not included in subsequent analysis. In addition, the region of the electropherogram before the first peak was not included in the analysis.

RESULTS AND DISCUSSION

Nomenclature. A number of names have been used to describe the separation of proteins in a sieving matrix. Neither SDS−PAGE nor capillary gel electrophoresis is an appropriate term; polyacrylamide is not always used as the sieving matrix, and non-cross-linked polymers do not form gels. We had proposed the use of the term capillary SDS-DALT electrophoresis to describe this separation method. However, the IUPAC has recently recommended the term "capillary sieving electrophoresis (CSE)" to describe separations in a capillary filled with a sieving medium, and we adopt that nomenclature.[25]

Protein Expression of a Single Cancer Cell. The analytical procedure to generate an electropherogram on the basis of the protein content of a single cancer cell requires four steps. First, a cell is injected into an electrophoresis capillary. Second, the cell is lysed inside the capillary. Third, released proteins are fluorescently labeled. Last, the labeled proteins are separated by capillary electrophoresis and detected by laser-induced fluorescence. We have applied this method to the analysis of proteins expressed in single cells on the basis of both capillary submicellar electrophoresis and capillary sieving electrophoresis.[18−21]

Figure 1 presents the electropherograms generated from 11 cells, along with the background electropherogram generated by analysis of the cellular suspension medium. Proteins will be present in the cellular supernatant. These proteins come from the cell culture medium and from lysed cells. Those proteins can contaminate the single-cell data. We wash the cells several times with isotonic buffer before analysis to remove those proteins, and the resulting background signal consists of a low-amplitude, broad peak that does not interfere in the measurement of the single cell protein fingerprint.

Between 20 and 30 components are partially resolved from each cell. Most peaks undoubtedly contain a number of proteins that comigrate. Intriguingly, some peaks, particularly the 80- and 100-kDa components, generated relatively sharp peaks (N > 140 000 plates). It is likely that those peaks are predominantly composed of a single protein with other components present at much lower levels.

(23) Krylov, S. N.; Starke, D. A.; Arriaga, E. A.; Zhang, Z.; Chan, N. W.; Palcic, M. M.; Dovichi, N. J. *Anal. Chem.* **2000**, *72*, 872−877.
(24) Li, X. F.; Ren, H.; Le, X.; Qi, M.; Ireland, I. D.; Dovichi, N. J. *J. Chromatogr.* **2000**, *869*, 375−384.
(25) Riekkola, M.-L.; Jönsson, J. Å. International Union Of Pure And Applied Chemistry, Terminology For Analytical Capillary Electromigration Techniques (http://www.iupac.org/reports/provisional/abstract01/riekkola_310102.html).

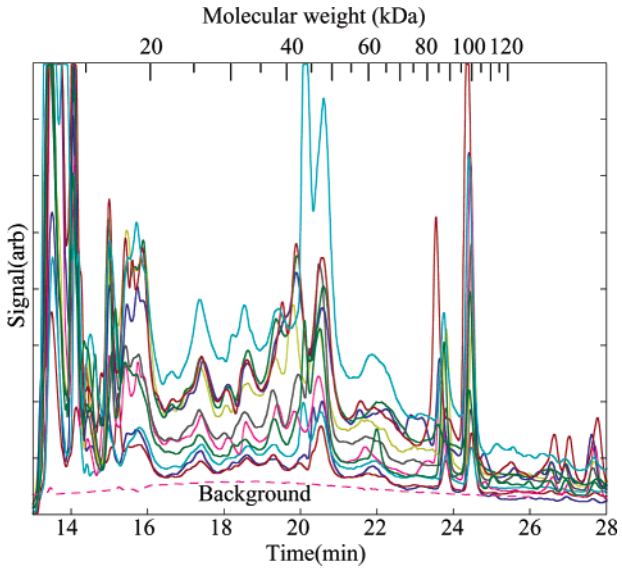

Figure 1. Capillary sieving electrophoresis analysis of 11 different HT-29 human cancer cells. The background signal, generated by analysis of the cellular supernatant, is the dashed curve.

Several peaks from a single cell saturated our detector, including the fastest migrating component, which may include peptides in addition to low molecular weight proteins. We performed CSE analysis of FQ-labeled insulin (2.5 kDa) and aprotinin (6.5 kDa), data not shown; those peptides migrate before other standard proteins. No peaks were observed for the analysis of most amino acids, presumably because of the low labeling efficiency of the α-amine.

The peaks that saturated our detector were not included in the statistical analysis of the data presented below. The noise was quite low when analyzing the reagent blank, and the dynamic range of the measurement, which is the ratio of the largest signal to the noise in the baseline, exceeded 5000.

There was a large cell-to-cell variation in fluorescence signal across the electropherogram. The relative standard deviation in fluorescence intensity ranged from a maximum of 100% to a minimum of 10%, with an average of 40%. It is important to identify the source of this cell-to-cell variation so that physiologically relevant differences can be distinguished from variability associated with the analytical procedure in single-cell protein studies.

Cell Cycle-Dependent Protein Expression of a Single Cancer Cell. We hypothesize that the phase of the cell in the cell cycle accounts for the majority of the variation in protein expression between cells. To test this hypothesis, we determined the phase of the cell in the cell cycle before generating the protein fingerprints. Figure 2 presents a micrograph of HT-29 cancer cells after incubation with Hoechst 33342. The fluorescence intensity is related to the phase of the cell in the cell cycle. A histogram was generated on the basis of the fluorescence intensity from the vital DNA stain and used to classify cells into the G1 and G2/M phases.[4] For the cells used in this paper, the DNA fluorescent signal was 0.51 ± 0.08 V for cells in the G1 phase and 1.13 ± 0.15 V for cells in the G2/M phase. Five G1- and five G2/M-phase cells were taken for single-cell protein analysis. One cell generated a DNA signal of 1.5 V, and that cell likely contained an additional copy of each chromosome; this cell was also taken for single-cell protein analysis. Like most cancer cells, the HT-29 cells have

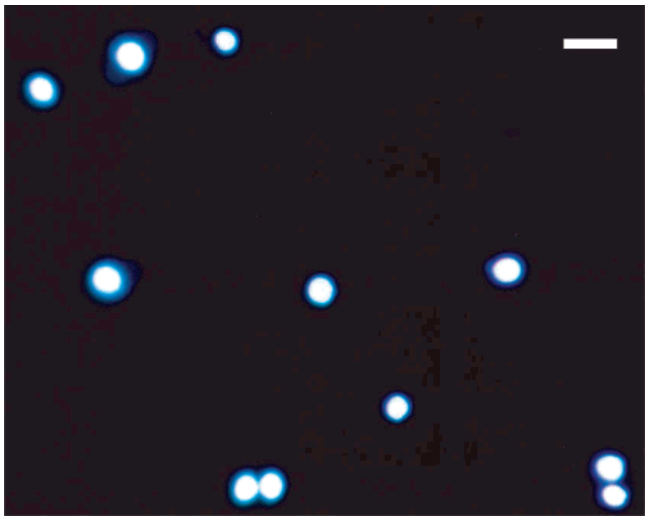

Figure 2. Photograph of HT-29 cancer cells after incubation with Hoechst 33342. The scale bar in the top right-hand corner is 20 μm.

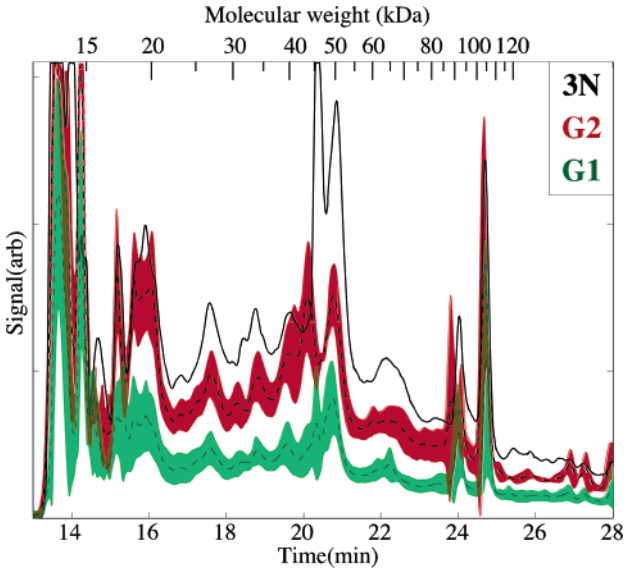

Figure 3. Mean (dashed trace) and 95% confidence interval (colored regions) for five cells in the G1 (green) and five cells in the G2/M (red) phases in the cell cycle. The signal generated by the cell whose DNA fluorescence signal was three times that of G1 phase cells is shown in black.

abnormal chromosomal content, and karyotyping sometimes reveals extra copies of chromosomes in this cell line.[2]

Cell-to-Cell Variation in Protein Expression within a Phase of the Cell Cycle. To summarize the distribution of protein expression within a phase in the cell cycle, Figure 3 presents the 95% confidence interval of peak amplitude for cells within the G1 phase (green) and G2/M phase (red) of the cell cycle. The mean amplitude of the electropherogram for G2/M phase cells was 1.9 ± 0.3 times larger than for cells in the G1 phase. This factor of 2 difference in signal amplitude arises because cells in the G2/M phase of the cell cycle are about to divide into two G1 daughter cells; G2/M phase cells are expected to have twice the protein content of the G1 phase cells.

There are a few components for which the two distributions overlap, including the components smaller than 15 kDa and near 105 kDa, where the detector was saturated for one or more cells,

and the components at 16, 18, and 95 kDa. This distribution graphically demonstrates the fleeting existence of some proteins as the cells progress through the cell cycle.

The relative standard deviation in protein expression for cells in either the G1 or the G2/M phase is half that of the overall standard deviation of the original population. The relative standard deviation in peak amplitude for cells in the G1 phase of the cell cycle ranged from 69 to 11%, with an average of 27%. The relative standard deviation in peak amplitude for cells in the G2 phase of the cell cycle ranged from 86 to 2%, with an average of 20%.

The residual cell-to-cell variation in protein expression for cells within the G1 and G2/M phases likely arises in part because the cells are captured at different points *within* the phase of the cell cycle. Cells that have recently entered the G1 phase are likely to have a protein content that is different from those cells that are about to leave that phase. Other sources of cell-to-cell variability will be associated with subtle differences in cell handling, including the confluence of the culture and the length of time between removal of the cells from the culture medium and their analysis, and the accumulated genetic damage of any particular cell. Finally, cells early in the S phase may have been misclassified as G1-phase cells; similarly, cells late in the S phase may have been classified as G2/M-phase cells.

Statistical Tests. The *t*-test was used to identify those regions of the electropherogram that differed significantly between cells at different phases of the cell cycle. The signal generated by cells in the G2/M phase of the cell cycle was larger than the signal generated by cells in the G1 phase of the cell cycle at the 99% confidence limit (8 degrees of freedom) throughout the electropherogram, except for regions noted in the previous section.

We investigated the contribution to the cell-to-cell variance associated with the phase of the cell cycle. As shown in Appendix 1, if equal numbers of cells are taken for analysis from the G1 and the G2/M phases of the cell cycle, the overall variance ($\sigma^2_{\text{overall}}$) can be written as

$$\sigma^2_{\text{overall}} = \frac{n/2 - 1}{n - 1}(\sigma^2_{\text{G1}} + \sigma^2_{\text{G2}}) + \frac{n/2}{n - 1}\sigma^2_{\text{phase}} \qquad (1)$$

where n is the total number of cells analyzed (10 G1 and G2/M phase cells in this paper), σ^2_{G1} and σ^2_{G2} are the variances for cells in the G1 and G2/M phases of the cell cycle, and σ^2_{phase} is the variance of the mean signal for G1 and G2/M phase cells.

The last term ($n/2/n - 1\sigma^2_{\text{phase}}$) is the contribution to the overall variance associated with the phase of the cell in the cell cycle. The fraction of the total variance that was associated with the phase of the cell in the cell cycle across the electropherogram ranged from 40 to 70%, with a mean of 55%. This result is vital, because it identifies the major contribution to cell-to-cell variation in protein expression, which is the phase of the cell in the cell cycle.

Single Cell Analysis of a Rare Cell. Single-cell protein analysis when combined with image cytometry allows study of rare cells. As an example, one unusual cell, which was larger than others in this culture, generated a DNA fluorescence signal that was three times that of a G1 phase cell, presumably because the cell contained an extra copy of each chromosome. The electropherogram for this cell is shown as the black curve labeled 3N in Figure 3. The average amplitude of the signal generated by this cell is 2.8 times larger than that of the G1 phase cells, which demonstrates that this cell had tripled its protein as well as its chromosomal content. The capillary sieving electropherogram produced two unusually large peaks, one at 14.5 min and one at 20 min, which saturated our fluorescence detector and which were present at much higher concentrations than in the G1 and G2/M single cell electropherograms.

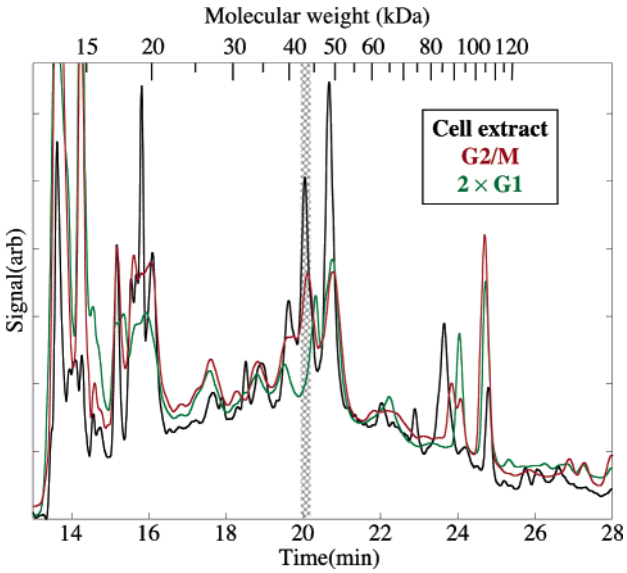

Figure 4. Mean fluorescence intensity for cells in the G2/M phase of the cell cycle (red), twice the mean intensity for cells in the G1 phase (green), and fluorescence signal generated by the cellular homogenate from an HT-29 cell culture (black). The hatched region denotes the one peak that differed at the 99% confidence limit between the 2× G1 and the G2/M electropherograms.

Cell-to-Cell Variation in Protein Expression Normalized for the Total Protein Content of the Cell. To identify cell cycle-dependent protein expression, it is necessary to correct for differences in cell mass. We multiplied the average G1-phase cell signal by 2 and plotted it with the average electropherogram for a G2/M phase cell, Figure 4. The two profiles are very similar; the difference in protein expression profiles for G1 and G2/M phase cells is dominated by the difference in the size of cells in the two phases.

Only the 45-kDa component differed between the two traces at the 99% confidence level, which corresponds to a candidate cell cycle-dependent protein. The peak appears to undergo a shift to faster migration time in G2/M-phase cells, as compared with G1 phase cells. As noted below, that peak fortuitously is also prominent in the electropherogram from the cellular homogenate. We address the tentative identification of this protein as cytokeratin 18 in a companion paper;[26] this protein is known from immunoblotting experiments to undergo a 3- to 4-fold increase in phosphorylation upon entering the G2/M phase of the cell cycle in the HT-29 cell line.[27] That additional negative charge would account for the faster migration time of the peak in the G2/M phase.

(26) Hu, S.; Zhang, L.; Newitt, R.; Aebersold, R.; Kraly, J.; Jones, M.; Dovichi, N. J. *Anal. Chem.* **2003**, *75*, 3502−3505.
(27) Liao, J.; Lowthert, L. A.; Ku, N. O.; Fernandez, R.; Omary, M. B. *J. Cell Biol.* **1995**, *131*, 1291−1301.

The similarity of the electrophoretic profiles is not surprising. In the only published two-dimensional gel electrophoretic analysis of cell cycle-dependent protein expression, over 1000 spots were resolved from proteins isolated from Chinese hamster ovary cells that had been treated to arrest cells at different phases of the cell cycle.[28] Only 1% of those proteins had significant changes in expression between the G1 and S phases of the cell cycle.

Protein Electropherogram Generated from Cellular Homogenate. Figure 4 includes the electropherogram of a cellular homogenate that contained several micrograms of proteins. The electropherogram from the homogenate is similar to that generated by single cells. However, there are notable differences, particularly the expression of an intense 19-kDa component and in the dramatic shift of the 80-kDa peak to lower molecular weight. There are several likely causes of these differences. As we have demonstrated elsewhere, single-cell analysis reduces the effects of hydrolytic enzymes, as compared with the preparation of bulk samples; the rapid lysis and subsequent electrophoresis separates hydrolytic enzymes from their substrate, eliminating degradation.[29] In addition, the more rigorous solublization protocol used in preparing the homogenate would disrupt protein−protein complexes that may survive the single-cell protocol.

We had been concerned that single-cell protein analysis only sampled the cytosolic fraction of the cell and that membrane-bound proteins were adsorbed to the capillary wall and lost from analysis. We earlier described the use of CSE to separate the cytosolic, membrane/organelle, nuclear, and cytoskeletal/nuclear matrix fraction of the HT29 cell line.[21] Each of those electropherograms was dramatically different from that of the cellular homogenate. On the basis of the similarity of the electropherogram generated from the cellular homogenate and from single G1- and G2/M-phase cells, single-cell protein analysis appears to sample a representative fraction of the entire cell. Of course, a higher resolution separation may show more subtle differences than can be resolved by the one-dimensional CSE data presented here.

CONCLUSION

Our motivation for this work is driven by cancer prognosis. We have hypothesized that the cell-to-cell variation in protein expression increases in later stages of progression. The study of the HT-29 cancer cell line serves as a baseline that defines the minimum detectable change in protein expression. Our demonstration that cell-to-cell variation in protein expression is dominated by differences in the cell cycle is important. By fingerprinting the protein content of cells at specific stages in the cell cycle, we eliminate the greatest source of variation in protein expression, which will allow greater precision in monitoring protein expression and facilitate detection of protein expression changes that correlate with progression of the disease. Our hypothesis may now be refined: the cell-to-cell variation in protein expression *at each stage of the cell cycle* increases in later stages of progression.

This experiment has several limitations. First, the amount of material contained within a cell is far too little to identify using classic methods. In a companion paper, we report on one method to tentatively identify proteins. Second, the separation efficiency of one-dimensional electrophoresis is not sufficient to resolve the large number of proteins present in a single cell. It will be necessary to develop a two-dimensional capillary electrophoresis system for single cell analysis. Such an instrument has been reported for the analysis of proteins extracted from a large number of cells;[30] work is in progress to adapt that technology for single cell analysis.

APPENDIX 1

The cell-to-cell variance in protein expression at some migration time is defined as

$$\sigma^2_{\text{overall}} = \frac{1}{n-1}\sum(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{\text{G1}}(x_{\text{G1}} - \bar{x})^2 + \frac{1}{n-1}\sum_{\text{G2}}(x_{\text{G2}} - \bar{x})^2 \quad \text{(A1)}$$

where x_i is the fluorescent intensity for all cells at some point in time, $\bar{x}$ is the average intensity for all cells, n is the total number of cells analyzed, x_{G1} is the fluorescence intensity for cells in the G1 phase, and x_{G2} is the fluorescence intensity for cells in the G2 phase. Each summation can be rewritten as

$$\frac{1}{n-1}\sum_{\text{G1}}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{\text{G1}}(x_{\text{G1}} - \overline{x_{\text{G1}}} + \overline{x_{\text{G1}}} - \bar{x})^2$$

$$\frac{1}{n-1}\sum_{\text{G2}}(x_{\text{G2}} - \bar{x})^2 = \frac{1}{n-1}\sum_{\text{G2}}(x_{\text{G2}} - \overline{x_{\text{G2}}} + \overline{x_{\text{G2}}} - \bar{x})^2 \quad \text{(A2)}$$

where $\overline{x_{\text{G1}}}$ is the mean intensity for G1-phase cells and $\overline{x_{\text{G2}}}$ is the mean intensity for G2/M-phase cells. Expanding the sums,

$$\sum_{\text{G1}}(x_{\text{G1}} - \overline{x_{\text{G1}}} + \overline{x_{\text{G1}}} - \bar{x})^2 = \sum_{\text{G1}}(x_{\text{G1}} - \overline{x_{\text{G1}}})^2 + \sum_{\text{G1}}(\overline{x_{\text{G1}}} - \bar{x})^2 + 2\sum_{\text{G1}}(x_{\text{G1}} - \overline{x_{\text{G1}}})(\overline{x_{\text{G1}}} - \bar{x})$$

$$\sum_{\text{G2}}(x_{\text{G2}} - \overline{x_{\text{G2}}} + \overline{x_{\text{G2}}} - \bar{x})^2 = \sum_{\text{G2}}(x_{\text{G2}} - \overline{x_{\text{G2}}})^2 + \sum_{\text{G2}}(\overline{x_{\text{G2}}} - \bar{x})^2 + 2\sum_{\text{G2}}(x_{\text{G2}} - \overline{x_{\text{G2}}})(\overline{x_{\text{G2}}} - \bar{x}) \quad \text{(A3)}$$

The middle term is simplified by $\sum_{\text{G1}}(\overline{x_{\text{G1}}} - \bar{x})^2 = n_{\text{G1}}(\overline{x_{\text{G1}}} - \bar{x})^2$ and $\sum_{\text{G2}}(\overline{x_{\text{G2}}} - \bar{x})^2 = n_{\text{G2}}(\overline{x_{\text{G2}}} - \bar{x})^2$, where n_{G1} is the number of cells in the G1 phase and n_{G2} is the number of cells in the G2/M phase. The last term of both summations equals 0 because the average deviations from the mean equal 0. The overall variance can be written as

(28) Naryzhny, S. N.; Lee, H. *Electrophoresis* **2001**, *22*, 1764−1775.
(29) Krylov, S. N.; Arriaga, E.; Zhang, Z.; Chan, N. W. C.; Palcic, M. M.; Dovichi, N. J. *J. Chromatogr., A* **2000**, *741*, 31−35.
(30) Michels, D.; Hu, S.; Schoenherr, R.; Dovichi, N. J. *Mol. Cell. Proteomics* **2002**, *1*, 69−74.

$$\sigma^2_{\text{overall}} = \frac{1}{n-1}\sum_{G1}(x_{G1} - \overline{x_{G1}})^2 + \frac{1}{n-1}\sum_{G2}(x_{G2} - \overline{x_{G2}})^2 + \frac{n_{G1}}{n-1}(\overline{x_{G1}} - \bar{x})^2 + \frac{n_{G2}}{n-1}(\overline{x_{G2}} - \bar{x})^2 \quad \text{(A4)}$$

The first sum is related to σ^2_{G1}; the second, to σ^2_{G2}. If equal numbers of G1 and G2/M cells are analyzed, $n_{G1} = n_{G2} = 1/2\ n$, and

$$\sigma^2_{\text{overall}} = \frac{n/2-1}{n-1}(\sigma^2_{G1} + \sigma^2_{G2}) + \frac{n/2}{n-1}^2[(\overline{x_{G1}} - \bar{x})^2 + (\overline{x_{G2}} - \bar{x})^2] \quad \text{(A5)}$$

where the term in brackets is simply the variance of the mean signal in the two phases, which we write as σ^2_{phase},

$$\sigma^2_{\text{overall}} = \frac{n/2-1}{n-1}(\sigma^2_{G1} + \sigma^2_{G2}) + \frac{n/2}{n-1}\sigma^2_{\text{phase}} \quad \text{(A6)}$$

ACKNOWLEDGMENT

This work was supported by funding from the National Institutes of Health Grant DA016286 and the Department of Defense Congressionally Directed Medical Research Program. We thank Dr. Rakesh Bhatnagar for assistance in taking the fluorescence micrograph.

Received for review February 17, 2003. Accepted April 11, 2003.

AC034153R