High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry

Ljiljana Paša-Tolić, Pamela K. Jensen, Gordon A. Anderson, Mary S. Lipton, Kim K. Peden, Suzana Martinović, Nikola Tolić, James E. Bruce, and Richard D. Smith*

Environmental Molecular Sciences Laboratory
Pacific Northwest National Laboratory
P.O. Box 999, Richland, Washington 99352

Received April 5, 1999

In contrast to a cell's virtually static genome, the proteome,[1] the protein complement expressed by an organism, continually changes in response to external stimuli and internal processes.[2] Global gene expression analysis at the mRNA level (i.e., transcriptome) has recently become feasible based on the serial analysis of gene expression[3] and oligonucleotide micro-array assays.[4] These techniques allow the activation states of thousands of genes to be polled simultaneously for a tissue or cell population. However, assays that measure mRNA abundances rather than the functional gene products (i.e., proteins) are uninformative with regard to protein modifications, and can poorly reflect protein abundances due to differences in stabilities, expression rates, etc., for both the mRNAs and proteins.[5]

The global study of protein abundances is now commonly accomplished using two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) methods capable of visualizing thousands of protein spots, generally with 1−100 fmol detection limits.[2] In recent years 2-D PAGE has been increasingly augmented by subsequent use of matrix-assisted laser desorption/ionization[6] or electrospray ionization (ESI)[7] mass spectrometry for protein characterization.[2,8] Proteome characterization remains constrained by both the speed and reproducibility of 2-D PAGE separations and the inability to obtain precise measurements of expression levels due to difficulties in controlling the range of variables associated with cell lysis, sample handling, 2-D PAGE separation and recovery from the gel, as well as variations in mass spectrometric response.

There are currently a number of methods being investigated to avoid the use of 2-D PAGE for the study of complex protein extracts.[9−12] One approach, reported by Lee and co-workers,[10] involves the use of capillary isoelectric focusing (CIEF) combined with ESI-MS, which has recently been extended by the use of on-line Fourier transform ion cyclotron resonance (FTICR) mass spectrometry.[11,12] FTICR[13] provides mass measurements of much higher resolution and accuracy than that for other types of mass spectrometers, as well as other capabilities useful for protein characterization (e.g., dissociation studies using tandem MS/MS to extract sequence and other structural information).[14] McLafferty and co-workers have obtained extensive sequence information from intact proteins using ESI-FTICR, and have demonstrated sub-attomole detection levels with on-line capillary electrophoresis.[15] Our recent studies have revealed up to 1000 putative proteins in the 2−100 kDa molecular mass (M_r) range from 200 to 300 ng of soluble *Escherichia coli* proteins (compared to 100 μg to > 1 mg total protein typically used for 2-D PAGE) in a single 30 min CIEF-FTICR analysis.[12] Assuming an average M_r of 30 kDa, the average protein detected is present at ∼10 fmol, and the lower abundant species obviously correspond to amol range levels. This is consistent with detection limits of 10 to 100 amol obtained in this work by spiking *E. coli* extracts with known proteins (e.g., carbonic anhydrase). The use of on-line separations with mass spectrometry would be greatly augmented by a broadly applicable quantitative measurement capability.

We have developed an approach utilizing organisms cultured in stable-isotope labeled media (e.g., rare-isotope depleted and normal) to provide effective "internal calibrants" for *all* detected proteins, thus enabling precise proteome-wide measurement of changes in protein abundances resulting from cellular perturbations. The two (or more) isotopically distinctive cell populations are mixed prior to sample processing steps, eliminating all experimental variables associated with cell lysis, separation, and mass spectrometric analysis. Changes in relative protein abundances are thus precisely reflected by the ratio of two isotopically different and resolvable versions of each protein.

To demonstrate this approach we used CIEF-FTICR to examine cadmium (Cd^{2+}) stress response in *E. coli* K-12 MG1655.[16] The high resolution of FTICR facilitated comparative proteome displays from mixtures of cells grown in both normal media and an otherwise identical ^{13}C-, ^{15}N- and ^{2}H-"depleted" media.[17] Marshall et al. have previously shown that improved accuracy, sensitivity, dynamic range, and detection limits are obtained for

* Address correspondence to: Richard D. Smith, Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, P.O. Box 999, Mail Stop K8-98, Richland, WA 99352. Telephone: (509) 376-0723. Fax (509) 376-7722. E-mail: rd_smith@pnl.gov.

(1) Wilkins, M. R.; Sanchez, J.-C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. *Biotechnol. Gene Eng. Rev.* **1995**, *13*, 19−50.

(2) Wilkins, M. R., Williams, K. L., Appel, R. D., Hochstrasser, D. F., Eds. *Proteome Research: New Frontiers in Functional Genomics*; Springer-Verlag: Berlin-Heildelberg, 1997.

(3) Velculescu, V. E.; Zhang, L.; Vogelstein, B.; Kinzler, K. W. *Science* **1995**, *270*, 484−487.

(4) Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995**, *270*, 467−70.

(5) Haynes, P. A.; Gygi, S. P.; Figeys, D.; Aebersold, R. *Electrophoresis* **1998**, *19*, 1862−1871. Anderson, N. L.; Anderson, N. G. *Electrophoresis* **1998**, *19*, 1853−1861. Humphery-Smith, I.; Cordwell, S. J.; Blackstock, W. P. *Electrophoresis* **1997**, *18*, 1217−1242.

(6) Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Anal. Chem.* **1991**, *63*, 1193−1202.

(7) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64−71.

(8) Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14440−14445. Wilm, M.; Shevchenko, A.; Houthaeve, L.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* **1996**, *379*, 466−469. Yates, J. R. *J. Mass Spectrom.* **1998**, 33, 1−19.

(9) Ogorzalek Loo, R. R.; Stevenson, T. I.; Mitchell, C.; Loo, J. A.; Andrews, P. C. *Anal. Chem.* **1996**, *68*, 1910−1917. Ogorzalek Loo, R. R.; Loo, J. A.; Andrews, P. C. *Methods Mol. Biol.* **1999**, *112*, 473−485. Chong, B. E.; Wall, D. B.; Lubman, D. M.; Flynn, S. J. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1900−1908.

(10) Tang, Q.; Harrata, A. K.; Lee, C. S. *Anal. Chem.* **1997**, *69*, 3177−3182.

(11) Yang, L.; Lee, C. S.; Hofstadler, S. A.; Paša-Tolić, L.; Smith, R. D. *Anal. Chem.* **1998**, *70*, 3235−3241.

(12) Jensen, P. K.; Paša-Tolić, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 2076−2084.

(13) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. *Mass Spectrom. Rev.* **1998**, *17*, 1−35.

(14) Williams, E. R. *Anal. Chem.* **1998**, *70*, 179A−185A.

(15) Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W. *Science* **1996**, *273*, 1199−1202.

(16) Ferianc, P.; Farewell, A.; Nyström, T. *Microbiology* **1998**, *144*, 1045−1050.

(17) *E. coli* K-12 strain MG1655 cells were cultured in either normal or isotopically depleted (∼99.95% ^{12}C, ∼99.99% ^{14}N, and >99.995% ^{1}H) Bioexpress media (Cambridge Isotope Laboratories, Andover, MA). The cultures were inoculated with a starter culture grown on depleted media at 37 °C while shaking at 225 rpm until mid-log phase. At OD_{600} nm ∼0.5 the isotopically depleted culture was "stressed" by the addition of $CdCl_2$ (to a final concentration of 300 μM). Aliquots from stressed and reference cultures were taken at 0, 45, and 150 min after Cd^{2+} addition, centrifuged to remove the broth, and frozen. Aliquots of cells were then mixed in various combinations, effectively providing internal calibrants for each protein and eliminating all sample processing and analysis variables. Cells were lysed and dialysed just prior to CIEF-FTICR analysis and kept on ice to minimize proteolysis.[12]

10.1021/ja991063o CCC: $18.00 © 1999 American Chemical Society
Published on Web 08/13/1999

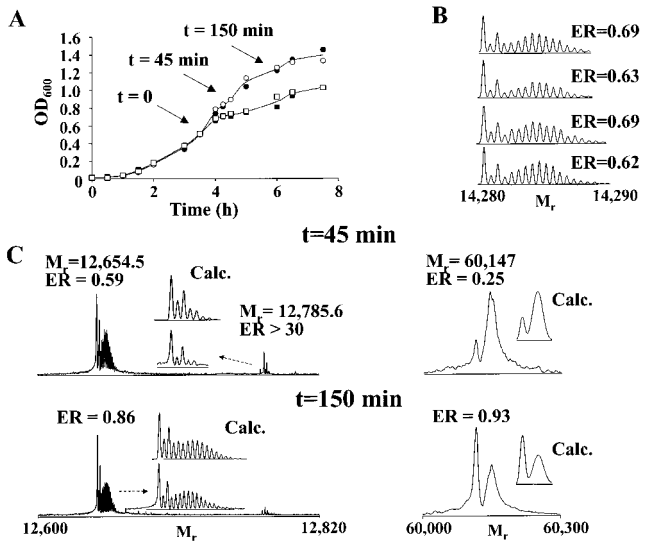

Figure 1. (A) Growth curves for *E. coli* grown on either normal [■ (+Cd^{2+}) and ● (w/o Cd^{2+})] or rare-isotope depleted media [□ (+Cd^{2+}) and ○ (w/o Cd^{2+})]. Arrows indicate $CdCl_2$ addition ($t = 0$) and times at which the aliquots were removed for analysis. (B) Reproducibility of protein expression ratios (ERs) was high due to the elimination of variables associated with sample handling and analysis, as indicated by results for replicate CIEF-FTICR analyses. (C) Examples of variations in protein expression observed for CIEF-FTICR analyses 45 and 150 min following Cd^{2+}-stress of *E. coli* grown on depleted media. The ER for each protein is obtained by fitting the experimental results with the predicted shapes of the calculated isotopic envelope to determine the relative contribution of the two known and isotopically distinctive protein versions. As examples, the ER for a protein of $M_r = 12{,}654.5$ was 0.59 at $t = 45$ min, and changed only slightly at 150 min after Cd^{2+}-stress (to $ER = 0.86$), suggesting an insignificant role in response to Cd^{2+}-stress. However, another protein of $M_r = 12{,}785.6$ was not detectable before Cd^{2+}-stress (i.e., the depleted version of the protein did not have a normal complement, $ER > 30$), and then displays significantly decreased abundance at $t = 150$ min. In another case a protein of $M_r = 60{,}147$ was significantly suppressed 45 min after Cd^{2+}-stress ($ER = 0.25$), but bounces back by $t = 150$ min ($ER = 0.93$). (M_r is given as the average mass for natural isotopic abundance.)

FTICR measurements of proteins expressed from depleted media.[18] The growth curves for *E. coli* grown in either normal or depleted media were similar, indicating no significant effects due to the small differences in isotopic composition (Figure 1 A). Immediately after Cd^{2+}-stress there is a period of growth arrest, followed by resumption of growth, presumably due to a production of Cd^{2+} adaptive proteins. Aliquots were removed from the untreated (normal media) and treated (depleted media) cultures at different time intervals after Cd^{2+}-stress, and cells mixed in various combinations prior to lysis, desalting, and analysis by CIEF-FTICR.[19]

The mass spectrometry analysis provides high precision measurements of the effects induced by Cd^{2+}-stress based on the protein expression ratio, ER (i.e., suppressed, ER < 1; induced, ER > 1; or showing more complex behaviors). The ER is defined as the ratio between the heights of the most abundant isotopes (or peak centroids for low resolution data) in rare-isotope depleted and natural abundance isotopic distributions normalized to the calculated value. To increase the precision and speed of the comparative analysis of the large number of spectra we developed computer software for automated identification of peaks due to proteins and their isotopically distinct complements (present at predictable mass differences). ERs were determined using least-squares fitting to isotopic envelopes or peak shapes calculated based on the known isotopic compositions of the growth media. The generated comparative display, corresponding to detected M_r vs scan number during the CIEF separation (which correlates with protein isoelectric point), uses spot color to represent relative expression of the protein (i.e., ER). The ERs for the 200 most abundant proteins detected for a Cd^{2+}-stress response experiment (at $t = 45$ min) ranged from <0.1 to >30. Since our approach circumvents the typical run-to-run variations for measurements of specific proteins, replicate analyses for different samples show that variations in ERs of <10% can be obtained (Figure 1B). Figure 1C shows portions of mass spectra from one analysis, illustrating types of variation in ERs observed at 45 and 150 min after Cd^{2+}-stress.[20]

In our initial proteome-wide comparative displays of protein expression, we have used CIEF-FTICR instrumentation, which facilitates the quantitative analysis of isotopic variants due to its speed, sensitivity, mass measurement accuracy, and resolution. It is anticipated however, that both alternative separation methods (e.g., capillary liquid chromatography) and different mass spectrometer technologies (e.g., time-of-flight) will be useful for the generation of comparative displays, within constraints dictated by achievable sensitivity, resolution, and mass accuracy. While the present results have involved the analysis of intact proteins, stable-isotope labeling is also well suited for use with approaches involving protein digestion prior to analysis, an approach that largely avoids limitations due to protein size and solubility, and allows protein identification based upon MS/MS of only a single polypeptide. In this regard, Yates and co-workers have demonstrated the rapid identification of complex protein mixtures using capillary liquid chromatography combined with ESI-MS analysis of the tryptic polypeptide mixtures.[21] Since both isotopic versions of the proteins would be digested simultaneously, the ability to make precise measurements of protein expression would be preserved. It is also important to note that highly isotopically enriched growth media are not required; it is only necessary to have isotopic envelopes or peak shapes with sufficient mass differences to be deconvoluted from those of the normal media. Isotopic labeling approaches have been developed for use with mammalian cells,[22] suggesting this approach could be extended to proteome-wide perturbation studies of complex cellular networks of higher organisms.

Acknowledgment. This research was supported by the U.S. Department of Energy Office of Biological and Environmental Research and internal Pacific Northwest National Laboratory (PNNL) Laboratory Directed Research and Development. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC06-76RLO 1830.

JA991063O

(18) Marshall, A. G.; Senko, M. W.; Li, W.; Li, M.; Dillon, S.; Guan, S.; Logan, T. M. *J. Am. Chem. Soc.* **1997**, *119*, 433−434.

(19) CIEF used a 50-cm long (50 μm i.d. and 192 μm o.d.) linear polyacrylamide coated capillary, filled with a sample solution (total protein concentration of ~0.3 μg/μL) containing 0.5% polyampholytes (Pharmalyte 3−10). Focusing was at 13 kV for ~10 min, followed by concurrent gravity and cathodic mobilization to the ESI source. High-resolution mass spectra were acquired with a 7 T FTICR mass spectrometer every ~3 s.[12]

(20) At present we have not attempted the broad identification of proteins. Although a small fraction of the proteins can be identified solely on the basis of accurate M_r measurements, available results indicate that most proteins are modified in some fashion and require additional steps for the initial protein identification of MS detected "spots", as discussed elsewhere.[12]

(21) McCormack, A. L.; Schieltz, D. M.; Goode, B.; Yang, S.; Barnes, G.; Drubin, D.; Yates, J. R. *Anal. Chem.* **1997**, *69*, 767−776. Ducret, A.; Vanoostveen, I.; Eng, J. K.; Yates, J. R.; Aebersold, R. *Protein Sci.* **1998**, *7*, 706−719.

(22) Hansen, A. P.; Petros, A. M.; Mazar, A. P.; Pederson, T. M.; Rueter, A.; Fesik, S. W. *Biochemistry* **1992**, *31*, 12713−12718.