Large Scale Protein Profiling by Combination of Protein Fractionation and Multidimensional Protein Identification Technology (MudPIT)*

Emily I. Chen‡§, Johannes Hewel¶¶, Brunhilde Felding-Habermann||, and John R. Yates, III‡

In the past decade, shotgun proteomic analysis has been utilized extensively to answer complex biological questions. New challenges arise in large scale proteomic profiling when dealing with complex biological mixtures such as the mammalian cell lysate. In this study, we explored the approach of protein separation prior to the shotgun multidimensional protein identification technology (MudPIT) analysis. We fractionated the mammalian cancer cell lysate using the PF 2D ProteomeLab system and analyzed the distribution of molecular weight, isoelectric point, and cellular localization of the eluted proteins. As a result, we were able to reduce sample complexity by protein fractionation and increase the possibility of detecting proteins with lower abundance in the complex protein mixture. Molecular & Cellular Proteomics 5:53–56, 2006.

Shotgun proteomics refers to the direct analysis of complex protein mixtures including biofluids, tissues, cells, organelles, or protein complexes. This approach has been facilitated by the use of multidimensional protein identification technology (MudPIT),1 which incorporates multidimensional high pressure LC/LC, MS/MS, and database-searching algorithms. As a result, shotgun proteomics continues to evolve and enable new areas of biological research. One challenge to the shotgun proteomic paradigm is the complexity of protein mixtures such as the mammalian cell lysate.

In this study, we combined chromatofocusing chromatography using the ProteomeLab PF 2D fractionation system and a shotgun proteomic method, MudPIT, to carry out large scale protein expression analysis of a metastatic breast cancer cell line, BCM2. Over 1,000 proteins were identified from 11 collected fractions. These proteins were further analyzed by the elution profile in relation to the expected pl, molecular weight, and cellular localization.

EXPERIMENTAL PROCEDURES

Breast Cancer Cell Lines

BCM2 metastatic breast cancer cell line was a gift from Dr. Felding-Habermann.

ProteomeLab PF 2D and MudPIT Analysis

First Dimension Chromatofocusing Fractionation and Sample Digestion—Material used to carry out first dimension protein separation was purchased from Beckman Coulter. Briefly cells used for this study were detached by trypsin/EDTA, and total protein extraction was carried out using the starting buffer provided by the ProteomeLab PF 2D kit. 1 mg of the cell lysate were resolved using the default method, and proteins were collected by interval of 0.3 pH unit from the first dimension separation using the ProteomeLab PF 2D fractionation system (Beckman Coulter). 11 fractions were generated and precipitated by TCA/acetone prior to in-solution trypsin digest. The protein pellet from each fraction was resuspended in trypsin digest buffer (50 mM ammonium bicarbonate + 0.1% Rapigest (Waters Corp., Milford, MA)) and digested by trypsin at 37 °C overnight.

Multidimensional Chromatography and Tandem Mass Spectrometry—Peptide mixtures were resolved by strong cation exchange liquid chromatography upstream of reversed phase liquid chromatography. The eluting peptides were electrospayed directly into an LTQ ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoFinnigan, San Jose, CA). Full MS spectra were recorded over a 400–1,600 m/z range followed by three MS/MS events sequentially generated in a data-dependent manner on the first, second, and third most intense ions selected from the full MS spectrum (at 35% collision energy). Mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system (ThermoFinnigan).

Interpretation of MS/MS Datasets—SEQUEST (1) was used to match MS/MS spectra to peptides in a database containing human sequences downloaded from the National Center for Biotechnology Information (NCBI) in July 2004. The validity of peptide/spectrum matches was hence assessed using the SEQUEST-defined parameters, cross-correlation score (XCorr), and normalized difference in cross-correlation scores (∆Cn). Spectrum/peptide matches were only retained if they had a ∆Cn of at least 0.08 and minimum XCorr of 1.8 for +1, 2.5 for +2, and 3.5 for +3 spectra. In addition, the minimum sequence length was seven amino acid residues. DTASelect (2) was used to select and sort peptide/spectrum matches passing this set of criteria. Peptide hits from multiple runs were compared using CONTRAST (2). Proteins were considered detected if they were identified by at least half-tryptic status and more than two peptides.

From the Departments of ‡Cell Biology and ||Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037

Received, June 13, 2005, and in revised form, November 3, 2005
Published, MCP Papers in Press, November 4, 2005, DOI 10.1074/mcp.T500013-MCP200

‡The abbreviation used is: MudPIT, multidimensional protein identification technology.

© 2006 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org
RESULTS AND DISCUSSION

Proteomic Profiling of Mammalian Cell Lysates

Protein mixtures generated from mammalian cell lysates have high complexity and create a challenge for large scale protein profiling. Here we show the advantage of prefractionating the complex protein mixture prior to the high resolution protein identification by mass spectrometry. 1 mg of cell lysate from BCM2, a human breast cancer cell line, were separated by the isoelectric point and collected in different fractions. Proteins collected from each fraction were precipitated and digested in solution by trypsin to generate complex peptide mixtures. Peptides produced by trypsin digestion were then separated and introduced into a mass spectrometer in which m/z ratios are measured and fragmentation spectra are created through tandem mass spectrometry. Peptide sequences and subsequently proteins are identified by unaided searching of uninterpreted tandem mass spectra through human NCBI protein database to identify proteins (3–6). Using the combination of chromatofocusing protein separation and MudPIT, we identified 1,160 proteins from 1 mg of cell lysate.

Protein Separation by ProteomeLab PF 2D Fractionation System

Comparison of pI and pH Elution Gradient—Protein fractionation by chromatofocusing chromatography is used to enrich proteins with similar isoelectric point and collect them in one fraction. By separating tumor cell lysates using the first dimension platform of the ProteomeLab PF 2D fractionation system, we observed that more than 50% of proteins identified from the cells were enriched in specific fractions (Fig. 1). Only a small percentage of proteins were eluted in many fractions. Obviously proteins in high abundance such as histones can render less optimal protein separation. In fact,

![Fig. 1. Protein elution profiles of BCM2 cell lysate after first dimension of ProteomeLab PF 2D fractionation system. The elution profile of the identified proteins was determined by matching the number of protein identification with the number of fractions from which they were identified. Over 50% of proteins were identified in specific fractions. Less than 10% of total proteins appeared in more than eight fractions.](image1)

![Fig. 2. Effect of histone and other abundant proteins on the protein separation profile. The average pI of proteins identified in each fraction was calculated and plotted against the expected elution gradient. The correlation between the average pI in each fraction and pH gradient was compared in three circumstances: total proteins (ALL), subtracting histones (No Histones), and plotting only proteins eluted in a specific fraction (Single Elution). Improvement of the correlation between average pI and expected pH gradient was found in the latter two circumstances.](image2)
when we analyzed the average pI of all proteins identified in each fraction against the expected pH elution gradient, we observed less correlation between the average pI and pH gradient than expected. However, correlation between pI and pH elution gradient was improved when histones were subtracted from each fraction, particularly in the lower pH gradient range. Further perfection of protein separation profile was noted when we graphed only proteins eluted in a single fraction against the pH gradient (Fig. 2). Although overall improvement of the protein separation profile can be achieved by removing the identification of abundant proteins from each fraction, it seems that the average pI of proteins eluted from the lower pH gradient consistently showed a poor correlation to the expected pI range even though proteins with the expected pI were also found in these fractions. It is possible that the last few fractions are enriched with proteins that carry the post-translational modifications, such as phosphorylation, and the modification causes a shift in their pI. In fact, this phenomenon is quite often observed in two-dimensional electrophoresis and is useful to identify proteins with different modification status. Therefore, further investigation of proteins identified in these acidic fractions will broaden the scope of our study.

Distribution of pI and Molecular Weight of Proteins Identified in the Cell Lysate—Using PF 2D ProteomeLab protein fractionation, we were able to identify proteins with a wide range of pI and molecular weight (Fig. 3, A and B). The pI and molecular weight profiles therefore show the benefit of protein fractionation. It is an important aspect of protein profiling to identify proteins with different biochemical properties to ensure a global coverage of the proteome.

Cellular Localization of Proteins Identified in the Cell Lysate—To further ensure a global coverage of the proteome, we investigated the cellular localization of the proteins identified in the cell lysate. Cellular localization of the identified proteins was obtained based on the classification of the GeneOntology database. Over 50% of identified proteins had unknown cellular localization. Cellular distribution of proteins with known cellular location was classified and is shown in Fig. 4. 41% of the identified proteins are localized to the cytoplasm. 28% of the identified proteins are localized to the membrane, and 27% of the identified proteins are localized in
nucleus. A small amount of extracellular proteins (4%) was also detected. Because membrane proteins are notoriously difficult to resolve, it is encouraging that we can detect a significant percentage of membrane proteins after protein fractionation.

In summary, we carried out the characterization of complex mammalian cell lysate by combining protein fractionation and the shotgun proteomic strategies. We demonstrated the potential of large scale proteome profiling by coupling protein fractionation to shotgun proteomic detection, MudPIT. Our results showed that proteins with a wide range of pI, molecular weight, and cellular localization could be resolved using this method. Furthermore the possibility of using a large amount of starting material increases the chance of detecting proteins with lower abundance. Finally we believe that protein fractionation can be an important addition to the methods of global proteome profiling.

* This work was supported in part by National Institutes of Health Grant P41RR11823 (to J. R. Y.) and NCI, National Institutes of Health Grants CA095458 and CA112287 and California Breast Cancer Research Program Grants 10YB-020 and 11IB-0077 (to B. F.-H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Supported by National Institutes of Health Training Grant T32 HL 07695 and later by the NIAID, National Institutes of Health Subcontract Grant UCSD/MCB0237059. To whom correspondence should be addressed: Dept. of Cell Biology, The Scripps Research Inst., 10550 N. Torrey Pines Rd., La Jolla, CA 92037. Tel.: 858-784-8876; Fax: 858-784-8883; E-mail: emilyc@scripps.edu.

¶ Supported by NIAID, National Institutes of Health Grant 5U19AI063603-02002.

REFERENCES

1. Eng, J. K., McCormack, A. L., and Yates, J. R., III (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989

2. Tabb, D. L., McDonald, W. H., and Yates, J. R., III (2002) DTASelect and Contrast: tools for assembling and comparing protein identification from shotgun proteomics. J. Proteome Res. 1, 21–26

3. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., III (1999) Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676–682

4. Larmann, J. P., Jr., Lemmo, A. V., Moore, A. W., Jr., and Jorgenson, J. W. (1993) Two-dimensional separations of peptides and proteins by comprehensive liquid chromatography-capillary electrophoresis. Electrophoresis 14, 439–447

5. Opitcek, G. J., and Jorgenson, J. W. (1997) Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. Anal Chem. 69, 2283–2291

6. Washburn, M. P., Wolters, D., and Yates, J. R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247