

Proteomic analysis

Michael J Dutt and Kelvin H Lee

The field of proteomics is becoming increasingly important as genome sequences are being completed and annotated. Recent advances in proteomics include experimental and mathematical proofs of the need to complement microarray analysis with protein analysis, improved sensitivity for mass spectrometric analysis of separated proteins, better informatic tools for gel analysis and protein spot annotation, first steps towards automated experimental procedures, and new technology for quantitation of protein changes.

Addresses

School of Chemical Engineering, Cornell University, Ithaca, NY 14853-5201, USA

Correspondence: Kelvin H Lee; e-mail: khlee@cheme.cornell.edu

Current Opinion in Biotechnology 2000, 11:176–179

0958-1669/00/\$ – see front matter

© 2000 Elsevier Science Ltd. All rights reserved.

Abbreviations

2DE	two-dimensional protein electrophoresis
ESI	electrospray ionization
IPG	immobilized pH gradient
MALDI	matrix-assisted laser desorption ionization
MS	mass spectrometry
PTM	post-translational modification
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

Introduction

Nucleic acid-based analysis of biological systems (e.g. DNA sequence information or mRNA expression microarrays) can begin to provide data on the nature of individual genes and on the coordinate regulation among many genes. Experimental evidence, however, clearly shows a disparity between the relative expression levels of mRNA and their corresponding proteins [1•,2]. Furthermore, it has recently been proven mathematically that expression information from both mRNA and proteins is required to understand a gene network [3•]. The need and desire to understand total protein expression is motivating the field of 'Proteomics'. The importance of protein-based analysis is its ability to study post-transcriptional control as well as post-translational modifications (PTMs) of proteins. Proteome analysis, which currently, but not exclusively, relies on the microchemical characterization of peptides separated by two-dimensional protein electrophoresis (2DE), can monitor synthesis rates, expression levels and PTMs of proteins. The field of proteome analysis has also largely been driven by technological developments. As with genomic information or data derived from microarray analysis, an informatic framework is required to organize proteomic data. A further informatic challenge is to establish effective connections between protein level and nucleic-acid level information about genes and gene networks.

Because of space limitations in this review, we focus only on technological developments in the areas of proteomics. In particular, we consider proteomics based on the electrophoretic separation of proteins followed by microchemical identification of resolved peptides and discuss recent technological advances. This subset area of proteomics has enjoyed some recent technological developments in the past 12–24 months, some of which are highlighted in this article and placed in the context of earlier work. We also cite key examples of the application of this technology by the biotechnology community.

Technology

Although the term 'proteomics' was first formalized in 1996 [4], the primary experimental tool to monitor genome-wide protein expression, 2DE, has been available since 1975 [5]. 2DE separates proteins first by charge using isoelectric focusing and second by size using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). An example of a 2DE image is given in Figure 1.

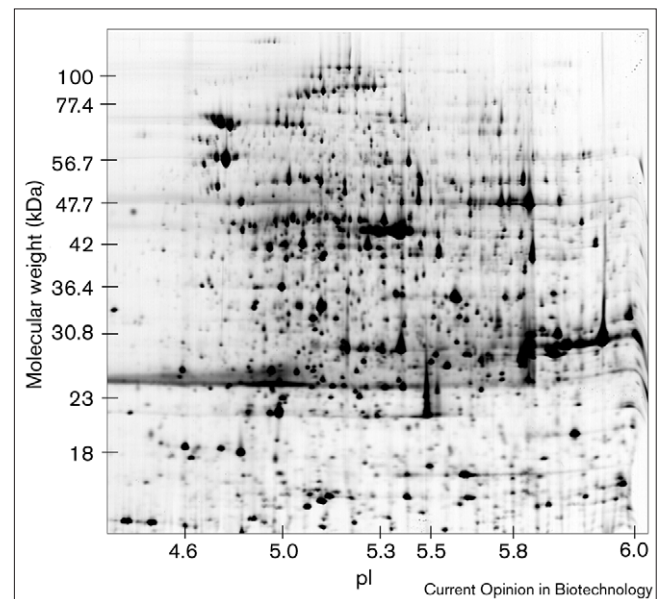
A key feature of 2DE-based analysis of protein mixtures is the ability to perform analyses at both an analytical scale as well as at a micropreparative scale without major modification of protocols, procedures, or equipment. Microgram quantities of protein can be studied in an initial experiment to identify proteins whose expression or PTMs change in an interesting manner. At this scale, the resolution of up to 11,200 proteins from a single mixture has been possible (J Klose, personal communication). Subsequently, the same sample can be applied in milligram quantities to purify individual peptides for further analysis by amino acid analysis, mass spectrometry, amino-terminal or internal amino acid sequencing, and other techniques [6]. The advent of immobilized pH gradient (IPG) strips has been particularly important in this regard [7]. IPG strips increase the reproducibility of isoelectric focusing by using pH gradients covalently grafted to the polyacrylamide-supporting matrix (which nearly eliminates pH drift) and facilitate the resolution of mixtures of milligram quantities of protein. An added benefit of IPG strips is the ability to tune the pH separation to any desired range.

Other important technological advances in 2DE include the development of sensitive protein stains, including the ammoniacal silver stain [8], which permit detection of proteins at or below nanogram quantities, and the use of in-gel sample application to IPG gradient gel strips [9]. In-gel sample loading, in contrast to loading at either the anodic or cathodic ends of the gel, permits the application of greater volumes and quantities of protein as well as reducing the observed focusing problems associated with protein precipitation.

Analytical 2DE has limited applicability because the resulting data (expression patterns of protein separated by charge and size) makes no connection to the actual identity of the protein in terms of the originating gene or genes. Indeed, the inability to more fully characterize the changes of interest is the probable reason why the original technique did not become more ubiquitous. Simple extensions to the technique, such as Western blotting and lectin affinity blotting [10,11], have provided only modest increases in information about proteins — such techniques are typically used to confirm a suspected identity. The demonstration by Aebersold *et al.* [12] that amino-terminal and internal protein sequence information could be obtained from 2DE-separated proteins blotted onto membranes was seminal. Direct amino acid sequencing provided an important connection between proteomic and genomic information — yielding a genetic basis for as yet uncharacterized proteins; however, equipment and reagent costs, as well as the limited sensitivity associated with direct sequencing, have been restrictive. More recently, peptides (digested proteins obtained from 2DE gels or blots) are characterized by mass spectrometry (MS) [13]. The two common ionization techniques for proteins are electrospray (or nanospray for smaller quantities) ionization (ESI) and matrix-assisted laser desorption ionization (MALDI). These techniques are coupled to different mass analyzers, all of which possess certain advantages and disadvantages. In general, MALDI-time-of-flight MS is more effective for the analysis of higher molecular weight proteins, whereas ESI-ion trap MS offers better sensitivity of detection, down to the femtomole level. For a more detailed review of ESI and MALDI techniques, the reader is referred to a review of the technologies [14]. A key feature of MS analysis of gel-separated proteins and peptides is the ability to generate different types of structural information about a particular peptide of interest. For example, the mass spectrometer can directly provide information on the mass of a particular peptide and can also be used to generate *de novo* amino acid sequence information from tandem mass spectra obtained either by postsource decay or collision-induced dissociation (as described in [15,16]). Further, the mass spectrometer can provide data on glycosylation patterns, phosphorylation and other PTMs by measuring mass shifts of peptide fragments. This flexibility, as well as the ability to study other non-protein small molecules, has motivated the current interest in mass spectrometry. Next we consider some of the recent developments in MS technology that apply specifically to proteomic analysis.

Recently, tandem MS has been used to identify proteins in macromolecular complexes [17]. This process uses two-dimensional chromatographic separations, peptide fragmentation, and then compares amino acid sequences to genomic sequences. This is a rapid method for protein identification that relies on the accuracy and predictive power of whole-genome sequences. Moreover, another form of tandem mass spectrometry seeks to replace the need for SDS-PAGE [18]. This approach offers sub-picomole sensitivity (0.3–3 pmol), mass measurement accuracy

Figure 1



Escherichia coli JM105 constitutively expressing green fluorescent protein was separated by two-dimensional protein electrophoresis and stained with ammoniacal silver. The pH gradient (horizontal direction) runs from pH 4–7 and the SDS-PAGE separation (vertical direction) was performed in a 9–17% T gradient gel. More than 2000 protein species are visible by this technique.

to $\pm 0.05\%$, and reduced sample-handling losses; however, it is limited in its ability to quantitate proteins.

Informatics

Proteomic analysis requires informatic tools at many levels. First, there is a desire to develop online databases of protein expression profiles. Such databases should be seamlessly linked to and integrated with databases of nucleic acid sequence and expressed sequence tags. An excellent resource for Internet-accessible proteome databases is the Expert Protein Analysis System (ExPASy), now available online at <http://www.expasy.ch/> [19]. Furthermore, there is the desire to develop software packages that can take multiple protein expression profiles and automatically identify quantitative changes of interest. Some pattern recognition software is commercially available from numerous vendors and one example is Melanie 3 (Swiss Institute for Bioinformatics) [20]. Key features of Melanie 3 include the ability to perform multivariate statistics on datasets, make comparisons with online databases and interface with MS spectrum.

There is currently a critical need to develop additional software tools that can combine biochemical data about an unidentified peptide from different experiments (e.g. isoelectric point, molecular weight and tandem mass spectra) and search protein sequence, nucleic acid sequence and expressed sequence tag databases to identify the most probable genetic basis.

A suite of proteomic tools [21,22] is now available at the ExPASy website [19••]. The Swiss Institute for Bioinformatics maintains this database and has released an excellent summary of the protein sequence data bank [23].

Applications

Differential proteome analysis compares the expression profile of 2DE-separated proteins from an arbitrary reference state of a cell, tissue, or organism, to the profile of a non-standard condition, such as a diseased state or after the addition of a toxin to the system. The differences in the two proteomes give an indication of the response mechanisms of that system to perturbations. There is an increasing variety of applications of differential proteome analysis to interesting problems. For example, exposure of rat kidneys to lead has been shown to alter the quantity of 76 proteins in the cortex and 13 in the medulla [24]. This work highlights the application of proteomics in identifying markers for toxicological studies. The phosphorylation state of membrane receptors after stimulation with platelet-derived growth factor [25] and the influence of temperature on Chinese hamster ovary (CHO) cell productivity [26••] are two examples that make use of the ability of proteome analysis to highlight PTMs to proteins in response to environmental perturbations. The low temperature CHO cell experiment is particularly noteworthy because it provides the first evidence that the cold response of mammalian cells includes changes in PTMs, specifically the phosphorylation of tyrosine residues of two proteins. Differential proteome analysis can also help identify post-transcriptional/pre-translational levels of control. By measuring the level of translationally controlled tumor protein and the abundance of its mRNA, it was shown that calcium levels in Cos-7 cells regulate protein expression at the level of transcription and at a post-transcriptional step [27]. Proteomic analysis has yielded clues to *Salmonella typhimurium* pathogenicity based on the analysis of a key regulon and its proteins [28]. The differential study of *Sinorhizobium meliloti* during early and late exponential growth phase has identified a set of novel proteins associated with growth control [29].

Future prospects

One goal in proteomics is the rapid and quantitative characterization of proteins. New developments in the use of mass spectrometry with isotope labeling hold much promise. Isotopic metabolic labeling with ¹⁵N can help identify and quantify proteins, including their modifications, from a purified subpopulation of proteins [30]. The technique is applicable to a variety of mass analyzers and ionization techniques. This approach has difficulty, however, measuring low-abundance proteins because of the sample loading limitations of analytical 2DE. Another approach that does not have the same limitations has recently been developed [31••]. This new approach is not limited to cells compatible with metabolic labeling, making it broadly applicable to almost any cell or tissue type. The sidechains of cysteinyl residues in a reduced protein sample are derivatized with either an isotopically light (cell-state one) or heavy (cell-state

two) form of a chemical reagent. The cells are combined, enzymatically cleaved, separated by affinity chromatography, and finally separated by tandem mass spectrometry. Operation of an LCQ™ ion trap mass spectrometer (Finnigan MAT, San Jose, CA) in dual mode allows both the quantity and sequence identity of the peptides from the different cell states to be determined because of their differential tags. This method is analogous to the combination of mRNA species from two differing cell states analyzed on DNA microarrays. This new protein approach should provide a widely applicable method for quantitatively comparing protein mixtures from cells and tissues. Still, there is a major limitation in the study of low abundance proteins because of the lack of an amplification strategy analogous to polymerase chain reaction for nucleic acids.

A particularly exciting development is the Molecular Scanner developed by Denis Hochstrasser and co-workers [32]. This new technology takes 2DE gels and combines protease digestion and electroblotting to a membrane in a single step, followed by the generation of MS fingerprints (by MALDI-MS) of individual regions of the electroblotted 'gel'. This approach theoretically permits the full characterization of all proteins on a gel. For a further discussion on the promise of automation in proteomics, the reader is referred to articles from the Australian Proteome Analysis Facility [33] and The Swiss Federal Institute of Technology [34].

Conclusions

Information about gene expression at the protein level is providing critical data on the genotype–phenotype relationship in a variety of settings. Such data is critical in understanding this nonlinear relationship and is necessary for metabolic and cellular engineering efforts. The past 12–24 months have seen several key advances in the characterization and quantitation of protein mixtures, however, there is much left to do. Given the need to study mRNA and protein expression profiles in parallel and the current focus on new technology for nucleic acid analysis; one might expect that the development of improved technology for proteome analysis will become increasingly important.

Acknowledgements

This work was supported in part by the National Science Foundation (BES-9874938), Intel (98–238), The New York State Science and Technology Foundation and DuPont. We thank Leila Choe for technical support in generating the data.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Gygi S, Rochon Y, Franza B, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19**:1720-1730.

This paper shows that the correlation between mRNA and protein levels is insufficient to predict protein expression levels from quantitative mRNA data.

2. Anderson L, Seilhamer J: **A comparison of selected mRNA and protein abundances in human liver.** *Electrophoresis* 1997, **18**:533-537.

3. Hatzimanikatis V, Lee K: **Dynamical analysis of gene networks**
 - **requires both mRNA and protein expression information.** *Metab Eng* 1999, **1**:275-281.

A nonlinear stability analysis is used to show that a combination of gene expression information at the message level and at the protein level is required to describe even simple models of gene networks.
 4. Wasinger V, Cordwell S, Cerpa-Poljak A, Yan J, Gooley A, Wilkins M, Duncan M, Harris R, Williams K, Humphery-Smith I: **Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*.** *Electrophoresis* 1995, **16**:1090-1094.
 5. O'Farrell P: **High resolution two-dimensional electrophoresis of proteins.** *J Biol Chem* 1975, **250**:4007-4021.
 6. Wilkins M, Gooley A: **Protein identification in proteome projects.** In *Proteome Research: New Frontiers in Functional Genomics*. Edited by Wilkins MR, Williams K, Appel R, Hochstrasser D. Tokyo: Springer; 1997:35-64.
 7. Bjellqvist B, Ek K, Richetti P, Gianazza E, Gorg A, Westermeier R, Postel W: **Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications.** *J Biochem Biophys Methods* 1982, **6**:317-339.
 8. Merrill C, Switzer R, VanKeuren M: **Trace polypeptides in cellular extracts and human body fluids detected by two-dimensional electrophoresis and a highly sensitive silver stain.** *Proc Natl Acad Sci USA* 1979, **76**:4335-4339.
 9. Sanchez J, Rouge V, Pisteur M, Ravier F, Tonella L, Moosmayer M, Wilkins M, Hochstrasser D: **Improved and simplified in-gel sample application using reswelling of dry immobilized pH gradients.** *Electrophoresis* 1997, **18**:324-327.
 10. Towbin H, Staehelin T, Gordon J: **Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications.** *Proc Natl Acad Sci USA* 1979, **76**:4350-4354.
 11. Golaz O, Gravel P, Walzer C, Turler H, Balant L, Hochstrasser D: **Rapid detection of the main human plasma glycoproteins by two-dimensional polyacrylamide gel electrophoresis lectin affino blotting.** *Electrophoresis* 1995, **16**:1187-1189.
 12. Aebersold R, Leavitt J, Saavedra R, Hood L, Kent S: **Internal amino acid sequence analysis of proteins separated by one- or two-dimensional electrophoresis after in situ protease digestion on nitrocellulose.** *Proc Natl Acad Sci USA* 1987, **84**:6970-6974.
 13. Haynes P, Gygi S, Figeys D, Aebersold R: **Proteome analysis: biological assay or data archive?** *Electrophoresis* 1998, **19**:1862-1871.
 14. Patterson S, Aebersold R: **Mass spectrometric approaches for the identification of gel-separated proteins.** *Electrophoresis* 1995, **16**:1791-1814.
 15. Keough T, Youngquist R, Lacey M: **A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry.** *Proc Natl Acad Sci USA* 1999, **96**:7131-7136.
- Here a new approach to postsource decay matrix-assisted laser desorption ionization mass spectrometry is presented that provides easy to interpret ion mass spectra.
16. Swiderek K, Davis M, Lee T: **The identification of peptide modifications derived from gel-separated proteins using electrospray triple quadrupole and ion trap analyses.** *Electrophoresis* 1998, **19**:989-997.
 17. Link A, Eng J, Schieltz D, Carmack E, Mize G, Morris D, Garvik B, Yates J: **Direct analysis of protein complexes using mass spectrometry.** *Nat Biotechnol* 1999, **17**:676-682.
- This paper details the identification of proteins in macromolecular complexes by the use of multidimensional liquid chromatography and tandem mass spectrometry to separate and fragment peptides.
18. Loo J, Brown J, Critchley G, Mitchell C, Andrews P, Loo R: **High sensitivity mass spectrophotometric methods for obtaining intact molecular weights from gel-separated proteins.** *Electrophoresis* 1999, **20**:743-748.
 19. EXPASY Molecular Biology Server on World Wide Web URL:
 - <http://www.expasy.ch/>

This is the website for the proteomics server of the Swiss Institute of Bioinformatics. It is dedicated to the analysis of protein sequences and structures, as well as 2D protein electrophoresis.
 20. Melanie 3 2D Gel Analysis Software on World Wide Web URL:
 - <http://www.expasy.ch/melanie/>
 21. Wilkins M, Gasteiger E, Wheeler C, Lindskog I, Sanchez J, Bairoch A, Appel R, Dunn M, Hochstrasser D: **Multiple parameter cross-species protein identification using Multident – a world-wide web accessible tool.** *Electrophoresis* 1998, **19**:3199-3206.
 22. Wilkins M, Gasteiger E, Tonella L, Ou K, Tyler M, Sanchez J, Gooley A, Walsh B, Bairoch A, Appel R, Williams K, Hochstrasser D: **Protein identification with N and C-terminal sequence tags in proteome projects.** *J Mol Biol* 1998, **278**:599-608.
 23. Bairoch A, Apweiler R: **The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999.** *Nucleic Acids Res* 1999, **27**:49-54.
 24. Witzmann F, Fultz C, Grant R, Wright L, Kornguth S, Siegel F: **Regional protein alterations in rat kidneys induced by lead exposure.** *Electrophoresis* 1999, **20**:943-951.
 25. Zimmermann J, Soskic V, Poznanovic S, Brianza F: **Functional proteomics of signal transduction by membrane receptors.** *Electrophoresis* 1999, **20**:952-961.
 26. Kaufmann H, Mazur X, Fussenegger M, Bailey J: **Influence of low temperature on productivity, proteome and protein phosphorylation of CHO cells.** *Biotechnol Bioeng* 1999, **63**:573-582.
- The influence of low temperature on the Chinese hamster ovary cell proteome is studied. For the first time, evidence for phosphorylation changes as a result of cold response are documented. The results suggest new opportunities for metabolic engineering.
27. Xu A, Bellamy A, Taylor J: **Expression of translationally controlled tumor protein is regulated by calcium at both the transcriptional and post-transcriptional level.** *Biochem J* 1999, **342**:683-689.
 28. Deiwick J, Hensel M: **Regulation of virulence genes by environmental signals in *Salmonella typhimurium*.** *Electrophoresis* 1999, **20**:813-817.
 29. Guerreiro N, Djordjevic M, Rolfe B: **Proteome analysis of the model micro symbiont *Sinorhizobium meliloti*: isolation and characterization of novel proteins.** *Electrophoresis* 1999, **20**:818-825.
 30. Oda Y, Huang K, Cross F, Cowburn D, Chait B: **Accurate quantitation of protein expression and site-specific phosphorylation.** *Proc Natl Acad Sci USA* 1999, **96**:6591-6596.
 31. Gygi S, Rist B, Gerber S, Turecek F, Gelb M, Aebersold R:
 - **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17**:994-999.

A new method of quantitative proteomics is described that uses a new class of chemical reagents termed isotope-coded affinity tags. This new technology promises the rapid quantitation of proteins needed for complete proteomic analysis.
 32. Binz A, Muller M, Walther D, Bienvenu W, Gras R, Hoogland C, Bouchet G, Gasteiger E, Fabbretti R, Gay S: **A molecular scanner to automate proteomic research and to display proteome images.** *Anal Chem* 1999, **71**:4981-4988.
 33. Walsh B, Molloy M, Williams K: **The Australian Proteome Analysis Facility (APAF): assembling large scale proteomics through integration and automation.** *Electrophoresis* 1998, **19**:1883-1890.
 34. Quadroni M, James P: **Proteomics and automation.** *Electrophoresis* 1999, **20**:664-677.