Proteomics: applications in basic and applied biology

N Leigh Anderson, Alastair D Matheson and Sandra Steiner

The rapid evolution of proteomics has continued during the past year, with a series of innovations in the core technologies of two-dimensional electrophoresis and mass spectrometry, and a diversity of productive research programmes. Well-annotated proteomics databases are now emerging in a number of fields to provide a platform for systematic research, with particularly promising progress in clinical applications such as cardiology and oncology. Large-scale quantitative research, comparable in power and sensitivity to that achieved for gene expression, is thus becoming a reality at the protein level.

Address
Large Scale Proteomics Corporation, 9620 Medical Center Drive, Rockville, MD 20850, USA

Current Opinion in Biotechnology 2000, **11**:408–412

0958-1669/00/$ – see front matter

Abbreviations
2DE two-dimensional electrophoresis
MS mass spectrometry

Introduction

If the 1990s were the decade of genomics, the first ten years of the new century are set to become the decade of proteomics. For the first time, the technologies of proteomics make it possible to generate quantitative protein expression data on a scale and sensitivity comparable to that achieved at the genetic level. This advance has major implications for our understanding of cellular organisation in health and disease, and for pharmaceutical and agricultural biotechnology. Indeed, proteomics is already yielding important findings across a wide range of applications. This article reviews new concepts, innovative technologies and biological applications in proteome research.

Concepts: structural versus quantitative regulation proteomics

The ultimate objectives of proteomics go beyond the simple cataloguing of the proteins that cells express in health and disease states. The eventual goal is to elucidate the organisation and dynamics of the metabolic, signalling and regulatory networks through which the life of the cell is transacted. Moreover, proteomics seeks to understand how these networks become dysfunctional in disease, and to predict how their function can be manipulated through interventions such as drugs and genetic manipulations.

These are ambitious goals that will require technologies of increased sensitivity and new concepts before they can be fully realised, but the task of mapping regulatory networks and diagnosing cell states is progressing. In microbes, VanBogelen *et al.* [1] from Parke-Davis/University of Michigan have identified proteomic signatures of specific cellular states, such as secretory dysfunction and the use of specific phosphorus sources, and are currently combining genomic and proteomic approaches in an attempt to map full regulatory networks [2•].

Most proteomics research, however, is directed towards the more proximal goal of investigating protein expression and function under specified physiological conditions. Here, two contrasting but complementary strategies have emerged. On the one hand, 'expression' or 'quantitative regulation' proteomics monitors the expression of large numbers of proteins within a cell or tissue and observes quantitatively how the pattern of expression changes under different circumstances, such as in the presence of a drug or in diseased tissue [3]. This makes it possible to identify disease-specific proteins, drug targets and markers of drug efficacy and toxicity, and also to make deductions about regulatory networks by identifying proteins that undergo co-ordinated changes of expression. This is currently the most widely used and productive model of proteomics, and at present is largely dependent on two-dimensional electrophoresis (2DE).

The second strategy has been referred to as 'cell map' or 'structural' proteomics [4]. Here, the immediate goal is to identify the structure of proteins and in particular to identify proteins that interact with and form complexes with other proteins. By documenting the physical interactions of proteins, this strategy is likely to prove highly effective for the detailed study of particular pathways, although without the breadth of focus of quantitative regulation proteomics. Although the two-hybrid and related assays (reviewed in [5]) are important in identifying specific protein–protein interactions, the most important technology for structural proteomics is mass spectrometry (MS).

Technologies: innovations in 2DE and MS

The distinction between quantitative regulation and structural proteomics is useful but it should be emphasised that most proteomics projects combine elements of both approaches and rely on both 2DE and MS. Both these core technologies have seen significant advances during the past year. There has been gradual progress on the vexed problem of isolating highly hydrophobic membrane proteins using 2DE through the continued development of sample preparation protocols based on zwitterionic detergents [6] and organic solvents [7], but this remains an important and unsolved problem. The challenge of obtaining samples of just a single cell type from a tissue has been successfully addressed using a laser capture microdissection system, in which a tissue sample is attached to a film and specific cells of interest detached using a laser beam [8].

The use of MS in proteomics continues to grow in power and versatility with a number of notable recent innovations.

A number of groups have introduced novel protein-tagging methodologies that improve the power and sensitivity of MS [9,10•]. Aebersold and co-workers [10•] have, for example, developed a method for labelling peptides with isotope-coded affinity tags (ICATs), which do not only support enhanced analysis of complex peptide mixtures but also allow accurate quantification of differences in the level of expression of proteins, a capability that MS-based proteomics did not previously have. The coupling of MS with a software tool, FindMod, has increased the power of this approach to identify post-translational modifications [11•]. Techniques for the characterisation of protein complexes have also been introduced, including a system based on liquid chromatography/tandem MS [12] and the development of a novel tandem affinity purification (TAP) tag for the rapid purification of complexes from cell samples [13]. Other developments in sampling protocols have recently allowed direct analysis by MS of the peptides from single secretory vesicles and it is anticipated that other organelles will be amenable to analysis by the same technique [14].

It is of course through the integration of 2DE- and MS-based approaches that proteomics achieves its greatest power. A representative example of the degree of integration now possible is the system developed by Hochstrasser's group [15•] in which whole 2DE gels are subject to *in situ* digests, electrotransferred onto membranes and directly scanned by MS, leading to the automated generation of an annotated (i.e. the proteins are characterised and identified) 2D map. Developments of this kind, which are also being pioneered within specialist proteomics companies, are now enabling proteomics to be conducted on a scale commensurate with the complexity of its subject matter.

Proteome databases

The proteome database, consisting of annotated 2DE data obtained under different cellular conditions, is the basic platform from which proteome research addresses specific biological and pharmaceutical questions. It is notable, however, that to date relatively few large-scale databases have emerged into the public domain. The Yeast Protein Database (YPD; http://www.proteome.com/databases/index.html) is one of the best and longest-established large-scale databases and its curators, Proteome Inc. (Beverly, MA), are now constructing WormPD [16], a proteome database for the nematode worm *Caenorhabditis elegans*, for which the complete genome was recently sequenced. One of the most important database developments during the past year was the construction by Hoffmann-La Roche of a *Haemophilus influenzae* database, consisting of over 1000 individual 2DE map protein spots from which more than 500 individual proteins have been identified to date [17••]. Klose and co-workers [18] have also recently reported their ongoing programme to develop a database for the mouse proteome, with an initial focus on brain proteins. In humans, the databases on bladder cancer constructed by Celis *et al.* [19] continue to be among the largest and best-annotated currently available, although the next few years are certain to see a major drive to generate major proteome databases for a wide variety of human tissues. The number of smaller proteome databases designed to address specific biological questions in humans and other organisms is now very large and beyond the scope of this review to address. The Expasy website (http://www.expasy.ch), home to the major SWISS-2D PAGE proteomics database and other resources, contains numerous links and remains the best platform from which to explore the world of public proteomics databases.

Basic biology

The number of proteomics projects addressing basic biological questions is now very large and continues to grow. In part, this reflects the broadening usage of the term 'proteomics' to embrace not only large-scale biotechnological research but also any project in which proteins are systematically studied by 2DE, MS or other techniques.

One of the most significant themes for general biology to emerge out of proteomics has been a deepening insight into the nature of the relationship between genes and proteins. In the mouse proteome project of Klose and co-workers described above [18], hundreds of mouse genes were mapped to chromosomal locations on the basis of protein polymorphisms. It emerged from these studies that many protein modifications are related to specific genes, such that a protein should be considered as the phenotype not of one but of many genes [20••]. Equally, a single genetic mutation may affect many proteins. Ultimately, studies of this kind will lead to a more sophisticated understanding of the relationship between genes and cell function in health and disease.

The combination of immunoaffinity techniques with 2DE and subsequent MS continues to be a major workhorse for studies of specific subsets of a cell's proteome. Numerous studies of basic cellular mechanisms are in progress using this approach. For example, immunoblotting using antiphosphotyrosine and antiphosphoserine antibodies was used to extract over 500 phosphorylated proteins from mouse fibroblasts, of which at least 100 were found to undergo alterations in phosphorylation following stimulation with platelet-derived growth factor (PDGF) [21]. These studies have revealed new putative signalling pathways downstream of the PDGF β receptor and are applicable to other signal transduction pathways. As a second example, immunoprecipitation was used to extract the chaperonin GroEL from *Escherichia coli*, in complex with over 300 newly translated polypeptides, which were subsequently separated by 2DE [22]. The technique enabled the function of GroEL in mediating protein folding to be investigated and is a typical example of how proteomics techniques can be used to target and study a specific subset of cellular components. Finally in this section, it is worth pointing out that proteomics is a burgeoning field in plant biology and

agricultural biotechnology [23]. Plant proteomics studies include characterization of individuals or lines, estimation of genetic variability within and between populations or establishment of genetic distances [23].

Clinical studies

The potential of proteomics, particularly quantitative regulation proteomics, in the clinical arena seems almost boundless, as the protein correlates of any given disease state can now be readily identified, provided sample preparation is adequate and the problem is studied on an appropriate scale. There are now numerous examples of proteomics-based clinical studies with the number of projects increasing monthly. One interesting set of data to emerge recently came from the University of Auckland, New Zealand, where the construction of a proteome map of the human hippocampus [24] has led to the identification of 18 proteins with abnormal expression in schizophrenics, several of which have been found to map to the same region of chromosome 6 [25]. The potential roles of these proteins in the pathogenesis of the disease are now being investigated.

For several years now, the group of Dunn and colleagues at Harefield Hospital, UK, have been developing one of the major clinical applications of proteomics in the study of heart disease, a programme that has recently been reviewed [26•]. Once again, this work is built on the solid foundation of good proteomics databases and recent work includes the identification of disease-specific proteins for dilated cardiomyopathy [27,28] and anti-endothelial cell antibodies that are being investigated as a potential predictive test for chronic heart transplant rejection [29]. But it is in oncology that proteomics has found its most widespread clinical applications to date, principally through the comparison of normal and tumour tissue using 2DE-based techniques. The past year has seen the publication of proteomics studies on a range of tumour types, among which two of the most notable have been studies of bladder and breast cancer. In squamous cell carcinoma of the bladder, Celis *et al.* [30•] used their database described above to identify several disease-specific proteins, which were used to raise antibodies capable of identifying metaplastic lesions. The same team has identified psoriacin as a protein marker of bladder squamous cell carcinoma, which may provide a useful, non-invasive approach to disease monitoring [31•].

During the past few years, several groups have described proteomics studies of breast cancer. The first stage of the most recent study reported the separation of over 1700 protein spots in normal luminal and myoepithelial breast cells, of which over 170 differed significantly in expression between the two cell types [32•]. The separation of the two cell types was an important feature of this study because ~95% of breast tumours are luminal in origin. The next stage of this detailed study will determine the proteome of tumour cells and identify differences with the proteome of healthy luminal cells.

Toxicology

Toxicology is likely to prove one of the most important applications of proteomics. 2DE is a highly sensitive means of screening for toxicity and probing toxic mechanisms. By comparing proteins expressed following treatment with a given drug with those present under untreated conditions, it is possible to identify changes in biochemical pathways via observed alterations in sets of proteins that may be related to the drugs efficacy or toxicity. When a large enough library of proteomic signatures has been compiled for compounds of known toxicity, it will be possible to use it to assess the toxicity of novel compounds. However, few large-scale proteomics studies in toxicology have yet entered the public domain. In one ongoing study, a group at Imperial College London/SmithKline Beecham recently reported an ongoing 2DE- and NMR-based study of glomerular nephrotoxicity in the rat following exposure to puramycin aminonucleoside [33]. By monitoring the proteins in urine, the study has permitted a more detailed understanding of the nature and progression of the proteinuria associated with glomerular nephrotoxicity than has previously been possible. In a second recent example, a study of lead toxicity in a rabbit model identified a number of proteins that change expression following increased lead exposure, of which several molecules, provisionally identified as glutathione-S-transferase variants, may be developed into valuable markers of lead toxicity in humans [34]. Other notable reports to emerge recently in proteomic toxicology include an analysis of cyclosporin A toxicity, which revealed a novel toxic mechanism involving the calcium-binding protein calbindin-D [35], and the ongoing studies of liver protein changes following exposure of rodents to peroxisome proliferators being conducted by Argonne National Laboratories and Large Scale Proteomics Corporation [36,37]. However, although numerous other toxicological studies have used proteomics tools such as 2DE and MS, the potential of proteomics in routine toxicology has yet to be realised.

Conclusions

The past year has seen continued technical advances in proteomics and the beginning of what is certain to be a major and sustained influx of data on the organisation and functions of proteins in living systems. In particular, there has been important progress in the compilation of protein expression databases on cell types ranging from microbes to human cancer, and the identification of a wide range of cell-state-specific proteins in numerous organisms. The technologies of 2DE and MS are the driving force in contemporary proteomics and are likely to remain so in the immediate future, although in the longer term new approaches, including antibody-based techniques and protein chips, may in time challenge their pre-eminence. Over the next few years, however, it will be through technologies that exist and are widely used today that proteomics will help to overcome many of the most challenging problems in contemporary biology and biotechnology.

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. VanBogelen RA, Schiller EE, Thomas JD, Neidhardt FC: **Diagnosis of cellular states of microbial organisms using proteomics.** *Electrophoresis* 1999, **20**:2149-2159.

2. • VanBogelen RA, Greis KD, Blumenthal RM, Tani TH, Matthews RG: **Mapping regulatory networks in microbial cells.** *Trends Microbiol* 1999, **7**:320-328.

A useful review of the combined use of proteomics and genomics to map global regulatory networks.

3. Anderson NL, Anderson NG: **Proteome and proteomics: new technologies, new concepts, and new words.** *Electrophoresis* 1998, **19**:1853-1861.

4. Blackstock WP, Weir MP: **Proteomics: quantitative and physical mapping of cellular proteins.** *Trends Biotechnol* 1999, **17**:121-127.

5. Vidal M, Legrain P: **Yeast forward and reverse 'n'-hybrid systems.** *Nucleic Acids Res* 1999, **27**:919-929.

6. Santoni V, Rabilloud T, Doumas P, Rouquié D, Mansion M, Kieffer S, Garin J, Rossignol M: **Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels.** *Electrophoresis* 1999, **20**:705-711.

7. Molloy MP, Herbert BR, Williams KL, Gooley AA: **Extraction of *Escherichia coli* proteins with organic solvents prior to two-dimensional electrophoresis.** *Electrophoresis* 1999, **20**:701-704.

8. Banks RE, Dunn MJ, Forbes MA: **The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis — preliminary findings.** *Electrophoresis* 1999, **20**:689-700.

9. Keough T, Youngquist RS, Lacey MP: **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.

10. • Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17**:994-999.

The authors describe a new technique that enables the relative quantities of proteins in complex mixtures to be assessed during MS-based protein characterisation.

11. • Wilkins MR, Gasteiger E, Gooley AA, Herbert BR, Molloy MP, Binz PA, Ou K, Sanchez JC, Bairoch A, Williams KL, Hochstrasser DF: **High-throughput mass spectrometric discovery of protein post-translational modifications.** *J Mol Biol* 1999, **289**:645-657.

This paper describes the use of the FindMod database coupled to MS to rapidly identify post-translational modifications.

12. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III: **Direct analysis of protein complexes using mass spectrometry.** *Nat Biotechnol* 1999, **17**:676-682.

13. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B: **A generic protein purification method for protein complex characterization and proteome exploration.** *Nat Biotechnol* 1999, **17**:1030-1032.

14. Rubakhin SS, Garden RW, Fuller RR, Sweedler JV: **Measuring the peptides in individual organelles with mass spectrometry.** *Nat Biotechnol* 2000, **18**:172-175.

15. • Binz P-A, Muller M, Walther D, Bienvenut WV, Gras R, Hoogland C, Bouchet G, Gasteiger E, Fabbretti R, Gay S *et al.*: **A molecular scanner to automate proteomic research and to display proteome images.** *Anal Chem* 1999, **71**:4981-4988.

A striking example of the levels of automation and high-throughput analysis that can now be achieved through the integration of technologies.

16. Costanzo MC, Hogan JD, Cusick ME, Davis BP, Fancher AM, Hodges PE, Kondu P, Lengieza C, Lew-Smith JE, Linger C *et al.*: **The Yeast Protein Database (YPD) and *Caenorhabditis elegans* Proteome Database (WormPD): comprehensive resources for the organization and comparison of model organism protein information.** *Nucleic Acids Res* 2000, **28**:73-76.

17. •• Langen H, Takacs B, Evers S: **Two-dimensional map of the proteome of *Haemophilus influenzae.*** *Electrophoresis* 2000, **21**:411-429.

An important addition to the array of proteomics databases currently available, with over 500 proteins identified from a total of over 1000 resolved.

18. Gauss C, Kalkum M, Lowe M, Lehrach H, Klose J: **Analysis of the mouse proteome. (I) Brain proteins: separation by two-dimensional electrophoresis and identification by mass spectrometry and genetic variation.** *Electrophoresis* 1999, **20**:575-600.

19. Human 2-D PAGE Databases for Proteome Analysis in Health and Disease on World Wide Web URL: http://biobase.dk/cgi-bin/celis

20. •• Klose J: **Genotypes and phenotypes.** *Electrophoresis* 1999, **20**:643-652.

The implications of the mouse proteome project for our understanding of the relationship between genes and proteins; proteins demonstrated to be polygenic entities with complex phenotypes.

21. Soskic V, Gorlach M, Poznanovich S, Boehmer FD, Godovac-Zimmermann J: **Functional proteomics analysis of signal transduction pathways of the platelet derived growth factor β receptor.** *Biochemistry* 1999, **38**:1757-1764.

22. Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU: **Identification of *in vivo* substrates of the chaperonin GroEL.** *Nature* 1999, **402**:147-154.

23. Thiellement H, Bahrman N, Damerval C: **Proteomics for genetic and physiological studies in plants.** *Electrophoresis* 1999, **20**:2013-2026.

24. Edgar PF, Douglas JE, Knight C, Cooper GJ, Faull RL, Kydd R: **Proteome map of the human hippocampus.** *Hippocampus* 1999, **9**:644-650.

25. Edgar PF, Douglas JE, Cooper GJ, Dean B, Kydd R, Faull RL: **Comparative analysis of the hippocampus implicates chromosome 6q in schizophrenia.** *Mol Psychiatry* 2000, **5**:85-90.

26. • Dunn MJ: **Studying heart disease using the proteomic approach.** *Drug Dev Therapeutics* 2000, **5**:76-84.

A useful introductory review of the Harefield group's wide-ranging studies in cardiology.

27. Warriach RS, Dunn MJ, Yacoub MH: **Subclass specificity of autoantibodies against myosin in patients with idiopathic dilated cardiomyopathy: pro-inflammatory antibodies in DCM patients.** *Biochem Biophys Res Commun* 1999, **259**:255-261.

28. Weekes J, Wheeler CH, Yan JX, Weil J, Eschenhagen T, Scholtysik G, Dunn MJ: **Bovine dilated cardiomyopathy: proteomic analysis of an animal model of human dilated cardiomyopathy.** *Electrophoresis* 1999, **20**:898-906.

29. Faulk WP, Rose M, Meroni PL, Del Papa N, Torry RJ, Labarrere CA, Busing K, Crisp SJ, Dunn MJ, Nelson DR: **Antibodies to endothelial cells identify myocardial damage and predict development of coronary artery disease in patients with transplanted hearts.** *Hum Immunol* 1999, **60**:826-832.

30. • Celis JE, Celis P, Ostergaard M, Basse B, Lauridsen JB, Ratz G, Rasmussen HH, Orntoft TF, Hein B, Wolf H, Celis A: **Proteomics and immunohistochemistry define some of the steps involved in the squamous differentiation of the bladder transitional epithelium: a novel strategy for identifying metaplastic lesions.** *Cancer Res* 1999, **59**:3000-3003.

Antibodies to disease-specific proteins identified using 2DE are used to identify metaplasia and clarify the stages in its development.

31. • Ostergaard M, Wolf H, Orntoft TF, Celis JE: **Psoriacin (S100A7): a putative urinary marker for the follow-up of patients with bladder squamous cell carcinomas.** *Electrophoresis* 1999, **20**:349-354.

One of the best candidates for a clinical disease marker yet produced by proteomics. This paper describes the discovery of this marker in urine of squamous cell carcinoma patients. The protein was identified earlier in keratinocytes.

32. • Page MJ, Amess B, Townsend RR, Parekh R, Herath A, Brusten L, Zvelebil MJ, Stein RC, Waterfield MD, Davies SC, O'Hare MJ: **Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties.** *Proc Natl Acad Sci USA* 1999, **96**:12589-12594.

The first stage of an ongoing proteomic study of breast cancer focusing on healthy breast tissue and distinguishing individual cell types.

33. Cutler P, Bell DJ, Birrell HC, Connelly JC, Connor SC, Holmes E, Mitchell BC, Monte SY, Neville BA, Pickford R *et al.*: **An integrated proteomic approach to studying glomerular nephrotoxicity.** *Electrophoresis* 1999, **20**:3647-3658.

34. Kanitz MH, Witzmann FA, Zhu H, Fultz CD, Skaggs S, Moorman WJ, Savage RE Jr: **Alterations in rabbit kidney protein expression following lead exposure as analyzed by two-dimensional gel electrophoresis.** *Electrophoresis* 1999, **20**:2977-2985.

35. Aicher L, Wahl D, Arce A, Grenet O, Steiner S: **New insights into cyclosporine A nephrotoxicity by proteome analysis.** *Electrophoresis* 1998, **19**:1998-2003.

36. Anderson NL, Esquer-Blasco R, Richardson F, Richardson F, Foxworthy P, Eacho P: **The effects of peroxisome proliferators on protein abundances in mouse liver.** *Toxicol Appl Pharmacol* 1996, **137**:75-89.

37. Giometti CS, Tollaksen SL, Liang X, Cunningham ML: **A comparison of liver protein changes in mice and hamsters treated with the peroxisome proliferator Wy-14,643.** *Electrophoresis* 1998, **19**:2498-2505.