Seminar

Proteomics: new perspectives, new biomedical opportunities

Rosamonde E Banks, Michael J Dunn, Denis F Hochstrasser, Jean-Charles Sanchez, Walter Blackstock, Darryl J Pappin, Peter J Selby

Proteomics-based approaches, which examine the expressed proteins of a tissue or cell type, complement the genome initiatives and are increasingly being used to address biomedical questions. Proteins are the main functional output, and the genetic code cannot always indicate which proteins are expressed, in what quantity, and in what form. For example, post-translational modifications of proteins, such as phosphorylation or glycosylation, are very important in determining protein function. Similarly, the effects of environmental factors or multigenic processes such as ageing or disease cannot be assessed simply by examination of the genome alone. This review describes the underlying technology and illustrates several areas of biomedical research, ranging from pathogenesis of neurological disorders to drug and vaccine design, in which potential clinical applications are being explored.

The 20th century saw massive progress in knowledge of processes underlying the genetic basis of our existence, ranging from the structure of DNA to the identification of specific disease-associated genetic abnormalities, and their clinical exploitation as the basis of trials of gene therapy or as prognostic markers. The perceived ultimate goal of identifying and sequencing the 40 000 to 60 000 genes of the human genome has reached the end of its first phase, and functional and structural genomics projects have now begun, using the genetic sequences to predict the resultant proteins and to investigate their structure and function. This information will contribute greatly to biology and medicine, but already the idea of one gene to one protein is known not to hold; organisms are far more complex than is indicated by analysis of genetic material alone. By the reverse approach, examination of the functional output, or proteome,[1,2] the expressed protein complement of a genome and also that of a tissue or cell type, additional complementary information can be obtained, which will allow the clinical benefits of the new knowledge to be fully realised.

The concept of mapping the human proteome was put forward almost 20 years ago,[3] but rapid advances in molecular-biology techniques and a change of government in the USA shifted the emphasis, leading ultimately to the human genome project. Today the pendulum is swinging back, with technological advances in protein separation and identification allowing the concept of proteomics-based approaches to become reality. Such approaches are likely to be increasingly adopted in the 21st century.

Lancet 2000; **356:** 1749–56

ICRF Cancer Medicine Research Unit, St James's University Hospital, Leeds LS9 7TF, UK (R E Banks PhD, Prof P J Selby MD); **Department of Cardiothoracic Surgery, National Heart and Lung Institute, Imperial College School of Medicine, Harefield Hospital, Harefield, UK** (M J Dunn PhD); **Central Clinical Chemistry Laboratory and Department of Medical Biochemistry, University Hospital and University of Geneva, Geneva, Switzerland** (Prof D F Hochstrasser MD, J-C Sanchez); **Cell Mapping Project, Biomolecular Structure Unit, Glaxo Wellcome Research and Development Ltd, Stevenage, UK** (W Blackstock PhD); **and ICRF Protein Sequencing Laboratory, London, UK** (D J Pappin PhD)

Correspondence to: Dr Rosamonde E Banks (e-mail: r.banks@leeds.ac.uk)

Why proteomics?

The major challenge now to biologists is to use the wealth of genetic information available from the genome-sequencing programme not just to decode the aminoacid sequence of the encoded proteins but also to find out their function. Genomics-based approaches initially use computer-based similarity searches against proteins of known function. The results may allow some broad inferences to be made about possible functions, which can then be explored experimentally. However, about a third of the sequences in organisms for which the genomes have been sequenced seem to code for proteins that are unrelated at this level to proteins of known function.[4] A major undertaking is the elucidation of the three-dimensional structure of such proteins after they have been expressed and purified in vitro, by techniques such as X-ray crystallography and nuclear magnetic resonance. These methods may allow the prediction of catalytic mechanisms, protein-protein associations, or protein-nucleic-acid interactions and provide further insight into function, which can then be investigated biologically.[4]

In addition to the genetic make-up of an individual or organism determining protein expression, many other factors determine gene and ultimately protein expression and affect proteins directly. These factors include cellular and environmental factors such as pH, hypoxia, and drug administration. Similarly, because the flow of information between genes and proteins is bidirectional, the cellular phenotype is influenced by the networks created by interaction between pathways that are regulated in a coordinated way or that overlap. Neither these effects, nor the biological basis of multigenic processes such as ageing, stress, and disease, can be identified solely from examination of the genome.[5] Further insights into possible roles of particular genes in disease can be obtained, as has already happened for many genes, from results of various techniques such as nucleic-acid-based microarray technology, which is increasingly used to characterise differences in mRNA populations between diseased and normal tissue.

There are several other compelling reasons to complement these studies and approach biological questions from the protein perspective to provide both qualitative and quantitative information about gene expression—ie, proteomics. Proteins are the functional output of the cell and therefore might be expected to

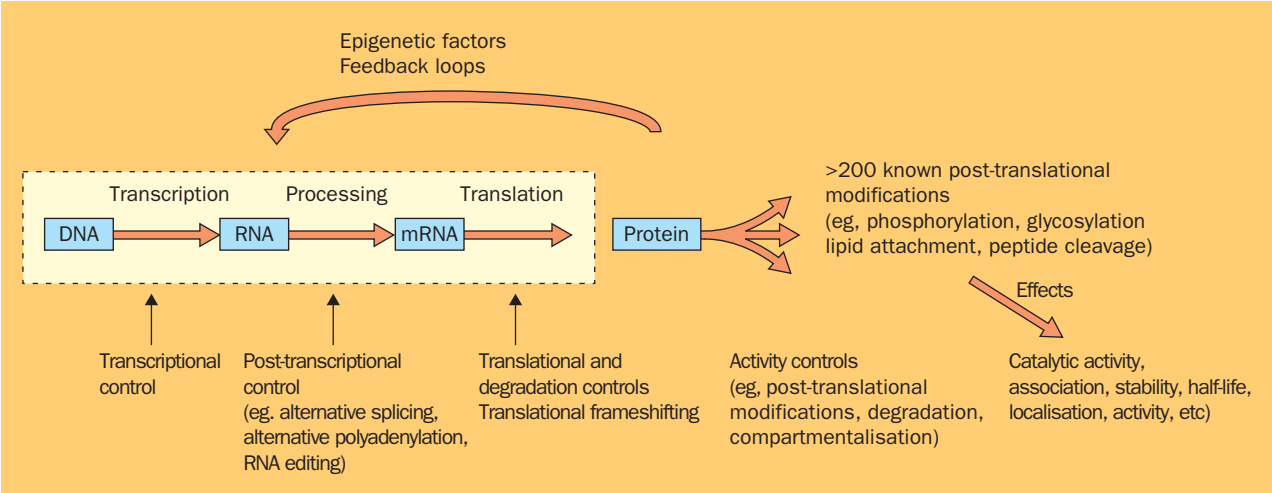

Figure 1: **The ways in which gene and protein expression can be regulated or modified from transcription to post translation**

provide the most relevant information, particularly when interpretation of their expression takes into account their dynamics in specific biological contexts. The expression or function of proteins is modulated at many points from transcription to post-translation (figure 1), which generally cannot be predicted from analysis of nucleic acids alone. There is poor correlation between the abundance of mRNA transcribed from the DNA and the respective proteins translated from that mRNA,[6] and the transcript can be spliced in various ways to yield different protein forms. Extensive changes can also be introduced during or after translation—for example, the addition of specific carbohydrate side-chains or phosphorylation—leading to multiple protein products from a single gene. Preliminary studies suggest an average number of protein forms per gene of one to two in bacteria, three in yeast, and three to more than six in human beings.[7] Thus, the human body may contain more than half a million modified proteins. Understanding of their interplay is a formidable undertaking, even at a cellular level, where only 5000–10 000 genes (although many more proteins) may be expressed, particularly with their functional state varying with post-translational modifications such as phosphorylation and dephosphorylation, glycosylation, cleavage, complex formation, and translocation within the cell.

Proteomic tools

Protein separation

The central tool for displaying the proteome is two-dimensional gel electrophoresis (figure 2).[8,9] Proteins are separated on the basis of charge in the first dimension and molecular mass in the second. Several improvements have been made to this method in the past few years, particularly in the first-dimension separation. The sample (eg, tissue, serum) is solubilised, and the proteins are denatured into their polypeptide subunits. This mixture is then separated by isoelectric focusing; on the application of a current, the charged polypeptide subunits migrate in a polyacrylamide gel strip that contains an immobilised pH gradient until they reach the pH at which their overall charge is neutral (isoelectric point or pI), hence producing a gel strip containing discrete protein bands along its length. This gel strip is then applied to the edge of a rectangular slab polyacrylamide gel containing sodium dodecyl sulphate, and the focused polypeptides migrate in an electric current into the second gel and are separated on the basis of molecular size. Typically 1000–3000 proteins per gel can be visualised, for example by staining with silver. Complementary approaches such as immunoblotting allow greater sensitivity for specific molecules. Multiple forms of individual proteins can be readily visualised (figure 2), and the particular subset of proteins examined from the proteome is determined by factors such as initial choice of sample solubilisation conditions and pH range of the gel strip used for the first dimension.

Protein identification and characterisation

Analysis of gel images with specialised software[10] allows comparisons of multiple gels both within a laboratory and, via links, to comprehensive proteome databases on the internet. Thus, by a process of subtraction, differences (eg, presence, absence, or intensity of proteins or different forms) between healthy and diseased samples can be revealed. Proteins of interest can then be identified on the basis of knowledge of the isoelectric point and apparent molecular size determined from the two-dimensional gels, supplemented by a combination of methods, generally applied hierarchically.[11] Increasingly these include highly sensitive mass-spectrometric methods, which require smaller amounts of material and have a higher throughput than conventional sequencing methods (sensitivity of femtomole to attomole concentrations). Proteins or peptides are ionised by electrospray ionisation from liquid state[12] or matrix-assisted laser desorption ionisation from solid state,[13] and the mass of the ions is measured very accurately by various coupled analysers.[11] For example, the time-of-flight analyser measures the time for ions to travel from the source to the detector (MALDI-TOF). If the excised protein spot is first digested with trypsin, which cleaves proteins at specific aminoacid sequences (if present), the protein can be broken into a mixture of peptides. The masses of the peptides can then be measured by mass spectrometry to produce a peptide mass fingerprint. This discriminating signature is compared with the peptide masses predicted from theoretical digestion of protein sequences currently contained within databases, and the protein can be identified. If necessary, actual sequence information can be obtained by tandem mass spectrometry, in which discrete peptide ions can be selected and fragmented, and complex algorithms are used to correlate the experimental data with data derived from peptide sequences in protein databases. If the protein or peptide of interest cannot be matched with any known sequence, generally, enough of

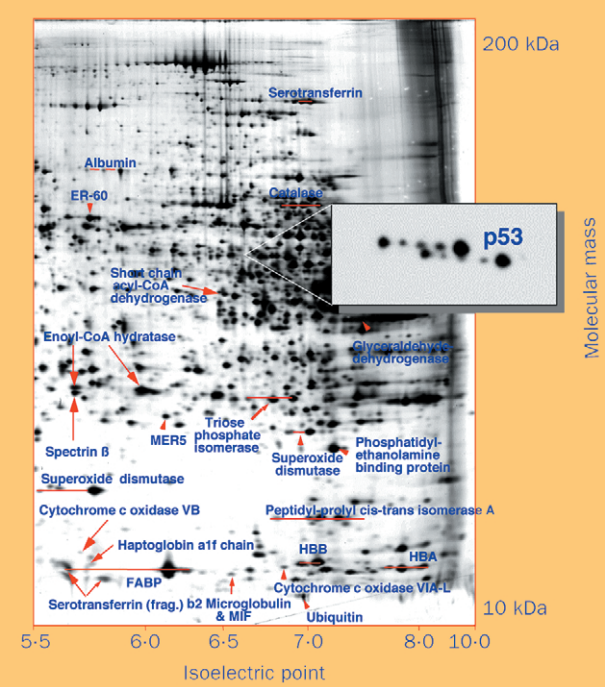

Figure 2: **Silver-stained two-dimensional polyacrylamide-gel electrophoresis pattern of a whole extract from a human liver biopsy sample**

Electrophoretic separation was carried out with immobilised pH gradients (pH 3·5–10·0, non- linear; the figure shows only pH 5·5–10·0) for isoelectric focusing in the first dimension and 9–16% T sodium dodecyl sulphate polyacrylamide-gel electrophoresis for molecular mass in the second dimension.
Inset shows enlargement of a western blot of one area of the gel, with multiple p53 polypeptides revealed by immunoblotting.

the sequence can be interpreted to allow the design of suitable nucleotide-based probes and subsequent identification of new proteins and genes.

Protein bioinformatics

Experiments done in a real laboratory need to be complemented by virtual experiments done on computer.[14] In addition to the software packages for analysing the electrophoretic separation, bioinformatic tools have been developed. Some of these are available via the internet with links to many provided from the ExPASy proteomics server (www.expasy.ch/www/tools.html).[15] These allow not only identification of proteins but further characterisation ranging from the calculation of basic physicochemical properties to the prediction of potential post-translational modifications and three-dimensional structures. Annotated protein and two-dimensional electrophoresis databases are the bioinformatic core of proteome research. SWISS-PROT is a typical example of such an annotated database (figure 3).[16] Many proteome projects are now underway, resulting in the generation of two-dimensional electrophoresis databases that are accessible on the internet (many can be accessed via links from the ExPASy server at www.expasy.ch) and can be browsed with interactive software and integrated with in-house results. These databases include protein maps of human plasma, urine, cerebrospinal fluid, and tissues such as breast, heart, and bladder transitional-cell and squamous-cell carcinomas, as well as various microorganisms. Ultimately, given the dynamic nature of the proteome, complex experimental details and related results should be displayed with the relevant biochemical pathways or disease implications highlighted.

Biomedical applications

The maximum effect of proteomics-based approaches on biomedical research has not yet been achieved, partly because of the lack of awareness in the research community about the technological advances that have made such an approach feasible on a large or small scale and partly because of the naturally occurring lead-in time after any technological advance. However, exciting progress is being made, and brief overviews of several biomedical areas are given below to illustrate the potential of this approach. Although not detailed here, proteomics-based approaches offer great potential in unravelling complex biological problems such as the nature of particular molecular complexes or pathways in disease pathogenesis. This process is best illustrated by examples from more fundamental biological research, such as the characterisation of the components of the multiprotein spliceosome complex[17] and the nuclear pore complex.[18]

Drug development

Drug development is generally based around the desire to upregulate or downregulate a specific activity implicated in disease pathogenesis or in treatment-associated side-effects. Most drugs exert their effects on proteins. The strategy of working forward from the gene has been used: a specific genetic lesion is identified and the resultant changes in protein structure, function, or expression are elucidated, so that a drug to counteract or correct such aberrations can be rationally designed. An example is the development of an inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukaemia.[19] Similarly, the identification on bioactivity grounds of a protein that is pivotal in a biological process has led to the specific design of drugs to manipulate these properties, as exemplified by the use of neutralising antibodies or inhibitors of the receptor tyrosine kinases to inhibit angiogenesis induced by vascular endothelial growth factor in tumours.[20] The challenge in proteomics-based approaches still lies in identifying the target molecules. A detailed analysis of 51 drugs showed that mechanistic-based grouping is reflected in the effects on protein,[21] and substantial investments have been made in proteomics technologies by pharmaceutical and biotechnology companies. Only a few thousand human genes are likely to be suitable targets,[22] and with any single company only able to work on a few hundred, selection is of key importance. Blackstock and Weir proposed[23] that pharmaceutical proteomics for target validation (although equally valid for other biomedical areas) be split into expression proteomics and cell mapping or interaction proteomics, each having a distinct role in the overall drug discovery process. Cell-mapping proteomics has a more defined goal of studying protein-protein interactions by systematically characterising the components of protein complexes and building up a map of cellular pathways and interactions that may be important either in a disease process or in the mechanism of action of a drug. By use of specific antibodies or artificially introduced tags, specific proteins can be isolated and any associated proteins can be identified rapidly by mass spectrometry. Targeting of analysis to multiprotein complexes may reveal likely functions of specific proteins more rapidly and indicate appropriate biological studies.[24]

Expression proteomics is the profiling of protein expression in a cell under various stimuli, probably of most use in the search for surrogate markers of drug responsiveness and in toxicology. Such an approach has formed part of the Developmental Therapeutics Program at the US National Cancer Institute, where 3989

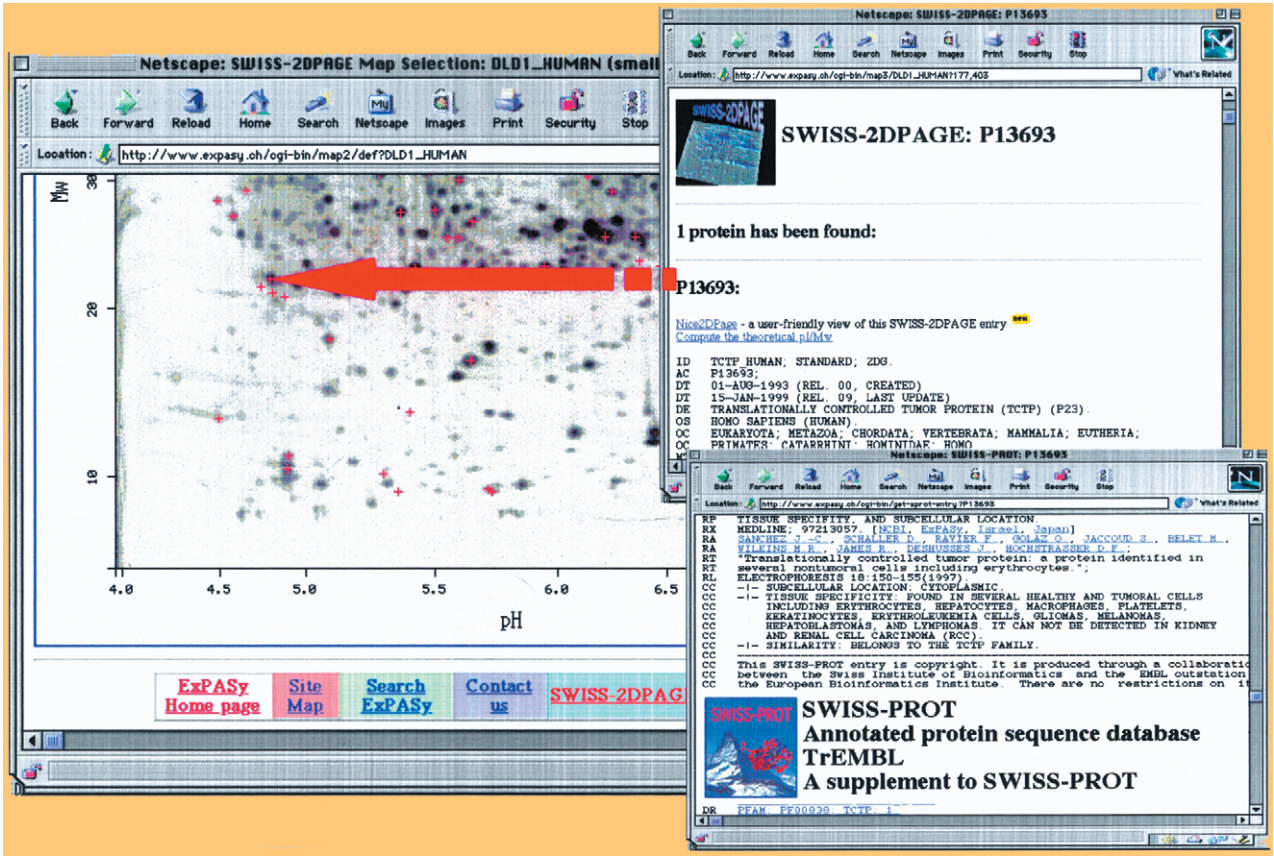

Figure 3: **Links and level of annotation detail of SWISS-2DPAGE and SWISS-PROT**
Arrow indicates the location of a growth-related protein, translationally controlled tumour protein (TCTP), on the two-dimensional map of the human colorectal adenocarcinoma cell line DLD-1, with the adjacent text containing part of the corresponding SWISS-PROT entry.

compounds have been screened against a panel of 60 cell lines, and their molecular pharmacology characterised by two-dimensional electrophoresis.[25] However, expression proteomics must compete with the genomic analogues of differential gene expression and chip-based hybridisation technologies. For example, with a DNA microchip,[26] the temporal expression changes in 8613 genes in human fibroblasts responding to serum were examined. Such technologies provide results more rapidly and are less labour-intensive than the corresponding proteomics-based approaches but, given the poor correlation between mRNA and protein product[6] and the higher value of the protein information, substantial investment is being made to automate large-scale proteomics operations. Clearly, genomic and proteomic approaches complement each other in terms of the information produced and their relative advantages and disadvantages.

In addition to target selection, equally important in progress to clinical use are target validation and toxicity studies. A classic example of validation is the study showing the actions of the cholesterol-lowering agent lovastatin in affecting proteins involved with cholesterol metabolism.[27] In assessments of the mechanisms of toxicity of ciclosporin, Aicher and colleagues showed decreased renal expression of the calcium-binding protein calbindin in human renal-transplant recipients with nephrotoxicity, which may be implicated in the tubular calcification side-effects of this drug.[28]

Neurological disorders

Post-translational modifications of proteins are not detectable by examination of nucleic acids, and nowhere have such modifications been implicated in disease processes as in the following two examples. Prion diseases, characterised by fatal degenerative encephalopathy, are rare in human beings. Creutzfeld-Jakob disease (CJD) affects about one person per million per year and may be inherited, acquired, or, as in most cases, sporadic. The infectious particle (prion) is unusual, apparently being a conformational isomer of a normal cellular glycoprotein (PrP^{C}) termed PrP^{Sc}, which is resistant to proteolytic degradation. Neuropathological diagnosis necessitates a brain biopsy or necropsy sample, although many sporadic cases have a typical clinical picture. However, with the emergence of a new variant form of the disease (vCJD) associated with the bovine spongiform encephalopathy epidemic, the search for diagnostic and screening tests that can be used before death has intensified. Analysis of cerebrospinal fluid by two-dimensional electrophoresis revealed two proteins, designated p130 and p131, the presence of which could be used to discriminate between CJD and other types of dementia with a sensitivity of 88% and specificity of 99%.[29] Sequencing identified them as members of the 14-3-3 family. The value of 14-3-3 proteins as discriminatory markers for CJD in patients with dementia has since been confirmed in several studies,[30,31] although clinical use in vCJD has yet to be clarified.[32,33] The proteins are also present in some patients with other neurological disorders without dementia, possibly reflecting neuronal damage rather than disease pathogenesis.

On electrophoresis, PrP is shown to exist as a diglycosylated, monoglycosylated, or non-glycosylated form. Protease treatment degrades all PrP^{C} but only partially degrades PrP^{Sc} leading to increased electrophoretic mobility of the three isoforms, with

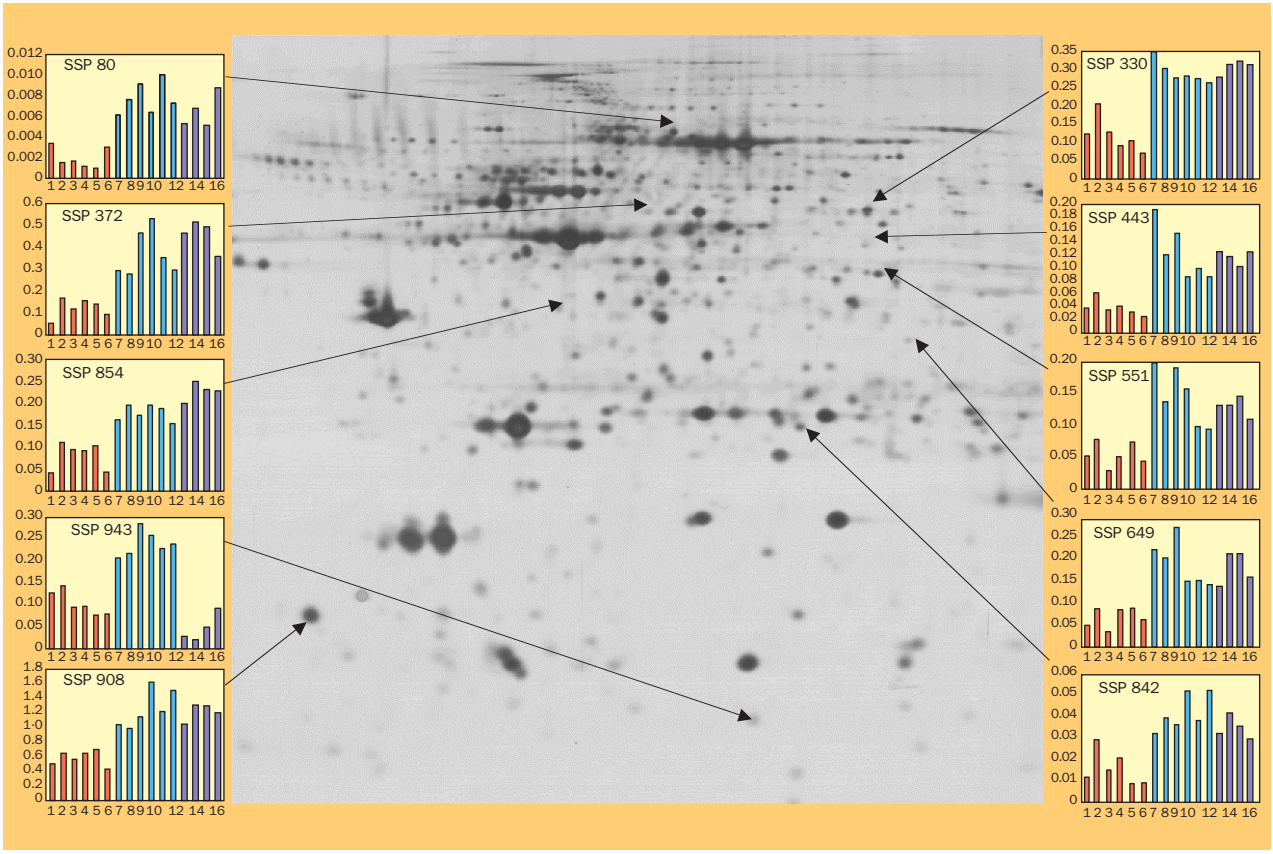

Figure 4: **Reference 'master' two-dimensional electrophoretic image of the proteins of the bovine heart**
Protein spots indicated with arrows are present in significantly reduced amounts in the hearts of animals with inherited dilated cardiomyopathy as shown in the quantitative histograms. Red bars (1–6) represent animals with clinical dilated cardiomyopathy, and blue bars (7–12) normal control animals. Purple bars (13–16) represent animals that are clinically normal but are offspring of two parents with clinical dilated cardiomyopathy (ie, the animals are genotypically affected but phenotypically normal). Most proteins are expressed in normal concentrations in the hearts of these animals, but protein spot 913 is expressed in low concentrations, as in the affected animals.

characteristic profiles in their relative abundances for different clinical variants of CJD.[34,35] The unique profile generated by vCJD brain samples resembled that seen in samples from animals with bovine spongiform encephalopathy.[35] The detection of PrP protein in all of nine tonsil samples from patients with vCJD, with a glycoform profile similar to that in brain samples of vCJD but differing slightly in the relative ratios of the isoforms,[33] raises the possibility that tonsil biopsy may be useful as a test to detect vCJD. Although obviously still at very early stages of development for this disease, such knowledge of specific protein abnormalities may lead to the development of therapeutic approaches. A recent study of peptides designed to reverse the conformational changes seen in PrP showed a greatly reduced infectivity of prion infectious material in mice with experimental scrapie.[36]

Alzheimer's disease is the commonest single cause of dementia. The characteristic neuropathological brain lesions are senile plaques and cytoplasmic neurofibrillary tangles. Aminoacid sequencing and identification of a common protein fragment of 4 kDa (β-amyloid) as the major extracellular protein component of senile plaques in both Down's syndrome and Alzheimer's disease[37,38] subsequently formed the basis for cloning of the full-length amyloid precursor protein (APP) gene.[39] This step was pivotal in highlighting the involvement of this protein. Normally cleaved into fragments and cleared, cleavage into amyloid $\beta_{1\text{-}42}$, in particular, results in increased deposition and ultimately neurotoxicity, although the exact mechanism is not known.[40] Further supportive evidence includes the identification of mis-sense mutations in APP, which result in increased deposition of amyloid β, mutations in the presenilin genes which are postulated to affect APP trafficking, and a specific polymorphism (ε4 allelle) of apolipoprotein E (ApoE) as a risk factor for Alzheimer's disease, possibly by differential binding to amyloid β and subsequent enhancement of aggregation.[40]

The major protein component of neurofibrillary tangles is the microtubule-associated protein tau, which is present as six alternatively spliced isoforms.[41] Immunochemical and electrophoretic analyses have shown that pathological tau differs from normal tau biochemically;[41] it is less soluble and particular post-translational modifications occur, including ubiquitination, glycosylation, glycation, and, most notably, hyperphosphorylation.[42] The relative contributions of these changes to the formation of the tangles is as yet unclear. Many studies of tau and APP have used two-dimensional electrophoretic and mass spectrometric techniques and, in view of the importance of protein-protein interactions and post-translational modifications in Alzheimer's disease and other neurological diseases, this is likely to remain a principal approach in the future.

Infectious diseases

Identification of proteins produced by microorganisms is facilitated by the small number of genes and the completion of genome sequencing for many micro-organisms.[43] The main aim of most studies has been the search for new diagnostic markers, candidate antigens for vaccines, and determinants of virulence. For some of the

microorganisms studied, two-dimensional electrophoresis databases are available on the internet (www.expasy.ch/).

Mycobacterium tuberculosis is probably the most studied microorganism. Tuberculosis still causes more than 2 million deaths each year, so an international priority is to improve diagnosis and develop a more effective vaccine. The current BCG vaccine, an attenuated avirulent form of *M bovis*, varies in efficacy from almost zero protection in southern India to about 70% protection in the UK. Culture filtrates of *M tuberculosis* provide effective vaccination in animal models,[44] and several research groups have been systematically characterising the mycobacterial proteomes, particularly the secreted proteins.[45,46] After further screening for reactivity in human beings, several of the antigens identified by these studies have now been incorporated into trial vaccines.[47] Studies examining the humoral response in vivo have also identified relevant antigens, with several having serodiagnostic potential in early tuberculosis.[48]

Examination of genetic differences between virulent *M tuberculosis*, virulent *M bovis*, and the avirulent *M bovis* BCG by genomic analysis identified three distinct regions, with one (RD1) being present in virulent strains but absent from all BCG sub-strains.[49] When RD1 was introduced into the BCG strain, expression of several proteins was altered, producing an expression profile almost identical to that of the virulent strain. This finding is not only useful in identifying potential regulators of virulence but also may provide the basis for discrimination between BCG immunisation and the presence of virulent tuberculosis.

Heart disease

Heart failure, arising from systemic disease or specific heart-muscle disease, is one of the leading causes of morbidity and mortality in more developed countries. The pathogenesis of the cardiac dysfunction is still largely unknown, but a proteomics-based approach in characterising overall changes in protein expression in heart disease and heart failure may provide new insights into the cellular mechanisms involved in cardiac dysfunction, together with new diagnostic markers and therapeutic opportunities. Federated two-dimensional electrophoresis databases of human cardiac proteins have been established (www.expasy.ch/ch2d/2d-index.html), and several hundred cardiac proteins have been identified.

The proteomic focus so far has been on dilated cardiomyopathy, which almost certainly has multiple causes. The combined results of these studies[50–52] have shown that the expression of some 100 cardiac proteins is significantly different from normal in dilated cardiomyopathy, and most of the proteins are less abundant in the diseased than in the normal heart. Most of the identified proteins can be grouped into three broad functional classes: cytoskeletal and myofibrillar proteins; proteins associated with mitochondria and energy production; and proteins associated with stress responses. With use of animal models to overcome inherent confounding factors such as genetic variability and effects of therapy, proteomic studies of dilated cardiomyopathy in cattle[53] and pacing-induced heart failure in dogs[54] have shown similarities with the proteomic analysis of human dilated cardiomyopathy. Most of the changes involve decreased protein abundance in the diseased heart (figure 4), and the proteins affected are mainly cytoskeletal or associated with mitochondria and energy metabolism. The most notable change in bovine dilated cardiomyopathy was a seven-fold increase in the enzyme ubiquitin C-terminal hydrolase. This finding accords with the suggestion that inappropriate ubiquitination and subsequent degradation of proteins by the proteasome may contribute to the development of heart failure.[55] In addition to the global proteomic approach (analogous to expression profiling), cardiac antigens that elicit specific antibody responses in vivo are being identified. Several of these antigens may be involved in processes of acute[56] and chronic[57] rejection after cardiac transplantation and are currently being investigated as potential non-invasive markers.

Cancer

Although studies concerned with the identification of novel antigens or markers for diagnostic, prognostic, or therapeutic use have been paramount, molecules and processes implicated in carcinogenesis per se are increasingly being investigated. Most tumour markers in current use were identified from protein-based approaches, from the identification in the 1800s of an abnormal urinary precipitate in myeloma (Bence-Jones protein) to the generation of tumour-specific antibodies against epithelial cancer cell lines (CEA and CA 125).[58] Genetic markers, detected cytogenetically or by mutation detection, are also now entering clinical practice,[58] but some changes likely to be important in carcinogenesis, diagnosis, and prognosis, such as abnormal expression of proto-oncogenes, may not be associated with a detectable genetic lesion. For proteins implicated in cancer, the use of multiple antibodies allows simultaneous characterisation of several proteins acting in a network. For example, overexpression and many post-translational modifications (largely phosphorylations) of several oncogene products and cell-cycle proteins such as p53 can be detected in transformed liver cells.[59] Aberrant glycosylation of many proteins with a known cancer association has been described,[60] and proteomics-based approaches are ideally placed to characterise such post-translational modifications, although much work still needs to be done to assess their clinical significance.

There have been many studies involving proteomics-based approaches into cancer,[61] and the potential of such work is exemplified by the studies in squamous-cell carcinoma of the bladder.[62,63] In Europe and America, about 90% of bladder cancers are transitional-cell carcinomas, but in regions where *Schistosoma haematobium* is endemic, such as parts of Africa, about 80% of bladder cancers are squamous-cell carcinomas. The histological evolution of these tumours is uncertain, and they may be difficult to distinguish from poorly differentiated transitional-cell carcinomas with areas of squamous differentiation. With two-dimensional electrophoresis, six squamous-cell carcinomas were correctly identified from 150 bladder tumours on the basis of characteristic protein-expression patterns. Psoriasin (also identified as a potential urinary marker), keratins 10 and 14, PA-FABP, galectin 7, and stratifin were particularly valuable in assessing the degree of differentiation.[62] In a novel approach,[63] the evolution of squamous-cell carcinomas is being explored by use of antibodies to proteins differentially expressed in these tumours and normal urothelium to stain serial sections ('immunowalking') from different areas of cystectomy specimens. This approach is based on the premise that the urothelium of these patients is likely to encompass a range of abnormalities from early metaplasia to invasive cancer. Proteomics-based studies of many tumour types are now underway, and they are likely to benefit from the use of complementary techniques such as laser capture microdissection[64,65] to isolate malignant cells for

electrophoretic analysis, thus facilitating the discovery of tumour-specific proteins.

Future perspective

Proteomics complements genomics-based approaches, providing additional information but presenting different technical challenges. For example, there is no protein equivalent of PCR for amplification of low-abundance proteins, so a range of detection from one to several million molecules per cell is needed. Proteins have properties arising from their folded structures, so generic methods are difficult to design and apply, and the analysis and significance of post-translational modifications provide a major challenge, both in normal and disease conditions. Certain technological processes, particularly protein separation and analysis, are inherently skill-based and remain difficult to automate. Separation techniques such as capillary electrophoresis may be more amenable to automation but are unlikely to replace two-dimensional electrophoresis with its superior resolving power. However, many complementary technologies are being developed and either alone or in combination will undoubtedly assume prominent roles in the armouries of proteomics and functional or structural genomics-based approaches, whether in expression profiling or molecular interaction screening. These include protein arrays,[66] the yeast two-hybrid system,[67] phage-display antibody libraries,[68] surface-enhanced laser desorption and ionisation,[69] and biological activity profiling of families of proteins such as proteases.[70] Parallel development of high-quality bioinformatic facilities is also essential, involving not only genetic and protein sequences but also the experimental conditions to unravel the biology. Clearly, the particular technological approach adopted will depend on the features of the proteins or proteome under investigation.

The proteome is dynamic, reflecting the conditions to which a cell is exposed or, for example, a specific disease process. There is therefore potentially a huge number of proteomes for each cell type. Hypothesis-driven research with careful selection of the specific features of a proteome that provide information relevant to the particular biomedical question is particularly important given that the bottleneck is likely to lie not in identifying the proteins but in their downstream characterisation. Proteomics benefits from integration with the genomics initiative—the gain from this interface as genome sequencing and functional genomics projects are completed, together with the introduction of national proteomic funding initiatives, should allow proteomics-based approaches to realise their potential in biomedical research and translation into clinical practice.

We thank the Imperial Cancer Research Fund, the British Heart Foundation, and Proteome Sciences plc for supporting our research in this area.

References

1 Wasinger VC, Cordwell SJ, Cerpa-Poljak A, et al. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 1995; **16:** 1090–94.

2 Wilkins MR, Sanchez JC, Gooley AA, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 1996; **13:** 19–50.

3 Anderson NG, Anderson L. The human protein index. *Clin Chem* 1982; **28:** 739–48.

4 Eisenstein E, Gilliland GL, Herzberg O, et al. Biological function made crystal clear: annotation of hypothetical proteins via structural genomics. *Curr Opin Biotechnol* 2000; **11:** 25–30.

5 Strohman R. Epigenesis: the missing beat in biotechnology. *Biotechnology* 1994; **12:** 156–64.

6 Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997; **18:** 533–37.

7 Wilkins MR, Sanchez JC, Williams KL, Hochstrasser DF. Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* 1996; **17:** 830–38.

8 O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 1975; **250:** 4007–21.

9 Gorg A, Obermaier C, Boguth G, et al. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 2000; **21:** 1037–53.

10 Appel RD, Palagi PM, Walther D, et al. Melanie II: a third-generation software package for analysis of two-dimensional electrophoresis images—features and user interface. *Electrophoresis* 1997; **18:** 2724–34.

11 James P. Protein identification in the post-genome era: the rapid rise of proteomics. *Q Rev Biophys* 1997; **30:** 279–331.

12 Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989; **246:** 64–71.

13 Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988; **60:** 2299–301.

14 Hochstrasser DF. Proteome in perspective. *Clin Chem Lab Med* 1998; **36:** 825–36.

15 Hochstrasser DF, Appel RD, Golaz O, et al. Sharing of worldwide spread knowledge using hypermedia facilities and fast communication protocols (Mosaic and World Wide Web): the example of ExPASy. *Methods Inf Med* 1995; **34:** 75–78.

16 Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1998. *Nucleic Acids Res* 1998; **26:** 38–42.

17 Neubauer G, King A, Rappsilber J, et al. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet* 1998; **20:** 46–50.

18 Rout MP, Aitchison JD, Suprapto A, et al. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J Cell Biol* 2000; **148:** 635–51.

19 Druker BJ, Lydon NB. Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest* 2000; **105:** 3–7.

20 Schlaeppi JM, Wood JM. Targeting vascular endothelial growth factor (VEGF) for anti-tumor therapy, by anti-VEGF neutralizing monoclonal antibodies or by VEGF receptor tyrosine-kinase inhibitors. *Cancer Metastasis Rev* 2000; **18:** 473–81.

21 Anderson NL, Anderson NG. Proteome and proteomics: New technologies, new concepts, and new words. *Electrophoresis* 1998; **19:** 1853–61.

22 Drews J. Genomic sciences and the medicine of tomorrow. *Nat Biotechnol* 1996; **14:** 1516–18.

23 Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* 1999; **17:** 121–27.

24 Lamond AI, Mann M. Cell biology and the genome projects: a concerted strategy for characterizing multiprotein complexes by using mass spectrometry. *Trends Cell Biol* 1998; **7:** 139–42.

25 Myers TG, Anderson NL, Waltham M, et al. A protein expression database for the molecular pharmacology of cancer. *Electrophoresis* 1997; **18:** 647–53.

26 Iyer VR, Eisen MB, Ross DT, et al. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999; **283:** 83–87.

27 Anderson NL, Esquer-Blasco R, Hofmann JP, Anderson NG. A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies. *Electrophoresis* 1991; **12:** 907–30.

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

29 Harrington MG, Merril CR, Asher DM, Gajdusek DC. Abnormal proteins in the cerebrospinal fluid of patients with Creutzfeldt-Jakob disease. *N Engl J Med* 1986; **315:** 279–83.

30 Zerr I, Bodemer M, Gefeller O, et al. Detection of 14-3-3 protein in the cerebrospinal fluid supports the diagnosis of Creutzfeldt-Jakob disease. *Ann Neurol* 1998; **43:** 32–40.

31 Beaudry P, Cohen P, Brandel JP, et al. 14-3-3 protein, neuron-specific enolase, and S-100 protein in cerebrospinal fluid of patients with Creutzfeldt-Jakob disease. *Dementia Geriatr Cogn Dis* 1999; **10:** 40–46.

32 Zeidler M, Stewart GE, Barraclough CR, et al. New variant Creutzfeldt-Jakob disease: neurological features and diagnostic tests. *Lancet* 1997; **350:** 903–07.

33 Hill AF, Butterworth RJ, Joiner S, et al. Investigation of variant Creutzfeldt-Jakob disease and other human prion diseases with tonsil biopsy samples. *Lancet* 1999; **353:** 183–89.

34 Parchi P, Castellani R, Capellari S, et al. Molecular basis of

phenotypic variability in sporadic Creutzfeldt-Jakob disease. *Ann Neurol* 1996; **39:** 767–78.
35 Collinge J, Sidle KC, Meads J, Ironside J, Hill AF. Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature* 1996; **383:** 685–90.
36 Soto C, Kascsak RJ, Saborio GP, et al. Reversion of prion protein conformational changes by synthetic b-sheet breaker peptides. *Lancet* 2000; **355:** 192–97.
37 Masters CL, Simms G, Weinman NA, et al. Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc Natl Acad Sci USA* 1985; **82:** 4245–49.
38 Glenner GG, Wong CW. Alzheimer's disease and Down's syndrome: sharing of a unique cerebrovascular amyloid fibril protein. *Biochem Biophys Res Commun* 1984; **122:** 1131–35.
39 Kang J, Lemaire H-G, Unterbeck A, et al. The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* 1987; **325:** 733–36.
40 Storey E, Cappai R. The amyloid precursor protein of Alzheimer's disease and the Ab peptide. *Neuropathol Appl Neurobiol* 1999; **25:** 81–97.
41 Tolnay M, Probst A. Review: tau prtoein pathology in Alzheimer's disease and related disorders. *Neuropathol Appl Neurobiol* 1999; **25:** 171–87.
42 Grundke-Iqbal I, Iqbal K, Tung YC, et al. Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc Natl Acad Sci USA* 1986; **83:** 4913–17.
43 Cash P. Proteomics in medical microbiology. *Electrophoresis* 2000; **21:** 1187–201.
44 Andersen P. Effective vaccination of mice against *Mycobacterium tuberculosis* infection with a soluble mixture of secreted mycobacterial proteins. *Infect Immun* 1994; **62:** 2536–44.
45 Andersen P, Andersen AB, Sorensen AL, Nagai S. Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. *J Immunol* 1995; **154:** 3359–72.
46 Weldingh K, Rosenkrands I, Jacobsen S, et al. Two-dimensional electrophoresis for analysis of *Mycobacterium tuberculosis* culture filtrate and purification and characterization of six novel proteins. *Infect Immun* 1998; **66:** 3492–500.
47 Orme IM. New vaccines against tuberculosis: the status of current research. *Infect Dis Clin North Am* 1999; **13:** 169–85.
48 Samanich KM, Belisle JT, Sonnenberg MG, et al. Delineation of human antibody responses to culture filtrate antigens of *Mycobacterium tuberculosis*. *J Infect Dis* 1998; **178:** 1534–38.
49 Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M bovis*. *J Bacteriol* 1996; **178:** 1274–82.
50 Corbett JM, Why HJ, Wheeler CH, et al. Cardiac protein abnormalities in dilated cardiomyopathy detected by two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis* 1998; **19:** 2031–42.
51 Knecht M, Regitz-Zagrosek V, Pleissner K-P, et al. Characterization of myocardial protein composition in dilated cardiomyopathy by two-dimensional gel electrophoresis. *Eur Heart J* 1994; **15** (suppl D): 37–44.
52 Pleissner KP, Soding P, Sander S, et al. Dilated cardiomyopathy-associated proteins and their presentation in a WWW-accessible two-dimensional gel protein database. *Electrophoresis* 1997; **18:** 802–08.
53 Weekes J, Wheeler CH, Yan JX, et al. Bovine dilated cardiomyopathy: proteomic analysis of an animal model of human dilated cardiomyopathy. *Electrophoresis* 1999; **20:** 898–906.
54 Heinke MY, Wheeler CH, Chang D, et al. Protein changes observed in pacing-induced heart failure using two-dimensional electrophoresis. *Electrophoresis* 1998; **19:** 2021–30.
55 Field ML, Clark JF. Inappropriate ubiquitin conjugation: a proposed mechanism contributing to heart failure. *Cardiovasc Res* 1997; **33:** 8–12.
56 Latif N, Rose ML, Yacoub MH, Dunn MJ. Association of pretransplantation antiheart antibodies with clinical course after heart transplantation. *J Heart Lung Transplant* 1995; **14:** 119–26.
57 Wheeler CH, Collins A, Dunn MJ, et al. Characterization of endothelial antigens associated with transplant-associated coronary artery disease. *J Heart Lung Transplant* 1995; **14:** S188–97.
58 Sell S. Cancer markers of the 1990s: comparison of the new generation of markers defined by monoclonal antibodies and oncogene probes to prototypic markers. *Clin Lab Med* 1990; **10:** 1–37.
59 Sanchez JC, Wirth P, Jaccoud S, et al. Simultaneous analysis of cyclin and oncogene expression using multiple monoclonal antibody immunoblots. *Electrophoresis* 1997; **18:** 638–41.
60 Taylor-Papadimitriou J, Epenetos AA. Exploiting altered glycosylation patterns in cancer: progress and challenges in diagnosis and therapy. *Trends Biotechnol* 1994; **12:** 227–33.
61 Alaiya AA, Franzen B, Auer G, Linder S. Cancer proteomics: from identificationb of novel markers to creation of artificial learning models for tumor classification. *Electrophoresis* 2000; **21:** 1210–17.
62 Ostergaard M, Rasmussen HH, Nielsen HV, et al. Proteome profiling of bladder squamous cell carcinomas: identification of markers that define their degree of differentiation. *Cancer Res* 1997; **57:** 4111–17.
63 Celis JE, Celis P, Ostergaard M, et al. 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:** 3003–09.
64 Banks RE, Dunn MJ, Forbes MA, et al. The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis: preliminary findings. *Electrophoresis* 1999; **20:** 689–700.
65 Emmert-Buck MR, Bonner RF, Smith PD, et al. Laser capture microdissection. *Science* 1996; **274:** 998–1001.
66 Walter G, Bussow K, Cahill D, Lueking A, Lehrach H. Protein arrays for gene expression and molecular interaction screening. *Curr Opin Microbiol* 2000; **3:** 298–302.
67 Uetz P, Giot L, Cagney G, et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae. Nature* 2000; **403:** 623–27.
68 Griffiths AD, Duncan AR. Strategies for selection of antibodies by phage display. *Curr Opin Biotechnol* 1998; **9:** 102–08.
69 Senior K. Fingerprinting disease with protein chip arrays. *Mol Med Today* 1999; **5:** 326–27.
70 McKerrow JH, Bhargava V, Hansell E, et al. A functional proteomics screen of proteases in colorectal carcinoma. *Mol Med* 2000; **6:** 450–60.