

# Disease proteomics

Sam Hanash

Department of Pediatrics, University of Michigan, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109, USA  
(e-mail: shanash@umich.edu)

The sequencing of the human genome and that of numerous pathogens has opened the door for proteomics by providing a sequence-based framework for mining proteomes. As a result, there is intense interest in applying proteomics to foster a better understanding of disease processes, develop new biomarkers for diagnosis and early detection of disease, and accelerate drug development. This interest creates numerous opportunities as well as challenges to meet the needs for high sensitivity and high throughput required for disease-related investigations.

**D**espite tremendous advances in our understanding of the molecular basis of diseases such as cancer, substantial gaps remain both in our understanding of disease pathogenesis and in the development of effective strategies for early diagnosis and for treatment. The current interest in proteomics is due in part to the prospects that a proteomic approach to disease investigations will overcome some of the limitations of other approaches<sup>1</sup>. The opportunities as well as the challenges facing disease proteomics are formidable. Particularly promising areas of research include: delineation of altered protein expression, not only at the whole-cell or tissue levels, but also in subcellular structures, in protein complexes and in biological fluids; the development of novel biomarkers for diagnosis and early detection of disease; and the identification of new targets for therapeutics and the potential for accelerating drug development through more effective strategies to evaluate therapeutic effect and toxicity.

The dynamic nature of the proteome of a cell or a tissue provides ample justification for studying gene expression in disease directly at the proteomic level. But capturing this dynamic state represents a technological challenge. Undoubtedly, tackling the numerous facets of disease proteomics requires implementation of multiple strategies and technology platforms.

Proteome profiling technologies are currently evolving in a manner that emphasizes the need for sensitivity and throughput. No one technology is likely to emerge that will meet the needs of all types of proteomics-based investigations, from expression proteomics to functional proteomics, particularly as they relate to disease.

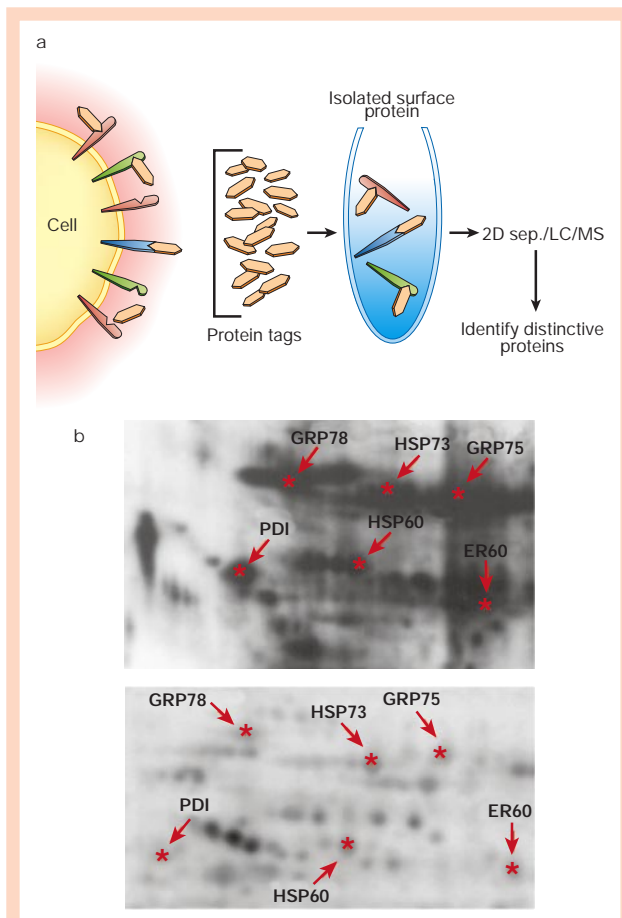
## The use of two-dimensional gels

During the early years of proteomics and until relatively recently, profiling of protein expression in disease relied primarily on the use of two-dimensional polyacrylamide gel electrophoresis (2D PAGE), which was later combined with mass spectrometry<sup>2</sup>. Most studies of this nature followed an approach in which a cocktail was used to solubilize the protein contents of an entire cell population, tissue or biological fluid, followed by separation of the protein contents of the lysate using 2D gels and visualization of the separated proteins using silver staining. It became clear that such an approach allows only a limited display of protein content that consisted of relatively abundant proteins. Nevertheless, profiling of disease tissues

using this approach has had some utility. For example, it was demonstrated long before the use of DNA microarrays that leukaemias could be classified into their different subtypes using 2D PAGE<sup>3</sup>.

Numerous other studies have also identified disease-related changes in protein expression, primarily using 2D PAGE and mass spectrometry. One such example is provided by studies of heart disease<sup>4</sup>, the spectrum of which encompasses a broad set of pathological conditions, some with acute onset of severe disease and others with slow, chronic progression. To assist in data gathering and mining, online 2D gel-derived databases of protein expression in the myocardium for human and other species were constructed<sup>5,6</sup>. These databases have allowed investigators to compare data and establish reference standards. Relevant findings have emerged from studies of changes in myocardial proteins associated with human heart failure as well as from studies of animal models of heart failure and of isolated rat myocytes<sup>4</sup>. Although only a small proportion of the proteome has been analysed, pronounced changes in the composition of the cardiac proteome have been found, affecting proteins with diverse functions. Altered overall levels of specific proteins or altered post-translational modifications of proteins such as myosin light chain 2 have been reported in the failing heart<sup>7</sup>. And protein expression studies have uncovered proteins that exhibited new disease-related post-translational modifications with predicted functional relevance<sup>8-11</sup>. This level of progress typifies that made in profiling disease tissue in a wide variety of diseases.

An important development in 2D PAGE is the use of immobilized pH gradients (IPGs) in which the pH gradient is fixed within the acrylamide matrix. IPGs also allow production of gels that cover a defined pH range, from wide to narrow<sup>12,13</sup>. A variation on this theme is the use of so-called 'zoom gels' in which the protein contents of an individual sample are first fractionated into narrow pH ranges under low resolution, and then each fraction undergoes high-resolution separation by 2D PAGE<sup>14</sup>. For example, the pre-fractionation of serum using this approach has enhanced the ability to detect low-abundance and potentially new circulating disease-marker proteins<sup>14</sup>. Yet another innovation in 2D gels is the use of differential in-gel electrophoresis (DIGE), in which two pools of proteins are labelled with different fluorescent dyes<sup>15</sup>. The labelled proteins are mixed and separated in the same 2D gel. In one study<sup>16</sup>, 2D DIGE was applied to quantify the differences in protein expression between oesophageal carcinoma cells and normal



**Figure 1** Affinity capture of surface-membrane proteins. Biotinylation reagents provide a 'tag' that transforms poorly detectable surface-membrane proteins into probes that can be recognized by a labelled detection reagent. **a**, Intact cells are tagged using lipid-insoluble biotin reagents. Tagged proteins are captured after cell lysis using avidin columns and subsequently eluted. Following a separation step using 2D gels or other means, tagged proteins are detected with a labelled avidin conjugate. Individual proteins are identified by mass spectrometry. **b**, Close-up section of a 2D pattern in which biotinylated proteins are selectively visualized (top) in contrast with the pattern of the same whole-cell lysate visualized by silver staining (bottom). Several selectively visualized proteins were found to have chaperone functions<sup>19</sup>.

epithelial cells. A large number of proteins were found to be either upregulated or downregulated in cancer cells.

### Sample fractionation prior to analysis

Some of the main challenges facing expression proteomics, be it using 2D PAGE or any other approach, include the great dynamic range of protein abundance and a wide range of protein properties, including mass, isoelectric point, extent of hydrophobicity and post-translational modifications. Reducing sample complexity prior to analysis — for example, by analysing protein subsets and subcellular organelles separately — improves the reach of 2D gels or other separation techniques for the quantitative analysis of low-abundance proteins. An elegant demonstration of the power of sub-proteome analysis is illustrated in studies of phagosomes<sup>17</sup>; these have led to the identification of over 250 proteins from this organelle and the demonstration that phagosomes are formed by direct association and fusion of endoplasmic reticulum to the plasma membrane during early phagocytosis.

The isolation of sub-proteomes may be combined with protein tagging to further enhance sensitivity, as in the case of surface-

membrane proteins, a compartment rich in diagnostic and therapeutic targets. Protein tagging technologies are currently being implemented for the comprehensive analysis of the cell-surface proteome (Fig. 1). A surface-protein biotinylation strategy, coupled with the use of mass spectrometry, was applied to the gastric pathogen *Helicobacter pylori*, leading to the identification of new surface-membrane proteins<sup>18</sup>. This strategy has also led to the detection and identification of many new proteins on the surface of cancer cells (Fig. 1)<sup>19</sup>.

### Beyond 2D gels for disease expression profiling

Even with all the improvements that could be introduced, 2D gels will probably remain a rather low-throughput approach that requires a relatively large amount of sample. The latter is particularly problematic for clinical samples, as such samples are generally procured in limited amounts. Furthermore, tissue heterogeneity complicates the analysis of clinical samples. Various tissue microdissection approaches are beneficial to reduce heterogeneity, but they further reduce the amount of sample available. In particular, the use of laser-capture microdissection, which allows defined cell types to be isolated from tissues, yields amounts of proteins that are difficult to reconcile with the need for greater amounts for 2D gels<sup>1</sup>.

Undoubtedly, various non-gel-based schemas that rely on liquid-based separations of proteins or peptides, with or without tagging, will have utility for disease proteomics, particularly given their potential for automation. Additionally, advances in microfluidic technology will likely allow automated separation of proteins in complex lysates using much reduced sample amounts. Microfluidic systems already have been integrated with mass spectrometry for protein digestion and identification<sup>20</sup>.

Non-separation-based strategies, including direct profiling using mass spectrometry or the use of protein microarrays, are important developments. Mass spectrometry has been applied to the *in situ* proteomic analysis of tissues, an approach that allows imaging of protein expression in normal and disease tissues<sup>21</sup>. By this method, frozen tissue is sliced and sections are applied on a matrix-assisted laser desorption/ionization (MALDI) plate and analysed at regular spatial intervals. The mass spectra obtained at different intervals are compared, yielding a spatial distribution of individual masses across the tissue section (Fig. 2). Mass profiles of tissue sections obtained from normal and disease tissues may be compared to detect altered protein expression. Tumour analyses using this approach have uncovered differences in protein expression between normal and tumour tissues that may have specificity for different tumour types<sup>21</sup>.

### DNA and protein microarrays in disease investigation

#### Cancer profiling using DNA microarrays

Profiling gene expression using DNA arrays has had a tremendous impact on biomedical research. Disease-related applications of DNA microarrays include uncovering unsuspected associations between genes and specific clinical features of disease that are helping devise new molecular-based classifications of disease. In relation to cancer, most published studies of tumour analysis using DNA microarrays have examined pathologically homogeneous sets of tumours to identify clinically relevant subtypes (for example, responders versus non-responders), pathologically distinct subtypes of tumours of the same lineage to identify molecular correlates (for example, high-stage versus low-stage tumours), or tumours of different lineages to identify molecular signatures for each lineage.

Published studies of breast cancer illustrate the potential contribution of DNA microarrays to uncover new disease subtypes. In one study, tumours could be classified into a basal epithelial-like group, an ErbB2-overexpressing group and a normal breast-like group<sup>22</sup>. In a later study, survival analyses on a sub-cohort of patients with locally advanced breast cancer showed significantly different outcomes for patients belonging to the various groups, despite uniform treatment<sup>23</sup>. In an independent study of 38 invasive breast cancers,

striking molecular differences between ductal carcinoma specimens were uncovered that led to a suggested new classification for oestrogen-receptor (ER)-negative breast cancer<sup>24</sup>. Similarly, a study of 58 node-negative breast carcinomas discordant for ER status also uncovered a list of genes that discriminated tumours according to ER status<sup>25</sup>. More recently, gene expression profiling was found to be a more powerful predictor of disease outcome in young patients with cancer than clinical- and histological-based classifications<sup>26</sup>.

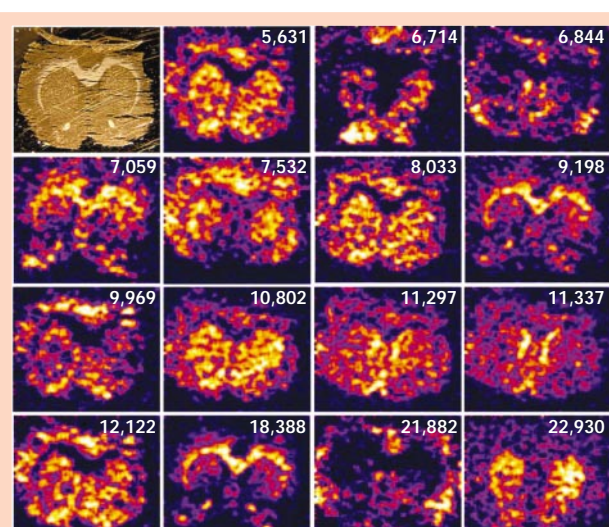
#### DNA versus protein microarrays

The DNA microarray studies described above, as well as numerous others in the literature, indicate the great utility of DNA microarrays for uncovering patterns of gene expression that are clinically informative. An important challenge for microarray analysis of disease tissues and cells is to understand at a mechanistic level the significance of associations observed between subsets of genes and clinical features of disease. Another challenge is to identify the smallest but most informative sets of genes associated with specific clinical features, which then can be interrogated using technologies available in clinical laboratories. Yet another challenge is to determine how well RNA levels of predictive genes correlate with protein levels. A lack of correlation may imply that the predictive property of the gene(s) is independent of gene function. For example, comparisons of messenger RNA and protein levels for the same tumours reported for lung cancer demonstrated that only a small percentage of genes had a statistically significant correlation between the levels of their corresponding proteins and mRNAs<sup>27</sup>.

Technologies for DNA microarray analysis are still evolving. There is a tendency by manufacturers to favour oligonucleotide- over complementary DNA-based microarrays, and progress has been made on adoption of data analysis standards<sup>28</sup>. Nevertheless, however perfected DNA microarrays and their analytical tools become for disease profiling, they will not eliminate a pressing need for other types of profiling technologies that go beyond measuring RNA levels, particularly for disease-related investigations. DNA microarrays have limited utility for the analysis of biological fluids and for uncovering assayable biomarkers directly in the fluid. Numerous alterations may occur in proteins that are not reflected in changes at the RNA level, providing a compelling rationale for direct analysis of gene expression at the protein level. As a result, there is substantial interest in developing microarrays or biochips that allow the systematic analysis of thousands of proteins (see review in this issue by Fields and co-workers, page 208).

Unlike DNA microarrays, which provide one measure of gene expression (namely RNA levels), there is a need to implement protein microarray strategies that address the many different features of proteins that can be altered in disease. These include, on the one hand, determination of their levels in biological samples and, on the other, determination of their selective interactions with other biomolecules, such as other proteins, antibodies, drugs or various small ligands. The compelling need for protein chips has led numerous biotechnology companies to devise new strategies for producing biochips that have utility for biomedical investigations. New classes of capture agents include aptamers (SomaLogic, <http://www.somallogic.com/>), ribozymes (Archemix, <http://www.archemix.com/>), partial-molecule imprints (Aspira Biosystems, <http://www.aspirabio.com>) and modified binding proteins (Phylos, <http://www.phylos.com>). For assays of protein interaction, biochips that contain either peptides or proteins are being produced. Peptides may be synthesized in very large numbers directly on the chip<sup>29</sup>. Alternatively, recombinant proteins may be arrayed and effort is underway to assemble large sets of purified recombinant proteins for microarrays and other applications.

Profiling studies of disease tissue that have used protein microarrays are beginning to emerge. As a model to better understand how patterns of protein expression shape the tissue microenvironment, Knezevic *et al.* analysed protein expression in tissue derived from squamous cell carcinomas of the oral cavity through an antibody

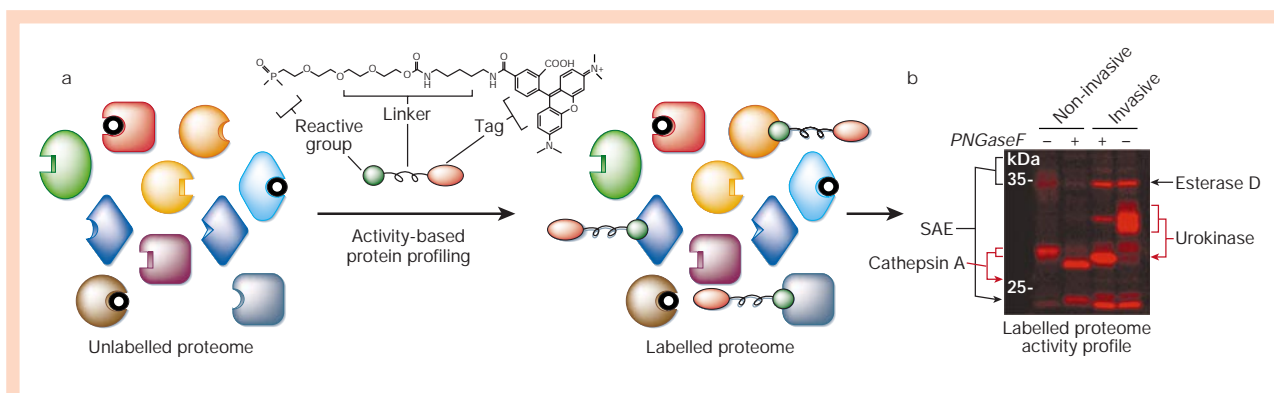


**Figure 2** Imaging mass spectrometry. Transverse sections of rat brain were cut, thaw-mounted on the target plate and coated with matrix<sup>21</sup>. A survey scan was performed first with data acquisition taken randomly across the section to generate an average protein profile. Over 200 individual mass peaks were detected in a mass-to-charge ( $m/z$ ) range of up to 40,000. The figure presents an optical image of the brain section prior to matrix deposition. The section was scanned by acquiring  $74 \times 75$  points with a resolution of  $180 \mu\text{m}$  by averaging spectra produced by 15 laser shots using an automated imaging computer algorithm. In this scan, the intensity of all of the different mass signals was monitored. Fifteen ion-density maps are shown, each obtained for different protein signals; some of these, in particular  $m/z$  6,844, have low intensities. As expected, some proteins were found to be highly specific for a given brain region. This is particularly striking for the density maps of the proteins detected at  $m/z$  5,631 and  $m/z$  18,388, which are almost 'negatives' of each other.

microarray approach for high-throughput proteomic analysis<sup>30</sup>. Using laser-capture microdissection to procure total protein from specific microscopic cellular populations, they showed that quantitative, and potentially qualitative, differences in expression patterns of multiple proteins within epithelial cells correlated reproducibly with oral-cavity tumour progression. Differential expression of multiple proteins was found in stromal cells surrounding and adjacent to regions of diseased epithelium that correlated directly with tumour progression of the epithelium. Most of the proteins identified in both cell types were involved in signal transduction pathways. Knezevic *et al.* hypothesized therefore that extensive molecular communications involving complex cellular signalling between epithelium and stroma play a key role in driving progression of oral-cavity cancer.

A reverse-phase protein array approach that immobilizes the whole repertoire of a tissue's proteins has been developed<sup>31</sup>. A high degree of sensitivity, precision and linearity was achieved, making it possible to quantify the phosphorylated status of signal proteins in subpopulations of human tissue cells. Using this approach, Paweletz *et al.*<sup>31</sup> performed a longitudinal analysis of the state of pro-survival checkpoint proteins at the microscopic transition stage from patient-matched, histologically normal prostate epithelium to prostate intraepithelial neoplasia and to invasive prostate cancer. Cancer progression was associated with increased phosphorylation of the serine/threonine kinase Akt, suppression of apoptosis pathways, and decreased phosphorylation of extracellular signal-regulated kinase (ERK). At the transition from histologically normal epithelium to intraepithelial neoplasia, a statistically significant surge in phosphorylated Akt was observed, together with a concomitant suppression of downstream apoptosis pathways preceding the transition into invasive carcinoma.

A clinically relevant application of protein microarrays is the identification of proteins that induce an antibody response in



**Figure 3** Activity-based protein profiling. This chemical strategy is used to monitor changes in the functional state of enzyme superfamilies directly in complex proteomes<sup>47</sup>. **a**, The method uses chemical probes that comprise three general elements: an active-site-directed reactive group (shown here as an ethoxy fluorophosphonate group that targets the serine hydrolase class of enzymes), a linker (shown here as a polyethylene glycol) and a tag for the visualization of active enzymes (shown here as a rhodamine group). Active

enzymes are denoted by shapes with open active sites; their inactive counterparts have their active sites shaded in black. **b**, Labelled proteome activity profile. Representative in-gel fluorescence analysis of the secreted proteome (labelled with fluorescent pigment) derived from invasive MUM-2B and non-invasive MUM-2C human melanoma cancer cells. Enzyme activities selectively associated with either invasive cells (for example, urokinase) or non-invasive cells (for example, sialic acid 9-*O*-acetyltransferase or SAE) are noted.

autoimmune disorders<sup>32</sup>. Microarrays were produced by attaching several hundred proteins and peptides to the surface of derivatized glass slides. Arrays were incubated with patient serum, and fluorescent labels were used to detect autoantibody binding to specific proteins in autoimmune diseases, including systemic lupus erythematosus and rheumatoid arthritis. Such microarrays represent a powerful tool to study immune responses in a variety of diseases, including cancer.

One of the main challenges in making biochips for global analysis of protein expression is the current lack of comprehensive sets of genome-scale capture agents such as antibodies. Another important consideration in protein microarrays is that proteins undergo numerous post-translational modifications that may be crucial to their functions. But these modifications are generally not captured using either recombinant proteins or antibodies that do not distinctly recognize specific forms of a protein. One approach for comprehensive analysis of proteins in their modified forms is to array proteins isolated directly from cells and tissues following protein fractionation schemes<sup>33</sup>. Fractions that react with specific probes are within the reach of chromatographic and gel-based separation techniques for resolving their individual protein constituents, and of mass spectrometric techniques for identification of their constituent proteins. Protein microarrays of different types are likely to become commercially available for clinically relevant assays of broad sets of proteins and may well rival DNA microarrays for introduction into the clinical laboratory.

### The quest for disease biomarkers using proteomics

There is substantial interest in applying proteomics to the identification of disease markers. Approaches include comparative analysis of protein expression in normal and disease tissues to identify aberrantly expressed proteins that may represent new markers, analysis of secreted proteins in cell lines and primary cultures, and direct serum protein profiling. The potential of mass spectrometry to yield comprehensive profiles of peptides and proteins in biological fluids without the need to first carry out protein separations has attracted interest. In principle, such an approach is highly suited for marker identification because of reduced sample requirements and high throughput.

This approach is currently popularized, particularly for serum analysis, by the technology referred to as surface-enhanced laser desorption/ionization<sup>1</sup>. Microlitre quantities of serum from many samples are applied to the surface of a protein-binding plate, with properties to bind a class of proteins. The bound proteins are treated

and analysed by MALDI. The mass spectra patterns obtained for different samples reflect the protein and peptide contents of these samples. Patterns that distinguish between cancer patients and normal subjects with remarkable accuracy have been reported for several types of cancer<sup>1</sup>. The main drawbacks of direct analysis of tissues or biological fluids by MALDI are the preferential detection of proteins with a lower molecular mass and the difficulty in determining the identity of proteins owing to post-translational modifications obscuring the correspondence of measured and predicted masses. Occasionally the masses observed match precisely the predicted masses of specific proteins. This was the case in a study of proteins secreted by stimulated CD8 T cells, which led to the identification of the small proteins  $\alpha$ -defensin 1, 2 and 3 as contributing to the anti-HIV-1 activity of CD8 antiviral factor<sup>34</sup>.

A productive approach for the identification of cancer markers has been the analysis of serum for autoantibodies against tumour proteins. There is increasing evidence for an immune response to cancer in humans, demonstrated in part by the identification of autoantibodies against a number of intracellular and surface antigens detectable in sera from patients with different cancer types<sup>35</sup>. The identification of panels of tumour antigens that elicit an antibody response may have utility in cancer screening, diagnosis or in establishing prognosis, and in immunotherapy against the disease.

There are several approaches for the detection of tumour antigens that induce an immune response<sup>35</sup>. A number of antigens have been detected by screening expression libraries with patient sera<sup>36–41</sup> or, more recently, by using a random peptide-library approach<sup>42</sup>. Multiple proteins that induce autoantibodies that are specific for different types of cancer have been identified using 2D gels to separate tumour proteins, followed by western blotting and incubation with patient sera<sup>43</sup>. For most antigenic proteins identified using this approach, post-translational modifications contributed to the immune response. In a study of lung cancer, sera from 60% of patients with lung adenocarcinoma and from 33% of patients with squamous-cell lung carcinoma, but from none of the non-cancer controls, exhibited immunoglobulin- $\gamma$ -based reactivity against proteins identified as glycosylated annexins I and II<sup>44</sup>. Microarrays that contain proteins derived from tumour cells have the potential of substantially accelerating the pace of discovery of tumour antigens and yielding a molecular signature for immune responses directed against protein targets in different types of cancer<sup>35</sup>.

The increased emphasis on proteomics for disease investigations is stimulating a reassessment of strategies for sample procurement and preservation to render them compatible with proteomics,

because of the inherent instability of proteins. For example, the manner in which biological fluids such as serum or plasma are collected is not ideally suited for proteomics. There is a need to reduce protein degradation and other forms of modifications that may substantially alter protein content and interfere with global profiling.

### Disease-related functional proteomics

Although data obtained by various expression proteomics strategies have functional relevance by uncovering altered levels or post-translational modification states of proteins in disease, additional technologies are needed for more direct functional analysis. This is exemplified by the need to analyse protein complexes and their disruption in disease, to assay in a high-throughput fashion the activity of various classes of proteins, and to manipulate the levels and activities of individual proteins, in a cellular context, to determine their role in different biological processes and disease states.

Various strategies are currently in use for studies of protein complexes and protein–protein interactions (see review in this issue by Fields and co-workers, page 208). So far, such strategies have been applied to disease investigations on a limited basis, with a relatively narrow focus on particular complexes. For example, using affinity pull-down assays, 2D gels and mass spectrometry, myocardial protein kinase C $\epsilon$  (PKC $\epsilon$ ) was found to be associated physically with at least 36 other proteins with a multitude of functions. Cardioprotection induced by activation of PKC $\epsilon$  was found to be coupled with dynamic modulation and recruitment of PKC $\epsilon$ -associated proteins<sup>45</sup>. Previously unrecognized functions of PKC $\epsilon$  have relevance to heart function and heart disease. Clearly, at the present time, most systematic studies of protein–protein interactions have dealt with normal or physiological states, as the field is still in its early stages and the merits of various technologies, particularly for disease investigations, have yet to be fully appreciated.

Likewise, there has been limited application of activity-based proteomics to disease investigations, although the field itself is still in its early stages<sup>46–48</sup>. Activity-based assays of normal and disease tissue have substantial relevance to the study of disease, but probably will be limited to the analysis of one class of proteins at a time. The potential contribution of this technique to the measurement of global dynamics in protein function is illustrated by the application of a chemical proteomics strategy to quantitatively compare enzyme activities across normal and disease tissue<sup>47</sup> (Fig. 3). A global analysis of the activity, subcellular distribution and glycosylation state for the serine hydrolase superfamily in a panel of human breast and melanoma cell lines resulted in the identification of a cluster of proteases lipases and esterases that distinguished cancer lines based on tissue of origin. Remarkably, the majority of these enzyme activities were downregulated in the most invasive cancer lines examined, which instead upregulated a distinct set of secreted and membrane-associated enzyme activities.

Approaches are being developed or implemented to allow direct protein manipulations other than by decreasing or increasing overall amounts of protein, as can be done by gene manipulations. One advantage of these approaches is the ability to inactivate a specific site in a protein in a time-dependent and localization-restricted manner, as with chromophore-assisted laser inactivation (CALI). Such inactivation allows assessment of the disease relevance of the site and provides a means to revert a cellular phenotype to a more normal state<sup>49</sup> (Fig. 4). Inactivation of protein targets involved in different signal-transduction pathways in cancer using CALI has been demonstrated<sup>49</sup>.

### Contributions of proteomics to studies of pathogens

Despite earlier predictions to the contrary, infectious diseases remain as a leading cause of death worldwide. A complicating factor in therapy for infectious disease is the development of resistance to commonly used drugs (for example, as has occurred in tuberculosis), which heightens the need for developing effective new therapies.

Table 1 Targeted proteomic approaches to microbial pathogens

- Characterization of submicrobial proteomes (for example, secreted proteins, surface proteins and immunogenic proteins)
- Comparative analysis of different strains
- Comparative analysis of different physiological states
- Identification of proteins related to pathogenicity
- Identification of proteins involved in host–pathogen interactions
- Evaluation of mechanisms of action of antimicrobials

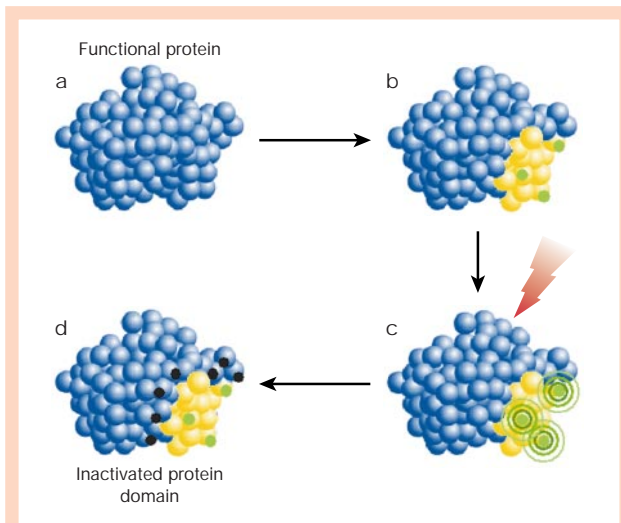
Interest in the application of proteomics to microbiology goes back at least two decades, with the pioneering work of Fred Neidhardt to characterize protein expression patterns in *Escherichia coli* under different growth conditions<sup>50</sup>. The complete sequencing of a number of microbial genomes has provided a framework for identifying proteins encoded in these genomes using mass spectrometry. A case in point is the sequencing of the genome of the malaria parasite *Plasmodium falciparum*, which has provided a basis for conducting comparative proteomics studies of this pathogen, leading to the identification of new potential drug and vaccine targets<sup>51,52</sup>. Aside from comprehensive identification of microbial proteins, proteomics is relevant to numerous aspects of microbial disease pathogenesis and treatment<sup>53–57</sup> (Table 1).

### Contribution of proteomics to drug development

There is currently a burgeoning interest in proteomics on the part of the pharmaceutical industry, evidenced by implementation of proteomics programmes by most major pharmaceutical companies. The notion has been advanced that, as the vast majority of drugs target proteins, proteomics should have substantial utility for drug development. But the industry has so far adopted a cautious attitude, and it is too early to make a critical assessment of the contributions of proteomics to drug development, relative to other approaches. The caution stems from the prior heavy investment in genomics and other approaches and some uncertainty surrounding the adequacy and scalability of proteomics to meet the needs of the pharmaceutical industry. Provided suitable technology platforms become available, the use of proteomics may permeate numerous aspects of drug development, by identifying new targets and facilitating assessment of drug action and toxicity both in the preclinical and clinical phases.

Several published studies illustrate the application of functional proteomics for identification of regulated targets in specific pathways. Lewis *et al.*<sup>58</sup> have combined functional proteomics with selective activation and inhibition of mitogen-activated protein kinase (MAPK) kinase (MKK), in order to identify cellular targets regulated by the MKK/ERK cascade. Twenty-five targets of this signalling pathway were identified, of which only five were previously characterized as MKK/ERK effectors. The remaining targets suggest new roles for this signalling cascade in cellular processes of nuclear transport, nucleotide excision repair, nucleosome assembly, membrane transport and cytoskeletal regulation.

In another study to identify proteases most suitable for drug targeting, an automated microtitre-plate assay was modified to allow detection of the four main classes of proteases in tissue samples (matrix metalloproteases, cathepsins, and the cell serine proteases, tryptase and chymase)<sup>59</sup>. Fifteen sets of colorectal carcinoma biopsies representing primary tumour, adjacent normal colon and liver metastases were screened for protease activity. Matrix metalloproteases were expressed at higher levels in the primary tumour than in adjacent normal tissue. The mast cell proteases, in contrast, were found at very high levels in adjacent normal tissue, but were not detectable in the metastases. Cathepsin B activity was significantly higher in the primary tumour, and highest in the metastases. The proteases detected by activity assays were then localized in biopsy sections by immunohistochemistry. Mast cell proteases were abundant in adjacent normal tissue, because of infiltration of the lamina propria by mast cells. Matrix metalloproteases were localized to the



**Figure 4** Chromophore-assisted laser inactivation. This technology involves the generation of short-lived radicals that induce covalent modifications at spatially restricted sites on a protein<sup>49</sup>. A transient functional inactivation occurs if the radicals modify amino acids of the target protein that have a functional role. A target protein (a) is complexed with a ligand (yellow; for example, antibody, antibody fragments such as scFv and Fab, peptide, nucleic acid aptamers or small molecules) labelled with dye molecules (green; for example, malachite green or fluorescein) (b). The complex is irradiated with laser or incoherent light (red lightning), leading to the generation of reactive species (concentric circles) that travel a short distance (c). The reactive species in turn lead to modifications (black) of nearby amino acids (d). If the modified amino acids are responsible for a function, that particular functional protein domain will be inactivated, leaving the other functional domains intact.

tumour cells themselves, whereas cathepsin B was expressed predominantly by macrophages at the leading edge of invading tumours.

Such activity-based screening provides a basis for selecting targets in the development of inhibitors to specific proteases. Protein biochips potentially could provide a high-throughput platform for target identification. Biochips developed for interaction studies could be important in lead-compound optimization and could accelerate drug development by allowing efficient evaluation of lead compounds for specificity and selectivity in binding to drug targets. Proteomics also may provide increased efficiency of clinical trials through the availability of biologically relevant markers for drug efficacy and safety.

### Organizing proteomics initiatives

It is clear that while some progress has been made in disease proteomics, the field is still in its infancy. Knowledge of the sequence of the human genome has provided a framework for genomic approaches to unravel disease processes. A similar knowledge of the human proteome is currently lacking. Developing a comprehensive knowledge framework of the proteome is considerably more complex than sequencing the human genome. Ideally, such a proteome framework would encompass knowledge of all human proteins, from their sequence to their post-translational modifications, to their interactions among each other, their cellular and sub-cellular distribution, and their temporal pattern of expression. Although such an exhaustive framework will not materialize in the foreseeable future, more modest goals may well be within reach.

To that effect, there is a need to begin an organized effort, the goals of which include developing an infrastructure in proteomics that would substantially facilitate unravelling the complexity of the proteome in health and in disease. The Human Proteome Organisation (HUPO, <http://www.hupo.org>) was founded to regroup scientists in the public and private sectors engaged throughout the world in

various aspects of proteomics. HUPO's mission is threefold: to consolidate national and regional proteome organizations into a worldwide organization; to engage in scientific and educational activities to encourage the spread of proteomics technologies and disseminate knowledge pertaining to the human proteome and that of model organisms; and to assist in the coordination of public proteome initiatives aimed at characterizing specific tissue and cell proteomes. Initiatives currently in the pilot phase include an international effort to identify proteins detectable in normal serum and plasma and their range of variation with age, ethnicity and physiological state, and a liver proteome study to identify proteins expressed in the liver. These initiatives have attracted substantial interest and will be integrated with efforts in protein informatics to achieve data standardization on the one hand, and data curation on the other.

### Concluding remarks

Proteome alterations in disease may occur in many different ways that are not predictable from genomic analysis, and it is clear that a better understanding of these alterations will have a substantial impact in medicine. A useful repertoire of proteomics technologies is currently available for disease-related applications, although further technological innovations would be beneficial to increase sensitivity, reduce sample requirement, increase throughput and more effectively uncover various types of protein alterations such as post-translational modifications. The use of these technologies will likely expand substantially, particularly to meet the need for better diagnostics and to shorten the path for developing effective therapy. □

doi:10.1038/nature01514

- Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C. & Liotta, L. A. Clinical proteomics: translating benchside promise into bedside reality. *Nature Rev. Drug Discov.* **1**, 683–695 (2002).
- Hanash, S. 2-D or not 2-D—is there a future for 2-D gels in proteomics? Insights from York proteomic meeting. *Proteomics* **1**, 635–637 (2001).
- Hanash, S. M., Madoz-Gurpide, J. & Misek, D. E. Identification of novel targets for cancer therapy using expression proteomics. *Leukemia* **16**, 478–485 (2002).
- Van Eyk, J. E. Proteomics: unraveling the complexity of heart disease and striving to change cardiology. *Curr. Opin. Mol. Therapeut.* **3**, 546–553 (2001).
- Li, X. P. *et al.* A two-dimensional gel electrophoresis database of rat heart protein. *Electrophoresis* **20**, 891–897 (1999).
- Evans, G., Wheeler, C. H., Corbett, J. M. & Dunn, M. J. Construction of HSC-2D PAGE: a two-dimensional gel electrophoresis database of heart proteins. *Electrophoresis* **18**, 471–479 (1997).
- van Der Velden, J. *et al.* Effects of calcium, inorganic phosphate, and pH on isometric force in single skinned cardiomyocytes from donor and failing human hearts. *Circulation* **104**, 1140–1146 (2001).
- Arrell, D. K., Neverova, I., Fraser, H., Marbán, E. & Van Eyk, J. E. Proteomic analysis of pharmacologically preconditioned cardiomyocytes reveals novel phosphorylation of myosin light chain 1. *Circ. Res.* **89**, 480–487 (2001).
- Ping, P., Zhang, J., Pierce, W. M. & Bolli, R. Functional proteomic analysis of protein kinase C $\epsilon$  signaling complexes in the normal heart and during cardioprotection. *Circ. Res.* **88**, 59–62 (2001).
- Heinke, M. Y. *et al.* Protein changes observed in pacing-induced heart failure using two-dimensional electrophoresis. *Electrophoresis* **19**, 2021–2030 (1998).
- Westbrook, J. A., Yan, J. X., Wait, R., Welson, S. Y. & Dunn, M. J. Zooming-in on the proteome: very narrow-range immobilized pH gradients reveal more protein species and isoforms. *Electrophoresis* **22**, 2865–2871 (2001).
- Hoving, S. *et al.* Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients. *Proteomics* **2**, 127–134 (2002).
- Langen, H. *et al.* Two-dimensional map of the proteome of *Haemophilus influenzae*. *Electrophoresis* **21**, 411–429 (2000).
- Zuo, X. & Speicher, D. W. Comprehensive analysis of complex proteomes using microscale solution isoelectrofocusing prior to narrow pH range two-dimensional electrophoresis. *Proteomics* **2**, 58–68 (2002).
- Patton, W. F. Detection technologies in proteome analysis. *J. Chromatogr. B* **771**, 3–31 (2002).
- Zhou, G. *et al.* 2D differential in-gel electrophoresis for the identification of esophageal cancer cell cancer-specific protein markers. *Mol. Cell. Proteomics* **1**, 117–124 (2001).
- Gagnon, E. *et al.* Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* **110**, 119–131 (2002).
- Sabarath, N. *et al.* Identification of surface proteins of *Helicobacter pylori* by selective biotinylation, affinity purification, and two-dimensional gel electrophoresis. *J. Biol. Chem.* **277**, 27896–27902 (2002).
- Shin, B. K. *et al.* Global profiling of the cell surface proteome of cancer cells uncovers an abundance of proteins with chaperone function. *J. Biol. Chem.* DOI: 10.1074/jbc.M210455200 (2002).
- Brivio, M. *et al.* Integrated microfluidic system enabling (bio)chemical reactions with on-line MALDI-TOF mass spectrometry. *Anal. Chem.* **74**, 3972–3976 (2002).
- Stoeckli, M., Chaurand, P., Hallahan, D. E. & Caprioli, R. M. Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nature Med.* **7**, 493–496 (2001).

22. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
23. Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
24. Brenton, J. D., Aparicio, S. A. & Caldas, C. Molecular profiling of breast cancer: portraits but not physiognomy. *Breast Cancer Res.* **3**, 77–80 (2001).
25. Gruvberger, S. *et al.* Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res.* **61**, 5979–5984 (2001).
26. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
27. Chen, G. *et al.* Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors. *Clin. Cancer Res.* **8**, 2290–2305 (2002).
28. Opinion. Microarray standards at last. *Nature* **419**, 323 (2002).
29. Pellois, J. P. *et al.* Individually addressable parallel peptide synthesis on microchips. *Nature Biotechnol.* **20**, 922–926 (2002).
30. Knezevic, V. *et al.* Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* **1**, 1271–1278 (2001).
31. Pawelz, C. P. *et al.* Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at cancer invasion front. *Oncogene* **20**, 1981–1989 (2001).
32. Robinson, W. H. *et al.* Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nature Med.* **8**, 295–301 (2002).
33. Madoz-Gurpide, J., Wang, H., Misek, D. E., Brichory, F. & Hanash, S. M. Protein based microarrays: a tool for probing the proteome of cancer cells and tissues. *Proteomics* **1**, 1279–1287 (2001).
34. Zhang, L. *et al.* Contribution of human  $\alpha$ -defensin 1, 2 and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science* **298**, 995–1000 (2002).
35. Hanash, S. Harnessing immunity for cancer marker discovery. *Nature Biotechnol.* **21**, 37–38 (2003).
36. Stockert, E. *et al.* A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. *J. Exp. Med.* **187**, 1349–1354 (1998).
37. Gourevitch, M. M. *et al.* Polymorphic epithelial mucin (MUC-1)-containing circulating immune complexes in carcinoma patients. *Br. J. Cancer* **72**, 934–938 (1995).
38. Gure, A. O. *et al.* Human lung cancer antigens recognized by autologous antibodies: definition of a novel cDNA derived from the tumor suppressor gene locus on chromosome 3p21.3. *Cancer Res.* **58**, 1034–1041 (1998).
39. Yamamoto, A., Shimizu, E., Ogura, T. & Sone, S. Detection of auto-antibodies against L-myc oncogene products in sera from lung cancer patients. *Int. J. Cancer* **22**, 283–289 (1996).
40. Soussi, T. The humoral response to the tumor-suppressor gene product p53 in human cancer: implications for diagnosis and therapy. *Immunol. Today* **17**, 354–356 (1996).
41. Old, L. J. & Chen, Y. T. New paths in human cancer serology. *J. Exp. Med.* **187**, 1163–1167 (1998).
42. Mintz, P. J. *et al.* Fingerprinting the circulating repertoire cancer patients. *Nature Biotechnol.* **21**, 57–63 (2003).
43. Le Naour, F. Contribution of proteomics to tumor immunology. *Proteomics* **1**, 1295–1302 (2001).
44. Brichory, F. M. *et al.* An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc. Natl Acad. Sci. USA* **98**, 9824–9829 (2001).
45. Vondriska, T. M. & Ping, P. Functional proteomics to study protection of the ischaemic myocardium. *Expert Opin. Therapeut. Targets* **6**, 563–570 (2002).
46. Borodovsky, A. *et al.* Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**, 1149–1159 (2002).
47. Jessani, N., Liu, Y., Humphrey, M. & Cravatt, B. F. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc. Natl Acad. Sci. USA* **99**, 10335–10340 (2002).
48. Greenbaum, D. *et al.* Chemical approaches for functionally probing the proteome. *Mol. Cell. Proteomics* **1**, 60–68 (2002).
49. Rubenwolf, S. *et al.* Functional proteomics using chromophore-assisted laser inactivation. *Proteomics* **2**, 241–246 (2002).
50. VanBogelen, R. A., Schiller, E. E., Thomas, R. D. & Neidhardt, F. C. Diagnosis of cellular states of microbial organisms using proteomics. *Electrophoresis* **20**, 2149–2159 (1999).
51. Lasonder, E. *et al.* Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419**, 537–542 (2002).
52. Florens, L. *et al.* A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526 (2002).
53. Nilsson, C. L. Bacterial proteomics and vaccine development. *Am. J. Pharmacogenomics* **2**, 59–65 (2002).
54. Eymann, C., Homuth, G., Scharf, C. & Hecker, M. *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *J. Bacteriol.* **184**, 2500–2520 (2002).
55. Haas, G. *et al.* Immunoproteomics of *Helicobacter pylori* infection and relation to gastric disease. *Proteomics* **2**, 313–324 (2002).
56. Reid, S. D. *et al.* Postgenomic analysis of four novel antigens of group A streptococcus: growth phase-dependent gene transcription and human serologic response. *J. Bacteriol.* **184**, 6316–6324 (2002).
57. Antelmann, H., Yamamoto, H., Sekiguchi, J. & Hecker, M. Stabilization of cell wall proteins in *Bacillus subtilis*: a proteomic approach. *Proteomics* **2**, 591–602 (2002).
58. Lewis, T. S. *et al.* Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. *Mol. Cell* **6**, 1343–1354 (2000).
59. McKerrow, J. H. *et al.* A functional proteomics screen of proteases in colorectal carcinoma. *Mol. Med.* **6**, 450–460 (2000).

**Acknowledgements** I thank P. Chaurand, R. Caprioli, G. Omenn, B. Cravatt, N. Jessani, R. Kuick, L. Ilag, E. Gulari for their insightful comments, stimulating discussions and other contributions to the preparation of this review.