Cancer proteomics: from signaling networks to tumor markers

Richard J. Simpson and Donna S. Dorow

Along with the great strides that have been made towards understanding cancer, has come a realization of the complexity of molecular events that lead to malignancy. Proteomics-based approaches, which enable the quantitative investigation of both cellular protein expression levels and protein–protein interactions involved in signaling networks, promise to define the molecules controlling the processes involved in cancer.

Cancer is a multi-faceted disease that presents many challenges to clinicians and cancer researchers searching for more-effective ways to combat its often devastating effects. Among the central challenges of this disease, are the identification of markers for improved diagnosis and classification of tumors, and the definition of targets for more-effective therapeutic measures. Although some cancer-related genes have been mainly identified by mutational analysis (Fig. 1), at present, tumor classification is a complex process based primarily on site and histological examination. However, tumors with a similar histological appearance can follow significantly different clinical courses and show different responses to therapy. Given the wide diversity of tumors, even those derived from the same tissue, additional methods of classification are urgently required. Furthermore, it is now clear that the genetic make-up of both the tumor and the individual patient can influence the outcome of a given treatment. Therefore, to be most effective, future treatments will need to be tailored not only for the specific tumor type but also, in some cases, for the individual as well.

Proteomics approaches to tumor marker identification hold the promise of identifying specific protein modifications in tumor tissues to assist in individualizing treatments for certain cancers. With the completion of the draft sequence of the human genome, there is a great deal of interest in the use of functional genomics, especially gene expression profiling techniques such as DNA microarrays and proteomics, to identify cancer-associated genes and their protein products. These two complementary technologies permit the analysis of thousands of genes or proteins simultaneously, and have the potential to identify markers for early detection, classification and prognosis of tumors, as well as pinpointing targets for improved treatment outcomes. Here, we aim to explore key themes in proteomics and their application to the study of cancer.

Cell mapping and global expression profiling
Currently, proteomics research embraces two contrasting but complementary strategies. The first strategy, known as cell-mapping proteomics, aims to define protein-protein interactions to build a picture of the complex networks that constitute intracellular signaling pathways. The second strategy, protein expression proteomics, monitors global expression of large numbers of proteins within a cell type or tissue, and quantitatively identifies how patterns of expression change in different circumstances.

Cell-mapping proteomics
Many genetic mutations associated with cancer progression affect genes encoding proteins in signaling pathways, highlighting the importance of defining signaling networks in tumor formation. The cell-mapping approach can be used in a wide variety of ways to answer basic questions about functional regulation in tumor cells. For example, Pandey and colleagues treated HeLa cells with either epidermal growth factor (EGF) or platelet-derived growth factor (PDGF), and used anti-phosphotyrosine immunoprecipitation to concentrate a range of proteins that were subsequently phosphorylated. The analysis revealed...
the role of Vav-2, as well as a number of other proteins, in growth factor signaling in these cells. Alternatively, Lewis et al. selectively activated or inhibited the mitogen-activated protein kinase (MAPK) pathway and used a proteomic approach to identify twenty novel targets of MAPK signaling. Affinity capture techniques have also been used to identify the anti-apoptotic protein DIABLO/SMAC (Ref. 8), binding proteins for suppressors of cytokine signaling (SOCS), MAPK pathway targets, and components of the PDGF receptor signaling pathway. Although the scope of this review does not cover the full range of possibilities available, these examples show the promise of cell-mapping proteomics to define the signaling molecules that regulate cellular growth.

A recent estimate based on the first draft sequence of the human genome2-3, suggests that the full complement of human genes is ~32 000; significantly lower than previous estimates, and much closer to the size of Caenorhabditis elegans (19 000 genes) and Drosophila melanogaster (13 600 genes) genomes. It is therefore evident that the functional evolution of proteins in higher eukaryotes results more from combinatorial diversification of regulatory networks than a substantial increase in gene number. Recent protein interaction studies in Saccharomyces cerevisae have revealed the architecture of interconnections in signaling pathways of a simple eukaryotic organism. This analysis revealed that many yeast signaling proteins interact with a small number of interconnecting partners, while a limited number of key proteins form ‘nodes’, connecting to a larger array of proteins in different pathways11. This complex connection scenario probably also holds true for signaling proteins in cells of higher eukaryotic organisms. For example, the p53 tumor-suppressor is one of the most highly connected proteins in human cells and p53 mutations can have severe consequences on a range of basic cellular functions12. One would therefore predict that removal of a highly connected node by a drug that affects the function of a key protein in a complex network, might be somewhat more disruptive to cellular function than targeting a protein that functions in a single pathway. Hence, understanding the connectivity of proteins within cellular signaling networks13 has important implications for selection of targets for anti-cancer therapies.

Proteomics versus two-hybrid analysis for studying protein interactions

The yeast two-hybrid (Y2H) system has been one of the most widely used and successful techniques to investigate protein-protein interactions on a genome-wide scale. Although this method claims considerable success in identifying protein interactions, it does have limitations: (1) the Y2H system is unsuitable for bait proteins, such as transcription factors, which can themselves activate the reporter system; and (2) the environment of the yeast cell might lack key elements required for the specific interaction under study, such as co-factors, members of certain multi-protein complexes, or proteins required for specific post-translational modifications (e.g. phosphorylation, methylation and glycosylation). These limitations do not apply to cell-mapping proteomics as the technique uses the environment of an isolated target cell and network of interest.

Protein-expression analysis

Global protein profiles can be produced for normal compared with tumor cells in a given tissue, or for cells before and after treatment with a specific drug. Currently, this is the most widely used model of proteomics and is largely dependent upon 2D gel electrophoresis (2DE) for visualization of protein profiles. Expression proteomics is the protein equivalent of DNA microarray analysis that define global patterns of RNA expression under various conditions.
conditions. Like DNA microarrays, it has the advantage of being non-prejudicial and could define unexpected ways in which known proteins regulate cellular responses. A major limitation of the 2DE system is the inability to detect proteins of medium to low abundance, and this has prompted much interest in non-2DE approaches for studying global protein profiles.

Why proteomics in addition to genomics?
In the search for tumor progression markers or anti-cancer drug targets, there has been a concerted effort to define gene expression profiles at the transcript level\(^14,15\). However, it is clear that mRNA expression data alone are insufficient to predict functional outcomes for the cell. For instance, mRNA expression data provide very little information about activation state, post-translational modification or localization of corresponding proteins. Moreover, there are numerous reports highlighting the disparity between mRNA transcript and protein expression levels\(^14\). Thus, at the very least, mRNA expression studies must be supported with proteomic information to provide a complete picture of how cells are altered during malignant transformation.

Enabling technologies
Sample preparation
Rigorous sample preparation is a crucial determinant of the success or failure of mRNA and protein expression profiling of tumor tissue or cells. Factors such as time interval between surgical removal and sample processing, and, in the case of fixed tissues, the type of fixative and embedding medium, the length of fixation time, and temperature of tissue processing, all impact significantly on the integrity of nucleic acids and proteins. Each step, from collection of tissues during biopsy or surgery, through to isolation and lysis of target cell populations and protein solubilization, must be carefully considered. In some instances, investigators have analyzed non-necrotic tumor tissues obtained within 30 min of surgical resection without further isolation of tumor cells\(^17,18\). However, unless the tissue is checked and sliced immediately by the pathologist, it is typically snap-frozen in liquid nitrogen within 30 min of surgical removal to avoid protein degradation. To ensure that experimental findings relate directly to changes in the tumor itself, rather than surrounding connective tissue, stroma or vasculature, it is generally desirable that cancer cells are rendered free of contaminating serum proteins, hemopoietic cells, stroma and necrotic tissue, before lysis. Depending upon the tumor type, target cells can be isolated by mechanical methods, such as fine needle aspiration, surface scraping or dissection and mincing of tumor tissue\(^19\). These approaches have been validated by the ability to detect predicted changes in expression of proliferation-associated proteins in breast tumors of different histopathological types and grades\(^20,21\).

Non-enzymatic methods, such as calcium starvation (to disrupt intercellular adhesion), have also been employed to release target cells from tissue sections. For example, immunomagnetic isolation techniques were used to purify normal human luminal and myoepithelial...
cells from calcium-starved primary cultures of tissues from reduction mammoplasties. This procedure yielded cell purity equivalent to that of fluorescent-activated cell sorting (FACS) and with sufficient quantity for proteomic characterization. In another series of studies, colon crypts were isolated from normal human and mouse colon mucosa and adenomatous polyps. Further purification of epithelial cells from the crypts has been accomplished using immunomagnetic beads coated with CAM 5.2 anti-cytokeratin antibodies. One of the most promising recent techniques for reliable isolation of tumor cells is laser capture microdissection (LCM), which allows accurate dissection of specific cells or cell populations from tissue sections under direct microscopic visualization. Cells recovered by LCM are suitable for a variety of molecular analysis methods, including high-throughput gene expression and proteomics technologies.

**Protein separation: 2D or not 2D?**

Traditionally, proteome profiling of complex mixtures, such as whole-cell lysates or enriched subcellular fractions, has involved the combination of 2D protein profiling with mass spectrometry (MS) for protein identification. 2D separates proteins by charge in the first dimension and molecular weight in the second, enabling visualization of a substantial array of proteins in a single experiment. Approximately 10^8 cells are required to analyze proteins of medium abundance using modern MS identification methods. However, owing to the complexity of whole cells or tissues, a single separation strategy such as 2D will probably reveal only a ‘thin slice’ or ‘window’ of the total proteome. The large dynamic range of protein expression levels, and the problems associated with solubility of various classes of proteins, can further reduce the window of proteins that can be recovered for identification. Nevertheless, advances in 2D technology, such as improved immobilized pH gradient strips (IPGs) for enhanced reproducibility, wide-range IPGs for separations up to pH 12, narrow-range IPGs to amplify regions of 1.0–1.5 pH units, and increased sample loading capacity, have helped 2D to remain the most popular technique for proteome analysis. Significant improvements in fluorescent and non-fluorescent visualization procedures have further increased 2D detection sensitivity, linear range, and compatibility with MS-based protein identification methods.

2D can also be used in conjunction with immunological detection techniques to identify a range of proteins for which suitable antibodies are available. However, conventional 2D is unsuitable for use with monoclonal antibodies that recognize conformational epitopes. In the case of the A33 monoclonal antibody, which selectively localizes to metastatic lesions in the colon, immunoblotting of non-reducing/ non-denaturing precast 2DE gels was required to visualize the A33 antigen, leading to its purification and biochemical characterization. Coupling 2D with methods for enrichment of specific groups of cellular proteins by fractionation of subcellular organelles, can be very useful to reduce complexity and increase identification power. This approach has led to the establishment of proteome databases for human placental mitochondria, lysosomes, and colonic epithelia membranes. As membrane proteins are not well recovered from 2DE gels, recent efforts have been directed towards optimization of solubilization and analysis procedures for this class of proteins.

**Non-2DE proteomic approaches**

Because of the inherent limitations of 2D methodologies, in particular, low sample load capacity and poor detection and recovery of particular classes of proteins, further separation techniques must be developed to enable a more comprehensive description of tumor cell proteomes. In this pursuit, Chong and colleagues describe a non-porous reversed-phase high-performance liquid chromatography (RP-HPLC) approach to monitor changes in protein expression as a function of neoplastic progression. They reported the on-line electrospray ionization (ESI)-MS identification of, typically, 70–80 proteins in mass range from 5 to 90 kDa from 35 µg of total cellular protein. The molecular weight profiles of these proteins can be displayed as a mass map or ‘virtual 1D gel’. A number of proteins identified by this method in lysates of pre-malignant and malignant breast cell lines were reported to be oncoproteins associated with breast or colon cancer progression. However, these designations were based on molecular mass alone and, in our experience, confirmatory MS-based sequencing of generated peptides is required for unambiguous identification.

Taking another approach, a 2D liquid-phase separation method was used to map cellular proteins from human erythroleukemia cell lysates with identification by matrix-assisted laser desorption–ionization time-of-flight (MALDI-TOF)-MS (Ref. 47). In this procedure, proteins were separated by liquid-phase isoelectric focusing (IEF) in the first dimension, and then by hydrophobicity using non-porous RP-HPLC in the second dimension. The authors also describe software that displays the data in a pattern similar to that of a 2DE gel image. Other recently described IEF fractionation techniques that have been applied to proteome analyses,
Figure 3. Protein identification and characterization by mass spectrometry

Using global protein expression approaches, proteins are separated by 2DE and (following visualization by Coomassie brilliant blue) or by fluorescence staining or silver staining. Proteins of interest are then excised, alkylated with the isotope-coded affinity tag (ICAT) reagent and subjected to in-gel proteolysis using trypsin. For cell-mapping proteomics or protein-interaction studies, proteins of interest are obtained by: (1) classical ‘pull-down’ experiments using an antibody directed towards the protein of interest; (2) tagging a gene of interest with a small-peptide epitope that can be used to affinity purify the protein following transfection of the tagged gene into cells; or (3) isolation of macromolecular complexes. Separation of protein complexes is carried out by either 1DE or, in the case of immunoaffinity approaches, by 2DE. Visualized proteins of interest are processed using the same method as described for global profiling. A hierarchical mass spectrometry (MS) approach is outlined, where tryptic digests are first analyzed in high-throughput using matrix-assisted laser desorption–ionization (MALDI-quadrapole-time of flight (qTOF) instrumentation. This approach should yield many protein identifications by interrogation of the databases using peptide ion masses and peptide fragment ion masses. For proteins that remain unidentified by this approach, tryptic digests are subjected to low-throughput ESI methods.
include the liquid-based IEF procedure of free flow electrophoresis (FFE)\(^48\), and the multicompartment electrolyzer that operates with IEF membranes\(^49\). Unlike the gel-based IPG strips used in 2DE methods, both FFE and the multicompartment electrolyzer are not restricted by sample load and are thus more amenable to the study of low-abundance proteins.

**Protein identification and characterization**

MS has been a significant enabling technology in the burgeoning field of proteomics, especially since the introduction of two soft ionization methods - MALDI and ESI - for MS analysis of proteins and peptides (Fig 3). Development of instruments with low femt mole (fmol) sensitivity and sub-10 ppm mass-resolution has helped to make MS the method of choice for peptide sequence analysis\(^50\). Public availability of ever increasing amounts of protein and nucleotide sequence information, coupled with computer software tools for on-line interrogation of sequence databases with either peptide or peptide-fragment ion masses\(^51\), has further increased the power of MS (Refs 50,52) in identifying proteins and their post-translational modifications.

**Using proteomics to study cancer**

Although claims regarding the potential of proteomics to define cancer-related molecules might outnumber reports of concrete achievement, both global-expression proteomics\(^53,54\) and, in particular, cell-mapping proteomics, have contributed to significant advances in understanding cancer\(^7-9\). Although it is impossible to cover each contribution here, a brief selection of reports in which a global expression proteomics approach has been applied to the study of cancer is given in Table 1.

**Table 1. Cancer biomarker studies using 2D electrophoresis\(^a\)**

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Specimen and/or cell line</th>
<th>URL</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bladder</td>
<td>Fresh tumors, random biopsies, cystectomies</td>
<td><a href="http://biobase.dk/cgi-bin/celis">http://biobase.dk/cgi-bin/celis</a></td>
<td>53,63,64</td>
</tr>
<tr>
<td>Colorectal</td>
<td>Colonic crypts, carcinoma cell lines LIM 1215 and LIM 1863</td>
<td><a href="http://www.ludwig.edu.au/jpsl/jpslhome.html">http://www.ludwig.edu.au/jpsl/jpslhome.html</a></td>
<td>23,36,65</td>
</tr>
<tr>
<td></td>
<td>Colorectal carcinoma, fresh tumors</td>
<td><a href="http://www.pmfa.pmfhk.cz/2d/2d.html">http://www.pmfa.pmfhk.cz/2d/2d.html</a></td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Colonic polyps</td>
<td>n.a.</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Colon carcinoma cell line HT 29</td>
<td>n.a.</td>
<td>66</td>
</tr>
<tr>
<td>Breast</td>
<td>Normal luminal and myoepithelial breast cells from reduction mammoplasties</td>
<td>n.a.</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Carcinoma cell line MDA-MB 231</td>
<td><a href="http://www.ludwig.edu.au/jpsl/jpslhome.html">http://www.ludwig.edu.au/jpsl/jpslhome.html</a></td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Breast ductal carcinoma and histologically normal tissue</td>
<td><a href="http://www.bio-mol.unisi.it/2d/2d.html">http://www.bio-mol.unisi.it/2d/2d.html</a></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Breast epithelial cells (HMEC, MB 231), adenocarcinoma, carcinoma</td>
<td><a href="http://www.anl.gov/BIO/PMG/projects/index_hbreast.html">http://www.anl.gov/BIO/PMG/projects/index_hbreast.html</a></td>
<td>68,69</td>
</tr>
<tr>
<td>Liver</td>
<td>Hepatocellular carcinoma (HCC) cell line HCC-M</td>
<td>n.a.</td>
<td>70</td>
</tr>
<tr>
<td>Leukemia</td>
<td>Jurkat T-cell line</td>
<td><a href="http://www.mpiib-berlin.mpg.de/2D-PAGE/">http://www.mpiib-berlin.mpg.de/2D-PAGE/</a></td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Lymphoproliferative disorders and leukemia</td>
<td>n.a.</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>B-cell chronic lymphocytic leukemia (B-CLL)</td>
<td>n.a.</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Lymphoma</td>
<td><a href="http://au.expasy.ch/ch2d/">http://au.expasy.ch/ch2d/</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Burkitt lymphoma cell line DG 75</td>
<td>n.a.</td>
<td>74</td>
</tr>
<tr>
<td>Prostate</td>
<td>Benign prostatic tissues and prostate carcinomas</td>
<td>n.a.</td>
<td>75</td>
</tr>
<tr>
<td>Ovarian</td>
<td>Fresh tumors</td>
<td>n.a.</td>
<td>76,77</td>
</tr>
<tr>
<td>Kidney</td>
<td>Renal cell carcinoma</td>
<td>n.a.</td>
<td>78</td>
</tr>
<tr>
<td>Melanoma</td>
<td>Melanoma cells (MeWo)</td>
<td>n.a.</td>
<td>79</td>
</tr>
<tr>
<td>Pancreatic</td>
<td>Adenocarcinoma of cell line EPP85-181P</td>
<td>n.a.</td>
<td>80</td>
</tr>
</tbody>
</table>

\(^a\)Abbreviation: n.a., not available.
The future
With the human genome in a first sequence draft, the focus of biologists will, undoubtedly, turn more towards definition of gene function and the roles of specific proteins or sets of proteins in a given biological context. The proteome of a cell is dynamic, reflecting the changing conditions during growth and differentiation, or in response to a specific disease process. Hence, information pertaining to spatial and temporal characteristics of a protein within a given cell, as well as its expression level and interaction or networking partners, will be of the utmost importance. In this regard, the more-targeted cell-mapping proteomics experiments will be an important adjunct to global protein-expression proteomics information. Reagents for high-throughput proteome quantitation that employ isotope-coded affinity tags (ICAT) are now commercially available. However, basing conclusions solely on changes in abundance overlooks the fact that numerous vital activities of proteins are modulated by post-translational modifications, such as phosphorylation. Indeed, it is estimated that as many as one-third of all cellular proteins are phosphorylated at some point. Therefore, three recently described chemical protocols for the rapid isolation and quantitative analysis of protein phosphorylation promise to expand the scope of functional proteomics.

In addition to expression profiling and protein-interaction mapping, many complementary technologies are currently being developed that will enhance the armory of available functional proteomics methods. These include protein arrays, phage-display antibody arrays, surface-enhanced laser desorption and ionization-based protein chips, and methods for the high-throughput profiling of families of proteins, such as proteases. Over the next few years, the integration of these proteomics-based approaches with data from genomics initiatives promises major advances in cancer research and translation into clinical practice.

Acknowledgements
We would like to thank James Eddes and Adam Cole for figure preparation.

References
17 Stulik, J. et al. (1999) Protein abundance alterations in matched sets of macroscopically normal colon mucosa and colorectal carcinoma. Electrophoresis 20, 3638-3646
19 Franzen, B. et al. (1993) Nonenzymatic extraction of cells from clinical tumor material for analysis of gene expression by 2-dimensional polyacrylamide-gel electrophoresis. Electrophoresis 14, 1045-1053
23 Ji, H. et al. (1994) 2-Dimensional electrophoretic analysis of proteins expressed by normal and cancerous human crypts - application of mass spectrometry to peptide-mass fingerprinting. Electrophoresis 15, 391–405
30 Banks, R.E. et al. (1999) The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis - preliminary findings. Electrophoresis 20, 689–700
44 Santoni, V. et al. (2000) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. Electrophoresis 21, 3329–3344
45 Santoni, V. et al. (1999) Towards the recovery of hydrophobic proteins on two-dimensional electrophoresis gels. Electrophoresis 20, 705–711
48 Hoffmann, P. et al. (2001) Continuous free-flow electrophoresis separation of cytosolic proteins from the human colon carcinoma cell line H1215: a non two-dimensional gel electrophoresis-based proteome analysis strategy. Proteomics 1, 807–818
49 Righetti, P.G. et al. (2001) Prefractionation techniques in proteome analysis Anal. Chem. 73, 320A–326A
54 Celis, J.E. et al. (2000) Bladder squamous cell carcinoma biomarkers derived from proteomics. Electrophoresis 21, 2115–2121
80 Sinha, P. et al. (1999) Increased expression of epidermal fatty acid binding protein, cofilin, and 14-3-3-sigma (stratifin) detected by two-dimensional gel electrophoresis, mass spectrometry and microsequencing of drug-resistant human adenocarcinoma of the pancreas. Electrophoresis 20, 2952–2960

Sciencejobs.com is a new website brought to you by BioMedNet, Cell, and New Scientist.
Search over one thousand positions by sector, discipline and geographic region. Register for the free e-mail alerts that update you about the latest jobs in your field.
All positions previously available on BioMedNet's Job Exchange are now available through http://www.sciencejobs.com