Monitoring the protein expression pattern in tumour cells by proteomics technologies offers opportunities to discover potentially new biomarkers for the early detection and diagnosis of cancer. Different proteomic tools such as 2D-PAGE, 2D-DIGE, SELDI-ToF-MS technology, protein arrays, ICAT, iTRAQ and MudPIT have been used for differential analysis of various biological samples, including cell lysates, cell secretome (conditioned medium), serum, plasma, tumour tissue and nipple aspirate fluid, to better understand the molecular basis of cancer pathogenesis and the validation and characterisation of disease-associated proteins. In recent years, there has been a large increase in cancer-related publications dealing with new biomarker discovery for cancer, therefore, in this review we have focused on the contribution of proteomics technologies in serum and conditioned medium-based oncology research particularly for lung, breast, melanoma and pancreatic cancer.

Although many effective therapies are present for early detection and diagnosis, cancer remains a major cause of death and disease. Cancer is a complex disease that reflects the genetic, as well as protein changes within a cell. Gene expression data gives us limited relevant information since proteins are the main functional units performing all biological process in the cell or organism and may have post-transcriptional event(s) and post-translational modification(s) that contribute to the biological activity of proteins. Protein expression patterns are also changed specifically and significantly in response to every disease (1). The first protein cancer marker, carcinoembryonic antigen (CEA), was identified in 1965 in patient serum for the detection of colorectal cancer (2). Other biomarkers discovered in the 1970s and 1980s include prostate-specific antigen (PSA) for prostate cancer, CA-19 for colorectal and pancreatic cancer, CA-15-3 for breast cancer and CA-125 for ovarian cancer. However, not all biomarkers are effective in all clinical situations. For example, PSA is well established in clinical practice, but approximately one third of patients with an elevated PSA level often undergo unnecessary medical procedures because they do not have a malignant form of prostate cancer (3). Many types of cancer, such as lung carcinoma and melanoma do not have any significant biomarkers available to screen at the early stage of disease. Identification of new tumour biomarkers with predictive value is necessary to allow early detection and treatment of cancer.

Proteomics Technologies

With the recent developments in electrophoresis, imaging, protein labelling, protein array-based approaches and mass spectrometric technologies, along with developments in genomic and protein bioinformatics, proteomics may provide powerful information for improved biomarker discovery. Several proteomics technologies including two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), surface enhanced laser desorption/ionisation time of flight (SELDI-ToF), protein arrays, isotope coded affinity tags (ICAT), iTRAQ and multidimensional protein identification technology (MudPIT) are the approaches being implemented in cancer research. 2D-PAGE and SELDI-ToF are the main technologies used in serum cancer research, however other technologies such as protein arrays, ICAT, iTRAQ and MudPIT also offer great potential for future biomarker discovery in cancer.

Two-dimensional electrophoresis. 2D-PAGE is the most widely used proteomics technique to study the proteome as well as cancer biomarkers (4-8). 2D-PAGE remains...
challenging mainly because of its low sensitivity and reproducibility. Modified 2D electrophoresis by fluorescent tagging to proteins, differential gel electrophoresis (DIGE), offers increased throughput, ease of use, reproducibility, and accurate quantitation of protein expression differences (9). This system enables the separation of two or three fluorescently labelled protein samples (Cy2, Cy3 and Cy5) on the same gel. Differential analysis software identifies the differentially expressed protein targets that can be trypsin-digested and readily identified using mass spectrometry by generating peptide mass fingerprints (PMF) using matrix assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-ToF MS), a technique that is both relatively easy to use and reasonably sensitive for identifying proteins. Additionally, other mass spectrometry techniques such as electrospray ionization (ESI-MS/MS) are capable of providing amino acid sequence information on peptide fragments of the parent protein (10).

Although 2D-PAGE-based techniques have a reasonable level of throughput, there are a number of difficulties inherent to the technique such as separation of low abundant proteins as it is difficult to enrich for these proteins. Membrane proteins are also difficult to separate due to poor solubility. Efforts have been made to overcome these limitations. For example, low abundant proteins can be identified using higher protein concentrations, and applying fractionation methods (7). Moreover, membrane proteins can be identified to some extent by using commercially available mild detergents such as oligooxyethylene, sulfobetaine, dodecyl maltoside or decaethylene glycol mono hexadecyl, as use of strong detergents like SDS interfere with the isoelectric focusing of proteins (11). Additional problems with 2D-electrophoresis include insufficient resolution to separate multiple species originating from a single protein with post-translational modifications, such as those with carbohydrates, difficulties in detecting proteins with molecular masses <120 kDa and those with pI values <4 or >9, low visualisation of less-abundant proteins, and co-migration of proteins to the same spots (12). Conventional 2D-electrophoresis shows only protein expression and cannot detect protein-protein interactions or protein function without using particular methods such as affinity electrophoresis (13).

**SELDI-ToF MS.** This technique allows proteins/peptides to be profiled from different biological samples on a variety of chemically (e.g. anionic, cationic, hydrophobic, hydrophilic, metal affinity capture) or biochemically (e.g. immobilized antibody, receptor, DNA, enzyme) defined chromatographic surfaces. A small amount of sample of interest is loaded onto ProteinChip™ arrays that selectively bind different subsets of proteins in crude samples by adsorption, partition, electrostatic interaction or affinity chromatography according to their surface chemistries. After a short incubation period, unbound proteins and unspecific substances are washed away with an appropriate buffer and water. The ToF reader records the time-of-flight and calculates the accurate molecular weight of proteins/peptides in the form of a spectral map containing mass to charge ratios (m/z) and intensities corresponding to each bound protein/peptide. Biomarker Wizard software analyses the spectral map and detects differentially expressed protein/peptides with statistical significance. An increase in published research using SELDI-ToF in the past few years itself has demonstrated its potential for the early detection of cancer, especially for low molecular weight cancer-associated proteins. For example, applications of SELDI-ToF have been demonstrated for the early detection of prostrate (14, 15), breast (16, 17), bladder (18) ovarian (19, 20), pancreatic (21), and lung (22) cancer biomarkers. However, there is some controversy over this technology such as its reproducibility, the bioinformatics used, the possibility of over-fitting, the potential bias in the samples, as well as how this could possibly fit into a routine diagnostic lab (23-25).

**Protein array technology.** Protein arrays are being used for drug discovery, biomarker identification and molecular profiling of cellular material (26-29). Protein arrays are generated by spotting antibodies (26) or other affinity reagents, such as aptamers, purified proteins (30), peptides (28) or fractionated proteins (27), onto some sort of matrix either on flat solid phases or in capillary systems to generate protein arrays. The sample is applied to the array to allow specific binding of the sample to the array and then the arrays are washed to remove the unbound fraction. The process can also be reversed whereby the protein samples of interest are spotted onto the matrix and then probed with different affinity reagents (31). The wider application of protein arrays in biomedical research is still limited, partly because of the cost of producing antibodies and the limited availability of antibodies with high specificity and high affinity for the target. Additionally, the difficulties associated with preserving proteins in their biologically active conformation before analysis with protein arrays further limits the application of this technology as a routine proteomic strategy.

**Isobaric Tag Labelling Technology**

**ICAT.** Isotope-coded affinity tags (ICAT) use stable isotope labelling to perform quantitative analysis of paired protein samples. They contain a protein-reactive group, an ethylene glycol linker and a biotin tag (32). Two different isotope tags are generated by using linkers that contain either eight hydrogen atoms (d0, light reagent) or eight deuterium atoms (d8, heavy reagent) which bind covalently to cysteine moieties of amino acid within protein(s). Both samples are
mixed, digested with trypsin, fractionated by avidin affinity chromatography and then these differentially tagged peptides are scanned in a mass spectrometer. Spectral peak analysis in single mass spectrometric (MS) mode of the isotopically resolved peptides from the two different sources enables quantitation of the relative amounts of the peptide and hence the protein levels. Differentially expressed proteins are then identified by tandem MS (MS/MS) sequencing. One weakness of ICAT is that only cysteine-containing peptides can be labelled. Approximately 10% of proteins do not have cysteine, therefore they will not be detected by ICAT.

iTRAQ. A similar isobaric labelling technology to ICAT, called iTRAQ, that labels amine residues in peptides has been recently developed (33). iTRAQ contains a set of four isobaric reagents and therefore can analyze up to four protein samples at one time. After trypsin digestion, samples are labelled with four independent iTRAQ reagents. The reporter groups of the iTRAQ reagents will split from the peptide and generate small fragments for each sample with mass/charge (m/z) of 114, 115, 116, and 117. The intensity of each of these peaks represents the quantity of small reporter group fragments and thus represents the quantity of a peptide sample. Peaks in the spectrum graph are used to identify peptide sequences and therefore protein sequences. By comparing the amounts of peptides labelled with each iTRAQ reagent, quantitative differences can be readily measured. A comparative analysis of iTRAQ and ICAT suggests that the information generated by the two methods is complementary. Both methods have some advantages and disadvantages over the other. Unlike iTRAQ, ICAT is preferred for low abundant proteins including signalling molecules, however overlapping peaks in the MS spectrum can compromise the quality of results. On the other hand, apart from non-specific nature of labelling, iTRAQ requires lengthy sample processing separately that increases the chances of experimental variation (34).

MudPIT. MudPIT is a non-gel approach that uses multidimensional high-pressure liquid chromatography (LC/LC) separation, tandem mass spectrometry and database searching (35). MudPIT permits a rapid and simultaneous separation and identification of proteins and peptides in a complex mixture without the need for pre- or post-separation labelling, which is not possible in 2D-electrophoresis, ICAT or iTRAQ (36). The complex protein mixture is digested with a specific protease, then peptide fragments are allowed to separate in parallel with two-dimensional liquid chromatography using a strong cation exchange (SCX) column that separate peptides based on charge, and then by a reverse phase (RP) column based on hydrophobicity. Eluted peptides are identified by tandem mass spectrometry and the technique is extremely sensitive and reproducible. One of the major weaknesses of MudPIT is in identifying quantitative differences in protein expression across protein mixtures (37). In a comparative study with 2D-electrophoresis, MudPIT has been reported to demonstrate superior detection efficiency (36).

Implementation of Proteomics Technologies in Serum Cancer Research

Lung cancer. Lung cancer is clinically divided into two major histological types, non-small cell (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for 75-85% of lung cancer patients and consists of several subtypes, predominantly squamous cell carcinomas, adenocarcinomas and large cell carcinomas, which are treated in the same manner. Small cell lung cancer accounts for 15-25% of lung cancer patients, often has neuroendocrine components and is primarily treated with chemotherapy and/or radiotherapy. Many lung cancers are composed of histologically mixed tumour types consisting of non-small cell and small cell components. Dyspnoea, cough and thoracic pain are early signs of lung cancer, however they are not specific to all cases, while haemoptysis often indicates advanced disease. There is no satisfactory marker for the early detection of lung cancer. However, many widely used biomarkers are available for differential diagnosis and sub typing, i.e. CEA, CYFRA 21-1, SCC, NSE and ProGRP (38-41), but offer low specificity and/or sensitivity.

Various putative biomarkers from serum and conditioned medium have been identified using proteomics as a searching tool. Dihydrodiol dehydrogenase (DDH) was shown to be secreted by the adenocarcinoma cell line, A549, by performing 2D electrophoresis on conditioned medium followed by mass spectrometry (42). Following this observation, the levels of DDH mRNA and protein expression were found to be significantly higher in 15 NSCLC cancer tissues and protein levels were also significantly increased in the serum of the NSCLC patients compared to non-malignant lung tumour and healthy controls. Since mRNA expression of DDH was also reported to be elevated in NSCLC tumour specimens, this raises the possibility of using DDH as a tissue marker and a novel serological marker for NSCLC detection (42, 43).

Protein gene product 9.5 (PGP 9.5) is a neurospecific polypeptide and was proposed as a marker for non-small cell lung cancer, based on its expression in tumour tissue (44, 45). PGP 9.5 and autoantibodies against PGP 9.5 were observed in sera from lung cancer patients through an antibody-based reactivity assay against lung adenocarcinoma proteins, resolved by 2D-PAGE (46). However PGP 9.5 is not limited to lung cancer, as it has also been reported in
pancreatic cancer (47). PGP 9.5 was also shown to be present at the cell surface, as well as secreted in conditioned medium by A549 (46).

In another study, approximately 250 and 100 different proteins were detected in HSA- and IgG-depleted serum samples from lung cancer patients using 2D-LCMS/MS and 1D-LCMS/MS respectively (48). Similarly, the serum samples from lung cancer patients were analyzed with 2D-PAGE after depletion of highly abundant plasma proteins by immunoaffinity chromatography and fractionation by anion-exchange chromatography; 58 gene products, including the classic plasma proteins and the tissue-leakage proteins catalase, clusterin, ficolin, gelsolin, lumican, tetranectin, triosephosphate isomerase and vitronectin, were identified from this study (49).

SELDI proteomic patterns of lung cancer patient serum have also been assessed to distinguish lung cancer patients from healthy individuals. Three protein peaks from the profiles of 30 lung cancer patients and 51 age- and sex-matched healthy individuals achieved 93.3% sensitivity in detection of lung cancer patients (50). Similarly, five protein peaks at 11493, 6429, 8245, 5335 and 2538 Da were chosen automatically as a biomarker pattern from the profile of a training set composed of 208 serum samples, including 158 lung cancer patients and 50 healthy individuals, and this pattern provided sensitivity of 91.4% in the detection of non-small cell lung cancers (NSCLC) (51).

Breast cancer. The classical pathological methods that are used to predict survival, development of metastatic disease or to guide selection of primary therapy (52, 53) in patients with breast cancer rely on anatomical staging of cancer and include the Nottingham Prognostic Index (54), Adjuvant Online (AO) (55) and the St. Gallen criteria (56). The current St. Gallen derived algorithm for selection of adjuvant systemic therapy for early breast cancer patients relies on tumour size and grade, nodal status, menopausal status, peritumoural vessel invasion, endocrine status and HER2 (epidermal growth factor receptor 2) status. HER2 is the most prominent and commonly used biomarker for breast cancer detection (57). However MMP-2 (matrix metalloproteinase-2 immuno-reactive protein), absence of oestrogen and progesterone receptors and high expression of Ki-67 (Mib-1) antigen, osteopontin (OPN), urikinses type plasmonogen activator and its inhibitors (PAI-1 and 2) and cathepsins (B and L) have also been indicated as prognostic biomarkers for breast cancer (58-66).

In a proteomics study of breast cancer serum, two proteins, HSP27 (up-regulated) and 14-3-3 sigma (down-regulated) were identified using 2D-PAGE coupled with MALDI-TOF-MS (67). Comparison of the expression patterns of these two proteins correctly classified 97% of the controls as not cancer and 100% of cancer samples as malignant. This result yielded 100% sensitivity. The positive predictive value for this sample set was 98%. In another study 2D-DIGE analysis of serum samples obtained from 39 patients with breast cancer and 35 controls revealed that prosoplipoprotein A-I, transferrin, and hemoglobin were up-regulated and three proteins, apoliprotein A-I, apolipoprotein C-III, and haptoglobin a2 were down-regulated in cancer patients (42). However, in cross validation with routine clinical immunochemical reactions, the serum levels of apolipoprotein A-I and haptoglobin could not be detected, indicating the superiority and sensitivity of the 2D-DIGE technique.

A pattern of three serum biomarkers (two up-regulated at 8.1 kDa and 8.9 kDa and one down-regulated at 4.3 kDa) were identified from the SELDI profiles of 169 serum samples from 103 breast cancer patients, 41 healthy women, and from 25 benign breast cancer patients that distinguished patients from controls (16). The sensitivity and specificity after cross-validation within the sample group (bootstrapping), using a random subset of the data to build the model and testing it with the remaining data, were found to be 93% and 91%, respectively. However, a new set of samples in a later study confirmed the up-regulation of the 8.1 kDa and 8.9 kDa biomarker species; subsequently the 8.9 kDa species was identified to be a complement component of C3a (desArg) and the 8.1 kDa species as a C-terminal-truncated form of C3a (desArg) (68). In a later study also using SELDI-TOF MS, the combination of an independent cancer biomarker, Ca 15.3, with the serum biomarkers, 4.286 kDa and 4.302 kDa possibly corresponding to the 4.3 kDa peak identified by Li et al. (16), and 8.919 kDa and 8.961 kDa possibly corresponding to the 8.9 kDa peak also identified by Li et al. (16), significantly improved breast cancer diagnosis (69). In another study, four peaks, CA1 (17.3 kDa), CA2 (26.2 kDa), CA3 (5.7 kDa), and CA4 (8.9 kDa), were chosen as potential biomarkers from the SELDI profile of 49 breast cancer patients, 51 patients with benign breast diseases and 33 healthy women to build a prediction model using artificial neural networks and discriminant analysis (70). One hundred percent sensitivity and specificity were observed in a training set while a blind test set, showed 76.47% sensitivity and 90.00% specificity.

Two genes, BRCA-1 and BRCA-2, have been observed to be strong tumour suppressors in men and women (71, 72). Fifteen serum samples from women with BRCA-1 mutations who developed breast cancer (BRCA-1 Ca) and 15 from those who did not (Carrier), 16 from normal volunteers (NL), and 16 from women with sporadic breast cancer (SBC) were profiled with SELDI-TOF enabling the differentiation between BRCA-1 Ca vs. Carrier with 87% sensitivity and 87% specificity and BRCA-1 Ca vs. SBC patients with 94% sensitivity and 100% specificity (73).
Similar results with ~86% discrimination between women with BRCA-1 breast cancers from the 15 non-cancer BRCA-1 carriers were later observed (74).

For optimization and individualization of therapeutic decisions, a multiprotein complex (including haptoglobin, C3a complement fraction, transferrin, apolipoprotein C1 and apolipoprotein A1)-based model developed on SELDI profiles of denatured and fractionated early postoperative serum samples from 81 high-risk early breast cancer patients was shown to correctly predict the outcome in 83% of patients for metastatic relapse (75).

Pancreatic cancer. Poor prognosis in pancreatic cancer is attributed to the fact that most patients do not develop overt symptoms until the disease has disseminated or caused local organ dysfunction (76). Currently CA 19-9 is the accepted serum marker for pancreatic cancer, but it was approved only for monitoring treatment response by the U.S. Food and Drug Administration (77, 78). Approximately 80-90% of people with pancreatic cancer have elevated levels of this marker in their blood (79). Additionally, current methods of diagnosis, including CA 19-9, are ineffective for identifying small, surgically resectable cancers.

To identify potential serum markers for pancreatic cancer, serum samples from 35 pancreatic cancer patients and 3 normal and healthy individuals were analysed using 2D-DIGE coupled with MALDI/TOF/TOF-MS and 24 unique up-regulated proteins and 17 unique down-regulated proteins were identified in cancer serum (8). Increased levels of apolipoprotein E, R-1-antichymotrypsin and inter-R-trypsin inhibitor in serum proteome analysis of 20 patients with pancreatic cancer and 14 controls were also found to be associated with pancreatic cancer by western blot analysis; the use of these proteins resulted in a sensitivity of 82.6% and a specificity of 100% in pancreatic cancer diagnosis. In another study, samples from 32 normal and 30 pancreatic cancer patients were analysed using 2D-PAGE and 9 spots out of 154 commonly overexpressed proteins discriminated 100% of pancreatic cancer samples and 94% of normal samples (80). Fibrinogen-γ, a protein associated with the hypercoagulable state of pancreatic cancer, was identified as one of these nine spots and was found to discriminate all cancer cases from normal sera, successfully indicating the potential of fibrinogen-γ in pancreatic cancer prediction.

Differential glycoprotein expression has also been observed in human cancer serum (81). Sialylated glycoproteins from highly abundant protein-depleted serum samples of normal and pancreatic cancer individuals were extracted using three different lectins [wheat germ agglutinin (WGA), elderberry lectin (SNA), *Maackia amurensis* lectin, (MAL)] after HPLC-based fractionation. 2D-PAGE was used for protein separation and approximately 130 sialylated glycoproteins were identified using μLC/MS/MS. The sialylated plasma protease C1 inhibitor was identified to be down-regulated in cancer serum. Changes in glycosylation sites in cancer serum were also observed by glycopeptide mapping using μLC-ESI-TOF-MS where the N83 glycosylation of R1-antitrypsin was found to be down-regulated.

Patients with cancer have been found to frequently develop autoantibodies and the identification of panels of tumour autoantigens may have a role in the early diagnosis of cancer and immunotherapy. DEAD-box protein 48 (DDX48), which is highly similar to eukaryotic initiation factor 4A, was observed in 63.64% of patients with newly diagnosed pancreatic cancer and in only 1.9% of normal controls using an antibody-based reactivity assay (82). Following on from this observation, a large sample set was analysed to evaluate DDX48 as a diagnostic antigen and 33.33% of pancreatic cancer patients, 10% of colorectal cancer patients, 6.67% of gastric cancer patients and 6.67% of hepatocellular cancer patients were positive for anti-DDX48 autoantibody reactivity using a purified DDX48 antigen-based ELISA assay, while none of the 20 chronic pancreatitis patients, 30 lung cancer patients, and 60 normal individuals were positive for this assay. Therefore, the detection of autoantibodies to DDX48 in serum may improve clinical diagnosis of pancreatic cancer (82).

Sera from 49 pancreatic cancer patients and 54 unaffected individuals were profiled using SELDI-ToF-MS (83). Based on the SELDI profiles, a classification model was generated with classification and regression tree and logistic regression methods and this resulted in differentiation of diseased patients with 100% sensitivity. The specificity was 93.5% using a decision tree algorithm and 100% using a logistic regression model. In another study, the two most discriminating protein peaks (m/z 3146 and 12861) from the SELDI profiles of serum samples from 60 patients with resectable pancreatic adenocarcinoma, 60 age- and sex-matched patients with non-malignant pancreatic diseases and 60 age- and sex-matched healthy controls, differentiated patients with pancreatic cancer from healthy controls with a sensitivity of 78% and a specificity of 97%. This was found to be significantly better than the current standard serum marker, CA19-9, on the same sample set (with a 65% sensitivity and 85% specificity) (84). The combination of the two peaks and CA19-9 resulted in slightly improved specificity. In a further study on pancreatic cancer, a set of four mass peaks at 8766, 17272, 28080 and 14779 m/z, whose mean intensities differed significantly in the SELDI profiles of 245 plasma samples including diseased and normal individuals, were selected as a biomarker profile using a support vector machine learning algorithm which accurately discriminated cancer patients from healthy controls in a
training cohort (sensitivity of 97.2% and specificity of 94.4%) and in a validation cohort (sensitivity of 90.9% and a specificity of 91.1%) (85). The combination of CA19-9 with these differentially expressed peaks resulted in 100% detection of pancreatic cancer cases (29/29 samples), including early-stage (stages I and II) tumours.

Melanoma. Currently no protein serum marker is available for surveillance of melanoma progression in early-stage melanoma (86, 87). A study of serum specimens of 49 early-stage (AJCC stage I and II) patients, including 25 patients with melanoma recurrence and 24 without evidence of disease, following resection were profiled using SELDI-ToF-MS (86). The differential pattern of peaks among patients with recurrence and without recurrence were used for predicting the chances of cancer recurrence and resulted in a sensitivity of 72% and a specificity of 75% that was significant (86). Similarly, 205 serum samples from 101 early-stage (American Joint Committee on Cancer (AJCC) stage I) and 104 advanced stage (AJCC stage IV) melanoma patients were analysed using SELDI-ToF and MALDI-ToF MS. SELDI profiles were used to train artificial neural networks (ANN). Based on these ANN algorithms, 88% of stage assignment was correctly predicted; 80% of stage III samples could be correctly as progressors or non-progressors and 82% of stage III progressors were correctly identified. The prediction accuracy was much improved compared to the cases predicted by the conventional marker S-100β (88) where only 21% of the stage III progressors were detected.

Future Directions

Significant progress has been made in proteomics technology development in the last decade and this has enabled researchers to move forward to a better understanding of the disease. Several potential cancer type-specific biomarkers for the early detection of the disease or for therapy decision-making have been proposed using 2D-electrophoresis or SELDI approaches on patient serum. Newer technologies, such as protein arrays, ICAT, iTRAQ and MudPIT, will increase this potential for biomarker discovery in cancer.

References


Received December 19, 2006
Accepted February 14, 2007