

# Proteomics in pathology research

Megan S Lim<sup>1,2</sup> and Kojo SJ Elenitoba-Johnson<sup>1,2</sup>

<sup>1</sup>Department of Pathology and <sup>2</sup>Associated Regional and University Pathologists Institute for Clinical and Experimental Pathology, University of Utah Health Sciences Center, Salt Lake City, UT, USA

**Proteomics is a multifaceted approach to study various aspects of protein expression, post-translational modification, interactions, organization and function at a global level. While DNA constitutes the ‘information archive of the genome’, it is the proteins that actually serve as the functional effectors of cellular processes. Thus, analysis of protein derangements on a proteome-wide scale will reveal insights into deregulated pathways and networks involved in the pathogenesis of disease. Although the field of proteomics has advanced tremendously in recent years, there are significant technical challenges that pose limitations to the routine application of mass spectrometry to clinical research. Despite these challenges, proteomic studies have yielded unparalleled information and understanding of the cellular biology of diseased states. The application of mass spectrometry to the study of diseases will ultimately lead to identification of biomarkers that are critical for the detection, diagnosis, prognosis and treatment of specific disease entities.**

*Laboratory Investigation* (2004) **84**, 1227–1244, advance online publication, 16 August 2004; doi:10.1038/labinvest.3700167

**Keywords:** pathology; proteomics; mass spectrometry; review

Proteomics is defined as the study of the proteome. As a term analogous to the genome, the proteome is defined as the total protein complement of an organelle, cell, tissue or an entire organism.<sup>1–3</sup> Proteomics is a multi-faceted endeavor to study protein expression and post-translational modification, and protein interactions, organization and function at a global level. Why study the proteome in disease? While the cause of a disease aberration may be at the genetic level, the functional consequences of such an aberration are expressed at the protein level. Hence, disruption of protein structure, function or interaction is the underlying mechanism of the majority of diseases. Further, several questions relating to the fundamental basis of diseases are best addressed at the protein level because they occur as post-translational events, or relate to abnormal regulation of protein function, or aberrant protein–protein or protein–DNA interactions. Additionally, covalent modification of proteins (undetectable by genomic studies) could confer abnormal activities to proteins critical to cell growth, survival, proliferation and apoptosis.<sup>4,5</sup>

Although the field has progressed rapidly in recent years, there are significant technical chal-

lenges which pose limitations to the application of mass spectrometry to the average pathologist–investigator. Despite these challenges, proteomic studies have yielded unparalleled information and understanding of the cellular biology of diseased states. The application of mass spectrometry to the study of diseases will ultimately lead to identification of biomarkers that are critical for the detection, diagnosis, monitoring, prognosis and treatment of specific disease entities. In this review, we provide an overview of the technical aspects involved in mass spectrometry-based proteomics, and illustrate some applications of these methods and technologies to the study of two distinct categories of lymphoid neoplasia.

## Principles and tools for proteomics

### Samples for Proteomic Analysis

#### *Protein isolation*

Most proteomic studies require proteins extracted from fresh or snap frozen cells, for example, biological fluids, such as serum, urine or tissue samples. Owing to cross-linking and degradation of proteins, paraffin-embedded formalin-fixed tissues are not ideal sources for proteomic studies. Nevertheless, proteins extracted from ethanol-fixed paraffin-embedded tissues can be utilized for two-dimensional gel electrophoresis–mass spectrometry (2D-MS) analysis.<sup>6</sup> Cellular proteins have to be

Correspondence: Dr MS Lim, MD, PhD and Dr KSJ Elenitoba-Johnson, MD, Division of Anatomic Pathology, University of Utah Health Sciences Center, 50 North Medical Drive, Salt Lake City, UT 84132, USA.

E-mail: megan.lim@path.utah.edu; kojo.elenitobaj@path.utah.edu  
Received 15 July 2004; accepted 16 July 2004; published online 16 August 2004

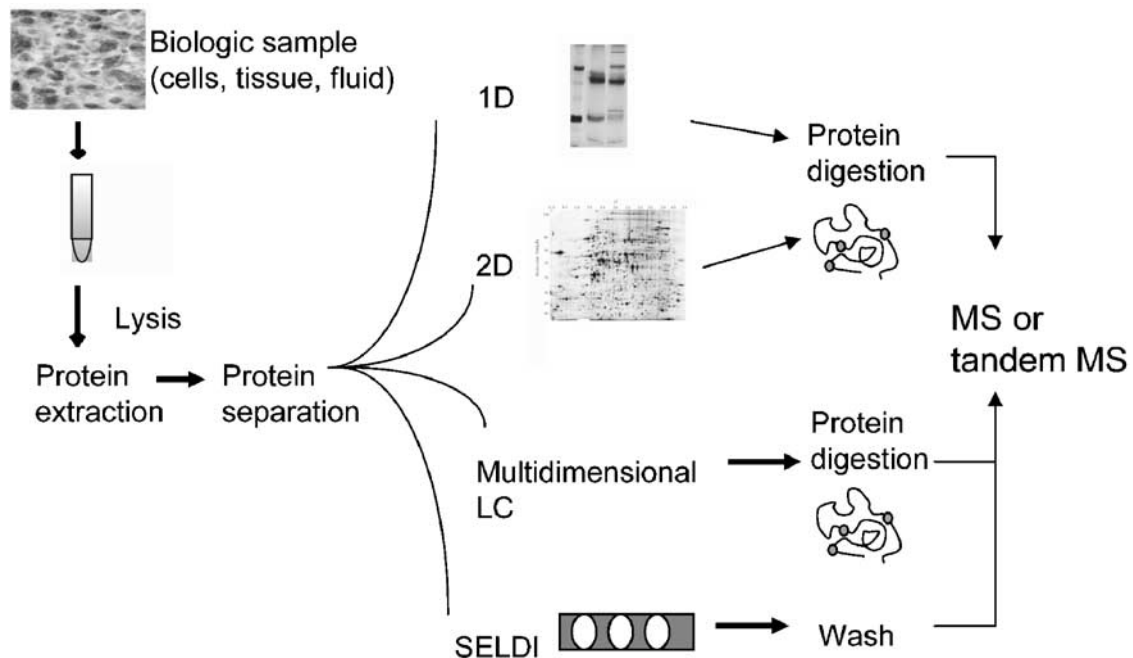
isolated from samples containing other biological molecules including carbohydrates, lipids and nucleic acids. Thus, protein extraction protocols entail; the homogenization of cells and tissues followed by application of detergents such as 3-(dimethylammonio)-1 propane sulfate (CHAPS),<sup>7</sup> Tween and sodium dodecyl sulfate (SDS), which help to dissolve the proteins and separate them from the lipid components, reducing agents such as dithiothreitol (DTT); denaturing agents such as urea which disrupt the bonds that are responsible for the formation of secondary and tertiary conformational structure; and enzymes which degrade nucleic acids such as DNAses and RNAses. Tissues obtained from laser capture microdissection of human colon cancer<sup>8</sup> and hepatocellular carcinoma have been successfully used for 2D gel electrophoresis followed by MALDI-TOF mass spectrometry analysis and 2D liquid chromatography tandem mass spectrometry.<sup>9</sup>

#### Separation of complex proteins into simpler components

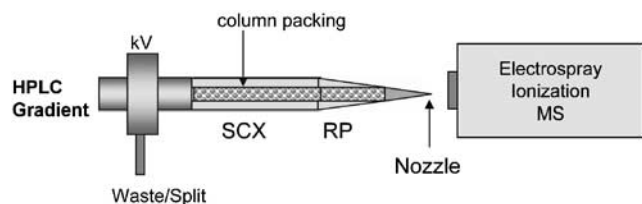
Several techniques are utilized for the analytical separation of proteins. Figure 1 illustrates different modalities commonly used for separation of proteins from complex mixtures. These include; one-dimensional gel electrophoresis (1D-GE) which achieves resolution of cellular proteins based on molecular weight, two-dimensional gel electropho-

resis (2D-GE), which involves initial separation of proteins based on isoelectric point (pI) followed by subsequent separation based on size; high-performance liquid chromatography (HPLC); ion exchange and different types of affinity chromatography.<sup>10,11</sup> The most powerful strategy entails the integration of the different protein and peptide separation methods as multidimensional combinations. In this regard, ion-exchange liquid chromatography (LC) in tandem with reverse-phase (RP)-HPLC is a powerful tool for resolving complex peptide mixtures, and has been automated to maximize efficiency.<sup>12–15</sup> When used with a 100  $\mu\text{m}$  2D strong cationic exchange-reverse-phase (SCX-RP) packed microcapillary, the MS system is capable of achieving a detection limit of 10 fmol for identifying a single tryptically digested protein.<sup>16</sup> The sensitivity of this technique (multidimensional protein identification technology, MudPIT, see Figure 2) is very attractive for the study of complex proteomes such as mammalian cellular samples. MudPIT has been applied to large-scale protein characterization and identification of up to 1484 proteins from yeast in a single experiment.<sup>13</sup>

Proteins isolated from gels or individual chromatography fractions can be subjected to proteolytic cleavage by enzymes such as trypsin with specific cleavage sites (at the carboxy-terminal of lysines or arginines), or nonspecific cleavage specificity such as elastase or subtilisin.<sup>3,15</sup>



**Figure 1** Overview of experimental design for mass spectrometry-based proteomic studies. Proteins are extracted from biologic samples and fractionated by a variety of separation methods. In 1D gel electrophoresis, proteins are separated by size. In 2D gel electrophoresis, proteins are separated based on isoelectric point (pI) and size. In multidimensional liquid chromatography, digested proteins are fractionated by 2D (strong cationic exchange (SCX) and reverse phase (RP)) or 3D (strong cationic exchange, avidin and reverse-phase) liquid chromatography. In the fourth method, proteins are separated based on functional properties according to physical, chemical or biochemical properties in surface-enhanced laser-desorption ionization (SELDI) technology.



**Figure 2** Multidimensional protein identification technology (MudPIT). Complex peptide mixtures from fractions obtained from whole cell lysates are separated onto a microcapillary column packed with strong cation exchange (SCX) and reverse-phase (RP) resins. Peptides are directly eluted into the tandem mass spectrometer via direct interfacing with the microcapillary high-pressure liquid chromatography (HPLC) column. The coupling of separation techniques allow for high-throughput, on-line analysis of peptide or protein mixtures.

### General Strategy for Mass Spectrometry-Based Proteomics

Proteomic studies require the simplification of a complex mixture of proteins into less complex components that are more amenable for analysis. In this regard, intact proteins with different biophysical characteristics such as molecular weight, hydrophobicity, and post-translational modifications may be present within a complex mixture intended for analysis. In 'top-down' proteomics, intact proteins are analyzed. In 'bottom-up' proteomics, the proteins are proteolytically cleaved using enzymes with or without cleavage specificity.

The development of sensitive instruments capable of analyzing larger biologic molecules such as proteins has greatly facilitated the analysis of the total complement of proteins in cells and tissues. Mass spectrometers measure the mass of the smallest of molecules with very high accuracy, and hence mass spectrometry can be considered as the smallest weighing scale. In parallel with the technological advancements in mass spectrometers, technological improvements in ionization methods have also enhanced the ability to analyze complex biologic molecules by mass spectrometry.<sup>4,17</sup> The final component that has greatly impacted the ability to conduct proteomic studies is development of translated genomic databases and specialized software algorithms that rapidly search mass spectrometric data against known or predicted proteins within the databases.<sup>4</sup>

In general, the measurement of fragmented peptide masses by mass spectrometry is more accurate than measuring the mass of intact proteins. Thus in 'bottom-up' proteomics, the typical work-flow involves initial simplification of a complex protein mixture followed by digestion into peptides which are subjected to mass spectrometric analysis. The mass spectrometric data are then analyzed using specialized software algorithms that identify the proteins from which the peptide sequences are derived. The ability to accurately determine the mass of a unique peptide that originates from a

particular protein greatly facilitates the identification of that protein. In essence, protein identification centers on the fact that a peptide sequence, composed of six amino-acid residues or greater, provides a unique opportunity for identification of the parent protein. This is because the probability that any one amino acid would occupy a particular position within a peptide sequence is  $1/20$ . For a sequence of six amino-acid residues, the theoretical probability is  $1$  in  $20^6 = 1$  in  $64\,000\,000$ . For a number of reasons, even these odds may be insufficient to unequivocally identify a protein from a single peptide. Identification of overlapping or longer peptide sequences with multiple 'hits' to a particular protein provides an even greater degree of certainty in the identification of the protein. Thus in many cases, it is possible to utilize database searches to identify a protein from only a few peptides.

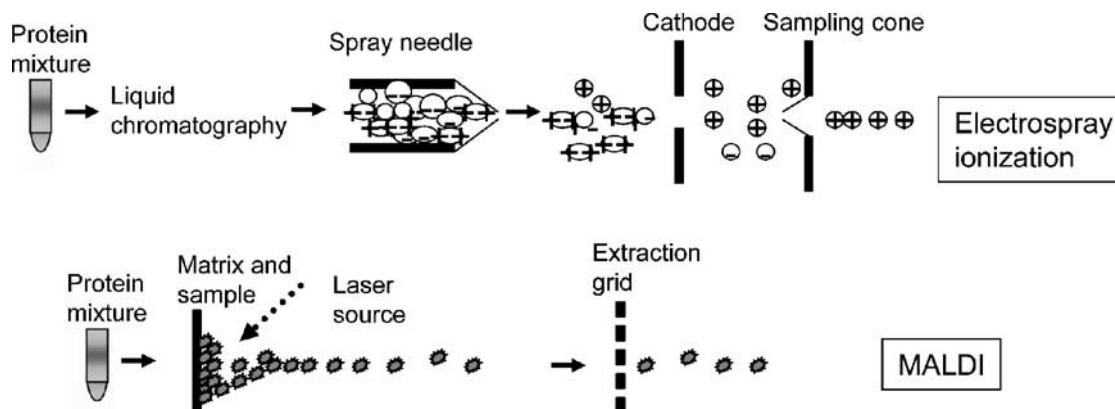
### Mass Spectrometers

A mass spectrometer is typically composed of three components: an ionization source, the mass analyzer and the detector. The ionization source creates ions from the sample to be analyzed. The mass analyzer resolves the ions by their mass-to-charge ratio ( $m/z$ ). The detector determines the mass of the ions. The most frequently utilized ionization sources and mass analyzers are discussed below. See Figure 3 for a schematic representation of mass spectrometers.

### Ionization Source

#### MALDI

The recent development of so-called 'soft' (low-energy) ionization techniques such as matrix-assisted laser desorption/ionization (MALDI)<sup>18,19</sup> and electrospray (ESI) techniques<sup>17</sup> have dramatically enhanced the possibility to analyze larger biomolecules in general, and proteins in particular by mass spectrometry. In MALDI, the sample intended for analysis is incorporated into a chemical matrix containing compounds such as 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid). Laser activation of the target by a laser within the ion source leads to the release from the target of peptide/protein ions into gas phase. More recently, a variation on the MALDI concept has been introduced; namely, surface-enhanced laser desorption/ionization (SELDI).<sup>20,21</sup> This latter format is embodied in the Ciphergen Chip™ system and is composed of several chip matrices, which exploit the differing biophysical and chromatographic characteristics of the different proteins for their preferential selection. The different surfaces include among others, a hydrophobic surface, a strong anionic exchange surface, and immobilized metallic ion with a strong affinity for phosphorylated proteins.



**Figure 3** Schematic representation of common protein ionization methods used for mass spectrometry-based proteomics. Two common ionization technologies are currently available for analysis of proteins. Electro spray ionization (ESI) volatilizes and ionizes peptides and proteins in solution and is coupled to two mass analyzers resulting in tandem MS. Matrix-assisted laser-desorption ionization (MALDI) uses analytes which are cocrystallized in a matrix composed of organic acid on a solid support. A pulse of ultraviolet laser is then used to evaporate the matrix and the analyte into gas phase resulting in generation of single charged ions.

### Electrospray ionization

In contrast to the MALDI wherein ions are generated from a solid matrix, electro spray ionization involves the generation of peptide ions from aqueous solution.<sup>17</sup> The solution containing the sample passed through a needle subjected to a high voltage. The solution stream is ejected from the needle orifice as a spray of droplets. The solvent is eliminated from the droplets by a heated capillary or an inert gas. Solutions with acidic pH favor protonation of the N-terminal amines and histidine nitrogens, and peptide fragmentation is facilitated when the peptide ions are positively charged. Thus, ESI protocols commonly include acidification steps prior to peptide ion analysis in the mass analyzer.

### Mass Analyzers

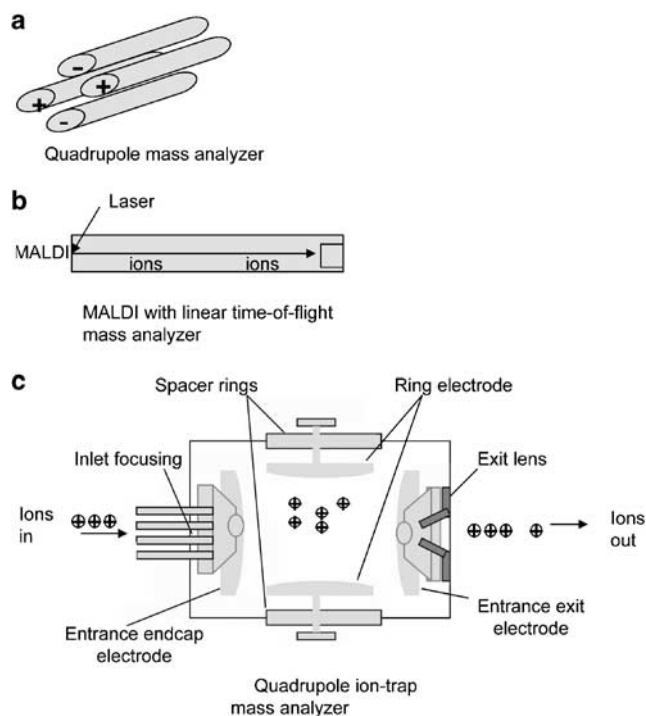
Mass analyzers measure the mass-to-charge ( $m/z$ ) ratio of gas-phase ions generated from the ionization source. Several types of mass analyzers are utilized for proteomic analysis. Examples include quadrupoles, ion traps, time-of-flight and Fourier transform ion cyclotron resonance (FTICR) type analyzers. In addition, hybrid instruments incorporating one type of analyzer with another (eg quadrupole-TOF) have also been utilized with great success.

Important performance characteristics that are relevant to mass analyzers are mass accuracy, mass resolution, and mass range. Mass accuracy refers to the extent to which a mass analyzer reflects the 'true'  $m/z$  values, and is measured in atomic mass units (amu) (parts per million (ppm) (eg 500 ppm) or percent accuracy (eg 0.05% accuracy). By comparison, mass resolution indicates the ability of the instrument to discriminate between ions with different  $m/z$  values. It is defined by the equation  $M/\Delta M$ , where  $M$  is the  $m/z$  ratio of a mass peak and  $\Delta M$  is the full width of a peak at half its maximum

height. The mass resolution of instrument often correlates with the accuracy of the instrument. High end mass spectrometers incorporate mass analyzers that employ variety of designs including reflectron time-of-flight and FTICR which exhibit exceptional mass accuracy and resolution. Mass range refers to the optimal  $m/z$  range at which a mass analyzer operates, and varies with the type of mass analyzer. For example, quadrupole mass analyzers exhibit a mass range of up to 4000  $m/z$ , and time-of-flight analyzers may have mass ranges extending up to 100 000 and beyond. Below, we briefly discuss the principles behind the operation of common analyzers that are utilized for proteomic mass spectrometry (illustrated in Figure 4).

### Quadrupoles

Quadrupole mass analyzers utilize electrical fields to separate ions based on their  $m/z$  values. A quadrupole is composed of four parallel poles or rods which have a fixed direct current and alternating radiofrequency potentials applied to them. Ions are either deflected into the poles or focused on the detector depending on the electrical field generated. Alteration of the electrical field frequencies and strengths would lead to the detection of different ions. The mass range and resolution that are achievable using the quadrupole assembly are determined principally by the physical dimensions (length and diameter) of the poles (Figure 4a). Quadrupole mass analyzers interface very well with ESI because quadrupole analyzers are tolerant of the relatively high pressures that are produced by ESI sources. Additionally, ESI generates multiply charged ions which widen the range of the  $m/z$  values of the ions to be detected by the mass analyzer. Since quadrupoles exhibit a wide mass range of detection, they are especially well-suited to the detection of ions produced by ESI. This complementary relationship has been exploited for



**Figure 4** Common types of mass analyzers used for proteomic studies. In quadrupole mass analyzers, ions are deflected into four parallel poles or rods which have fixed direct current and alternating radiofrequency potentials are applied to them (a). In time-of-flight mass analyzers (b), the ions are accelerated linearly down a flight tube until they impact a detector at the opposite end. The lighter ions will travel faster and reach the detector ahead of the heavier slower ions. The ion-trap mass analyzers (c) consist of the ring electrode, the entrance endcap electrode and the exit endcap electrode. They are unique in their ability to trap ions in a three-dimensional electrical field. The electrodes form a cavity in which ions are trapped and analyzed.

the design of several instruments, wherein ESI and quadrupole mass analyzers have been combined in commercial mass spectrometers. The two kinds of quadrupole mass spectrometers that are made for analytical work are single-stage and triple quadrupoles. The single-stage instruments have only limited application in proteomics studies because of limited mass resolution and lack of tandem mass spectrometry capability. By comparison, the triple quadrupoles instruments comprising three quadrupole analyzers in tandem are capable of true tandem mass spectrometry, and have been successfully utilized for a number of proteomics applications including product ion, precursor ion, and neutral loss scanning. These features are particularly the latter, are very useful in the analysis of post-translational modifications.<sup>22–25</sup>

#### Time-of-flight

In time-of-flight mass analyzers, the ions from the ion source are accelerated linearly down a flight tube until they impact a detector at the other end of the tube. Given the fact that the ions contain the same amount of energy but have different masses,

the lighter ions travel faster and reach the detector ahead of the heavier (slower) ions. The relationship between the mass and velocity of the ions is given by the equation  $KE = 1/2 mv^2$ , where ( $KE$ ) represents the kinetic energy, ( $m$ ) represents the mass and ( $v$ ) represents the velocity. Hence, TOF analyzers derive their name from the concept that the ‘time of flight’ ( $t$ ) of ions is related both to their  $m/z$  ratio and velocity ( $v$ ) within a fixed distance ( $d$ ) as expressed in the equation  $t = d/v$ . Although TOF analyzers have been frequently coupled with MALDI ion sources, some recently developed mass spectrometers have combined ESI with TOF. While linear mode TOF analyzers do not provide the best resolution, the ‘reflectron’ analyzers (Figure 4b) overcome this shortcoming and provide excellent resolution. For proteomics applications, the quadrupole time-of-flight (QqTOF) hybrid instruments with their superior mass accuracy, mass range and mass resolution are of much greater utility than simple TOF instruments and are very popular.<sup>22,26</sup>

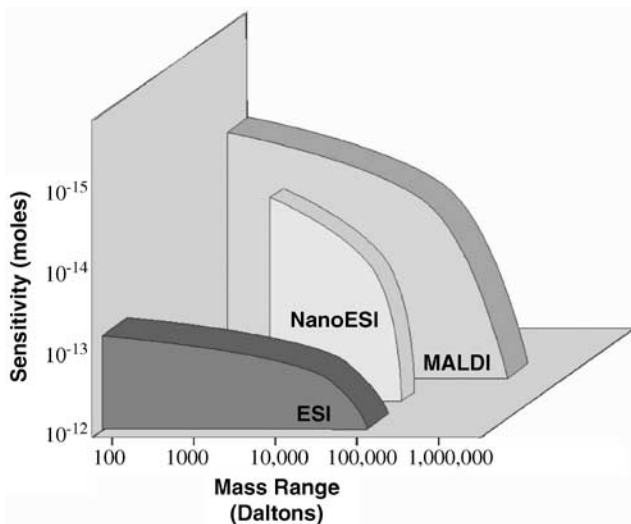
#### Ion-trap

The distinctive feature of ion trap mass analyzers is their ability to trap ions in a three-dimensional electrical field. The ion trap consists of a ring electrode and end cap electrodes (Figure 4c). The ion trap differ significantly in design and in operation from the quadrupoles in that, while triple quadrupoles perform tandem mass analysis on peptide ions as they pass through the analyzer, ion traps are capable of isolating and retaining specific ions for fragmentation upon collision with an inert gas in the same cell. The discovery of the mass-selective instability mode transformed the ion trap from a storage device into a powerful instrument capable of sequential ejection of selected or all ions out of the trapping cell to a detector.<sup>27</sup> The ion trap instruments are quite versatile, interface readily with liquid chromatography, ESI<sup>4</sup> and more recently, MALDI<sup>28</sup> and are very popular for many applications in proteomics.

Figure 5 illustrates the differential sensitivities characteristically observed for ESI, NanoESI and MALDI mass spectrometry. ESI and nanoESI have similarly high mass ranges, while ESI exhibits lower sensitivity. MALDI allows a high sensitivity as well as high practical mass range.

#### Fourier transform ion cyclotron resonance mass spectrometry

The Fourier transform ion cyclotron resonance mass spectrometer (FTICR) can be considered to be a special type of ion-trap mass analyzer.<sup>29</sup> However, the distinctive property of an FTICR is that the ion-trapping cell is surrounded by a magnetic field within which entrapped ions can resonate at their cyclotron frequencies. Additionally, an electrical field may be applied to excite the ions into a larger radius that can be measured as they travel past the detector plates situated on opposite sides of the trap.



**Figure 5** Typical sensitivity and mass ranges characteristic of different ionization techniques. Electrospray ionization (ESI) and nanoelectrospray ionization (nanoESI), have similarly high mass ranges, while ESI has lower sensitivity. Matrix-assisted laser desorption ionization (MALDI) allows a high sensitivity as well as high practical mass range.

The detector measures the cyclotron frequencies of all the entrapped ions and calculates mass-to-charge values of all the ions using a fast Fourier transform operation. The magnetic field strength (eg 7, 9.4 or 12 T) correlates with the performance properties of the FTICR instruments. FTICR instruments are very powerful and provide exquisite sensitivity, high mass accuracy and high resolution; which are especially desirable properties in the analysis of complex protein mixtures. The FTICR instruments are especially compatible with ESI,<sup>30</sup> but may also be used with MALDI as an ionization source.<sup>31</sup> Despite these advantages, FTICR instruments require considerable expertise for successful operation, thus limiting somewhat their widespread utilization for routine proteomics studies.

### Tandem Mass Analysis

Tandem analysis refers to the capacity of the analyzer to isolate an ion, subject it to further fragmentation and to perform analysis of the fragment ions. Fragmentation of the parent ion contains structural information that can be used to determine the amino-acid sequence of peptides, and to characterize post-translational modifications of proteins.

In tandem mass spectrometers, peptide ions analyzed in an initial mass analyzer are directed into a collision cell, where individual peptide ions can be isolated and subjected to further fragmentation upon collision with an inert gas (eg helium) in a process known as collision-induced dissociation (CID). Fragment ions exit the collisional cell and are separated in a second mass analyzer before

scanning out to the detector. The predictable nature of peptide fragmentation in CID makes it possible to predict an expected pattern of fragmentation for any peptide.

### Protein Identification Strategies

#### *Peptide mass fingerprinting*

Peptide mass fingerprinting (PMF) is the simplest method for protein identification. Peptide mass fingerprinting combines enzymatic digestion, mass spectrometry and computer algorithm-aided data analyses. Sequence-specific enzymes (eg trypsin) or chemicals (eg cyanogen bromide) with specific proteolytic cleavage activity are used to generate a series of peptides from a protein of interest. The peptides generated are analyzed by MS and the masses obtained are compared with theoretical mass spectra of the proteins in a sequence database.

The experimentally determined peptide masses are matched with the theoretical peptide masses of proteins within a database. The proteins with the highest number of experimental and theoretical peptide mass matches are ranked highest in probability of identification. The statistical significance of each of the matches is calculated and the matches are ranked in descending order of probability, with the protein showing the highest number of matches to the experimental data being considered the most likely match. Software algorithms that facilitate peptide mass mapping include PeptIdent/Multi-Ident and ProFound.<sup>32,33</sup> Peptide mass fingerprinting algorithms may assign higher scores to larger proteins which contain more peptides, thereby leading to an increased propensity for misassignment of peptide matches which could result in erroneous identifications.

Peptide mass fingerprinting is most suitable for the identification of proteins from species in which complete genome sequences are available. Identifications by PMF are most reliable when high mass accuracy mass spectrometers are utilized. In addition, PMF is optimally utilized for the identification of proteins that have been previously resolved by 2-D GE, such that additional information including the molecular weight and isoelectric points can be used to supplement PMF identification. Thus, PMF is not well-suited for searching expressed sequence tag (EST) databases which contain incomplete gene coding information for any particular EST. Further, PMF is also not ideal for the analysis of complex protein mixtures, since it would be difficult to determine which proteins the peptides originated from.

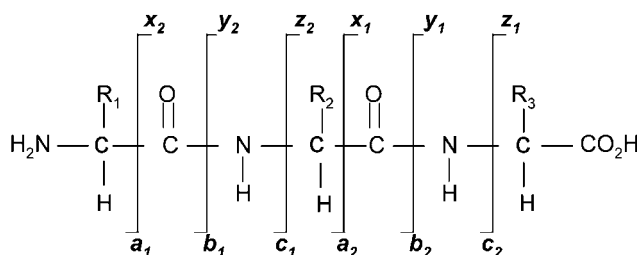
#### *Peptide sequencing by tandem mass spectrometry*

Peptide sequencing by tandem mass spectrometry is based upon the random cleavage of the peptide bonds between adjacent amino-acid residues in a peptide sequence achieved by collisional-induced

dissociation (CID). CID of the peptides yields ion series that are important for the identification of the amino-acid residues in a peptide. While several of the bonds along the peptide molecule are susceptible to cleavage, the most critical cleavages relevant to peptide sequencing by tandem mass spectrometry are those occurring along the peptide backbone. The most frequently observed cleavage occurs at the bond between the carbonyl oxygen and the amide nitrogen, fragmented to form a 'b-ion' and a 'y-ion' (Figure 6). In the 'b-ion', the positive charge is retained on the amino-terminal, and in the y ion the positive charge is retained on the carboxy-terminus of the parent peptide ion. Ions containing positive charge at both the N- and C-terminal ends are doubly charged. Fragmentation of doubly charged peptide ions results in 'b- and y-ions' that are complementary in peptide sequence information. Fortunately, these are also the most informative ions for sequence analysis.

The extent to which the predicted fragment ions for each of the amino-acid sequences match the peptide masses generated from the experimental tandem mass spectra is used to calculate a cross-correlation (XCorr) score 15 such that higher the cross-correlation score, the higher the quality of the match. The difference between a normalized cross-correlation score and the next best match is reported as the ( $\Delta C_n$ ) and indicates the quality of the top match in comparison to the next ranked sequences in the database.<sup>15</sup>

Recent developments in software algorithms facilitating improved identification include scoring routines that normalize XCorr values to be independent of peptide and database size (SEQUEST-NORM)<sup>34</sup> and PeptideProphet, INTERACT and ProteinProphet,<sup>35–38</sup> which utilize statistical modeling algorithms that permit assignment of confidence values for each identification. In particular, a very useful aspect of the latter routine is that it allows the determination of false-positive rates of identification in specific datasets.



**Figure 6** Peptide ion fragmentation by collision-induced dissociation (CID). Fragmentation of peptide ions occurs predominantly along the peptide backbone. In the most commonly observed cleavage, the bond between the carbonyl oxygen and the amide nitrogen is cleaved to form a 'y-ion' and a 'b-ion'. The peptide fragments retaining the positive charge on the N-terminal are designated as a, b, c and the fragments retaining the positive charge on the carboxy-terminal end are designated x, y, z.

It should be noted that because of sequence conservation among similar or related proteins, a single tandem mass spectrum should not be construed as providing a unique identification of a particular protein. Rather, multiple peptide 'hits' corresponding to a specific protein sequence provide unequivocal evidence of identification of that protein.

As compared to PMF, MS/MS is advantageous in that it can provide unambiguous identification of proteins in complex mixture; indeed, matching of multiple MS/MS spectra to peptide sequences within the same protein increases the confidence of protein identification. MS/MS data can be used to search EST databases with reliable matches. A flow chart comparison of PMF and tandem MS for protein identification is summarized in Figure 7.

## Applications of mass spectrometry-based proteomics

There are two major groups of applications of proteomics to the study of human disease. One is 'expression proteomics' which deals with the identification/quantification of proteins expressed, and the levels of proteins expressed in a given sample such as body fluids, normal or disease tissues. This protein expression 'signature' is conceptually similar to that obtained by genomic microarray analyses and would allow investigators to identify biomarkers or disease-specific proteins that may represent therapeutic targets. The second application is 'functional proteomics' which encompasses the study of protein in their functional environment. This includes the analysis of protein interactions with other proteins, interactions with DNA or RNA, and post-translational modifications such as phosphorylation and glycosylation. The latter approach allows investigators to obtain information regarding function of a protein, for example, identifying networks of signaling pathways that are characteristic of physiologic and pathologic states.

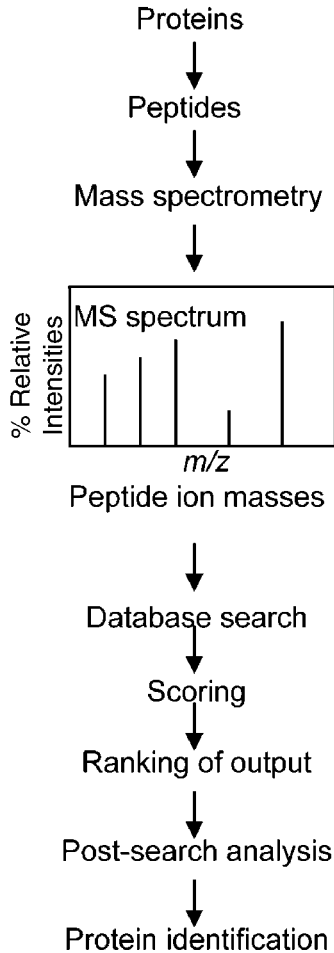
## Protein Expression Profiling

The application of proteomics to the identification of disease markers from body fluids and tissue has received significant attention. The potential of obtaining mass spectral profiles of peptides and proteins without the need to carry out protein separation is highly suitable for biomarker discovery, especially as it would have reduced sample requirements and represent a high throughput approach. A popular method is SELDI-TOF.<sup>39–44</sup> The advantage of this method is that only small quantities of biological fluid or tissue material are needed. The mass spectral patterns reflect the protein and peptide content of the samples. In this regard, artificial-intelligence-based bioinformatics

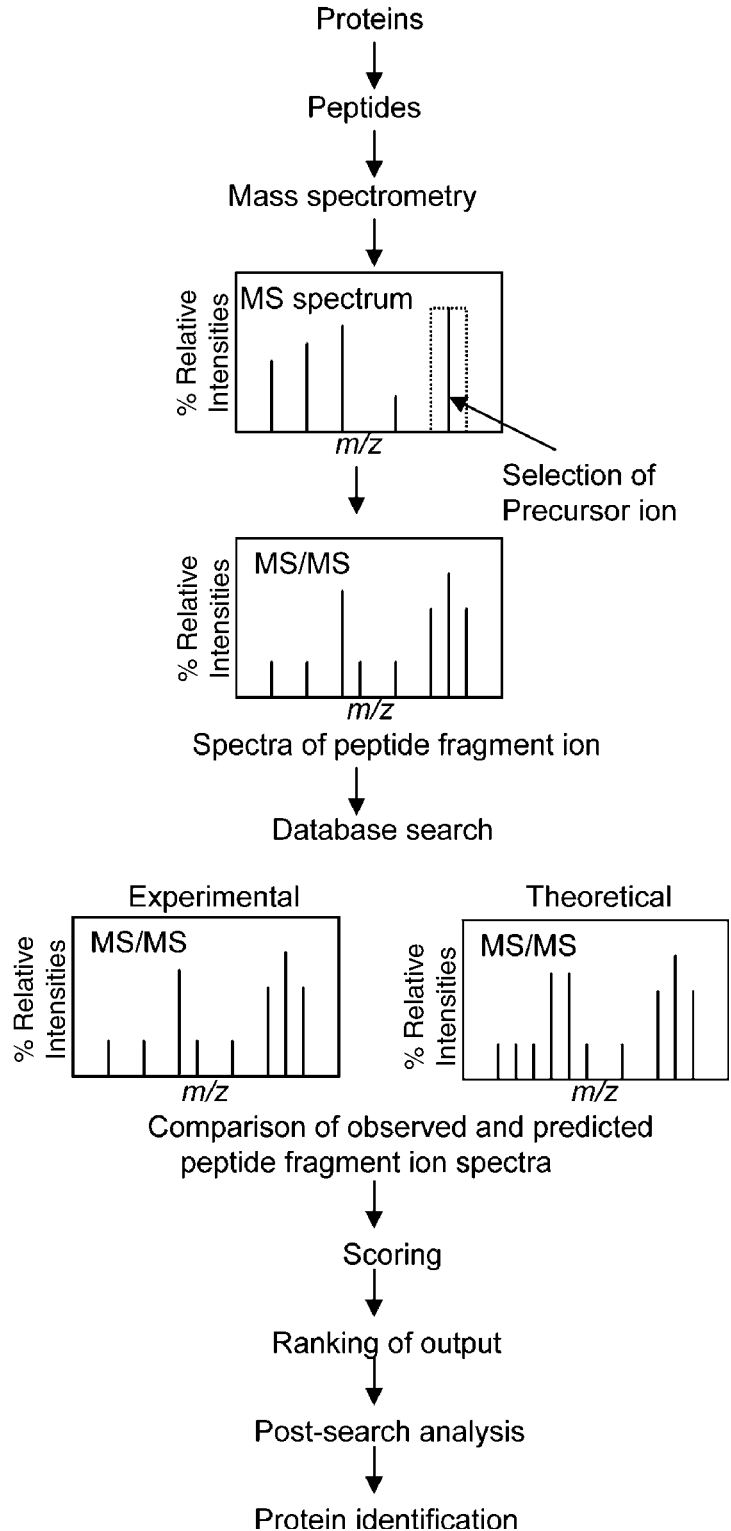
analyses of mass spectral patterns have been proposed to distinguish serum of normal patients with that of a number of neoplastic conditions including ovarian,<sup>42</sup> breast,<sup>45</sup> prostate<sup>46</sup> and liver

cancer.<sup>47</sup> Such analyses require extensive pre- and postacquisition procedures including mass calibration, baseline correction and noise subtraction to facilitate identification of *bona fide* features that are

### Peptide mass fingerprinting



### MS/MS-based protein identification



robust and biologically relevant discriminators of the normal, benign and malignant states.<sup>48,49</sup> Proteomic patterns of nipple aspirate fluids,<sup>50</sup> cytologic specimens<sup>51</sup> and tissue biopsies<sup>52</sup> using SELDI-TOF have also revealed a potential for utility in discovery of novel biomarkers that aid in diagnosis.

Our laboratory sought to identify proteomic changes that accompany histologic transformation in follicular lymphoma. To this end, we used SELDI-TOF mass spectrometry (ProteinChip™, CIPHERGEN Biosystems) to perform pairwise comparisons of low-grade and transformed follicular lymphomas in order to identify proteins that may be involved in the transformation process. In preliminary experiments, we demonstrated that the MS data obtained from SELDI-TOF MS were reproducible, and that reduction in sample complexity improved the ability to detect lower abundance proteins. We subjected crude cell lysates obtained from snap-frozen archived follicular lymphoma and patient-matched diffuse large B-cell lymphoma to analysis by SELDI-TOF mass spectrometry. Reproducible mass spectral profiles generated using a strong anionic exchange (SAX2)/immobilized metallic affinity-capture surface chemistry revealed upregulation of a protein with a molecular weight of 32 kDa (Figure 8a). Protein database searches revealed several candidates, among them cyclin D3 (32.5 kDa) whose differential expression were confirmed by immunohistochemical analysis on the primary tissue specimens (Figure 8b).<sup>52</sup> These studies demonstrate the utility of SELDI-TOF-mass spectrometry for the rapid discovery of differentially expressed proteins using femtomolar quantities of crude protein derived from biopsy material. The versatility of this methodology supports its application to the rapid discovery of potential biomarkers in a variety of cellular systems.<sup>39-44</sup>

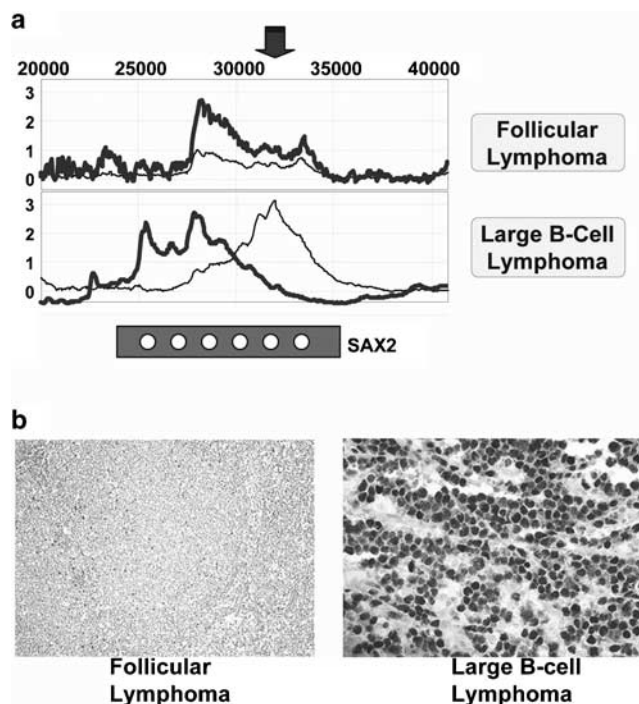
### Imaging Mass Spectrometry

Mass spectrometry has been used for the *in situ* analyses of proteins in tissue sections and those obtained by laser capture microdissection,<sup>53</sup> thereby allowing imaging and comparison of protein expression between normal and disease tissues.<sup>54-61</sup> In this strategy, frozen tissue sections are applied to a MALDI plate and analyzed at regular spatial intervals. The mass spectral data obtained at different intervals are compared to yield a spatial distribution of masses (proteins) across the tissue section. Analyses using this approach have revealed up to

1600 protein peaks from histologically selected 1 mm diameter regions of single frozen sections.<sup>57</sup> Using this approach, investigators have been able to distinguish glial neoplasms from benign brain tissues, and differentiate tumors of different histological grades.<sup>61</sup>

### Differential Protein Profiling (Quantitative Proteomics)

Most quantitative proteomic studies are designed to determine the proteomic differences between one cellular state and another, and these are 'relative'. In this regard, 2D-GE has been extensively utilized with great success.<sup>62</sup> Two-dimensional gel electrophoresis (2D-GE) although a traditionally popular



**Figure 8** SELDI-TOF mass spectrometry for the identification of differentially expressed proteins in follicular lymphoma transformation. A matched pair of follicular lymphoma and its transformed counterpart from the same patient was assessed using strong anionic exchange surface chemistry (SAX2). A protein with a molecular weight of 32 kDa was seen to be upregulated (arrow) in the large B-cell lymphoma compared to the follicular lymphoma (a). The molecular weight, the pI and chip binding characteristics suggested the identity of the protein to be cyclin D3, which was validated by its differential expression as evaluated by tissue immunohistochemistry of a follicular lymphoma and large B-cell lymphoma (b).

**Figure 7** Comparison of peptide mass fingerprinting and tandem mass spectrometry for protein identification. In PMF, the measured masses of intact peptide ions are compared with the calculated masses of peptides derived from *in silico* proteolysis of proteins from a database. PMF is useful for analyzing relatively pure samples containing few proteins. In tandem MS/MS, CID-induced fragmentation yields peptide ions with *m/z* values corresponding to specific amino-acid residues in the parent peptide. CID spectra are matched to peptide sequences with an intact peptide ion mass similar to that of the observed precursor ion. The calculated peptide fragment ion spectrum is compared to the observed CID spectrum. Unambiguous identification of a protein can be made with fewer matching peptides than required by PMF. Thus, tandem MS/MS is highly suitable for confident protein identification of samples composed of complex protein mixtures.

method for determining relative protein expression between two cellular states, is a relatively low throughput approach, which requires a large amount of starting material (in the order of 50  $\mu\text{g}$ ) with low sensitivity for detection of low abundance proteins such as cytokines and signaling molecules. Furthermore, proteins at both extremes of pI and molecular weight and those associated with the membrane fractions are not well represented by 2D-GE.<sup>11</sup> Despite these limitations, due to highly automated and roboticized spot picking technologies, numerous reports have successfully used 2D-GE followed by MALDI-TOF analysis to determine differential expression of protein profiles in normal vs tumor tissues including breast, colon,<sup>59</sup> lung, esophageal tissues<sup>63</sup> and cellular responses to stimulating and differentiating agents such as LPS<sup>64</sup> and response to Fas,<sup>65</sup> cytotoxic agents such as butyrate.<sup>66</sup> The average number of differentially expressed proteins identified by 2D-GE followed by MS analysis is in the range of 30–70 proteins.

#### *Stable isotope labeling in cell culture (SILAC)*

Stable isotope labeling with amino acids in cell culture (SILAC) is another useful global quantification strategy for the evaluation of differential expression of proteins from two distinct cellular populations.<sup>67,68</sup> In essence, the SILAC procedure entails culturing of cells from two different biologic conditions in parallel culture media that are deficient in a natural amino acid, but supplemented with monoisotopically labeled amino acid (eg <sup>12</sup>C, <sup>13</sup>C; <sup>14</sup>N, <sup>15</sup>N, respectively). The two cell populations metabolically incorporate the corresponding 'light' or 'heavy' isotopes in the synthesis of their respective cellular proteins. The proteins from each sample can thus be isolated, mixed at a 1:1 ratio and subjected to proteolytic digestion and mass spectrometric analysis. Corresponding peptides from each sample coelute during liquid chromatography and relative quantification of a particular peptide represented in both samples can be performed by measuring the ratios of the peptide mass peak intensities from matching isotopic peak pairs. The sequence of the peptide is subsequently obtained from tandem mass spectrometric analysis greatly facilitating identification proteins with differential expression in the two conditions. Of necessity, isolated cells must be capable of protein synthesis *in vitro*.

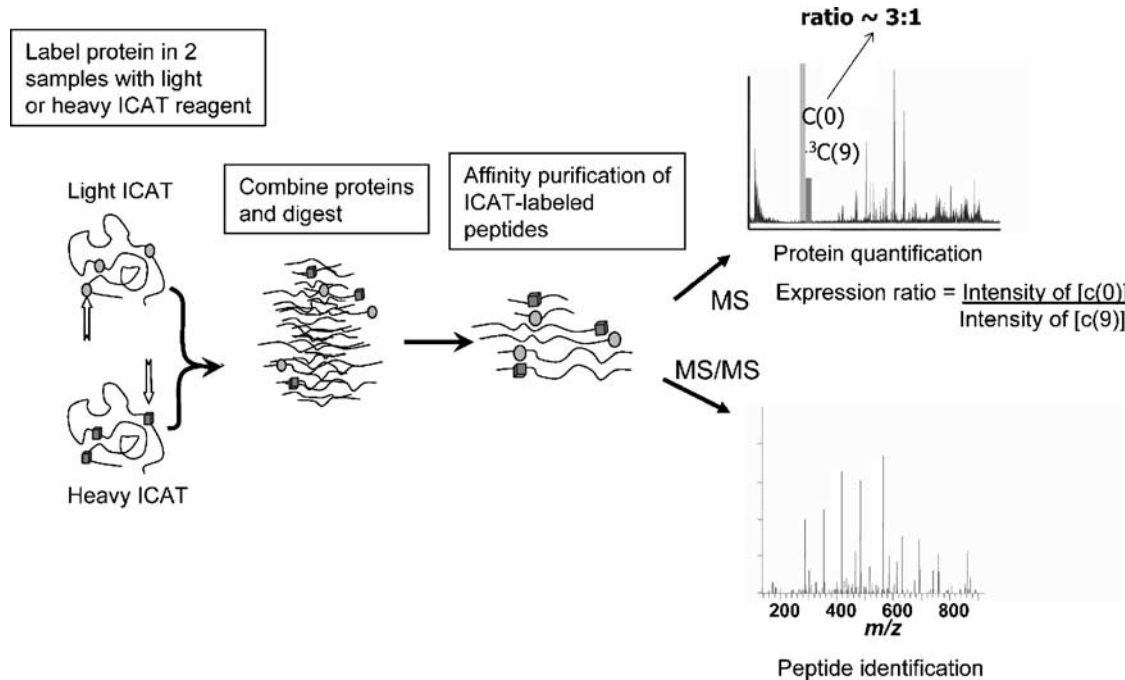
#### *Isotope-coded affinity tag method*

The isotope-coded affinity tags (ICAT) strategy is a relatively new technology for relative protein quantification, relying on postharvested, stable isotope labeling. The ICAT method uses the ICAT<sup>TM</sup> reagent to differentially label protein samples on their cysteine residues. The ICAT<sup>TM</sup> method is advantageous in that it permits the evaluation of low-abundance proteins and proteins at both extremes of molecular weight and isoelectric point.<sup>69</sup> The ICAT<sup>TM</sup>

reagent is composed of (i) a thiol-reactive group that reacts with the cysteine residues; (ii) a linker in which stable isotopes have been incorporated; and (iii) a biotin tag that enables affinity isolation and detection of peptides labeled with either the heavy or light versions of the ICAT<sup>TM</sup> reagent. In this system, one sample is labeled with a tag containing a light isotope, and the other sample to which it is being compared is labeled with a tag heavy isotope. The two samples are combined, proteolytically digested and analyzed by mass spectrometry (Figure 9). Because ICAT-labeled peptides elute as pairs from a reverse-phase column, calculating the ratio of the areas under the curve for identical peptide peaks labeled with the light and heavy ICAT<sup>TM</sup> reagent will allow determination of the relative abundance of that peptide in each sample. Advantages of the ICAT<sup>TM</sup> strategy include internal quantitation, automation, and reduced complexity of the peptide mixture. The original ICAT<sup>TM</sup> reagents featured either eight deuterium or hydrogen atoms at particular positions in the linker. The recently improved cleavable (c)ICAT<sup>TM</sup> reagent contains nine <sup>13</sup>C in the heavy version of the linker and nine <sup>12</sup>C in the light version of the linker. The resultant database search is constrained by the requirement of cysteinyl-group and is compatible with analysis of low abundance proteins and can be performed from proteins obtained from snap-frozen archival tissues. See Table 1 for comparison of 2D-SDS-PAGE, SILAC and ICAT<sup>TM</sup> methods for differential protein expression analysis.

The ICAT<sup>TM</sup> methodology has been applied to the quantitative protein expression analysis of *myc* overexpression, which revealed novel effects of *myc* on cytoskeletal function<sup>70</sup> as well as providing a global proteomic view of the intestinal epithelial cell response to enteropathogenic *Escherichia coli*.<sup>71</sup> Quantitative analysis of proteome changes that occur during neoplastic transformation from normal hepatocytes to hepatocellular carcinoma has also revealed hundreds of differentially expressed proteins<sup>9</sup> even from tissues obtained by laser capture microdissection.

We have utilized the ICAT<sup>TM</sup> method to determine the quantitative proteomic changes associated with inhibition of p38 mitogen-activated protein kinase (MAPK) in transformed follicular lymphomas.<sup>72</sup> Our previous work showed evidence for p38 (MAPK) activation in a subset of transformed follicular lymphomas.<sup>73</sup> The p38 MAPK is a key mediator of stress, extracellular-, growth factor-, and cytokine-induced signaling and has been implicated in the development of some cancers.<sup>74,75</sup> Using a selective inhibitor of p38 MAPK (SB203580), we demonstrated that p38 MAPK inhibition resulted in dose- and time-dependent caspase-3-mediated apoptosis in transformed follicular lymphoma-derived cell lines. In order to further elucidate the basis of the cellular effects of SB203580, we have employed a systems biologic approach involving quantitative



**Figure 9** Outline of experimental protocol used for differential protein expression profiling by ICAT™. Protein mixtures obtained from two cell populations are either labeled with light or heavy isotopic versions of the cleavable ICAT™ reagent. Labeled proteins are combined, subject to multidimensional separation by SCX, RP and avidin affinity chromatography and analyzed by tandem mass spectrometry for peptide and protein identification. Based on the relative ratio of the two isotopically labeled peptides, a relative abundance of protein expression can be determined.

**Table 1** Comparison of methods for quantitative protein expression profiling

Parameters	2-Dimensional gel	Isotope-coded affinity tag	Stable isotope labeling in cell culture
Time	1–2 days for 2D gel, followed by spot identification and MS	Up to 1 week for labeling, multidimensional separation followed by MS	1–2 weeks to grow cells in special media
Samples	One sample analyzed per gel	Pairs of samples analyzed	Pairs of samples analyzed
Amount of protein required	50–100 µg	10 µg–1 mg	1 mg
Potential for automation	Yes	Yes	Limited
Type of samples	Cell lines, biopsy tissues	Cell lines, biopsy tissues	Viable cells grown in culture
Sensitivity	Moderate	High	High
Average number of proteins identified and quantitated per experiment	50–70	300–500	300–500
Limitations	Poor sensitivity for proteins with extremes of pI, and size. Membrane proteins are poorly detected	ICAT reagent is costly	Direct quantitative comparisons are most feasibly done using viable cells

proteomic analysis of transformed follicular lymphoma-derived cells treated with SB203580. Quantitative proteomic analysis using ICAT™-LC/MS/MS identified 277 differentially expressed proteins at 3 h and 350 proteins at 21 h of treatment with SB203580, the majority of which were also down-regulated. Analysis of functional groups of the differentially expressed proteins implicated components of diverse overlapping pathways including the IL-6/PI-3 K, IGF-2/Ras/Raf, WNT8d/WNT5a pathways affecting diverse cellular processes such as

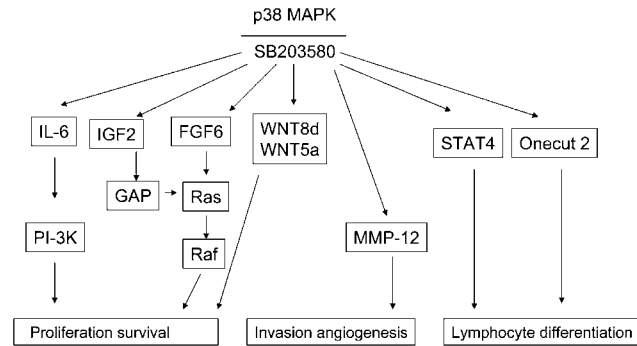
cell proliferation, survival, invasion, angiogenesis and lymphocyte differentiation (see Figure 10). Our proteomic approach revealed the global cellular consequences of SB203580 treatment and provides insights into its growth inhibitory effect on transformed follicular lymphoma cells.

The ability to perform global quantitative proteomics has been significantly enhanced by the advent of the isotope-coded affinity tag-based technology (ICAT™) which is efficient in simplifying the proteome, and in combination with LC-MS/MS

permits detection and quantification of proteins and peptides from very complex samples.

### Identification of Protein-Protein Interactions

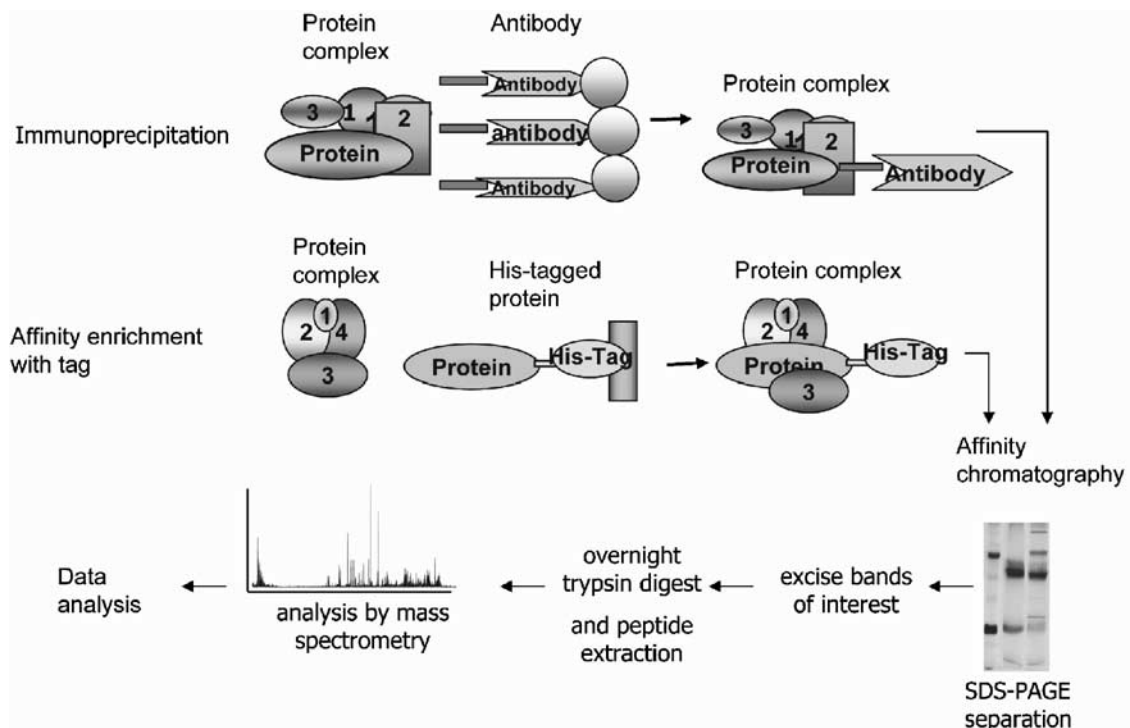
Until recently, assessment of interacting proteins has only been feasible via laborious and time-



**Figure 10** Quantitative proteomic analysis of the response to the p38 MAPK inhibitor (SB203580) in transformed follicular lymphoma cells using ICAT™ strategy. Proteins belonging to cell surface growth factor receptors, signaling molecules, and transcription factors were shown to be downregulated following exposure of transformed follicular lymphoma cells to the p38 MAPK inhibitor SB203580. The downregulation of these proteins would lead to alterations of multiple cellular effects including proliferation, survival, invasion, angiogenesis and lymphocyte development.

consuming molecular biologic techniques such as the yeast-two-hybrid system. Using a variety of approaches including protein complex purification, immunoprecipitation, affinity chromatography<sup>76</sup> followed by high-performance liquid chromatography (HPLC) and electrospray ionization and tandem mass spectrometry (ESI-MS/MS), interacting proteins of CD4 receptor complex, PCNA,<sup>77</sup> nonmuscle myosin heavy chain II<sup>78</sup> and protein kinase C signaling complex<sup>79</sup> have been identified. The advantage of such copurification protocols is that the fully processed protein which serves as the bait can allow interactions in a native environment and cellular location to allow multicomponent complexes to be isolated. A limitation is the availability of an antibody which allows specific immunoprecipitation. Proteins can also be tagged with a specific motif such as 6-Histidine or GST, which allows purification using affinity characteristics to Nickel or GSH beads, respectively. See Figure 11 for immunoprecipitation and affinity tag-based approaches to identify interacting protein partners.

Using a functional proteomic approach, our laboratory has determined the identity of proteins that interact with the NPM-ALK oncogenic tyrosine kinase by immunoprecipitation with anti-ALK antibody followed by electrospray ionization (ESI) and tandem mass spectrometry (MS/MS). The NPM-ALK fusion protein results from the t(2;5)(p23;q35)



**Figure 11** Methods used to determine proteins within multi-protein complexes. Antibody specific for a protein can be used for isolation of immunocomplexes. An affinity tag such as 6-Histidine (His6), glutathione S-transferase (GST) or FLAG can be attached to a target protein of interest and used as a 'bait'. Affinity chromatography is used to precipitate proteins associated with the affinity tagged proteins. Purified protein complexes are subsequently resolved by 1D SDS-PAGE, proteins excised, digested with enzyme and analyzed by mass spectrometry.

chromosomal aberration, which is characteristic of a subtype of T-cell lymphoma known as anaplastic large cell lymphoma.<sup>80–82</sup> Proteins that interact with ALK tyrosine kinase play important roles in mediating downstream cellular signals, and are potential targets for novel therapies. A total of 46 proteins were identified as unique to the ALK immunocomplex using monoclonal and polyclonal antibodies, while 11 proteins were identified in the NPM immunocomplex. Previously reported proteins in the ALK signal pathway were identified including PI3-K, Jak2, Jak3, Stat3, Grb2, IRS and PLC $\gamma$ 1. More importantly, many proteins previously not recognized to be associated with NPM-ALK, but with potential NPM-ALK interacting protein domains were identified. Proteins identified by MS were confirmed by Western blotting and reciprocal immunoprecipitation.<sup>83</sup>

Proteomic-based methods for the identification and quantitation of proteins associated with the chromatin<sup>84</sup> or specific transcription factor complexes have been developed. Using the ICAT technique of quantitative mass spectrometry, proteins associated with a transcription factor NF-E2p18/Mafk, essential for  $\beta$ -globin expression in murine erythroleukemia cells were determined<sup>85</sup> during two differentiation states.

### Post-Translational Modifications

Proteins are modified to their mature functional form through a highly regulated and complex sequence of post-translational processing. Most modifications are reversible and play important roles in regulating the biologic function of the protein. There are over 200 reported modes of post-translational modifications of proteins such as phosphorylation, glycosylation, ubiquitination to mention a few. Of these, methods to determine the type and sites of protein phosphorylation have received significant attention. Reversible phosphorylation of proteins is a key event in signal transduction from extracellular stimulus via a transmembrane receptor to the nucleus.<sup>86</sup> Phosphorylation occurs mainly on serine, threonine and tyrosine residues with the ratio being 1800:200:1 in vertebrates.<sup>87</sup> Although the phosphorylation of tyrosine residues is much less frequent in the proteome, it has been most extensively studied.

More recently, attempts have been made to define the phosphorylation status of proteins at a global scale.<sup>88</sup> Most approaches involve the use of phospho-specific antibodies to enrich for proteins with phosphorylated residues. Owing to the availability of excellent antibodies that react with phosphotyrosines the analysis of tyrosine phosphoproteins far outnumber those for serine and threonine phosphoproteins. Using phospho-specific antibodies, serine/threonine-phosphorylated proteins were enriched by immunoprecipitation and identi-

fied by mass spectrometry. Functional studies led to the identification of a novel protein which was demonstrated to be a substrate of protein kinase A.<sup>89</sup> Similarly, a novel phosphotyrosine protein was identified and characterized to be an ITAM-containing signal transduction protein induced by epidermal growth factor stimulation.<sup>90</sup> Using 2-D gel electrophoresis and mass spectrometry Lim *et al*,<sup>91</sup> identified over 50 distinct tyrosine phosphoproteins induced by epidermal growth factor in a human epidermoid carcinoma cell line which may represent novel targets for therapy. Similarly, phosphotyrosine proteomes have been characterized from thrombin-activated platelets,<sup>92</sup> B-lymphoblasts<sup>93</sup> and in response to heat shock.<sup>94</sup> More large-scale studies of phosphoproteins have taken advantage of commercially available immobilized metal-ion affinity chromatography (IMAC), which allows enrichment of phosphopeptides<sup>95</sup> and also allows identification of several phosphorylation sites on single proteins.<sup>96</sup>

Peng *et al*,<sup>10</sup> have used affinity purification to enrich for ubiquitinated proteins of yeast and have identified over 1000 proteins that were ubiquitinated. Their ability to identify the site of ubiquitination of over 100 cases provides validation of their high throughput proteomic-based methodology.

### Cellular Subproteomics

One of the major initiatives of the Human Proteome Organization (HUPO) is the comprehensive characterization of the complete subproteome of each cell type. Defining the global fingerprint of proteins expressed in a given cell type will aid in the identification of deregulated proteins that are characteristic of certain disease states and aid in the diagnostic and prognostication.

#### Secretome

Protein secretion by diseased cell types may provide a means for earlier detection of cancer. Systematic approaches to purify and identify secreted proteins from a variety of cell types have been reported. Martin *et al*,<sup>97</sup> have used a combination of ICAT and tandem mass spectrometry to identify and quantify a comprehensive list of over 500 proteins, which were secreted from neoplastic prostate cancer cell line (LNCaP) in the presence or absence of androgen receptor stimulation. Similar studies have identified numerous secreted proteins during differentiation of 3T3-L1 preadipocytes to adipocytes,<sup>98</sup> in response to calcium-dependent secretion in human neutrophils,<sup>99</sup> during osteoclast differentiation<sup>100</sup> and astrocytes.<sup>101</sup>

#### Cell surface proteome

Proteomic approaches for the comprehensive profiling and identification of proteins on cell surface or membranes have been reported. Cell surface

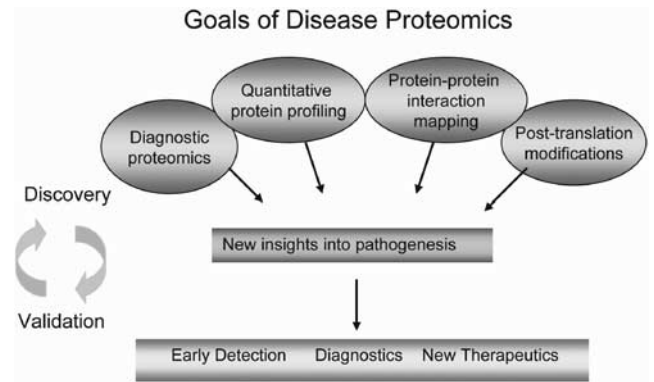
membranes may be subjected to biotinylation, affinity-captured and purified on avidin columns. Biotinylated intact proteins which were eluted can be separated by 2-D gel electrophoresis and protein spots from matching gels analyzed by mass spectrometry. Alternatively plasma-membrane fractions can be purified by sucrose-gradient centrifugation and analyzed by mass spectrometry. Using these approaches cell surface membrane proteins of human chronic lymphocytic leukemia,<sup>102</sup> acute leukemia<sup>103</sup> have been characterized. Analysis of the cell surface proteome of cancer cells of a variety of histologies has revealed an abundance of proteins with chaperone function such as GRP78, HSP70 and protein disulfide isomerase.<sup>104</sup>

### Organellar proteomics

Systematic analysis of the proteins expressed by specific intracellular organelles such as the cancer-cell mitochondria,<sup>105</sup> lymphocyte-derived exosomes,<sup>7</sup> the lipofuscin in human retinal pigment epithelial cells,<sup>106</sup> the phagosome,<sup>107</sup> microsomes from cancer cells,<sup>108</sup> lipid rafts<sup>109</sup> and human nucleolus<sup>110</sup> have provided extensive insights into the organellar proteomes that was previously impossible to obtain. These studies also demonstrate that even intracellular organelles as small as nucleoli are much more complicated in their protein expression profiles, and express hundreds of proteins some of which are not previously associated with their function. They also show that the protein composition of organelles is not static and can respond significantly to changes in the states of the cells.

## Conclusions

The recent advances in protein separation techniques, mass spectrometry and completion of the genome sequences of several organisms are all critical developments that facilitate proteomic studies. At the present time, substantial progress has been made in the development of high-throughput technologies for analysis of proteins including protein and antibody arrays. The challenges for the future lie in the archiving and integration of the vast amounts of data derived from mass spectrometry experiments into cohesive information relevant to physiologic and disease processes. To this end, it will be necessary to establish universally accepted criteria for positive identification of proteins by mass spectrometry. Further, it will be of considerable benefit to establish guidelines and standards for the reporting of proteomic data with adequate provisions for open access to proteomic data sets by investigators worldwide. It will also be critical to integrate the various aspects of proteomics such as diagnostic proteomics, high throughput analyses of post-translational modifications, imaging mass spectrometry, protein-protein interaction mapping and



**Figure 12** Overview of goals and challenges of disease proteomics. A complementary approach in which a variety of high throughput methodologies are utilized will be critical in elucidating new insights into the pathogenesis of diseases. Ongoing discovery and validation phases of clinical samples are necessary steps required to identify proteins that may function as protein biomarkers with diagnostic, prognostic and therapeutic relevance.

quantitative protein expression profiling to formulate coherent hypotheses regarding the pathogenesis of disease entities. These studies will have to occur in concert with large-scale validation studies on clinical samples obtained from well-controlled patient populations before they can be considered for potential diagnostic, prognostic or therapeutic biomarkers (Figure 12). In the near future, proteomic studies will impact the discovery of novel biomarkers and targets for the practical diagnosis and treatment of various diseases including malignancies.

## Acknowledgements

This work was financially supported by the ARUP Institute for Clinical and Experimental Pathology.

## References

- 1 Wasinger VC, Cordwell SJ, Cerpa-Poljak A, *et al.* Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 1995; 16:1090–1094.
- 2 Wilkins MR, Sanchez JC, Gooley AA, *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 1996;13:19–50.
- 3 Blackstock WP, Weir MP. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* 1999;17:121–127.
- 4 Yates III JR. Mass spectrometry and the age of the proteome. *J Mass Spectrom* 1998;33:1–19.
- 5 Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 2001;70:437–473.

- 6 Ahram M, Flaig MJ, Gillespie JW, *et al.* Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications. *Proteomics* 2003;3:413–421.
- 7 Wubbolts R, Leckie RS, Veenhuizen PT, *et al.* Proteomic and biochemical analyses of human B cell-derived exosomes. Potential implications for their function and multivesicular body formation. *J Biol Chem* 2003;278:10963–10972.
- 8 Lawrie LC, Curran S, McLeod HL, *et al.* Application of laser capture microdissection and proteomics in colon cancer. *Mol Pathol* 2001;54:253–258.
- 9 Li C, Hong Y, Tan YX, *et al.* Accurate qualitative and quantitative proteomic analysis of clinical hepatocellular carcinoma using laser capture microdissection coupled with isotope-coded affinity tag and two-dimensional liquid chromatography mass spectrometry. *Mol Cell Proteomics* 2004;3:399–409.
- 10 Peng J, Schwartz D, Elias JE, *et al.* A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* 2003;21:921–926.
- 11 Gygi SP, Corthals GL, Zhang Y, *et al.* Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci USA* 2000;97:9390–9395.
- 12 Link AJ, Eng J, Schieltz DM, *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 1999;17:676–682.
- 13 Washburn MP, Wolters D, Yates III JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 2001;19:242–247.
- 14 Washburn MP, Ulaszek R, Deciu C, *et al.* Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* 2002;74:1650–1657.
- 15 MacCoss MJ, McDonald WH, Saraf A, *et al.* Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci USA* 2002;99:7900–7905.
- 16 Gatlin CL, Kleemann GR, Hays LG, *et al.* Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal Biochem* 1998;263:93–101.
- 17 Fenn JB, Mann M, Meng CK, *et al.* Electrospray ionization for mass spectrometry of large biomolecules. *Science* 1989;246:64–71.
- 18 Tanaka K, Waki H, Ido Y, *et al.* Protein and polymer analyses up to  $m/z$  100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spec* 1998;2:151–153.
- 19 Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 1988;60:2301–2303.
- 20 Kuwata H, Yip TT, Yip CL, *et al.* Bactericidal domain of lactoferrin: detection, quantitation, and characterization of lactoferrin in serum by SELDI affinity mass spectrometry. *Biochem Biophys Res Commun* 1998;245:764–773.
- 21 Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 2000;21:1164–1177.
- 22 Yost RA, Boyd RK. Tandem mass spectrometry: quadrupole and hybrid instruments. *Methods Enzymol* 1990;193:154–200.
- 23 Li J, Kelly JF, Chernushevich I, *et al.* Separation and identification of peptides from gel-isolated membrane proteins using a microfabricated device for combined capillary electrophoresis/nanoelectrospray mass spectrometry. *Anal Chem* 2000;72:599–609.
- 24 Dainese P, Staudenmann W, Quadroni M, *et al.* Probing protein function using a combination of gene knockout and proteome analysis by mass spectrometry. *Electrophoresis* 1997;18:432–442.
- 25 Steen H, Kuster B, Mann M. Quadrupole time-of-flight versus triple-quadrupole mass spectrometry for the determination of phosphopeptides by precursor ion scanning. *J Mass Spectrom* 2001;36:782–790.
- 26 Chernushevich IV, Loboda AV, Thomson BA. An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* 2001;36:849–865.
- 27 Stafford G, Kelley P, Syka J, *et al.* Recent improvements in and analytical applications of advanced ion-trap technology. *Int J Mass Spectrom Ion Processes* 1984;60:85–98.
- 28 Krutchinsky AN, Kalkum M, Chait BT. Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. *Anal Chem* 2001;73:5066–5077.
- 29 Marshall AG, Hendrickson CL, Jackson GS. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass Spectrom Rev* 1998;17:1–35.
- 30 Bruce JE, Anderson GA, Wen J, *et al.* High-mass-measurement accuracy and 100% sequence coverage of enzymatically digested bovine serum albumin from an ESI-FTICR mass spectrum. *Anal Chem* 1999;71:2595–2599.
- 31 Solouki T, Fort Jr RC, Alomary A, *et al.* Gas phase hydrogen deuterium exchange reactions of a model peptide: FT-ICR and computational analyses of metal induced conformational mutations. *J Am Soc Mass Spectrom* 2001;12:1272–1285.
- 32 Wilkins MR, Gasteiger E, Wheeler CH, *et al.* Multiple parameter cross-species protein identification using MultiIdent—a world-wide web accessible tool. *Electrophoresis* 1998;19:3199–3206.
- 33 Zhang W, Chait BT. ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal Chem* 2000;72:2482–2489.
- 34 MacCoss MJ, Wu CC, Yates III JR. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem* 2002;74:5593–5599.
- 35 Nesvizhskii AI, Keller A, Kolker E, *et al.* A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75:4646–4658.
- 36 Von Haller PD, Yi E, Donohoe S, *et al.* The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: I. Statistically annotated datasets for peptide sequences and proteins identified via the application of ICAT and tandem mass spectrometry to proteins copurifying with T cell lipid rafts. *Mol Cell Proteomics* 2003;2:426–427.
- 37 Von Haller PD, Yi E, Donohoe S, *et al.* The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: II. Evaluation of tandem mass spectrometry methodologies for large-scale protein analysis, and the application of statistical tools for data analysis and interpretation. *Mol Cell Proteomics* 2003;2:428–442.

- 38 Keller A, Nesvizhskii AI, Kolker E, *et al.* Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* 2002;74:5383–5392.
- 39 Petricoin EF, Zoon KC, Kohn EC, *et al.* Clinical proteomics: translating benchside promise into bedside reality. *Nat Rev Drug Discov* 2002;1:683–695.
- 40 Petricoin III EF, Ornstein DK, Paweletz CP, *et al.* Serum proteomic patterns for detection of prostate cancer. *J Natl Cancer Inst* 2002;94:1576–1578.
- 41 Petricoin III EF, Hackett JL, Lesko LJ, *et al.* Medical applications of microarray technologies: a regulatory science perspective. *Nat Genet* 2002;32(Suppl):474–479.
- 42 Petricoin EF, Ardekani AM, Hitt BA, *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 2002;359:572–577.
- 43 Rosenblatt KP, Bryant-Greenwood P, Killian JK, *et al.* Serum proteomics in cancer diagnosis and management. *Annu Rev Med* 2004;55:97–112.
- 44 Rui Z, Jian-Guo J, Yuan-Peng T, *et al.* Use of serological proteomic methods to find biomarkers associated with breast cancer. *Proteomics* 2003;3:433–439.
- 45 Li J, Zhang Z, Rosenzweig J, *et al.* Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. *Clin Chem* 2002;48:1296–1304.
- 46 Adam BL, Qu Y, Davis JW, *et al.* Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer Res* 2002;62:3609–3614.
- 47 Poon TC, Yip TT, Chan AT, *et al.* Comprehensive proteomic profiling identifies serum proteomic signatures for detection of hepatocellular carcinoma and its subtypes. *Clin Chem* 2003;49:752–760.
- 48 Coombes KR, Fritsche Jr HA, Clarke C, *et al.* Quality control and peak finding for proteomics data collected from nipple aspirate fluid by surface-enhanced laser desorption and ionization. *Clin Chem* 2003;49:1615–1623.
- 49 Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments. *Bioinformatics* 2004;20:777–785.
- 50 Paweletz CP, Trock B, Pennanen M, *et al.* Proteomic patterns of nipple aspirate fluids obtained by SELDI-TOF: potential for new biomarkers to aid in the diagnosis of breast cancer. *Dis Markers* 2001;17:301–307.
- 51 Fetsch PA, Simone NL, Bryant-Greenwood PK, *et al.* Proteomic evaluation of archival cytologic material using SELDI affinity mass spectrometry: potential for diagnostic applications. *Am J Clin Pathol* 2002;118:870–876.
- 52 Lin Z, Jenson S, Lim M, *et al.* Application of SELDI-TOF mass spectrometry for the identification of differentially expressed proteins in transformed follicular lymphoma. *Mod Pathol* 2004;17:670–678.
- 53 Xu BJ, Caprioli RM, Sanders ME, *et al.* Direct analysis of laser capture microdissected cells by MALDI mass spectrometry. *J Am Soc Mass Spectrom* 2002;13:1292–1297.
- 54 Stoeckli M, Chaurand P, Hallahan DE, *et al.* Imaging mass spectrometry: a new technology for the analysis of protein expression in mammalian tissues. *Nat Med* 2001;7:493–496.
- 55 Chaurand P, Schwartz SA, Caprioli RM. Imaging mass spectrometry: a new tool to investigate the spatial organization of peptides and proteins in mammalian tissue sections. *Curr Opin Chem Biol* 2002;6:676–681.
- 56 Chaurand P, Caprioli RM. Direct profiling and imaging of peptides and proteins from mammalian cells and tissue sections by mass spectrometry. *Electrophoresis* 2002;23:3125–3135.
- 57 Yanagisawa K, Shyr Y, Xu BJ, *et al.* Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet* 2003;362:433–439.
- 58 Yanagisawa K, Xu BJ, Carbone DP, *et al.* Molecular fingerprinting in human lung cancer. *Clin Lung Cancer* 2003;5:113–118.
- 59 Friedman DB, Hill S, Keller JW, *et al.* Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* 2004;4:793–811.
- 60 Chaurand P, Schwartz SA, Billheimer D, *et al.* Integrating histology and imaging mass spectrometry. *Anal Chem* 2004;76:1145–1155.
- 61 Schwartz SA, Weil RJ, Johnson MD, *et al.* Protein profiling in brain tumors using mass spectrometry: feasibility of a new technique for the analysis of protein expression. *Clin Cancer Res* 2004;10:981–987.
- 62 Rabilloud T. Detecting proteins separated by 2-D gel electrophoresis. *Anal Chem* 2000;72:48A–55A.
- 63 Zhou G, Li H, DeCamp D, *et al.* 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol Cell Proteomics* 2002;1:117–124.
- 64 Lian Z, Wang L, Yamaga S, *et al.* Genomic and proteomic analysis of the myeloid differentiation program. *Blood* 2001;98:513–524.
- 65 Gerner C, Frohwein U, Gotzmann J, *et al.* The Fas-induced apoptosis analyzed by high throughput proteome analysis. *J Biol Chem* 2000;275:39018–39026.
- 66 Tan S, Seow TK, Liang RC, *et al.* Proteome analysis of butyrate-treated human colon cancer cells (HT-29). *Int J Cancer* 2002;98:523–531.
- 67 Ong SE, Blagoev B, Kratchmarova I, *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 2002;1:376–386.
- 68 Ong SE, Kratchmarova I, Mann M. Properties of <sup>13</sup>C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* 2003;2:173–181.
- 69 Gygi SP, Rist B, Gerber SA, *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 1999;17:994–999.
- 70 Schirmer EC. MudPIT: A powerful proteomics tool for discovery. *Discov Med* 2003;3:38–39.
- 71 Hardwidge PR, Rodriguez-Escudero I, Goode D, *et al.* Proteomic analysis of the intestinal epithelial cell response to enteropathogenic *Escherichia coli*. *J Biol Chem* 2004;279:20127–20136.
- 72 Lin Z, Crockett DK, Jenson SD, *et al.* Quantitative proteomic and transcriptional analysis of the response to the p38 MAPK inhibitor SB203580 in

- transformed follicular lymphoma cells. *Mol Cell Proteomics* 2004 (in press).
- 73 Elenitoba-Johnson KS, Jenson SD, Abbott RT, *et al.* Involvement of multiple signaling pathways in follicular lymphoma transformation: p38-mitogen-activated protein kinase as a target for therapy. *Proc Natl Acad Sci USA* 2003;100:7259–7264.
  - 74 Xu Q, Karouji Y, Kobayashi M, *et al.* The PI 3-kinase-Rac-p38 MAP kinase pathway is involved in the formation of signet-ring cell carcinoma. *Oncogene* 2003;22:5537–5544.
  - 75 Hardwick JC, van den Brink GR, Offerhaus GJ, *et al.* NF-kappaB, p38 MAPK and JNK are highly expressed and active in the stroma of human colonic adenomatous polyps. *Oncogene* 2001;20:819–827.
  - 76 Kumar A, Snyder M. Protein complexes take the bait. *Nature* 2002;415:123–124.
  - 77 Ohta S, Shiomi Y, Sugimoto K, *et al.* A proteomics approach to identify proliferating cell nuclear antigen (PCNA)-binding proteins in human cell lysates. Identification of the human CHL12/RFCs2-5 complex as a novel PCNA-binding protein. *J Biol Chem* 2002;277:40362–40367.
  - 78 Rey M, Vicente-Manzanares M, Viedma F, *et al.* Cutting edge: association of the motor protein non-muscle myosin heavy chain-IIA with the C terminus of the chemokine receptor CXCR4 in T lymphocytes. *J Immunol* 2002;169:5410–5414.
  - 79 Edmondson RD, Vondriska TM, Biederman KJ, *et al.* Protein kinase C epsilon signaling complexes include metabolism- and transcription/translation-related proteins: complimentary separation techniques with LC/MS/MS. *Mol Cell Proteomics* 2002;1:421–433.
  - 80 Morris SW, Kirstein MN, Valentine MB, *et al.* Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* 1994;263:1281–1284.
  - 81 Shiota M, Fujimoto J, Semba T, *et al.* Hyperphosphorylation of a novel 80 kDa protein-tyrosine kinase similar to Ltk in a human Ki-1 lymphoma cell line, AMS3. *Oncogene* 1994;9:1567–1574.
  - 82 Shiota M, Fujimoto J, Takenaga M, *et al.* Diagnosis of t(2;5)(p23;q35)-associated Ki-1 lymphoma with immunohistochemistry. *Blood* 1994;84:3648–3652.
  - 83 Crockett DK, Lin Z, Elenitoba-Johnson KS, *et al.* Identification of NPM-ALK interacting proteins by tandem mass spectrometry. *Oncogene* 2004;23:2617–2629.
  - 84 Shiio Y, Eisenman RN, Yi EC, *et al.* Quantitative proteomic analysis of chromatin-associated factors. *J Am Soc Mass Spectrom* 2003;14:696–703.
  - 85 Brand M, Ranish JA, Kummer NT, *et al.* Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol* 2004;11:73–80.
  - 86 Pawson T, Scott JD. Signaling through scaffold, anchoring, and adaptor proteins. *Science* 1997;278:2075–2080.
  - 87 Hunter T. The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philos Trans R Soc Lond B Biol Sci* 1998;353:583–605.
  - 88 Mann M, Ong SE, Gronborg M, *et al.* Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* 2002;20:261–268.
  - 89 Gronborg M, Kristiansen TZ, Stensballe A, *et al.* A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics* 2002;1:517–527.
  - 90 Pandey A, Fernandez MM, Steen H, *et al.* Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. *J Biol Chem* 2000;275:38633–38639.
  - 91 Lim YP, Diong LS, Qi R, *et al.* Phosphoproteomic fingerprinting of epidermal growth factor signaling and anticancer drug action in human tumor cells. *Mol Cancer Ther* 2003;2:1369–1377.
  - 92 Maguire PB, Wynne KJ, Harney DF, *et al.* Identification of the phosphotyrosine proteome from thrombin activated platelets. *Proteomics* 2002;2:642–648.
  - 93 Imam-Sghiouar N, Laude-Lemaire I, Labas V, *et al.* Subproteomics analysis of phosphorylated proteins: application to the study of B-lymphoblasts from a patient with Scott syndrome. *Proteomics* 2002;2:828–838.
  - 94 Kim HJ, Song EJ, Lee KJ. Proteomic analysis of protein phosphorylations in heat shock response and thermotolerance. *J Biol Chem* 2002;277:23193–23207.
  - 95 Ficarro SB, McClelland ML, Stukenberg PT, *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* 2002;20:301–305.
  - 96 Fuglsang AT, Visconti S, Drumm K, *et al.* Binding of 14-3-3 protein to the plasma membrane H(+)-ATPase AHA2 involves the three C-terminal residues Tyr(946)-Thr-Val and requires phosphorylation of Thr(947). *J Biol Chem* 1999;274:36774–36780.
  - 97 Martin DB, Gifford DR, Wright ME, *et al.* Quantitative proteomic analysis of proteins released by neoplastic prostate epithelium. *Cancer Res* 2004;64:347–355.
  - 98 Kratchmarova I, Kalume DE, Blagoev B, *et al.* A proteomic approach for identification of secreted proteins during the differentiation of 3T3-L1 preadipocytes to adipocytes. *Mol Cell Proteomics* 2002;1:213–222.
  - 99 Boussac M, Garin J. Calcium-dependent secretion in human neutrophils: a proteomic approach. *Electrophoresis* 2000;21:665–672.
  - 100 Kubota K, Wakabayashi K, Matsuoka T. Proteome analysis of secreted proteins during osteoclast differentiation using two different methods: two-dimensional electrophoresis and isotope-coded affinity tags analysis with two-dimensional chromatography. *Proteomics* 2003;3:616–626.
  - 101 Lafon-Cazal M, Adjali O, Galeotti N, *et al.* Proteomic analysis of astrocytic secretion in the mouse. Comparison with the cerebrospinal fluid proteome. *J Biol Chem* 2003;278:24438–24448.
  - 102 Boyd RS, Adam PJ, Patel S, *et al.* Proteomic analysis of the cell-surface membrane in chronic lymphocytic leukemia: identification of two novel proteins, BCNP1 and MIG2B. *Leukemia* 2003;17:1605–1612.
  - 103 Jang JH, Hanash S. Profiling of the cell surface proteome. *Proteomics* 2003;3:1947–1954.
  - 104 Shin BK, Wang H, Yim AM, *et al.* Global profiling of the cell surface proteome of cancer cells uncovers an

- abundance of proteins with chaperone function. *J Biol Chem* 2003;278:7607–7616.
- 105 Verma M, Kagan J, Sidransky D, *et al.* Proteomic analysis of cancer-cell mitochondria. *Nat Rev Cancer* 2003;3:789–795.
- 106 Schutt F, Ueberle B, Schnolzer M, *et al.* Proteome analysis of lipofuscin in human retinal pigment epithelial cells. *FEBS Lett* 2002;528:217–221.
- 107 Garin J, Diez R, Kieffer S, *et al.* The phagosome proteome: insight into phagosome functions. *J Cell Biol* 2001;152:165–180.
- 108 Wright ME, Eng J, Sherman J, *et al.* Identification of androgen-coregulated protein networks from the microsomes of human prostate cancer cells. *Genome Biol* 2003;5:R4.
- 109 Foster LJ, De Hoog CL, Mann M. Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc Natl Acad Sci USA* 2003;100:5813–5818.
- 110 Andersen JS, Lyon CE, Fox AH, *et al.* Directed proteomic analysis of the human nucleolus. *Curr Biol* 2002;12:1–11.