ELSEVIER

Journal of Chromatography B, 771 (2002) 33–48

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

Proteomic tools for biomedicine

Valerie C. Wasinger*, Garry L. Corthals

Mass Spectrometry and Protein Analysis Laboratory, Garvan Institute of Medical Research, 384 Victoria Street, Sydney, NSW 2010, Australia

Abstract

Proteomic tools measure gene expression, protein activity and interactions of biological events at the protein level. Proteins are the major catalysts of biological functions and contain several dimensions of information that collectively indicate the actual rather than the potential functional state as indicated by mRNA analysis. Measurements can be made in terms of protein quantity, location, and time-point. For the future we see a further integration of existing and new technologies for proteomics from a wide range of areas of biochemistry, chemistry, physics, computing science and molecular biology. This will further advance our knowledge of how biological systems are built up and what mechanisms control these systems. However, the potential of proteomics to comprehensively answer all biological questions is limited as only protein activity is measured. A unification of genomics, proteomics, and other technologies is needed if we are to start to understand the complexity of biological function in the context of disease and health.

Keywords: Reviews; Proteomic; Biomedicine

Contents

*Corresponding author. Tel.: +61-2-9295-8195; fax: +61-2-9295-8194.
E-mail address: v.wasinger@garvan.org.au (V.C. Wasinger).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00125-3

1. Introduction

Until recently, proteomics was almost exclusively associated with the display of large numbers of proteins using two-dimensional electrophoresis (2-DE). Now however, the field of proteomics is undergoing rapid ''evaluation'', where old technologies are being refined to become downstream compatible as part of integrated analytical systems. We are also witnessing an enormous growth of new technologies allowing faster, more reliable, more comprehensive studies than ever before. This rapid growth and rise in applications of proteomic research has been enabled through the already impressive work done with genomic technologies. Genomic and proteomic research now complement each other and are partners for comprehensive large-scale studies that will contribute to the understanding of disease and health processes and drug action. Proteomics now incorporates multiple technologies including: differential display analysis by gels or chips; protein and peptide separation and prefractionation to extend coverage of analysis, or specifically select classes of proteins; measurement of relative abundance through affinity based quantitation; and sophisticated algorithms for database interrogation to allow accurate and rapid determination of proteins and validate and determine possible protein–protein interactions. The focus now is to apply cutting-edge proteomic technologies so that a robust analysis of complex biological systems becomes timely, inexpensive and widespread.

2. Defining proteomics

After decades of measuring protein expression and investigating function, using a variety of techniques to separate, visualise and identify proteins, the field of proteomics has emerged as a robust and global approach to analyse protein expression. There are hundreds of ''proteome projects'' (PubMed search revealed 828; for some examples: http://au.expasy.org/ch2d/) where one performs relative quantitative measurements of the phenotype and correlates these with the genotype, or where the level and quality of protein interactions are defined. Proteome applications are wide and now span all areas of biology from agricultural to medical science.

While the ''proteome'' was originally defined as the total protein complement of the genome [1,2], it was clear from an early stage that a total analysis of the proteome was impractical to realise because:

(i) The use of two-dimensional (2-D) gels combined with current technologies would not support the visualization and or identification of all proteins [3].

(ii) Much of the genetic code of organisms is translated at differing stages of cellular development [4,5].

(iii) Complexity is increased by regulation most of which is done post RNA translation, post-translational modifications of proteins, and cleavage and splice possibilities of genes increase the complexity of the protein observed making complete analysis difficult, if not (currently) impossible [6–9].

(iv) There are differences in the expression of proteins between different organelles, cell types, organs and whole organisms [10].

(v) Similar phenotypes can be attained from alternative protein pathways within cellular networks, which can be influenced by disease, environmental, internal, and biochemical stimuli [11–14].

A schematic of phenotype in relation to cellular processes and the proteome is provided in Fig. 1. Due to the inability to capture all expressed proteins within a cell for a particular time point, proteomics has undergone a logical rationalisation into three main research areas, defined by various applications where technology development is an enabling factor in all areas. These areas target: expression

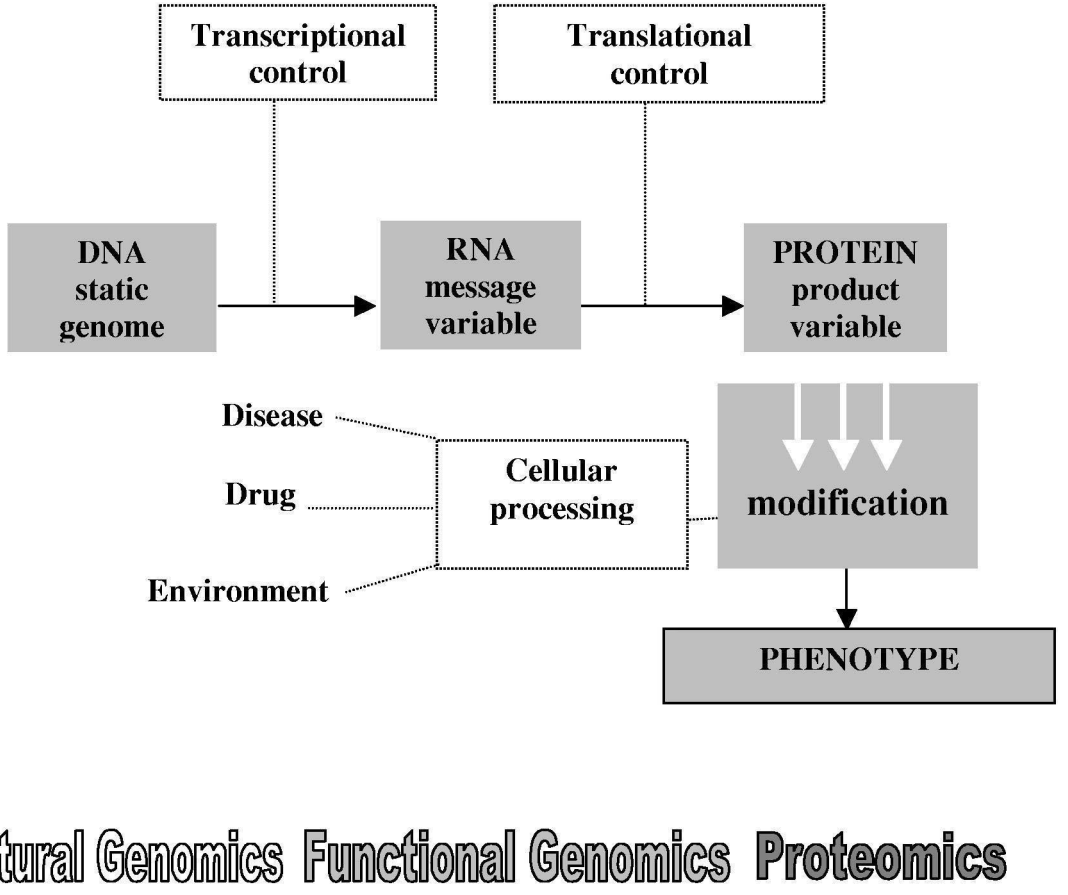

Fig. 1. Schematic of relationship between genome, transcriptome and proteome.

proteomics; subcellular map proteomics; and modular proteomics.

2.1. Applications of expression proteomics

Expression proteomics is conceptually similar to DNA microarray data. Microarrays allow rapid measurements on global changes in gene expression. However, the information content for mRNA expression and protein expression can be quite different. Recently there have been several reports that show a poor correlation between RNA expression and protein expression [15,16]. This implies that the information obtained for a pathological condition derived at the level of gene activity is incomplete [16,17], or at most only reflects the potential state of a system, whereas measurements of proteins reflect the actual state of a system.

The study of global changes in protein expression between diseased and healthy states, or cells perturbed by drug or other stimuli, is called expression proteomics [18–20]. Multigenic diseases such as cancer, diabetes and heart disease, account for around 98% of diseases [21]. Diagnosis and characterization of multigenic disease is difficult due to the complex interactions and epigenetic events that circumvent affected pathways [22]. Discovery of disease markers therefore requires a global approach, and comparative expression maps allow proteins with increased or decreased abundance to be identified and quantitated. By measuring the expression of large numbers of proteins, the coordinated changes in expression of disease-specific proteins

can be monitored. This is currently most commonly performed by a combination of 2-DE and mass spectrometry (MS). However, recently isotope coded affinity tags (ICATs) [23] and difference gel electrophoresis (DIGE) [24] have been used to accurately and sensitively display the differential expression of proteins on a large scale (discussed below). Expression proteomics is not limited to these technologies. Various tools and technologies are being used to screen for new markers and mechanisms in many diseases. These include 2-DE, MS based approaches, serial analysis of gene expression (SAGE), cell-line specific expressed sequence tags (EST) and cDNA array analysis.

One of the leading causes of cancer deaths in men stems from prostate cancer [25]. Diagnosis of prostate cancer can be made via blood stream released prostate specific antigen (PSA). However, there are many patients with clinically localized cancer who do not have an elevated serum PSA level, and others who have high serum PSA and do not have prostate cancer. Therefore, more accurate and sensitive biomarkers are being sought. Studies that have used proteomic approaches include 2-DE and MS [26], and SAGE [27] combined with 2-DE and MS [28,29]. The well established procedure undertaken to identify differentially expressed proteins by 2-DE and MS is shown in Fig. 2. Androgen deprivation is a common therapy for prostate cancer. By comparing the protein profile of two cell states differing only in their stimulation with a synthetic androgen, the androgen-mediated regulation of protein expression could be established using 2-DE and MS. Nucleoside diphosphate kinase A (NDKA/nm23) was differentially expressed and identified. This protein has tumor metastasis suppressor activity both in prostate

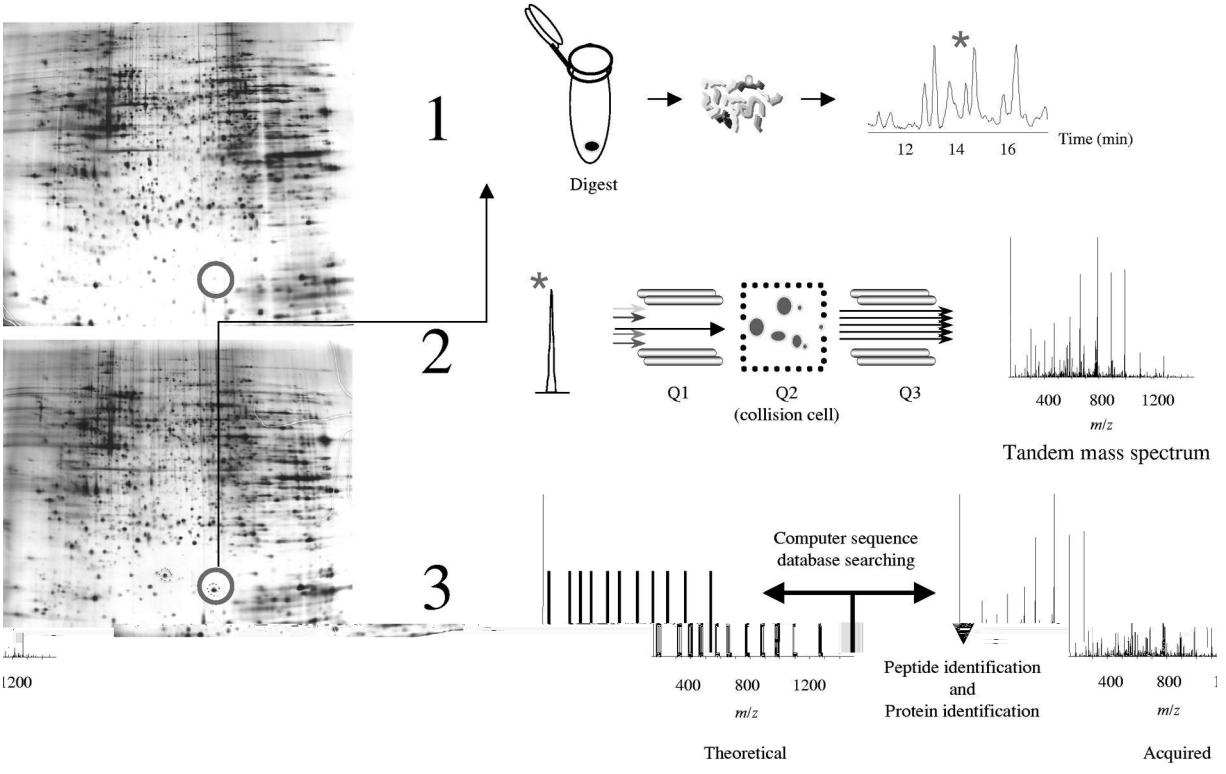

Fig. 2. Identification of androgen-regulated protein from metastatic prostate cancer cells by 2-DE and MS. M12AR cells were (A) starved or (B) stimulated with synthetic androgen. Protein expression profiles of whole cell lysates were compared and proteins undergoing qualitative expression differences were digested with trypsin (1). Peptides were separated using μLC–MS/MS (2) and identified by sequence database searching (3). Modified from Nelson et al. [28].

and in several different human tumors [28]. In a second study on androgen-regulated genes, 32 proteins from 1032 were found to be altered in their expression, while 351 genes (from 16 570) were found to have changed in expression [29], making obvious the necessity to obtain many levels of information by applying multiple technologies.

2.2. Applications of subcellular maps

Organelles such as mitochondria, lysozyme, endoplasmic reticulum, golgi and microsome, have numerous specific functions within the cell. The presence and activity of proteins in organelles reflects these functions and mapping their protein distribution will give insight into the spatial distribution and localized protein activity. From a technological perspective, an important benefit of subcellular fractionation is that sample complexity is decreased, enhancing the ability to comprehensively analyse selected organelles. The discovery of novel proteins and the potential for localising disease pathways that are specific to different organelles has prompted research using 2-DE. Examples include human placental mitochondria [30], placental lysozymes [31], a mitochondrial ribosome [32], and golgi [33]. The golgi complex is involved in sorting, packaging, redistributing, secreting and recycling proteins. Such a composite role is reflected in the protein repertoire. Wu et al. [33] have identified several classes of proteins involved in the regulation of membrane fusion and secretion using 2-DE and liquid chromatography (LC)–MS to obtain a global functional screen between basal and maximal secretory golgi states in rat mammary glands.

The microsomal fraction consists of proteins that are integral or associated with lipid membranes. These proteins make up pores, transporters and receptors and therefore mediate cell-specific interactions. Han et al. [34] have quantitated the differentially expressed microsomal proteins following exposure of the HL-60 cell line to 12-phorbol 13-myristate acetate (PMA). Here, ICAT, multi-dimensional chromatography and automated MS/MS were used to record the intracellular distribution of proteins during cellular differentiation and attachment. Membrane proteins are of considerable therapeutic value as has been exemplified by Her2/neu, a gene whose over expression is implicated in breast cancer tumorigenesis and whose modulated activity is commonly used as a diagnostic marker for breast cancer [35]. A humanised monoclonal antibody, called *Herceptin* provides passive immunity by recognising Her2/neu receptors. The Her2/neu receptor and other receptors are exposed on the membrane and thus can become therapeutic targets.

2.3. Applications of modular maps

An emerging paradigm in protein biology involves the systematic identification of proteins that interact with each other at a ‘‘biological level’’. Proteins can be defined in the context of their interactions within a network to provide a unique characterisation of associations within whole proteomes. Protein domains and their interactions and associations increase the complexity in higher organisms. Many biological regulatory events result through protein–protein interactions. Even mis-aggregation and incorrect interaction such as the formation of amyloid beta protein can cause the onset and progression of illnesses such as Alzheimer’s [7,36,37]. Establishing protein interaction maps is still difficult but is not impossible. The complexity of this task can be appreciated by the ~300 000 interactions possible within a single yeast cell [20,38,39]. A representation of a protein interaction map is pictured in Fig. 3 [40].

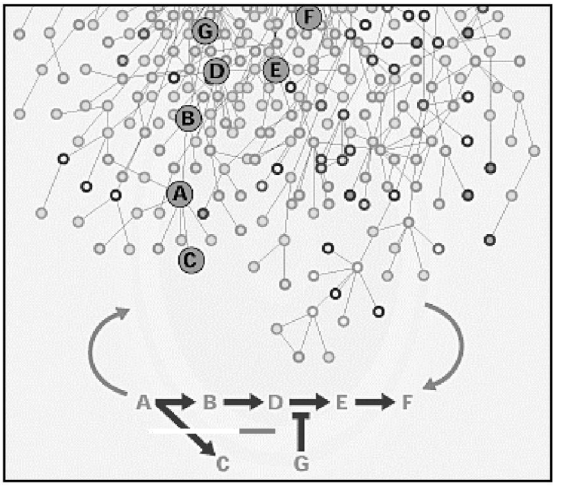

Fig. 3. A protein–protein interaction map showing many additional interactions for each of the proteins in the pathway [40].

Modular proteomics relies on many technologies. Two reasons for this are the relatively low physiological concentration of proteins inside the cells, and the difficulty of attaining a multi-protein state in its ''biological form''. The use of high affinity antibodies and tags [41] for enrichment of complex components to study modular proteomics has been reviewed recently [42]. The combination of a number of techniques have been used to identify proteins and peptides, and include the use of: 2-D chromatography and MS for the direct analysis of protein complexes from mixtures to characterize the 80S ribosomal subunits [43,44], and to identify peptides displayed on major histocampatibility complex (MHC) 1 associated with *Trypanasoma cruzi* infection and the related disease ''Chagas'' [45]; solid-phase extraction (SPE) coupled with capillary electrophoresis (CE) and MS for the elucidation of ~90% of the yeast ribosomal multi-protein complex [46]; and several different chromatographic steps followed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and MS used to identify 40 components of the nuclear pore complex (NPC) [47].

Recently, two similar methods that are particularly effective at identifying protein complexes were reported [48,49]. Both have been used in ''large-scale'' identification projects and can provide information on three or more proteins in complexes. The methods involve ''co-precipitation/mass spectrometry''. With these methods ''affinity tags'' are attached to target protein(s) (bait). The bait proteins are then purified on an affinity column and then separated by SDS–PAGE. Proteins are then treated as described in Fig. 2.

In another study that involved measuring perturbations in the yeast galactose utilisation pathway using DNA microarrays and quantitative proteomics, and taking information from databases of known physical interactions, a refined model was suggested for the yeast galactose utilisation pathway [15]. This work did not use affinity tags for protein associations, however does offer high-throughput analysis through the use of ICATs (for the proteomic component). The study is rich in proteomic and genomic data and integration of this data reveals a physical-interaction network. Thus, while no co-precipitation was used, important information was obtained for protein–DNA binding and protein–protein interactions, and the effect of biological function is suggested.

Protein–protein interactions can also be predicted via computational methods, as protein families originate from a common ancestor and therefore have similar features. Domain fusion or ''Rosetta stone'' sequences of fused genes have been used to computationally predict protein–protein interactions within signaling pathways, structural complexes and metabolic pathways [50,51]. Here, fused domains of a protein in one organism are used to predict interaction of these domains separated through evolution in another organism. Other computational methods for studying protein associations include: (i) phylogenetic profiling, based on the co-occurrence of proteins in different genomes and the presence or absence of a pair of proteins to predict physical associations [52]; (ii) mRNA expression and association, where proteins with the same expression levels over a series of conditions may be functionally linked; (iii) gene clustering in one organism where the genes have been separated by evolution in another organism [53,54].

3. Molecular analysis tools

In all areas of proteomics, sophisticated computer programs and highly sensitive and robust technology is required to analyse proteins and their biochemical properties and how these proteins interact. The field of proteomics is therefore now incorporating techniques previously only associated with other fields, and there has been a rush of new ''integrated'' technologies. An overview of current proteomics technologies is given in Fig. 4.

3.1. Separation/prefractionation technologies

3.1.1. Multi-dimensional separation: liquid based

Alternative approaches to 2-DE such as chromatography are being used for proteomics studies [55–58]. Most notably, Link et al. [56] described an integrated system that employed a biphasic two-dimensional micro (μ)LC column packed with strong cation-exchange (SCX) and reversed-phase (RP) materials. This enabled a step-wise analysis of peptides initially by charge and then by hydropho-

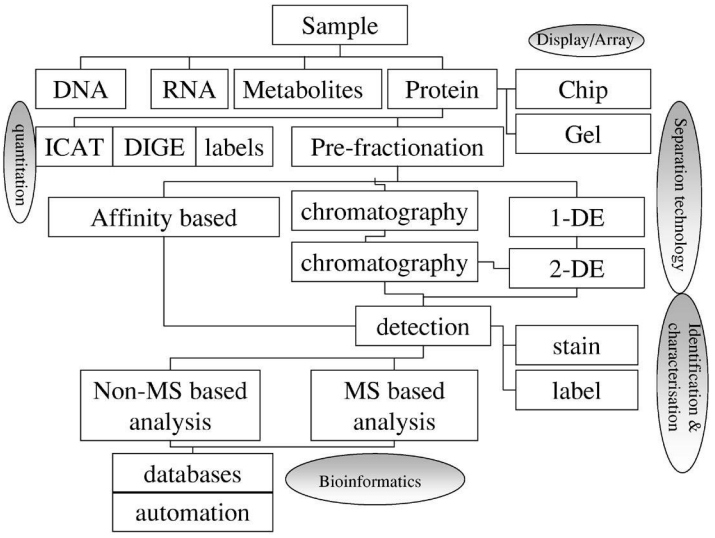

Fig. 4. Schematic of the integration of technologies for proteomics including fractionation, display, quantitation, identification and informatics. The unification of DNA, mRNA, and metabolite information is also essential for a complete understanding of biological function.

bicity. The multi-dimensional protein identification technology (MuDPIT) system was built in-line with a tandem mass spectrometer. With this approach proteins of high (>100 kDa)/low (<10 kDa) mass and of relatively low abundance (codon bias <0.2) and solubility could be identified; thereby extending the analysis range possible because proteins and peptides are separated by different physicochemical properties than with 2-DE. Recently, Washburn et al. [57] accounted for 1,484 yeast proteins using the MuDPIT approach. Multi-dimensional approaches can combine size-exclusion [59], ion-exchange [60], immobilised metal affinity [61], reversed-phase chromatography or capillary electrophoresis [62] and are either directly integrated with an MS instrument or run off-line prior to MS identification [63]. The process is represented in Fig. 5. By coupling these high-resolution separation techniques directly to a tandem MS system, the sampling losses are decreased allowing for greater sequence coverage.

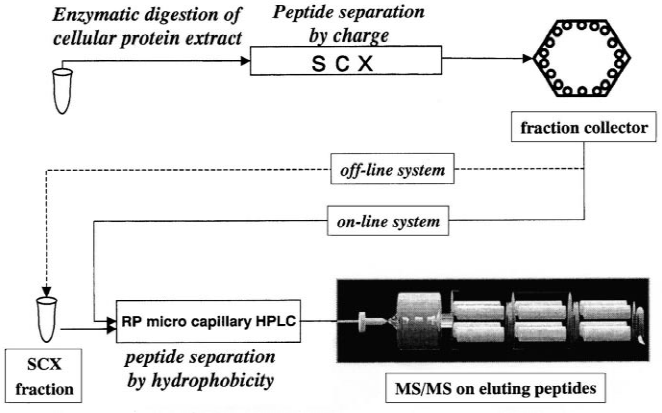

Fig. 5. Schematic of two-dimensional chromatography coupled to mass spectrometry for the identification of complex protein mixtures.

Integrated approaches are widely used to identify phosphorylation sites as this type of work often requires enrichment of phosphopeptides prior to LC–MS/MS analysis. In the past we have identified nitric-oxide synthase (eNOS) phosphopeptides by using immobilised metal affinity chromatography (IMAC) for pre-enrichment of phosphopeptides and performing subsequent solid-phase extraction–capillary electrophoresis (SPE–CE) connected on-line with an electrospray ionization (ESI) MS/MS instrument [64]. In this particular study we found ''peak parking'' during CE (i.e., data dependent modulation of electrophoretic voltage) to be extremely advantageous for the detection of very low abundance phosphopeptides by data-dependent MS/MS [65]. This advantage is achieved because a reduction of the voltage reduces the electrophoretic mobility and therefore enhances the time available for the mass spectrometer to detect and select for collision induced dissociation (CID) peptides without adversely affecting sensitivity [66]. We have applied this technique to the analysis of in vivo phosphorylation sites of endothelial nitric oxide synthase. For a review of the use of integrated IMAC system see elsewhere [67].

3.1.2. Function based selection of proteins

The number and extent of protein modifications is unclear for proteome studies, although the percentage of modified proteins is expected to be very high. It is of extreme importance to find out which modifications play a role in biological function. Of the 200 different types of protein modifications that have been described [68] to date only a few have been shown to be reversible and thus of regulatory importance in biological processes. Protein phosphorylation receives the most attention because the protein phosphorylation/de-phosphorylation process alters the structure and consequently affects the function of target proteins [69–71]. Recently we have been developing a new technique which involves double isoelectric focusing (dIEF). With this method isoelectric focusing is performed in both first and second dimensions. In the first dimension samples are separated in immobilised pH gradients (IPGs) according to conventional methods. Between the first and the second dimension, the sample is (bio) chemically treated to selectively alter the isoelectric point (p*I*) for a particular class of proteins. An IPG gel plate rather than an SDS–PAGE gel is used in the second dimension. Thus, instead of an array of protein spots as with IEF/SDS–PAGE methods, proteins will focus in a diagonal line across the gel plate, except for those proteins that have undergone a shift in p*I* from the first to the second dimension (Fig. 6). This method allows proteins containing specific modifications to become outliers [72]. A similar principle using mass separation in

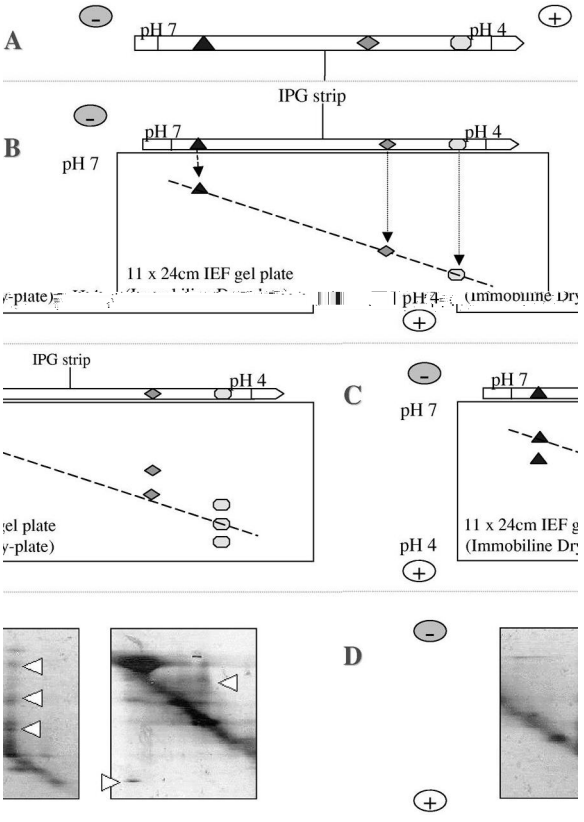

Fig. 6. The principles of dIEF. Panel A: A schematic diagram showing the first dimension of dIEF where proteins are separated according to their p*I* in an IPG strip; each protein is represented by circle, diamond and triangle. Panel B: Proteins are also separated according to their p*I* in the second dimension of dIEF. Proteins move from the IPG strip to the Immobiline dry-plate and form a diagonal line. Panel C: Following biochemical treatment, such as denaturing or reducing conditions, proteins that have undergone change from the first- to second-dimension will move off the resulting protein diagonal. Panel D: Native/reducing dIEF of yeast lysate. The arrows indicate where proteins focused away from the diagonal. Proteins were visualised by silver staining.

two dimensions has also been developed and has been particularly useful for isolating disulfide-bonded proteins in complex samples [73].

3.2. Protein arrays

3.2.1. Protein chip displays

In addition to high-throughput MS based technologies for proteome analyses, emerging microarray technologies are allowing for systematic analysis on a proteome-wide scale. Protein micro-arrays now exist that are chemically robust and stable, have a high binding efficiency and specificity [74], and are compact [75]. The number of bound peptides that can be synthesized per unit area has grown from the early applications involving 96-microtitrer plates [76] to the use of photolithographic techniques capable of over 250 000 elements/cm^2 [75]. Protein-to-protein based arrays include antibody, phage displayed antibody or polypeptide recognition moieties [74,77]. The global analysis of entire proteomes under carefully defined and controllable conditions such as pH, temperature and test binding molecules [78] is very attractive and facilitates the large scale screening of biochemical activity and protein interactions [79,80]. Recently, proteomic chips have been used to characterise the protein binding partners of calmodulin; a conserved membrane protein and secondary messenger for growth, differentiation and membrane trafficking. Zhu et al. [80] arrayed 5800 different proteins comprising ~93% of yeast open reading frames (ORF's) and found six known and 33 novel calmodulin binding partners.

Proteins have also been successfully characterised using a combination of chip-based analyte capture and MS [81]. This combination allows variations in proteins following different treatments to be quantitated and these proteins identified using surface enhanced laser desorption/ionisation time-of-flight mass spectrometry (SELDI-TOF-MS). SELDI retains proteins on a solid-phase binding surface as an array using affinity capture with direct detection of retained proteins by TOF-MS [82].

3.2.2. 2-DE–protein "macroarrays"

2-DE is a powerful technology for simultaneously resolving large numbers of proteins and has remained a key component in proteome analysis. The information obtainable from 2-DE is p*I* and mass, relative abundance, protein isoforms and modifications, subcellular localisation, rate of turnover, protein associations, and co-regulation of proteins. This information and the resolving power achieved using narrow range IPGs (1 pH unit/24 cm IPG) are being used to assemble a "three-dimensional view" of the proteome, where the third dimension relates to the subcellular location, temporal or spatial expression, or sequential (detergent) solubilisation of proteins [83]. The use of 2-DE for expression proteomics is particularly targeted at those proteins that meet the specifications of mass (<10 kDa, >100 kDa) and p*I* (>pH 3, <pH 9). Proteins that have a low solubility or high hydrophobicity, are of low abundance, and have a detection limiting number of dye-staining groups can fall beyond the scope of 2-DE. Some membrane proteins are difficult or unable to be analysed by 2-DE. Significant progress has been made in the separation of intra-membrane, membrane-associated and other hydrophobic proteins using alkaline (carbonate) incubation [84,85], or organic solvent (chloroform and methanol) extraction [86]. However, although these 2-DE methods have been successful, a one-dimensional SDS–PAGE approach maybe more suitable for some membrane proteins because of the harsher conditions employed to solubilise proteins [87]. Protein mixtures with large differences in copy number (7 to 8 orders of magnitude in human cells [88]) may also present a challenge for the 2-DE approach. Areas of concern are the IPG loading capacity, and staining that offers only 4–5 orders of magnitude for silver and fluorescent stain detection. The amount of sample required to detect low abundance proteins is displayed in Fig. 7. Some of the challenges faced by gel electrophoresis may be readily solved by the use of alternative technologies discussed above.

3.3. Protein and peptide quantitation

Quantitative and global protein analysis is an essential complement to the study of gene expression. Until recently, quantitative measurements of protein expression consisted of using either radioactive [16], or stable isotope [89] labelling followed by 2-D electrophoresis, high-performance liquid chromatography (HPLC) [90] or (f-moc) based amino

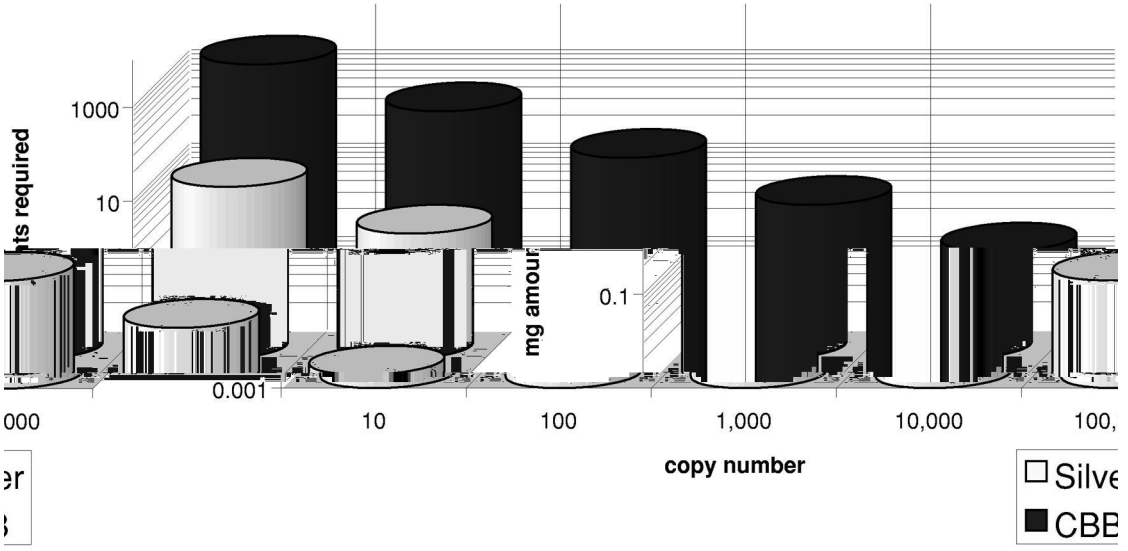

Fig. 7. Representation of total protein amounts required for individual protein visualisation by silver and Coomassie staining. These calculations are based on 1 mg of yeast protein being derived from harvesting $6{\cdot}10^7$ cells, protein mass of 50 kDa and 100% efficiency.

acid analysis [91] of proteins. 2-D gels have successfully been used for this approach in the past [92]. However, multiple 2-DE gels are required to track reproducible and statistically relevant changes and this is constrained by the speed of 2-DE separations and software analysis. There are several new quantitative technologies, which will be discussed below.

3.3.1. Isotopic labels

Proteolytic stable isotope labelling provides quantitative and concurrent comparisons between individual proteins from two entire proteome pools. This allows the relative differences of signal intensities of heavy and light paired peptides to be determined. Two ^{18}O atoms are incorporated into the carboxyl termini of all tryptic peptides during the cleavage of all proteins in the first pool, while proteins from the second pool contain two ^{16}O atoms. Both peptide mixtures are combined and the masses and isotope ratios of each peptide pair (differing by 4 Da) are measured by high-resolution MS. Relative signal intensities of paired peptides quantify the expression levels of proteins and sequence analysis enables protein identification. Some recent applications include: the comparison of virion proteins between two serotypes of adenovirus [93]; wild-type versus mutant protein expression in yeast cells [94]; and expression differences for some micro-organisms [95].

Recently, much attention has been given to the quantitative profiling of differentially induced proteins, tagged and affinity captured using ICATs and MS [96], see Fig. 8A. With this approach the expressed proteins (from a cell grown under two different physiological conditions) are labelled separately on the side chains of their reduced cysteine residues using one of two isotopically different but chemically identical sulfhydryl-reactive ICAT reagents (Fig. 8B). One of the ICAT reagents has hydrogen atoms on the carbon backbone (the d0 reagent), and the other is an isotopically heavy (d8) reagent, where the hydrogen atoms have been replaced with deuterium atoms. As the pair of peptides labelled with the d0 and d8 versions of the ICAT reagent are chemically identical, according to the stable isotope dilution theory [97], they serve as mutual internal standards for accurate protein quantification. The relative quantity of each protein present in the two biological samples is determined by measuring the relative signal intensities of the concurrently eluting isotopically labelled peptides using an initial mass spectral scan. The identification of the

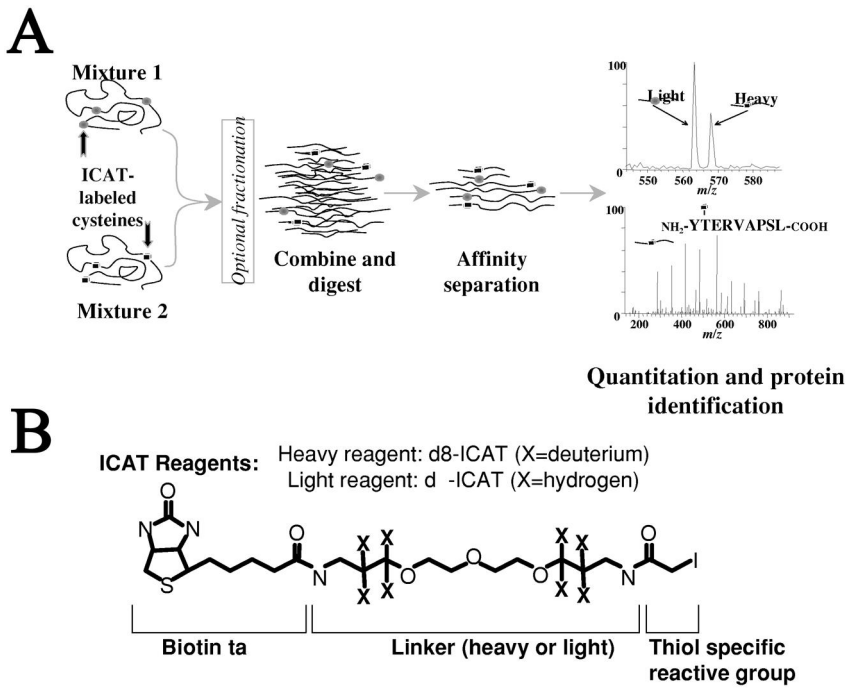

Fig. 8. The ICAT approach uses protein mixtures of two cell states that have been treated with the isotopically labelled light and heavy reagent. (A) The protein mixtures are combined and enzymatically digested. ICAT labelled peptides are isolated using the tag and separated by LC. Labelled peptides are chemically identical and elute almost simultaneously. They are easily recognised because they have a mass difference of 8 Da. The ratios of the two peptides can then be calculated and the protein identified. (B) This reagent is made up of an affinity tag (biotin), a linker to incorporate stable isotopes, and a thiol reactive group specific toward the thiol group of cysteines [96].

proteins is accomplished by switching the instrument to MS/MS mode in which it selects peptides for CID.

3.3.2. Two-dimensional difference gel electrophoresis (DIGE) gels

Experimental variability between 2-DE gels requires reference images to be built up of multiple (~6) gels before statistically significant comparisons can be made [98]. This requirement can be circumvented if comparisons between two states are achieved on a single gel. 2-D DIGE [99] achieves this by using different amine reactive fluorescent dye labels for two samples which are then combined and run on a single 2-D gel. The gel is then compared for variations in expression using different wavelengths for image analysis. Modified expression can be quantified because of differences in the ratio of dye label. This technology has been applied to study the differences in expression induced by benzoic acid in *Escherichia coli*. In *E. coli*, benzoate disrupts the transmembrane electrical potential and uncouples ATP synthesis. Lambert et al. were able to show an increase in benzoate induced expression of 33 proteins using the DIGE technology [100]. These proteins were subsequently identified using MALDI-TOF.

3.4. Protein identification and characterisation

The technologies that enable the identification of proteins and peptides target specific physico chemical properties. Some techniques are more amenable than others to the analysis of particular classes of proteins. Whilst amino acid composition analysis (AAC) [101–105] is now used less often because of accurate and sensitive MS based techniques, it has been used successfully for many years. AAC bases identifications on similarity of composition of amino acids and can provide quantitative as well as quali-

tative information. N-Terminal sequencing is also used successfully providing proteins are not blocked at the amino terminus and are in pmol quantities [106]. N-Terminal sequencing is extremely useful for establishing sequence information where no database correlation is possible.

3.4.1. MS based analysis

MS has become a powerful, rapid and sensitive tool for the analysis of proteins. All mass spectrometers are built up of three components: (i) an ionisation source such as MALDI or ESI; (ii) a mass analyser such as quadrupoles, ion trap, TOF tube; (iii) an ion detector such as electron multipliers, photomultipliers or conversion dynode [107,108]. The flexibility to combine ionisation sources with mass analysers has brought forward numerous types of MS instruments with varying mass accuracy, sensitivity and importantly, applications. Ions produced in the ion source are separated in the mass analyzer by their mass-to-charge (m/z) ratio. MS data are recorded as ''spectra'' which display ion intensity versus the m/z value. The two techniques that have become preferred methods for ionization of peptides and proteins are ESI and MALDI, due to their effective application on a wide range of proteins and peptides [109,110]. Generally, two types of MS data have been used for protein identification by correlation with sequence databases: (i) the accurate mass of peptides (within 5 ppm resolution) derived by specific emzymatic cleavage of the isolated protein; and (ii) CID spectra from individual peptides isolated after proteolysis of the target protein. The methods to generate these data have been described elsewhere [111]. An important caveat to these correlative analyses is that the protein sequence, or protein sequence translated in six frames from nucleotide sequence, is contained within the database being searched.

Besides MS hardware, development of MS software to assist in data acquisition directly relates to the success of MS based protein identification. Identification of peptides in complex solutions is limited by the quality and amount of data generated during an MS experiment. Factors that contribute to the quality and abundance of data are: dynamic exclusion; ''data dependent'' acquisition; and ''peak-parking''. Dynamic exclusion refers to the exclusion of previously fragmented ion masses for a selected time (e.g., 45 s) or the exclusion of a user defined ion list. This is done to eliminate the (redundant) fragmentation of peptides already fragmented during an experiment. Consequently a greater number of peptide ions are fragmented. In addition, one can program an instrument to only perform MS/MS in a data dependent manner; MS/MS information is obtained only for those peptides that yield ''quality'' data, that can be used post-acquisition for database interrogation. CID can be performed in order of relative abundance of ions, which in many cases rapidly elute from the HPLC or following SPE–CE (discussed above) and exceed the scan rate of the mass spectrometer. These tools therefore allow selective data acquisition and extended analysis time for peptides in complex solutions or of low abundance.

4. Bio databases and software for interrogation

The continuing exponential increase of sequence information and annotation of databases, as well as the large amount of data that is generated in single high-throughput experiments requires current databases to rely on sophisticated database management systems. A database is a store of information in any number of particular formats for which content can be probed via a user interface to make inferences about experimental data. The maintenance of a robust infrastructure of biological data is very important as the majority of identifications are via search engines probing protein and nucleotide databases. For example, the EST database [112,113] is the largest store of information and contains 9 184 900 entries (October 2001). This information can be manipulated in different ways for the identification of known proteins or peptides. Databases can be grouped into three broad categories defined by their information content and storage format. These categories are summarised with examples in Fig. 8, and consist of literature (citations and journals), factual (sequences and structure), and knowledge (motifs and biochemical pathways) databases.

Protein identification from MS experimental data requires computer algorithms that correlate MS and MS/MS spectra with database information [114–116]. These algorithms are specific for MS data interrogation applications and some of the web addresses of these applications have been included in Fig. 9. Protein identification requires either; accurate peptide mass (~5–25 ppm resolution) for MS, or MS/MS from peptides following enzymatic cleavage. The analysis of PMF results are not always straightforward, as experimental input masses do not always match theoretical database masses. These unexpected masses [117] can result in the incorrect identification, and may result from modification and/or non-specific cleavage or multiple proteins within the sample. There are however useful search engines that take known protein modifications into account [118]. In comparison, the information content of MS/MS spectra can provide rapid and unambiguous identification of known peptides from databases because more information is present in a CID spectrum of a peptide such as; mass of the intact peptide, sequence specific fragmentation of ions and internal fragment ions. The benefits of the MS/MS approach is that complex mixtures of proteins can be identified as well as some post-translational modifications [119].

5. Conclusion

Proteomic tools measure gene expression, protein activity and interactions of biological events at the protein level. Proteins are the major catalysts of biological function and contain several dimensions of information that collectively indicate the actual rather than the potential functional state as indicated by mRNA analysis. Measurements can be made in terms of protein quantity, location, and time-point. For the future we see a further integration of existing and new technologies for proteomics from a wide range of areas of biochemistry, chemistry, physics, computing science and molecular biology. This will further advance our knowledge of how biological systems are built up and what mechanisms control these systems. However, the potential of proteomics to comprehensively answer all biological questions is limited as only protein activity is measured. A unification of genomics, proteomics, and other technologies is needed if we are to start to understand the

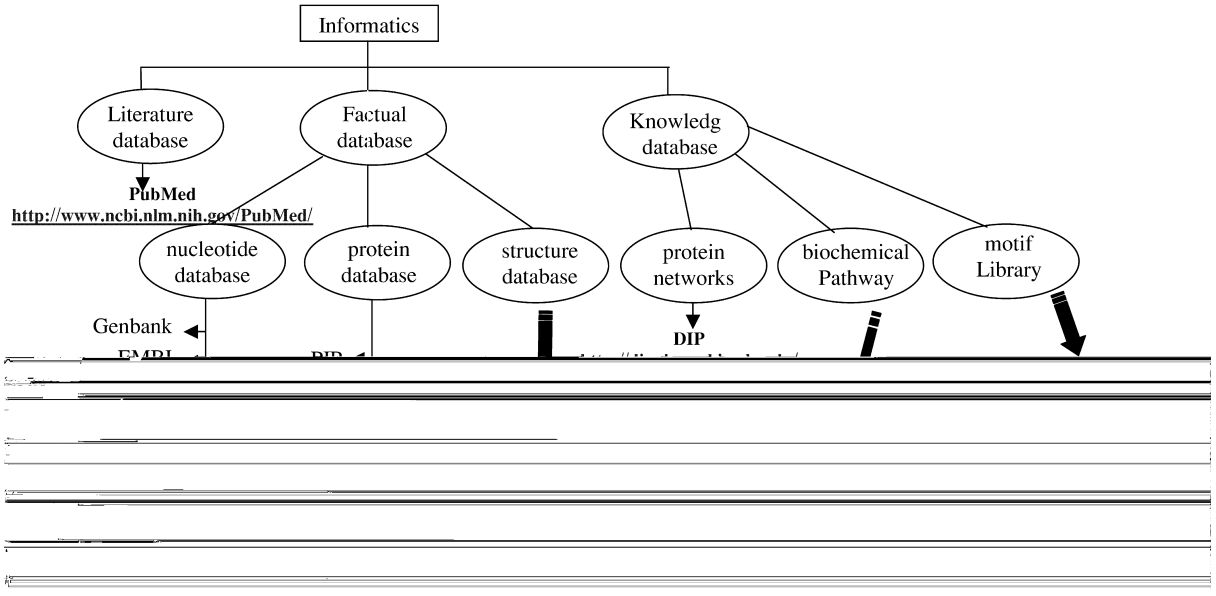

Fig. 9. Representation of information content of databases.

complexity of biological function in the context of disease and health.

References

[1] V.C. Wasinger, S.J. Cordwell, A. Cerpa-Poljak, J.X. Yan, A.A. Gooley, M.R. Wilkins, M.W. Duncan, R. Harris, K.L. Williams, I. Humphery-Smith, Electrophoresis 16 (1995) 1090.

[2] M.R. Wilkins, J.C. Sanchez, A.A. Gooley, R.D. Appel, I. Humphery-Smith, D.F. Hochstrasser, K.L. Williams, Biotechnol. Genet. Eng. Rev. 13 (1996) 19.

[3] G.L. Corthals, V.C. Wasinger, D.F. Hochstrasser, J.C. Sanchez, Electrophoresis 21 (2000) 1104.

[4] M. Fountoulakis, R. Hardmaier, E. Schuller, G. Lubec, Electrophoresis 3 (2000) 673.

[5] P.A. Haynes, S.P. Gygi, D. Figeys, R. Aebersold, Electrophoresis 19 (1998) 1862.

[6] G.M. Rubin, M.D. Yandell, J.R. Wortman, G.L. Gabor Miklos, C.R. Nelson, I.K. Hariharan, M.E. Fortini, P.W. Li, R. Apweiler, W. Fleischmann, J.M. Cherry, S. Henikoff, M.P. Skupski, S. Misra, M. Ashburner, E. Birney, M.S. Boguski, T. Brody, P. Brokstein, S.E. Celniker, S.A. Chervitz, D. Coates, A. Cravchik, A. Gabrielian, R.F. Galle, W.M. Gelbart, R.A. George, L.S. Goldstein, F. Gong, P. Guan, N.L. Harris, B.A. Hay, R.A. Hoskins, J. Li, Z. Li, R.O. Hynes, S.J. Jones, P.M. Kuehl, B. Lemaitre, J.T. Littleton, D.K. Morrison, C. Mungall, P.H. O'Farrell, O.K. Pickeral, C. Shue, L.B. Vosshall, J. Zhang, Q. Zhao, X.H. Zheng, S. Lewis, Science 287 (2000) 2204.

[7] H. Hernández, C.V. Robinson, J. Biol. Chem. 276 (2001) 46685.

[8] B.H. ter Kuile, H.V. Westerhoff, FEBS Lett. 500 (2001) 169.

[9] K. Arrell, I. Neverova, J.E. Van Eyk, Circ. Res. 88 (2001) 763.

[10] V.M. Hoang, R. Foulk, K. Clauser, A. Burlingame, B.W. Gibson, S.J. Fisher, Biochemistry 40 (2001) 4077.

[11] J. Vohradsky, J. Biol. Chem. 276 (2001) 36168.

[12] M.T. Laub, H.H. McAdams, T. Feldblyum, C.M. Fraser, L. Shapiro, Science 290 (2000) 2144.

[13] J.M. Etcher, S.P. Nair, J.M. Ward, B. Henderson, M. Wilson, Microb. Pathog. 30 (2001) 359.

[14] J.W. Thatcher, J.M. Shaw, W.J. Dickinson, Proc. Natl. Acad. Sci. USA 95 (1998) 253.

[15] T. Ideker, V. Thorsson, J.A. Ranish, R. Christmas, J. Buhler, J.K. Eng, R. Bumgarner, D.R. Goodlett, R. Aebersold, L. Hood, Science 292 (2001) 929.

[16] S.P. Gygi, Y. Rochon, B.R. Franza, R. Aebersold, Mol. Cell Biol. 19 (1999) 1720.

[17] L.D. Greller, F.L. Tobin, Genome Res. 9 (1999) 282.

[18] N.L. Anderson, N.G. Anderson, Electrophoresis 19 (1998) 1853.

[19] W.P. Blackstock, M.P. Weir, Trends Biotechnol. 17 (1999) 121.

[20] G.L. Corthals, P.S. Nelson, Pharmacogenomics J. 1 (2001) 15.

[21] R.C. Strohman, Biotechnology 12 (1994) 156.

[22] R.C. Strohman, Integr. Physiol. Behav. Sci. 30 (1995) 273.

[23] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.

[24] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.

[25] S.L. Parker, T. Tong, S. Bolden, P.A. Wingo, CA Cancer J. Clin. 47 (1997) 5.

[26] J.P. Charrier, C. Tournel, S. Michel, S. Comby, C. Jolivet-Reynaud, J. Passagot, P. Dalbon, D. Chautard, M. Jolivet, Electrophoresis 22 (2001) 1861.

[27] V.E. Velculescu, L. Zhang, W. Zhou, J. Vogelstein, M.A. Basrai, D.E. Bassett, P. Hieter, B. Vogelstein, K.W. Kinzler, Cell 88 (1997) 243.

[28] P.S. Nelson, D. Han, Y. Rochon, G.L. Corthals, B. Lin, A. Monson, V. Nguyen, B.R. Franza, S.R. Plymate, R. Aebersold, L. Hood, Electrophoresis 21 (2000) 1823.

[29] A. Waghray, F. Feroze, M.S. Schober, F. Yao, C. Wood, E. Puravs, M. Krause, S. Hanash, Y. Q Chen, Proteomics 10 (2001) 1327.

[30] T. Rabilloud, S. Kieffer, V. Procaccio, M. Louwagie, P.L. Courchesne, S.D. Patterson, P. Martinez, J. Garin, J. Lunardi, Electrophoresis 19 (1998) 1006.

[31] T.K. Chataway, A.M. Whittle, M.D. Lewis, C.A. Bindloss, R.C. Davey, R.L. Moritz, R.J. Simpson, J.J. Hopwood, P.J. Meikle, Placenta 19 (1998) 643.

[32] T. Suzuki, M. Terasaki, C. Takemoto-Hori, T. Hanada, T. Ueda, A. Wada, K. Watanabe, J. Biol. Chem. 276 (2001) 33181.

[33] C.C. Wu, J.R. Yates III, M.C. Neville, K.E. Howell, Traffic 1 (2000) 769.

[34] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Nat. Biotechnol. 19 (2001) 946.

[35] E. Liu, Oncogene 7 (1992) 1027.

[36] M. Okamoto, Y. Nakajima, T. Matsuyama, M. Sugita, Neuroscience 104 (2001) 653.

[37] E.H. Koo, P.T. Lansbury Jr., J.W. Kelly, Proc. Natl. Acad. Sci. USA 96 (1999) 9989.

[38] A. Šali, Nature 402 (1999) 23.

[39] B. Schwikowski, P. Uetz, S. Fields, Nat. Biotechnol. 18 (2000) 1257.

[40] E.M. Marcotte, Nat. Biotechnol. 19 (2001) 626.

[41] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, Nat. Biotechnol. 17 (1999) 1030.

[42] D. Figeys, L.D. McBroom, M.F. Moran, Methods 24 (2001) 230.

[43] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates III, Nat. Biotechnol. 17 (1999) 676.

[44] J.R. Yates III, A.L. McCormack, D. Schieltz, E. Carmack, A. Link, J. Protein Chem. 16 (1997) 495.

[45] A.H. Brockman, R. Orlando, R.L. Tarleton, Rapid Commun. Mass Spectrom. 13 (1999) 1024.

[46] W. Tong, A. Link, J.K. Eng, J.R. Yates III, Anal. Chem. 71 (1999) 2270.

[47] M.P. Routa, J.D. Aitchison, A. Supraptoa, K. Hjertaasb, Y. Zhaoa, B.T. Chait, J. Cell Biol. 148 (2000) 635.
[48] A.-C. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.-M. Michon, C.-M. Cruciat, M. Remor, C. Höfert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.-A. Heurtier, R.R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, G. Superti-Furga, Nature 414 (2001) 141.
[49] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.-L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier, L. Yang, C. Wolting, I. Donaldson, S. Schandorff, J. Shewnarane, M. Vo, J. Taggart, M. Goudreault, B. Muskat, C. Alfarano, D. Dewar, Z. Lin, K. Michalickova, A.R. Willems, H. Sassi, P.A. Nielsen, K.J. Rasmussen, J.R. Andersen, L.E. Johansen, L.H. Hansen, H. Jespersen, A. Podtelejnikov, E. Nielsen, J. Crawford, V. Poulsen, B.D. Sørensen, J. Matthiesen, R.C. Hendrickson, F. Gleeson, T. Pawson, M.F. Moran, D. Durocher, M. Mann, C.W.V. Hogue, D. Figeys, M. Tyers, Nature 414 (2002) 180.
[50] E.M. Marcotte, M. Pellegrini, H.-L. Ng, D.W. Rice, T.O. Yeates, D. Eisenberg, Science 285 (1999) 751.
[51] A.J. Enright, I. Illiopoulos, N.C. Kyrpides, C.A. Ouzounis, Nature 402 (1999) 86.
[52] M. Pellegrini, E.M. Marcotte, M.J. Thompson, D. Eisenberg, T.O. Yeates, Proc. Natl. Acad. Sci. USA 96 (1999) 4285.
[53] T. Dandekar, B. Snel, M. Huynen, P. Bork, Trends Biochem. Sci. 23 (1998) 324.
[54] R. Overbeek, M. Fonstein, M. D'Souza, G.D. Pusch, N. Maltsev, Proc. Nat. Acad. Sci. USA 96 (1999) 2896.
[55] A.L. McCormack, D.M. Schieltz, B. Goode, S. Yang, G. Barnes, D. Drubin, J.R. Yates III, Anal. Chem. 69 (1997) 767.
[56] A.J. Link, J.K. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates III, Nat. Biotechnol. 17 (1999) 676.
[57] M.P. Washburn, D. Wolters, J.R. Yates III, Nat. Biotechnol. 19 (2001) 242.
[58] G.J. Opiteck, S.M. Ramirez, J.W. Jorgenson, M.A. Moseley 3rd, Anal. Biochem. 258 (1998) 349.
[59] G.J. Opiteck, J.W. Jorgenson, Anal. Chem. 69 (1997) 2283.
[60] M.T. Davis, J. Beierle, E.T. Bures, M.D. McGinley, J. Mort, J.H. Robinson, C.S. Spahr, W. Yu, R. Luethy, S.D. Patterson, J. Chromatogr. B 752 (2001) 281.
[61] P. Cao, J.T. Stults, Rapid Commun. Mass Spectrom. 14 (2000) 1600.
[62] W. Tong, A. Link, J.K. Eng, J.R. Yates III, Anal. Chem. 71 (1999) 2270.
[63] D. Figeys, S.P. Gygi, Y. Zhang, J. Watts, M. Gu, R. Aebersold, Electrophoresis 19 (1998) 1811.
[64] B. Gallis, G.L. Corthals, D.R. Goodlett, H. Ueba, F. Kim, S.R. Presnell, D. Figeys, D.G. Harrison, B. C Berk, R. Aebersold, M. Corson, J. Biol. Chem. 274 (1999) 30101.
[65] D. Figeys, G.L. Corthals, B. Gallis, D.R. Goodlett, A. Ducret, M.A. Corson, R. Aebersold, Anal. Chem. 71 (1999) 2279.
[66] D.R. Goodlett, J.H. Wahl, H.R. Udseth, R.D. Smith, J. Microcol. Sep. 5 (1993) 57.
[67] G.L. Corthals, D.R. Goodlett, V.C. Wasinger, R. Aebersold, Methods Enzymol., in press.
[68] R.G. Krishna, F. Wold, Adv. Enzymol. Relat. Areas Mol. Biol. 67 (1993) 265.
[69] H. Charbonneau, N.K. Tonks, Annu. Rev. Cell. Biol. 8 (1992) 463.
[70] E.H. Fischer, E.G. Krebs, Biochim. Biophys. Acta 1000 (1989) 297.
[71] T. Hunter, Cell 50 (1987) 823.
[72] T. Liaw, V.C. Wasinger, G.L. Corthals, manuscript in preparation.
[73] J. Goverman, Methods Mol. Biol. 112 (1999) 265.
[74] R.E. Jenkins, S.R. Pennington, Proteomics 1 (2001) 13.
[75] S.P. Fodor, J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, D. Solas, Science 251 (1991) 767.
[76] H.M. Geysen, R.H. Meloen, S.J. Barteling, Proc. Natl. Acad. Sci. USA 81 (1984) 3998.
[77] A.Q. Emili, G. Cagney, Nat. Biotechnol. 18 (2000) 393.
[78] G. MacBeath, Nat. Biotechnol. 19 (2001) 828.
[79] H. Zhu, J.F. Klemic, S. Chang, P. Bertone, A. Casamayor, K.G. Klemic, D. Smith, M. Gerstein, M.A. Reed, M. Snyder, Nat. Genet. 26 (2000) 283.
[80] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R.A. Dean, M. Gerstein, M. Snyder, Science 293 (2001) 2101.
[81] F. von Eggeling, K. Junker, W. Fiedler, V. Wollscheid, M. Dürst, U. Claussen, G. Ernst, Electrophoresis 14 (2001) 2898.
[82] http://www.ciphergen.com
[83] S.J. Cordwell, A.S. Nouwens, N.M. Verrills, D.J. Basseal, B.J. Walsh, Electrophoresis 21 (2000) 1094.
[84] M.P. Molloy, N.D. Phadke, J.R. Maddock, P.C. Andrews, Electrophoresis 22 (2001) 1686.
[85] M.P. Molloy, B.R. Herbert, M.B. Slade, T. Rabilloud, A.S. Nouwens, K.L. Williams, A.A. Gooley, Eur. J. Biochem. 267 (2000) 2871.
[86] M.P. Molloy, B.R. Herbert, K.L. Williams, A.A. Gooley, Electrophoresis 20 (1999) 701.
[87] R.J. Simpson, L.M. Connolly, J.S. Eddes, J.J. Pereira, R.L. Moritz, G.E. Reid, Electrophoresis 21 (2000) 1707.
[88] N.L. Anderson, N.G. Anderson, Electrophoresis 19 (1998) 1853.
[89] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. USA 96 (1999) 6591.
[90] D.R. Goodlett, A. Keller, J.D. Watts, R. Newitt, E.C. Yi, S. Purvine, J.K. Eng, P. von Haller, R. Aebersold, E. Kolker, Rapid Commun. Mass Spectrom. 15 (2001) 1214.
[91] J.X. Yan, W.C. Kett, B.R. Herbert, A.A. Gooley, N.H. Packer, K.L. Williams, J. Chromatogr. A 1 (1998) 187.
[92] M. Goldfarb, Electrophoresis 20 (1999) 870.
[93] X. Yao, A. Freas, J. Ramirez, P.A. Demirev, C. Fenselau, Anal. Chem. 73 (2001) 2836.
[94] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. USA 96 (1999) 6591.

[95] R.D. Smith, L. Pasa-Tolic, M.S. Lipton, P.K. Jensen, G.A. Anderson, Y. Shen, T.P. Conrads, H.R. Udseth, R. Harkewicz, M.E. Belov, C. Masselon, T.D. Veenstra, Electrophoresis 22 (2001) 1652.
[96] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[97] A.P. De Leenheer, L.M. Thienpont, Mass Spectrom. Rev. 11 (1992) 249.
[98] V.C. Wasinger, J.D. Pollack, I. Humphery-Smith, Eur. J. Biochem. 267 (2000) 1571.
[99] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[100] L.A. Lambert, K. Abshire, D. Blankenhorn, J.L. Slonczewski, J. Bacteriol. 179 (1997) 7595.
[101] C. Eckerskorn, P. Jungblut, W. Mewes, J. Klose, F. Lottspeich, Electrophoresis 9 (1988) 830.
[102] G. Shaw, Proc. Natl. Acad. Sci. USA 90 (1993) 5138.
[103] U. Hobohm, T. Houthaeve, C. Sander, Anal. Biochem. 222 (1994) 202.
[104] A. Galat, F. Bouet, S. Riviere, Electrophoresis 16 (1995) 1095.
[105] O. Golaz, M.R. Wilkins, J.C. Sanchez, R.D. Appel, D.F. Hochstrasser, K.L. Williams, Electrophoresis 17 (1996) 573.
[106] H. Chen, J. Higgins, E. Kondorosi, A. Kondorosi, M.A. Djordjevic, J.J. Weinman, B.G. Rolfe, Electrophoresis 21 (2000) 3823.
[107] G. Suizdak, in: Mass Spectrometry For Biotechnology, Academic Press, London, 1996, p. 6;
G. Suizdak, in: Mass Spectrometry For Biotechnology, Academic Press, London, 1996, p. 35;
G. Suizdak, in: Mass Spectrometry For Biotechnology, Academic Press, London, 1996, p. 48.
[108] D. Figeys, G.L. Corthals, B. Gallis, D.R. Goodlett, A. Ducret, M.A. Corson, R. Aebersold, Anal. Chem. 71 (1999) 2279.
[109] M. Karas, F. Hillenkamp, Anal. Chem. 60 (1988) 2299.
[110] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Science 246 (1989) 64.
[111] G.L. Corthals, S. Gygi, R. Aebersold, S.P. Patterson, in: T. Rabilloud (Ed.), Proteome Research: 2D Gel Electrophoresis and Detection Methods, Springer-Verlag, Heidelberg, 1999, p. 197.
[112] M.S. Boguski, T.M.J. Lowe, C.M. Tolstoshev, Nat. Genet. 4 (1993) 332.
[113] M.S. Boguski, C.M. Tolstoshev, D.E. Bassett, Science 265 (1994) 1993.
[114] W.J. Henzel, T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, C. Watanabe, Proc. Natl. Acad. Sci. USA 90 (1993) 5011.
[115] M. Mann, M. Wilm, Anal. Chem. 66 (1994) 4390.
[116] K.R. Clauser, S.C. Hall, D.M. Smith, J.W. Webb, L.E. Andrews, H.M. Tran, L.B. Epstein, A.L. Burlingame, Proc. Natl. Acad. Sci. 92 (1995) 5072.
[117] S.D. Patterson, R. Aebersold, Electrophoresis 16 (1995) 1791.
[118] http://ch.expasy.org/tools/
[119] J.R. Yates III, J.K. Eng, A.L. McCormack, Anal. Chem. 67 (1995) 1426.