ELSEVIER

Journal of Chromatography B, 782 (2002) 165–181

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Review

The functional proteomics toolbox: methods and applications

Thomas C. Hunter, Nancy L. Andon, Antonius Koller, John R. Yates III, Paul A. Haynes*

Protein and Metabolite Dynamics Department, Torrey Mesa Research Institute, Syngenta Genomics Research and Technology, 3115 Merryfield Row, San Diego, CA 92121, USA

Keywords: Reviews; Proteomics

Contents

1. Introduction

1.1. The current state of proteomics

During the 20th century advances in biomedical research have led to remarkable improvements in the prevention, diagnosis and treatment of human diseases. However, current epidemics of human diseases such as AIDS, tuberculosis and malaria, and our inability to devise effective treatments for chronic diseases such as Alzheimer's disease or multiple sclerosis, serve to highlight that our understanding of disease biology is still incomplete [1]. These diseases remain major killers, largely because researchers have previously not been able to investigate the complex relationship between proteins and genes on a genomic scale. The recent publication of

*Corresponding author. Fax: +1-858-812-1349.
E-mail address: paul.haynes@syngenta.com (P.A. Haynes).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00570-6

the human genome sequence [2], and the completion of the genomic sequences for many infectious organisms, has opened up new opportunities for more detailed and complete understanding of human disease that will inevitably lead to new strategies for treatment and prevention. Proteomics is an emerging field that is poised to have a significant impact on the future of research into human diseases.

Proteomics is the global analysis of complex protein mixtures for the purpose of qualitative, quantitative and functional analysis of all the proteins present in a sample. The role of a protein is reflected by interaction with other proteins. Therefore, the identification of a protein in the context of its cellular environment is necessary to understand function and regulation. Furthermore, alternative splicing of transcripts, and post-translational modification (phosphorylation, glycosylation, etc) of a protein can lead to multiple forms. Thus, the estimated 35 000–80 000 genes predicted in the human genome could easily produce several hundred thousand or more different proteins [3,4]. Proteomic studies are imperative since neither genomic or transcriptomic data is able to provide a complete picture of the organism. Systematic analysis of gene function is preferable at the protein level rather than at the genetic level, since it is proteins that perform most of the reactions necessary for the cell. Functional proteomics, the global characterization of functional features of proteins, is necessary to better understand these events which constitute the metabolic and structural signals that control growth, development, replication and stress response of cells.

In this review, we will describe the current methodologies available for the global analysis of protein expression, focusing on four main areas: the role of mass spectrometry in proteomics; complex mixture characterization; quantitative proteomics; and protein fractionation.

1.2. The role of mass spectrometry in proteomics

Since the late 1980s advances in several areas of mass spectrometer design have led to the current generation of high-performance machines. In the field of proteomics, three key areas of technology have all undergone rapid advances simultaneously: improvements in mass spectrometry design, including the recent advent of commercial hybrid instruments such as the matrix-assisted laser desorption ionization MALDI–TOF–TOF and the electrospray- and MALDI–Q–TOF (reviewed in Ref. [5]); tremendous increases in available computing power, with a concomitant decrease in per unit cost; and the ever-increasing availability of genomic sequence data for a variety of organisms. These developments have fueled the current rapid growth in the field of proteomic analysis, and have served to enhance the utility and applicability of mass spectrometric techniques in the analysis of biological molecules.

One of the most significant developments in the last decade for proteomics has been the completion of many genome sequencing projects. Complete genome sequences have already been reported for a number of organisms, among them *Haemophilus influenzae* [6], *Saccharomyces cerevisiae* [7], *Escherichia coli* [8], *Caenorhabditis elegans* [9], *Drosophila melanogaster* [10], *Homo sapiens* [2] and *Oryza sativa* (rice) [11]. A complete list of sequenced genomes and current sequencing projects can be found at the Institute for Genomic Research (TIGR) website (www.tigr.org).

Information contained within these very large databases would be of little use without the means to determine the function of the uncharacterized open reading frames. Previous methods for identifying the gene from which a protein originated lacked the sensitivity, reliability and speed to look at samples on a genomic scale. However, it was realized in the early 1980s that tandem MS methods could be used to sequence oligopeptides from complex mixtures by collision-activated dissociation on a triple quadrupole mass spectrometer at picomole levels [12]. The approach involves enzymatic and/or chemical digestion of the protein to the resulting peptides. The peptides are then fractionated by high-performance liquid chromatography and are then analyzed directly by secondary-ion–collision-activated dissociation mass spectrometry (tandem MS–MS) on a multi-analyzer instrument [12,13]. The data generated by these experiments contain highly specific information such as sequence specific fragmentation patterns as well as peptide mass information. The most notable computer algorithm for analysis of tandem mass spectra is SEQUEST [14,15]. The SEQUEST program uses information on the digestion of a protein

with a site-specific protease, the molecular masses of the resulting peptides, and the sequence specific fragmentation spectra to identify the protein from translated protein sequences, raw nucleotide sequences, or expressed sequence tags (ESTs) [16–18]. The ability to search nucleotide databases is an advantage when analyzing data obtained from organisms whose genomes are not yet completely sequenced, but a large amount of expressed gene sequence data is available.

2. Techniques for complex mixture characterization

No single protocol is adequate to address the varied issues for analysis in complex protein mixture characterization, which stem from the diverse nature of the cellular environment. For example, a particular set of proteins not amenable to separation by one technique may be amenable to characterization by an alternate method. Therefore, each protocol for characterization must be customized to meet the needs of the specific biological question to be addressed. For this reason, techniques for analysis of complex protein mixtures will often be novel, evolving and difficult to generalize. Many of the initial qualitative and quantitative analytical techniques for mass spectrometric based proteomics have evolved from classical biochemical techniques and have come back into vogue as they have found application in proteomics. The following sections will discuss the use of two-dimensional gel electrophoresis and chromatographic methodologies as strategies for separating complex mixtures.

2.1. The 2D electrophoresis approach

The classical technique that is most widely used in global proteome analysis is two-dimensional electrophoresis (IEF–SDS–PAGE) (2DE) [19]. In 2DE, proteins are first separated by isoelectric focusing and then further resolved by SDS–PAGE in the second, perpendicular, dimension. Separated proteins can then be visualized by numerous staining methods, or by autoradiography, to produce a two-dimensional image array that can contain thousands of proteins [20,21]. The identification of individual proteins from polyacrylamide gels, with one or two dimensions, has traditionally been carried out using comigration with known proteins [22], immunoblotting, N-terminal sequencing [23,24] or internal peptide sequencing [25,26]. In recent years there has been a fundamental shift in the ways such experiments are performed, principally due to the rapid growth of large-scale genomic databases. The current widely used method relies on excising spots from gels, proteolytically digesting the spots, and then extracting the peptides produced. The final stage involves analyzing these peptides by mass spectrometry (MS) or tandem mass spectrometry (MS–MS) and then correlating the mass spectral data derived from the peptides with information contained in databases of protein sequence, genomic sequence or expressed sequence tags (ESTs) [14,15,27,28].

The disadvantages of 2D electrophoresis are that it is extremely time-consuming, has a limited dynamic range, does not work well for hydrophobic proteins, and is essentially non-quantitative. Large format gels typically require at least 24 h to complete, and for reasons of practicality are usually completed over the course of several days. Staining of individual 2DE spots can be measured and compared using scanning densitometry, but there are so many caveats attached to the data that the results are of questionable value unless the differences are quite large. Many staining techniques, such as silver staining, suffer from a limited dynamic range, so that the intensity of less abundant spots is not linearly correlated to that of more abundant spots [29]. Moreover, some types of proteins, especially those that are post-translationally modified, can give quantitatively and qualitatively different staining in comparison to similar amounts of other proteins. However, protein staining sensitivity is constantly being improved, enabling researchers to better visualize the proteome of their system. Comparative studies between colloidal Coomassie blue, Daiichi silver, Sypro orange, Sypro red, Sypro ruby and Sypro tangerine concluded that the newest generation of fluorescent protein stains, compared with traditional staining methods, are more compatible with MALDI and LC–MS methods, and have a greater dynamic range [30–33]. Hydrophobic proteins, especially those of high molecular mass, are especially problematic in 2D gels because the presence of SDS is incompatible with successful first

dimension IEF. Thus, most IEF sample buffers solubilize a wide range of proteins by including high concentrations of chaotropic salts, such as urea, and lower levels of mild detergents such as CHAPS. It should be noted, however, that significant progress in overcoming this particular limitation has been made in recent years, including the development of new detergents with greater solubilizing power [34,35], and the selective application of organic solvents to aid in solubilizing hydrophobic proteins [36].

Several studies have shown that the majority of proteins identified in 2DE are the more abundant and the more long-lived proteins in the cell. In a study of more than 150 proteins identified in 2DE of yeast cells, for example, no proteins were identified with a codon bias value of less than 0.1, an arbitrarily defined cutoff indicating low abundance [37]. In contrast, calculated values indicate that over half of the 6000 genes in yeast [7] have a codon bias index of less than 0.1 and thus are unlikely to be seen in 2DE without prior enrichment. Several techniques have been proposed as generic sample pretreatment strategies to increase the total number of spots that can be visualized in 2DE. These include sequential extraction of a sample with buffers of increasing solubilizing power, which generates fractions on the basis of hydrophobicity [38], and using very narrow range pH gradients for the first dimension IEF, which expands the resolution in a given range [39,40].

Despite these disadvantages, 2DE remains the method of choice for displaying proteins as the front end of a proteomics project for two principal reasons: firstly, because it can be used to visualize a very large number of proteins simultaneously; and secondly, because it can be used in a differential display format. The ability to study complex biological systems in their entirety rather than as a multitude of individual components makes it far easier to discover the many complex relationships between proteins in functioning cells. This type of experiment, where the aim is to catalog as many of the expressed proteins as possible and build up a database of expressed proteins, is often referred to as a ''proteome project''.

There are now too many proteome projects under way to list them all. Relatively large scale proteome characterization projects which have been reported in recent years include those of microbial organisms such as *Saccharomyces cerevisiae* [41], *Escherichia coli* [42], *Haemophilus influenzae* [43], *Mycobacterium tuberculosis* [44], *Ochrobactrum anthropi* [45], *Salmonella enterica* [46], *Spiroplasma melliferum* [47], *Synechocystis* spp. [48], *Dictyostelium discoideum* [49] and *Rhizobium leguminosarum* [50], and tissues including human liver [51], human plasma [51], human fibroblasts [52], human keratinocytes [52], human bladder squamous cell carcinomas [52], mouse kidney [52], rat serum [53–55], and the roots of *Medicago truncatula* [56].

Following the announcement of the completion of the human genome sequence [2], considerable attention has been focused on efforts to ''map the human proteome'' [57]. This is a potentially endless undertaking, given the inherent changeability of the proteome of even one specimen of such a multicellular organism, and the added difficulty of defining what constitutes a ''normal human''. A major issue with establishing any proteome characterization project is defining the proteome in question. A single genome can give rise to an essentially infinite number of qualitatively and quantitatively different proteomes, depending on such variables as the stage of the cell cycle, growth and nutrient conditions, temperature and stress response, pathological conditions, and strain differences, to name but a few. Another way of expressing the same problem is that genomes are essentially static, while proteomes are by their very nature dynamic, and therefore a 2DE based proteome project can only represent a snapshot rather than the whole constantly moving picture.

Nonetheless, efforts are moving ahead in this field, with a public consortium recently meeting to initiate the planning of a human proteome project [58], and a privately funded effort well on the way to completion of an initial draft of a study involving a very large number of discrete tissues, known as the human proteome index [59]. Also, an ambitious attempt to undertake a complete human proteome project has been announced by the same research group responsible for one of the major successful efforts in the human genome project [57]. It remains to be seen whether this effort will be quite so successful.

2.2. Chromatographic separation methods

To purify and identify individual peptides, chromatographic strategies have been devised to separate peptides by physiochemical properties such as size, charge, and hydrophobicity. The same purification techniques are available for peptide purification as are used for protein purification, and include size-exclusion chromatography, ion-exchange chromatography, and reversed-phase HPLC, which is the technique most commonly used for peptide purification in proteomics. In reversed-phase HPLC the peptides are mainly retained due to hydrophobic interactions with the stationary silica phase that is chemically bonded with an alkylsilyl compound. Polar mobile phases, such as water mixed with methanol or acetonitrile, are used to elute the bound peptides. Peptides are eluted in order of decreasing polarity (increasing hydrophobicity). The most popular reversed-phase packing is C_{18} in which octadecasilyl groups are bonded to the silica surface [60,61].

Reversed phase chromatography can be used as the sole separation procedure for moderately complex peptide mixtures prior to tandem mass spectrometric analysis, but it is generally considered to have insufficient resolution for the analysis of more complex mixtures. While the mass spectrometer can perform mass measurements on several co-eluting peptides, if many peptides co-elute then the MS instrument can not fragment them all before they finish eluting from the column, and therefore valuable information is irretrievably lost. One interesting approach that has been used to circumvent this problem is the iterative analysis of aliquots of the same sample, but using very narrow mass ranges for the initial peptide mass measurements of an MS–MS scanning regime [62]. This technique, therefore, uses the mass selection of peptides in the mass spectrometer as an extra ''dimension of separation'' in the analysis of a complex mixture. This approach is necessarily limited in application to those cases where large amounts of sample are available for analysis. It has been used, for example, in LC–MS–MS analysis of the proteome of normal human urine, in a study in which 124 different gene products were identified [63].

One new methodology that represents a significant step forward in proteome analysis is the use of multidimensional liquid chromatography coupled to tandem mass spectrometry (LC–LC–MS–MS). The LC–LC–MS–MS method as recently reported for use in the analysis of complex mixtures of peptides [64], is now commonly known by the acronym MudPIT, for multi-dimensional protein identification technique. Multi-dimensional chromatography of proteins has been reported previously, involving the off-line coupling of reversed-phase columns with cation-exchange columns [65], or the coupling of size exclusion columns with reversed-phase columns and online detection [66]. The true power of this method for proteomics applications was demonstrated, however, when the technique was employed with a mixed bed microcapillary column containing strong cation-exchange (SCX) and reversed-phase (RP) resins [64]. This chromatographic technique contains several steps, as outlined below.

Firstly, a reduced and denatured protein mixture is proteolytically digested to produce peptide fragments. The mixture of peptides is loaded onto a microcapillary column containing SCX resin upstream of RP resin, eluting directly into a tandem mass spectrometer. A discrete fraction of the absorbed peptides are displaced from the SCX column onto the RP column using a step gradient of salt, causing the peptides to be retained on the RP column while contaminating salts and buffers are washed through. Peptides are subsequently eluted from the RP column using a solvent gradient, and analyzed by MS–MS. This process is applied in an iterative manner, typically involving 10–20 steps using increasing salt concentration to displace additional fractions from the SCX column, and the MS–MS data from all of the fractions are analyzed by database searching [14,15]. The combined data output gives a comprehensive picture of the protein components present in the initial sample.

There are several advantages of the MudPIT technique: it avoids the need for time-consuming 2DE; and can be run in a fully automated manner. The use of two dimensions for chromatographic separation also greatly increases the number of peptides that can be identified from very complex mixtures. For example, analysis of a total yeast cell

lysate identified 1484 unique proteins in a single series of MudPIT experiment [67], which is far more than would be expected from a conventional LC–MS–MS experiment. In addition, the method has a very wide dynamic range, and none of the protein solubility problems associated with 2DE since the proteins are all proteolytically digested at the same time. This difference is graphically demonstrated by the example of the analysis of the yeast ribosomal 80s complex, which was found to contain 64 proteins by analysis of 56 discrete spots visible in a 2DE experiment, while an additional 11 proteins were identified by analyzing the same sample using MudPIT [64].

The main disadvantages of this approach are concerned with post-experimental data processing. The extremely large volume of data collected in a MudPIT experiment consisting of 10–20 cycles of reversed-phase chromatography presents a significant problem in terms of both the time required to collate and assemble the data into a useable format and the computing power needed to complete database searching. This problem may be alleviated over time, as computing resources continue to steadily increase in performance and become more affordable. One other disadvantage of this approach is that it is generally limited to use with organisms that have complete genome sequence data available for searching. Mass spectrometric instrumentation and de novo sequencing algorithms will surely improve, however, making de novo sequencing on such a large scale a more practical proposition. Also, at some point in the future complete genomic sequence data will be available for all the major research organisms and therefore this will no longer represent a problem. There is also considerable scope for improvement in this technique, as it could, for example, be combined with a specific peptide or protein enrichment strategy, or it could be coupled with some of the strategies used for improving success in MS–MS based de novo sequencing experiments, such as employing proteolytic digestion in ^{18}O enriched water to provide an isotopic end label [68].

MudPIT represents a technique that is best suited to rapidly building a proteomic database rather than being applied in a differential display proteomic assay, and it is a viable alternative to 2DE for the analysis of certain complex mixtures. This approach is clearly going to become increasingly attractive as a means of extracting as much information as possible in a short time from a protein sample that could represent a protein complex [64], a relatively simple whole organism [67], or a tissue or other sample from a more complex organism [69].

3. Quantitative proteomics

3.1. Comparative 2D electrophoresis

The technique that is still the most widely used in global proteome analysis is two-dimensional electrophoresis (IEF–SDS–PAGE) (2DE), as explained in Section 2.1. In 2DE, proteins are first separated by isoelectric focusing and then further resolved by SDS–PAGE in the second, perpendicular, dimension. Separated proteins can then be visualized by numerous staining methods, or by autoradiography, to produce a two-dimensional image array that can contain thousands of proteins [20,21].

Although 2DE is not strictly quantitative, the presence or absence of one or more spots in a gel, when compared to a similar gel, is often readily detectable. Using this approach, the state of a cellular system in response to a particular treatment can be assessed using 2DE of samples from each state. This approach enables the simultaneous assessment of the effect of the treatment on many proteins at once, rather than measuring, for example, levels of a single marker protein. This type of experiment is often referred to as differential display proteomics [70]. Examples where this can be used to directly visualize physiologically relevant proteins include the characterization of changes in growth or nutrient conditions [71], characterization of tumor specific proteins [72–75], cell differentiation studies [76], treatment of cultured cells with a potential therapeutic drug [77], or identifying mechanisms of drug resistance in cultured cells [78]. Once these protein spots have been visualized and identified, knowledge of the proteins that are directly affected during such a treatment can identify the biochemical pathways involved, and therefore be of great value in deciding the direction of future research. In this implementation, proteome analysis is used as a biological assay

rather than as a database of protein identification information as described previously.

Thus, despite numerous drawbacks and limitations, 2DE remains an important tool in proteome analysis [19]. It is clear, however, that there is room for improvement in the efficiency of analysis, and either incremental advances in current methods or development of new technologies may achieve this.

One of the major technical limitations in comparative 2D gel electrophoresis is the difficulty in identifying matching protein spots between gel images [79,80]. The inherent variability of 2DE separation often makes it difficult to be certain with a high degree of confidence that a particular 2DE spot on one gel actually represents the same protein on a different gel. This variability between 2DE gels can include spots streaking, gels bending or warping in broad or localized areas, or different spots being visualized. All of these factors combine to make it a difficult, laborious task to compare gel images, despite the development of numerous commercial software packages designed to assist in this task.

One experimental approach that has been developed in recent years with the aim of overcoming this inherent inter-gel variability is known as difference gel electrophoresis (DIGE). This involves labeling two protein samples for comparison with two different fluorescent dyes prior to the first dimension of 2DE [81]. The combined samples are separated using the same first and second dimension gels, thus minimizing any inter-gel variations. The gel images are then visualized using fluorescent scanning at two separate wavelengths specific to the two fluorescent dyes. This enables the proteins present in each of the original samples to be viewed separately, and makes even relatively subtle differences in protein expression levels between the two samples immediately apparent. However, it is has already been found to be reproducible and sensitive enough for use in identification of protein expression level changes in tumor tissues [82], and is sure to become more widely used as the technology becomes more readily available. A variation to this approach has also been recently reported, in which two different samples are both metabolically radiolabeled in vivo, one with ^{14}C and one with ^{3}H. The samples are then combined and run together on a single 2DE gel. The $^{3}H/^{14}C$ ratio of each protein spot is determined by exposure to two types of imaging plates, one sensitive to ^{14}C and the other to both ^{14}C and ^{3}H. This technique was used to compare the cellular levels of several hundred proteins in yeast cells [83].

3.2. In vivo metabolic labeling for quantitative proteomics

Quantitative analysis of protein expression can employ either stable-isotopes or radioisotope methods. In a typical radiolabeling approach the cells are cultured in the presence of ^{35}S-methionine. The cell lysate is then separated by a high-resolution method such as two-dimensional gel electrophoresis, the changes in protein concentration are determined by autoradiography, and the identity of the proteins are determined by mass spectrometry [84]. Limitations of this method include both the need to first gel separate the proteins by electrophoresis, and the practical difficulties inherent in the use of radioisotopes.

Another promising method is the use of stable-isotopes for the purposes of protein quantitation by mass spectrometry [85–89]. The basis of this method is utilizing the ability of the mass spectrometer ability to differentiate the change in mass of a protein or peptide that is introduced by a non-abundant stable-isotope during cell culture [90,91]. One pool of cells is grown on medium containing the naturally occurring abundance of the stable-isotopes ^{14}N (99.6%), and ^{15}N (0.4%), while a second pool is grown on the same medium enriched in ^{15}N ($>$96%). The two sample pools are combined and the resulting proteins or peptides are separated by either two-dimensional gel electrophoresis [85] or two-dimensional chromatography [88]. The proteins are proteolyzed, the resulting peptides are analyzed by MS–MS, and the resulting spectra are used to both identify the protein and determine the relative abundance in the two cellular protein extracts. The MS analysis is able to differentiate between the peptides occurring from the two protein pools because incorporation of the ^{15}N increases the mass of the peptides leading to a pair of peaks for each peptide. The ratios of the intensities of the lower and upper mass components can then be compared to provide an accurate measurement of the relative

abundances of the proteins from the original lysates [85,86]. This approach has recently been reported in the analysis of more than 800 unique proteins from a yeast lysate, demonstrating the utility of this method for large-scale quantitative proteomic analysis [88]. The obvious limitations of this method are, however, the restriction to cell culture experiments, and the possibility that introduction of non-natural isotope ratios may perturb cellular systems in an undesired manner.

3.3. In vitro stable isotope labeling of protein mixtures

Several methods exist for post-translational labeling of proteins and peptides for quantitative analysis [89,92]. One of the most interesting emerging technologies in proteomics is known as isotope-coded affinity tag (ICAT) peptide labeling [93]. This is an approach that combines accurate quantification and concurrent sequence identification of the individual proteins in complex mixtures. When used for the pairwise comparison of protein samples isolated from two different cell states, this method can provide simultaneous identification and quantification of up- or down-regulated proteins. The initial incarnation of the ICAT reagents contained a biotin affinity tag and a thiol specific reactive group, which are joined by a spacer domain which is available in two forms; regular and isotopically heavy, which includes eight deuterium atoms [93]. The method is based on combining peptide labeling with ICAT reagents with analysis of the labeled peptides by HPLC and tandem mass spectrometry.

The method consists of four major steps. Firstly, a mixture of reduced proteins representing one cell state is derivatized with the isotopically light version of the ICAT reagent, while the corresponding reduced protein mixture representing a second cell state is derivatized with the isotopically heavy version of the ICAT reagent. Secondly, the labeled samples are combined and digested with a protease to produce peptide fragments. Thirdly, the tagged cysteine containing peptide fragments are isolated by avidin affinity chromatography. Finally, the isolated tagged peptides are separated and analyzed by capillary HPLC–tandem mass spectrometry, which provides both identification of the peptides by fragmentation in MS–MS mode and relative quantitation of labeled pairs by comparing signal intensities in MS mode. This is similar in analytical approach to metabolic labeling, with the crucial distinction being that this method is not limited in scope to organisms that can be successfully cultured in metabolic labeling media.

There are several advantages of this approach when compared to the more traditional differential display 2DE method. This approach is much faster than is time-consuming 2DE experiments, it is scaleable so that, in theory, a large enough amount of sample can be used to enable analysis of low abundance proteins, and since it is based on stable isotope labeling of isolated protein samples, it does not require the use metabolic labeling or radioactivity. Most important of all, however, is that it provides accurate relative quantification of each peptide identified. For example, if a protein is present at the same level in the two original samples, the amount of each peptide detected will be the same. If, however, a protein is present at a fivefold higher level in the sample derivatized with the heavy ICAT reagent, then the amount of heavy ICAT labeled peptide detected will be five times greater than the amount of light ICAT labeled peptide detected. Although mass spectrometry is an essentially non-quantitative technique, in this case the peptides act as mutual internal standards, since they are chemically identical and differ only by eight neutrons, and thereby eliminate potential problems due to differing ionization efficiencies or other physicochemical properties.

This method has already been shown to be applicable to the identification and quantitation of proteins in cellular systems undergoing different perturbations. These include the characterization of protein expression in in vitro differentiated human myeloid leukemia tissue culture cells [94], and the identification and quantification of galactose and glucose repressed proteins in yeast harvested under different growth conditions [95]. Also, this approach has recently been refined by using several MS techniques in combination to produce a system in which only those proteins undergoing significant changes in expression levels are subjected to tandem mass spectrometry and thereby identified [96]. With fur-

ther development, this technology may be able to identify and quantify even the most subtle protein expression level changes between samples. This type of data is usually lost in the context of large amounts of very abundant background proteins.

There are also several obvious disadvantages to this technique as it is currently employed, but all of them appear to be surmountable in the course of future development. The proteins must first of all contain cysteine, which is true for an estimated 92% of yeast proteins, for example, and appropriately spaced protease cleavage sites must also flank those cysteines. Moreover, the ICAT tag is a large moiety when compared to the size of some small peptides and thus may interfere with peptide ionization and can greatly complicate mass spectral interpretation [97]. It seems likely that all of these problems may be overcome by designing different reagents with specificity for other peptide side-chains, using a smaller tag group, and using different proteases. A good example of how this approach could be expanded to include other functional groups was described in a recent report in which a version of the ICAT reagents was employed for the specific analysis of phosphopeptides, albeit with fairly limited sensitivity [98]. In addition, many other research groups are now producing variations on the original theme, such as the differential lysine guanidation technique known as mass-coded abundance tagging (MCAT), which was shown to be useful in both protein quantitation and de novo peptide sequence identification [99].

The recently published report using ICAT methodology in the identification and quantification of galactose and glucose repressed proteins in yeast harvested under different growth conditions represents a significant advance in the field of proteomics [95]. This analysis combined protein identification and quantification data with DNA microarray data to produce a comprehensive picture of changes in cellular pathways caused by an experimentally-induced stress. This type of integrated approach, combing both genomic and proteomic data, has given rise to a new field, known as systems biology [100]. It is to be hoped that further research and development in this area will yield even more promising data in the future.

4. Protein fractionation techniques in proteomics

4.1. Protein enrichment by tissue and subcellular fractionation

The inability of the currently available proteomics methods to analyze all of the proteins in a complex organism or cell can be partially overcome by appropriate purification steps before a sample is analyzed by mass spectrometry. This can be done by purification of specific tissues from an organism, purification of subcellular fractions, purification of protein complexes, or purification of protein fractions by chromatographic steps.

A good example of tissue fractionation occurs in the proteomic analysis of plants. Most studies begin by separating the plant, at the very least, into the three major tissues; leaf, root and seed [28,56,101]. Although these may all be from the same plant, they produce very different protein expression patterns as each tissue exists in a very different environment from the others, and each tissue performs specific biological functions. This paradigm also holds true in the analysis of protein samples from animals, which are generally taken from a particular organ, such as the kidney [102–104], or a particular physiological fluid such as plasma [105] or serum [106].

One application where fractionation occurs at the cellular rather than tissue level is in the identification of markers of diseased tissue, where the affected cells must be separated from the surrounding normal cells. This can be done by manual microdissection [107,108], by laser capture microdissection [109], or by immunoisolation of cell type using antibodies to specific markers on the specified cells [110–112].

The next stage in sample complexity reduction after individual cell separation is the isolation and analysis of subcellular fractions using a combination of standard biochemical techniques and proteomics methodologies. This has been reported for many different subcellular fractions and organelles, including such recent examples as the proteomic characterizations of complete golgi [113], the golgi membrane [114], the mitochondria [115], the chloroplast [116], the chloroplast membrane [117], and the nuclear envelope [118].

Sample complexity can be reduced still further by the isolation of protein complexes. This can be achieved by analysis of endogenous cellular protein complexes, including such examples as the nuclear pore of yeast cells [119], the proteasome [120], the preprotein translocase [121], the ribosome [64,122–124], and the human nucleolus [125,126]. An alternative approach to the analysis of cellular protein complexes is to insert an epitope tag into a protein via recombinant methods, then isolate and subsequently identify all of the proteins that interact with the tagged protein [127]. This technique was employed on a grand scale in a pair of recent reports that uncovered a wealth of new information regarding protein–protein interactions in yeast [128,129].

Complex protein mixtures assembled into cellular material can also be fractionated still further by separating the proteins according to their chromatographic properties, as an initial step prior to further analysis by other proteomics methods [130]. This has been applied, for example, in: the separation of proteins from *E. coli* by anion-exchange chromatography prior to further separation by 2DE [131]; the profiling of bacterial proteins by reversed-phase chromatography prior to identification by MALDI–TOF–MS and capillary electrophoresis–electrospray ionization MS [132]; and in the differential screening and mass mapping of proteins from premalignant and cancer cell lines using nonporous reversed-phase HPLC coupled with ESI–TOF and MALDI–TOF MS analysis [133].

4.2. Protein fractionation by structural groups

Complex protein samples may also be simplified by fractionation on the basis of structural groups such as carbohydrate chains, phosphate groups, or lipid modifications such as acylation, prenylation, and GPI-anchoring. This will not necessarily assign the purified proteins a known function, but can be useful for comparative analysis of samples. Most gene products are either co- or post-translationally modified. The two most prevalent modifications are glycosylation and phosphorylation. The extent and type of protein glycosylation has long been known to effect protein localization, binding, and activity [134–136]. The glycosylation profile of many proteins is also known to shift dramatically in a number of disease states, including cancer [137], diabetes [138], and autoimmune diseases such as arthritis [139], Sjogren's syndrome [140], and Alzheimer's disease [141]. Lectin affinity chromatography in combination with proteomics techniques may be used in the large-scale capture and characterization of glycoproteins [142]. Glycosylation mapping has recently been used to identify post-translational modifications associated with breast cancer [143], and this approach is readily adaptable to a wide range of disease studies. 2DE gel separation of glycoproteins results in the classic pattern of a ''train'' of protein spots separated on the basis of different isoelectric points. However, typical methods of glycan analysis are not sensitive enough for the levels generally separated by 2DE, and the massive complexity of glycan structures does not lend itself well to most currently available proteomics tools for studying post-translational modification. Recent developments in mass spectrometry technologies, however, have enabled the characterization of carbohydrate structures at picomolar levels [144–146]. These advances should soon allow detailed structural analysis of differentially glycosylated proteins.

Analysis of the phosphoproteome is considerably more advanced. Reversible protein phosphorylation has long been considered one of the most important factors controlling protein activity and function. It is possible to detect phosphorylated proteins by 2DE analysis; as with glycosylation, candidate proteins are present as a trail of spots of slightly differing p*I*. 2DE has been used in a number of studies, whether by using computer assisted differential spot analysis, or in combination with phospho-labelling and Western blotting, to look for specific phosphorylated target proteins [147–149].

Global detection of the majority of phosphorylated proteins present in a complex sample can be accomplished by screening 2DE gels with highly specific anti-phosphotyrosine and anti-phosphoserine antibodies [150,151]. This method may be extended to compare changes in the pattern or extent of phosphorylation between samples. Phage-display techniques provide an alternative method for screening complex samples for specific phosphorylated pro-

teins or groups of proteins such as kinases. As an alternative to large-scale display techniques, protein samples may be enriched for phosphoproteins or phosphopeptides prior to standard analysis by matrix-assisted laser desorption–ionization (MALDI–TOF) MS or nanoelectrospray tandem mass spectrometry. For example, affinity chromatography with anti-phosphotyrosine or anti-phosphoserine antibodies has been used in combination with mass spectrometry to identify specific phosphorylated targets [152,153]. Immobilized metal ion affinity columns of either Fe^{3+} or Ga^{3+} can also be used to selectively enrich for phosphopeptides. Notwithstanding the fact that strongly negatively charged peptides can interfere in this analysis, this procedure enables the subsequent identification and mapping of phosphorylation sites in the enriched proteins [154,155]. Chemical replacement of phosphate moieties by affinity tags can also be used to enrich for phosphoproteins in complex protein samples [156], and this methodology has already been successfully applied to whole cell lysates from yeast [157].

Each of these various approaches can be used to reduce sample complexity in order to enable the analysis of less abundant cellular components. The added advantage of such methods is that key structural information is obtained along with the identification of the resultant subset of proteins.

4.3. Protein fractionation by functional groups

One approach that can be used for high-throughput purification of functional classes of proteins is to combine chromatography using specific affinity matrices with mass spectrometry. Most high-throughput proteomic methods result in the isolation of a number of proteins for which no function is known. The function is usually deduced using sequence similarities to proteins with known functions or the identification of motifs with a known function. The process can be time-consuming and may not result in the identification of the correct function. Affinity chromatography in combination with mass spectrometry can be used to isolate, analyze and identify both known and novel proteins. By careful selection of the affinity ligand, protein function can be assigned as well as protein identity. This method of isolation allows an immediate function to be deduced for the molecule by its ability to bind specific molecules on the affinity column.

A variety of chromatographic approaches have been discussed for fractionating complex protein mixtures in order to make them more manageable for mass spectrometric analysis. These rely on separation by size, charge, hydrophobicity, or specific affinity.

Activity-based protein fractionation may be used as a strategy that reduces sample complexity, while at the same time assigning a known function to those proteins that are isolated. That function may be broad-based, for example the use of lectin affinity chromatography to isolate the ‘‘glycome’’ of a species [158,159], the use of protein kinase chips [160,161] to look for phosphorylation targets for a diverse array of kinases or the use of polysaccharide matrices to isolate carbohydrate binding proteins [158,162]. Whole functionally related families of proteins may be isolated through the use of an appropriate affinity matrix, for example the isolation of calcium-binding proteins with calmodulin [163–165] or use of the FK2 antibody, which recognizes the conjugated ubiquitin molecule, to purify poly-ubiquitinated proteins [166,167]

Increasing specificity may be obtained by careful selection of the affinity ligand. One such method is to screen for reactivity to specific chemical probes. Studies of this kind have largely been limited to the isolation of one specific protein based on a known activity. Recently, however, several studies have been carried out [168,169] that utilize a library of chemical probes to isolate families of proteins based on their specific reactivity. In one report, a library of biotinylated sulfonate esters was applied to the screening of a complex mixture of proteins. A number of different patterns of reactivity were found, for example the irreversible inhibition of an alcohol dehydrogenase, and the partial inhibition of catalytic activity for another group of target enzymes. In another study [170], the authors were able to use differential binding to a biotinylated fluorophosphonate in order to distinguish between subsets of serine hydrolase family members.

All of the examples presented here provide relevant functional information in addition to protein

sequence identification. A great deal of flexibility is available in the specificity of any particular method for isolating target proteins, and the application of such methodology is well suited to the increasing sensitivity of proteomic technology.

4.4. Antibody technologies

The importance of antibody technology to the field of functional proteomics is growing, particularly as the field of antibody engineering reaps the benefits of the post-genomic era [171]. Antibodies may be designed with selective reactivity to an almost unlimited variety of epitopes [172–175]. One antibody-based approach that has been investigated a number of times, albeit with limited success thus far, is the use of a specific monoclonal antibody to remove the vast excess of albumin present in human serum and plasma, with the aim of improving the visualization of previously obscured less abundant proteins [176].

However, one of the limiting factors for antibody-based methods has been the ability to produce antibodies on a scale compatible with genome- or proteome-wide analysis. Hybridoma technology, a slow and cumbersome process, has been used to produce monoclonal antibodies (mAbs) for such applications. Separate immunizations are required for each antigen, and the cell fusion process required to generate the hybridoma is slow and inefficient.

Recent advances using antibody display technologies in both phage and yeast now make it possible to overcome many of the limitations of hybridoma-based systems and generate mAbs that recognize any desired antigen. The method of phage display has been developed as a means of making single-chain variable fragment (scFv) antibodies in vitro [177–179]. This technology links the antigen-specific immunoglobulin (Ig) variable (V) domains from both heavy (V_H) and light chain (V_L) chains into a single DNA coding sequence. Functional scFv antibodies, fused to a minor coat protein of a bacteriophage particle containing the gene encoding the antibody, can be isolated against any desired target from libraries containing over 10^9 different antibody specificities. The phage antibody that binds specifically to a target can be separated from non-binding antibodies and then amplified, eliminating the need for an immunization regime. The advantages of phage display include the ability to generate antibodies in a relatively short time compared to traditional hybridoma technology, and the compatibility with recent advances in the areas of laboratory automation.

One limitation in the selection procedure for this methodology is the requirement for sufficient quantities of purified protein targets. Purification is often difficult and time-consuming, and would not be feasible on a global scale. As an alternative, peptides have long been used to immunize animals for the production of antibodies against the native protein from which the peptide was derived. An obvious appeal to using peptides as immunogenic targets is the ability, in theory, to obtain scFv antibodies against any given open reading frame without needing to purify the protein.

With this technology in place, several groups are currently developing antibody-based protein chips [172,180]. Protein chips based on recombinant antibody fragments may be used in combination with mass spectrometry to detect structural modifications of single proteins, or capture specific proteins or protein complexes via the use of high affinity sites. Protein–protein interactions may be examined on a large scale, through the use of epitope directed analysis. Antibody arrays can be used to detect differentially expressed proteins, allowing for the possibility of comparative analysis between disease states, tissue types, and treatment conditions.

5. Conclusions

Functional proteomics can potentially provide qualitative, quantitative, and functional information on all proteins present within a biological system. The ability to identify a large number of proteins, and characterize their differential expression and post-translational modifications, will contribute significantly to the understanding of disease and drug discovery. Furthermore, since it is proteins rather than genes that are directly responsible for most reactions within a cell under any given condition, this information promises to have a greater impact on the biomedical field than could be derived from genomic information alone. The promise of the

discoveries yet to be made in this field, the recent technological advances in mass spectroscopy, and the completion of many different genome-sequencing projects, have all contributed to the explosive growth in the field of proteomics.

In this review, we have shown that there are a very wide range of biological problems to be solved in this area, and an equally diverse range of technological solutions available. Since many functional proteomics applications involve extremely complicated interacting networks of genes and proteins, it is clear that no single methodology will be sufficient in most cases.

Mass spectrometry is still one of the most important tools in the functional proteomics toolbox. Continued technological advances, such as the development of new MALDI–QTOF [181–183] and MALDI–TOF–TOF [184] MS instruments, will provide lower detection limits, improved specificity and higher throughput for analysis of complex mixtures. Rapid advances are also occurring in other technological areas such as protein fractionation, quantitative protein labeling and chromatographic separation techniques. Taken together, these advances will provide the post-genomic biologist with the ability to exploit the wealth of available genomic data in order to answer questions of biological significance.

6. Nomenclature

2DE	two-dimensional electrophoresis
AIDS	acquired immune deficiency syndrome
ESI	electrospray ionization
ESI–TOF	electrospray ionization–time of flight
EST	expressed sequence tag
ICAT	isotope coded affinity tagging
IEF	isoelectric focusing
IEF–SDS–PAGE	isoelectric focusing–sodium dodecyl sulfate–polyacrylamide gel electrophoresis
HPLC	high pressure liquid chromatography
Ig	immunoglobulin
LC–LC–MS–MS	liquid chromatography–liquid chromatography–tandem mass spectrometry
LC–MS–MS	liquid chromatography tandem mass spectrometry
mAb	monoclonal antibody
MALDI–QTOF	matrix-assisted laser desorption ionization–quadrupole-time of flight
MALDI–TOF	matrix-assisted laser desorption ionization–time of flight
MALDI–TOF–TOF	matrix-assisted laser desorption ionization–time of flight–time of flight
MCAT	mass-coded abundance tagging
MS	mass spectrometry
MS–MS	tandem mass spectrometry
MudPIT	multi-dimensional protein identification technique
RP	reversed-phase
scFv	single chain variable fragment
SCX	strong cation-exchange
V	variable chain
V_H	heavy chain
V_L	light chain

Acknowledgements

The authors wish to acknowledge the continued support of Steve Briggs and Karl Strauss. PH would like to dedicate this paper to the memory of Gethyn Allen, whose spirit and enthusiasm will be with me always.

References

[1] C. Ash, B. Jasny, Science 295 (2002) 2035.
[2] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt et al., Science 291 (2001) 1304.
[3] C. Nock, C. Gauss, L.C. Schalkwyk, J. Klose, H. Lehrach, H. Himmelbauer, Electrophoresis 20 (1999) 1027.
[4] C. Janke, M. Holzer, J. Klose, T. Arendt, FEBS Lett. 379 (1996) 222.
[5] A.J. Sender, in: Genome Technology, 2002, p. 50.

[6] R.D. Fleischmann, M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.-F. Tomb, B.A. Dougherty, J.M. Merrick et al., Science 269 (1995) 496.
[7] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston et al., Science 274 (1996) 546.
[8] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew et al., Science 277 (1997) 1453.
[9] Consortium, Science 282 (1998) 2012.
[10] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle et al., Science 287 (2000) 2185.
[11] S.A. Goff, D. Ricke, T.-H. Lan, G. Presting, R. Wang, M. Dunn, J. Glazebrook, A. Sessions, P. Oeller, H. Varma et al., Science 296 (2002) 92.
[12] D.F. Hunt, A.M. Buko, J.M. Ballard, J. Shabanowitz, A.B. Giordani, Biomed. Mass Spectrom. 8 (1981) 397.
[13] D.F. Hunt, J.R. Yates, J. Shabanowitz, S. Winston, C.R. Hauer, Proc. Natl. Acad. Sci. USA 83 (1986) 6233.
[14] J. Eng, A.L. McCormack, J.R. Yates, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[15] J.R. Yates, J.K. Eng, A.L. McCormack, D. Schieltz, Anal. Chem. 67 (1995) 1426.
[16] J.R. Yates, Electrophoresis 19 (1998) 893.
[17] J.R. Yates, J. Mass Spectrom. 33 (1998) 1.
[18] A.R. Dongre, J.K. Eng, J.R. Yates, Trends Biotechnol. 15 (1997) 418.
[19] T. Rabilloud, Proteomics 2 (2002) 3.
[20] A. Gorg, W. Postel, S. Gunther, Electrophoresis 9 (1988) 531.
[21] J. Klose, U. Kobalz, Electrophoresis 16 (1995) 1034.
[22] B. Honore, H. Leffers, P. Madsen, J.E. Celis, Eur. J. Biochem. 218 (1993) 421.
[23] P. Matsudaira, J. Biol. Chem. 262 (1987) 10035.
[24] R.H. Aebersold, D.B. Teplow, L.E. Hood, S.B. Kent, J. Biol. Chem. 261 (1986) 4229.
[25] R.H. Aebersold, J. Leavitt, R.A. Saavedra, L.E. Hood, S.B. Kent, Proc. Natl. Acad. Sci. USA 84 (1987) 6970.
[26] J. Rosenfeld, J. Capdevielle, J.C. Guillemot, P. Ferrara, Anal. Biochem. 203 (1992) 173.
[27] M. Mann, M. Wilm, Anal. Chem. 66 (1994) 4390.
[28] L. Porubleva, K. Vander Velden, S. Kothari, D.J. Oliver, P.R. Chitnis, Electrophoresis 22 (2001) 1724.
[29] T. Rabilloud, Methods Mol. Biol. 112 (1999) 297.
[30] T.H. Steinberg, L.J. Jones, R.P. Haugland, V.L. Singer, Anal. Biochem. 239 (1996) 223.
[31] T.H. Steinberg, R.P. Haugland, V.L. Singer, Anal. Biochem. 239 (1996) 238.
[32] W.F. Patton, J. Chromatogr. B 771 (2002) 3.
[33] W.M. Lauber, J.A. Carroll, D.R. Dufield, J.R. Kiesel, M.R. Radabaugh, J.P. Malone, Electrophoresis 22 (2001) 906.
[34] V. Santoni, T. Rabilloud, P. Doumas, D. Rouquie, M. Mansion, S. Kieffer, J. Garin, M. Rossignol, Electrophoresis 20 (1999) 705.
[35] V. Santoni, P. Doumas, D. Rouquie, M. Mansion, T. Rabilloud, M. Rossignol, Biochimie 81 (1999) 655.
[36] M.P. Molloy, B.R. Herbert, K.L. Williams, A.A. Gooley, Electrophoresis 20 (1999) 701.
[37] S.P. Gygi, Y. Rochon, B.R. Franza, R. Aebersold, Mol. Cell Biol. 19 (1999) 1720.
[38] M.P. Molloy, B.R. Herbert, B.J. Walsh, M.I. Tyler, M. Traini, J.C. Sanchez, D.F. Hochstrasser, K.L. Williams, A.A. Gooley, Electrophoresis 19 (1998) 837.
[39] A. Di Luccia, L. Iannibelli, E. Addato, B. Masala, L. Manca, L. Ferrara, Comp. Biochem. Physiol. B 99 (1991) 887.
[40] J.S. Esteve-Romero, I.M. Yman, A. Bossi, P.G. Righetti, Electrophoresis 17 (1996) 1380.
[41] P.E. Hodges, W.E. Payne, J.I. Garrels, Nucleic Acids Res. 26 (1998) 68.
[42] R.A. VanBogelen, K.Z. Abshire, B. Moldover, E.R. Olson, F.C. Neidhardt, Electrophoresis 18 (1997) 1243.
[43] A.J. Link, L.G. Hays, E.B. Carmack, J.R. Yates, Electrophoresis 18 (1997) 1314.
[44] B.L. Urquhart, T.E. Atsalos, D. Roach, D.J. Basseal, B. Bjellqvist, W.L. Britton, I. Humphery-Smith, Electrophoresis 18 (1997) 1384.
[45] V.C. Wasinger, B. Bjellqvist, I. Humphery-Smith, Electrophoresis 18 (1997) 1373.
[46] C.D. O'Connor, M. Farris, R. Fowler, S.Y. Qi, Electrophoresis 18 (1997) 1483.
[47] S.J. Cordwell, D.J. Basseal, I. Humphery-Smith, Electrophoresis 18 (1997) 1335.
[48] T. Sazuka, O. Ohara, Electrophoresis 18 (1997) 1252.
[49] J.X. Yan, L. Tonella, J.C. Sanchez, M.R. Wilkins, N.H. Packer, A.A. Gooley, D.F. Hochstrasser, K.L. Williams, Electrophoresis 18 (1997) P 491.
[50] N. Guerreiro, J.W. Redmond, B.G. Rolfe, M.A. Djordjevic, Mol. Plant Microb. Interact. 10 (1997) 506.
[51] R.D. Appel, J.C. Sanchez, A. Bairoch, O. Golaz, M. Miu, J.R. Vargas, D.F. Hochstrasser, Electrophoresis 14 (1993) 1232.
[52] J. Celis, P. Gromov, M. Ostergaard, P. Madsen, B. Honore, K. Dejgaard, E. Olsen, H. Vorum, D.B. Kristensen, I. Gromova et al., FEBS Lett. 398 (1996) 129.
[53] P.A. Haynes, I. Miller, R. Aebersold, M. Gemeiner, I. Eberini, M. Rosa Lovati, C. Manzoni, M. Vignati, E. Gianazza, Electrophoresis 19 (1998) 1484.
[54] I. Miller, P.A. Haynes, M. Gemeiner, R. Aebersold, I. Eberini, M. Rosa Lovati, C. Manzoni, M. Vignati, E. Gianazza, Electrophoresis 19 (1998) 1493.
[55] I. Miller, P.A. Haynes, I. Eberini, M. Gemeiner, R. Aebersold, E. Gianazza, Electrophoresis 20 (1999) 836.
[56] U. Mathesius, G. Keijzers, S.H. Natera, J.J. Weinman, M.A. Djordjevic, B.G. Rolfe, Proteomics 1 (2001) 1424.
[57] D. Butler, Nature 403 (2000) 815.
[58] E.F. Petricoin, Proteomics 1 (2001) 637.
[59] N.G. Anderson, A. Matheson, N.L. Anderson, Proteomics 1 (2001) 3.
[60] P. Bohlen, F. Castillo, N. Ling, R. Guillemin, Int. J. Pept. Protein Res. 16 (1980) 306.
[61] J. Rivier, R. McClintock, R. Galyean, H. Anderson, J. Chromatogr. 288 (1984) 303.

[62] M.T. Davis, C.S. Spahr, M.D. McGinley, J.H. Robinson, E.J. Bures, J. Beierle, J. Mort, W. Yu, R. Luethy, S.D. Patterson, Proteomics 1 (2001) 108.
[63] C.S. Spahr, M.T. Davis, M.D. McGinley, J.H. Robinson, E.J. Bures, J. Beierle, J. Mort, P.L. Courchesne, K. Chen, R.C. Wahl et al., Proteomics 1 (2001) 93.
[64] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mize, D.R. Morris, B.M. Garvik, J.R. Yates, Nature Biotechnol. 17 (1999) 676.
[65] G.J. Opiteck, K.C. Lewis, J.W. Jorgenson, R.J. Anderegg, Anal. Chem. 69 (1997) 1518.
[66] G.J. Opiteck, S.M. Ramirez, J.W. Jorgenson, M.A. Moseley, Anal. Biochem. 258 (1998) 349.
[67] M.P. Washburn, D. Wolters, J.R. Yates, Natute Biotechnol. 19 (2001) 242.
[68] A. Shevchenko, I. Chernushevich, W. Ens, K.G. Standing, B. Thomson, M. Wilm, M. Mann, Rapid Commun. Mass Spectrom. 11 (1997) 1015.
[69] A. Koller, M. Washburn, N. Andon, C. Deciu, P.A. Haynes, L. Hays, D. Schieltz, R. Ulaszek, J. Wei et al., Proc. Natl. Acad. Sci. USA (2002) in press.
[70] S.J. Cordwell, A.S. Nouwens, N.M. Verrills, D.J. Basseal, B.J. Walsh, Electrophoresis 21 (2000) 1094.
[71] B. Franzen, S. Becker, R. Mikkola, K. Tidblad, A. Tjernberg, S. Birnbaum, Electrophoresis 20 (1999) 790.
[72] M. Ahram, C.J. Best, M.J. Flaig, J.W. Gillespie, I.M. Leiva, R.F. Chuaqui, G. Zhou, H. Shu, P.H. Duray, W.M. Linehan et al., Mol. Carcinog. 33 (2002) 9.
[73] S.O. Lim, S.J. Park, W. Kim, S.G. Park, H.J. Kim, Y.I. Kim, T.S. Sohn, J.H. Noh, G. Jung, Biochem. Biophys. Res. Commun. 291 (2002) 1031.
[74] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn et al., Lancet 359 (2002) 572.
[75] A.A. Alaiya, M. Oppermann, J. Langridge, U. Roblick, L. Egevad, S. Brindstedt, M. Hellstrom, S. Linder, T. Bergman, H. Jornvall et al., Cell. Mol. Life Sci. 58 (2001) 307.
[76] A. Alaiya, U. Roblick, L. Egevad, A. Carlsson, B. Franzen, D. Volz, S. Huwendiek, S. Linder, G. Auer, Anal. Cell. Pathol. 21 (2000) 1.
[77] C.Z. Shi, H.W. Collins, C.W. Buettger, W.T. Garside, F.M. Matschinsky, S. Heyner, Mol. Reprod. Dev. 37 (1994) 398.
[78] P. Sinha, G. Hutter, E. Kottgen, M. Dietel, D. Schadendorf, H. Lage, J. Biochem. Biophys. Methods 37 (1998) 105.
[79] J.M. Corbett, M.J. Dunn, A. Posch, A. Gorg, Electrophoresis 15 (1994) 1205.
[80] A. Blomberg, L. Blomberg, J. Norbeck, S.J. Fey, P.M. Larsen, M. Larsen, P. Roepstorff, H. Degand, M. Boutry, A. Posch et al., Electrophoresis 16 (1995) 1935.
[81] R. Tonge, J. Shaw, B. Middleton, R. Rowlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie, M. Davison, Proteomics 1 (2001) 377.
[82] F. Von Eggeling, A. Gawriljuk, W. Fiedler, G. Ernst, U. Claussen, J. Klose, I. Romer, Int. J. Mol. Med. 8 (2001) 373.
[83] C. Monribot-Espagne, H. Boucherie, Proteomics 2 (2002) 229.
[84] T. Kanamoto, U. Hellman, C.H. Heldin, S. Souchelnytskyi, EMBO J. 21 (2002) 1219.
[85] Y. Oda, K. Huang, F.R. Cross, D. Cowburn, B.T. Chait, Proc. Natl. Acad. Sci. USA 96 (1999) 6591.
[86] T.D. Veenstra, S. Martinovic, G.A. Anderson, L. Pasa-Tolic, R.D. Smith, J. Am. Soc. Mass Spectrom. 11 (2000) 78.
[87] T.P. Conrads, K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. Udseth, W.B. Chrisler et al., Anal. Chem. 73 (2001) 2132.
[88] W. Washburn, R. Ulaszek, C. Deciu, D.M. Schieltz, J.R. Yates, Anal. Chem. 74 (2002) 1650.
[89] F.E. Regnier, L. Riggs, R. Zhang, L. Xiong, P. Liu, A. Chakraborty, E. Seeley, C. Sioma, R.A. Thompson, J. Mass Spectrom. 37 (2002) 133.
[90] B.W. Patterson, D.L. Hachey, G.L. Cook, J.M. Amann, P.D. Klein, J. Lipid Res. 32 (1991) 1063.
[91] B.W. Patterson, F. Carraro, R.R. Wolfe, Biol. Mass Spectrom. 22 (1993) 518.
[92] M. Munchbach, M. Quadroni, G. Miotto, P. James, Anal. Chem. 72 (2000) 4047.
[93] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nature Biotechnol. 17 (1999) 994.
[94] D.K. Han, J. Eng, H. Zhou, R. Aebersold, Nature Biotechnol. 19 (2001) 946.
[95] 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.
[96] T.J. Griffin, D.K. Han, S.P. Gygi, B. Rist, H. Lee, R. Aebersold, K.C. Parker, J. Am. Soc. Mass Spectrom. 12 (2001) 1238.
[97] R. Zhang, C.S. Sioma, S. Wang, F.E. Regnier, Anal. Chem. 73 (2001) 5142.
[98] H. Zhou, J.D. Watts, R. Aebersold, Nature Biotechnol. 19 (2001) 375.
[99] G. Cagney, A. Emili, Nature Biotechnol. 20 (2002) 163.
[100] L. Bonetta, Nature Med. 8 (2002) 315.
[101] K. Gallardo, C. Job, S.P. Groot, M. Puype, H. Demol, J. Vandekerckhove, D. Job, Plant Physiol. 126 (2001) 835.
[102] S. Balabanov, U. Zimmermann, C. Protzel, C. Scharf, K.J. Klebingat, R. Walther, Eur. J. Biochem. 268 (2001) 5977.
[103] P. Cutler, D.J. Bell, H.C. Birrell, J.C. Connelly, S.C. Connor, E. Holmes, B.C. Mitchell, S.Y. Monte, B.A. Neville, R. Pickford et al., Electrophoresis 20 (1999) 3647.
[104] C.S. Klade, T. Voss, E. Krystek, H. Ahorn, K. Zatloukal, K. Pummer, G.R. Adolf, Proteomics 1 (2001) 890.
[105] H.M. Georgiou, G.E. Rice, M.S. Baker, Proteomics 1 (2001) 1503.
[106] S. Hanash, F. Brichory, D. Beer, Dis. Markers 17 (2001) 295.
[107] L. Whetsell, G. Maw, N. Nadon, D.P. Ringer, F.V. Schaefer, Oncogene 7 (1992) 2355.
[108] Z. Zhuang, P. Bertheau, M.R. Emmert-Buck, L.A. Liotta, J. Gnarra, W.M. Linehan, I.A. Lubensky, Am. J. Pathol. 146 (1995) 620.
[109] R.A. Craven, N. Totty, P. Harnden, P.J. Selby, R.E. Banks, Am. J. Pathol. 160 (2002) 815.

[110] P.C. Baer, W.A. Nockher, W. Haase, J.E. Scherberich, Kidney Int. 52 (1997) 1321.
[111] M.J. Page, B. Amess, R.R. Townsend, R. Parekh, A. Herath, L. Brusten, M.J. Zvelebil, R.C. Stein, M.D. Waterfield, S.C. Davies et al., Proc. Natl. Acad. Sci. USA 96 (1999) 12589.
[112] M.A. Reymond, J.C. Sanchez, G.J. Hughes, K. Gunther, J. Riese, S. Tortola, M.A. Peinado, T. Kirchner, W. Hohenberger, D.F. Hochstrasser et al., Electrophoresis 18 (1997) 2842.
[113] C.C. Wu, J.R. Yates, M.C. Neville, K.E. Howell, Traffic 1 (2000) 769.
[114] A.W. Bell, M.A. Ward, W.P. Blackstock, H.N.M. Freeman, J.S. Choudhary, A.P. Lewis, D. Chotai, A. Fazel, J.N. Gushue, J. Paiement et al., J. Biol. Chem. 276 (2001) 5152.
[115] M.F. Lopez, B.S. Kristal, E. Chernokalskaya, A. Lazarev, A.I. Shestopalov, A. Bogdanova, M. Robinson, Electrophoresis 21 (2000) 3427.
[116] J.B. Peltier, G. Friso, D.E. Kalume, P. Roepstorff, F. Nilsson, I. Adamska, K.J. van Wijk, Plant Cell 12 (2000) 319.
[117] M. Hippler, J. Klein, A. Fink, T. Allinger, P. Hoerth, Plant J. 28 (2001) 595.
[118] M. Dreger, L. Bengtsson, T. Schoneberg, H. Otto, F. Hucho, Proc. Natl. Acad. Sci. USA 98 (2001) 11943.
[119] M.P. Rout, J.D. Aitchison, A. Suprapto, K. Hjertaas, Y. Zhao, B.T. Chait, J. Cell Biol. 148 (2000) 635.
[120] R. Verma, S. Chen, R. Feldman, D. Schieltz, J.R. Yates, J. Dohmen, R.J. Deshaies, Mol. Biol. Cell 11 (2000) 3425.
[121] W. Werhahn, A. Niemeyer, L. Jansch, V.V. Kruft, U.K. Schmitz, H.P. Braun, Plant Physiol. 125 (2001) 943.
[122] K. Yamaguchi, A.R. Subramanian, J. Biol. Chem. 275 (2000) 28466.
[123] K. Yamaguchi, K. von Knoblauch, A.R. Subramanian, J. Biol. Chem. 275 (2000) 28455.
[124] R.J. Arnold, J.P. Reilly, Anal. Biochem. 269 (1999) 105.
[125] J.S. Andersen, C.E. Lyon, A.H. Fox, A.K.L. Leung, Y.W. Lam, H. Steen, M. Mann, A.I. Lamond, Curr. Biol. 12 (2002) 1.
[126] A.H. Fox, Y.W. Lam, A.K.L. Leung, C.E. Lyon, J. Andersen, M. Mann, A.I. Lamond, Curr. Biol. 12 (2002) 13.
[127] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, Nature Biotechnol. 17 (1999) 1030.
[128] A.C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat et al., Nature 415 (2002) 141.
[129] Y. Ho, A. Gruhler, A. Heilbut, G.D. Bader, L. Moore, S.L. Adams, A. Millar, P. Taylor, K. Bennett, K. Boutilier et al., Nature 415 (2002) 180.
[130] S. Krishnan, J.E. Hale, G.W. Becker, in: H.C. Kirst, W.C. Yeh, M.J. Zmijewski (Eds.), Enzyme Technology for Pharmaceutical and Biotechnological Applications, Marcel Dekker, New York, 2001, p. 575.
[131] A. Butt, M.D. Davison, G.J. Smith, J.A. Young, S.J. Gaskell, S.G. Oliver, R.J. Beynon, Proteomics 1 (2001) 42.
[132] B.E. Chong, J. Kim, D.M. Lubman, J.M. Tiedje, S. Kathariou, J. Chromatogr. B 748 (2000) 167.
[133] B.E. Chong, R.L. Hamler, D.M. Lubman, S.P. Ethier, A.J. Rosenspire, F.R. Miller, Anal. Chem. 73 (2001) 1219.
[134] F.J. Castellino, Semin. Thromb. Hemost. 10 (1984) 18.
[135] M. Gonzalez-Gronow, J.M. Edelberg, S.V. Pizzo, Biochemistry 28 (1989) 2374.
[136] S.R. Pirie-Shepherd, E.A. Jett, N.L. Andon, S.V. Pizzo, J. Biol. Chem. 270 (1995) 5877.
[137] H. Kuniyasu, N. Oue, A. Wakikawa, H. Shigeishi, N. Matsutani, K. Kuraoka, R. Ito, H. Yokozaki, W. Yasui, J. Pathol. 196 (2002) 163.
[138] P. Odetti, I. Aragno, R. Rolandi, S. Garibaldi, S. Valentini, L. Cosso, N. Traverso, D. Cottalasso, M.A. Pronzato, U.M. Marinari, Diabetes Metab. Res. Rev. 16 (2000) 74.
[139] P.M. Rudd, T. Elliott, P. Cresswell, I.A. Wilson, R.A. Dwek, Science 291 (2001) 2370.
[140] V. Riccieri, A. Spadaro, L. Saso, G. Valentini, E. Taccari, B. Silvestrini, Clin. Exp. Rheumatol. 19 (2001) 53.
[141] F. Liu, T. Zaidi, K. Iqbal, I. Grundke-Iqbal, R.K. Merkle, C.X. Gong, FEBS Lett. 512 (2002) 101.
[142] N. Taniguchi, A. Ekuni, J.H. Ko, E. Miyoshi, Y. Ikeda, Y. Ihara, A. Nishikawa, K. Honke, M. Takahashi, Proteomics 1 (2001) 239.
[143] M.V. Dwek, H.A. Ross, A.J. Leathem, Proteomics 1 (2001) 756.
[144] P.M. Rudd, C. Colominas, L. Royle, N. Murphy, E. Hart, A.H. Merry, H.F. Hebestreit, R.A. Dwek, Proteomics 1 (2001) 285.
[145] M. von Der Ohe, S.F. Wheeler, M. Wuhrer, D.J. Harvey, S. Liedtke, M. Muhlenhoff, R. Gerardy-Schahn, H. Geyer, R.A. Dwek, R. Geyer et al., Glycobiology 12 (2002) 47.
[146] N. Kawasaki, Y. Haishima, M. Ohta, S. Itoh, M. Hyuga, S. Hyuga, T. Hayakawa, Glycobiology 11 (2001) 1043.
[147] C. Colvis, D. Garland, Arch. Biochem. Biophys. 397 (2002) 319.
[148] Y. Iwafune, H. Kawasaki, H. Hirano, Electrophoresis 23 (2002) 329.
[149] M. Goldfarb, Electrophoresis 20 (1999) 870.
[150] K. Marcus, D. Immler, J. Sternberger, H.E. Meyer, Electrophoresis 21 (2000) 2622.
[151] M. Yanagida, Y. Miura, K. Yagasaki, M. Taoka, T. Isobe, N. Takahashi, Electrophoresis 21 (2000) 1890.
[152] L.F. Stancato, E.F. Petricoin, Electrophoresis 22 (2001) 2120.
[153] H. Steen, B. Kuster, M. Fernandez, A. Pandey, M. Mann, J. Biol. Chem. 277 (2002) 1031.
[154] A. Stensballe, S. Andersen, O.N. Jensen, Proteomics 1 (2001) 207.
[155] M.C. Posewitz, P. Tempst, Anal. Chem. 71 (1999) 2883.
[156] Y. Oda, T. Nagasu, B.T. Chait, Nature Biotechnol. 19 (2001) 379.
[157] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, F.M. White, Nature Biotechnol. 20 (2002) 301.
[158] J. Hirabayashi, Y. Arata, K. Kasai, Proteomics 1 (2001) 295.
[159] M. Geng, X. Zhang, M. Bina, F. Regnier, J. Chromatogr. B 752 (2001) 293.

[160] D.M. Williams, P.A. Cole, Trends Biochem. Sci. 26 (2001) 271.
[161] D.S. Lawrence, Genome Biol. 2 (2001) 1007.
[162] T. Larsson, J. Bergstrom, C. Nilsson, K.A. Karlsson, FEBS Lett. 469 (2000) 155.
[163] H. Liu, L. Xue, C. Li, R. Zhang, Q. Ling, Biochem. Biophys. Res. Commun. 285 (2001) 633.
[164] A.S. Reddy, I.S. Day, S.B. Narasimhulu, F. Safadi, V.S. Reddy, M. Golovkin, M.J. Harnly, J. Biol. Chem. 277 (2002) 4206.
[165] S.M. Pitson, J. D'Andrea R, L. Vandeleur, P.A. Moretti, P. Xia, J.R. Gamble, M.A. Vadas, B.W. Wattenberg, Biochem. J. 350 (2000) 429.
[166] K. Takada, T. Hirakawa, H. Yokosawa, Y. Okawa, H. Taguchi, K. Ohkawa, Biochem. J. 356 (2001) 199.
[167] T. Usuba, Y. Ishibashi, Y. Okawa, T. Hirakawa, K. Takada, K. Ohkawa, Int. J. Cancer 94 (2001) 662.
[168] G.C. Adam, B.F. Cravatt, E.J. Sorensen, Chem. Biol. 8 (2001) 81.
[169] S.N. Savinov, D.J. Austin, Comb. Chem. High Throughput Screen. 4 (2001) 593.
[170] D. Kidd, Y. Liu, B.F. Cravatt, Biochemistry 40 (2001) 4005.
[171] J.S. Huston, A.J. George, Hum. Antibodies 10 (2001) 127.
[172] C.A. Borrebaeck, S. Ekstrom, A.C. Hager, J. Nilsson, T. Laurell, G. Marko-Varga, Biotechniques 30 (2001) 1126.
[173] H.L. Brooks, A.M. Sorensen, J. Terris, P.J. Schultheis, J.N. Lorenz, G.E. Shull, M.A. Knepper, J. Physiol. 530 (2001) 359.
[174] R. Hashimoto, Y. Nakamura, K. Imamura, K. Nakadate, Y. Kashiwagi, N. Matsumoto, M. Takeda, Eur. J. Neurosci. 14 (2001) 1237.
[175] M.A. Knepper, S. Masilamani, Acta Physiol. Scand. 173 (2001) 11.
[176] B.A. Lollo, S. Harvey, J. Liao, A.C. Stevens, R. Wagenknecht, R. Sayen, J. Whaley, F.G. Sajjadi, Electrophoresis 20 (1999) 854.
[177] A. Bradbury, Trends Biotechnol. 17 (1999) 137.
[178] B. Liu, D. James, Anal. Biochem. 286 (2000) 119.
[179] R.W. Siegel, B. Allen, P. Pavlik, J.D. Marks, A. Bradbury, J. Mol. Biol. 302 (2000) 285.
[180] D.J. Cahill, J. Immunol. Methods 250 (2001) 81.
[181] A. Shevchenko, A. Loboda, W. Ens, K.G. Standing, Anal. Chem. 72 (2000) 2132.
[182] A.V. Loboda, A.N. Krutchinsky, M. Bromirski, W. Ens, K.G. Standing, Rapid Commun. Mass Spectrom. 14 (2000) 1047.
[183] A. Shevchenko, S. Sunyaev, A. Loboda, P. Bork, W. Ens, K.G. Standing, Anal. Chem. 73 (2001) 1917.
[184] K.F. Medzihradszky, J.M. Campbell, M.A. Baldwin, A.M. Falick, P. Juhasz, M.L. Vestal, A.L. Burlingame, Anal. Chem. 72 (2000) 552.