Mini review

New frontiers in proteomics research: A perspective

Vikas Dhingra [a,*], Mukta Gupta [b], Tracy Andacht [c], Zhen F. Fu [a]

[a] *Department of Pathology, University of Georgia, Athens, GA 30602, USA*
[b] *Department of Pharmaceutical and Biomedical Sciences, University of Georgia, Athens, GA 30602, USA*
[c] *Proteomics Resource Facility, Integrated Biotech Laboratories, University of Georgia, Athens, GA 30602, USA*

Abstract

Substantial advances have been made in the fundamental understanding of human biology, ranging from DNA structure to identification of diseases associated with genetic abnormalities. Genome sequence information is becoming available in unprecedented amounts. The absence of a direct functional correlation between gene transcripts and their corresponding proteins, however, represents a significant roadblock for improving the efficiency of biological discoveries. The success of proteomics depends on the ability to identify and analyze protein products in a cell or tissue and, this is reliant on the application of several key technologies. Proteomics is in its exponential growth phase. Two-dimensional electrophoresis complemented with mass spectrometry provides a global view of the state of the proteins from the sample. Proteins identification is a requirement to understand their functional diversity. Subtle difference in protein structure and function can contribute to complexity and diversity of life. This review focuses on the progress and the applications of proteomics science with special reference to integration of the evolving technologies involved to address biological questions.

1. Introduction

The completion of human genome sequence has been one of the most important scientific achievements of this century. One of the most surprising elements of the whole research was that humans have barely more genes than fly and/or worm. The human complexity cannot be solely explained by its genomics, but rather in the way these gene products interact, initiating a study on proteins (Grant and Blackstock, 2001). The concept of proteome was first proposed by Wilkins (1995) and refers to the protein equivalent of the genome. Proteomics is a systematic analysis of proteins expressed by a genome. Genome now only represents the first step in the complexity of understanding biological function. While traveling from genome to proteome, craters have been encountered. Gene protein relationship may not be monogamous. A single gene can encode for

* Corresponding author. Tel.: +1 706 542 6455;
fax: +1 706 542 5828.
E-mail address: vdhingra@vet.uga.edu (V. Dhingra).

more than one protein species. The recent compilation of the human genome database has given rise to a number of predicted proteins, which require primary structural and functional identities. Hence, the gene sequence information and pattern of gene activity inside the cell do not provide a complete picture. One has to now look beyond the genomic level to establish a link between the gene and the protein. Measurement of transcript or mRNA levels cannot give complete information on cellular regulations, as gene expression is regulated post-transcriptionally and by the time the myriad of the post-translational modifications are considered, the number of proteins expressed in a cell is many times more than the coding potential of the organism (Cobon et al., 2002). Hence, it is necessary to determine the protein expression levels directly. The proteome of the cell will reflect the immediate environment of the protein. However, protein dynamics does keep changing in accordance to the intracellular and/or extra-cellular stress. One distinct advantage of proteomics is the potential of detecting changes that occur after the messenger RNA step, thus giving quantitative analysis of protein expression profiles. It can also aid in studying the interactions between proteins, hence aiding us to understand the mechanisms of specific processes and pathways inside the cells. The aim would not only be to identify the proteins, but also to create a map representing where the protein is located. These ambitious goals require involvement of different disciplines such as molecular biology, bioinformatics, and biochemistry. Hence, the earlier definition of proteomics (Wilkins, 1995) does not take into account the highly dynamic nature of proteome. The definition of Proteomics should be the protein complement of a given cell at a given time, including the set of all protein isoforms and modifications (de Hoog and Mann, 2004).

In biomedical applications, proteomics promises a new dawn by providing insight into the causes of disease or in identifying an early marker in the initiation of a disease process. Thus, by understanding the pathological mechanisms, novel disease pathways can be known. This, however, is not yet a reality due to the sheer complexity. Proteomics is also specially well-matched to drug discovery because most drugs inhibit function of specific proteins. Many elegant tools have been designed in an approach to study proteins using this approach. The major ones include, high quality separation of proteins in two dimensions, characterization of separated proteins by mass spectrometry, and information mining using Bioinformatics tools (Fig. 1). There is, at present, no substitute to the two-dimensional electrophoresis. However, there are number of areas where this technology can be improved (Celis and Gromov, 1999) i.e., sample preparation and detection methods. The challenge is substantial, and with the knowledge along with the tools available, looks achievable. Proteomics can, thus, revolutionize the health and commercial arenas by providing functional identification to such proteins using its multifaceted approaches.

2. Sample preparation

Sample preparation is one of the most important steps in the whole proteome analysis. Pretreatment of samples for 2D electrophoresis involves solubilization, denaturation and reduction to completely break up the interactions between the proteins (Rabilloud, 1996) and, removal of all interfering compounds to ensure efficient separation. One must ensure that there are no artifacts due to the method of protein preparation. Galvani et al. (2001a, 2001b) and Herbert et al. (2001) have recently showed that initial steps of sample preparation can have profound effect on the final outcome of protein separation and their subsequent analysis.

Several methods have also been suggested to improve the sample preparation to get an unbiased reliable map that gives the accurate representation of all proteins in the sample. Lysis of the cells or tissue may be achieved by flash freezing the cells in liquid nitrogen and homogenizing in mortar with pestle (Dhingra et al., 2000; Davidsson et al., 2001) or in potters homogenizer (Carboni et al., 2002), with either ultrasonic disintegrators or enzymatic lysis (e.g. with lyzozyme), detergents, or repeated freezing and thawing or a combination of these methods. After lysis, it is necessary to remove the interfering substances (phenolic compounds, nucleic acids), and insoluble components (by high speed centrifugation) (Gorg et al., 2000). There can be no general method of sample preparation. The lysis procedure for each kind of tissue or cell needs to be optimized to minimize proteolysis and modification of proteins. The ultimate goal would be to preserve the proteins in the same primary structure found in the
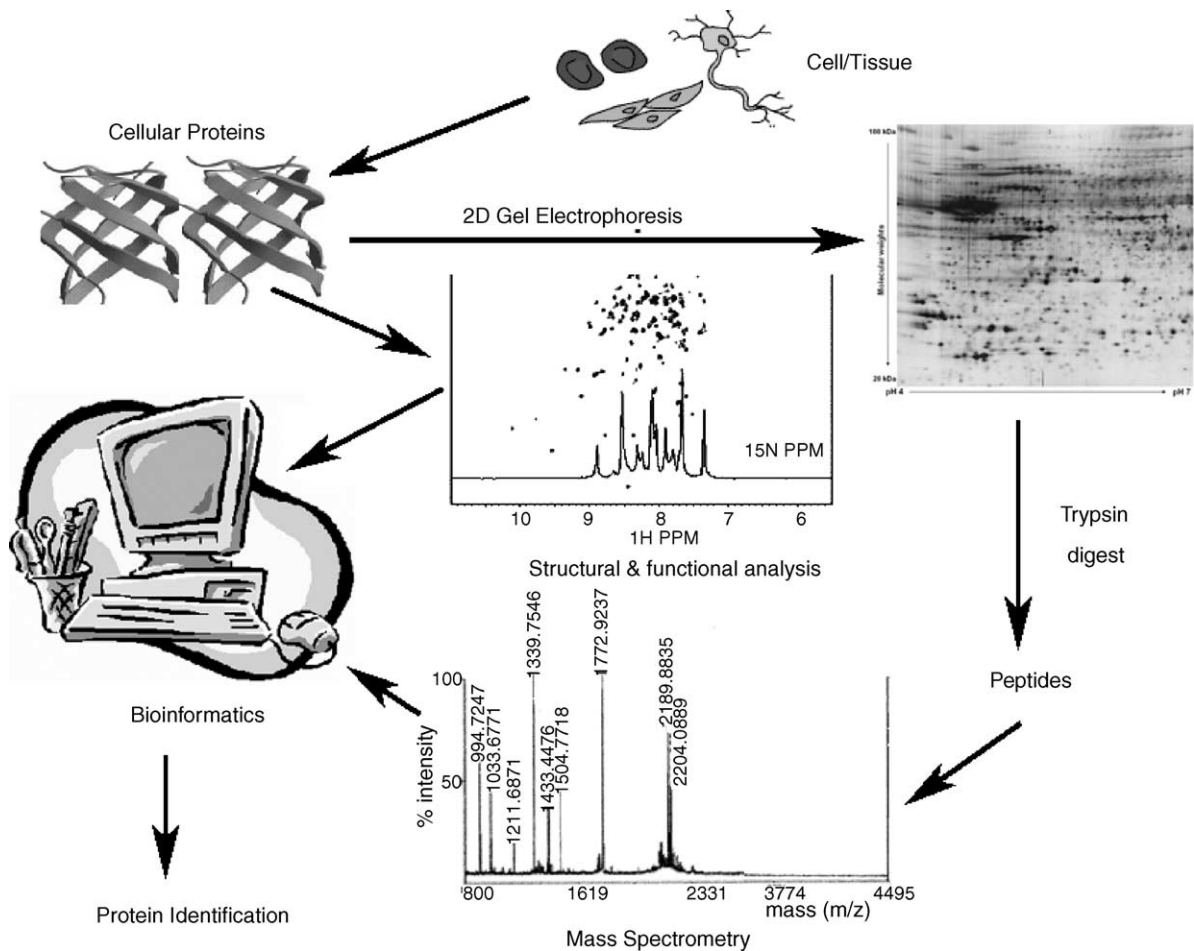
Fig. 1. Outline of a strategy to perform proteomics (an overview).

cell. Proteases present in the sample tend to produce artifacts on the 2D map. These can be silenced by addition of proteases inhibitors. However, these proteases inhibitors can sometime too aid in adding artifactual spots on the two-dimensional profile (Dunn, 1993). Proteases can also be inactivated by boiling the sample in 1% SDS solution. However, SDS has a negative charge and therefore, not compatible with isoelectric focusing. SDS also has a substantial negative effect on alkylation efficiency of the protein (Galvani et al., 2001b). PMSF or protease cocktails can be added to deactivate proteases. However, they need to be added before the addition of reductant and can sometimes lead to charge modifications of some proteins (Dunn, 1993). The proteins can be cleaned by either precipitat-

ing them in ice-cold acetone or TCA or TCA/acetone. After cell lysis and removal of interfering components, the samples should be resolubilized (Damerval et al., 1986). Sample solubility can be improved by using appropriate mixture of Chaotropic agents (Rabilloud et al., 1997), and new efficient detergents (e.g. NP-40, Triton X-100, CHAPS or SDS) (Rabilloud et al., 1990; Chevallet et al., 1998). Luche et al. (2003) evaluated various non-ionic and zwitter-ionic detergents as solubilizers for membrane protein. They found that non-ionic detergents like dodecyl maltoside, decaethylene glycol mono hexadecyl ether, Triton X-100 and detergents with oligooxyethylene sugar or sulfobetaine polar heads were quite efficient in solubilizing membrane proteins. Stasyk et al. (2001) has suggested the

use of 2D clean up kit (Amersham Biosciences, USA) to clean the protein solution. The principle is similar to the method suggested by Damerval et al. (1986). There is an addition of a detergent co-precipitant by which the proteins are better precipitated from the solution. The wash buffer contains some organic additives that allow rapid and complete resuspension of the proteins with solubilizing or lysis buffer. A typical lysis buffer could be a mixture of 9 M Urea, 1% DTT, 2% CHAPS and 0.8% carrier ampholytes. Urea denatures sample proteins, rapidly inactivating any enzyme activity, which could lead to protein modification. 3-(3-Chloramidopropyl)dimethylammonio-1-propanesulphonate (CHAPS) is a zwitter-ionic detergent and one of the most effective solubilizing agent. It increases the solubility of hydrophobic proteins. Thiourea increases the solubility of proteins in presence of urea (Rabilloud et al., 1997; Pasquali et al., 1997). DTT maintains proteins in a reduced state, breaking disulphide bonds and enhancing solubility (Cleland, 1964). 2-Mercaptoethanol is not recommended, as it is required in higher concentration and impurities can result in artifacts (Marshall and Williams, 1984). In more recent times, thiophosphine is also used as it is more powerful than DTT and, also is required in much lower concentrations (3–5 mM) (Herbert et al., 1998). However, its low solubility and poor stability in rehydration buffer can limit its use. IPG buffer (carrier ampholytes) ensures even charge distribution along the pH gradient in the first dimension run. They also tend to improve separation by enhancing protein solubility. Bromophenol blue (an anionic dye) is generally added to control the running conditions. To minimize protein modifications and ensure high resolution, there must be an even charge distribution on the proteins and, the reagent should not contribute significantly to the conductivity of the sample (Herbert et al., 1997).

3. Two-dimensional polyacrylamide gel electrophoresis

A critical requirement of proteomics research is high quality separation of cellular proteins. A path breaking innovation by O'Farrel (1975) led to the development of a classical proteomics tool, two-dimensional gel electrophoresis. The proteins are resolved in two dimensions: the first dimension separates proteins in a pH gradient according to their pI, whilst in the second dimension the proteins are separated according to their molecular weight by SDS-PAGE. Combination of these two separation techniques resolves protein to a spot on the two-dimensional map that is fixed by the coordinates as determined by its isoelectric point and molecular weight (Fig. 2). This map can be considered as the protein fingerprint of the sample. Two such fingerprints derived from two different cellular stages of an organism can be compared to see the up/down regulation or appearance/disappearance of protein spots. This can be extrapolated in determining the protein function in a particular cellular stage of the organism.

The first dimension of electrophoresis involves denaturing isoelectric focusing using 3 mm wide Immobilized pH gradient gels. IPG strips have a gradient of charge imbedded in acrylamide significantly improving the reproducibility and reliability. They also overcome problems of pH gradient instability, discontinuous and restricted pH gradients, and the difficulty of standardizing batches of carrier ampholytes (Bjellqvist et al., 1982; Humphrey-Smith et al., 1997). Also, fewer proteins are lost during equilibration in SDS buffer because the fixed charged groups of the gradient hold the proteins back like a weak ion exchanger (Righetti and Gelfi, 1984). Samples are applied either by cup loading or by in-gel rehydration. In cup loading method, the strips are pre-rehydrated with rehydration buffer and the samples are applied into the loading cup at either acidic or basic end. While in in-gel rehydration, the sample in lysis buffer is diluted with the rehydration buffer. Rehydration of a strip occurs in an individual strip holder or in a reswelling tray. The dry gel matrix takes up the fluid together with the proteins (Westermeier and Naven, 2002). Focusing (IEF) is carried out on a first dimension electrophoresis unit (either Multiphor or IPGPhor from Amersham Biosciences, USA) consisting of five phases of stepped voltage from 500 to 3500 V (Multiphor) or 500 to 8000 V (IPGPhor). Immobilized dry strips come in a variety of lengths (from 7 to 24 cm) and pH ranges (wide gradients for an overview of the entire protein spectrum and narrow gradients for increased resolution).

However, proteome profiles generated by the two-dimensional maps do not always represent the entire proteome. Most proteins seen on the gel tend to be of higher abundance or "housekeeping" proteins. Proteins which are in low abundance (low copy numbers) may
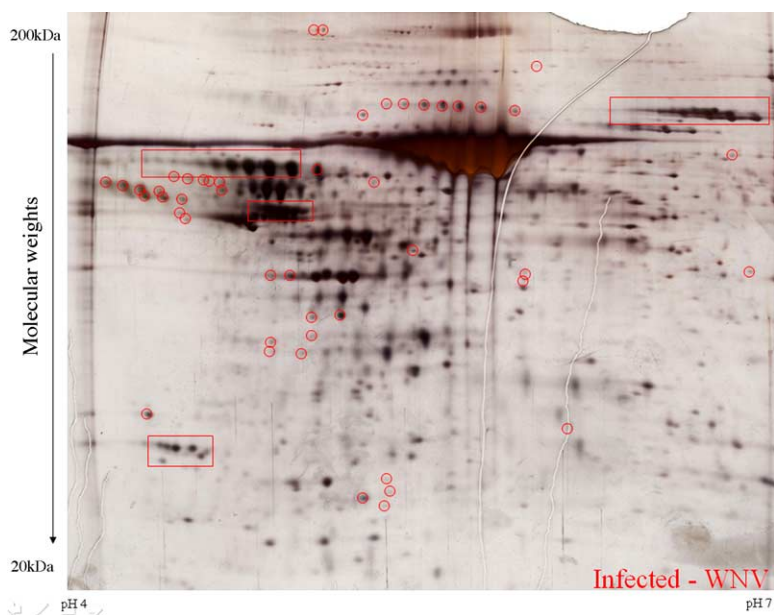
Fig. 2. A 2D map of proteins from mice neurons. The 100 µg of proteins were focused on an IPG strip at a pH range of 4–7 and subsequently run in second dimension. The spots circled represent the protein expression that have been up-regulated due to viral infection. The gel was silver stained on Hofer Processor plus (Dhingra et al., in press).

not be visible on the two-dimensional gel (Gygi et al., 2000). To improve the loading capacity and generate a more representative profile, different strategies are adopted.

Samples are fractionated and/or narrow range pH gradient (single pH range) can be used. More samples (∼500 µg) can be loaded on narrow range pH. This can increase the detection of protein spots by 5–10 folds (Gygi et al., 2000). Conventional separation techniques, viz. ion-exchange chromatography can be incorporated to fractionate the samples prior to 2D electrophoresis (Butt et al., 2001). Recently, Coquet et al. (2004) showed that the use of 24 cm immobiline dry strip gels also improve protein separation. Another approach would be to use differential in-gel electrophoresis (DIGE), a technology developed by Amersham Biosciences, USA. Conventional two-dimensional method relies on comparing profiles from at least two gels. This can sometimes translate to a state highly analogous to frustration, as no two gels can be identical due to differences in gel composition, pH and electric fields, and thermal fluctuations. Hence, no two-gel images are directly super imposable and warping is always required in order to overlay and compare them.

A bad run leading to smearing of spots can always engender false positive. Also, to be sure of the differences observed, numerous gels have to be run and compared.

In differential in-gel electrophoresis (DIGE), one can quantitate the protein differences within the same 2D gel. An equal amount of each protein preparation can be differentially labeled with reactive Cy2, Cy3 and Cy5 dyes. These are identical in molecular weight and isoelectric point, but different in excitation and emission wavelength. These are then mixed and run on same 2D gels (Fig. 3). These proteins are linked to the dyes by lysine labeling, which label the proteins via the terminal amine group of lysine (Unlu et al., 1997). The ratio of dye to protein is kept very low to ensure that the protein visualized on the gel contains a single dye molecule. In differential in-gel electrophoresis, the labeled samples are run on the same 2D gel. These samples are then imaged separately (at different wavelengths). Since they have originated from the same gel, the images can be overlaid and compared directly without warping (Tonge et al., 2001) (Fig. 4). They also do not compromise the mass spectrometry analysis and lead to enhanced recovery of peptides.
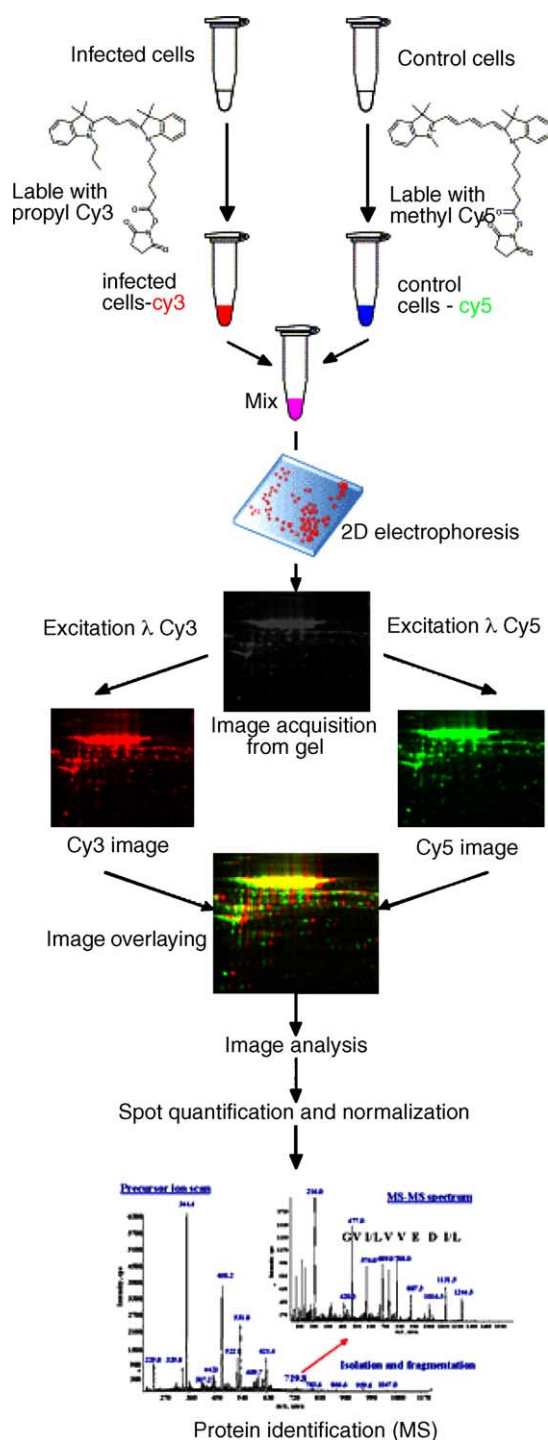
Fig. 3. Principle of 2D-DIGE (Dhingra et al., in press). Protein differences can be quantified within the same 2D gel.

Visualization of proteins following 2D electrophoresis is usually done by staining with either coomassie brilliant blue stain, which will detect proteins present in amount greater than 100 ng (Neuhoff et al., 1988). For greater sensitivity, proteins in the range of 2–5 ng can be detected by silver staining (Shevchenko et al., 1996). However, silver stains can be disparaging to mass spectrometry analysis as many of the fixation methods contain glutaraldehyde, which can cross-link proteins and lead to inefficient digestion by trypsin.

Glutaraldehyde can be omitted from the silver stain formulation to make it more compatible to the mass spectrometer. However, the detection sensitivity is then compromised. Moreover, silver stain, though sensitive, does not have dynamic linearity and has a large protein-to-protein variability making quantification more difficult and less reproducible (Steinberg et al., 1996). Proteins have also been radioactive labeled to get quantitative information using a Phosphorimager (Vercoutter-Edouart et al., 2001). However, metabolic labeling can be operationally hazardous and, can sometimes change the biology of the system under study (Tonge et al., 2001). Non-covalent fluorescent stains such as SYPRO orange, SYPRO ruby (Molecular Probes, OR) have a larger linear range and, are as sensitive as silver stains (Berggren et al., 1999; Steinberg et al., 1996). Western blotting or immuno staining can also be done to stain single protein or a group of proteins against which antibodies are available.

4. Protein identification

The desired end point of any proteomics expression would be to identify and characterize the proteins. One of the most common methods of identifying proteins is through peptide-mass fingerprinting (Henzel et al., 1993; Mann et al., 1993). The proteins are digested with a proteolytic enzyme such as trypsin, to produce a set of tryptic fragments unique to each protein (Fig. 5). The masses of these peptides are then determined by mass spectrometry. It has been possible to determine the molecular structure of proteins that show up as spots on two-dimensional gels using mass spectrometry. This novel technology determines the mass to charge ratio (m/z) of individual molecules in the gas phase by observing their flight in electric and/or
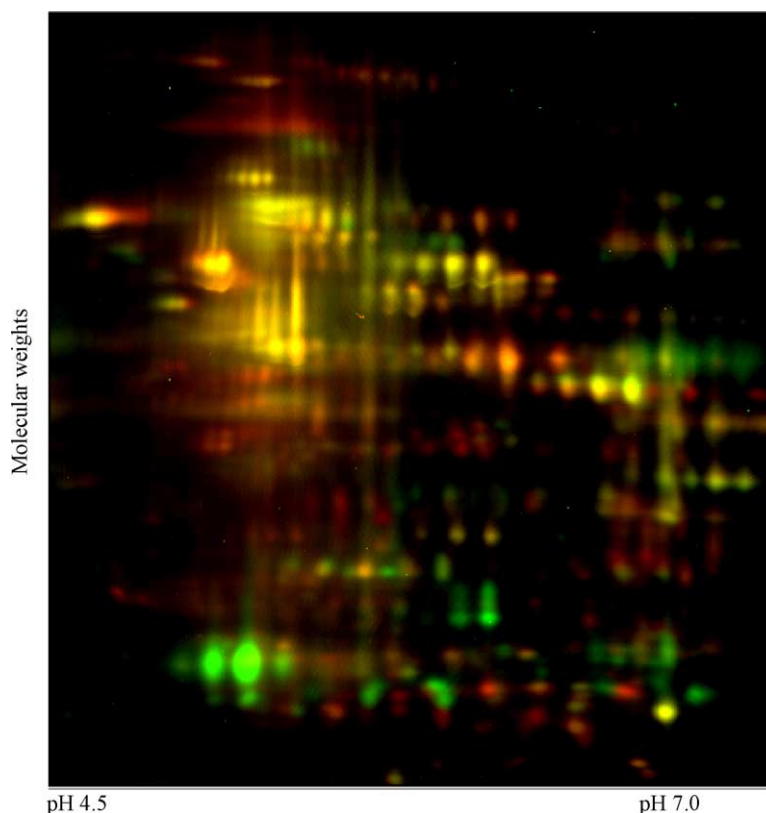
Fig. 4. 2D-DIGE showing protein profiles of *Fundulus heteroclitus* liver extract, between p*I* of 4.5 and 7.0 Green spots represent proteins that are up regulated in Fundulus liver from clean site whereas red spots indicate proteins up regulated in Fundulus liver from creosote-contaminated site (Atlantic Wood, VA). Yellow spots represent proteins common in both extracts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

magnetic fields. Once the ions are formed, they can be separated according to their m/z and detected. The most common system for doing this is matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF). MALDI allows the analysis of high molecular weight compounds with high sensitivity and, soft ionization with little or no fragmentation is observed. It is also tolerant to salts in millimolar concentration. However, there is always a possibility of photo degradation by laser desorption/ionization and Matrix masks m/z below 500. The resolution is low, and MALDI is quite intolerant to detergents. MALDI uses a solid matrix and a laser light as its ionizing beam. The matrix is typically a small energy absorbing molecule such as α-cyano-4-hydroxycinnamic acid or 2,5,-dihydroxybenzoic acid. The non-volatile matrix plays an important role by absorbing the laser radi-

ation resulting in vaporization of matrix and sample embedded in matrix. Once in the gas phase, the analyte molecules are ionized, these are then directed towards the mass-analyzer, calculating the time of flight. The molecular weight values of the trypsinized peptides obtained by MALDI-TOF are then used to identify the predicated proteins using Web-based search engines such as MASCOT and PROFOUND (Perkins et al., 1999; Zhang and Chait, 2000) (Fig. 6). The mass spectrometry technology can also be utilized to identify differences in the post-translational modification of various proteins. It is often clear from the gel profiles if there are significant changes in glycosylation or phosphorylation of proteins in the extract (Cobon et al., 2002).

A minority of the proteins may be recalcitrant to MALDI-TOF analysis. Moreover, the information
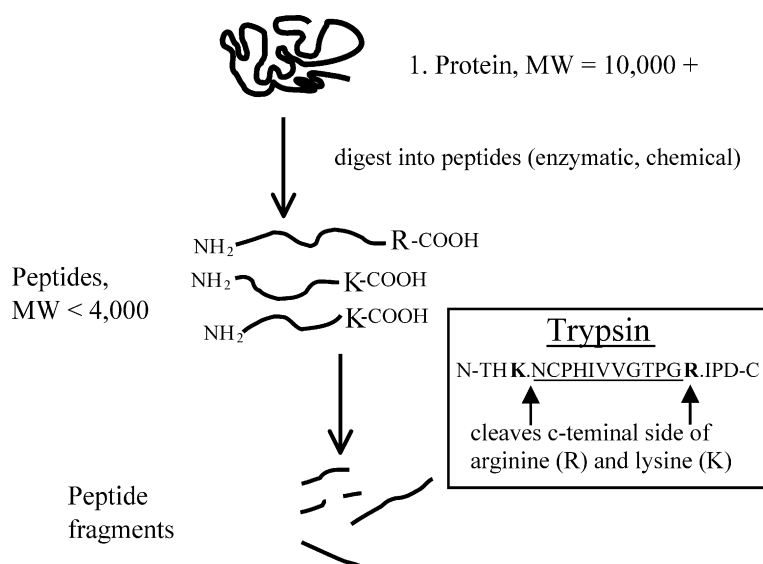
Fig. 5. Trypsinization of proteins to peptide fragments (from Gygi et al., 2000). The proteins are digested with a proteolytic enzyme such as trypsin, to produce a set of trypic fragments unique to each protein.

obtained from MALDI-TOF is simply a function of signal intensity and hence, non-uniform nature of matrix and/or variation in detector response can lead to non-reproducible signal response. Also, a given peptide map may not have sufficient information to identify the protein or the protein may not be present in the database. In such cases, the proteins may be analyzed by nanoelectrospray tandem mass spectrometry (Shevchenko et al., 1996) or quadruple mass spectrometer (Q-TOF) (Morris et al., 1997), which provides MS/MS spectra that can be used to deduce the order of the amino acids in the tryptic peptides. The types of fragment ions observed in an MS/MS spectrum

depend on many factors including primary sequence, the amount of internal energy, how the energy was introduced, charge state, etc. The accepted nomenclature for fragment ions was first proposed by Roepstorff and Fohlman (1984), and subsequently modified by Johnson et al. (1987). Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either a, b or c. If the charge is retained on the C terminal, the ion type is either x, y or z. A subscript indicates the number of residues in the fragment (Table 1). The difference in the mass between adjacent y- or b-ions corresponds to that of an amino acid. This can be
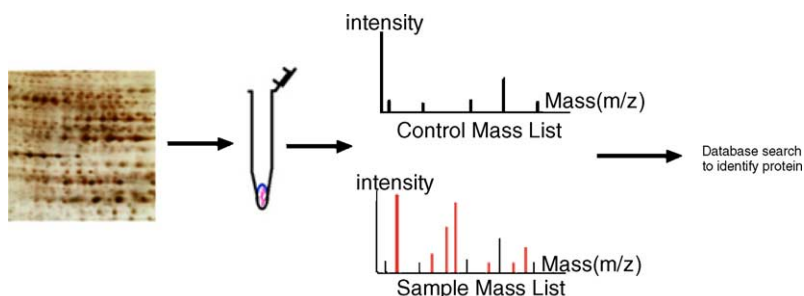

Fig. 6. A typical mass spectrometry experiment. The proteins are in-gel digested and applied to MALDI-TOF to calculate the time-of-flight of ions that are formed when the ionizing beam of laser hits the peptides. The molecular weight of the trypsinized peptides are then used to identify predicted proteins using web-based search engines.

Table 1
Formulae to calculate fragment ion masses

Ion type	Ion mass
a	$[N] + [M]$-CO
a*	a-NH_3
a°	a-H_2O
A^{++}	(a + H)/2
B	$[N] + [M]$
B*	b-NH_3
B°	b-H_2O
B^{++}	(b + H)/2
C	$[N] + [M] + NH_3$
D	a: partial side chain
V	y: complete side chain
W	z: partial side chain
X	$[C] + [M] + CO$
Y	$[C] + [M] + H_2$
Y*	y-NH_3
Y°	y-H_2O
Y^{++}	(y + H)/2
z	$[C] + [M]$-NH

$[N]$ is the mass of N-terminal group, $[C]$ is the mass of C-terminal group, $[M]$ is mass of the sum of the neutral amino acid residue masses.

used to identify the amino acid and, hence the peptide sequence.

5. Bioinformatics in proteome analysis

Genes and proteins interact structurally, evolutionarily, functionally, metabolically, resulting in huge flow of information due to biomolecular interactions. Bioinformatics, thus, serves as a bridge between observations (data) in diverse biologically related disciplines and the derivations of understanding (information) about how the systems or processes function, and subsequently the application (knowledge). It may be defined as conceptualizing biology in terms of macromolecules and then applying "informatics" techniques (derived from disciplines such as applied mathematics, computer science, and statistics) to understand and organize the information associated with these molecules on a large-scale (Luscombe et al., 2001). Bioinformatics has become an integral part of proteomics. Proteomics has become more dependent on bioinformatics for storing proteomic data in databases as well as for protein extraction and interpretation (de Hoog and Mann, 2004). Develop-

ment of sophisticated software for an efficient analysis of data that is acquired from running multiple 2D gels and mining from mass spectrometry is needed to meet challenges put up by genomics research. A friendly interface, interactive data acquisition and networking are few of the important aspects that need to be addressed to define a subset of variables from a number of possibilities that may be generated from processing a two-dimensional gel image. Over the past few years, the field has renewed interest due to the boom in software and its orientation towards the genomic industry. The decrease in cost of computing and data storage has had its own role to play in defining the role of software in biological studies (Patterson, 2003).

Databases are at the core of bioinformatics, which involve building a satisfactory system that is able to capture, store, retrieve and analyze all data concerned. The principle idea of these databases is to enable the scientific user to get a quick idea about the current knowledge that has been gathered about a particular subject. Identifying proteins by mass requires access to protein sequence database. The most commonly used databases are SWISS-PROT, TrEMBL and non-redundant collection of protein sequences at the US National Centre for Biotechnology Information (NCBI). Swiss-Prot (Bairoch and Apweiler, 2000) is a collection of curated protein sequence database, which provides a high level of annotations, a minimal level of redundancy and high level of integration with other databases. Each entry corresponds to a single contiguous sequence as contributed to the bank or reported in the literature. TrEMBL is a computer annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot. This helps in making data available to users as quickly as possible after publication. The NCBI database contains translated protein sequences kept at GenBank. GenBank is the largest of the biological public database, which collects all known nucleotide and protein sequences with supporting bibliographic and biological annotation (Benson et al., 2000). It is built by the National Center for Biotechnology Information (NCBI) at NIH, along with its two partners, the DNA database of Japan (DDBJ) (Tateno et al., 2000) and the European Molecular Biology Laboratory (EMBL) (Baker et al., 2000), a nucleotide database from the European Bioinformatics Institute. GenBank depends on its contributors

to help keep the database as comprehensive, current, and accurate as possible. Recently, RESID Database of Protein Modifications has been developed. It is a comprehensive collection of annotations and structures for protein modifications and cross-links including pre-, co-, and post-translational modifications. The database provides: systematic and alternate names, atomic formulas and masses, enzymatic activities that generate the modifications, keywords, literature citations, gene ontology (GO) cross-references, protein sequence database feature table annotations, structure diagrams, and molecular models. This database is freely accessible on the internet through resources provided by the European Bioinformatics Institute (http://www.ebi.ac.uk/RESID), and by the National Cancer Institute, Frederick Advanced Biomedical Computing Center (http://www.ncifcrf.gov/RESID). Each RESID database entry presents a chemically unique modification and shows how that modification is currently annotated in the protein sequence databases, Swiss-Prot and the Protein Information Resource (PIR). The RESID database provides a table of corresponding equivalent feature annotations that is used in the UniProt project, an international effort to combine the resources of the Swiss-Prot, TrEMBL and PIR (Garavelli, 2004). Several such proteomic databases exist today and are listed in Table 2. The public availability of such libraries is likely to result in a dramatic progress in the ability to correlate the expression of specific protein structures with biological functions. Three-dimensional structure databases can be utilized for finding novel lead-structures of medicaments, pesticides, and other biologically active compounds. The Protein Data Bank is the sole archive of the three-dimensional structures of biological macromolecules and, its entries have recently increased dramatically. Another three-dimensional structure database (3DPSD) developed by the Akira Dobashi's group of Tokyo University of Pharmacy and Life Sciences provides a fully optimized structures, molecular motion trajectories, conformational distribution and electrostatic potential mapped on the electron density surfaces (http://www.ps.toyaku.ac.jp/dobashi/). These databases provide fundamental resource for drug development. The molecular modeling database (MMBD) available on NCBI contains 3D macromolecular structures, including proteins and polynucleotide. MMDB contains over 28,000 structures and is linked to the rest of the NCBI databases, including sequences, bibliographic citations, taxonomic classifications, and sequence and structure neighbors.

Data analysis is also an Achilles heel of proteomics. Proteomics is a substrate limited science and, proteins tend to exist over a wide range of concentration in a biological sample. This can generate a huge amount of data and keeping pace with the ability to analyze the voluminous data can be difficult. Proteomics relies heavily on protein sequence database. When protein sequence is available, tools such as BLAST (Altschul et al., 1990), FASTA (Pearson and Lipman, 1988) or TagIdent (Wilkins et al., 1998) can be used. The two main technologies used in proteomics, two-dimensional electrophoresis (quantification through image analysis) and mass spectrometry (quantification of mixture of proteolytic peptides) can pose a challenge to data analysis. The two-dimensional software continues to improve, but still need manual intervention.

Computer programs for the analysis of 2D gels were first described 20 years ago (Anderson et al., 1981). Most of the packages were developed in academic institutions and later commercialized (e.g. Elsie, Gellab, Melanie, Quest, etc.) (Weissig and Bourne, 1999; Rosengren et al., 2003). The software available for identification of peptides selects mass spectrometric generated fragment ions automatically and engenders large amount of data for analysis. This sometimes gives a problem when the majority of spectra are due to instrument noise or minor contaminants. This data analysis can create huge loss of computing time. The development of new algorithms has solved problems in few cases, but these produce scores that give only significant matches. This can lead to false positives or false negatives; unless a manual check has been made (Patterson, 2003). A major challenge to protein bioinformatics, however, is the integration between the huge diversity of data generated and, the database and the tools currently available.

6. Post-translational modifications

After transcription from DNA to RNA, the gene transcript can be spliced in different ways prior to translation into protein. Following translation, most proteins are chemically changed through post-translational modifications, mainly through the

Table 2
Protein and protein–protein interaction databases and tools

Swiss-Prot/TrEMBL	http://www.expasy.org	Analysis of protein sequences and structures as well as 2D PAGE
PROSITE	http://au.expasy.org/prosite/	Protein families and domains
SWISS-2D PAGE	http://au.expasy.org/ch2d/	Data on proteins identified on various 2D PAGE and SDS-PAGE reference maps
ENZYME	http://au.expasy.org/enzyme/	Repository of information relative to the nomenclature of enzymes
SWISS-3DIMAGE	http://au.expasy.org/sw3d/	Image database which strives to provide high quality pictures of biological macromolecules with known 3D structure
SWISS-MODEL Repository	http://swissmodel.expasy.org/repository/	Contains annotated 3D comparative protein structure models generated by the fully automated homology-modeling pipeline SWISS-MODEL
GermOnLine	http://www.germonline.unibas.ch/	Cross-species community annotation knowledgebase that provides microarray data relevant for the mitotic and meiotic cell cycle as well as gametogenesis
ChloroP	http://www.cbs.dtu.dk/services/ChloroP/	Prediction of chloroplast transit peptides
LipoP	http://www.cbs.dtu.dk/services/LipoP/	Prediction of lipoproteins and signal peptides in Gram negative bacteria
AACompIdent	http://au.expasy.org/tools/aacomp/	Identify a protein by its amino acid composition
MultiIdent	http://au.expasy.org/tools/multiident/	Identify proteins with pI, M_w, amino acid composition, sequence tag and peptide mass fingerprinting data
Unigene	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene	Experimental system for automatically partitioning GenBank sequences into a non-redundant set of gene-oriented clusters
Ensembl	http://www.ebi.ac.uk/ensembl/	Joint project between the EMBL-EBI and the Wellcome Trust Sanger Institute that aims at developing a system that maintains automatic annotation of large eukaryotic genomes
RefSeq	http://www.ncbi.nlm.nih.gov/RefSeq/key.html	The Reference Sequence (RefSeq) collection aims to provide a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products, for major research organisms
IPI	http://www.ebi.ac.uk/IPI/IPIhelp.html	Maintains database that describe the proteomes of higher eukaryotic organisms
BIND	http://bind.ca/	Collection of records documenting molecular interactions
BRITE	http://www.genome.ad.jp/brite	Database of binary relations for network computation and logical reasoning involving genes, proteins, and other biological molecules
Cellzome	http://yeast.cellzome.com	Yeast protein complex database
DIP	http://dip.doe-mbi.ucla.edu	Database catalogs experimentally determined interactions between proteins
GRID	http://biodata.mshri.on.ca/grid/index.html	Database of genetic and physical interactions developed in The Tyers Group at the Samuel Lunenfeld Research Institute at Mount Sinai Hospital
HPRD	http://www.hprd.org	Centralized platform to visually depict and integrate information pertaining to domain architecture, post-translational modifications, interaction networks and disease association for each protein in the human proteome
InterPreTS	http://www.russell.embl.de/interprets	Prediction of protein interaction though protein structure
MIPS	http://mips.gsf.de/genre/proj/yeast/index.jsp	The MIPS Comprehensive Yeast Genome Database (CYGD) contains information on the molecular structure and functional network of the entirely sequenced, well-studied model eukaryote, the budding yeast *S. cerevisiae*
PIM-Hybrigenics	http://www.hybrigenics.fr	Database dedicated to the exploration of protein pathways

Table 2 (*Continued*)

Riken-PPI	http://fantom21.gsc.riken.go.jp/PPI/	In searching interactions of proteins of interest
PPID	http://www.ppid.org	Unifies molecular entries across three species, namely human, rat and mouse and is footed on sequence databases such as SwissProt, EMBL, TrEMBL (translated EMBL sequences) and Unigene and the literature database PubMed
Wormbase	http://www.wormbase.org	Comprehensive data resource for Caenorhabditis biology
Flybase	http://www.flybase.org	Comprehensive data resource for Drosophila biology
MGI	http://www.informatics.jax.org	Integrated database on the genetics, genomics, and biology of the laboratory mouse
Prodom	http://protein.toulouse.inra.fr/prodom.html	Comprehensive set of protein domain families automatically generated from the SWISS-PROT and TrEMBL sequence databases
SWALL	http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+LibInfo+-id+1THFp1PWcEH+-lib+SWALL	Comprehensive protein sequence database that combines the high quality of annotation in Swiss-Prot with the completeness of the weekly updated translation of all protein coding sequences from the EMBL nucleotide sequence database
SMART	http://smart.embl-heidelberg.de	Contains information on Swiss-Prot, SP-TrEMBL and stable Ensembl proteomes

addition of carbohydrate and phosphate groups. Such modifications play a vital role in modulating the function of many proteins. Analysis of these modifications presents the most challenging aspect in proteomic research. These modifications can affect protein turnover, localization, protein–protein interactions and activity. The most common post-translational modifications include glycosylation, phosphorylation, ubiquitination, methylation, acetylation, and lipidation. Ubiquitination and sumoylation modification targets the protein for degradation, while farnesylation tethers proteins to membranes (de Hoog and Mann, 2004). These modifications have been shown to play an important role in development, physiology and diseases of animals and plants. Recently, Denison et al. (2005) used a proteomics strategy to gain insight into the complex process of sumoylation. SUMO-conjugated proteins were isolated by a double-affinity purification procedure from a *Saccharomyces cerevisiae* strain engineered to express tagged SUMO. The sumoylated proteins were identified by subsequent (LC–MS/MS) analysis using an LTQ FT mass spectrometer. Methylation of proteins plays an important role in differentiation of progenitor cells to mature and functional cells. Protein methylation is mediated by protein methyl transferases. Post-translational modifications is mediated by specific enzymes that catalyze binding of the groups to proteins (e.g. kinases), or removal of the groups from the proteins (e.g.

phosphatases). These protein modifications are highly specific and tightly regulated. Analysis and comparison of subproteomes (e.g. the phosphoproteome) of biological material (cells, tissues) under different conditions (stimulated versus control cells, tissues at different stages during development, cancer cells versus control cells) provides new insights into cell growth and differentiation, development and cancer.

Phosphorylation of proteins is one of the most studied Post-translational modifications and, it has been shown to be essential for the function of numerous proteins. The most common sites of phosphorylation in proteins are tyrosine, serine and threonine. These modifications resulting in change in the molecular mass of the affected protein are characterized by MALDI-MS and MS/MS (Mann and Jensen, 2003). Kjellstrom and Jensen (2003) suggested the use of in situ liquid–liquid extraction before MALDI-MS and MS/MS for the separation of hydrophobic and hydrophilic peptides. This would tremendously aid in analyzing peptide mixtures containing phosphorylated, glycosylated, or acylated peptides. Ibarrola et al. (2003) have used stable isotope labeled amino acid incorporation in mammalian cell culture, which has been used as a tool for relative quantitation of phosphopeptides by mass spectrometry. Western blot, Gel shift assays, Selenocysteine insertion, HPLC techniques are the other common detection methods used on characterizing these modifications (Nakamura and Goto, 1996; Ying et

al., 2004; Kryukov et al., 2003; Dean et al., 1997). Affinity surface enhanced laser desorption ionization (SELDI) has also been used to validate proteins to monitor the disease-induced post-translational modification and the ternary status of myocyte-originating protein, cardiac troponin I in serum (Stanley et al., 2004). This method relies on surface enhanced laser desorption ionization time-of-flight mass spectrometry and, uses a gold coated chip with 8 or 16, 2 mm spots that are modified with chromatographic surfaces to allow selective adsorption of polypeptides directly from sample of interest. Different phosphorylated forms of a protein can be visualized by two-dimensional electrophoresis by the 'pearls-on-string' pattern (Bykova et al., 2003). However, identification of post-translationally modified proteins can be quite challenging, since the fraction of peptide bearing a particular modification can be a small fraction of the total amount of peptide present in the sample. Also, lack of specific antibodies tends to exploit the limitation to a problem (Gronborg et al., 2002). Analysis is also intricate due to low abundance and speedy removal of modified proteins. Strategies are now being developed to characterize individual modified proteins and to map the sites to their molecular detail. Several methods have also been developed that chemically tag phosphopeptides and make it possible to isolate them from complex mixtures (McLachlin and Chait, 2001). One method identifies novel phosphoproteins involved in intracellular signaling following immunoprecipitation and blotting of phosphoserine and phosphothreonine proteins using specific antibodies (Gronborg et al., 2002). Oda et al. (2001) and Goshe et al. (2001) used β-elimination chemistry to cause loss of H_3PO_4 from phosphoserine and derivatization with ethanedithiol. The phosphopeptides were then labeled with biotin and pulled down by affinity chromatography. However, with this approach the phosphotyrosine-containing tryptic peptides cannot be enriched. An alternate approach would be to use immobilized metal affinity column based phosphopeptide enrichment approach as described by Ficarro et al. (2002). In contrast to Ficarro's study, Zhou et al. (2001) found predominately single phosphorylated peptides in their analysis. They modified the phosphopeptide using carbodiimide condensation. Recent development of fluorescent dyes may also be combined with two-dimensional electrophoresis to label phosphoproteins (Steinberg et al., 2003). The

phosphoprotein can then be identified by MALDI-MS. Edman or MS/MS sequencing can then be used to confirm the identification of phosphorylated peptide. Yamagata et al. (2002) identified 5% of proteins visible on a 2-DE gel in rat skin fibroblast culture as phosphorylated. Kimura et al. (2003) investigated the post-translational modification of 26S proteasome and identified the N-acetylated subunits using a proteomics approach. Soskic et al. (1999) also successfully identified phosphorylated proteins using antiphosphoserine and anti-phosphotyrosine antibodies following two-dimensional electrophoresis. Although these studies involved cell lines, they highlight possible directions for proteomic research in two-dimensional protein mapping and blotting (Aldreda et al., 2004). An alternative approach utilizing phospho-affinity step to isolate intact phosphoproteins can also be considered. The phosphoproteins can then be subsequently characterized by electrophoresis and, identified by direct de novo sequencing using tandem mass spectrometry. Metodiev et al. (2004) applied this technique to probe signal induced changes in the phosphoproteome of human U937 cells and found that the pools of two cancer-related phosphoproteins implicated in intracellular hormones signaling are dramatically altered in the course of monocyte to macrophage differentiation.

Glycosylation is an important post-translational modification, which involves different carbohydrate groups. Enormous heterogeneity is found among these modifications due to different lengths of sugar chains that are involved. The composition of the glycans is crucial for the function of many proteins in cell signaling and host–pathogen interactions. Glycosylation increases the complexity of protein molecules and causes them to migrate as diffuse bands or spots on SDS-PAGE gels to complicate efforts to identify protein expression patterns that correlate with disease states (Zaia, 2004). Recently, Halligan et al. (2004) have developed a web-based tool for mapping protein modifications on two-dimensional gels (ProMoST). It calculates the effect of single or multiple post-translational modifications (PTMs) on protein isoelectric point (p*I*) and molecular weight, and displays the calculated patterns, as two-dimensional (2D) gel images. Thus, proteomics, combined with separation technology and mass spectrometry, makes it possible to dissect and characterize the individual parts of post-translational modifications and provide a systemic

analysis. However, detection of post-translational modification does not provide enough information for estimating the function of protein. To our knowledge, there is no database and software, which can predict the function from the information of post-translational modification that is detected by mass spectrometry.

7. Structural proteomics

3D structures of proteins are a key element in understanding the biological processes and play a vital role in the discovery of new lead drugs. Structural proteomics now plays a central role in various biomedical and pharmaceutical research. It attempts to identify all the proteins within a protein complex, characterize their interactions and provide information on protein signaling, disease mechanisms or protein–drug interactions. Since the 3D structure of a protein is more conserved than sequence, these initiatives also open up the possibility of biochemical or biophysical functional characterization via structure. Understanding protein function at molecular level and its mechanism at genetic level necessitates elucidation of its detailed interaction with its genetic counterpart. Unique contributions have been made on the mechanisms for control of genetic information, packaging, repair and other DNA–protein interactions both by X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR) (Jamin and Toma, 2001). X-ray crystallography has been used effectively in generating the first structure-based drugs, derived from structure–function studies of neuraminidase inhibitors such as zanamivir, and the HIV-protease inhibitors amprenavir and nelfinavir (Blundel et al., 2002). Although X-ray crystallography has the advantage of defining ligand-binding sites with more certainty, the ability of NMR to measure proteins in their native state is an important distinction. Furthermore, NMR is increasingly being recognized as a valuable tool, not only in 3D structure determination, but in many more upstream parts of the drug discovery process (Renfrey and Featherstone, 2002).

The availability of a 3D structure has been used to add supporting evidence to a functional assignment made on the basis of sequence similarity. TM0423 from *Thermotoga maritima* was annotated as a glycerol dehydrogenase and subsequent structure determination showed a Tris buffer molecule (which appeared to mimic a glycerol substrate) bound in the active site, confirming the earlier annotation (Lesley et al., 2002). NMR can also be used as a tool for determination of structure and dynamics of protein and protein–ligand complex. It can identify and characterize interactions between protein–protein and/or protein–DNA. One of the major advantage of using NMR spectroscopy in structural proteomic study is that it needs little or no samples preparation, is rapid and non-destructive and uses small sample sizes (Nicholson and Wilson, 1989). The recent sequencing of human genome suggests that there are about 600–3000 genes that can code for potential drug targets for human disease (Hopkins and Groom, 2002). With human proteome being much larger and diverse that genome, there can be many more potential drug targets (Kubinyi, 2003). Both X-ray crystallography and NMR can aid in determining the structure of the potential drug target (Clore and Gronenborn, 1998; Staunton et al., 2003). Determination of the 3D protein structure provides an insight to its binding site and biological function (Minn et al., 1997). This technology was first used in characterizing the binding of penicillin to serum albumin in a nuclear magnetic relaxation study (Fischer and Jardetzky, 1965). Recently, Peti et al. (2004) used NMR for structural proteomics of 141 small recombinant proteins of *T. maritima* proteins for globular folding. NMR structure determination is currently limited by size constraints and lengthy data collection and analysis time. However, in spite of the harmonizing limitations, NMR spectroscopy can play a momentous role in structural proteomics. Yee et al. outlined a strategy for the use of NMR spectroscopy for structural proteomics of small proteins based on data from 513 proteins from five microorganisms. These microorganisms includes both thermophilic and mesophilic species and, representatives from the prokaryotes, archaea, and eukaryotes (Yee et al., 2002). Chemometric analysis of biological NMR spectra is currently being given high-priority in the pharmaceutical industry with respect to development of efficient high throughput toxicity screening systems for lead drug candidate selection. Advances in bioinformatics will also have an impact on structural proteomics in the drug development chain. Structural proteomics is the first step to move away from traditional hypothesis driven research and to better understand the relationship between protein sequence, structure and function.

8. Conclusions

Proteomics is in its exponential growth phase. It provides an excellent tool to study variations in protein expression between different states and conditions. The fact remains that the gene sequence and protein function cannot be correlated. However, sequencing of human and other important genomes has opened the door for proteomics to provide a skeleton to protein mining and to develop a catalogue of proteins, which can be beneficial in taking science to the next generation. It has changed the whole approach to a biological problem. Proteome is the next natural step after genome. Extensive progress has been made on interaction studies using experimental technologies such as mass spectrometry, nuclear magnetic resonance, and computational biology. Post-translational modifications generate tremendous diversity, complexity and heterogeneity of gene products and, their determination is one of the main challenges in proteomics research. Combinations of affinity-based enrichment and extraction methods, multi-dimensional separation technologies and mass spectrometry are particularly attractive for systematic investigation of post-translationally modified proteins.

It is apparent that the current 2D technology has its limitations for proteome analysis. However, a technology superior to 2D electrophoresis for global profiling is yet to emerge. Taking into account the factors such as cost, availability and ease of use, we believe that in present times, 2D electrophoresis is one of the most apposite approaches towards the methodical characterization of proteomes. Moreover, cell proteomes are complex and would need both 2D based and non-2D based technologies to help decipher protein function. This will remain a tremendous and exciting challenge in proteomics research for years to come. Finally, the interpretation of data generated would give meaningful conclusions with successful integration of computer routines. Structural proteomics along with the high-throughput chemistry and screening forms an integrated platform to investigate the mechanisms that underpin the modern drug discovery process. The eventual goal of proteomics is to typify the information tide through protein networks. This information can be a cause, or a corollary, of disease processes. Together these would provide us with a complete picture to improve our understanding of health and disease.

References

Aldreda, S., Melissa, M., Grant, M.M., Griffiths, H.R., 2004. The use of proteomics for the assessment of clinical samples in research. Clin. Biochem. 37, 943–952.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anderson, N.L., Taylor, J., Scandora, A.E., Coulter, B.P., Anderson, N.G., 1981. The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. Clin. Chem. 27, 1807–1820.

Bairoch, A., Apweiler, R., 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL. Nucl. Acids Res. 28, 45–48.

Baker, W., vanden Broek, A., Camon, E., Hingamp, P., Sterk, P., Stoesser, G., Tuli, M.A., 2000. The EMBL nucleotide sequence database, Nucl. Acids Res. 28, 19–23.

Benson, D.A., Karsch-Mizrachi, I.K., Lipman, K.D., Ostell, J., Rapp, B.A., Wheeler, D.L., 2000. GenBank. Nucl. Acids Res. 28, 15–18.

Berggren, K., Steinberg, T.H., Lauber, W.M., Carroll, J.A., Lopez, M.F., Chernokalskaya, E., Zieske, L., Diwu, Z., Haugland, R.P., Patton, W.F., 1999. A luminescent ruthenium complex for ultra sensitive detection of proteins immobilized on membrane supports. Anal. Biochem. 276, 129–143.

Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Gorg, A., Westermeier, R., Postel, W., 1982. Isoelectric focusing in immobilized pH gradients: principle, methodology, and some applications. J. Biochem. Biophys. Meth. 6, 317–339.

Blundel, T.L., Jhothi, H., Abell, C., 2002. High-throughput crystallography for lead discovery in drug design. Nat. Rev. Drug Discov. 1, 45–54.

Butt, A., Davison, M.D., Young, J.A., Gaskell, S.J., Oliver, S.G., Benyon, R.J., 2001. Chromatographic sepration as a prelude to two-dimensional electrophoresis in proteomics analysis. Proteomics 1, 42–53.

Bykova, N.V., Stensballe, A., Egsgaard, H., Jensen, O.N., Moller, I.M., 2003. Phosphorylation of formate dehydrogenase in potato tuber mitochondria. J. Biol. Chem. 278, 26021–26030.

Carboni, L., Piubelli, C., Righetti, P.G., Jansson, B., Domenici, E., 2002. Proteomic analysis of rat brain tissue: comparison of protocols for two-dimensional gel electrophoresis analysis based on different solubilizing agents. Electrophoresis 23, 4132–4141.

Celis, J.E., Gromov, P., 1999. 2D electrophoresis: can it be perfected? Curr. Opin. Biotechnol. 10, 16–21.

Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kiefer, S., Rossignol, M., Lunardi, J., Garin, J., Rabilloud, T., 1998. Electrophoresis 19, 1901–1909.

Cleland, W.W., 1964. Dithiothreitol, a new protective reagent for SH groups. Biochemistry 3, 480–482.

Clore, G.M., Gronenborn, A.M., 1998. NMR structure determination of proteins and protein complexes larger than 20 kDa. Curr. Opin. Chem. Biol. 2, 564–570.

Cobon, G.S., Verrills, N., Papakostopoulos, P., Eastwood, H., Linnane, A.W., 2002. The proteomics of ageing. Biogerontology 3, 133–136.

Coquet, L., Cosette, P., Jouenne, T., 2004. The use of 24 cm, pH 3–11 NL immobiline drystrip gels improves protein separation. Life Sci. News 18, 12–13.

Damerval, C., DeVienne, D., Zivy, M., Thiellement, H., 1986. Electrophoresis 7, 53–54.

Davidsson, P., Paulson, L., Hesse, C., Blennow, K., Nilsson, C.L., 2001. Proteome studies of human cerebrospinal fluid and brain tissue using a preparative two-dimensional electrophoresis approach prior to mass spectrometry. Proteomics 1, 444–452.

Dean, R.T., Fu, S., Stocker, R., Davies, M.J., 1997. Biochemistry and pathology of radical-mediated protein oxidation. J. Biochem. 324, 1–18.

de Hoog, C.L., Mann, M., 2004. Proteomics. Annu. Rev. Genomics Hum. Genet. 5, 267–293.

Denison, C., Rudner, A.D., Gerber, S.A., Bakalarski, C.E., Moazed, D., Gygi, S.P., 2005. A proteomic strategy for gaining insights into protein sumoylation in yeast. Mol. Cell Proteomics 4, 246–254.

Dhingra, V., Chakrapani, R., Narasu, M.L., 2000. Partial purification of proteins involved in the bioconversion of Arteannuin B to Artemisinin. Bioresour. Technol. 73, 279–282.

Dhingra, V., Li, Q., Allison, A.B., Stallknecht, D.E., Fu, Z.F., in press. Proteomic profiling and neurodegeneration in West Nile virus infected neurons. J. Biomed. Biotechnol.

Dunn, M.J., 1993. Gel Electrophoresis of Proteins. BIOS Scientific Publishers Ltd./Alden Press, Oxford.

Ficarro, S., McCleland, M., Stukenberg, P., Burke, D., Ross, M., Shabanowitz, J., Hunt, D., White, F., 2002. Phosphoproteaome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. Nat. Biotechnol. 20, 301–305.

Fischer, J.J., Jardetzky, O., 1965. Nuclear magnetic relaxation study of intermolecular complexes. The mechanism of penicillin binding to serum albumin. J. Am. Chem. Soc. 87, 3237–3244.

Galvani, M., Rovatti, L., Hamdan, M., Herbert, B., Righetti, P.G., 2001a. Protein alkylation in the presence/absence of thiourea in proteome analysis: a matrix assisted laser desorption/ionization-time of flight mass spectrometry investigation. Electrophoresis 22, 2066–2074.

Galvani, M., Hamdan, M., Herbert, B., Righetti, P.G., 2001b. Alkylation kinetics of proteins in preparation of two-dimensional maps: a matrix assisted laser desorption/ionization mass spectrometry investigation. Electrophoresis 22, 2058–2065.

Garavelli, J.S., 2004. The RESID database of protein modifications as a resource and annotation tool. Proteomics 4, 1527–1533.

Goshe, M.B., Conrads, T.P., Panisko, E.A., Angell, N.H., Veenstra, T.D., Smith, R.D., 2001. Phosphoprotein isotope coded affinity tag approach for isolating and quantitating phosphopeptides in proteome wide analyses. Anal. Chem. 74, 2578–2586.

Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., Weiss, W., 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 21, 1037–1053.

Grant, S.G.N., Blackstock, W.P., 2001. Proteomics in neuroscience: from protein to network. J. Neurosci. 21, 8315–8317.

Gronborg, M., Kristiansen, T.Z., Stensballe, A., Andersen, J.S., Ohara, O., Mann, M., et al., 2002. A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. Mol. Cell Proteomics 1, 517–527.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., Aebersold, R., 2000. Evaluation of two-dimensional electrophoresis based proteome analysis technology. Proc. Natl. Acad. Sci. U.S.A. 97, 9390–9394.

Halligan, B.D., Ruotti, V., Jin, W., Laffoon, S., Twigger, S.N., Dratz, E.A., 2004. ProMoST (protein modification screening tool): a web-based tool for mapping protein modifications on two-dimensional gels. Nucl. Acids Res. 1, W638–W644 (Web Server Issue).

Henzel, W., Billeci, T.M., Stults, J.T., Wong, S.C., Grimley, C., Watanabe, C., 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc. Natl. Acad. Sci. U.S.A. 90, 5011–5015.

Herbert, B.R., Sanchez, J.C., Bini, L., 1997. Two-dimensional electrophoresis: the state of art and future directions. In: Wilkins, M.R., Williams, K.L., Appel, R.D., Hochstrasser, D.F. (Eds.), Proteome Research: New Frontiers in Functional Genomics. Springer, Berlin, pp. 13–33.

Herbert, B.R., Molloy, M.P., Gooley, B.A.A., Walsh, J., Bryson, W.G., Williams, K.L., 1998. Improved protein solubility in two-dimensional electrophoresis using tributyl phosphine as reducing agent. Electrophoresis 19, 845–851.

Herbert, B., Galvini, M., Hamdan, M., Olivieri, E., MacCarthy, J., Perdson, S., Righetti, P.G., 2001. Reduction and alkylation of proteins in preparation of two-dimensional map analysis: why, when and how? Electrophoresis 22, 2046–2057.

Hopkins, A.L., Groom, C.R., 2002. The druggable genome. Nat. Rev. Drug Discov. 1, 727–730.

Humphrey-Smith, I., Cordwell, S.J., Blackstock, W.P., 1997. Proteome research: complementarily and limitations with respect to the RNA and DNA worlds. Electrophoresis 18, 1217–1242.

Ibarrola, N., Kalume, D.E., Gronborg, M., Iwahori, A., Pandey, A., 2003. A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. Anal. Chem. 75, 6043–6049.

Jamin, N., Toma, F., 2001. NMR studies of protein–DNA interactions. Prog. NMR Spectrosc. 38, 83–114.

Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T., Watson, J.T., 1987. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. Anal. Chem. 59, 2621–2625.

Kimura, Y., Saeki, Y., Yokosawa, H., Polevoda, B., Sherman, F., Hirano, H., 2003. N-terminal modifications of the 19S regula-

tory particle subunits of the yeast proteasome. Arch. Biochem. Biophys. 409, 341–348.

Kjellstrom, S., Jensen, O.N., 2003. In situ liquid–liquid extraction as a sample preparation method for matrix-assisted laser desorption/ionization MS analysis of polypeptide mixtures. Anal. Chem. 75, 2362–2369.

Kryukov, G.V., Castellano, S., Novoselov, S.V., Lobanov, A.V., Zehtab, O., Guigo, R., et al., 2003. Characterization of mammalian selenoproteomes. Science 300, 1439–1443.

Kubinyi, H., 2003. Drug research: myths, hype and reality. Nat. Rev. Drug Discov. 2, 665–668.

Lesley, S.A., Kuhn, P., Godzik, A., Daecon, A.M., Mathews, I., Kreusch, A., Spraggon, G., Klock, H.E., et al., 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. U.S.A. 99, 11664–11669.

Luche, S., Santoni, V., Rabilloud, T., 2003. Evaluation of non-ionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. Proteomics 3, 249–253.

Luscombe, N.M., Greenbaum, D., Gerstein, M., 2001. What is bioinformatics? A proposed definition and overview of the field. Meth. Inf. Med. 40, 346–358.

Mann, M., Jensen, O.N., 2003. Proteomic analysis of post-translational modifications. Nat. Biotechnol. 21, 255–261.

Mann, M., Hojrup, P., Roepstorff, P., 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol. Mass Spectrosc. 22, 338–345.

Marshall, T., Williams, K.M., 1984. Artifacts associated with 2-mercaptoethanol upon high resolution twodimensional electrophoresis. Anal. Biochem. 139, 502–505.

McLachlin, D.T., Chait, B.T., 2001. Analysis of phosphorylated proteins and peptides by mass spectrometry. Curr. Opin. Chem. Biol. 5, 591–602.

Metodiev, M.V., Timanova, A., Stone, D.E., 2004. Differential phosphoproteome profiling by affinity capture and tandem matrix-assisted laser desorption/ionization mass spectrometry. Proteomics 4, 1433–1438.

Minn, A.J., Velez, P.A., Schendel, S.L., Liang, H., Muchmore, S.W., Fesik, S.W., et al., 1997. Bcl-xL forms an ion channel in synthetic lipid membranes. Nature 385, 353–357.

Morris, H.R., Paxton, T., Panico, M., McDowell, R., Dell, A., 1997. A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing. J. Prot. Chem. 16, 469–479.

Nakamura, A., Goto, S., 1996. Analysis of protein carbonyls with 2,4-dinitrophenyl hydrazine and its antibodies by immunoblot in two dimensional gel electrophoresis. J. Biochem. 119, 768–774.

Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., 1988. Improved staining of poteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using coomassie brilliant blue G-250 and R-250. Electrophoresis 9, 255–262.

Nicholson, J.K., Wilson, I.D., 1989. High resolution proton magnetic resonance spectroscopy of biological fluids. Prog. NMR Spectrosc. 21, 444–501.

Oda, Y., Nagasu, T., Chait, B.T., 2001. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. Nat. Biotechnol. 19, 379–382.

O'Farrel, P.H., 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250, 4007–4021.

Pasquali, C., Fialka, I., Huber, L.A., 1997. Preparative two-dimensional gel electrophoresis of membrane proteins. Electrophoresis 18, 247–257.

Patterson, S.D., 2003. Data analysis—the Achilles heel of proteomics. Nat. Biotechnol. 21, 221–222.

Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. U.S.A. 85, 2444–2448.

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., 1999. Probability based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3555.

Peti, W., Etezady-Esfarjani, T., Herrmann, T., Klock, H.E., Lesley, S.A., Wuthrich, K., 2004. NMR for structural proteomics of *Thermotoga maritima*: screening and structure determination. J. Struct. Funct. Genomics 5, 205–215.

Rabilloud, T., 1996. Solubilization of proteins for electrophoretic analysis. Electrophoresis 17, 813–818.

Rabilloud, T., Gianazza, E., Catto, N., Righetti, P.G., 1990. Amidosulfobetaines, a family of detergents with improved solubilization properties: application for isoelectric focusing under denaturing conditions. Anal. Biochem. 185, 94–102.

Rabilloud, T., Adessi, C., Giraudel, A., Lunardi, J., 1997. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 18, 307–316.

Renfrey, S., Featherstone, J., 2002. Structural proteomics. Nat. Rev. Drug Discov. 1, 175–176.

Righetti, P.G., Gelfi, C.J., 1984. Immobilized pH gradients for isoelectric focusing. III. Preparative separations in highly diluted gels. J. Biochem. Biophys. Meth. 9, 103–119.

Roepstorff, P., Fohlman, J., 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spectrom. 11, 601–603.

Rosengren, A.T., Salmi, J.M., Aittokallio, T., Westerholm, J., Lahesmaa, R., Nyman, T.A., Nevalainen, O.S., 2003. Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. Proteomics 3, 1936–1946.

Shevchenko, A., Wilm, M., Vorm, O., Mann, M., 1996. Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 68, 850–858.

Soskic, V., Gorlach, M., Poznanovic, S., Boehmer, F.D., Godovac-Zimmermann, J., 1999. Functional proteomics analysis of signal transduction pathways of the platelet-derived growth factor beta receptor. Biochemistry 38, 1757–1764.

Stanley, B.A., Gundry, R.L., Cotter, R.J., Van Eyk, J.E., 2004. Heart disease, clinical proteomics and mass spectrometry. Dis. Markers 20, 167–178.

Stasyk, T., Hellman, U., Souchelnytskyi, S., 2001. Life Sci. News 9, 9–12.

Staunton, D., Owen, J., Campbell, I.D., 2003. NMR and structural genomics. Acc. Chem. Res. 36, 207–214.

Steinberg, T.H., Jones, L.J., Haugland, R.P., Singer, V.L., 1996. Sypro orange and sypro red protein gel stains: one step fluorescent staining of denaturing gels for detection of nanogram level protein. Anal. Biochem. 239, 223–237.

Steinberg, T.H., Agnew, B.J., Gee, K.R., Leung, W.Y., 2003. Global quantitatibe phosphoprotein analysis using multiplex proteomics technology. Proteomics 3, 1128–1144.

Tateno, Y., Miyazaki, S., Ota, M., Sugawara, H., Gojobori, T., 2000. DNA data bank of Japan (DDBJ) in collaboration with mass sequencing teams. Nucl. Acids Res. 28, 24–26.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., Davison, M., 2001. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. Proteomics 1, 377–396.

Unlu, M., Morgan, M.E., Minden, J.S., 1997. Difference gel electrophoresis: a single gel method of detecting changes in protein extracts. Electrophoresis 18, 2071–2076.

Vercoutter-Edouart, A.S., Czeszak, X., Crepin, M., Lemoine, J., Boilly, B., Le Bourhis, X., Peyrat, J.P., Hondermarck, H., 2001. Proteomic detection of changes in protein synthesis induced by fibroblast growth factor-2 in MCF-7 human breast cancer cells. Exp. Cell Res. 262, 59–68.

Weissig, H., Bourne, P.E., 1999. An analysis of the Protein Data Bank in search of temporal and global trends. Bioinformatics 15, 807–831.

Westermeier, R., Naven, T., 2002. Expression proteomics. In: Proteomics in Practice—A Laboratory Manual of Proteome Analysis. Wiley/VCH, Germany, 11–159.

Wilkins, M., 1995. Government backs proteome proposal. Nature 378, 653.

Wilkins, M.R., Gasteiger, E., Tonella, L., Ou, K., Tyler, M., Sanchez, J.C., Gooley, A.A., Walsh, B.J., Bairoch, A., Appel, R.D., Williams, K.L., Hochstrasser, D.F., 1998. Protein identification with N and C terminal sequence tags in proteome projects. J. Mol. Biol. 278, 599–608.

Yamagata, A., Kristensen, D.B., Takeda, Y., Miyamoto, Y., Okada, K., Inamatsu, M., et al., 2002. Mapping of phosphorylated proteins on two dimensional polyacrylamide gels using protein phosphatase. Proteomics 2, 1267–1276.

Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G.M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C.H., Cort, J.R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L.P., Gehring, K., Kennedy, M.A., Edwards, A.M., Arrowsmith, C.H., 2002. An NMR approach to structural proteomics. Proc. Natl. Acad. Sci. U.S.A. 99, 1825–1830.

Ying, W., Hao, Y., Zhang, Y., Peng, W., Qin, E., Cai, Y., et al., 2004. Proteomic analysis on structural proteins of Severe Acute Respiratory Syndrome coronavirus. Proteomics 4, 492–504.

Zaia, J., 2004. Mass spectrometry of oligosaccharides. Mass Spectrom. Rev. 23, 161–227.

Zhang, W., Chait, B.T., 2000. Profound: an expert system for protein identification using mass spectrometric peptide mapping information. Anal. Chem. 72, 2482–2488.

Zhou, H., Watts, J.D., Aebersold, R., 2001. A systematic approach to the analysis of protein phosphorylation. Nat. Biotechnol. 19, 375–378.