Annu. Rev. Biomed. Eng. 2002. 4:349–73
doi: 10.1146/annurev.bioeng.4.020702.153443

ADVANCES IN PROTEOMIC TECHNOLOGIES

Martin L. Yarmush and Arul Jayaraman
Center for Engineering in Medicine/Surgical Services, Massachusetts General Hospital, Harvard Medical School and Shriners Burns Hospital, Boston, Massachusetts 02114; e-mail: ireis@sbi.org

Key Words two-dimensional gel electrophoresis, mass spectrometry, protein maps, protein-protein interactions

■ **Abstract** Proteomics is a rapidly emerging set of key technologies that are being used to identify proteins and map their interactions in a cellular context. With the sequencing of the human genome, the scope of proteomics has shifted from protein identification and characterization to include protein structure, function and protein-protein interactions. Technologies used in proteomic research include two-dimensional gel electrophoresis, mass spectrometry, yeast two-hybrids screens, and computational prediction programs. While some of these technologies have been in use for a long time, they are currently being applied to study physiology and cellular processes in high-throughput formats. It is the high-throughput approach that defines and characterizes modern proteomics. In this review, we discuss the current status of these experimental and computational technologies relevant to the three major aspects of proteomics—characterization of proteomes, identification of proteins, and determination of protein function. We also briefly discuss the development of new proteomic technologies that are based on recent advances in analytical and biochemical techniques, engineering, microfabrication, and computational prowess. The integration of these advances with established technologies is invaluable for the drive toward a comprehensive understanding of protein structure and function in the cellular milieu.

CONTENTS

INTRODUCTION

Recent progress in the sequencing of genomes has had significant impact on biomedical research leading to the development of a new field called functional genomics, which utilizes global approaches to study gene function and interaction at the mRNA, protein, and various functional levels (1). Although rapid advances in our ability to obtain genomic information have been made, less progress has been seen in our ability to relate this information to the gene products, i.e., proteins. It is becoming increasingly clear that knowing the genome sequence will not be sufficient to provide insights into complex cellular, organ, and system level phenomena. Thus, an adjunct enterprise has emerged that encompasses the study of the structure, function, and interaction of proteins, and this enterprise forms the realm of "proteomics."

The term proteome refers to the protein equivalent of the genome. It comprises the complete set of gene products that are synthesized by the genome, and the analysis of these gene products has been termed proteomics. The field of proteomics is quite daunting as the proteome is exceedingly dynamic, extremely complex, and largely defines cellular behavior and function. A vast array of analytical methods including two-dimensional gel electrophoresis (2DE) and amino acid sequencing has been used in the past for the identification and characterization of proteins. Although these techniques have been extremely useful, they suffer from both technical limitations and the lack of high-throughput formats that have become mandatory for protein characterization. Hence, there has been a recent interest in the adaptation of these traditional proteomic technologies to high-throughput formats as well as in the development of new technologies for studying protein function and interactions.

This review provides an overview of technologies that are important in studying three aspects of proteomics—characterization of proteomes, identification of proteins and their modifications, and determination of protein function. Because proteomics is primarily a technology-driven area of research, our discussion of proteomics is based on the advances and developments in the enabling technologies required for studying these three aspects of proteomics. Our approach is focused on the amenability of the individual technologies to high-throughput formats required for understanding cellular behavior and function. We restrict our discussion on well-established technologies such as 2DE to their limitations, whereas for more recently developed technologies such as mass spectrometry (MS), yeast two-hybrids (Y2H), and computational techniques, we concentrate on both existing limitations as well as innovative improvements. We also briefly discuss several emerging proteomics methods that are being developed to address these limitations. A list of the proteomic technologies described in this review is given in Table 1.

SCOPE OF PROTEOMICS

The importance of proteomics in all areas of biomedical research is evident as proteins are the main functional agents inside the cell. Transcript levels as measured

TABLE 1 Overview of proteomic technologies discussed in this review

	Advantages	***Problems***
Proteome profiling		
2D electrophoresis	Good resolution of proteins Detection of posttranslational modifications	Poor protein solubility Limited dynamic range of detection Analysis and quantitation are difficult
Protein identification		
mass spectrometry	Determination of molecular weight and amino acid sequence information Detection of posttranslational modifications High-throughput capability	High capital costs Requires sequence databases for analysis
Protein function		
yeast two hybrid	Pair-wise interaction of proteins Functional assignment to unknown protein High-throughput capability	False positives and negatives Secondary assays needed for confirming interactions
Immunoaffinity	Identification of protein complexes Functional assignment to proteins without *a priori* knowledge of function	High affinity needed between all proteins in the complex Protein loss during purification
Homology	Detailed information on specificity and function	Homologous protein structures needed
Nonhomology	Interaction with unknown proteins Functional assignment to unknown proteins	Complete genome sequences required Limited coverage of genome Inaccuracy of predictions
Molecular scanner	Annotated 2D images obtained High-throughput capability Minimal sample handling and processing steps	Time required for scanning
ICAT	Sensitive, quantitative method Unbiased profile of proteome	Cysteine residues must be present for labeling Analysis of spectra is difficult
Protein chips	High-throughput capability Easily scalable	Not yet fully standardized

by microarrays and PCR-based methods do not provide comprehensive information on the proteome, as gene expression can be regulated at the transcriptional, translational, or posttranslational levels. For example, studies in two very different systems, namely liver (2) and yeast (3, 4), have been unable to establish an exclusive direct relationship between the transcriptome and the proteome. This is because mRNA transcripts can be assembled or spliced together in multiple ways to yield several protein isoforms (5), and preliminary studies have suggested the presence of at least three to six protein forms per human mRNA transcript (6). In addition, transcript profiles do not provide any information on posttranslational modifications of proteins (such as phosphorylation and glycosylation) that are crucial for protein transport, localization, and function (7). Lastly, proteins are constantly changing in response to the intracellular and extracellular environment, and this dynamic nature of protein expression make them better suited than the transcriptome for describing complex cell behavior and function.

Blackstock & Weir (8) have proposed that the scope of proteomics encompasses two distinct areas of research—the expression of proteins and their interactions. The goal of expression proteomics is to establish quantitative maps of protein expression under specific conditions (such as environmental stimuli, disease states, etc.). This approach is especially useful in drug development or toxicological studies where it is of interest to profile the entire proteome in response to a particular perturbation and identify biomarker proteins whose expression levels can be used to diagnose the onset of a disease or to evaluate the efficiency of drug targets (6). An example of this approach is the work of Anderson et al. (9) in which proteins involved in cholesterol metabolism affected by the drug lovastatin were probed. For an expression proteomics approach to be successful, it is mandatory to generate a truly representative proteome map that can resolve a complex mixture of proteins for the quantitation of relevant proteins. In contrast to the open-ended nature of expression proteomics, interaction proteomics is focused towards the specific characterization of protein-protein interactions and the formation of protein complexes (10). Identification of interacting proteins would facilitate the study of specific processes and pathways in the cell and lead to the development of "interaction maps" for various cell types and conditions, thereby increasing our understanding of intracellular networks (8, 10).

PROTEOMIC TECHNOLOGIES

The complete characterization of a proteome is a formidable task and the degree of success achieved depends on the methods available and their amenability to automation and high throughput formats (11, 12). Parameters such as the complexity of the protein mixture, levels of expression and modification, and intracellular localization all impact the choice of proteomics technology to be used (13). Proteomic technologies (Figure 1) can be broadly classified into those used for protein mapping and characterization, and those for studying protein-protein interactions and protein function.

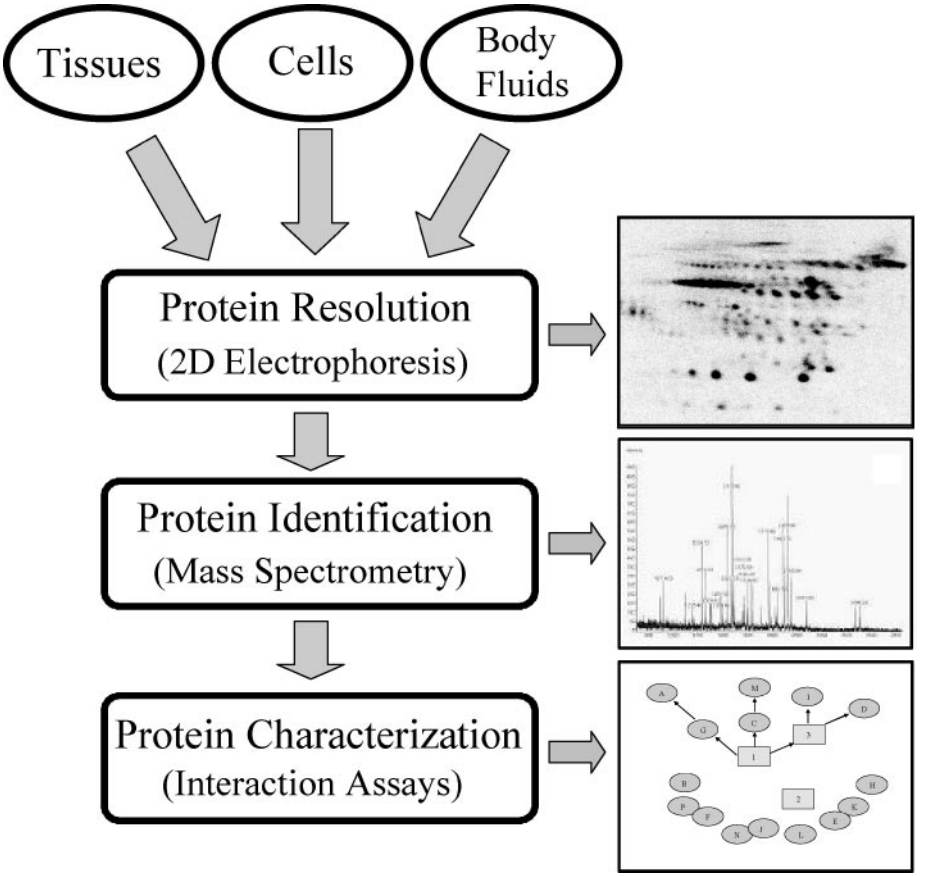

Figure 1 Generalized proteomics scheme involving proteome profiling, protein identification, and characterization of protein-protein interactions.

Technologies for Generating Protein Maps and Protein Characterization

Proteome maps or fingerprints are obtained using two-dimensional gel electrophoresis (2DE) (14), a technology that is synonymous with proteomics and has been the method of choice for the analysis of proteomes for nearly 25 years. Protein characterization, including detection and identification of posttranslational modifications, is achieved in proteomics using mass spectrometry (MS).

TWO-DIMENSIONAL ELECTROPHORESIS FOR PROTEOME PROFILING In 2DE, proteins are separated first by their isoelectric point (pI), followed by their separation based on molecular weight (MW) (Figure 2). The combination of these two orthogonal separation techniques resolves proteins into spots (each spot is a

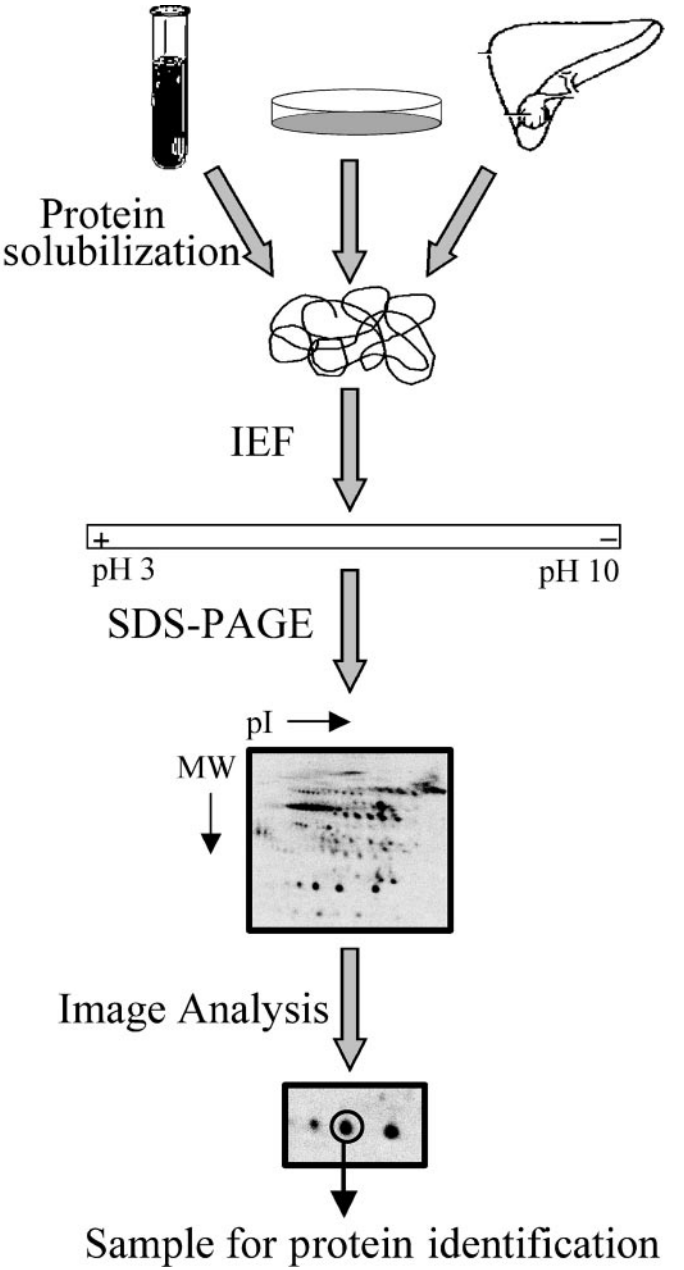

Figure 2 Typical two-dimensional electrophoresis (2DE) scheme for generating protein maps. Proteins are solubilized and separated based on their isolelectric point (first dimension) and molecular weight (second dimension). Protein profiles are compared using image analysis and protein spots of interest excised for identification and characterization.

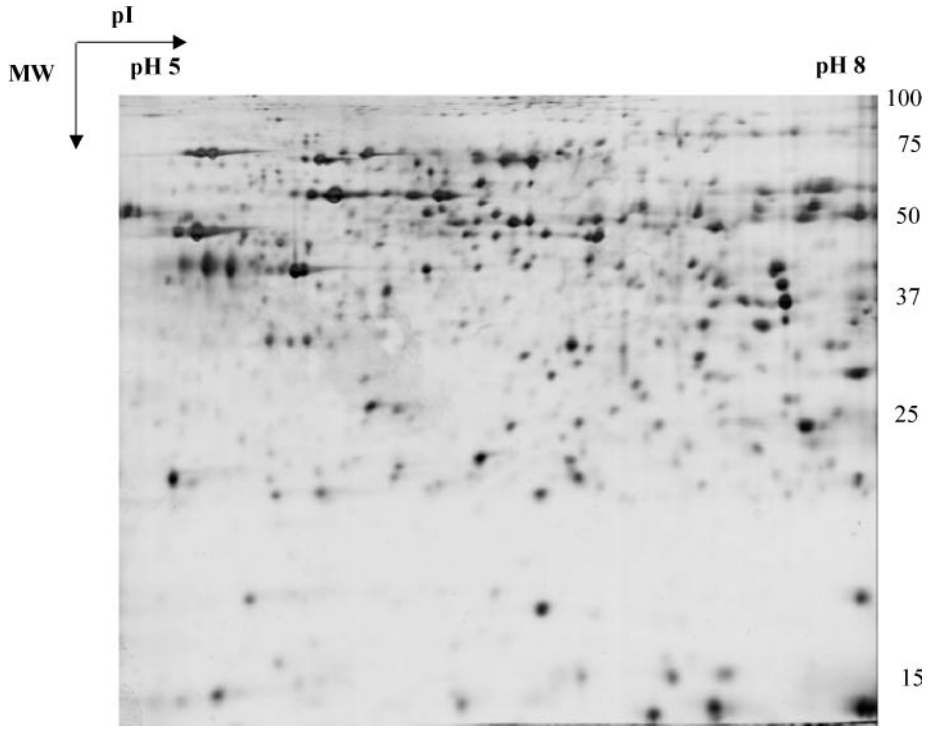

Figure 3 Silver stained two-dimensional electrophoresis profile of rat liver proteins. Approximately 40 μg of protein was separated first by isoelectric focusing on pH 5–8 strips for 22,000 volt-hours followed by 2nd dimension SDS-PAGE on a 12% gel and silver staining. The gel was imaged and analyzed using the Melanie II software (Bio-Rad, Hercules, CA).

protein isoform with specific pI and MW coordinates), and this map of protein spots can be considered as the "protein fingerprint" of that sample (Figure 3). Two such fingerprints from a normal and abnormal cellular state can be compared to identify proteins of relevance to that particular cellular state or phenotype. Typically, 2DE is used in expression proteomics studies where the focus is on studying alterations in protein expression profiles due to the appearance of a new protein spot, or the disappearance of a protein spot, or changes in the intensity of an existing protein spot. While the resolution of complex protein mixtures obtained with 2DE is far superior to that with conventional one-dimensional protein electrophoresis, current 2DE methods have several technical drawbacks that limit their widespread application. Some of these issues are discussed below.

Protein solubility One of the major limitations of 2DE is the difficulty in obtaining a protein profile that is representative of the entire proteome. Cellular proteins

have widely varying solubilities, and while the typical urea concentration in 2DE solubilization buffers is sufficient for the initial solubilization of most proteins, it is often not enough to solubilize all proteins as well as maintain their solubility during isoelectric focusing (15). Membrane proteins are difficult to solubilize and this results in significant amounts of protein being lost during the sample preparation as well as isoelectric focusing (16). Hydrophobic proteins that constitute nearly 30% of certain cellular proteomes (17–19), basic (pI > 8) proteins, and high-molecular-weight (∼200 kDa) proteins can all precipitate during isoelectric focusing and are poorly resolved by 2DE. While the use of zwitterionic detergents such as ASB 14 (15) has improved the solubilization of hydrophobic proteins, the presence of these detergents may also affect the resolution of other classes of proteins (e.g., hydrophilic proteins) present in the protein mixture.

Fractionation of proteins based on common biophysical properties is an alternate approach that has been used to increase protein solubility in 2DE. An example of this approach is the sequential extraction of proteins, where different buffers are used to generate protein fractions that are enriched for certain proteins (hydrophobic, basic, etc.) (20, 21). Solution-phase isoelectric focusing prior to 2DE (22) and adsorption chromatography (23, 24) has also been used to generate enriched basic and hydrophobic protein fractions. The establishment of subcellular proteomes is another approach where proteins are fractionated based on their intracellular localization so that a single buffer can be used for solubilizing a majority of proteins in that fraction (25, 26).

Protein capacity Even though 2DE is still considered the workhorse for most global proteome analysis studies, its applicability is often limited owing to the poor resolution of complex protein mixtures. Consequently, the proteome profile generated by 2DE does not always represent the entire proteome. Approximately 5,000–10,000 genes are expressed in a cell at any given time, which could result in the synthesis of at least 20,000–30,000 distinct proteins because of mRNA splicing and posttranslational modification (11, 27). However, using current 2DE methods, it is only possible to detect approximately 3,000 protein spots on an 18×20 cm^2 gel (28). Therefore, only the most abundant proteins (e.g., housekeeping proteins, metabolic enzymes) are detected on a 2D gel, yielding a distinct bias against lower abundance proteins (e.g., signaling molecules, transcription factors) (29, 30).

Gygi et al. (29) have demonstrated this bias against low abundance proteins by measuring the codon bias value (a measure of the tendency to preferentially utilize the same codon during protein synthesis, which results in the same amino acid being incorporated in the synthesized polypeptide chain) for the 6,139 known genes of the yeast genome. Low abundance proteins have a codon bias value of less than 0.1, and of all the yeast proteins detected on a 2D gel, almost no protein had a codon bias value of less than 0.1. Because nearly 50% of the yeast genome has a codon bias value of less than 0.1, it is evident that protein profiling with 2DE cannot represent the entire proteome. Sample prefractionation is the simplest approach

that has been adopted for improving the loading capacity of 2DE and generating a more representative proteome profile. The use of very narrow pH gradients (1 pH unit over an 18-cm gel) for isoelectric focusing has been adopted as a means to improve the resolution and detect low abundance proteins (31). This pI-based prefractionation method allows the separation of protein loads as high as 15 mg and detection of proteins that may have been difficult to achieve otherwise (16, 32). However, the costs and time associated with large gels make it impractical for routine 2DE. Conventional analytical separation techniques have also been linked to 2DE to enhance the resolution of proteins. One such example is the use of non-denaturing anion exchange column chromatography by Butt et al. (33) to prefractionate an *E. coli* soluble protein sample prior to 2DE. This approach yielded up to 13-fold enrichments for individual protein spots and also facilitated subsequent protein characterization.

Protein detection and quantitation The dynamic range of detection methods used with 2DE also limits their applicability. In a cell, proteins are expressed over a broad range and the abundance may cover nearly seven orders of magnitude (34–36). However, even the most sensitive gel staining method available has a dynamic detection range of only four to five orders of magnitude (36), which introduces errors in protein detection as well as quantitation. The linear range of detection has been improved with the development of fluorescent protein stains. Noncovalent stains such as the SYPRO dye from Molecular Probes (Eugene, OR) have also considerably improved the sensitivity of staining methods and provide a broader linear dynamic range of detection than Coomassie brilliant blue and silver staining (37, 38).

Another drawback of current 2DE technology is the inability to reliably quantify protein levels from proteome profiles. Protein spot quantitation depends to a large extent on the detection and resolution of each protein spot using image analysis. Spot resolution is the first step of most commercially available 2D image analysis programs like MELANIE (39, 40) and PDQUEST (41), which align the gel images for determining differences in expression patterns. Accurate spot assignment is also crucial for protein identification, and often, protein spots are excised from the gel for subsequent analysis.

The inherent variability in 2DE and limitations in both simple imaging and quantification systems have led to a recent upsurge in the development of more sophisticated image analysis systems. While improvements to spot detection algorithms have been made, almost all 2D gel image analysis programs still require some degree of manual intervention to verify the accuracy of detected spots. This verification process is extremely tedious, time consuming, and is a major bottleneck in the efforts to automate proteomic methods (42). A recently developed 2D image analysis algorithm Z3 (42) uses unprocessed gel images for image registration and the subsequent spot detection steps. Essentially, multiple features present in the gel (such as streaks and spot smears) are used for spot detection to enhance

the accuracy of matching of spots across gels and quantitation. This enhanced image registration technique has been shown to be useful for identifying very subtle shifts in protein spot location due to posttranslational modifications (42).

MASS SPECTROMETRY FOR PROTEIN CHARACTERIZATION The desired end-point of expression proteomics is the characterization of all cellular proteins in terms of their abundance, modification, localization, and function. Such a comprehensive database of cellular proteins could serve as a tool for studying the proteomic response of cells or tissues in response to a variety of conditions (30), and depends to a great extent on the methods available for protein identification and characterization. A relatively new and rapidly evolving development in proteomics research has been the application of mass spectrometry (MS) (Figure 4), which, in conjunction with the development of comprehensive protein databases (43) and advances in computational methods (44), is being used for high-throughput characterization and identification of proteins. MS can be used to determine the molecular weight as well as the amino acid composition of proteins (43–46) at low concentrations (attomole to femtomoles). MS is also easily adaptable to high-throughput formats, which has made it the method of choice for protein identification (3, 44, 47) and characterization (46).

Ionization sources for mass spectrometry Matrix-assisted laser desorption-ionization (MALDI) and electrospray ionization (ESI) are two technologies that are commonly used for protein ionization. In MALDI (48), the protein samples to be analyzed are incorporated in an energy-absorbing matrix (44). The matrix is typically a small molecule such as dihydrobenzoic acid or 4-hydroxycinnamic acid that is deposited onto a metal substrate along with the sample (46), and ions are generated when the matrix is struck and excited by a laser. MALDI generates single charged ion species so that a direct correlation between the mass spectra and the levels of the corresponding protein in the sample is obtained (45). In ESI mass spectrometry (49), the protein sample is in solution, and a potential is applied to create a fine mist of charged droplets that are subsequently dried and introduced into the mass analyzer (44). In contrast to MALDI, ESI produces highly charged ions without fragmentation of the ions in the gas phase (46). Proteins, oligonucleotides, sugars, and polar lipids of all molecular weights can be analyzed with ESI MS (46) as multiple charges can be added to the molecules lowering the mass-to-charge (m/z) ratio to a range detectable by mass analyzers (45, 46).

Mass analyzers for separation Time-of-flight (TOF) and quadrupole mass analyzers have been developed for use in mass spectrometers (45, 50), and of these, TOF analyzers are more common because of their ease of operation. TOFs are commonly used with a MALDI ion source, whereas quadrupole analyzers are often combined with an ESI source (51). A TOF analyzer measures the time taken for the gas-phase ions to travel from the ionization source to the detector (44, 50),

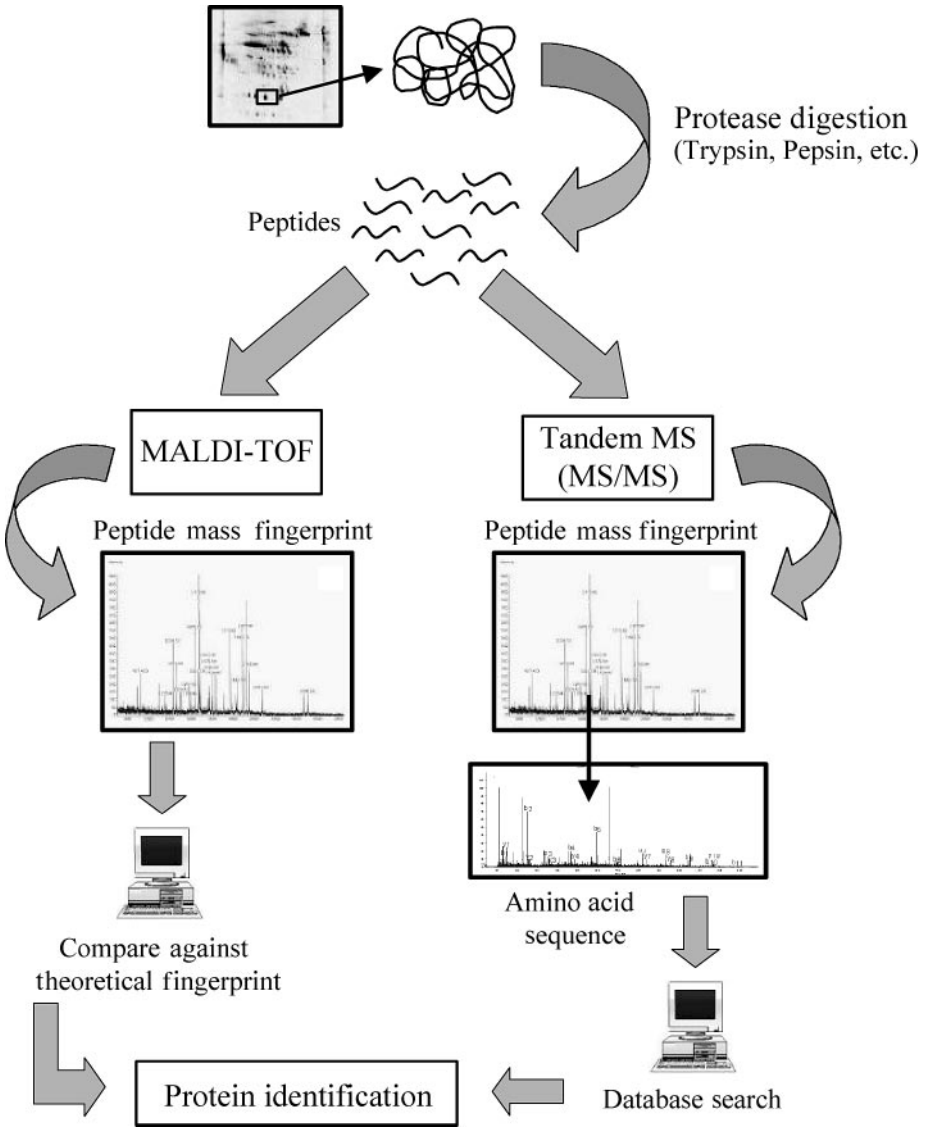

Figure 4 Typical mass spectrometry scheme for protein identification using peptide mass mapping and tandem mass spectrometry. Protein spots from a 2D gel are digested and the peptides analyzed using either a MALDI-TOF or a tandem MS. The peptide mass fingerprint or amino acid sequence are then compared against databases to identify the protein.

which is then related to the m/z ratio. A quadrupole mass analyzer consists of four parallel metal rods that are arranged lengthwise and can be manipulated to allow ions of a specific m/z ratio to pass between them for detection (46, 51). Coupling multiple quadrupoles together has further enhanced the power of this mass filtering technique so that ions with a range of characteristics can be selected with the first quadrupole scanning a range of m/z values and a single m/z value being selected with the second quadrupole (45, 46).

The accurate determination of protein molecular weights is mainly achieved using a MALDI-TOF instrument. Proteins are proteolytically digested to generate 1–2 kDa peptides, and are applied to the mass spectrometer to generate a mass spectrum or peptide mass fingerprint (PMF) (52). These PMFs are compared with a database of virtual PMFs generated by the theoretical digestion of known proteins by specific proteases (53). Matches from at least three to six peptides derived from the same protein are required to positively identify a protein (53, 54).

The success of mass mapping depends on the ability to generate peptides by digestions and the availability of sequences for generating theoretical PMFs. A 50%–90% success rate has been reported for the identification of proteins from fully sequenced organisms (46). However, not all proteins (e.g., small acidic proteins) yield the required number of peptides from a single protease digestion (44). Therefore, multiple proteases with unique substrate recognition sequences are being used to generate enough peptides for accurate mass mapping (45). Mass mapping also requires the presence of a large fraction of each protein sequence in databases for generating theoretical PMFs, which is crucial for the detection of protein isoforms in complex protein samples.

Tandem MS (MS/MS) has been used to obtain the amino acid sequence of peptides and is proving to be more useful in protein identification than mass mapping. Peptide ions generated from an ESI source are separated based on the m/z ratio and further dissociated by collision with an inert gas (46). The resultant tandem spectra of amino acid composition can be searched against protein, expressed sequence tags (ESTs), and genome databases to identify the protein (3, 55). The breadth and coverage of the available databases are extremely crucial as database-searching strategies can be applied only if the protein sequence exists in the database. Sequest, developed at the University of Washington (56), is the most widely used tool for searching protein databases (57). Sequest is ideal for high-throughput proteomics as it automatically extracts and searches the MS/MS data against the database (19, 57). Other programs that are available on the web for the analysis of tandem spectra include PROWL (58), Protein Prospector (59), and MASCOT (60). In the absence of sufficient protein sequence information in databases, de novo sequencing of peptides (51) and hydrogen/deuterium exchange (61) have also been used for the deduction of complete peptide sequence from MS/MS spectra.

EMERGING TECHNOLOGIES Several new proteomic technologies are being developed for proteome profiling and protein characterization that utilize advances in

biochemical techniques, analytical methods, and computational prowess. Examples include the molecular scanner (62) that has been developed to enable high-throughput protein characterization after 2DE and the gel-free systems (29) to quantitatively characterize protein expression without any bias against low abundance proteins.

Molecular scanner Often, 2DE and MS are used in combination to obtain proteome profiles and identify specific proteins of interest. However, this approach is not amenable to high-throughput formats as protein spots excised from the 2D gels cannot be easily processed in parallel, and manually annotating and updating expression databases used for protein identification is extremely tedious. Recently, Binz et al. (62) have reported the development of a highly automated method that can be used to generate a fully annotated 2D map. This method still relies on the use of 2DE as the primary separation tool, wherein the proteins in a 2D gel are transferred to a membrane by electroblotting without staining. During the transfer process, the proteins pass through an interface whose surface is coated with immobilized trypsin so that the proteins are proteolytically digested during the transfer. The resultant tryptic peptides are blotted onto a membrane, which now contains a set of tryptic peptides at each location corresponding to a protein spot on the gel (63). The membrane is then sprayed with a MALDI matrix solution, scanned by mass spectrometry, and the mass spectra are used for protein identification and the generation of an annotated image. Binz et al. (62) have demonstrated the utility of the molecular scanner by obtaining a 2D scan map of human plasma.

The molecular scanner is a high-throughput method for identifying all the protein spots on a gel in parallel. The amount of protein lost due to sample handling and manipulation is minimized, and there are no issues with linearity of protein detection as staining of the gel and membrane is completely avoided. Further, the high-resolution mass fingerprint is capable of resolving protein spots that overlap and comigrate on 2D gels. Currently, the method is primarily limited by the time taken to scan the membrane with the mass spectrometer (nearly 36 hours for a $4 \times 4\,\text{cm}^2$ membrane) (62), but this is sure to be addressed as the mass spectrometry technology develops further. Nevertheless, the development of a molecular scanner represents a significant step towards high-throughput resolution and identification of proteins.

Gel-free proteomics The quantitative measurement of protein expression in cells and tissues under different conditions is a major goal of proteomics, for which proteome profiling needs to be truly unbiased and global. As mentioned earlier, current 2DE methods are biased against low abundance proteins in the cell and sample prefractionation has been adopted to increase the coverage of proteins in a 2D profile; however, this renders protein quantitation nearly impossible (29). Therefore, there is a real need for techniques that allow the quantitative measurement of proteins in a proteome.

Recently, Gygi et al. (64) have developed isotope-coded affinity tags (ICAT) for accurate quantitation of protein expression. ICAT reagents with a light or heavy isotope (differing by eight mass units) are used for differential labeling of the two protein samples by alkylation of cysteine residues and analysis by microcapillary LC-ESI-MS/MS. Changes in expression are determined from the difference in the intensities of the two peaks produced by the light and heavy isotopes. This method is not biased towards a particular protein type and can be used for global quantification of all cellular proteins. ICAT labeling is dependent on the presence of cysteine residues in all the proteins, and the ICAT label can lead to complications with the mass spectra analysis and database searching (29). In spite of these drawbacks, the gel-free separation technique offers improvement in protein identification and quantification, and has been demonstrated in the comparison of protein profiles between yeast grown on either ethanol or galactose as the carbon source (64).

Technologies for Studying Protein-Protein Interactions and Protein Function

Protein function can be described based on its role either in the behavior of the organism (phenotypic function) or a cell type (cellular function) and in terms of the interactions with other molecules (molecular function) (65). The assignment of functional annotations to proteins is no trivial task. Sali (66) estimates between 5 and 50 functional interaction links per protein in yeast, which adds up to between 30,000 and 300,000 total links for the entire yeast proteome. For complex eukaryotic systems, this number is expected to be even larger, thus driving the development of experimental and computational approaches to study protein interactions.

EXPERIMENTAL METHODS Experimental approaches to derive functional linkages between proteins have been based on advances in molecular biology, biochemistry, and analytical chemistry (67). Pair-wise protein interactions are often studied using the yeast two-hybrid system (68–70), whereas protein complexes are typically identified with immunoaffinity methods (3, 11).

Yeast two-hybrid systems The yeast two-hybrid (Y2H) screen has been used to identify pairs of interacting proteins (11) and is based on the modular nature of transcription factors and the transcriptional activation of reporter genes (3, 69, 70). Each gene sequence encoding for one element of a protein pair to be tested is fused to either the transcription factor binding or activation domain and expressed as fusion proteins in yeast (3, 70) (Figure 5). If the two proteins interact, the fused transcription factor binding and activating domains will also interact, initiating transcription of a reporter gene such as β-galactosidase (Figure 5). The clones yielding a positive interaction are then isolated, sequenced, and searched against genome databases to identify the interacting protein pair (70).

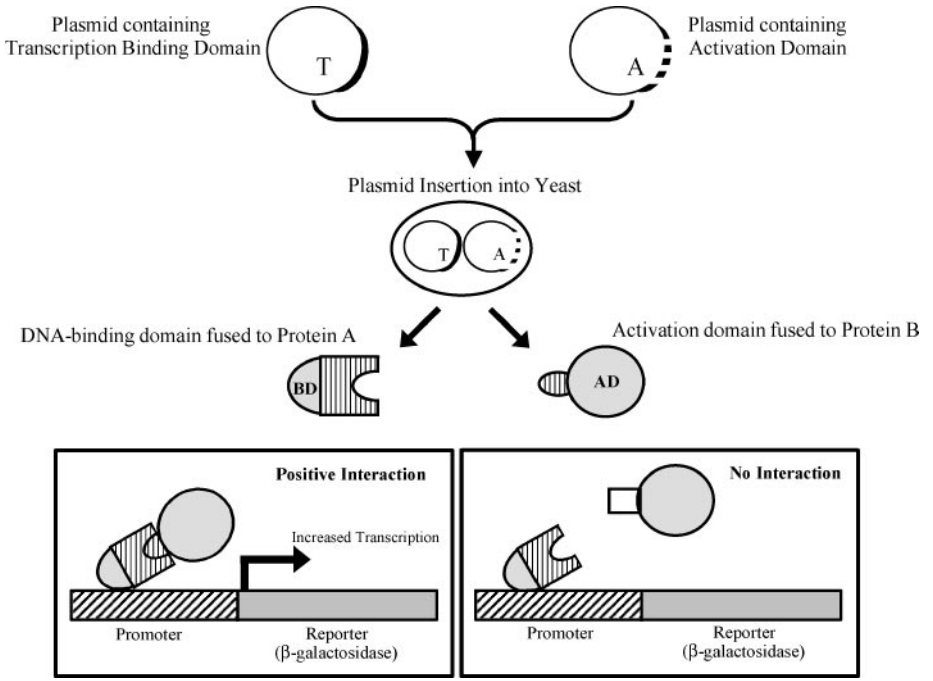

Figure 5 Yeast two-hybrid (Y2H) scheme for identifying protein-protein interactions. Putative interaction partners are expressed as fusions to either the DNA-binding domain or its activation domain in yeast. If the two proteins interact, the fusion proteins are brought into proximity and results in transcription of the reporter gene.

Y2H screens can be used to detect novel interactions between proteins of known and unknown function and identify protein-protein interactions to define a cellular phenotype (1). A classic example of Y2H is the study by Ito et al. (71) who identified 183 independent two-hybrid interactions among the yeast open reading frames (ORFs), of which nearly 50% were previously unknown. Y2H screens have also been applied to numerous other systems (72, 73) in either a labor-intensive matrix format or high-throughput array format (72, 73).

The tremendous potential of Y2H in identifying interaction partners is tempered by a few limitations. A major problem associated with Y2H screens is the generation of false positives and negatives. A positive identification by the Y2H screen merely suggests the existence of an interaction without providing any insight into its biological relevance. Quite often, potential protein interactions identified by two-hybrid screens may not be physiologically meaningful and may represent interactions that are not real, i.e., false positives. Similarly, if the experiment is performed under conditions of high stringency, weak interactions could be excluded, resulting in false negatives (74). Therefore, any putative interaction

identified by Y2H needs to be verified by rigorous experimentation, which limits the applicability of Y2H to high-throughput proteomics.

Immunoaffinity methods The interaction of several proteins in a multi-protein complex has been studied using immunoaffinity methods (11). Several methods have been used for isolating protein complexes, of which the most common is the use of Glutathione *S*-transferase (GST) fusions (3) where the protein of interest whose interaction partners need to be determined is expressed as a fusion protein with a cleavable GST tag and immobilized to a solid support. The immobilized protein is incubated with a cell extract containing all the target proteins so that multi-protein complexes can be formed. This protein complex can then be cleaved off from the GST tag, purified, and the bound proteins identified by 2DE and MS (3). Using this approach, the human spliceosome multi-protein complex has been assembled using biotinylated RNA as the bait and resolved on a 2D gel to identify 19 new factors (75). Other protein complexes that have been purified and identified using this approach include the GroEL chaperonin system and proteins found in the yeast nuclear-pore complex (76).

Because no a priori knowledge of the interacting proteins is assumed, this approach can be used to identify new interactions between known proteins in a cellular pathway or novel interacting proteins. The purification of protein complexes not only helps to identify its protein constituents, but also provides insights on the relationship between apparently unrelated pathways in the cell (3). A crucial requirement for the identification of protein complexes is high affinity between the bait protein and the other complex proteins to ensure that protein elements of the complex are not lost during purification (11).

COMPUTATIONAL METHODS The problems associated with experimental methods such as Y2H have resulted in the development of computational approaches to predict protein-protein interactions. These computational methods primarily seek to utilize information embedded in patterns of gene and protein expression amongst multiple genomes to derive functional links between different proteins (77). Current approaches to predict protein-protein interactions are either based on homology between proteins (78, 79) or on evolutionary information about the proteins (77, 80, 81).

Homology methods Sequence homology is the standard method to establish links for a protein of unknown function with a proteins of known function. This approach is based on the hypothesis that homologous proteins that have evolved from a common ancestor protein and have similar sequence and structure would also have similar function (81–83). Several homology-based schemes used to predict the potential interaction partners of a protein have been based on protein surface properties, solvent accessibility (65, 84), and amino acid sequence (82). Bock & Gough (85) outlined a variation to this theme by developing a Support Vector Machine (SVM) learning system to predict protein interactions among

eukaryotic proteins. Protein data from a database of interacting proteins have been used to train the program to predict protein interactions based solely on protein primary structure and physiochemical properties. This method has been able to predict four out of five potential protein interactions correctly and is a promising development towards the rapid determination of protein interactions following their identification (85). The number of fully characterized homologous protein structures available for comparisons limit the prediction of protein interactions based on homology (66). Further, these predictions are based on protein structures in ideal solutions and are independent of the cellular context in which the proteins function (81).

Non-homology methods Alternate methods for predicting protein interactions use functional similarity instead of sequence similarity for the prediction of protein interactions. Non-homology methods can be used to assign function to uncharacterized proteins and also establish networks of protein interactions. Proteins that belong to the same pathway or cellular process interact with each other and are classified as functionally linked. The functions of uncharacterized proteins are typically inferred from the functionally characterized proteins with which they share functional links (77, 81). Phylogenetic profiling, domain fusions, and gene ordering are three methods that have been used for determining protein interaction maps and functional linkages (81).

Phylogenetic profiling for functional linkage of two proteins is based on their expression being conserved over a set of genomes under similar conditions. If this condition is satisfied, the two proteins are said to share a similar phylogenetic profile and are assigned a functional link (86) as the expression pattern of two proteins will not be identical in two genomes under the same conditions unless they are functionally related (77, 81). This property has been used to develop a comprehensive interaction of all protein-protein interactions in fully sequenced genomes. Pellegrini et al. (86) have computed phylogenetic profiles for all 4,290 *E. coli* proteins by comparing them with 16 fully sequenced genomes. This is an extremely powerful method for predicting protein interactions as up to 2^n phylogenetic profiles can be derived from n fully sequenced genomes. Therefore, with the number of fully sequenced genomes expected to increase rapidly, the utility of phylogenetic profiling in predicting protein interactions can also be expected to increase.

The domain fusion method identifies functionally related proteins by analyzing domain fusion patterns (87). Proteins such as GyrA and GyrB that are expressed as non-homologous proteins in *E. coli* are expressed in yeast as a single fused topoisomerase II protein, implying that GyrA and GyrB are linked in their function in yeast. The fused protein that reveals the functional link is called the Rosetta stone protein (77, 81, 87). Marcotte et al. (87) have searched 4,290 *E. coli* proteins pairwise for similarity to a single protein in another genome to discover 6,809 protein pairs that were linked by domain fusions, many of which did not have any sequence homology. Therefore, the Rosetta stone method can predict protein pairs that have similar function and interactions and not predicted by homology searches.

The ordering of related genes in genomes has also been used for identifying interacting proteins (88). This approach is especially useful in analyzing protein interactions in prokaryotes where proteins are more often clustered in operons. If two genes are located next to each other in several genomes, a functional linkage may be inferred between them. The gene neighbor method has been used to identify functional links between enzymes in the arginine biosynthetic pathway in *Mycobacterium* (81).

While computational prediction methods for protein interactions are rapidly evolving, they often need to be corroborated by another comparable method (66). The accuracy of non-homology-based predictions and the range of interactions they cover vary, which are the major sources of error from these approaches. Marcotte et al. (87) have estimated that up to 30% of the pair-wise predictions in yeast that were obtained by phylogentic profiles and used for assigning functions to 2,557 uncharacterized proteins were erroneous (89). On the other hand, Enright et al. (80) were able to obtain only 215 functional links to *E. coli* proteins using the Rosetta stone domain fusion method, but with fewer errors. Despite the limitations in the accuracy of pair-wise prediction of interactions, computational methods for protein classification are extremely useful in functional genomics as a preliminary means to minimize the number of interactions to be studied. Moreover, as the number of genomes whose sequences are fully available increases, the number of predictions of protein function as well as the coverage and accuracy of protein function predictions will also increase.

EMERGING TECHNOLOGIES The bottlenecks in current approaches to study protein-protein interactions have led to the development of protein microarrays or protein chips (90) for the high-throughput characterization of protein interactions as well as other adjunct technologies for proteomics that utilize advances in microfabrication and microfluidics.

Protein chips Chip-based systems for large-scale analysis of proteins have lagged behind that of mRNA because of the lack of a standard method (such as the PCR for DNA) to amplify and generate large amounts of protein. In addition, for proteins to maintain their activity upon arraying, their three-dimensional structure has to be maintained. The diversity observed in protein structures also means that, unlike DNA, a single binding strategy for attaching proteins to surfaces will not work for all classes of proteins (91). Advances in microfabrication (92) and surface chemistry methods (93) have all enabled the development of several protein chip formats, including the commercially available ProteinChip® technology (94), microfluidic protein arrays (90), antibody arrays (95), and tissue arrays (96).

In the ProteinChip® process, proteins are exposed to chips with different surface chemistries in parallel so that sets of proteins with common properties (hydrophilic, hydrophobic, electrostatic interactions, etc.) will adsorb to a particular type of surface. After washing to remove unbound proteins, proteins of interest can be enriched on the chip surface by selective washing, and protease digestion (91).

ProteinChips® require very small sample volumes and can be used directly with biological fluids to provide information on protein structure, properties, and post-translational modifications (91, 94). Currently, ProteinChips® are being used for a wide variety of applications including the quantitative assessment of prostate-specific membrane antigen levels (97) and the identification of collagen binding protein in *Lactobacillus* species (98). Surface plasmon resonance-biomolecular interaction analysis (SPR-BIA) is another chip-based method that has been used for kinetic and thermodynamic studies of protein interactions. SPR-BIA has also been combined with MALDI-TOF to develop an integrated device for the non-destructive study of protein structure and function (93).

Adjunct technologies Microfabrication and microfluidics principles are also being increasingly applied in proteomics for the development of cost-effective, high-throughput strategies (99). Proof-of-concept applications of microfabricated and microfluidic devices for the microdialysis of small volumes of proteins (100), microfluidics-based isoelectric focusing systems (101), and mass spectrometry ionization sources (102) have all been recently reported. Li et al. (90) have also reported the development of a microfabricated capillary electrophoresis/nano-electrospray mass spectrometer that includes capillary electrophoresis channels and an interface to a nanoelectrospray ionization tip. This microfabricated device is connected with an autosampler so that up to 25 samples of tryptic digests can be processed each hour. It should be noted that these concepts have been individually applied to all aspects of proteomics research including sample separation, protein purification, enzymatic digestion, and mass spectrometry. A fully integrated microfabricated proteomics system has not been developed thus far; however, the potential of these approaches for delivering improvements in proteomics work is clearly evident.

SUMMARY AND FUTURE DIRECTIONS

A comprehensive proteomics approach is required for developing an understanding of complex cellular processes and networks. It is evident that a single proteomics technology (2DE, MS, Y2H, or ProteinChip®) is not capable of addressing all the different facets of the field. The central role 2DE has occupied in proteomics has somewhat diminished over the past few years; however, improvements are being made to expand its utility in expression proteomics. Mass spectrometry is rapidly evolving and is being used for a variety of protein characterization, identification, and structural studies. New experimental approaches for determining protein interactions as well as computational methods for predicting protein function are also leading to the development of cell interaction maps necessary to understand cellular processes. Advances also need to be considered in the framework of an integrated strategy where several of these technologies can be interfaced for a single scheme. A good example of this is the development of fluorescent dye stains for

2DE where one of the additional benefits is the compatibility of stained proteins with MS protocols so that sample handling and processing is minimized.

A big thrust in proteomics will be the adaptation of current methods to high-throughput formats and large-scale analysis. Microfabrication and microfluidic technologies are spurring the miniaturization of some proteomics modules with an eye towards this goal. The integration of these modules into a single device will also be a major area of focus in the future. However, proteomics alone cannot provide all the information required for understanding cellular processes. Complementary approaches in genomics and bioinformatics will have to be used together with proteomics so that the maximum benefit can be realized.

ACKNOWLEDGMENTS

The authors thank Drs. Jeffrey Morgan and Charles Roth for their critical comments and helpful discussions, Dr. Xunbao Duan for Figure 3, and Dr. Deanna Thompson and Kristina Roberts for assistance with the illustrations. Financial support from the National Science Foundation (Grant BES 9910186) and use of the Functional Genomics and Proteomics Core Facility at the Shriners Burns Hospital are also acknowledged.

The *Annual Review of Biomedical Engineering* is online at http://bioeng.annualreviews.org

LITERATURE CITED

1. Oliver S. 2000. Guilt-by-association goes global. *Nature* 403:601–3
2. Anderson L, Seilhamer J. 1997. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18:533–37
3. Pandey A, Mann M. 2000. Proteomics to study genes and genomes. *Nature* 405: 837–46
4. Gygi SP, Rochon Y, Franza BR, Aebersold R. 1999. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol.* 19:1720–30
5. Wilkins MR, Sanchez J-C, Williams KL, Hochstrasser DF. 1996. Current challenges and future applications for protein maps and post-translational vector maps in proteome projects. *Electrophoresis* 17: 830–38
6. Banks RE, Dunn MJ, Hochstrasser DF, Sanchez J-C, Blackstock WP, et al. 2000. Proteomics: new perspectives, new biomedical opportunities. *Lancet* 356:1749–56
7. Steiner S, Witzmann FA. 2000. Proteomics: applications and opportunities in preclinical drug development. *Electrophoresis* 21:2099–104
8. Blackstock WP, Weir MP. 1999. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17:121–27
9. Anderson NL, Blasco-Esquer R, Hofmann JP, Anderson NG. 1991. A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies. *Electrophoresis* 12:907–30
10. Anderson NL, Matheson AD, Steiner S. 2000. Proteomics: applications in basic and applied biology. *Curr. Opin. Biotechnol.* 11:408–12

11. Naaby-Hansen S, Waterfield MD, Cramer R. 2001. Proteomics—post-genomic cartography to understand gene function. *Trends Pharm. Sci.* 22:376–84
12. Chambers G, Lawrie L, Cash P, Murray GI. 2000. Proteomics: a new approach to the study of disease. *J. Pathol.* 192:280–88
13. Patterson SD. 2000. Proteomics: the industrialization of protein chemistry. *Curr. Opin. Biotechnol.* 11:413–18
14. Anderson L, Anderson NG. 1977. High resolution two-dimensional electrophoresis of human plasma proteins. *Proc. Natl. Acad. Sci. USA* 74:5421–25
15. Rabilloud T, Chevallet M. 2000. Solubilization of proteins in two-dimensional electrophoresis. See Ref. 104, pp. 9–30
16. Harry JL, Wilkins MR, Herbert BR, Packer NH, Gooley AA, Williams K. 2000. Proteomics: capacity versus utility. *Electrophoresis* 21:1071–81
17. Paulsen II, Sliwinski MK, Nelissen B, Goffeau B, Saier MH Jr. 1998. Unified inventory of established and putative transporters encoded within the complete genome of *Saccharomyces cerevisiae*. *FEBS Lett.* 430:116–25
18. Wallin E, von Heijne G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukarytoic organisms. *Protein Sci.* 7:1029–38
19. Fey SJ, Larsen PM. 2001. 2D or not 2D. *Curr. Opin. Chem Biol.* 5:26–33
20. Cordwell SJ, Nouwens AS, Verrills NM, Basseal DJ, Walsh BJ. 2000. Subproteomics based upon cellular location and relative solubilities in conjunction with composite two-dimensional electrophoresis gels. *Electrophoresis* 21:1094–103
21. Molloy MP, Herbert BR, Walsh BJ, Tyler MI, Traini M, et al. 1998. Extraction of membrane proteins by differential solubilization of proteins for separation using two-dimensional electrophoresis gels. *Electrophoresis* 19:837–44
22. Zuo X, Speicher DW. 2000. A method for global analysis of complex proteomes using sample prefractionation by solution isoelectrofocusing prior to two-dimensional electrophoresis. *Anal. Biochem.* 284:266–78
23. Fountoulakis M, Takacs MF, Berndt P, Langen H, Takacs B. 1999. Enrichment of low abundance proteins of *Escherichia coli* by hydroxyapatite chromatography. *Electrophoresis* 20:2181–95
24. Harrington MJ, Coffman JA, Calzone FJ, Hood LE, Britten RJ, Davidson EH. 1992. Complexity of sea urchin embryo nuclear proteins that contain basic domains. *Proc. Natl. Acad. Sci. USA* 89:6252–56
25. Pasquali C, Fialka I, Huber LA. 1999. Subcellular fractionation, electromigration analysis and mapping of organelles. *J. Chromatogr. B Biomed. Sci. Appl.* 722: 89–102
26. Jung E, Heller M, Sanchez J-C, Hochstrasser DF. 2000. Proteomics meets cell biology: the establishment of subcellular proteomics. *Electrophoresis* 21:3369–77
27. Celis JE, Ostergaard M, Jensen M, Gromova I, Rasmunssen HH, Gromov P. 1998. Human and mouse proteomic databases: novel resources in the protein universe. *FEBS Lett.* 430:64–72
28. Celis JE, Kruhoffer M, Gromova I, Fredriksen C, Ostergaard M, et al. 2000. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* 480:2–16
29. Gygi SP, Rist B, Aebersold R. 2000. Measuring gene expression by quantitative proteome analysis. *Curr. Opin. Biotechnol.* 11:396–401
30. Gygi SP, Aebersold R. 2000. Mass spectrometry and proteomics. *Curr. Opin. Chem. Biol.* 4:489–94
31. Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, et al. 2000. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 21:1037–53

32. Hanash S. 2000. Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. *Electrophoresis* 21:1202–9
33. Butt A, Davison MD, Young JA, Gaskell SJ, Oliver SG, Benyon RJ. 2001. Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis. *Proteomics* 1:42–53
34. Anderson L, Anderson NG. 1998. Proteome and proteomics: new technologies, new concepts. *Electrophoresis* 19:1853–61
35. Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez J-C. 2000. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* 21:1104–15
36. Pradet-Balade B, Boulme F, Beug H, Mullner EW, Garcia-Sanz JA. 2001. Translation control: bridging the gap between genomics and protemics? *Trends Biochem. Sci.* 26:225–29
37. Matsui NM, Smith-Beckerman DM, Epstein LB. 1999. Staining of preparative 2-D gels: coomassie blue and imidazole-zinc negative staining. See Ref. 103.
38. Rabilloud T. 1999. Silver staining of 2-D electrophoresis gels. See Ref. 103, pp. 297–306
39. Appel RD, Palagi PM, Walther D, Vargas JR, Sanchez J-C, et al. 1997. Melanie II—a third generation software package for analysis of two-dimensional electrophoresis images: I. Features and user interface. *Electrophoresis* 18:2724–34
40. Appel RD, Vargas JR, Palagi PM, Walther D, Hochstrasser DF. 1997. Melanie II—a third generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. *Electrophoresis* 18:2735–48
41. Sinha P, Hutter G, Kottgen E, Dietel M, Schadendorf D, Lage H. 1998. Increased expression of annexin I and thioredoxin detected by two-dimensional gel electrophoresis of drug resistant human stomach cancer cells. *J. Biochem. Biophys. Methods* 37:105–16
42. Smilansky Z. 2001. Automatic registration for images of two-dimensional protein gels. *Electrophoresis* 22:1616–26
43. Keough T, Lacey MP, Fieno AM, Grant RA, Sun Y, et al. 2000. Tandem mass spectrometry methods for definitive protein identification in proteomics research. *Electrophoresis* 21:2252–65
44. Lahm H-W, Langen H. 2000. Mass spectrometry: a tool for the identification of proteins separated by gels. *Electrophoresis* 21:2105–14
45. Yates JR III. 1998. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* 33:1–19
46. Mann M, Hendrickson RH, Pandey A. 2001. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* 70:437–73
47. Yates JR III. 2000. Mass spectrometry: from genomics to proteomics. *Trends Genet.* 16:5–8
48. Karas M, Hillencamp F. 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* 60:2299–301
49. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64–71
50. McDonald WH, Yates JR III. 2000. Proteomic tools for cell biology. *Traffic* 1: 747–54
51. Corthals GL, Gygi SP, Aebersold R, Patterson SD. 2000. Identification of proteins by mass spectrometry. See Ref. 104, pp. 196–231
52. Henzel WJ, Billeci T, Stults JT, Wong SC. 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* 90:5011–15
53. Yates III JR. 1998. Database searching using mass spectrometry data. *Electrophoresis* 19:893–900

54. Pappin DJ, Hojrup P, Bleasby AJ. 1993. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* 3: 327–32
55. Kuster B, Mortensen P, Andersen JS, Mann M. 2001. Mass spectrometry allows direct identification of proteins in large genomes. *Proteomics* 1:641–50
56. Eng J, McCormack AL, Yates JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5:976–89
57. Quadroni M, James P. 1999. Proteomics and automation. *Electrophoresis* 20:664–77
58. Qin J, Fenyo D, Zhao Y, Hall WW, Chao DM, et al. 1997. A strategy for rapid, high-confidence protein identification. *Anal. Chem.* 69:3995–4001
59. Clauser KR, Baker P, Burlingame AL. 1999. Role of accurate mass measurement (+/−10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Biochem.* 71: 2871–82
60. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–67
61. Spengler B, Lutzenkirchen F, Kaufmann R. 1993. On-target deuteration for peptide sequencing by laser mass spectrometry. *Org. Mass Spectrom.* 28:1482–90
62. Binz PA, Muller M, Walther D, Bienvenut WV, Gras R, et al. 1999. A molecular scanner to automate proteomic research and display proteome images. *Anal. Biochem.* 71:4981–88
63. Bienvenut WV, Sanchez J-C, Karmime A, Rouge V, Rose K, et al. 1999. Toward a clinical molecular scanner for proteome research: parallel protein chemical processing before and during western blot. *Anal. Biochem.* 71:4800–7
64. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. 1999. Quantitative analysis of protein mixtures using isotope coded affinity tags. *Nat. Biotechnol.* 17:994–99
65. Moult J, Melamud E. 2000. From fold to function. *Curr. Opin. Struct. Biol.* 10:384–89
66. Sali A. 1999. Functional links between proteins. *Nature* 402:23–26
67. Figeys D, Pinto D. 2001. Proteomics on a chip: promising developments. *Electrophoresis* 22:208–16
68. Bartel PL, Fields S. 1997. *The Yeast Two Hybrid System*. Oxford, UK: Oxford Univ. Press
69. Legrain P, Selig L. 2000. Genome-wide protein interaction maps using two-hybrid systems. *FEBS Lett.* 480:32–36
70. Stephens DJ, Banting G. 2000. The use of yeast two-hybrid screens in studies of protein: protein interactions involved in trafficking. *Traffic* 1:763–68
71. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, et al. 2000. Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between yeast proteins. *Proc. Natl. Acad. Sci. USA* 97:1143–47
72. Janoueix-Lerosey I, Jollivet F, Camonis J, Marche PN, Goud B. 1995. Two-hybrid system screen with the small GTP-binding protein Rab6. Identification of a novel mouse GDP dissociation inhibitor and two other potential partners of Rab6. *J. Biol. Chem.* 273:31401–7
73. Uetz P, Glot L, Cagney G, Mansfield TA, Judon RS, et al. 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623–27
74. Toby GG, Golemis EA. 2001. Using the yeast interaction trap and other two-hybrid-based approaches to study protein-protein interactions. *Methods* 24:201–17
75. Neubauer G, King A, Rappsilber J, Calvio C, Watson M, et al. 1998. Mass spectrometry and EST-database searching allows

characterization of the multi-protein spliceosome complex. *Nat. Genet.* 20:46–50
76. Blobel G, Wozniak RW. 2000. Proteomics for the pore. *Nature.* 403:835–36
77. Marcotte EM. 2000. Computational genetics: finding protein function by nonhomology methods. *Curr. Opin. Struct. Biol.* 10:359–65
78. Huynen M, Snel B, Lathe W, Bork P. 2000. Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res.* 10:1204–10
79. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. 1997. Correlated mutations contain information about protein-protein interactions. *J. Mol. Biol.* 271:511–23
80. Enright AJ, Iliopoulos I, Krypides NC, Ouzonis CA. 1999. Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402:86–90
81. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO. 2000. Protein function in the post-genomic era. *Nature* 405:823–26
82. Kini RM, Evans HJ. 1996. Prediction of potential protein-protein interaction sites from amino acid sequence. Identification of a fibrin polymerization site. *FEBS Lett.* 385:81–86
83. Aloy P, Querol E, Aviles FX, Sternberg MJ. 2001. Automated structure-based prediction of functional sites in proteins: applications to assessing the validity of inheriting protein function from homology in genome annotation to protein docking. *J. Mol. Biol.* 311:395–408
84. Jones S, Thornton JM. 1997. Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.* 272:133–43
85. Bock JR, Dough DA. 2001. Predicting protein-protein interactions from primary structure. *Bioinformatics* 17:455–60
86. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. 1999. Assigning protein function by comparitive genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. USA* 96:4285–88
87. Marcotte EM, Pellegrini M, Ng H-L, Rice DW, Yeates TO, Eisenberg D. 1999. Detecting protein function and protein-protein interactions from genome sequences. *Science* 285:751–53
88. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. 1999. The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci. USA* 96:2896–901
89. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. 1999. A combined algorithm for genome-wide prediction of protein function. *Nature* 402:83–86
90. Li J, Tremblay T-L, Wang C, Attiya S, Harrison JD, Thibault P. 2001. Integrated system for high-throughput protein identification using a microfabricated device coupled to capillary electrophoresis/nanospray mass spectrometry. *Proteomics* 1: 975–86
91. Fung ET, Thulasiraman V, Weinberger SR, Dalmasso EA. 2001. Protein biochips for differential profiling. *Curr. Opin. Biotechnol.* 12:65–69
92. Folch A, Toner M. 2000. Microengineering of cellular interactions. *Annu. Rev. Biomed. Eng.* 2:227–56
93. Nelson RW, Nedelkov D, Tubbs KA. 2000. Biosensor chip mass spectrometry: a chip-based proteomics approach. *Electrophoresis* 21:1155–63
94. Merchant M, Weinberger SR. 2000. Recent advances in surface-enhanced laser desorption/ionization-time of flight mass spectrometry. *Electrophoresis* 21:1164–77
95. de Wildt RMT, Mundy CR, Gorick BD, Tomilson IM. 2000. Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat. Biotechnol.* 18: 989–94
96. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Scraml P, et al. 1998. Tissue microarrays for high-throughput molecular profiling of tumour specimens. *Nat. Med.* 4:844–47
97. Xiao Z, Jiang X, Beckett ML, Wright

GL Jr. 2000. Generation of a baculovirus recombinant prostate-specific membrane antigen and its use in the development of a novel protein biochip quantitative immunoassay. *Protein Expr. Purif.* 19:12–21

98. Howard JC, Heinemann C, Thatcher BJ, Martin B, Gan BS, Reid G. 2000. Identification of collagen-binding proteins in *Lactobacillus* spp. with surface-enhanced laser desorption/ionization-time of flight ProteinChip technology. *Appl. Environ. Microbiol.* 66:4396–400
99. Lee KH. 2001. Proteomics: a technology-driven and technology-limited discovery science. *Trends Biotechnol.* 19:217–22
100. Xu N, Lin Y, Hofstadler SA, Matson D, Call CJ, Smith RD. 1998. A microfabricated dialysis device for sample cleanup in electrospray ionization mass spectrometry. *Anal. Chem.* 70:3553–56
101. Hofmann O, Che D, Cruickshank KA, Muller UR. 1999. Adaptation of capillary isoelectric focusing to microchannels on a glass chip. *Anal. Chem.* 71:678–686
102. Licklider L, Wang XQ, Desai A, Tai YC, Lee TD. 2000. A micromachined chip-based electrospray source for mass spectrometry. *Anal. Chem.* 72:367–75
103. Link AJ, ed. 1999. *2-D Proteome Analysis Protocols*. Totowa, NJ: Humana
104. Rabilloud T, ed. 2000. *Two-Dimensional Gel Electrophoresis and Identification Methods*. New York, NY: Springer