1

PROTEOMICS: THE BASIC OVERVIEW

Daniel Figeys

Department of Biochemistry, Microbiology, and Immunology, University of Ottawa, Ottawa, Canada

Industrial Proteomics: Applications for Biotechnology and Pharmaceuticals, edited by Daniel Figeys
ISBN 0-471-45714-0

List of Abbreviations

2D	Two dimensional
β-gal	β-galactosidase
bp	base pair
CZE	capillary zone electrophoresis
ESI	electrospray ionization
EST	expressed sequence tag
FTMS	Fourier transform mass spectrometer
HPLC	high-performance liquid chromatography
i.d.	inner diameter
ICAT	isotope-coded affinity tag
IPG	immobilized pH gradient
MALDI	matrix-assisted laser desorption ionization
MHC	major histocompatibility complex
Micro-ESI	microelectrospray ionization
MS/MS	tandem mass spectrum
MS	mass spectr(um/ometer)
MW	molecular weight
Nano-ESI	nanoelectrospray ionization
o.d.	outer diameter
ORF	open reading frame
ppm	parts per million
psi	pound per square inch
PVP-40	polyvinylpyrrolidone
RF	radio frequency
SAGE	serial analysis of gene expression
SNP	single-nucleotide polymorphism
SPE	solid-phase extraction
TOF	time of flight

INTRODUCTION

The astonishing pace of scientific discovery in the twentieth century has had direct implications on our daily lives. The increase in quality of life and in life expectancy

Figure 1.1. Life expectancy in the United States from 1900 to 1999. The data were averaged according to gender and race and smoothed for a 3-year period. Data were obtained from the *National Vital Statistics Report*, Vol. 47, No. 28, December 13, 1999. US Department of Health and Human Services.

seen in the industrialized countries correlate with the discoveries made in medicine and biological sciences (Fig. 1.1). At the onset of the twentieth century, medicine and biology were descriptive sciences aimed at understanding macroscopic phenomena. By the end of the century, these sciences had accessed the micro and nanoscopic worlds with an increased understanding of the implications involved in the biological processes of our lives. The dawn of the new millennium is ushering in a new paradigm for understanding system biology and complex multigenic diseases, as well as the accumulation of massive amounts of genomic information.

In the last quarter of the twentieth century, a few visionaries realized that all the parts were available to create a deoxyribonucleic acid (DNA) sequencing engine that would become capable of sequencing the human genome. Numerous sequencing projects were started with the promise of creating technology to accelerate the pace of sequencing. It took longer than expected to fill this promise, but by the end of the twentieth century new technologies were available to accelerate the pace of sequencing (Fig. 1.2).

In particular, instrumentation was developed to sequence large genomes in reasonable time periods (Dovichi, 1997). This instrumentation provided the framework for developing comprehensive approaches to heighten our understanding of diseases and biology. Furthermore, it was quickly realized that the complexity of biological processes could not be resolved only with genomic sequencing. It became apparent that the diseases remaining to be cured were increasingly more complex, were often trig-

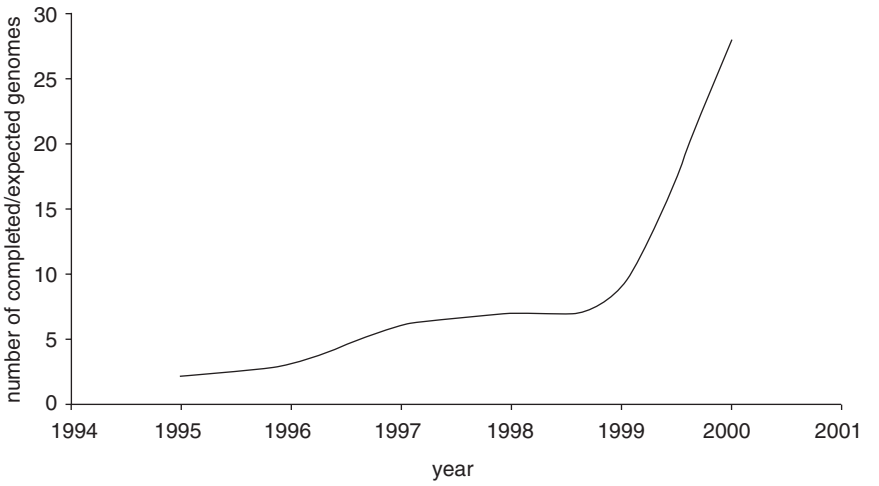


Figure 1.2. Number of completed and expected to be completed genome sequencing projects from 1995 to 2004. The intensification of sequencing efforts is visible from 1999 forward.

gered by a series of genes, and were dependant on the genetic makeup of individuals. Genomic sequencing can rapidly provide the list of the parts of the genome, but it cannot provide the instruction on how these parts fit together. This insufficiency created a disproportion between the amount of information available and the tools available for large-scale studies of molecules involved in biological processes.

Conventional biochemical approaches were relied on to access the functions and interactions of translated material. The key to capturing the value of the human genome is to find approaches that can rapidly link the genomic information to drug discovery and diagnostics. The development, however, of high-throughput tools for assessing the expression levels and functions of transcripted and translated genetic materials has globalized the study of cellular processes that may provide the link between genomic information and drug discovery and diagnostics.

Deoxyribonucleic Acid to Ribonucleic Acid

A series of tools became available for the rapid and quantitative analysis of expression levels in ribonucleic acid (RNA). For example, the emergence of the DNA/RNA array technology and the serial analysis of gene expression (SAGE) have facilitated the acquisition of quantitative expression profiles for complete sets or subsets of RNA (Desprez et al., 1998; Marshall and Hodgson, 1998; Ruan et al., 1998; Service, 1998; Velculescu et al., 1995; Madden et al., 1997; Matsumura et al., 1999, Neilson et al., 2000; Lal et al., 1999; Stein et al., 2004; Weeraratna et al., 2004). Different RNA expression profiles can be acquired using these techniques, and genes that are “up”- or “down”-regulated when comparing different cells or different cell states can also be detected.

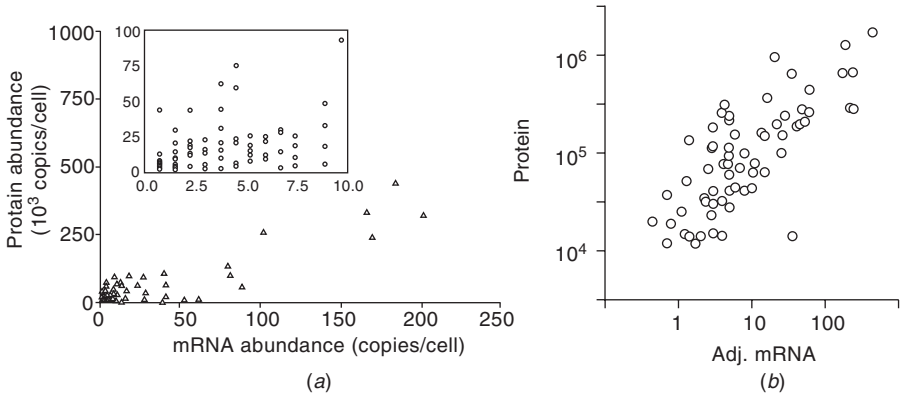


Figure 1.3. Correlation between protein and mRNA levels in *S. Cerevisiea*. (*a*) Correlation between protein and mRNA levels for 106 genes in yeast. The inset shows the low-end portion of the main figure. The Pearson product moment correlation for the entire data set was 0.935 and 0.356 for the inset. Reproduced with permission from Gygi et al. (1999). (*b*) Correlation of protein abundance with adjusted mRNA abundance in yeast. The Pearson product moment is 0.76. Reproduced with permission from Futcher et al. (1999).

Already, with one step away from the genomic sequence, a tremendous amount of information is extracted by studying RNA.

It is clear that these high-throughput DNA/RNA screening technologies provide a rapid and quantitative overview of the genes that are differentially expressed. It turns out, however, that the RNA expression on its own is not sufficient for understanding biological processes and gene functions. Evidence has been collected from different research groups (Gygi et al., 1999b; Gygi and Aebersold, 1999; Anderson and Seilhamer, 1997; Futcher et al., 1999) indicating that the expression of RNA has a poor linear relationship with the changes happening at the protein level (Fig. 1.3) while other experiments indicate a good correlation for higher abundance proteins (Kern et al., 2003). This is due to the numerous regulation mechanisms in place during protein expression and postexpression.

Furthermore, a recent large-scale study of the yeast genome has clearly demonstrated a poor relationship between DNA chip results and protein expressions (Ross-Macdonald et al., 1999). This study, performed by transposon tagging and gene disruption, found 31 meiotic genes, which were detected at the protein level by "in-frame" *lacZ* fusion and assay for β-galactosidase (β-gal) activity. Out of the 31 meotic genes, only 17 had been previously reported to be induced by at least twofold during sporulation. This was achieved by detection based on the DNA microarray representing all the annotated open reading frames (ORFs) in yeast. For the remainder of the meotic genes, the DNA microarray analysis failed to find any significant induction during meiosis. Therefore, not only is the relationship between the expression level of RNA and the expression level of proteins a complex matter, but it is also misleading

to rely solely on the RNA expression patterns to predict cellular functions. Clearly, the amount of protein related to a gene can dramatically change without any change in the RNA expression level if this is achieved through downstream regulation mechanisms.

Ribonucleic Acid to Protein

Although serious efforts have been made to develop efficient genomic technologies, the study of proteins cannot be avoided in the quest to understand biological processes. The justification for studying proteins goes even deeper than just the lack of correlation between the expression levels of RNA and proteins. The presence of posttranslational modification, posttranslational truncation of proteins, and protein–ligand interactions are a few examples that illustrate the complexity at a protein level.

The study of proteins has always been done on a relatively small scale, partly because of the lack of methods to unambiguously and easily verify the protein identity. Experiments had to be carried out with great care to ensure that only the protein of interest was isolated. All of this has changed over the last 10 years with the development of technology capable of performing large-scale analyses and identification of proteins (Issaq et al., 2002; Wang and Hanash, 2003). This achievement has opened the door for comprehensive studies of proteins related to a genome (proteome) (Wilkins et al., 1996a).

PROTEOME HANDLING IN CLASSICAL AND FUNCTIONAL PROTEOMICS

Two-dimensional (2D) gel electrophoresis is typically used in profiling proteomic studies, and its most popular implementation is the differential displays of proteins expressed under different conditions. It turns out that conserving the proteome to obtain a truly representative 2D gel pattern is not trivial, and this is probably the most important experimental step in proteomic studies. For example, great care must be taken during the extraction of cells from their environment and during cell lysis to reduce the influence of the sample extraction protocol on the observed state of the proteome. Mistakes are often made while manipulating a proteome, thus seriously affecting the conclusions from the experiments. Therefore, the history of the sample is a prerequisite in order to assess the validity of a sample.

PROTEIN PURIFICATION

The extraction of protein from a cell lysate is a critical step for establishing a stable proteome. It is well known that once the cells are lysed the enzymes that would normally be compartmentalized are brought in contact with other proteins and then rapidly degrade the proteome. Fortunately, sets of well-characterized methods are available to

cover the majority of the needs in protein extraction from cells. The harvesting of soluble proteins is simply performed by lysing the cells and collecting the supernatant. Different cell lysis methods are easily accessible (see *www.expasy.ch*). It is best to choose the simplest approach that is directly compatible with the immobilized pH gradient (IPG) isoelectric (i.e., minimum salt and ionic surfactant contents).

The current protein extraction protocols are not universally applicable to all biological samples, and they have limitations in terms of the protein representation in the extracted proteome. First, hydrophobic proteins are not easily extracted and represented on a 2D gel pattern (Wilkins et al., 1998). Figure 1.4 shows the number of proteins visualized on a 2D gel of *Saccharomyces Cerevisiae* versus the gravy hydrophobicity scale. Clearly, a significant portion of the predicted proteins in a proteome is hydropho-

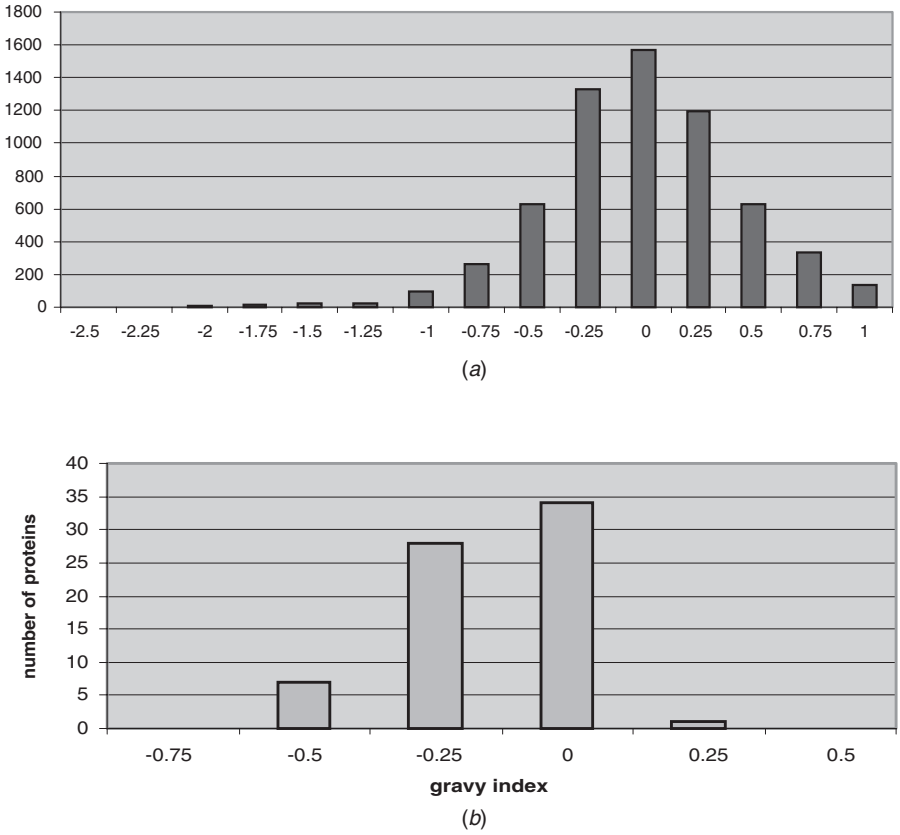


Figure 1.4. Hydrophobicity vs. protein observed. (*a*) The gravy index was calculated for all the known yeast ORF. (*b*) The gravy index was calculated from some of the observed proteins on a 2D gel electropherogram. Data were compiled from Gygi et al. (2000) and Perrot et al. (1999). A negative gravy index indicates hydrophilicity while a positive gravy index indicates hydrophobicity.

bic in nature; however, the compilation of the proteins that have been identified by 2D gel electrophoresis and mass spectrometry indicate a serious lack of hydrophobic proteins. This means that 2D gel electrophoresis fails to display the hydrophobic portion of the proteome. In recent years, significant efforts from the group of Rabilloud (Adessi et al., 1997; Blisnick et al., 1998; Chevallet et al., 1998; Goldberg et al., 1996; Rabillout, 1998; Rabillout et al., 1999; Santoni et al., 1999; Tastet et al., 2003; Luche et al., 2003) provided improved protocols and chemicals for the recovery of hydrophobic proteins. The extraction of hydrophobic proteins, however, is still a considerable challenge in proteomics. Comprehensive reviews have been published on the extraction and solubilization of proteins from biological samples for 2D gel electrophoresis (Dunn and Corbett, 1996; Rabillout, 1996, 2002; Ramagli, 1999).

Current protein extraction protocols often fail to provide a unique proteome. This is particularly a problem when dealing with tissues that are composed of many different cells at different stages and each having its own proteome. Classical protein extraction techniques would lyse the whole tissue or tissue fraction generating a scrambled proteome composed of all the original proteomes. Recent proteomic studies have clearly indicated that the cellular diversity within a biological sample, and the protein localization within cells greatly impact the conclusion of a proteomic study. The present challenges related to the level of cellular mass complexity are presented below.

Cell Culture

Cultured cells have been the most popular source of protein for proteomic studies. They are mainly used because they can provide a wealth of information that is often difficult to obtain from tissues and primary cultures. Furthermore, cell cultures are easily accessible at large curated collections of cells, such as the American Type Culture Collection (*www.atcc.org*).

Another significant advantage of cell culture over tissue material is the controllability of the environment and the growth conditions, as well as the potential to use different stimulations. Furthermore, cultures are usually obtained from single colonies and are homogeneous, thus greatly reducing the complexity of the proteome. Moreover, the culture growth and the lysis conditions are also controllable, which significantly minimize batch-to-batch fluctuations and artifacts in the proteome. To date, cell cultures provide the most controllable environment to perform proteomic experiments.

Other Source of Samples

Proteomes have also been analyzed from more complex samples, such as clinical samples, plants, and animals (Ostergaard et al., 1997; Wimmer et al., 1996; Lubec et al., 2003; Aebersold and Mann, 2003). The conclusions, however, that can be reached from such samples are often affected by many parameters that are not easily controllable. Typically, large numbers of samples need to be studied for any hope of finding the relevant proteins. Many factors increase the complexity of the analysis. The first

factor is the methodology used to extract and store the clinical samples. The failure of clinical samples in proteomics is often traced back to the first step in the study, that is, the extraction and storing of the clinical samples. The proteome can often be seriously changed due to extended storage at room temperature. Even though large collections of clinical samples have been established, they are often of limited value for proteomic studies because they were not stored properly shortly after their extraction. Proteomic studies often require the revision of methodology for clinical and other samples.

Once proper methodologies are in place to limit the effects of sample isolation on the proteome, other factors can still make the study difficult. Tissue samples are heterogeneous in nature (i.e., composed of different cell types, volumes, and cell ratio compositions). This can cause large variations in the protein contents within the same tissues and for tissues of different sources. In reality, tissue samples are composed of many different proteomes. The lysis of these samples generates a scrambled mixture of proteomes. Therefore, multiple experiments need to be performed in order to extract the relevant proteins, or else the scrambling of proteomes will shadow the important proteins, thus making it impossible to interpret the results.

Examples are available that demonstrate the feasibility of extracting valuable information from scrambled proteomes. In a proteomic study of bladder cancer, Ostergaard and co-workers (Ostergaard et al., 1997; Wimmer et al., 1996; Celis et al., 1999) have demonstrated that new disease markers associated with the different cancer stages can be determined. They processed hundreds of well-preserved bladder cancer. By applying a proteomic approach, they found a handful of proteins that were differentially expressed during the different progression stages of the disease. Antibodies were raised against these protein markers (mainly keratinocyte markers) and were then used as a diagnostic tool to ascertain the different cancer stages within cryostat sections of biopsies from bladder cystectomies. Although the study was successful, the identified disease markers were actually high abundance proteins, and their changes in expression would be noticeable in scrambled proteomes. In any case, it would have been significantly more difficult to identify low- to mid-abundance proteins involved in the disease.

Two-dimensional gel-based analysis also introduces a third limiting factor that is related to the sample capacity of the technique (See Table 1.1). Typically, up to 100 µg of proteins are used during 2D gel electrophoresis. It is known that 2D gel electrophoresis can at least separate up to 11,000 proteins. That would represent about 9 ng for every protein, which is in the low subfemtomole level for most proteins displayed by 2D gel electrophoresis. In most cells, however, about 10 percent of the proteins represent 90 percent of the protein mass, while the remaining 90 percent of proteins represent only 10 percent of the protein mass. Assuming that cells express about 10,000 proteins, then the mid- to low-abundance proteins would only be, in the best case scenario, at about 1 ng on the gel. In reality the number would be even less because of the sample losses associated with 2D gels and low-abundance proteins. In a recent study, Gygi et al. (2000) demonstrated that for a typical 2D gel of yeast, protein identification could not be achieved for proteins with a codon bias of less than 0.1, thus indicating the lack of identification of low-abundance proteins. Furthermore, they demonstrated that they could discover low- to mid-abundance proteins on a 1D gel, but

TABLE 1.1. Effect of Different Cell Composition and Protein Concentration on the Level of Protein Observed

	[ ] in Cell *X*	[ ] in Cell *Y*	% of Cell That Are *X*				
			10	30	50	70	90
			Observed protein concentration				
Protein 1	1	1	1	1	1	1	1
Protein 2	2	1	1.1	1.3	1.5	1.7	1.9
Protein 3	5	1	1.4	2.2	3	3.8	4.6
Protein 4	10	1	1.9	3.7	5.5	7.3	9.1
Protein 5	100	1	10.9	30.7	50.5	70.3	90.1

only once the sample load had been significantly increased. The difference in sensitivity of different proteomics approached applied to yeast was recently illustrated (Ghaemmaghami et al., 2003).

The technique of 2D gel separation of proteins isolated from complex tissue samples has been performed for over 30 years, and recently, tools have been developed to allow the identification of these proteins. These tools have now been integrated in high-throughput platforms to allow serious study of proteomes. As proteomics moves from cell culture to primary cultures and tissues, it is important to keep in mind that these samples when used with conventional solubilization approaches produce complex mixtures of proteomes, each corresponding to an individual subpopulation of cells. The resulting scrambled proteome can be very difficult to make any sense of even with large population studies.

Subpopulation of Cells

The fractionation of subpopulation of cells from tissues can reduce the complexity of the proteome. Without cellular fractions it is very likely that the information generated will be meaningless. This is clearly illustrated in Table 1.1. The different cell compositions drastically affect the observed protein levels. Interestingly, the proteins that are not significantly different in concentration are not scrambled, while the ones that are significantly different are seriously scrambled.

Thus, it is important to know the cellular composition of the tissue being handled and if necessary to use fractionation of subpopulations of cells. Laser capture microdissection is increasingly being used in proteomics studies (Banks et al., 1999; Sirivatanauksorn et al., 1999; Emmert-Buck et al., 2000; Ornstein et al., 2000). As well, immunomagnetic techniques using immobilized antibodies specific to certain cell types have been investigated to separate cellular populations for proteomic studies (Clarke et al., 1994; Gomm et al., 1995). For example, Page et al. (1999) utilized the technique of immunomagnetic beads to separate cells based on immobilized antibodies that bind to known cell surface markers and followed by a magnetic pull down of the beads.

TABLE 1.2. Cell Culture and Estimated Protein Levels[a]

Cell Culture	Copy/Cell	Number of Mole	Nanogram Expressed for a Protein of 70 kDa	Total Mass of Protein[b]
10^7	10	0.16 fmol	0.01	6.5 µg
	100	1.6 fmol	0.1	
	1,000	16 fmol	1	
	10,000	160 fmol	10	
10^8	10	1.6 fmol	0.1	65 µg
	100	16 fmol	1	
	1,000	160 fmol	10	
	10,000	1.6 pmol	100	
10^9	10	16 fmol	1	650 µg
	100	160 fmol	10	
	1,000	1.6 pmol	100	
	10,000	16 pmol	1,000	

[a] Gray area indicates level of proteins that can be detected on a silver-stained 2D gel.
[b] Assumption 1: 90% of the protein mass is representing 10% of proteins highly expressed (>1000 copy/cell). About 10,000 proteins are expressed.
Assumption 2: 60% of the protein mass is representing 10% of proteins. About 10,000 proteins are expressed.

Subcellular Components

The complexity of specific proteome can also be reduced to the subcellular level. The enrichment of subcellular components, such as the organelles and plasma membrane, reduces the complexity of the proteome while increasing the likelihood of observing lower abundance proteins (Howell et al., 1989; Jung et al., 2000; Brunet et al., 2003). We have come to realize that in many instances it would not be possible to distinguish compartment proteins using conventional whole-cell protein displays by 2D gel electrophoresis due to dynamic range limitation.

Free-flow electrophoresis and density gradients are the classical approaches to subcellular fractionations. An interesting example of the utilization of density gradient separation in proteomics was provided by Fialka et al. (1997). They isolated subcellular compartments from murine mammary epithelial cells (EpH4) by continuous sucrose gradient centrifugation. This was then followed by high-resolution 2D gel electrophoresis on the proteins recovered from the sucrose gradient centrifugation. They were able to obtain 2D gel electropherograms of the late endosomes, early endosomes, and the majority of the rough endoplasmic reticulum. This study did not utilize mass spectrometry to systematically identify the isolated proteins.

Garin et al. (2001) utilized latex beads that are incubated with mouse macrophage-like cell line generating phagosomes that contain these latex beads. The phagosome can then be easily purified using a simple flotation approach. The proteins contained in the phagosome were then separated by 2D gel electrophoresis and the isolated proteins analyzed by mass spectrometry. The end result was the discovery of over 140 proteins associated with the latex bead containing phagosomes.

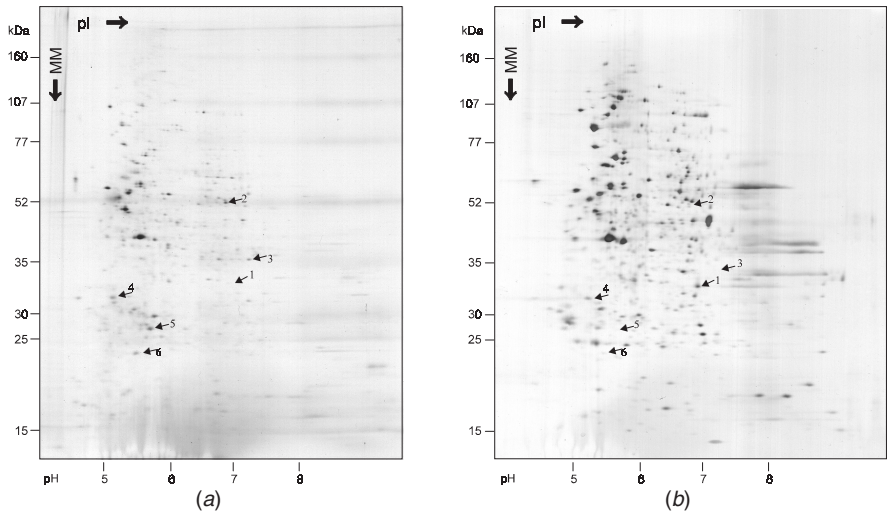


Figure 1.5. Subcellular localization: Magnetic purification. Tetanus-toxin-coated microbeads were incubated with U937 cells for 15 min at 37°C with chased 2 h. Then, the compartments that contained the magnetic beads were magnetically isolated. Two-dimensional gel electropherograms were obtained for the magnetic fraction (*a*) and for the whole-cell lysate. Reproduced with permission from Perrin-Cocon et al. (1999).

An alternative approach was also presented for the recovery of specific compartments based on immunomagnetic purification (Perrin-Cocon et al., 1999; Sarto et al., 2002; Himeda et al., 2004). In their specific application, Perrin-Cocon et al. (1999) isolated intracellular compartments containing endocytosed antigens. The magnetic beads were first covalently attached to a tetanus toxin. The beads were then incubated with U937 cells, allowing the antigen to bind to its receptor. The internalization of the microbeads by the cell was achieved by means of pinocytosis or receptor-mediated endocytosis. The compartments containing the beads were then isolated by the application of a magnetic field. After extensive washes, they were lysed, and their protein content was displayed by 2D gel electrophoresis. Figure 1.5 shows the 2D gel electropherogram obtained for the magnetic purified fraction and for the postnuclear supernatant. About 20 different proteins appeared to be enriched in the isolated compartments.

PROTEIN SEPARATION BY TWO-DIMENSIONAL GEL ELECTROPHORESIS

After carefully assessing the best method to extract the proteome from its medium, the second challenge is to extract information from the proteome, which generally starts

by separating the proteins contained in the proteome. Some proteomes are relatively small, containing a few hundred proteins or less, while others contain thousands of proteins. One-dimensional (1D) separation techniques do not have the resolving power to separate complex mixtures. The combination, however, of orthogonal separation techniques can provide the required resolving power. Furthermore, it is important that the quantitative aspect of the proteome be conserved through the separation technique. About 30 years ago, a technique called 2D gel electrophoresis was introduced, and this technique satisfied the resolving power requirement while conserving the quantitative aspect of the proteome. Two-dimensional gel electrophoresis has been the method of choice for the large-scale purification of proteins in proteomic studies. The 2D gel electrophoresis method can potentially separate several thousand proteins in a single experiment (Gorg et al., 1988; Klose and Kobalz, 1995). Although the predictions for the number of genes in some genomes are high, it is generally believed that the number of genes expressed is, on average, between 5000 and 15,000 per cell type. However, these genes can lead to many forms of proteins, greatly increasing the complexity of the proteome. Protein separation can now be achieved as low as 0.1 isoelectric point (pI) unit and 1 kDa in molecular weight (MW).

Principles of Two-Dimensional Gel Electrophoretic Separation

Most people are not aware that 2D gel electrophoretic separation can be performed in different modes. In proteomics, the overwhelmingly popular implementation of 2D gel electrophoresis is the separation of the proteins according to their pI in the first dimension, followed by their separation according to the MW in the second dimension. Clearly, the combination of pI and MW separation offers a truly orthogonal separation technique, which is reflected by the unsurpassed resolving power of 2D gel electrophoresis.

Although 2D gel electrophoresis is a powerful separation technique, it was also initially very tedious and irreproducible. In fact, it was not clear in the late 1970s and early 1980s if the technique would become widely applied and if it would survive. Fortunately, the problems related with reproducibility were greatly reduced in the mid-1980s when the immobilized pH gradient was introduced, and it became commercially available (IPG strip) (Righetti and Gianazza, 1987; Righetti and Bossi, 1997a, 1997b; Righetti et al., 1983; Gorg, 1993; Fichmann, 1999; Matsui et al., 1999b; Sanchez et al., 1999). In the mid-1980s it was realized that the pH gradient could be immobilized and stabilized by copolymerizing different acrylamide monomers carrying ampholyte properties with acrylamide and low levels of cross-linkers. Although the mixture of ampholytes can be relatively complex, a simple system and a computer algorithm were used to, respectively, cast the IPG strip and to accurately predict the pH profile across the strip. IPG strips with various pH gradients and integrated instruments to perform isoelectric focusing became commercially available. The tediousness, however, involved in the technique is still present today, and, although the technique is definitely applied in proteomic groups, it has not been fully implemented in daily biological experiments.

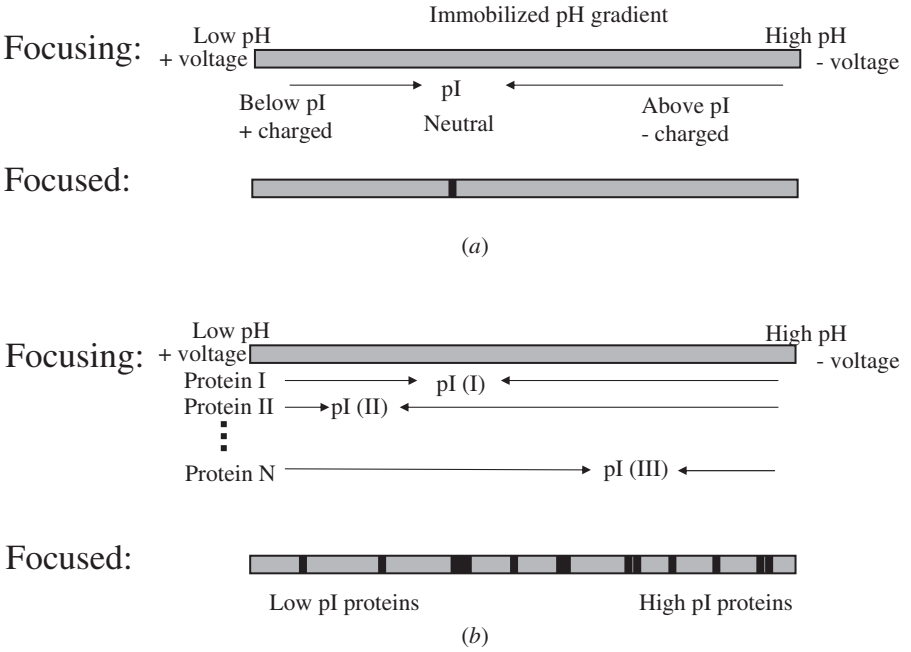


Figure 1.6. Isoelectric focusing on an immobilized pH gradient (IPG). (*a*) Focusing of a single protein. (*b*) Focusing of a mixture of proteins.

The commercial IPG strips are typically available in a dried form, sandwiched between two polymer cover sheets. The first step in using these strips is to remove one of the polymer covers and to inspect the strip for uniformity of the IPG. The peeling of sections of the strips has caused a higher rate of IPG strip rejection in recent years. Once a suitable IPG strip is found, it is first reswelled using a mixture of the protein extract with a loading buffer (Sanchez et al., 1999). This ensures that the protein mixture is uniformly distributed across the whole IPG and limits precipitation due to the excess of proteins in specific areas of the gel. The addition of a small electric field across the strip improves the transfer of the proteins to the IPG strip. Once the strip is reswelled, then an increasingly larger electric field is gradually applied across the IPG (Righetti and Gianazza, 1987) (Fig. 1.6). The proteins that are positively charged (i.e., in the area of the strip with pH below their pI) move toward the cathode and encounter an increasing pH until reaching their pI, at which point they will be neutral. The proteins that are negatively charged (i.e., in the area of the strip with pH above their pI) will move toward the anode and encounter a decreasing pH until reaching their pI. Every protein is concentrated and constantly focused at their respective pI. In theory, the higher the electric field and the longer the focusing is performed, the better the separation. Practically, the power supply, the presence of salts in the sample, and electroosmotic pumping will limit the performance of the separation. We have found that a separation involving 50,000 to 75,000 Vt × hour is sufficient for the types of samples we have handled. These

numbers were typically achieved after 12 to 20 h of focusing. It is important to remember that these numbers need to be adjusted for different proteomes and to keep the time frame of the experiment reasonable.

The second dimension for the 2D separation is typically prepared during the focusing of the first dimension; however, it can also be precast and stored in a fridge with an appropriate buffer for 2 to 3 weeks with no separation problems. The second dimension is generally formed by pouring and polymerizing an acrylamide solution in between two glass plates, spaced using 1- to 1.5-mm spacers. After proper equilibration of the IPG strip performed by in-gel reduction and alkylation of the proteins, the strip is applied to the second dimension. Different systems are available for the second dimension. Some systems provide gels that are slightly larger than a sheet of paper or even larger. The strip can also be cut in different parts and applied to minigels. Regardless of the size of the gel, it is necessary that a tight contact between the strip and the gel be maintained. An electric field is then applied across the gel and the proteins migrate into the second dimension where they are separated according to their MW. Again, systems are commercially available to run the second dimension, and precast, larger gels were recently introduced.

Detection of Proteins Separated by Two-Dimensional Gel Electrophoresis

The ability to detect proteins separated by 2D gel electrophoresis is crucial to its application in proteomics. Over the years different methods have been developed to visualize proteins separated by gel electrophoresis (Rabilloud, 2000) (Fig. 1.7). Although different chemistries are used, the methodology involves either labeling the proteins prior to the separation, labeling the proteins after the first dimension (Jackson et al., 1988; Urwin and Jackson, 1991), or labeling the proteins after the second dimension. Labeling the protein before and after 2D gel electrophoresis has been the dominant approach to the visualization of proteins.

Prelabeling of Proteins. Prelabeling of proteins prior to 2D gel electrophoresis is often performed by adding radioisotopically labeled amino acids to the growth medium of cells to provide in vivo labeling of proteins (O'Farrell, 1975). Typically, ^{35}S-methionine is incorporated in the culture medium for the radiolabeling of proteins. The protein mixture is then separated by 2D gel electrophoresis and visualized using a film or a phosphor imager screen. The radioactivity level used in the in vivo labeling approach cannot be applied for the study of human and animals.

Another common approach to the prelabeling of proteins is the derivatization of the proteins using neutral covalently attached fluorescent dyes. These dyes are covalently attached to the proteins prior to 2D gel electrophoresis (Urwin and Jackson, 1993). Because they are not charged, these dyes have the advantage of minimally disturbing the isoelectric properties of the proteins. The sensitivity of the approach, however, suffers because of poor absorption and inadequate quantum yields. Novel fluorescent dyes provide nanogram levels of sensitivity for proteins separated by 2D

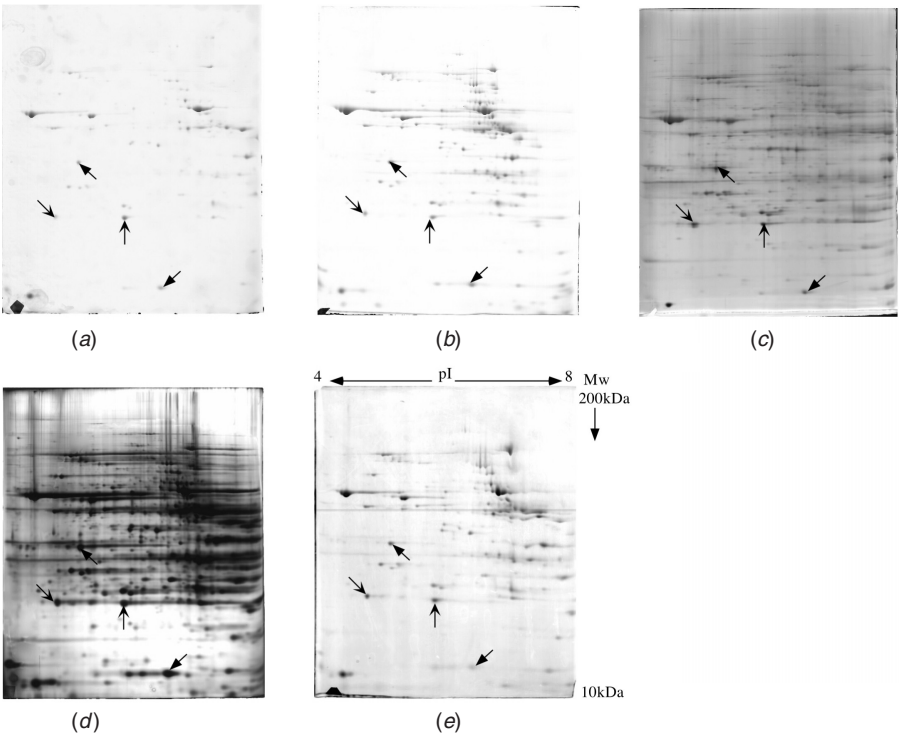


Figure 1.7. Staining approaches for 2D gel electropherogram. Comparison of sensitivities for different staining methods. A series of 2D gel electropherograms of 400 μg of mitochondrial proteins were stained with (*a*) Brilliant Blue G in acid alcohol medium, (*b*) colloidal Brilliant Blue G, (*c*) imidazole zinc, (*d*) silver, and (*e*) Sypro Orange (detected with a fluorescence laser scanner). Homologous spots are marked with arrows. Reproduced with permission from Rabilloud (2000).

gel electrophoresis (Unlu et al., 1997; Tonge et al., 2001; Yan et al., 2002). One disadvantage of this approach is the requirement for a fluorescence detection system and the requirement for an automated spot picker for any further processing of the individual protein spots.

Postlabeling of Proteins. The postseparation detection of proteins has been by far the most preferred route for the visualization of 2D gel-separated proteins. In particular, colloidal coomassie staining and silver staining are the methods of choice because of their ease of use and sensitivity. Colloidal coomassie staining typically provides a limit of detection (LOD) at about 25 to 50 ng of protein (Smith, 1994; Matsui, 1999a), while silver staining routinely provides limits of detection of about 5 ng of protein (Rabilloud, 1990, 1999; Rabilloud et al., 1994; Blum et al., 1987; Richert et al., 2004).

Silver staining is the most sensitive, direct visualization tool for proteins separated by 2D gel electrophoresis; however, it is tedious to manually perform. Instruments have been developed to perform automated silver staining and to significantly reduce the tediousness of the approach. Recently, an improved protocol for silver staining has been reported to provide limits of detection down into the subnanogram range. This was achieved by carefully selecting the chemical utilized for silver staining and by performing an extended rinse of the gel to reduce the background. Each spot on the gel can contain one or more proteins and can be manually or automatically excised from the gel for further analysis.

Fluorescence staining has also been reported for postseparation detection of gel-separated proteins. This technique typically provides a better sensitivity and a wider dynamic range than conventional colloidal coomassie and silver staining (Steinberg et al., 1996a, 1996b, 2000; Steinberg, 1997). The fluorescence-based approach, however, requires the access to a specialized fluorescence-based detection system and access to gel cutting robots when postprocessing of the proteins is necessary. Systems are now commercially available to perform the scanning of fluorescently labeled gels and the extraction of spots for gels.

Software to Handle Two-Dimensional Gel Electropherogram

Differential displays by 2D gel electrophoresis can be extremely tedious and frustrating when manually performed. The gel can display up to 10,000 spots, and from one experiment to another the spots can slightly shift on the pI and MW scale, depending on post-translational modification (PTM), sample processing, and sample composition. These shifts can become a serious problem for studies that require differential expression of proteins on an extended number of samples. To approach this problem, a few software packages have been developed for the alignment of multiple spots. The first challenge is to accurately detect and localize the positions of thousands of spots on a gel. Poor spot detection will affect the number of spots detected and their quantitation. The second challenge is to align multiple gels. This is typically achieved by selecting marker spots that are common to all the gels, and they are then used to realign the gels and the position of the spots. Obviously, when more deviations are present from gel to gel, then the number of markers required is higher. The issues of quantitation of multiple spots and the display of massive amounts of information also need to be addressed.

Commercial software packages are now available to solve some of these problems. In particular, the Melanie package (Wilkins et al., 1996b) from the Swiss Institute of Bioinformatics (*http://www.expasy.ch/*), the Phoretix 2D software from Phoretix (*http://www.nonlinear.com/*), and Gellab II from Scanalytics (*http://www.scanalytics.com/*) can be used to analyze 2D gel patterns.

PROTEIN PROCESSING

The information provided by 1D and 2D gel electrophoresis is far from being adequate to identify proteins. The observed MW on a gel can be easily off by a few thousand

Daltons. Furthermore, the posttranslation modifications of proteins and the truncation of proteins can drastically affect the observed MW and pI. Therefore, the information contained in gels is not conclusive for the identification of a protein, and more information needs to be extracted for unambiguous identification. How this extra information could be extracted from a protein stock in a gel spot was another challenge to overcome.

Blot Digestion of Proteins

Obviously, the first approach to the problem would be to remove the protein from the gel. A technique called electrotransfer, or electroblotting, was specifically designed for this purpose (Patterson et al., 1996; Aebersold et al., 1986, 1987). Once proteins had been separated by 1 or 2D gel electrophoresis, they could be readily transferred to a binding membrane, such as nitrocellulose, by simply sandwiching the gel with a receiving membrane and applying an electric field perpendicular to the plane of the gel. The proteins contained in the gel migrate toward the anode and encounter the binding membrane to which they become attached. The pattern of proteins present on the blot can then be visualized using a staining method such as coomassie or silver staining. In this manner, the separation profile on a gel can be transferred to a blot membrane with fidelity and be available for chemical/enzymatic treatment.

Electroblotting was initially used for protein identification using Edman degradation. The works of Aebersold pioneered this approach (Aebersold et al., 1986, 1987). Edman degradation will be introduced below. Basically, the electroblotting approach linked the gap between Edman degradation and the mainstream approaches for protein separation, that is, gel electrophoresis.

In the early to mid-1990s mass spectrometry (MS) became an attractive approach for the identification of proteins using peptides derived from the protein. The MS techniques will be described below. Electroblotting became the method of choice for processing proteins for MS analysis. The method was modified to allow the enzymatic digestion of the protein present on the blot. Once the spots of interest on a blot were excised, they were treated with a blocking agent such as polyvinglpyrrolidone-40 (PVP-40). The PVP-40 coats the membrane and allows the enzyme solution to freely access the blotted protein without binding to the membrane. Interestingly, the action mechanism of the enzyme is a combination of two reactions: a solid-phase reaction while the enzyme digests the attached proteins and a solution reaction for the further digestion of the released fragments. The end result is a solution of peptides that has been derived from the blotted protein. This peptide solution is then directly compatible with MS analysis.

The utilization of blots for protein identification was the technique of choice in the mid-1990s; however, it had some obvious limitations. Its foremost limitation was the tediousness and time required to perform the experiment. First, the 2D gel of the lysate of interest had to be produced and then electrotransferred to a nitrocellulose followed by staining. It was then followed by the excision of the stained spots and the

blocking of the spots using PVP-40, which involves a large number of repetitive washes. Finally, the blotted proteins were digested with trypsin. Altogether, the process could take about 3 to 4 days depending on the number of spots. To curb the tediousness, fluidic stations could be used to perform the majority of the postblot processing of the proteins.

The second limitation was the inconsistent quantitative transfer of proteins from the gel to the blot. This was apparent when comparing stained 2D gel with a similar stained blot. The rate of transfer of some proteins was high while the rate for other proteins was almost nil. The last important limitation was the carryover of PVP-40 during the analysis when not enough rinse steps were included after the blocking of the membrane.

In-gel Digestion of Proteins

The application of in-gel digestion of proteins for Edman sequencing had been developed in the early 1990s (Rosenfeld et al., 1992). Its application to the field of protein analysis by mass spectrometry was first introduced by Wilm et al. (1996). It was an instant success, and the method rapidly replaced the electroblotting approach. Its main advantage was the significant reduction in the number of processes required after gel electrophoresis to obtain digest suitable for analysis. As in the blot approach, a lysate is separated by gel electrophoresis. The difference now is that the gel is fixed and stained to highlight the proteins, the spots of interest are excised from the gels, properly rinsed, and then the enzyme solution is added to the gel pieces. The trick to introduce the enzyme into the gel is to shrink the gel pieces and let them swell in an enzyme solution. After digestion, only a few extraction steps are needed to obtain the peptides. Obviously, the reduction in labor and the reduction in processes offered by the in-gel digestion method made it the method of choice for the generation of peptides from proteins separated by gel electrophoresis.

PROTEIN IDENTIFICATION

Edman Degradation

In the early 1960s a technique called *the protein sequenator* was presented by Edman and Begg (1967) for the N-terminal sequencing of proteins. This chemical degradation technique allowed the extraction of individual amino acids in a cycle-dependant manner from the N-termini of proteins. The retention time observed by high-performance liquid chromatography (HPLC) was then used to identify the individual amino acid sequences. By repeating the chemical degradation cycle, it was possible to obtain the amino acid sequence at the N-terminus of typically up to 20 amino acids. Using this approach, it is possible to painfully de novo sequence a protein or to sequence a sufficient length of the protein to be able to clone the gene. The availability of protein and DNA sequence

TABLE 1.3. Performances of Different Mass Spectrometers for Protein Identification

	Sensitivity in MS/MS	Resolution	Cost	ID Based on MS	ID Based on M/MS
MALDI-TOF	Low	5,000–10,000	Medium	Yes	Sometimes by PSD
MALDI-Pulsar	High	10,000–15,000	Medium	Yes	Yes
ESI-Triple Quadrupole	Low	1,000	Medium	No	Yes
Esi-Ion Trap	Medium	1,000–3,000 higher in zoom scan	Low	No	Yes
ESI-FTMS	High	>50,000	High	Yes	Yes but slow
ESI-Pulsar and Qtof	High	10,000–15,000	Medium	No	Yes

databases has now facilitated the work and allowed protein identification using limited N-terminal amino acid sequencing.

Protein Identification by Mass Spectrometry

Mass spectrometry has been used in its different forms for the analysis of proteins as long as the Edman degradation technique. By today's standard, mass spectrometry was tedious, slow, and required large amounts of samples. Over the years, however, the idea of using mass spectrometers to perform protein identification has evolved with the improvement in instrumentation and in the changes occurring in genomic databases. The introduction of effective matrix-assisted laser desorption ionization and time-of-flight (MALDI-TOF) mass spectrometry and electrospray ionization (ESI) tandem mass spectrometry has revolutionized the field. Manufacturers of these instruments have "beefed-up" their efforts to produce more instruments for proteomic purposes. Therefore, when assessing previous reports, it is important to keep in perspective the tremendous amount of change that has occurred in recent years (Table 1.3).

MALDI-TOF Mass Speetrometry

MALDI-TOF: AN EVOLVING INSTRUMENT. Simultaneous development in the field of mass spectrometry has allowed rapid and accurate mass measurements of analytes by a technique called *matrix-assisted laser desorption ionization* (MALDI) and *time-of-flight* (TOF) mass spectrometry. This technique allows the transfer of peptides from a solid state to the gas phase, while the TOF mass spectrometer rapidly separates peptides according to their *m*/*z* ratio (Fig. 1.8).

Proteins isolated from 2D gel electrophoresis are digested, desalted, and then spotted on a MALDI plate for co-crystallization with a saturated matrix solution. Alternatively, protein digests can be separated by HPLC and eluting peptides deposited on a MALDI plate. The target plate is then introduced in the vacuum chamber of the mass

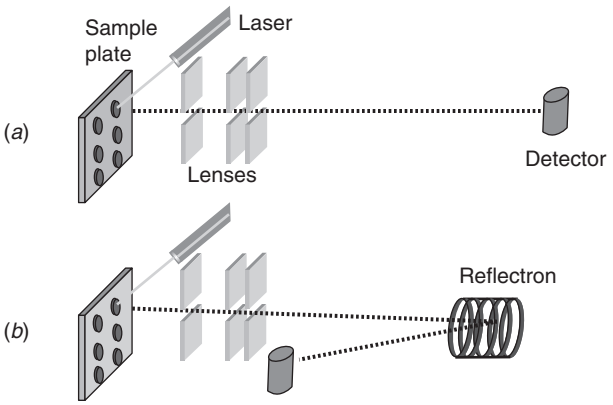


Figure 1.8. Schematic of a MALDI-TOF mass spectrometer. (*a*) Schematic of the linear mode. In the linear mode a portion of the sample is transferred to the gas phase by a pulsed laser. Once in the gas phase, the charge molecules are accelerated for a short distance and enter the field-free region of the tube, where they separate according to their *m*/*z* ratio. (*b*) Schematic of the reflectron mode. The mechanisms are the same as in the linear mode, except that a reflectron is placed to refocus the ions toward the detector leading to a higher resolution.

spectrometer. An automated translation stage positions individual spots on the axis of the mass spectrometer as well as at the focal point of a laser. A light pulse from a laser beam (1- to 10-ns pulse) of a wavelength tuned to the absorbance of the matrix is focused onto a limited area of the spot. The rapid transfer of energy ejects an ionization plume of material from the plate surface and brings along the peptides into the gas phase. The MALDI plate is biased to high voltages of +20 to +30 kV, with respect to a grounded orifice. This induces the positively charged peptides to accelerate toward the orifice of the flight tube. The peptides are affected by the electric field during the time that they are between the plate and the orifice. They reach the orifice with a velocity proportional to $(z/m)^{1/2}$. Once they pass the orifice, the peptides have all the same kinetic energy but not the same velocity because of their different mass. Thus, molecules that have an identical charge will therefore have a velocity proportional to their masses. Generally, in MALDI the resulting peptide ions are singly charged. Once in the field-free region of the flight tube, the peptides fly through the tube only according to their initial velocity at the orifice. The vacuum in the flight tube is such that the likelihood of a collision with another molecule while in the flight tube is low. Because the peptides are usually singly charged, they traverse the flight tube according to their mass and hit the detector at different time intervals. The mass analyzer, triggered by the laser pulse, records the signal detected versus the time of flight, which can be readily transformed into *m*/*z* ratios if the mass spectrometer is properly calibrated.

In principle, a longer tube infers a better separation of the different masses. Collisions with gas molecules in the path of the analytes would effectively destroy the separation. Therefore, the longer the path, the larger the vacuum requirement is

imposed to maintain a path free of possible collisions. This is why you will see these mass spectrometers equipped with sufficient pumping to achieve 10^{-9} to 10^{-10} torr.

A second factor that affects the resolution of the separation is the initial kinetic energy distribution due to the burst of analytes off the MALDI plate. The technique of *delay extraction* of the ion and the reflectron were added to the MALDI-TOF to improve the resolution, based on a more uniform kinetic energy. The delay extraction acts on the kinetic energy prior to the separation, while the reflectron acts on the kinetic energy distribution during the time of flight. The delay extraction is simply achieved by introducing a time delay prior to the high-voltage biasing of the MALDI plate. This allows the ions ejected from the plate to kinetically cool down providing improved resolution.

The reflectron allows the kinetic refocusing of the ions, which also results in better resolution. It is normal to obtain a peptide mass accuracy down to 10 to 50 ppm and to obtain a 5000 to 15,000 resolution. The reflectron is a piece of hardware that is placed in the path of the ions. Therefore, the design of a MALDI-TOF equipped with a reflectron is very different than the conventional MALDI-TOF instrument. Furthermore, different designs have been fabricated; however, the principle remains the same. The reflectron is a set of ring lenses that are stacked together. Increasingly higher potentials are applied to the rings. In the time of flight each peptide is represented by a pocket of ions that travel with a small range of kinetic energy (i.e., speed). The ions that are at the front edge of the pocket have a higher kinetic energy, while the ions at the trailing edge of the pocket have a lower kinetic energy. The fast moving ions penetrate the reflectron first and start to feel the repulsing electric field imposed by the reflectron. Because of their high kinetic energy, they travel further into the reflectron before reaching a point of zero kinetic energy. Then they are reaccelerated by the repulsing field toward the entrance of the reflectron. Meanwhile, the other ions also enter the reflectron in order of their decreasing kinetic energy. Those ions will travel less of a distance into the reflectron before their directions are reversed. As a result, the slow moving ions are now at the front of the traveling pocket, while the fast moving ions are behind. All the ions then travel back in the field-free region of the TOF tube. While traveling, the pockets of ions become focused by the higher speed trailing ions catching up to the slower leading ions. Obviously, the resolution that is achieved is very dependant on the positioning of the detector at the focal point of the reflectron. The pocket of ions will not have time to focus if the detector is positioned too close to the reflectron. Also, the pocket of ions will start defocusing if the detector is positioned too far from the reflectron.

Different Approaches to Sample Preparation for MALDI. Sample preparation for MALDI is often an art. Numerous protocols have been reported for the spotting of the samples. Here we review the most utilized approaches.

Cleanup Prior to Spotting. The utilization of reverse-phase material, either in a column format or loosely packed into a pipet tip, has proven to be a useful approach for the desalting of samples and their separation from polymers prior to MALDI-TOF

analysis. The properties, however, of the reverse-phase material and the packing need to be carefully assessed. For example, Ziptips are pipet tips that have a small plug of extraction material at their distal end. The plug of material is obtained by embedding dispersed reverse-phase material in a polymeric support. To maintain easy flow of liquid through the tip, the reverse-phase beads are well dispersed into the embedding polymer resulting in a low back-pressure. The tips have proven useful for the cleanup of high to mid levels of peptide mixtures. The dispersion, however, of the reverse beads causes the extraction efficiency to be diffusion limited, that is, the linear flow is too high compared with the distance that the peptides have to diffuse to reach the packing material. Therefore, the extraction efficiency drops significantly for lower concentrations of peptide mixtures, and it can therefore be a challenge to analyze lower amounts (<100 fmol) of peptide mixtures.

Spotting Approaches. Over the years different sample preparation methods have been developed for the analysis of minute amounts of peptides. The most popular method has been the *dried droplet method* (Karas and Hillenkamp, 1988). In this method the peptide mixture of interest is dried down and resuspended in an acidified water : organic (acetonitrile or methanol) solution saturated with α-cyano-4-hydroxycinnamic acid. One microliter from this mixture is then deposited onto a MALDI plate and allowed to dry. Although this is a rapid way of preparing samples, it does not provide for the most sensitive analysis of peptides.

A second approach termed the *two-layer method* was recently introduced for the analysis of peptides (Dai et al., 1999). First, a solution of α-cyano-4-hydroxycinnamic acid is pipetted on a MALDI plate and allowed to dry to form a microcrystal layer. Then a solution with the analytes and the matrix is pipetted on top of the first layer and allowed to dry. Although this approach is more tedious, it offers a better representation of the peptides, a cleaner spectra, and it can be automated.

Cleanup Post Spotting. A cleanup method can be directly applied on the MALDI plate either with the dried droplet method or the two layer method. The sample cleanup method consists of performing a cleanup procedure right after the spotting. Briefly, a drop of water is added to individual spots on the MALDI plate for a set time period. The water droplet is then blown away, and the spot is allowed to dry again before being placed in the MALDI-TOF mass spectrometer. This rinse step allows the extraction of salts and other hydrophilic molecules from the dried crystal. This often leads to a decrease in background, therefore an increase in signal/noise ratio. The timing of the rinse is important to avoid unnecessary loss of hydrophilic peptides.

On-Plate Concentration. As mentioned previously, mass spectrometers are concentration-dependant devices. MALDI is more complicated because the ionization process and the involvement of the matrix are not clearly understood. It appears that an increase in the analyte concentration on the surface of the MALDI plate might improve the signal observed; this, however, would only work well if the sample can be sufficiently cleaned up. The amount of sample is limited, so the only way to increase

the concentration is to limit the size of the MALDI spot. MALDI plates were developed with spotting sections of small diameters that are hydrophilic, while the rest of the plate is hydrophobic. When a drop of peptide solution is deposited onto the spot, the hydrophobic surroundings and the surface tension force the droplet to take a shape that limits its contact with the surface. As the droplet dries, it focuses into the small hydrophilic patch. The end result is the concentration of a large volume into a small area on the MALDI plate. In itself, this is not sufficient to provide significant improvement because the contaminants, usually salts, are similarly concentrated. The addition of an on-plate cleanup (drop of water method) helps to remove salts while retaining peptides in the crystal. The drop of water is then blown away.

Protein Identification by MALDI-TOF: A Moving Target

Using Accurate Protein Mass Measurement. It was believed that protein identification could be achieved based on the accurate measurement of the protein MW. But, it was quickly realized that the growing size of databases and the accuracy of the mass measurement limited the unambiguous identification of proteins based on their MW. The information carried by the protein MW was insufficient to identify the protein. It was then realized that unambiguous identification of a protein could be readily achieved by accurately measuring the masses of the peptides contained in a protelytic digestion of the protein. Clearly, in itself, the mass of an individual peptide derived from a protein is inadequate to identify the protein, although, the masses of a large set of peptides derived from the same protein is often sufficient to identify the protein. MALDI-TOF MS is the method of choice for the measurement of peptide masses; however, MALDI-TOF MS experiments must be carefully designed to preserve the mass accuracy. In MALDI-TOF, the ions are accelerated for a short distance, and then they freely fly in the TOF tube, sometimes up to a few meters. It is generally assumed that the acceleration space is constant; however, any errors on the acceleration space are propagated in the time of flight and thus affect the accuracy of the measured masses. In reality, the plate fidelity, the positioning, and tilt of the plate holder can slightly change the acceleration space and, therefore, can affect the mass accuracy. Approaches have been developed to reduce this phenomenon. This is normally achieved by adding an internal standard to the sample or correcting the masses by scanning an external calibrant near the sample spot on the MALDI plate.

Protein Identification by MALDI-TOF and Accurate Peptide Masses. The identification of proteins by MALDI-TOF mass spectrometry is generally achieved by measuring the m/z ratio of the peptides predominantly of charge +1. The combination of accurate peptide mass measurement with the availability of protein sequence databases forms the basis of protein identification by MALDI-TOF. Figure 1.9 describes the principle behind the identification of proteins by MALDI-TOF. This is often called peptide mass fingerprinting. The measured masses present in tryptic digests are tabulated, and the known contaminants are deleted from the list. The reduced mass list is then used to search protein databases.

Different software packages have been developed for the identification of proteins based on the accurate measurement of peptide masses. The simplest method for scoring

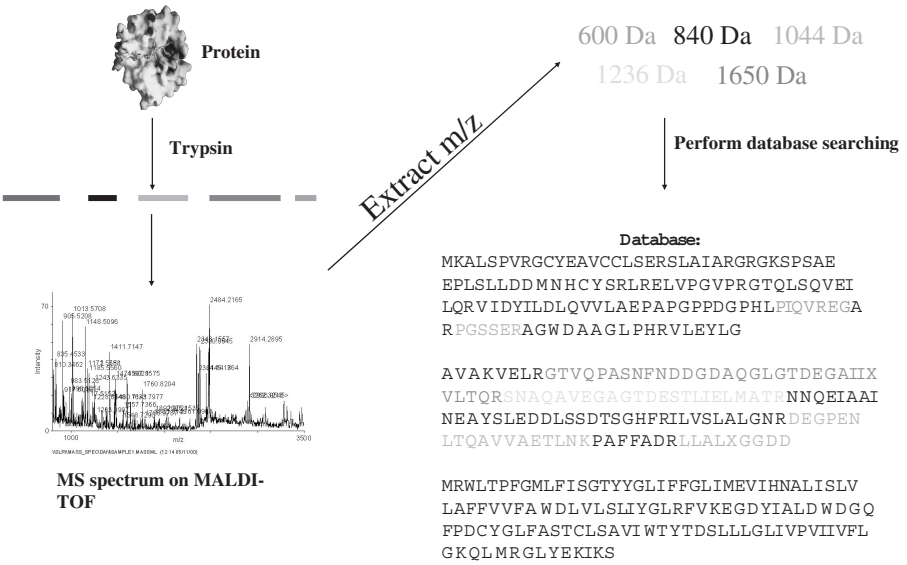


Figure 1.9. Scheme employed for the identification of proteins based on MALDI-TOF and peptide mass fingerprinting. A protein of interest is digested with trypsin and the resulting peptide mixture is analyzed by MALDI-TOF mass spectrometry. The measured *m/z* ratios are then used to search protein/DNA databases leading to the identification of the protein. The matching proteins are ranked according to the number of observed *m/z* ratios that match to their predicted tryptic peptide patterns.

is to add the number of peptide masses that match with the predicted masses for each entry in a protein database. The database entries are then ranked according to the number of hits. This forms the basis behind software such as PepSea and MS-Fit (*http://prospector.ucsf.edu*). Typically, these software packages work well for quality experimental data.

Mascot originated from the software called MOWSE (Pappin et al., 1993). MOWSE uses more information to make its decision on the score by taking into account the protein size and the relative abundance of peptides in the databases. Mascot further incorporates probability scoring for the probability that the match between the data and the entry in the database will be a random event. This score is calculated for every entry in the database. The identification is then established by ordering the proteins with a decreasing probability of being a random match. Mascot can be freely accessed over the web (*http://www.matrixscience.com/*) and is also available for commercial purposes through MatrixScience.

ProFound (Zhang and Chait, 2000) uses a different approach based on Bayesian theory to rank the protein sequences in a database by their probability of occurrence. It is an expert system (i.e., it simulates what an expert in the field would do) that uses detailed information about each protein sequence and empirical information about the

distribution of proteolytic peptides that are included in the scoring scheme. ProFound is also free over the Internet (*http://prowl.rockefeller.edu/*).

Generally, all these software packages perform well when good-quality spectra are available. The ones that provide more advanced scoring schemes perform better when less information is available or when the quality of the MS spectra is reduced. Regardless of the software, the identification of the protein depends on the number of peptides observed, the accuracy of the measurement, and the size of the genome of the particular species. For smaller genomes, such as yeast and *Escherichia coli* (*E. coli*), protein identification using the MALDI-TOF mass spectrometer is generally successful. For larger genomes the rate of success drops significantly using MALDI-TOF.

Postsource Decay. MALDI-TOF mass spectrometry also offers the possibility of recording the fragmentation patterns obtained from a peptide. This is achieved using a technique called *postsource decay* (PSD). PSD is achieved by increasing the laser power beyond the value needed to generate ions. The precursor ions are transferred from the MALDI plate to the gas phase. The excess energy induces the precursor ions (peptides) to fragment along their backbone. Generally, these ions are not seen in conventional MALDI-TOF analysis because of their lower kinetic energy. Fortunately, on the MALDI-TOF equipped with a reflectron, the lower kinetic energy of peptide fragments can be compensated by changing the settings on the reflectron. PSD is typically achieved by acquiring spectra for specific mass ranges with different settings on the reflectron. All the spectra are then stitched together to make a full PSD spectrum.

Although PSD seems to be a rapid way of obtaining fragmentation patterns of peptides, it seriously suffers in terms of sensitivity. Furthermore, the fragmentation patterns are often difficult to discern and are of poor quality. Therefore, PSD has not been the method of choice for the generation of peptide fragmentation patterns.

Hybrid Instruments

Conventional MALDI-TOF mass spectrometers can also provide fragmentation patterns related to the amino acid sequence of a peptide. This is also done using postsource decay; however, the mass selection of the peptide to be fragmented, the sensitivity of the approach, and the quality of the MS/MS spectra make it difficult to utilize postsource decay for the routine and rapid generation of MS/MS spectra. Recently, a set of novel MS/MS-capable mass spectrometers have been developed based on MALDI ionization. These instruments combine the MALDI ionization technique with the fragmentation of ions by collision-induced dissociation (Shevchenko et al., 2000; Loboda et al., 2000). For example, the recently introduced MALDI-Pulsar from Sciex is illustrated in Figure 1.10. This instrument, as its name alludes to, includes a MALDI ionization interface followed by a set of quadrupoles and a collision cell. The first quadrupole is used as an ion guide. It is followed by a second quadrupole, which is either used as an ion guide, a precursor scan, or for a parent ion selection. The first quadrupole is followed by a collision cell that can be used to fragment ions by collision-induced dissociation. A pulsing grid set at a 45° angle is positioned after the collision cell and deflects the ions upward into a TOF mass analyzer equipped with a

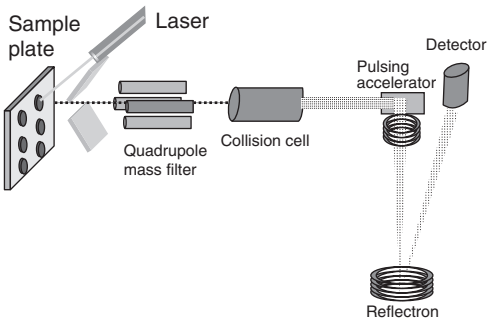


Figure 1.10. Schematic of a MALDI-Pulsar mass spectrometer. A portion of the sample is transferred from the MALDI plate to the gas phase by a pulsed laser. Once in the gas phase, the charged molecules go through a quadrupole, which can be used as a mass filter, lenses, or a mass analyzer. They then enter a collision cell followed by a pulsing plate, which pulse part of the beam of ions into the perpendicular TOF. They enter the field-free region of the TOF in which they separate according to their *m/z* ratio. A reflectron is employed to refocus the ions and to provide better resolution. This instrument is capable of providing MS spectra and good quality MS/MS spectra generated by collision-induced dissociation.

reflectron. Although a pulsing laser is still used to generate ions off the MALDI plate, it is not used for the timing and correction of the masses of the observed ions.

In this design, the ionization of the sample is decoupled from the acceleration of the ions into the TOF tubes. Furthermore, significant collisional cooling is present in the first quadrupole, reducing the distribution of the kinetic energy of the ions. Therefore, the geometry of the plates does not affect the mass accuracy of the instrument. This is a significant difference over conventional MALDI-TOF in which great care must be taken regarding the plate geometry. This means that internal or close external calibrants are not necessary to maintain the mass accuracy of the MALDI-Pulsar system. It also means that totally different plate designs can be constructed.

The Pulsar also allows peptide selection and efficient fragmentation by collision-induced dissociation in the collision cell. Collision-induced dissociation is more reliable than postsource decay for generating fragmentation patterns. MALDI ionization predominantly provides peptides of charge 1+. The fragmentation of 1+ ions generates lower quality fragmentation patterns than what is usually obtained for 2+ and 3+ ions. Therefore, this hybrid approach offers significant advantages over the conventional MALDI-TOF in terms of its efficient fragmentation patterns. It also offers advantages over the ESI-MS/MS for the rapid screening of samples.

Electrospray Ionization Mass Spectrometry. Electrospray ionization is also a very popular approach to introduce protein and peptide mixtures to mass spectrometers. Typically, ESI is used in conjunction with a triple quadrupole, an ion trap, or a hybrid quadrupole-TOF mass spectrometer. ESI is mainly popular because it provides a direct interface between the mass spectrometer and the atmospheric pressure. It can

also be readily coupled to separation techniques such as HPLC, liquid chromatography (LC), and capillary zone electrophoresis (CZE), or it can be used for the continuous infusion of samples. Until the introduction of the MALDI-Pulsar, the ESI-based mass spectrometers were the only viable approach for the generation of MS/MS spectra. Furthermore, ESI produces more multiply charged ions that provide richer MS/MS spectra.

Principle. Electrospray ionization (Fenn, 1990) allows analytes to transfer from the liquid phase to the gas phase at atmospheric pressure. Generally, the ionization process is achieved by applying an electric field between the tip of a small tube and the entrance of a mass spectrometer. The electric field forces the charged liquid at the end of the tip to form a cone, called *Taylor cone*, that minimizes the charge/surface ratio. Droplets form at the end of the cone and move toward the entrance of the mass spectrometer. Different theories have been put forward to explain what happens after the droplets have been formed and how the transfer of the analytes to the gas phase is achieved. The most popular explanation is that the liberated droplets go through a repetitive process of solvent evaporation, whereby they fragment into smaller droplets to reduce the charge density on them. This leads to a large number of droplets of shrinking sizes until the solvent has disappeared; thus the analytes are left in the gas phase. Furthermore, as the droplets shrink, the pH in the droplets decreases and facilitates the protonation of the analytes.

Mass Spectrometer and Information. The most important advantage of ESI is that it can be readily coupled online with separation techniques. This has been illustrated over the years by the numerous development of methodologies based on separation techniques coupled to electrospray mass spectrometers. The variety of applications, separation techniques, and mass spectrometers utilized with ESI can be confusing to those new to the field. Therefore, in the next few sections we will introduce the different mass spectrometers with their respective limitations, the different separation techniques that are specifically used for proteomics, and the generation of tandem mass spectra.

Mass Analyzers Utilized with Electrospray Ionization in Proteomics. There has been a flurry of novel commercial mass spectrometers over the last 5 years. This is in part related to the constant demand for newer and better mass spectrometers imposed by the biotechnology/pharmaceutical sectors. Even though the number of mass spectrometers is increasing, and the confusion related to their application in proteomics is building even faster, it is still possible to make a logical assessment of their functions, advantages, and limitations.

Triple Quadrupole Mass Spectrometer. The first mass analyzer that was utilized for proteomic studies is the triple quadrupole mass spectrometer. A quadrupole mass analyzer consists of four rods placed at equidistance as if they were placed on the surface of a cylinder. The Mathieu equation was derived to describe the motion of a charged molecule in an electric field. It is also applicable to describe the motion of charge molecules in the triple quadrupole mass spectrometer. Electric fields can be con-

stant [direct current (DC)] or variable [alternating current (AC); different amplitude and frequency]. The Mathieu equation takes into account the combination of DC and AC fields. The combinations that are possible are such that the quadrupole can be used for three functions. The combination of a DC and radio frequency (RF) potential transforms the quadrupole into a mass filter to transmit a specific *m*/*z* ratio. The application of a RF-only mode will set the quadrupole as an ion guide, while the application of a DC-only mode transforms the quadrupole into a lens element. For every mass, there is a region of stability for the DC, RF amplitude, and frequency. Therefore, the resolution of the instrument and the ion transmission (sensitivity) are intimately linked. Increasing the resolution decreases the number of ions transmitted. Changing the shape of the quadrupole rods affects the electric field and can also improve the resolution. The mass filter mode and the RF-only mode are the two most common applications of the quadrupole. Furthermore, most triple quadrupoles have been fabricated using cylindrical rods for simplicity. More recently, designs for hyperbolic rods have allowed improved resolution, and they have been introduced in the TSQ-quantum (Thermo-Finnigan):

$$d^2x/dt^2 + [\alpha + \beta f(t)]x = 0 \tag{1.1}$$

where α and β are constant and the $f(t)$ is the sinusoidal function of time. This equation is called the Mathieu equation. In a quadrupole mass analyzer only two axis of direction are being affected by the field.

The triple quadrupole mass spectrometer consists of three subsequent sets of quadrupoles, positioned one after the other in a linear fashion toward the detector (Fig. 1.11). The functions of the first and third quadrupoles are either as mass filters or as ion guides. The second quadrupole is typically run in the RF-only mode (i.e., ion guide) and modified to allow the introduction of gas. It is used for collision-induced dissociation of charged molecules. This combination of quadrupole (q_0), collision cell (q_1), and quadrupole (q_2) allows a multiplicity of experiments to be performed.

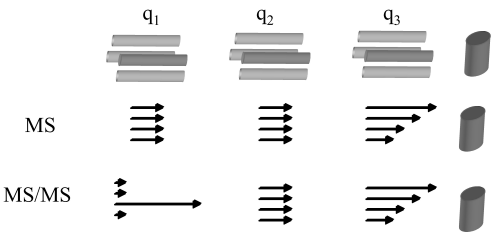


Figure 1.11. Triple quadrupole mass spectrometer for the analysis of peptides. A set of two experiments is performed on the triple quadrupole mass spectrometer for peptide analysis. In the first pass an MS spectrum is acquired to detect the *m*/*z* ratio of the analytes. Then, a peptide corresponding to a specific *m*/*z* ratio is selected in a second experiment, isolated in q_1 and fragmented in the collision cell. The resulting fragments are separated in q_3 leading to an MS/MS spectrum.

Although numerous experiments can be concocted with a triple quadrupole mass spectrometer, only two serial experiments are performed for peptide identification. The first experiment consists of running q_0 and q_1 in RF-only mode while scanning the m/z on q_2 (DC and RF). In this fashion, q_0 and q_1 let the positively charged analytes pass through while q_2 continuously scans the m/z, generating an MS spectra (signal versus m/z) for the positively charged ions present in a proteolytic digest.

The second set of experiments, called MS/MS, consists of selecting and fragmenting an ion followed by the separation of the generated daughter fragments. This is achieved using q_0 as a mass filter to allow only a narrow m/z window around the selected ion to pass. A small amount of neutral gas is then added to q_1, while the potentials, applied to lenses, are changed to provide increased kinetic energy to the ions entering q_1. The ions collide with the small gas molecules and fragment by collision-induced dissociation. The generated ions are then separated by q_2. The end result is an MS/MS spectrum that contains the selected ion fragmentation patterns.

The combination of both experiments (MS and MS/MS) allows peptides to be detected, selected, and fragmented. Typically, the triple quadrupole mass spectrometer is either manually or automatically cycled through MS and MS/MS acquisition.

The application of the triple quadrupole mass spectrometer in proteomics was supplanted by the introduction of more sensitive instrumentation such as ion trap mass spectrometers and hybrid mass spectrometers. The advantage of the triple quadrupole mass spectrometer with collision-induced dissociation is the efficient conversion of precursor ions into product ions. The disadvantages of the triple quadrupole mass spectrometer are its low resolution, the mass discrimination (i.e., peak height dependency on mass), the narrow window, and the strong mass dependency on collision energy for efficient fragmentation of peptides. Generally, the information generated in MS mode is too low of a resolution to be useful for database searches. Furthermore, the mass accuracy is generally too low to allow adequate de novo sequencing of proteins when the identification by database searching fails.

Ion Trap Mass Spectrometer. The second analyzer introduced for proteomic studies was the ion trap mass spectrometer. The ion trap mass spectrometer utilizes a combination of electrodes for the accumulation of ions in a space defined by the shape of the electric field present in the trap. Similar to the triple quadrupole mass spectrometer, the ion trap mass spectrometer follows the Mathieu equation. Because of the shape of the electrodes and the definition of the electric fields, the resolution obtained on the trap can be higher than on conventional quadrupole instruments. An example of a complete ion trap mass spectrometer is described in Figure 1.12. It consists of a heated stainless steel capillary followed by a skimmer, an ion guide, by the ion trap, and a detector. All of these parts are maintained under vacuum. The ion guides have a dual function: to guide the ions toward the ion trap and to gate the ions into the trap. The number of ions present in the trap is critical and needs to be controlled. Otherwise, a phenomenon called *space charging* can occur that causes the performance of the instrument to be distorted. The ion optic is used to control the injection of ions into the trap. The reversal of the polarity applied to the optics deviates the ions from their paths and prevents the ions from entering the trap. In reality, a scan from an ion trap consists

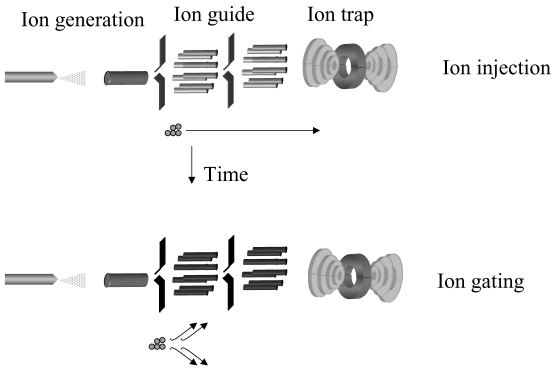


Figure 1.12. Schematic of the ion trap mass spectrometer. The schematic of the trap illustrates how the ions are guided to the trap and how the gating approach is used to sufficiently fill the trap prior to MS and MS/MS experiments.

of at least two experiments. In the first experiment a short burst of ion is injected into the trap and then ejected to the detector. The signal observed by the detector is used to automatically calculate the maximum allowable injection time at which space charging can still be avoided. Then the ion optics are open to inject the ions into the trap for the calculated injection time. The ions injected into the trap are not stationary but have an orbit that resembles a boomerang shape. Depending on the *m*/*z* ratio of the ions, the orbit will occupy more or less space in the trap. Once enough ions have been accumulated, they can be sequentially ejected from the trap by changing the RF amplitude applied to the ring electrode. This creates instabilities in the orbits of specific ions, pushing them further away from the center of the trap, and finally drawn to the ejection point toward the detector. The lower *m*/*z* ions are ejected at a lower RF amplitude, while the higher *m*/*z* ions require higher RF amplitudes. Typically, the RF amplitude is calibrated using analytes of well-defined masses.

Ion ejection and ion fragmentation can be achieved by *resonance excitation*. This approach consists of creating a resonance effect in the orbit of the ions to increase its axial kinetic energy and to increase the axial dimension of the orbit. Resonance excitation is achieved by applying a small DC voltage to the endcap electrodes at the same time as applying an RF to the ring electrode. This creates a cumulative increase in the kinetic energy of the selected ions, thus leading to an unstable orbit and ejection. Furthermore, the increased kinetic energy can be funneled to fragment the selected ions by introducing a small amount of inert gas into the trap.

The operation of the ion trap mass spectrometer in MS mode consists of a rapid scan for automatic gain control, followed by the timed injection of ions, and finally the scanning of the RF amplitude to successively eject the ions from the trap. The end result is an MS spectrum with the RF amplitude on the bottom axis, which then can be converted into *m*/*z* by calibration.

The operation of the ion trap mass spectrometer in MS/MS mode consists of a rapid injection/ejection for automatic gain control, a timed injection of ions and an ejec-

tion of undesirable ions (except a preselected mass window), followed by a resonance excitation of the selected peptide in the presence of small gas molecules, and, finally, by a scanning of the RF amplitude to successively eject the fragmented ions from the trap. The end result is an MS/MS spectrum. Although the fragmentation of ions is also obtained by collision-induced dissociation, the quality of the MS/MS spectrum is different than the one obtained on a triple quadrupole mass spectrometer. This can be explained by the fact that only the selected ions are sensitive to the RF and to the resonance excitation; therefore, once fragmentation occurs, the resulting fragments are of different m/z and are invisible to the RF/resonance excitation. This is not the case for triple quadrupoles because all the generated charged fragments are still subjected to the acceleration voltage and can gain sufficient kinetic energy to proceed into further collision-induced dissociation.

The ion trap mass spectrometer has seen wide acceptance in proteomic application mainly due to its level of automation, sensitivity, and cost. The information provided in MS mode is insufficient for protein identification based solely on peptide masses. The information provided by MS/MS spectra is, however, generally sufficient to identify proteins by database searching. De novo sequencing of peptides is also difficult on ion trap mass spectrometers due to its low resolution and its lower mass cutoff in MS/MS mode.

More recently, new linear-type ion trap mass spectrometers have been introduced. These offer a larger trap space allowing wider dynamic range as well as better sensitivity and resolution over conventional traps. Furthermore, some of these instruments can be operated either as a trap or as a triple quadrupole mass spectrometer (QTrap from MDS-Sciex).

Ion Cyclotron Mass Spectrometer. The third analyzer is the *ion cyclotron Fourier transform* mass analyzer (Comisarow and Marshall, 1974). At the basic level this analyzer is also an ion trap mass spectrometer. It typically consists of an ion source, followed by an ion guide, and, finally, by a Fourier transform cell contained in a superconducting magnet (Fig. 1.13). As in the ion trap mass spectrometer, the purpose of the ion guide is to focus the ions and transfer them into the Fourier transform cell. In some designs the ion guide is also utilized to perform the accumulation and selection of ions, for example, using a fragmented quadrupole before pulsing them into the trap. The purpose of the superconducting magnet is to provide a uniform magnetic field in the trapping region of the cell. The cell itself forms a box made of two detector electrodes facing each other, another set of two RF electrodes facing each other, and a third set consisting of trapping plates. In the cyclotron, the ions have to pass through significant fringing fields created by the superconducting magnet before progressing into the trap. The ions that enter the Fourier transform cell are trapped by the magnetic field and the trapping plates. Furthermore, once in the trap, they will be limited only to circular orbits. It turns out that every ion of a specific mass and charge has a different cyclotron frequency, defined by:

$$w = \beta z m^{-1} \tag{1.2}$$

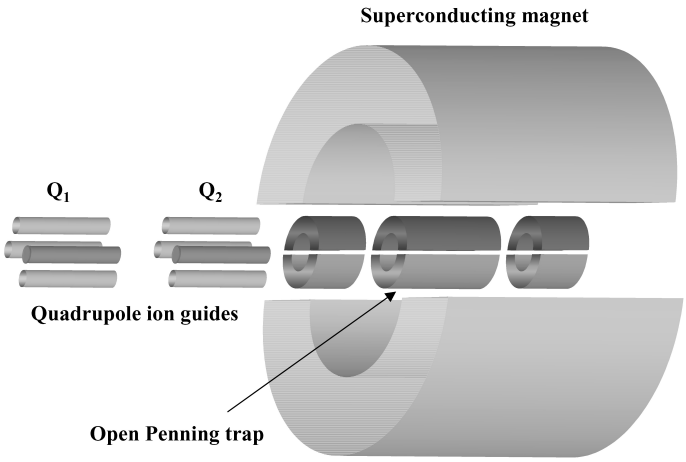


Figure 1.13. Schematic representation of a Fourier transform mass spectrometer. The schematic illustrates the utilization of the quadrupole as ion guide. The trap consists of an open penning trap inserted into the bore of a superconducting magnet. A set of electrodes in the trap is used to bring the ions to their cyclotron frequency. A second set of electrodes is used to measure the frequencies.

where w is the cyclotron frequency in radians per second, β is the magnetic field strength in tesla, z is the charge, and m is the mass of the ion. This equation is only valid once the ions are in orbits, but the ions do not spontaneously spring into the orbits. By applying an oscillating electric field that corresponds to the cyclotron frequency of the ions to the two opposite electrodes of the cell, a cyclotron resonance can be induced. This cyclotron resonance induces the ions to travel into an increasing circular orbit. After a short time, the electric field is turned off, and the ions reach a stable orbit determined by their m/z ratios and the magnetic field. The stabilization and resolution of different circular orbits require a very homogeneous magnetic field created by a superconducting magnet. Typically, the cyclotron resonance is provided by a sine wave signal generator. Ions of different m/z ratios have different resonance frequencies, and they will only be accelerated if the right resonance frequency is applied to the electrodes.

The ions in orbit pass by the electrodes and induce a small, but detectable, current at the electrode, called the *image current*. The end result of the motion is an alternative current in phase with the frequency of the ion orbits. Furthermore, the amplitude of the image current is proportional to the number of ions in orbit. An amplification system is used to transform the small AC signal into a larger and easier signal to process.

Obviously, this would be a relatively useless approach if only one ion could be trapped and detected at a time. It turns out that in the beginning of the ion cyclotron design, the ions were individually detected, but it could take up to 20 min to acquire a full MS scan. The twist around this problem is to careful look at Eq. (1.2). If all the

ions could be accelerated into their respective orbits, they would all have a resonance frequency related to their m/z ratio. Therefore, the induced current would be a composite of all the currents induced by the ions at their respective frequencies. Clearly this is a signal processing problem that can be handled by Fourier transform analysis to extract each individual signal from the composite.

Today, Fourier transform mass spectrometry (FTMS) spectra are rapidly acquired by first activating all the ions with a sweep of the RF to accelerate all the ions. The transient image current (due to the multitude of m/z) is accumulated, and the MS spectrum is reconstructed by the Fourier transform. Although the FTMS instrument is the ultimate choice in terms of resolution and sensitivity, it has not seen widespread acceptance in proteomics. In the past, most cells were designed to trap a defined fraction of the ion beam. This means that the dynamic range of the cell was limited. Methods were developed to increase the ion density in the ion beam by trapping ions and by the pulsed injection of ions using fragmented quadrupoles prior to the cyclotron. Again, the challenge is to have enough ions for detection while avoiding space charging. For most real-life samples this can still be an issue forcing the mass accuracy to be above the 1 to 1.5 ppm level. The cost of FTMS instruments and their robustness have also been serious issues. The information, however, provided in MS mode is often enough for protein identification based solely on accurate peptide masses. The information provided in MS/MS mode can be used for database searches or for de novo sequencing of proteins; however, the generation of MS/MS fragments by collision-induced dissociation in the trap is relatively slow in FTMS and precludes its utilization for rapid analysis.

Recently, hybrid instruments that combine external selection and fragmentations of ions have been developed. For example, Thermo-Finnigan developed a hybrid linear ion trap coupled to an FTMS. The advantage of this approach is the rapid fragmentation and analysis of peptides. As the fragmentation is performed in the linear trap, there is no need to wait for long ion cooling period in the FTMS. Other types of hybrid FTMS have been introduced by IonSpec and Bruker.

Hybrid Mass Spectrometers. The last analyzer is the hybrid quadrupole time-of-flight mass spectrometer. The Pulsar mass spectrometer from Sciex and the Qtof mass spectrometer from Micromass are two examples of hybrid instruments. The Pulsar utilized with ESI is the same as the one previously described for the MALDI-Pulsar. It utilizes an ESI source instead of a MALDI source. Please refer to the MALDI section for the description of the Pulsar instrument.

Techniques to Introduce Peptide Mixtures to ESI-MS/MS. The MALDI-TOF mass spectrometry is definitely a high-throughput platform for the identification of proteins. So why rely on the more cumbersome and lower-throughput ESI approaches? ESI provides multiply charged peptides that can be efficiently fragmented by collision-induced dissociation. The MALDI-TOF provides singly charged peptides that are more difficult to fragment even when collision-induced dissociation is available. Furthermore, ESI can be readily coupled with infusion and concentration/separation techniques. Hence, intense efforts have been focused in recent years on the

development of infusion and separation techniques for efficient analysis of protein digests. These techniques differ in sensitivity, the handling of contamination, the quality of the MS/MS spectra, and the throughput. In the following sections we will review the different approaches to sample introduction, their disadvantages, and their advantages.

Continuous-Infusion ESI-MS/MS. Within the last 10 years, ESI has become compatible with the analysis of low-level protein digests. The main challenge was to make the flow and concentration detection of conventional electrospray approaches compatible with the volume and concentration that were obtained with blot, or in-gel digestion of separated proteins. Conventional electrospray requires a few microliters/minute flow rate to be stable, while the final volume of a digest is in the low microliter levels. Therefore, efforts were placed in developing new "lower-flow" versions of the electrospray process. The microelectrospray (100 to 500 nL/min), the nanoelectrospray (1 to 100 nL/min), and the picoelectrospray (<1 nL/min) methods were developed. They all have the reduction of the flow requirement in common by reducing the internal and external diameter of the sprayer. The introduction of these novel electrospray approaches has greatly facilitated the analysis of protein digests by providing longer analytical time windows, better ionization, and better transfers to the mass spectrometer.

In particular, the "low-flow" requirement of the nanoelectrospray became rapidly attractive for the identification of proteins based on the continuous infusion of a sample (Shevchenko et al., 1996a, 1996b, 1997). From as little as 1 µl, an analytical time of 30 min to 2 h could be easily obtained. This would leave plenty of time to acquire good-quality MS spectra, to select large number of ions, and to successively generate their MS/MS spectrum. Furthermore, when the nanoelectrospray technique was introduced, most mass spectrometers did not provide automated control of the different experiments (i.e., data-dependant experiments), thus frequent user interventions were required for the analysis of peptides (Shevchenko et al., 1997). This means that on the mass spectrometers, the operator had to generate and interpret the MS spectrum, select the peptides from the MS spectrum, switch the mass spectrometer to collision-induced dissociation, and then successively generate the MS/MS spectrum of the ions by manually changing the parameters for every ion. Even for the most experienced user, this can easily take 5 to 15 min to perform per peptide.

In the continuous-flow infusion nanoelectrospray, the peptide mixture of interest is inserted into a glass needle of 1 mm in diameter that has been pulled to a closed and tapered end using a pipette puller. Furthermore, the pulled needles are also coated with a layer of conductive material such as gold, to which a high voltage is applied (Fig. 1.14). The needle is installed in front of the mass spectrometer, and a slight pressure is applied. Then, the end of the needle is opened by touching it against the front plate of the MS, producing an open tip of only 1- to 10-µm inner diameter and a slightly bigger outer diameter. Alternatively, a nanospray needle that is tapered down, but not closed, can also be purchased. The peptide mixture is then continuously delivered at the end of the needle by a gentle gas pressure. The peptide solution that reaches the tip of the nanoelectrospray needle is electrosprayed into the mass spectrometer at low flow rates

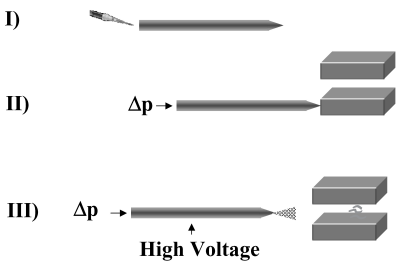


Figure 1.14. Schematic representation of the nanoelectrospray analysis of peptide mixtures. The samples of interest are pipetted into a nanoelectrospray needle. The needle is then mounted in front of the mass spectrometer. The distal end of the needle is open by applying a slight gas pressure at the other end of the column and gently touching the front plate of the mass spectrometer. Once, the open needle is positioned to electrospray into the mass spectrometer.

(few nanoliters/minute). Roughly 30 min to 1.5 h analysis can be performed on 1 μl of sample.

Separation Coupled to ESI-MS/MS. Electrospary ionization can also be coupled with separation techniques. This is often preferable for low levels of samples and for contaminated samples. Furthermore, separation techniques are readily automatable, while the nanoelectrospray approach has evaded automation due to its inherent skill requirement. Separation techniques offer the advantage of concentrating the analytes into shorter and separated analytical windows. While nanoelectrospray presents all the analytes at the same time to the mass spectrometer, separation systems subsequently present the analytes to the mass spectrometer; however, a very narrow analytical window (typically 1 to 30 s) is available for every analyte, during which all mass spectrometric measurements for an analyte must be performed.

Automated Mass Spectrometers. Evidently, the manual triggering of the mass spectrometer, as conventionally done in nanoelectrospray, is not possible with separation techniques. Therefore, the utilization of an automated mass spectrometer is essential for the success of these experiments. A set of standard data-dependent features has emerged. (1) The mass spectrometer should offer the automated selection of precursor ions based on a threshold or a signal-to-noise ratio. (2) The *n*th-most intense ion should be selectable for fragmentation as predefined by the user. (3) A static exclusion list of ions should be predefinable. The static exclusion list allows the exclusion of known contaminants, such as clusters, trypsin autolytic peptides, and other known peptides. (4) The ions for which *m* number of MS/MS spectra have been generated should be added to a dynamic exclusion list until their intensity in MS mode falls below a threshold value or until a time period has elapsed. The dynamic exclusion of ions allows the selection of the next most intense ion for the next round of MS/MS spectra. (5) The dynamic isotope exclusion is also an important option, especially for higher resolution instruments. This feature ensures that the isotope peaks from an ion do not trigger the

acquisition of an MS/MS spectrum. (6) The exclusion of charge states, such as the 1+ ions, can also be convenient because it avoids the generation of MS/MS spectra for 1+ ions, which are typically not very informative. (7) The automated selection of fragmentation energy is an important aspect for adequate fragmentation of peptides. Although most mass spectrometers utilize collision-induced dissociation to fragment peptides, the ways in which the kinetic energy is ascribed to the precursor ion are different. In triple quadrupole and hybrid quadrupole TOF mass spectrometers, the fragmentation pattern is only adequate in a limited kinetic energy range. Excess energy increases the internal fragmentation of the precursors and causes poor transmission of ions through the collision cell, thus resulting in a poor-quality MS/MS spectrum. Therefore, the automated adjustment of the collision energy is an important feature on these mass spectrometers. The fragmentation pattern is less sensitive to the kinetic energy on the ion trap and cyclotron mass spectrometers, as long as sufficient energy is provided. Therefore, the automated adjustment of the collision energy is less important in ion trap and cyclotron mass spectrometers.

High-Performance Liquid Chromatography–MS/MS. High-performance liquid chromatography has been the technique of choice for the separation of analytes online with ESI mass spectrometers (LaCourse, 2000). In reverse-phase mode, HPLC concentrates and separates analytes according to their hydrophobicity. Different approaches have been developed to introduce peptides into a mass spectrometer by HPLC. They all have in common peptide solutions that are loaded and separated on an HPLC column made of C18-like material. They differ, however, in terms of their flow rates, their sample paths, robustness, and sensitivity.

Conventional HPLC. Conventional microbore HPLC coupled to ESI-MS is a well-established approach for the identification of protein levels higher than picomoles (Hunt et al., 1981, 1986; Gibson and Biemann, 1984). Microbore HPLC is usually achieved using a 0.3-mm inner diameter (i.d.) HPLC column installed on a low-flow HPLC system. The flow rate generated through these columns is typically in the low microliter per minute range. The integration of an autosampler capable of handling less than 10 μL of sample completes the automation of the procedure. Therefore, sequential introduction, separation, and analysis of protein digests on this automated system can be routinely performed.

Microflow HPLC. High-performance liquid chromatography is a concentrating technique, while mass spectrometry is a concentration-dependent device for constant flow rates. Therefore, improving the concentration of analytes delivered to the mass spectrometer provides better sensitivity. To improve the concentration when a limited amount of analytes is available, only two approaches can be taken. The first approach consists of improving the separation and decreasing the peak width of the eluting analytes. Although this is feasible, practically it does not provide sufficient improvement in sensitivity. The second approach consists of decreasing the size and flow rate of the column. For example, reducing the size of an HPLC column from 1 mm to 100 μm in diameter offers an improvement by a factor of 100 in the concentration of the analytes.

This is a very attractive alternative for low-level samples. It is important to keep in mind that the HPLC theory also predicts that the volume amount of injected analytes should be correspondingly reduced. In reality, there is a tendency to disregard this warning and to inject as much volume as possible at the price of losing the lower affinity analytes.

Although attractive, low-flow HPLC (<200 nL/min) was not directly compatible with conventional electrospray. Earlier attempts to couple low-flow HPLC with conventional electrospray interfaces by using sheath liquid were successful; however, they did not demonstrate the expected improvement in sensitivity. Fortunately, microelectrospray interfaces compatible with the low-flow technique provided by μ-HPLC columns were developed, and, these greatly improved the usefulness of this technique (Chervet et al., 1996; Yates et al., 1996).

Fabrication of Columns. Recently, microcolumns (150 μm i.d. or less) have been commercially available; however, the range of available packing materials and the cost of these columns are often limited. Therefore, it is still common to fabricate in-house columns. Two standard techniques have been developed for the fabrication of microcolumns.

The first technique consists of fabricating a column in a capillary tubing with a constant inner diameter. In this design, small borosilicate beads are dry packed at the end of a capillary tubing of 50 to 150 μm i.d. by tapping the capillary in an aliquot of beads. Typically, about 1 mm of the capillary tubing is filled with beads. The end that contains the beads is rapidly sintered in a Bunsen burner. This results in the formation of a small porous frit at the end of the capillary tubing. It takes some practice to appropriately sinter the beads. The packing material for the column is pressure forced at the other end of the capillary in the form of a slurry. This requires the utilization of a bomb that can be pressurized up to a few 1000 psi. Alternatively, the slurry can be installed in a large volume union and pushed into the column using an HPLC pump. The pressure is operated until the required length of column packing is achieved; then it is slowly reduced to atmospheric pressure. It also helps to place the part of the column being packed in a sonification bath. The end result is a packed capillary column (typically 5- to 10-cm of packing) terminated by a frit. The disadvantage of this approach is that it does not include a microelectrospray needle. Typically, the fritted end of the column is installed in a low dead-volume union connected to a micro-ESI needle. To minimize the band broadening, a short needle that has a smaller inner diameter than the capillary column and a reduced diameter at the tip (5- to 15-μm opening) is utilized to generate the electrospray (see Fig. 1.15).

The column can be directly formed in a microelectrospray needle to avoid the dead-volume issues of the packed capillary approach. This requires a frit of some sort to be fabricated in the tip of a microelectrospray needle. Typically, a needle of 50 to 150 μm i.d., sharply terminated at the tip to 5 to 15 μm i.d., is used for the fabrication of the column. These needles are made from a capillary tubing that is stretched using a laser pipette puller. Then again, the needle can be pulled by placing a capillary tubing in a vertical position, attaching a weight to it, and heating a section of the capillary with a flame. Both of these needle pulling techniques can be tedious to perform, while creating a frit at the end of the needle can be even more challenging. Therefore, the best

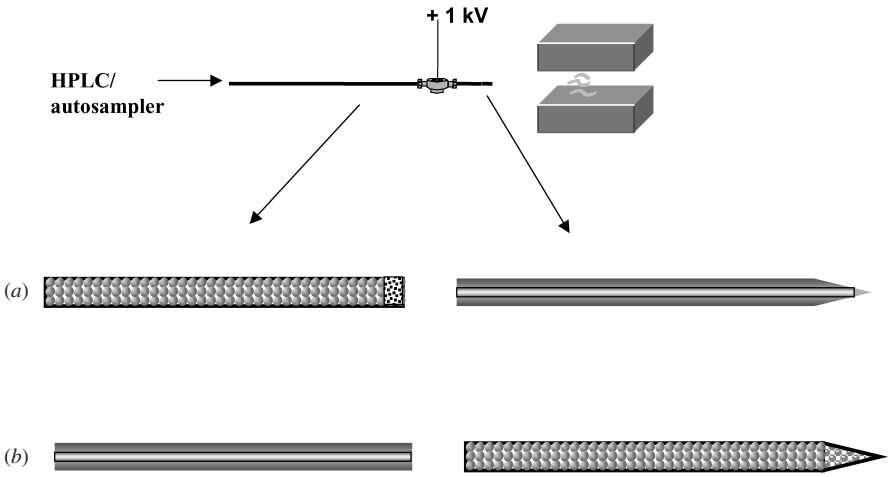


Figure 1.15. Schematic of nanoflow HPLC-ESI-MS/MS. (*a*) In this design an HPLC column is fabricated in a capillary tubing (50 to 150 μm i.d. × 5 cm long) terminated with a sintered glass frit. This capillary column is then connected to a low inner diameter nanoelectrospray tip by a low-dead-volume union. (*b*) In this design a low inner diameter capillary tubing is used to connect an HPLC pump to a low-dead-volume union. An HPLC column is then fabricated in a (50 to 100 μm i.d.) nanospray needle with a tip inner diamater of 5 to 15 μm. The HPLC needle is then installed at the other end of the low-dead-volume union.

alternative is to buy microelectrospray needles (e.g., New Objectives needles) with a frit already fabricated at the tip, and then to pack the column with the reverse-phase material of choice. The packed needle is then installed on a low-dead-volume union also connected to a nanoflow HPLC system (see Fig. 1.15).

More exotic techniques have been developed for the fabrication of columns (Fig. 1.16). In particular, needles without frits have been packed with a reverse-phase material. This was done either by pulling long needles of progressively decreasing inner diameters or by creating an isthmus by differentially pulling the capillary tubing (Martin et al., 2000). These two methods allow the needles to be packed without having to create a frit. Davis and Lee (1998) used a laser-pulled microelectrospray needle (150 μm i.d.) with a 5-μm orifice and introduced into the needle a short piece of fused silica capillary [25 μm i.d. × 150 μm outer diameter (o.d.)]. The piece of capillary was pushed as far as possible in the needle and then backed with a membrane frit. The reverse-phase material was then packed behind that frit. Although these approaches are interesting alternatives for HPLC needle fabrication, they have not yet seen widespread application and thus remain the forte of only a few groups.

More recently, we have seen the appearance of monolithic columns that are fabricated by direct polymerization of the stationary phase inside the capillary tubing. These monolithic columns are obtained by the in situ polymerization of a mixed polymer such as styrene and divinylbenzene. The polymerization occurs in an oven at a controlled temperature. The advantage of the monolithic column is its lower "multiple-path" effect, which provides a good number of theoretical plates and a good resolution for

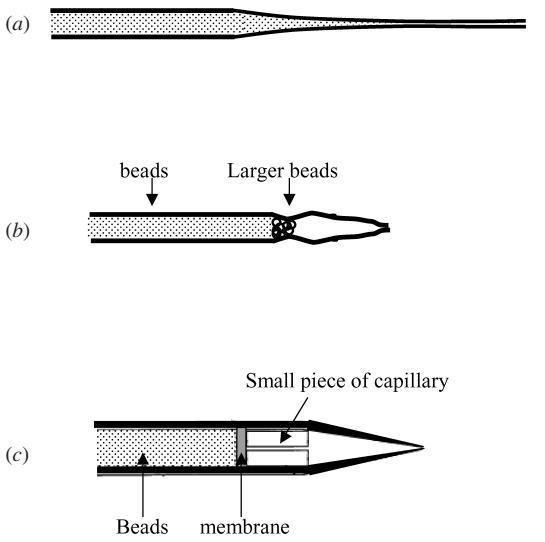


Figure 1.16. Schematics of different packed nanospray needles that have been fabricated. (*a*) Needle with no frit obtained by gradually narrowing down the inner diameter of the needle. (*b*) Needle with an isthmus to retain the reverse-phase beads (Martin et al., 2000). (*c*) Needle with a capillary frit to retain the beads (Davis and Lee, 1998).

short column beds. Comparisons with conventional packed columns have indicated similar, if not better, separations. Other advantages of the monolithic columns are that they are easier to fabricate and to mass produce than the conventional packed HPLC columns. The disadvantage, however, of the monolithic column is that it requires an organic synthesis, which has to be carried out with caution. Therefore, this approach is less amenable for the in-house and widespread fabrication of columns by research laboratories.

High-Performance Liquid Chromatography Setup. The microcolumn system would not be complete without the addition of an HPLC gradient delivery system. The reverse-phase microcolumn typically requires a flow rate of 100 to 200 nL/min. Although conventional HPLCs can deliver such flow rates in isocratic mode, they cannot deliver gradients or provide efficient mixing at that flow rate. The simplest solution to this problem has been to "flow split" the eluent from an HPLC system that pumps at 5 to 50 nL/min down to 100 to 200 nl/min. This is typically achieved by adding a tee after the HPLC system and by having one of the two exits going to waste while the other exit is connected to the microcolumn. The split ratio is obtained by carefully gauging the tubing at both exits. The advantage of this method is that it can be easily adapted to conventional HPLC. Its disadvantage is that the split ratio is affected by the back pressure on the microcolumn. Therefore, as the column ages and the back pressure increases, the split ratio changes. Regardless of the disadvantages, the splitting method is still the preferred approach for the generation of nanoliter per minute gradi-

ents. In fact today's commercially available nanoflow HPLCs are based on the splitting principle.

High-performance liquid chromatography systems can be divided into two broad types: low-pressure mixing systems that mix the solvents prior to a single pump and high-pressure mixing systems that mix the solvents in a static or dynamic mixer placed after the pumps, or they can mix the solvents directly into the pump head. For both systems, the important parameters for μ-HPLC-MS/MS are the delay time in the gradient, the prepressurization performance of the pump, and their performance at lower pressures.

The low-pressure mixing systems can have low void volumes following the pump because no mixers are placed after the pump. Therefore, flow splitting can be performed right after the pump, reducing the dead time of the experiment; however, low-pressure mixing systems have the disadvantage of obtaining the solvent ratios by mechanically proportioning the solvents. This happens either by switching valves or by proportional valves, using the vacuum action of the pump refill stroke. These approaches can perform poorly for low and high solvent ratios, generating less reproducible gradients.

The more common HPLC systems are the ones that perform high-pressure mixing of solvents and have the mixer placed after the pump or directly in the pump head. In these systems, the pumps are driven at a few microliters per minute, and the flow splitting required to achieve 100 to 200 nL/min occurs after the mixer and the pressure damper. Generally, the void volume is fairly large from the pumps to the column. For example, a void volume of 100 μL is not unusual and can represent 20 min delay in the gradient, assuming a flow rate of 5 μL/min. Furthermore, the back pressure can be as little as 100 to 1000 psi, which can cause some check valves used in HPLC to fail to operate properly.

Coupling to the Electrospray Mass Spectrometers. In recent years three approaches have emerged for the coupling of μ-HPLC columns to ESI-MS. The first approach consists of establishing an electrical contact directly at the end of the electrospray needle. This is typically achieved by gold coating a part of the tip. This allows the direct application of a high-voltage potential to the tip of the column, which is necessary for the generation of the ESI process. The direct application of the high voltage generates, by far, the most stable electrospray process; however, the thin gold layer applied to the tip will be removed over time, and the electrical connection can be lost.

Alternatively, the high voltage can be indirectly applied to the electrospray needle by using a liquid junction before the needle. This is typically achieved using an ultra-low-dead-volume connector placed between the transfer line and the electrospray needle. This approach has the advantage of being more robust in terms of maintaining the electrical connection; however, it has the disadvantage of generating a less stable, although still adequate, electrospray process.

The last approach is a combination of the previous two approaches, and it consists of using a fully external gold-coated needle coupled to a ultra-low-dead-volume connector to which a high voltage is applied. In this fashion, the electric connection is always maintained while an optimum electrospray process is obtained.

The electrical liquid junction in commercial connections or home-made connections can cause debilitating problems when used improperly. It is important to understand that electrolysis does occur in the liquid junction, generating a small amount of gas. Generally, the linear flow is strong enough to carry the gas forward with no formation of bubbles; however, in improperly fitted connections, the linear flow could be reduced when entering the liquid junction to a level that allows the formation of gas bubbles.

Application to Protein Identification Based on Peptide Mass Spectrometry. Once all the elements of the system are in place, the protein identification experiment can be performed. In a typical experiment, a few microliters of the protein digest is pressure loaded onto the μ-HPLC column. Then, the HPLC system delivers a solvent gradient of increasing hydrophobicity to the column. The peptides are eluted in order of hydrophobicity, transferred to the microelectrospray interface, and then moved to the MS. The eluting analytes successively trigger the MS to select one of the analytes and to generate the MS/MS spectra. The MS/MS spectra generated for the analytes are then used to search protein databases.

This method has established itself over the years as the method of choice for protein identification by mass spectrometry. Although, the continuous infusion by nanoelectrospray has been shown to provide exquisite sensitivity, it still remains a slow and manual technique. HPLC offers the advantages of being rapid and automatable, while still providing comparable limits of detection. The most popular combination has been to couple μ-HPLCs with ion trap mass spectrometers. Figure 1.17 shows the analysis of 50 fmol of bovine serum albumin (BSA) on such a system. The analysis was completed in about 35 min. The insert shows the typical quality of fragmentation patterns that can be obtained at that level of protein. It has been reported that approximately 100 fmol of protein present on the gel can be analyzed with this system.

Two-Dimensional HPLC-MS/MS Techniques. One-dimensional HPLC-MS/MS does not provide a high enough resolution for complex samples. Although longer gradients can be performed, one-dimensional HPLC has an inherent separation limitation. Historically, it has not been a significant problem because the complexity of the samples was always reduced through 1D and 2D gel electrophoresis. The limitations, however, of gel separation have pushed research toward novel two-dimensional separation systems.

Jorgenson's group has been pioneering the multidimensional separation from the analytical point of view (Hooker and Jorgenson, 1997; Lewis et al., 1997; Moseley et al., 1989; Opiteck et al., 1997, 1998). Techniques have been recently developed for online multidimensional chromatographic separation of peptides and their subsequent analysis by mass spectrometry. In particular, multidimensional chromatographic approaches, based on ion exchange chromatography, have been coupled to reverse-phase chromatography online with an ESI mass spectrometer for protein identification (Washburn et al., 2001). This system allows the analysis of complex peptide mixtures, such as the mixtures obtained by proteolytic digestion of whole-cell lysates.

Setup. A bed of reverse-phase material is packed in a capillary tubing followed by a second packing of a bed of ion exchange material. The end result is a column con-

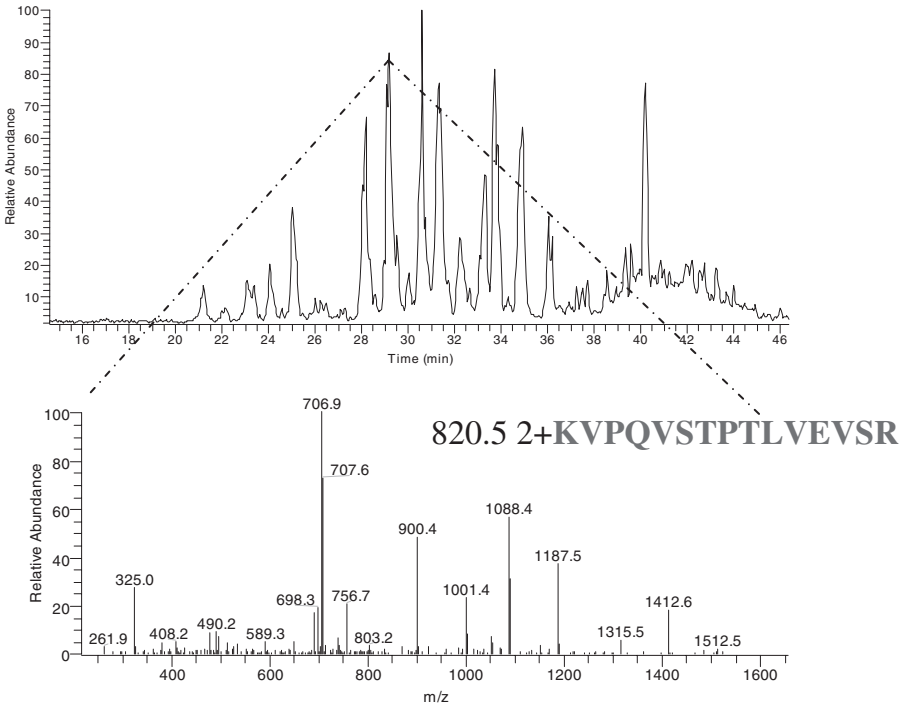


Figure 1.17. Example of the analysis of protein digest by HPLC-ESI-MS/MS. (*a*) Base peak chromatogram for 50 fmol of BSA tryptic digest analyzed by HPLC-ESI-MS/MS on an electrospray LCQ ion trap mass spectrometer using a 75 μm i.d. × 360 μm o.d. × 5 cm long C18 packed needle column. (*b*) The inserts show an MS/MS spectra automatically generated for $m/z = 820.5$ of charge 2+. Sufficient MS/MS spectra were generated to identify the protein by searching protein sequence databases.

sisting of two beds of different materials. This system is then used to address the complexity of large-scale protein digests, such as the ones obtained from the proteolytic digestion of whole-cell lysates.

The peptide mixture is then loaded on the dual-mode column and accumulates on the ion exchange resin. Groups of peptides are then subsequently eluted off the ion exchange resin by a ionic step gradient. In reality, the ion exchange material is used only for the extraction of the peptides, and it does not provide efficient separation; however, during each step of the ionic gradient, the eluted peptides are captured at the head of the reverse-phase bed. Each step in the ionic gradient is followed by a rinse and an organic-phase gradient to perform the reverse-phase separation of the peptides. The peptides are then separated on the reverse-phase column and eluted to the ESI mass spectrometer. The mass spectrometer is used to detect the peptides and perform collision-induced dissociation of selected peptides.

Microfabricated Devices Coupled to Mass Spectrometer. The application of microfabricated devices has dramatically increased in the 1990s, and it has grown from an

esoteric technique only applied in a few laboratories to a recognized technique. Its application in the field of proteomics is more recent because of the difficulty to couple microfabricated devices to an external detector such as a mass spectrometer. In this section, we will describe some of the microfabricated systems that have been used for the identification of protein by mass spectrometry. It is, however, important to remember that all these systems are far from being robust and still have some serious hurdles to overcome. Therefore, at this point, use the information provided below as an eye opener to a novel technology that might, one day, be a strong platform for proteomics.

Microfabricated devices are typically made by photolithographic techniques to create various patterns of reservoirs, channels, and reaction chambers on a planar substrate such as glass or polymer. Then a second plate, with access ports for the reservoirs, is bonded to seal the channels. These devices were originally designed to have in situ detection of the analytes. The analysis, however, by mass spectrometry requires the transfer of the analytes to the mass spectrometer. In the case of ESI mass spectrometers the standard microelectrospray interface can be used for the transfer; however, the interface still needs to be connected to the microfabricated device through a transfer capillary. This is still the most challenging aspect of interfacing microfabricated devices to MS analyzers. Various ways of making the connectivity have been published; however, they are mostly incompatible with the mass production of the devices.

Continuous Infusion of Peptides for Protein Identification by Mass Spectrometry. The simplest application of microfabrication is the continuous infusion of peptide digests to a mass spectrometer and the fragmentation of the individual peptides. For example, Figeys et al. (1997) performed protein identification via infusion from a microfabricated device into an MS. In their design, a 12-cm pump capillary was connected to a simple three-reservoir device and to a micro-ESI source. The device contained two independent sample reservoirs and a buffer reservoir. The sample mobilization and their direction were done by controlling the voltage at each of the three reservoirs. In this fashion, one sample can be mobilized toward the MS, while the other sample is retained in its reservoir through an unfavorable electric field. The mass spectrometer is then used to provide MS/MS analysis of the eluting peptides. They demonstrated that reasonable levels of standard peptides (33 fmol/μL) can be analyzed through this system while retaining a high signal-to-noise ratio.

In a subsequent implementation, Figeys et al. (1998b) put into practice a more complex nine-reservoir device with computer-controlled voltage applied to each reservoir (Fig. 1.18). Furthermore, a software was developed to control the chip device and to trigger the mass spectrometer for automated acquisition of the peptide MS/MS. The only user intervention that was required was the manual pipetting of samples into their individual reservoirs on the microfabricated device before triggering the software. The system sequentially mobilized the sample using directed electroosmotic pumping, while the MS software controlled the data acquisition and subsequent database searching. The system was successfully used to identify yeast proteins previously separated by 2D gel electrophoresis.

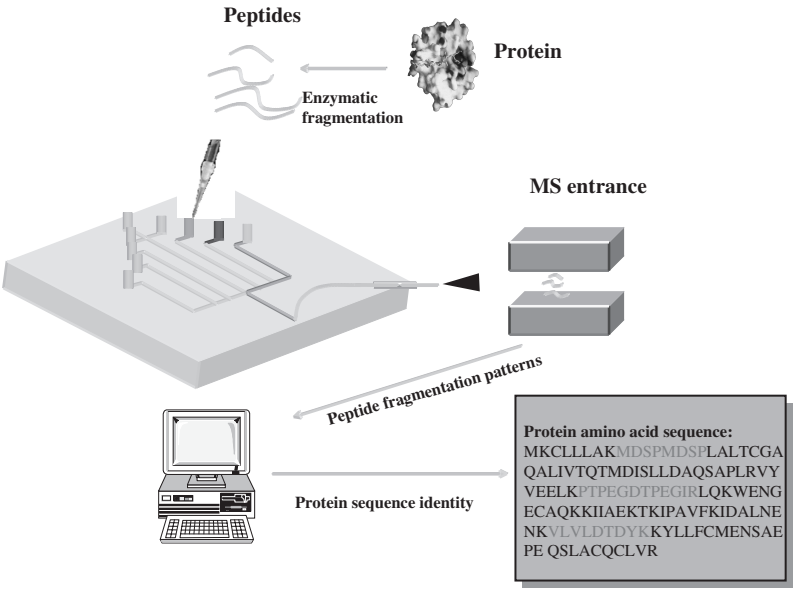


Figure 1.18. Schematic of a nine-position microfabricated device coupled to an ESI-MS/MS mass spectrometer for the automated analysis of protein digests. The protein digests of interest are pipetted into the different reservoirs. Then, the samples are successively mobilized by the application of a directed electric field from their specific reservoir toward the mass spectrometer where the peptides are analyzed by MS and MS/MS. The analysis proceeds until all the samples have been successively analyzed. Reproduced with permission from Figeys et al. (1998b).

Separation of Peptides on Microfabricated Devices Coupled to Mass Spectrometer. Electrophoresis has been the method of choice for the separation of analytes on chips. Furthermore, it is also compatible with the microelectrospray interface. The same limitations, however, that were present in CZE are also present in the microfluidic system. Therefore, discontinuous methods such as sample stacking, isotachophoresis, and solid-phase extraction (SPE) are required to increase the amount of sample injected on the microfabricated device.

Electrophoretic separations with offline sample concentration (Zhang et al., 1999) and online sample concentration (Figeys et al., 1998a; Li et al., 1999; Wang et al., 2000) have been reported. In particular, a microfabricated system for the separation of peptides by electrophoresis was coupled to a mass spectrometer by Li et al. (2000). In this approach, the protein digests of interest are introduced one at a time on the microfabricated device. An electric field is applied to fill an injection cross with the sample. Once the cross has been filled, an electric field is applied from the buffer reservoir to the microelectrospray needle driving the separation of analytes toward a microelectrospray interface. Rapid separation of protein digests has been obtained through this approach; however, the limit of detection for real samples still needs to be improved (Fig. 1.19).

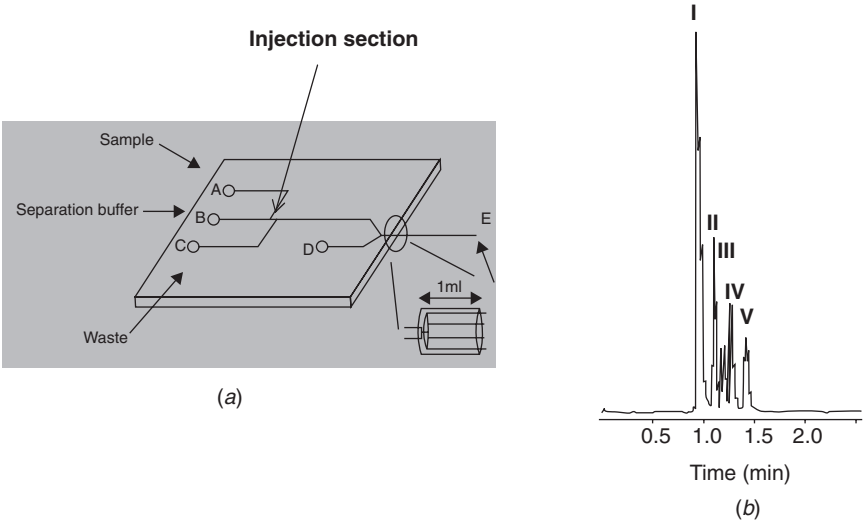


Figure 1.19. Schematic of a four-position device coupled to an ESI-MS/MS mass spectrometer for the separation of peptide mixtures and their analysis by mass spectrometetry. (*a*) Schematic of the device coupled to a triple quadrupole mass spectrometer. A portion of the sample is mobilized from the sample reservoir toward the waste, filling a small injection line. Once the injection section is filled, a potential is applied between reservoir B and the nanospray emitter driving the separation and the electrosrpay process. (*b*) Example of a separation for a peptide mixture in less than 2 min. Separation of leu-enkephalin (I), LHRH (II), somatostatin (III), angiotensin II (IV), and bradykinin (V) all at 10 μg/mL. Reproduced with permission from Li et al. (2000).

The development in the field of microfluidics coupled to mass spectrometry has been tremendous in the late 1990s. There are still, however, some serious challenges to the technology that still need to be addressed before it can become routinely applicable in a robust format.

Bioinformatics for the Identification of Proteins Based on MS/MS

All the ESI-based tandem mass spectrometers are generally used to perform MS and MS/MS spectra. Although the mass accuracy, resolution, sensitivity, and quality of the fragmentation patterns are different, they contain information that can be used to identify a peptide provenance by protein/DNA database searching. The MS/MS spectra contain the fragmentation patterns related to the amino acid sequence of specific peptides. The analysis of MS/MS spectra is more intensive than the interpretation of MS data. The approaches that are used for the interpretation of these spectra can be classified into three subgroups according to the level of user intervention required.

No Interpretation. Clearly one would like to be able to identify proteins without having to do any interpretation of the MS/MS spectra. For high-throughput analysis,

this becomes essential. A few algorithms have been developed to search protein/DNA databases with uninterpreted MS/MS spectra. They all have in common the requirement for the partial or full sequence of the protein to be already known and included in a database. All of the software packages provide a list of possible matches between individual MS/MS spectrum and peptide sequences obtained from the database; however, the scoring algorithms used to determine the validity of the matches are very different. These algorithms are explained in more detailed below.

Mascot by Matrix Sciences (*www.matrixscience.com*) can be freely accessed over the web (for noncommercial entities) (Pappin et al., 1993). It was built based on the MOWSE scoring algorithm from Papin. The MOWSE algorithm computes a matrix of probability in which each row represents an interval of 100 Da in peptide mass, and each column represents an interval of 10 kDa in intact protein mass. As each sequence entry is processed, the appropriate matrix elements are increased to accumulate statistics on the size distribution of peptide masses as a function of the protein mass. Therefore, this matrix represents the peptide distribution by protein MW present in the database. The matrix will change as the database increases. From this matrix, a score is ascribed based on the measured masses. The same algorithm is used for peptide MS/MS. In Mascot they have further incorporated a probability based on the probability that a match is a random event. This allows the probabilistic evaluation and ordering of all the potential matches, as well as a probabilistic evaluation of the best match.

ProteinProspector from UCSF (*http://prospector.ucsf.edu/*) is also an example of a web-based MS/MS search engine. The identification of the protein is typically unambiguous, achieved by the number of peptides that matches to the same protein.

Another algorithm that is popular is Sequest (Eng et al., 1994; McCormack et al., 1997; Yates et al., 1995). Sequest is not freely accessible over the web and must be purchased. This algorithm performs a two-pass search of the database. In the first pass, for every MS/MS spectra submitted, Sequest searches protein/DNA databases for the top 500 isobaric peptides using some of the information on immonium ions. In the second pass, the 500 predicted spectra corresponding to the isobaric peptides are generated and are rapidly matched against the measured spectra. This is achieved by multiplying in the frequency domain of the fast-Fourier transformation, the measured spectrum with each individual predicted spectrum for the 500 isobaric peptides, and by pairwise multiplication of the measured and predicted transforms.

The multiplication in the Fourier domain is the same as doing a convolution in the time domain. This means that this approach verifies how well each predicted spectrum matches to the measured spectrum. Correlation parameters that indicate the quality of the match between predicted and measured spectra are then deduced from the results of the transform multiplication. A high cross-correlation indicates a good match between the predicted spectrum and the measured spectrum. More importantly, the cross-correlation value is independent of the size and nature of the database. Therefore, the same peptide sequence found in a small database or a large database will return the same correlation coefficient when matched against the measured spectrum. This is an important aspect of Sequest that makes it unique. All the other approaches provide scoring schemes that are very dependent on the size and nature of the database.

Furthermore, although protein identification has been performed with as little as one peptide using this algorithm, unambiguous identification of the provenance of a

protein is often achieved by the multitude of peptides that matches with the same entry in a database. The Sequest software is computing intensive, and for high-throughput demand, it can rapidly paralyze the best dual-CPU server. In reality, the slowness of Sequest is due to the recurrent scans of the selected database to find the top 500 isobaric peptides. The larger the database, the longer it takes to scan the databases. An improved version of the software, called Turbo-Sequest, predigests and orders the databases and has greatly improved the search speed.

Sonars by Proteometrics can search uninterpreted MS/MS spectra of peptides against protein and DNA databases (Field et al., 2002). This software uses Bayesian theory to rank the protein sequences in a database by their probability of occurrence. It also provides a scoring scheme for the differentiation of the match. In our hands, it has proven to be the most rapid system for the analysis of peptide digests. X!Tandem (*http://www.proteome.ca/x-bang/tandem/tandem.html*) (Graig and Beaivis, 2004).

Partial Interpretation. Although the fully automated approaches are favorable for automation, historically the computing power was not available for the rapid identification of proteins.

Another subgroup of database search engines, based on the partial interpretation of the MS/MS spectra, were developed to perform faster searches of databases, while requiring human intervention. The most popular partial interpretation of MS/MS spectra is the "sequence-tag" approach (Wilkins et al., 1996a; Mann and Wilm, 1994) (Fig. 1.20). It consists of reading the mass spacing between specific fragments of an MS/MS spectrum. This allows the generation of a short peptide sequence (tag). The tag is then used to pinpoint the possible peptide sequences from isobaric peptides in databases, while the residual mass information before and after the tag confirms one or a few peptide matches. Every MS/MS spectrum requires the generation of a tag followed by database searching. Unambiguous identification of the protein is established by the multitude of peptides that matches to the same protein.

De novo Sequencing. Finally, the last option is the full interpretation of the MS/MS spectra, often called de novo sequencing (Papayannopoulos, 1995; Shevchenko et al., 1997). Obviously, the other automated and semiautomated approaches are used prior to this approach. Although many genomes have been recently sequenced, the genomic data only represent a fraction of the world's genomic pool, and often no databases are available for specific organisms. The requirements for de novo sequencing are more stringent. The MS/MS spectra must be of good quality in terms of intensity and coverage of the peptide sequence. The MS/MS spectra of peptides contain ladder-type information, which, in principle, indicates their amino acid sequence. Experienced mass spectrometrists can manually extract the peptide sequence from the MS/MS spectra.

In large-scale proteomic studies, the throughput of analysis is a critical factor. Therefore, once enough MS/MS spectra have been generated to unambiguously identify a protein, generating MS/MS spectra on the residual peptides can be a waste of time. In some cases, however, like when dealing with expressed sequence tag (EST) or protein mixtures, it is important to increase the number of MS/MS generated.

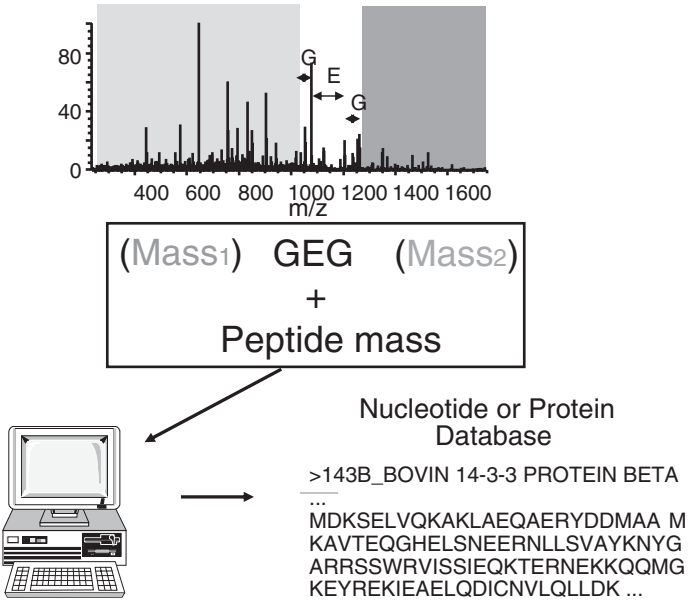


Figure 1.20. Schematic of protein identification by sequence tag. A small stretch of an individual peptide sequence is read from the MS/MS spectrum. The peptide mass, the short amino acid sequence, and the residual masses before and after the sequence are used to search protein/DNA databases.

CURRENT CHALLENGES IN PROTEOMICS

Classical Proteomics

There has been a lot of confusion in the field of proteomics on the issue of the complete proteome coverage and low-abundance proteins. This confusion is, in part, due to a bit of propaganda about the power of the technology, as well as some confusion related to the definition of *proteome* and *proteomics*. The idea of the comprehensive study of the proteome is more an idealism than a reality.

Also, 2D gel-based proteomics is still facing a serious number of challenges. In particular, although 2D gel electrophoresis is a powerful technique, it still lacks the capability for discovering hydrophobic proteins, basic proteins, and low-abundance proteins.

The presence or absence of low-abundance proteins on 2D gels has been controversial. Fortunately, for *S. cerevisiae* and a few other species, indexes can be calculated for all the ORF entries in their corresponding database that reflects the abundance at the protein level. One such index is the codon bias index. The codon bias index is calculated based on a compiled index for codon usage in messenger RNA (mRNA). The codon usage is typically measured for a few ORF that are translated into highly expressed proteins. The codon bias index is thought to be a good indication of the expression levels of proteins. Codons that correspond to low-level transfer RNAs

(tRNAs) during translation decrease the potential yield of expression; on the other hand, codons that utilize higher abundance tRNA during translation will increase the potential yield of the protein expression. Figure 1.21*a* shows the expected range of codon bias in yeast calculated using the predicted ORF. Clearly, a good portion of the yeast proteome is expected to fall in the low codon bias (<0.1) and therefore, has a low expression. Figure 1.21*b* shows a compilation of the codon bias index for the proteins that have been identified by 2D gel electrophoresis of *S. cerevisiae* [compiled from Gygi et al. (1999b) and Perrot et al. (1999)]. At a first glance, low-abundance proteins appear to be detected; however, if we only plot the proteins that were identified by mass spectrometry, the result is different. Furthermore, some of the reported proteins with 0.1 codon biases are suspicious because they represent proteins that were present at the picomole level on the yeast 2D gel, and this level is far from being low abundance. Therefore, they probably represent erroneous identifications (Fig. 1.21*c*). The lack of identification of low-abundance proteins has been reported to be caused by a limited load capacity on 2D gel electrophoresis. In order to see low-abundance proteins, the amount of sample that needs to be loaded exceeds the capacity of 2D gel electrophoresis.

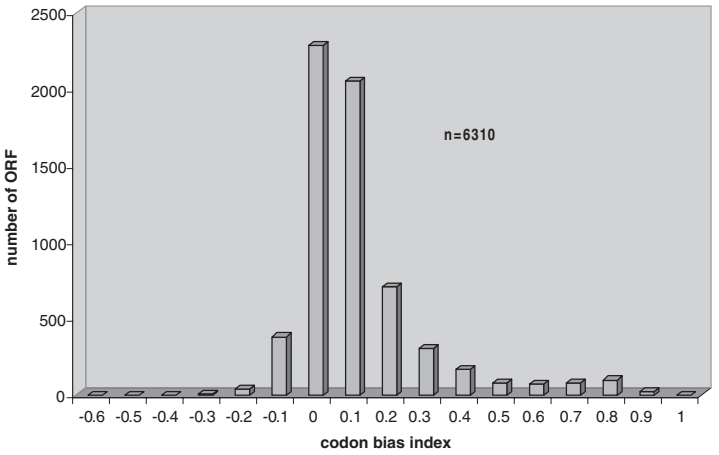
Hydrophobic proteins have also been a challenge for proteomics. Hydrophobic proteins have the tendency to precipitate once they reach their pI during the electrofocusing on an IPG, and therefore they do not transfer to the second dimension. Although, efforts have been made to improve the solubilization of hydrophobic proteins (Rabilloud, 1999; Santoni et al., 1999), the display of hydrophobic proteins by 2D gel electrophoresis still remains a challenge.

The efficient gel digestion of low-level proteins can, as well, be a challenge for proteomics. The kinetic of digestion is dependent on the concentration of the substrate (protein to be digested). Therefore, as the concentration of substrate decreases, the efficiency of the digestion also decreases. Furthermore, sample loss is also a critical factor for low-level proteins.

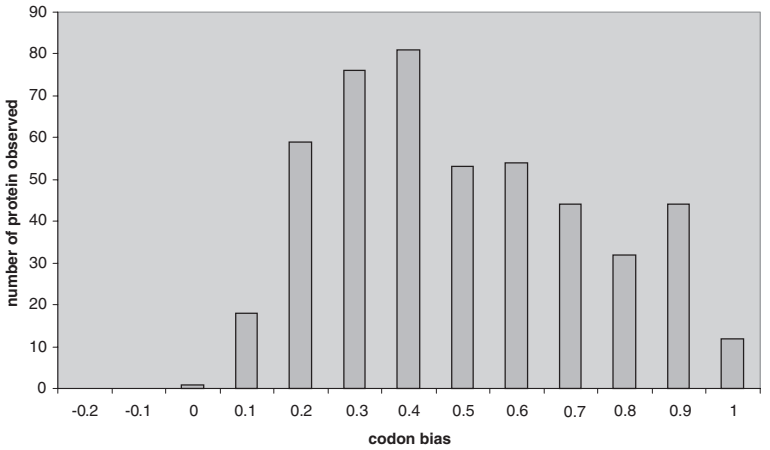
Gel-Free Analysis: An Alternative Differential Display

The limitations of 2D gel electrophoresis can be alleviated with the gel-free analysis of complex protein mixtures. The analysis of complex proteolytic mixtures is now possible due to the improvement of online peptide separation and the improvement of mass spectrometry software. In this approach, the lysate of interest is directly digested to provide a complex mixture of proteins. This offers the advantage of reducing sample lost and contamination, therefore expanding the dynamic range of protein analysis toward lower abundance proteins. Characteristic hydrophilic peptides can also be found from proteins that have a sequence in preponderance of hydrophobic amino acids. The

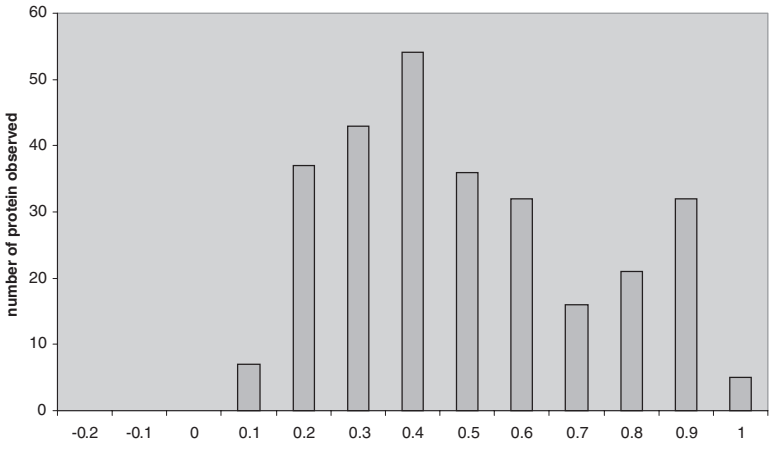
Figure 1.21. Codon bias and the yeast proteome. (*a*) Distribution of codon bias calculated for all the yeast ORF. (*b*) Compilation of the codon bias for the proteins that have been observed on 2D gel electropherograms of yeast. (*c*) Compilation of the codon bias for the proteins identified from a 2D gel electropherogram of yeast using only mass spectrometry.


2500
2000
1500
1000
500
0
number of ORF
n=6310
-0.6 -0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
codon bias index

(*a*)


90
80
70
60
50
40
30
20
10
0
number of protein observed
-0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
codon bias

(*b*)


60
50
40
30
20
10
0
number of protein observed
-0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
codon bias

(*c*)

complex mixture of proteins is then separated online either by one- or two-dimensional HPLC and ESI mass spectrometry.

Quantitation. The relative quantitation of the proteins that are differently expressed between two sets of proteins is an important aspect of proteomics. The gel-based separation of proteins provides direct visualization and relative quantitation of separated proteins based on the staining intensities on the gel. Therefore, choosing the proteins of interest is as simple as comparing two lanes, or two gels, and looking for changes in staining intensity.

The gel-free analysis of proteins is typically a dynamic process, which also requires the relative quantitation of peptides. No staining is performed in gel-free approaches, and the quantitation relies on the mass spectrometric signal. Mass spectrometers have been used for many years to do quantitative studies; however, this is always done by establishing response curves for particular analytes. This is hardly possible for a constantly changing set of analytes, such as peptides obtained from a proteolytic digestion of proteins. Furthermore, it is often difficult to compare even two experiments run one after the other in terms of peptide signal intensity.

The differential analysis of complex protein mixtures in a gel-free approach can be achieved by using isotope tagging of peptides to allow a direct comparison of the changes in peptide levels between two proteomes in a single experiment. Gygi et al. (1999a) have demonstrated a method called ICAT (isotope-coded affinity tag) that allows the differential quantitation of proteomes using cystein-containing peptides. Munchbach et al. (2000) proposed a more general way for N-terminal labeling of peptides for the differential quantitation. Although the method by Munchbach et al. was only demonstrated for gel-separated proteins, it appears that the reaction would just be as valid for gel-free analysis of proteins.

Briefly, the ICAT (Gygi et al., 1999a) method consists of labeling cystein-containing peptides from different samples with a light and heavy form of a reactive chemical. The ICAT reagent consists of a biotin group followed by a linker and is terminated with a cystein reactive group. The only difference between the light and the heavy tag is the presence of hydrogen or deuterium. This results in a similar response at the mass spectrometer while a diagnostic m/z spacing occurs on the MS spectrum. Furthermore, because the heavy/light peptides are quasi co-elute and are identical in sequence, their relative quantities can be obtained by the ratio of their intensity.

Typically, in the ICAT approach, a lysate sample first is reduced and reacted with the light form of the ICAT reagent. A second lysate sample is also reduced and reacted with the heavy form of the ICAT reagent. Both ICAT reagents preferentially react at the sulfhydryl group present on the cysteins. Once both labeling reactions have been performed, the two protein lysates are combined and enzymatically cleaved to generate peptide fragments. Then, the cystein-containing peptides are purified using a monomeric avidin column. The purified cystein-containing peptides are separated on a nanoflow HPLC system, online with an ESI mass spectrometer. Pairs of peptides with identical amino acid sequences that are tagged, respectively, with the light and heavy form of the ICAT are different in mass by 8 mass units. These are easily differentiated on a mass spectrometer and eluted with just a slight delay from each other. Further-

more, the automated generation of MS/MS fragmentation patterns of the peptides identifies the peptide sequence and its protein provenance by database searching. The relative quantitation is extracted from the MS spectrum using the ratio of the signal intensity observed for the light/heavy labeled peptides.

Munchbach et al. (2000) have introduced a more generic approach for the labeling of peptides at their N-termius using 1-(nicotinoyloxy)succinimide esters. They have synthesized a heavier form of the 1-(nicotinoyloxy)succinimide ester and introduced four deuteriums, which provide a 4 mass unit difference with the normal 1-(nicotinoyloxy)succinimide ester. The light/heavy pair of chemicals can be used for the differential labeling of proteins and their relative quantitation. Briefly, both samples are enzymatically digested to generate peptide fragments. The peptide fragments are then succinylated to block the ε-amino group of lysines. Then the first sample is reacted with a light form of 1-(nicotinoyloxy)succinimide ester, and the second sample is labeled with the heavy form of the 1-(nicotinoyloxy)succinimide ester. The two samples are then combined and analyzed by MALDI-TOF mass spectrometry.

It is to be expected that many more pairs of chemicals will become available for the relative quantitation of proteomes. It is important to realize that the expression profile of a proteome covers a few orders of magnitude. All the chemical tagging approaches have a defined dynamic range. Therefore, it is important to keep in mind that for low levels of proteins, these reactions can fail to provide linear responses. Therefore, although both of the labeled heavy and light forms of the peptide might be observed, the calculated relative quantitation for the low-level proteins might be inaccurate. Side reactions can also provide significant challenges.

Fourier Transform Mass Spectrometery for Gel-Free Analysis and Single-Peptide-Based Identification

The FTMS has attracted renewed attention from the proteomic community. The instrumentation is now more robust, it offers fragmentation, it has a larger dynamic range, an improved sensitivity, and it has recently been coupled to MALDI. The FTMS is by far the best mass spectrometer in terms of mass accuracy, resolution, and sensitivity. Peptide masses can be measured down to 1 ppm or less, depending on the sample. We have come to appreciate the advantages that accurate mass measurement offers in terms of protein identification.

Depending on the size of the genome, it is possible to unambiguously identify a protein solely based on the accurate mass measurement of one of its peptides (mass tag). For example, Figure 1.22*a* shows the plot of the number of isobaric peptides versus the peptide masses for a 10 ppm tolerance and a 1 ppm tolerance for the yeast proteome. This was predicted for all of the known yeast open-reading frames. Figure 1.22*b* also shows the percentage of unique peptides versus the peptide masses. It is clear that for small and medium genomes, protein identification based on a single peptide seems to be a possibility. It is also clear that the homology level is low in yeast. We have performed an $N \times N$ blast of the yeast genome and obtained about 20% of the protein in yeast with at least one homolog at an homology level of 80%. We expect this number to be significantly larger for more complex organisms. Furthermore, the idea of identifying a protein solely based on a single peptide does not take into account splicing

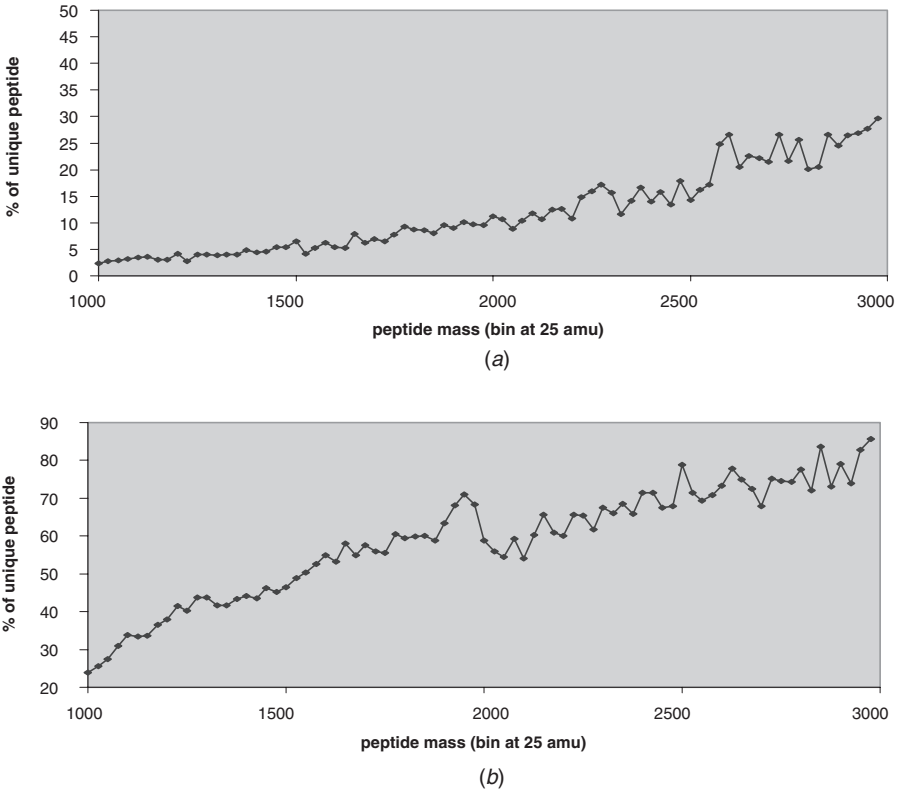


Figure 1.22. Effect of mass accuracy observed on the mass spectrometer on the probability of unique identification.

and point mutation. Also, it is one thing to say that a specific peptide is unique to a protein; however, it is another thing to find it.

One of the possible applications of this approach is the identification of protein contained in complex mixtures in a gel-free manner based on HPLC-ESI-MS/MS. Smith et al. have demonstrated the potential of the protein identification based on a single-peptide analysis by ESI-FTMS. Furthermore, they have also demonstrated that the addition of the ICAT reagent improves identification of peptides by restricting the search to the cystein-containing peptides. This also means that proteins can be identified in complex mixtures, based only on the presence of one of their peptides. It is not clear yet if this approach will be useful for larger genomes, such as the human genome.

REFERENCES

Adessi, C., Miege, C., Albrieux, C., and Rabilloud, T. (1997). Two-dimensional electrophoresis of membrane proteins: A current challenge for immobilized pH gradients. *Electrophoresis* **18**(1):127–135.

Aebersold, R., and Mann, M.(2003). Mass spectrometry-based proteomics. *Nature* **422**:198–207.

Aebersold, R., Leavitt, J., Saavedra, R., Hood, L., and Kent, S. (1987). Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl Acad. Sci. USA* **84**(20):6970–6974.

Aebersold, R., Teplow, D., Hood, L., and Kent, S. (1986). Electroblotting onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis. *J. Biol. Chem.* **261**(9):4229–4238.

Anderson, L., and Seilhamer, J. (1997). A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**:533–537.

Banks, R., Dunn, M., Forbes, M., Stanley, A., Pappin, D., Naven, T., Gough, M., Harnden, P., and Selby, P. (1999). The potential use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis—preliminary findings. *Electrophoresis* **20**(4–5):689–700.

Bartel, P.L., Roecklein, J.A., SenGupta, D.J., and Fields, S. (1996). A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.* **12**:72–77.

Blanchard, A. (1998). Synthetic DNA arrays. *Genet. Eng. (N Y)* **20**:111–23.

Blisnick, T., Morales-Betoulle, M.E., Vuillard, L., Rabilloud, T., and Braun Breton, C. (1998). Non-detergent sulphobetaines enhance the recovery of membrane and/or cytoskeleton-associated proteins and active proteases from erythrocytes infected by plasmodium falciparum. *Eur. J. Biochem.* **252**(3):537–541.

Blum, H., Beier, H., and Gross, H.J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis* **8**:93–99.

Brunet, S., Thibault, P., Gagnon, E., Kearney, P., Bergeron, J.J., and Desjardins, M. (2003). Organelle proteomics: Looking at less to see more. *Trends Cell Boil.* **13**:629–638.

Celis, J.E., Celis, P., Ostergaard, M., Basse, B., Lauridsen, J.B., Ratz, G., Rasmussen, H.H., Orntoft, T.F., Hein, B., Wolf, H., Celis, A. (1999). Proteomics and immunohistochemistry define some of the steps involved in the squamous differentiation of the bladder transitional epithelium: A novel strategy for idenfying metaplastic lesions. *Cancer Res.* **59**(12):3003–3009.

Chervet, J.P., Ursem, M., and Salzmann, J.B. (1996). Instrumental requirements for nanoscale liquid chromatography. *Anal. Chem.* **68**(9):1507–1512.

Chevallet, M., Santoni, V., Poinas, A., Rouquie, D., Fuchs, A., Kieffer, S., Rossignol, M., Lunardi, J., Garin, J., and Rabilloud, T. (1998). New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis* **19**(11):1901–1909.

Clarke, C., Titley, J., Davies, S., and O'Hare, M. (1994). An immunomagnetic separation method using superparamagnetic (Macs) beads for large-scale purification of human mammary luminal and myoepithelial cells. *Epithelial Cell Biol.* **3**(1):38–46.

Comisarow, M.B., and Marshall, A.G. (1974). Fourier transform ion cyclotron resonance spectroscopy. *Chem. Phys. Lett.* **25**:282–283.

Conrods, T.P., Anderson, G.A., Veenstra, T.D., Pasa-Tolic, L., and Smith, R.D. (2000). Utility of accurate mass tags for proteume-wide protein identification. *Anal. Chem.* **72**:3349–3354.

Craig, R., and Beavis, R.C. (2004) TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* Feb. 19 [Epub ahead of print].

Dai, Y., Whittal, R.M., and Li, L. (1999). Two-layer sample preparation: A method for MALDI-MS analysis of complex peptide and protein mixtures. *Anal. Chem.* **71**:1087–1091.

Davis, M.T., and Lee, T.D. (1998). Rapid protein identification using a microscale electrospray LC/MS system on an ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **9**(3):194–201.

Desprez, T., Amselem, J., Caboche, M., and Hofte, H. (1998). Differential gene expression in Arabidopsis monitored using CDNA arrays. *Plant J.* **14**(5):643–652.

Dovichi, N.J. (1997). DNA sequencing by capillary electrophoresis. *Electrophoresis* **18**(12–13): 2393–2399.

Dunn, M.J., and Corbett, J.M. (1996). In *Two-Dimensional Polyacrylamide Gel Electrophoresis. High Resolution Separation and Analysis of Biological Macromolecules. Part B, Applications.* Karger, B.L., and Hancook, W.S. (Eds.). San Diego, Academic Press, Vol. 271, pp. 177–203.

Edman, P., and Begg, G. (1967). A protein sequenator. *Eur. J. Biochem.* **1**(1):80–91.

Emmert-Buck, M.R., Gillespie, J.W., Paweletz, C.P., Ornstein, D.K., Basrur, V., Appella, E., Wang, Q.H., Huang, J., Hu, N., Taylor, P., Petricoin, E.F. 3rd (2000). An approach to proteomic analysis of human tumors. *Mol. Carcinog.* **27**(3):158–165.

Eng, J., McCormack, A.L., and Yates, J.R. III (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**:976–989.

Fenn, J.B., Mann, M., and Meng, C.K. (1990). Electrospray ionization—principles and practice. *Mass Spectro. Rev.* **9**(1):37.

Fialka, I., Pasquali, C., Lottspeich, F., Ahorn, H., and Huber, L.A. (1997). Subcellular fractionation of polarized epithelial cells and identification of organelle-specific proteins by two-dimensional gel electrophoresis. *Electrophoresis* **18**(14):2582–2590.

Fichmann, J. (1999). Advantages of immobilized pH gradients. *Methods Mol. Biol.* **112**:173–174.

Field, H.I., Fenyo, D., and Beavis, R.C. (2002). RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in relational database. *Proteomics* **2**:36–47.

Figeys, D., Zhang, Y., and Aebersold, R. (1998a). Optimization of solid phase microextraction–capillary zone electrophoresis–mass spectrometry for high sensitivity protein identification. *Electrophoresis* **19**(13):2338–2347.

Figeys, D., Gygi, S., McKinnon, G., and Aebersold, R. (1998b). An integrated microfluidics–tandem mass spectrometry system for automated protein analysis. *Anal. Chem.* **70**(18):3728–3734.

Figeys, D., Ning, Y., and Aebersold, R. (1997). A microfabricated device for rapid protein identification by microelectrospray ion trap mass spectrometry. *Anal. Chem.* **69**(16):3153–3160.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., and Garrels, J.I. (1999). A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**(11):7357–7368.

Garin, J., Diez, R., Kieffer, S., Demine, J.-F., Duclos, S., Gagnon, E., Sadoul, R., Rondeau, C., and Desjardsins, M. (2001). The phagosome proteome: Insight into phagosome functions, *J. Cell Biol.* **152**(1):165–180.

Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in year, *Nature* **425**:737–741.

Gibson, B.W., and Biemann, K. (1984). Strategy for the mass spectrometric verification and correction of the primary structures of proteins deduced from their DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **81**:1956–1960.

Goldberg, M.E., Expert-Bezancon, N., Vuillard, L., and Rabilloud, T. (1996). Non-detergent sulphobetaines: A new class of molecules that facilitate in vitro protein renaturation. *Fold Des.* **1**(1):21–27.

Gomm, J.J., Browne, P., Coope, R., Liu, Q., Buluwela, L., and Coombes, R. (1995). Isolation of pure populations of epithelial and myoepithelial cells from the normal human mammary gland using immunomagnetic separation with dynabeads. *Anal. Biochem.* **226**(1):91–99.

Gorg, A. (1993). Two-dimensional electrophoresis with immobilized PH gradients: Current state. *Biochem. Soc. Trans.* **21**(1):130–132.

Gorg, A., Postel, W., and Gunther, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**(9):531–546.

Gygi, S., and Aebersold, R. (1999). Absolute quantitation of 2-D protein spots. *Methods Mol. Biol.* **112**:417–421.

Gygi, S., Corthals, G., Zhang, Y., Rochon, Y., and Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Nat. Acad. Sci.* **97**(17):9390–9395.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999a). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotech.* **17**:994–997.

Gygi, S., Rochon, Y., Franza, B., and Aebersold, R. (1999b). Correlation between protein and mRNA abundance in yeast. *Mol. Cel. Biol.* **19**(3):1720–1730.

Himeda, C.L., Ranish, J.A., Angello, J.C., Maire, P., Aebersold, R., and Hauschka, S.D. (2004). Quantitative proteomic identification of six4 as the trex-binding factor in the muscle creatine kinase enhancer. *Mol. Cell Biol.* **24**:2132–2143.

Hooker, T.F., and Jorgenson, J.W. (1997). A transparent flow gating interface for the coupling of microcolumn LC with CZE in a comprehensive two-dimensional system. *Anal. Chem.* **69**:4134–4142.

Howell, K.E., Devaney, E., and Gruenberg, J. (1989). Subcellular fractionation of tissue culture cells. *Trends Biochem. Sci.* **14**(2):44–47.

Hunt, D.F., Yates, J.R., Shabanowitz, J., Winston, S., and Hauer, C.R. (1986). Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **83**:6233–6237.

Hunt, D.F., Bone, W.M., Shabanowitz, J., Rhodes, J., and Ballard, J.M. (1981). Sequence analysis of oliyopeptides by second ion/collision activated dissociation mass sepectrometry. *Anal. Chem.* **53**:1704–1706.

Issaq, H.J., Conrads, T.P., Janini, G.M., and Veenstra, T.D. (2002). Methods for fractionation, separation and profiling of proteins and peptides. *Electrophoresis* **23**:3048–3061.

Jackson, P., Urwin, V.E., and Mackay, C.D. (1988). Rapid imaging, using a cooled charge-coupled-device, of fluorescent two-dimensional polyacrylamide gels produced by labelling proteins in the first-dimensional isoelectric focusing gel with the fluorophore 2-methoxy-2,4-diphenyl-3(2h)furanone. *Electrophoresis* **9**(7):330–339.

Jung, E., Heller, M., Sanchez, J.C., and Hochstrasser, D.F. (2000). Proteomics meets cell biology: The establishment of subcellular proteomes. *Electrophoresis* **21**(16):3369–3377.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceding 10,000 daltons. *Anal. Chem.* **60**:2299–2301.

Kern, W., Kohlmann, A., Wuchte, C., Schnittger, S., Schoch, C., Mergenthaler, S., Ratei, R., Ludwig, W.D., Hiddemann, W., and Haferlach, T. (2003). Correlation of protein expression and gene expression in acute leukemia. *Cytometry* **55B**:29–36.

Klose, J., and Kobalz, U. (1995). "Two-dimensional electrophoresis of proteins: An updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16**(6):1034–1059.

LaCourse, W.R. (2002). Column liquid chromatography: Equipment and instrumentation. *Analy. Chem.* **72**:37R–51R.

Lal, A., Lash, A.E., Altschul, S.F., Velculescu, V., Zhang, L., McLendon, R.E., Marra, M.A., Prange, C., Morin, P.J., Polyak, K., Papadopoulos, N., Vogelstein, B., Kinzler, K.W., Stansberg, R.L., and Riggins, G.L. (1999). A public database for gene expression in human cancers. *Cancer Res* **59**(21):5403–5407.

Lewis, K.C., Opiteck, G.J., Jorgenson, J.W., and Sheeley, D.M. (1997). Comprehensive online RPLC-CZE-MS of peptides. *J. Am. Soc. Mass Spectrom.* **8**(5):495–500.

Li, J., Kelly, J.F., Chernushevich, I., Jed Harrison, D.J., and Thibault, P. (2000). Separation and identification of peptides from gel-isolated membrane proteins using a microfabricated device for combined capillary electrophoresis/nanoelectrospray mass spectrometry. *Anal. Chem.* **72**:599–609.

Li, I., Thibault, P., Bings, N.H., Skinner, C.D., Wang, C., Colyer, C., and Harrison, J. (1999). Integration of microfabricated devices to capillary electrophoresis–electrospray mass spectrometry using a low dead volume connection: Application to rapid analyses of proteolytic digests. *Anal. Chem.* **71**(15):3036–3045.

Loboda, A.V., Krutchinsky, A.N., Bromirski, M., Ens, W., and Standing, K.G. (2000). A tandem quadrupole/time-of-flight mass spectrometer with a matrix-assisted laser desorption/ionization source: Design and performance. *Rapid Commun. Mass Spectrom.* **14**(12):1047–1057.

Lubec, G., Krapfenbauer, K., and Fountoulakis, M. (2003) Proteomics in brain research: Potentials and limitations. *Prog. Neurobiol.* **69**:193–211.

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

Madden, S.L., Galella, E.A., Zhu, J.S., Bertelsen, A.H., and Beaudry, G.A. (1997). Sage transcript profiles for P53-dependent growth regulation. *Oncogene* **15**(9):1079–1085.

Mann, M., and Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**(24):4390–4399.

Marshall, A., and Hodgson, J. (1998). DNA chips: An array of possibilities. *Nature Biotech.* **16**:27–31.

Martin, S.E., Shabanowitz, J., Hunt, D.F., and Marto, J.A. (2000). Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry. *Anal. Chem.* **72**:4266–4274.

Matsui, N.M., Smith-Beckerman, D.M., and Epstein, L.B. (1999a). Staining of preparative 2-D gels. Coomassie blue and imidazole-zinc negative staining. *Methods Mol. Biol.* **112**:307–311.

Matsui, N.M., Smith-Beckerman, D.M., Fichmann, J., and Epstein, L.B. (1999b). Running preparative carrier ampholyte and immobilized pH gradient IEF gels for 2D. *Methods Mol. Biol.* **112**:211–219.

Matsumura, H., Nirasawa, S., and Terauchi, R. (1999). Technical advance: Transcript profiling in rice (*Oryza sativa L.*). *Plant J.* **20**(6):719–726.

McCormack, A.L., Schieltz, D.M., Goode, B., Yang, S., Barnes, G., Drubin, D., and Yates, J.R. (1997). Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**(4):767–776.

Moseley, M.A., Deterding, L.J., Tomer, K.B., and Jorgenson, J.W. (1989). Coupling of capillary zone electrophoresis and capillary liquid chromatography with coaxial continuous-flow fast atom bombardment tandem sector mass spectrometry. *J. Chrom.* **480**:197–209.

Munchbach, M., Quadroni, M., Miotto, G., and James, P. (2000). Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal. Chem.* **72**:4047–4057.

Neilson, L., Andalibi, A., Kang, D., Coutifaris, C., Strauss, J.F., 3rd, Stanton, J.A., and Green, D.P. (2000). Molecular phenotype of the human oocyte by PCR-SAGE. *Genomics* **63**(1):13–24.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**(10):4007–4021.

Opiteck, G.J., Jorgenson, J.W., Moseley III, M.A., and Anderegg, R.J. (1998). Two-dimensional microcolumn HPLC coupled to a single-quadrupole mass spectrometer for the elucidation of sequence tags and peptide mapping. *J. Microcolumn Separations* **10**(4):365–376.

Opiteck, G.J., Lewis, K.C., Jorgenson, J.W., and Anderegg, R.J. (1997). Comprehensive on-line LC/LC/MS of proteins. *Anal. Chem.* **68**:1518–1524.

Ornstein, D.K., Gillespie, J.W., Paweletz, C.P., Duray, P.H., Herring, J., Vocke, C.D., Topalian, S.L., Bostwick, D.G., Linehan, W.M., Petricoin III, E.F., and Emmert-Buck, M.R. (2000). Proteomic analysis of laser capture microdissected human prostate cancer and in vitro prostate cell lines. *Electrophoresis* **21**(11):2235–2242.

Ostergaard, M., Rasmussen, H.H., Nielsen, H.V., Vorum, H., Orntoft, T.F., Wolf, H., and Celis, J.E. (1997). Proteome profiling of bladder squamous cell carcinomas: Identification of markers that define their degree of differentiation. *Cancer Res.* **57**(18):4111–4117.

Page, M.J., Amess, B., Townsend, R.R., Parekh, R., Herath, A., Brusten, L., Zvelebil, M.J., Stein, R.C., Waterfield, M.D., Davies, S.C., and O'Hare, M.J. (1999). Proteomic definition of normal human luminal and myoepithelial breast. *Proc. Natl. Acad. Sci. U.S.A.* **96**(22):12589–12594.

Papayannopoulos, I.A. (1995). The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* **14**:49–73.

Pappin, D.J.C., Hojrup, P., and Bleasby, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**(6):327–332.

Patterson, S.D., Thomas, D., and Bradshaw, R.A. (1996). Application of combined mass spectrometry and partial amino acid sequence to the identification of gel-separated proteins. *Electrophoresis* **17**(5):877–891.

Perrin-Cocon, L.A., Marche, P.N., and Villiers, C.L. (1999). Purification of intracellular compartments involved in antigen processing: A new method based on magnetic sorting. *Biochem. J.* **338**:123–130.

Perrot, M., Sagliocco, F., Mini, T., Monribot, C., Schneider, U., Shevchenko, A., Mann, M., Jeno, P., and Boucherie, H. (1999). Two-dimensional gel protein database of *Saccharomyces cerevisiae*. *Electrophoresis* **20**:2280–2298.

Rabilloud, T. (2002). Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, bit it still climbs up the mountains. *Proteomics* **2**:3–10.

Rabilloud, T. (2000). Detecting proteins. Separated by 2-D gel electrophoresis. *Anal. Chem.* **72**(1):48A–55A.

Rabilloud, T. (1999). Silver staining of 2-D electrophoresis gels. *Methods Mol. Biol.* **112**: 297–305.

Rabilloud, T. (1998). Use of thiourea to increase the solubility of membrane proteins in two-dimensional electrophoresis. *Electrophoresis* **19**(5):758–760.

Rabilloud, T. (1996). Solubilization of proteins for electrophoretic analyses. *Electrophoresis* **17**(5):813–829.

Rabilloud, T. (1990). Mechanisms of protein silver staining in polyacrylamide gels: A 10-year synthesis. *Electrophoresis* **11**(10):785–794.

Rabilloud, T., Blisnick, T., Heller, M., Luche, S., Aebersold, R., Lunardi, J., and Braun-Breton, C. (1999). Analysis of membrane proteins by two-dimensional electrophoresis: Comparison

of the proteins extracted from normal or *Plasmodium falciparum*-infected erythrocyte ghosts. *Electrophoresis* **20**(18):3603–3610.

Rabilloud, T., Vuillard, L., Gilly, C., and Lawrence, J.J. (1994). Silver-staining of proteins in polyacrylamide gels: A general overview. *Cell Mol. Biol.* (*Noisy-le-grand*) **40**(1):57–75.

Ramagli, L.S. (1999). Quantifying protein in 2-D page solubilization buffers. *Methods Mol. Biol.* **112**:99–103.

Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., and Rabilloud, T. (2004). About the mechanism of interference of silver staining with peptide mass spectrometry. *Proteomics* **4**:909–916.

Righetti, P.G., and Bossi, A. (1997a). Isoelectric focusing in immobilized pH gradients: An update. *J. Chromatogr. B Biomed. Sci. App.* **699**(1–2):77–89.

Righetti, P.G., and Bossi, A. (1997b). Isoelectric focusing in immobilized pH gradients: Recent analytical and preparative developments. *Anal. Biochem.* **247**(1):1–10.

Righetti, P.G., and Gianazza, E. (1987). Isoelectric focusing in immobilized pH gradients: Theory and newer methodology. *Methods Biochem. Anal.* **32**:215–278.

Righetti, P.G., Gianazza, E., and Bjellqvist, B. (1983). Modern aspects of isoelectric focusing: Two-dimensional maps and immobilized pH gradients. *J. Biochem. Biophys. Methods* **8**(2): 89–108.

Rosenfeld, J., Capdevielle, J., Guillemot, J.C., and Ferrara, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **203**(1):173–179.

Ross-Macdonald, P., Coelho, P.S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K.H., Sheehan, A., Symoniatis, D., Umansky, L., Heidiman, M., Nelson, F.K., Iwasaki, H., Hager, K., Gerstein, M., Miller, P., Roeder, G.S., and Snyder, M. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**(6760):413–418.

Ruan, Y., Gilmore, J., and Conner, T. (1998). Towards arabidopsis genome analysis: Monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J.* **15**(6):821–833.

Sanchez, J.C., Hochstrasser, D., and Rabilloud, T. (1999). In-gel sample rehydration of immobilized pH gradient. *Methods Mol. Biol.* **112**:221–225.

Santoni, V., Rabilloud, T., Doumas, P., Rouquie, D., Mansion, M., Kieffer, S., Garin, J., and Rossignol, M. (1999). Towards the recovery of hydrophobic proteins on two-dimensional. *Electrophoresis* **20**(4–5):705–711.

Sarto, C., Valsecchi, C., and Mocarelli, P. (2002). Renal cell carcinoma: Handling and treatment. *Proteomics* **2**:1627–1629.

Service, R.F. (1998). Microchip arrays put DNA on the spot. *Science* **282**(5388):396–401.

Shevchenko, A., Loboda, A., Shevchenko, A., Ens, W., and Standing, K.G. (2000). MALDI quadrupole time-of-flight mass spectrometry: A powerful tool for proteomic research. *Anal. Chem.* **72**(9):2132–2141.

Shevchenko, A., Chernushevich, I., Ens, W., Standing, K.G., Thomson, B., Wilm, M., and Mann, M. (1997). Rapid "de novo" peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Comm. Mass Spectrom.* **11**(9):1015–1024.

Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. (1996a). Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc. Nat. Acad. Sci. U.S.A.* **93**(25):14440–14445.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996b). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**(5):850–858.

Sirivatanauksorn, Y., Drury, R., Crnogorac-Jurcevic, T., Sirivatanauksorn, V., and Lemoine, N. (1999). Laser-assisted microdissection: Applications in molecular pathology. *J. Pathol.* **189**(2):150–154.

Smith, B.J. (1994). Quantification of proteins on polyacrylamide gels (nonradioactive). *Methods Mol. Biol.* **32**:107–111.

Stein, W.D., Litman, T., Fojo, T., and Bates, S.E. (2004). A serial analysis of gene expression (SAGE) database analysis of chemosensitivity: Comparing solid tumors with cell lines and comparing solid tumors from different tissue origins. *Cancer Res.* **64**:2805–2816.

Steinberg, T.H., Chernokalskaya, E., Berggren, K., Lopez, M.F., Diwu, Z., Haugland, R.P., and Patton, W.F. (2000). General—ultrasensitive fluorescence protein detection in isoelectric focusing gels using a ruthenium metal chelate stain. *Electrophoresis* **21**(3):486–496.

Steinberg, T.H., White, H.M., and Singer, V.L. (1997). Optimal filter combinations for photographing Sypro orange or Sypro red dye-stained gels. *Anal. Biochem.* **248**(1):168–172.

Steinberg, T.H., Haugland, R.P., and Singer, V.L. (1996a). Applications of Sypro orange and Sypro red protein gel stains. *Anal. Biochem.* **239**(2):238–245.

Steinberg, T.H., Jones, L.J., Haugland, R.P., and Singer, V.L. (1996b). Sypro orange and Sypro red protein gel stains: One-step fluorescent staining of denaturing gels for detection of nanogram levels of protein. *Anal. Biochem.* **239**(2):223–237.

Tan, D., Deeb, G., Wang, J., Slocum, H.K., Winsto, J., Wiseman, S., Beck, A., Sait, S., Anderson, T., Nwogu, C., Ramnath, N., and Loewen, G. (2003). HER-2/neu protein expression and gene alteration in stage I-IIIA non-small-cell lung cancer: A study of 140 cases using a combination of high throughput tissue microarray, immunohistochemistry, and fluorescent in situ hybridization. *Diagn. Mol. Pathol.* **12**:201–211.

Tastet, C., Lescuyer, P., Diemer, H., Luche, S., van Dorsselaer, A., and Rabilloud, T. (2003). A versatile electrophoresis system for the analysis of high- and low-molecular-weight proteins. *Electrophoresis* **24**:1787–1794.

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

Unlu, M., Morgan, M., and Minden, J. (1997). Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis* **18**(11):2071–2077.

Urwin, V.E., and Jackson, P. (1993). Two-dimensional polyacrylamide gel electrophoresis of proteins labeled with the fluorophore monobromobimane prior to first-dimensional isoelectric focusing: Imaging of the fluorescent protein spot patterns using a cooled charge-coupled device. *Anal. Biochem.* **209**(1):57–62.

Urwin, V., and Jackson, P. (1991). A multiple high-resolution mini two-dimensional polyacrylamide gel electrophoresis system: Imaging two-dimensional gels using a cooled charge-coupled device after staining with silver or labeling with fluorophore. *Anal. Biochem.* **195**(1):30–37.

Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. *Science* **270**(5235):484–487.

Wang, H., and Hanash, S. (2003). Multi-dimensional liquid phase based separations in proteomics. *J. Chromatogr. B Analyt Technol. Biomed Life Sci.* **787**:11–18.

Wang, C., Oleschuk, R., Ouchen, F., Li, J., Thibault, P., and Harrison, J. (2000). Integration of immobilized trypsin bead beds for protein digestion within a microfluidic chip incorporating capillary electrophoresis separations and an electrospray ionization interface. *Rap. Comm. Mass Spectrom.* **14**:1377–1383.

Washburn, M.P., Wolters, D., and Yates, J.R.I. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotech.* **19**:242–247.

Weeraratna, A.T., Becker, D., Carr, K.M., Duray, P.H., Rosenblatt, K.P., Yang, S., Chen, Y., Bittner, M., Strausberg, R.L., Riggins, G.J., Wagner, U., Kallioniemi, O.P., Trent, J.M., Morin, P.J., and Meltzer, P.S. (2004). Generation and analysis of melanoma SAGE libraries: SAGE advice on the melanoma transcriptome. *Oncogene* **23**:2264–2274.

Wilkins, M.R., Gasteiger, E., Sanchez, J.C., Bairoch, A., and Hochstrasser, D.F. (1998). Two-dimensional gel electrophoresis for proteome projects: The effects of protein hydrophobicity and copy number. *Electrophoresis* **19**(8–9):1501–1505.

Wilkins, M.R., Ou, K., Appel, R.D., Sanchez, J.C., Yan, J.X., Golaz, O., Farnsworth, V., Cartier, P., Hochstrasser, D.F., Williams, K.L., and Gooley, A.A. (1996a). Rapid protein identification using N-terminal "sequence tag" and amino acid analysis. *Biochem. Biophys. Res. Commun.* **221**(3):609–613.

Wilkins, M.R., Hochstrasser, D.F., Sanchez, J.C., Bairoch, A., and Appel, R.D. (1996b). Integrating two dimensional gel databases using the Melanie II software. *Trends Biochem. Sci.* **21**(12):496–497.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996). Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**(Feb. 1):466–469.

Wimmer, K., Kuick, R., Thoraval, D., and Hanash, S.M. (1996). Two-dimensional separations of the genome and proteome of neuroblastoma cells. *Electrophoresis* **17**(11):1741–1751.

Yan, J.X., Devenish, A.T., Wait, R., Stone, T., Lewis, S., and Fowler, S. (2002). Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli. *Proteomics* **2**:1682–1698.

Yates, J.R., McCormack, A.L., Link, A.J., Schieltz, D., Eng, J., and Hays, L. (1996). Future prospects for the analysis of complex biological systems using micro-column liquid chromatography electrospray tandem mass spectrometry. *Analyst* **121**(7):R65–R76.

Yates, J.R. III, Eng, J.K., McCormack, A.L., and Schieltz, D. (1995). Method to correlate tandem mass-spectra of modified peptides to amino-acid-sequences in the protein database. *Anal. Chem.* **67**:1426–1436.

Zhang, W., and Chait, B.T. (2000). Profound: An expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem.* **72**:2482–2489.

Zhang, B., Liu, H., Karger, B.L., and Foret, F. (1999). Microfabricated devices for capillary electrophoresis–electrospray mass spectrometry. *Anal. Chem.* **71**(15):3258–3264.