Chapter 7
Mass Spectrometry Based Proteomics in Cancer Research

Mohamad A. Abbani, Parag Mallick, and Maryann S. Vogelsang

Abstract Proteomics has become an important component of biological and clinical research. Numerous proteomics methods have been developed to identify and quantify the proteins present in biological and clinical samples (Gerber et al., Proc Natl Acad Sci U S A 100:6940–6945, 2003; Ong et al., Methods 29:124–130, 2003). Differences among cell types or treatment groups have been used to identify cellular functions and pathways affected by disease or perturbations (Wright et al., Genome Biol 5:R4, 2003; Durr et al., Nat Biotechnol 22:985–992, 2004), new components and changes in the composition of protein complexes and organelles (Andersen et al., Nature 426:570–574, 2003; Blagoev et al., Nat Biotechnol 21:315–318, 2003; Ranish et al., Nat Genet 36:707–713, 2004), and putative disease biomarkers (Marko-Varga et al., J Proteome Res 4:1200–1212, 2005). Despite widespread success, the application of these approaches to discovery of relevant protein markers from clinical samples has been hampered by sample complexity and variability. To begin to broach this challenge, complex experimental protocols for enrichment, separation, and quantification have been developed for selective or comprehensive proteome analysis. In this chapter, we describe techniques for enrichment, separation, quantification, fundamentals of mass spectrometry, and the computational analysis of data generated by these processes within the context of using these approaches for asking and answering biologically and clinically important questions.

P. Mallick (✉)
Center for Applied Molecular Medicine, University of Southern California,
Los Angeles, CA, USA
and
Department of Chemistry and Biochemistry, University of California,
Los Angeles, CA, USA
e-mail: parag@ucla.edu

S. Yegnasubramanian and W.B. Isaacs (eds.), *Modern Molecular Biology*,
Applied Bioinformatics and Biostatistics in Cancer Research,
DOI 10.1007/978-0-387-69745-1_7, © Springer Science+Business Media, LLC 2010

7.1 Introduction

Proteomics has become an important component of biological and clinical research. Numerous proteomics methods have been developed to identify and quantify the proteins present in biological and clinical samples (Gerber et al. 2003; Ong et al. 2003). Differences among cell types or treatment groups have been used to identify cellular functions and pathways affected by disease or perturbations (Wright et al. 2003; Durr et al. 2004), new components and changes in the composition of protein complexes and organelles (Andersen et al. 2003; Blagoev et al. 2003; Ranish et al. 2004), and putative disease biomarkers (Marko-Varga et al. 2005). Despite widespread success, the application of these approaches to discovery of relevant protein markers from clinical samples has been hampered by sample complexity and variability. To begin to broach this challenge, complex experimental protocols for enrichment, separation, and quantification have been developed for selective or comprehensive proteome analysis. In this chapter, we describe techniques for enrichment, separation, quantification, fundamentals of mass spectrometry, and the computational analysis of data generated by these processes within the context of using these approaches for asking and answering biologically and clinically important questions.

Generally, we can define proteomics as the systematic study of the many and diverse properties of the proteins in a system with the aim of providing detailed descriptions of the structure, function, and control of biological systems in health and disease. Advances in methods and technologies have catalyzed an expansion of the scope of biological studies from the reductionist biochemical analysis of single proteins to proteome-wide measurements. Proteomics, like other high-throughput "discovery" approaches, such as genomic sequencing, microarray analysis, and metabolite profiling, has been catalyzed by mapping and sequencing of the complete genomes of many species. Through the sequencing of a genome, we are able to generate an approximate estimate the scale of the proteome. However, it is critical to recognize that the proteome is fundamentally different in nature than the genome. The proteome is significantly more complex than the genome as multiple protein isoforms can be synthesized from a single gene and as proteins have greater chemical diversity (due to post-translational modifications (glycosylation, methylation, proteolytic cleavage, etc.). Proteins are found over a wide dynamic range of concentration (10^8 per cell to 10^{12} in biological fluids) (Anderson and Anderson 2002). In addition, the genome is relatively static over the lifespan of an organism, whereas the proteome is highly dynamic, changing on rapid timescales in response to both environmental and chemical perturbations to the system.

Over the years, mass spectrometry has become the method of choice for proteomic analysis. This technology has enabled us to characterize complex mixtures and probe for detailed information about individual proteins (e.g., covalent structures and post-translational modifications). A variety of proteomics applications exist, including the study of protein–protein interactions via affinity-based isolations on a small and proteome-wide scale, the mapping of numerous organelles, the concurrent description of the genome and proteome of small organisms (e.g., tuberculosis and malaria), and the generation of differential quantitative protein profiles of diverse species.

The ability of mass spectrometry to identify and, increasingly, to precisely quantify thousands of proteins from complex samples can be expected to broadly impact biology and medicine. Furthermore, proteomics has become an integral part in the emerging field of systems biology.

The essence of systems biology approaches pre-supposes that for any given system, the space of possible biomolecules and their organization into pathways and processes is large but finite. Consequently, the biological systems operating in an organism can be described comprehensively if a sufficient density of observations on all of the elements that constitute the system can be obtained. Proteomics is a particularly rich source of biological information because proteins are involved in almost all biological activities and they also have diverse properties, which collectively contribute greatly to our understanding of biological systems.

A standard proteomics process has three main components: (1) sample preparation, (2) mass spectrometric analysis, (3) data analysis and interpretation. In this chapter we initially describe these three aspects of the process. We conclude by describing specific examples of proteomics applications, including protein-protein interaction characterization, protein post-translational modification characterization, and quantitative differential analyses (both unbiased and targeted). In the interest of space we do not specifically describe the cannon of techniques for extraction of a sample from a patient or from in vitro experiments. Instead, we focus on the techniques for sample generation using metabolic labeling, used in quantitative proteomics, and on the downstream analysis once a protein mixture has been obtained.

7.2 Sample Preparation

7.2.1 Proteome Analysis Challenged by the Large Concentration Range

Evaluation of the human proteome provides opportunities to improve disease diagnosis and therapeutic monitoring. However, before these goals are realized, challenges must be overcome. For example, the human plasma proteome has inherent properties that complicate analysis by mass spectrometry. It contains a densely concentrated number of proteins and has a large dynamic concentration range of proteins that exceeds 10 orders of magnitude (Anderson and Anderson 2002). This large dynamic range is further complicated as a large fraction of the proteins are albumin (55%) and glycoproteins. In addition, it contains subsets of proteins from other tissues. Regardless of these challenges, human plasma can be analyzed. Advantages include abundance and ease of collection in the clinic.

The massive complexity of the plasma proteome requires that it be divided by fractionation into manageable smaller parts prior to analysis by currently available analytical methods. Although the physiochemical properties of proteins offer scientists a rich basis for many separation techniques (Table 7.1), protein solubility is considered the bottleneck problem in fractionation. To solubilize proteins, detergents

Table 7.1 A summary of fractionation techniques

Method		Properties based separations	Application	Separation	% Reproducibility
Chromatography					80–90
	Gel filtration	Size	Fractionation	Poor	
	Ion exchange	Electrostatic	Fractionation	Moderate	
	Chromatofocusing	Isoelectric point	Fractionation	Good	
	RPLC	Hydrophobicity	Fractionation	Moderate	
	Metal chelate	Electrostatic	Enrichment	Good	
	Affinity	Structure/ligand binding capacity	Fractionation/enrichment	Good	
Electrophoresis		Size and P*I*	Fractionation	Good	40–60

are often used; the choice of detergent often puts limits on the types of fractionation that can be used. Sample recovery and reproducibility are of great importance. Several methods of protein fractionation are discussed in the following section. The most common paradigm in proteomics, "bottom-up" proteomics, operates on peptide fragments of proteins. These peptides are typically generated by digestion of a protein mixture with a protease (e.g., trypsin). The fractionation methods below can be applied either on protein mixture prior to digestion, or on peptide mixtures following digestion.

7.2.2 Fractionation Using Chromatographic Techniques

In general, for chromatographic separations, proteins interact with a solid phase and are then released into a liquid phase and recovered (Fig. 7.1). In general, chromatographic

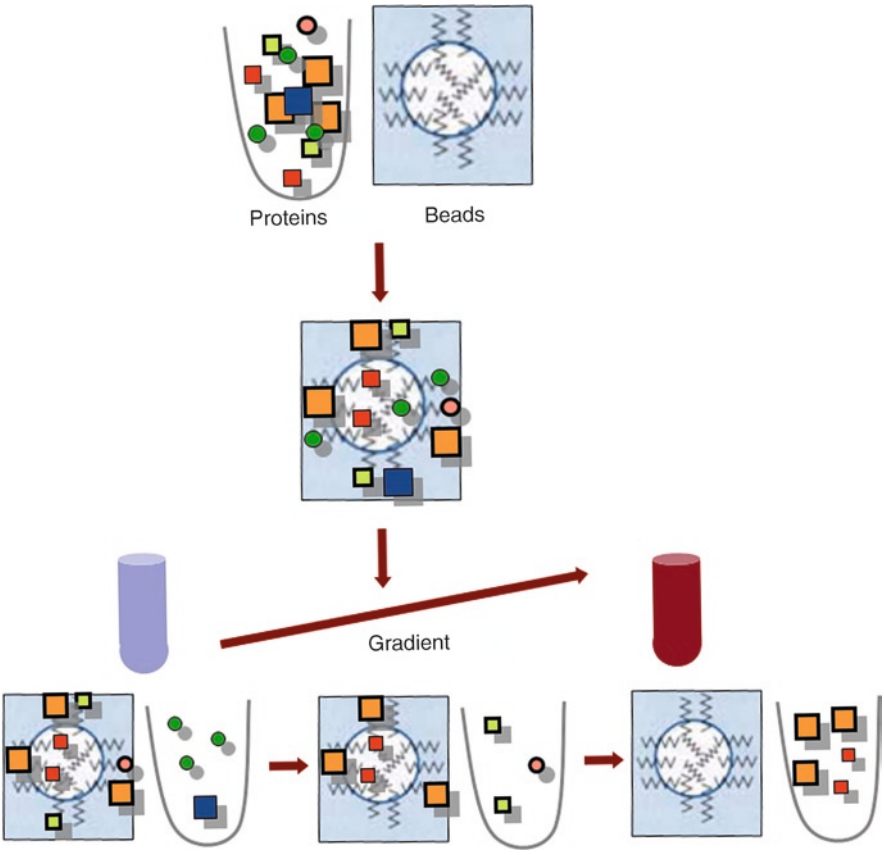

Fig. 7.1 Schematic diagram of the general process of chromatography. First, proteins are loaded on the column, where they interact with the solid phase. Second, proteins are released gradually by a gradient to be recovered into the liquid phase separated into different fractions

methods offer reproducible separation with high recovery of proteins as compared to electrophoresis techniques. Based on protein separation parameters, chromatographic techniques are categorized as follows:

7.2.2.1 Gel Filtration

The gel filtration process separates proteins based on size. In this method, the proteins travel through a column with a solid phase made of permeable beads. Large proteins cannot enter the pores of the beads; they pass around the beads and travel a shorter distance than smaller proteins to elute first. Smaller proteins penetrate the pores and pass through the beads, resulting in retardation of migration through the column. In gel filtration, proteins elute in the order of decreasing size. Although this technique offers good recovery and reproducibility, it suffers from low resolution as compared to other procedures.

7.2.2.2 Ion Exchange Chromatography

On ion exchange resin, proteins are separated based on their electrical charge. Proteins bind to the solid phase ion exchange resin through electrostatic interactions between charges on the proteins and of the opposite charges on the resin. The mobile phase used for loading the proteins onto the column is electrically neutral, so no interference with protein binding is introduced. Proteins are eluted as the ionic strength of the mobile phase is increased; this increases the competition for binding of proteins to the solid phase and proteins are eluted in order of affinity for the ionic resin. Less commonly, the pH of the solution can be adjusted such that the charge on either the proteins or the solid phase is altered to dislodge the analytes. Although the ion exchange procedure is limited by protein solubility, it offers higher resolution separation than gel filtration. For instance, charged detergents cannot be used in binding solution when applying this technique, but neutral ones offer a great advantage in solubilizing proteins. In addition, native or denatured proteins can be separated in the presence or absence of non-interfering detergents. A limitation of this technique is that a multicharge difference is required between species for good resolution, rendering it unpractical for post-translational modification studies. However, it is still considered a very powerful and useful tool for protein fractionation. This approach has been refined for efficient peptide separation by the use of strong ion exchange chromatography (Burke et al. 1989).

7.2.2.3 Chromatofocusing

Another valuable technique is chromatofocusing, a variant of ion exchange chromatography. In this technique, fractionation is based on the protein's isoelectric point (P*I*). The binding and elution of proteins is controlled solely by the pH of

the mobile phase. In anion chromatofocusing, binding is achieved at high pH and elution is accomplished by introducing a pH gradient. As the pH on the column is decreased, the positive charge on the proteins becomes more pronounced, the negative charge of the column becomes weaker and proteins are dislodged. The opposite process is followed for cation chromatofocusing. As the proteins are dissociated from the top of the column due to the change in pH, they re-bind at a lower region where the pH is still favorable. This process is repeated until the proteins reach the bottom of the column and elute in a very concentrated volume; this leads to greater resolution than is obtained on ion exchange. Like ion-exchange, protein solubility poses a challenge to the chromatofocusing process. However, unlike ion-exchange, this technique can be used to distinguish post-translational modifications to proteins, since even a single modification can lead to a different P*I*.

7.2.2.4 Reversed Phase Chromatography

Hydrophobicity is another parameter used in protein fractionation. The most common technique in this class is high-performance liquid chromatography (HPLC), a special type of reversed phase chromatography. In general, a non-polar stationary phase and an aqueous, moderately polar mobile phase are employed. Proteins that are non-polar have a longer retention time, whereas polar molecules elute quickly. By increasing the non-polar character of the mobile phase, adsorbed proteins are eluted. For proteomic studies, the proteins from whole cell lysates or biological fluids are denatured prior to fractionation.

7.2.2.5 Metal Chelate Chromatography

Proteins can also be fractionated by metal chelate chromatography, in which a metal ion is affixed to the solid phase via an immobilized iminodiacetic acid resin. All metal binding compounds are attracted to this surface; elution is carried out using solutions containing competing molecules such as immidazole or by varying the pH. For proteomic studies, this method is commonly applied for enrichment. For example, immobilized metal affinity chromatography (IMAC) can be used to extract phospho-proteins/peptides from complex mixtures (Porath 1992). In this approach, the phospho-peptides are captured by a trivalent cation complex and then eluted by either high pH or a phosphate buffer. To minimize non-specific binding from peptides rich in carboxylate groups, tryptic peptides are converted to methyl esters using methanolic HCl prior to enrichment (Ficarro et al. 2002).

7.2.2.6 Affinity Chromatography

A variety of proteins can be selectively captured by a matrix immobilized ligand. The ligand–protein complex is then destabilized by salts or by competition through

another ligand-binding entity. Although, this technique has been applied successfully to fractionation, it is used also for complex sample depletion of abundant proteins to reduce mixture complexity. For example, this strategy has been utilized in proteomic studies of serum/plasma and other body fluids to enhance the detection of low abundance proteins and achieve broader proteome coverage. In human plasma, there are 22 most abundant proteins responsible for ~99% of the total protein mass and these proteins mask the detection of hundreds of thousands of other proteins (Anderson and Anderson 2002).

In addition, affinity chromatography is used for enrichment of glycoproteins and glycopeptides. One method of glyco-capture is the use of lectin affinity chromatography (Cummings and Kornfeld 1982; Hirabayashi 2004). In this method, *N*-glycoproteins are captured through their binding to immobilized concanavalin A (Con A) (Bunkenborg et al. 2004; Fan et al. 2004). This method has been refined by combining Con A with wheat germ agglutinin (WGA) for the capture *O*-glycopeptides/proteins (Bunkenborg et al. 2004; Yang and Hancock 2004). Enrichment of glycoproteins or peptides can be achieved by other approaches discussed above. For example, size exclusion chromatography is utilized since most tryptic glycopeptides in a complex mixture have a relatively high mass (Alvarez-Manilla et al. 2006). Hydrophilicity of the glycan moiety is also utilized for enrichment through the hydrophilic interaction with a carbohydrate-based matrices (Wada et al. 2004).

7.2.3 Fractionation by Gel Electrophoresis

This type of protein separation exploits the size and charge or isoelectric point of proteins as parameters for separation. The proteins are driven through the gel matrix by an electric current. The most common method that uses size and charge as parameters for separation is the sodium dodecyl sulfate (SDS) electrophoresis. In general, the matrix is made up of acrylamide crosslinked to produce differently sized porous networks. Initially, proteins are uniformly charged with SDS (1.4 g SDS/g protein) (Reynolds and Tanford 1970). This SDS–protein coupling prevents aggregation, as all proteins are highly negatively charged, and produces a uniform charge to mass ratio for all proteins such that separation occurs based solely on size. Denatured proteins are loaded in wells in the gel and when the electric current is on they move toward the anode. Gels are usually sectioned into two steps to achieve better resolution. In the first step (stacking), proteins are concentrated by an isotachophoresis process. In the second step, proteins are separated based on their ability to navigate the matrix and recovered by excision for MS analysis. One disadvantage of SDS electrophoresis is its inability to separate post-translationally modified proteins. Another drawback is that some proteins, including glycoproteins and very basic proteins, have poor detergent binding. Also, the high pH conditions of electrophoresis may remove or introduce some post-translational modifications. BAC-SDS was developed by Macfarlane et al. to

avoid alteration of post-translational modifications. The first dimension of BAC-SDS is a cationic detergent based electrophoresis and the second dimension is SDS-based.

Isoelectric focusing (IEF) uses the isoelectric point of proteins as a separation parameter. In IEF, separation is achieved by placing proteins in a pH gradient and driving movement by an electric current. Once proteins reach their isoelectric point zone, their charge becomes zero and they cease movement. Similar to chromatofocusing discussed above, this technique provides high resolution, even when proteins differ slightly (as little as 0.1 pH unit in their P*I*). This makes IEF a very useful tool for separating proteins and their post-translationally modified forms. Isoelectric focusing is widely used in proteomics in a stepwise combination with SDS-PAGE resulting in a two-dimensional electrophoresis (2DE). First, proteins are fractionated using IEF and next are resolved on SDS-PAGE gels. Although gel-based separation techniques have been widely used, they suffer major drawbacks, including poor reproducibility and low protein recovery.

Given the specificity and restrictions of each of the above mentioned methods, scientists often use a combination of these techniques to achieve separations that can be used for proteomics studies. Routinely, proteomists combine these techniques in a stepwise fashion. Also liquid chromatography (of peptides) has been coupled directly to tandem MS in a further attempt to achieve additional resolution.

7.2.4 Quantitative Proteomics

Although, mass spectrometry has been a very successful tool for studying proteins in complex mixtures, these studies have been so far dominated by qualitative results (Fig. 7.2). To complement this, proteomics researchers have developed two approaches to attain quantitative proteomic information. In general, a rough approximation of relative protein amounts between two samples can be extracted by comparison of the same peptide signals derived from samples prepared under different conditions. Alternatively, a predictable mass difference can be artificially introduced to more accurately accomplish quantitation. In order to prevent spectral overlap, incorporation of stable isotope labels should result in at least a 3 Da mass shift. These labels can be added using chemical "post-biosynthetic" or metabolic "pre-biosynthetic" approaches. The addition of the label allows for mixing of samples originating under different conditions for simultaneous analysis. When samples are mixed early in the workflow, less bias is introduced during sample processing, resulting in high reproducibility (Fig. 7.3). Therefore, methods that incorporate the stable isotope label at the protein level have higher reproducibility than those that introduce it at the peptide level. Different isotopic labeling techniques will be discussed in the following section.

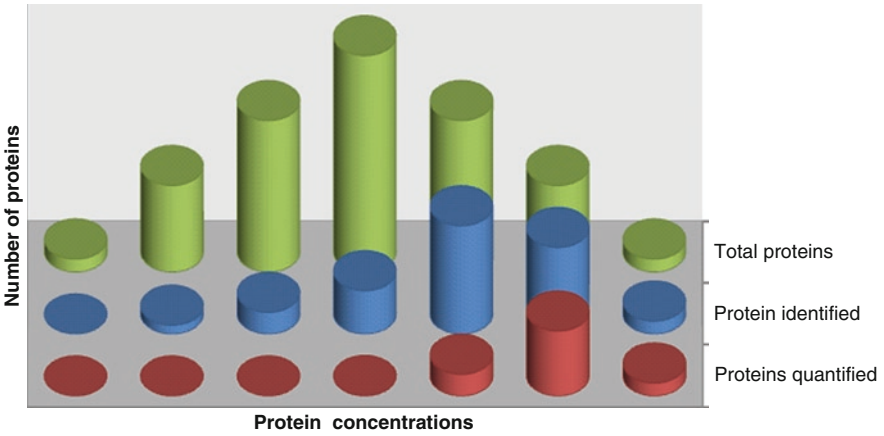

Fig. 7.2 Schematic depiction of identification and quantification of a proteome by mass spectrometry. Because of the limitations of current technology, only a fraction of the proteome can be identified, whereas a subset of that can be quantified

7.2.4.1 Pre-biosynthetic Labeling

In vivo labeling takes advantage of cell metabolism to effectively incorporate a stable isotope into proteins via the process of translation during cell growth and division. There are two approaches, the less practical is global labeling of proteins by growing cells by in ^{15}N-supplemented cell culture medium (Gygi et al. 1999). Although in vivo ^{15}N metabolic protein labeling of *C. elegans*, *Drosophila melanogaster* (Krijgsveld et al. 2003), rat (Wu et al. 2004), and plants (Gruhler et al. 2005) is feasible, it is not widely applied because it is time consuming and very expensive.

The most popular approach is the stable isotope labeling with amino acids in cell culture (SILAC) (Ong et al. 2002). In general, the stable isotope is incorporated by supplementing the cell growth medium with $^{13}C_6$-arginine and $^{13}C_6$-lysine. A second set of cells are grown in a label free environment. This approach guarantees that the resultant peptides from the tryptic cleavage of a protein do not overlap in the MS spectrum, excluding its C-terminus, since they contain no less than one labeled amino acid per peptide (heavy) with a constant mass increase of 6 Da as compared to the non-labeled corresponding peptides (light). The two cell populations are pooled, lysed, and proteins are isolated, denatured, reduced, and digested. The peptides are then quantified by MS. Protein identification is determined from either the "heavy" or the "light" peptide by MS, while relative quantitation is achieved by taking the ratios of the intensities of the two isotopes of the specific peptide in the MS spectrum. The advantage of SILAC over full metabolic protein labeling by ^{15}N lies in more straightforward data analysis since the labels in SILAC are specifically incorporated, defined, and not peptide-sequence dependent. SILAC is widely used for metabolic labeling of higher eukaryotic cells. Near complete incorporation of labels occurs after six doubling of cells grown in SILAC media (Ong et al. 2002).

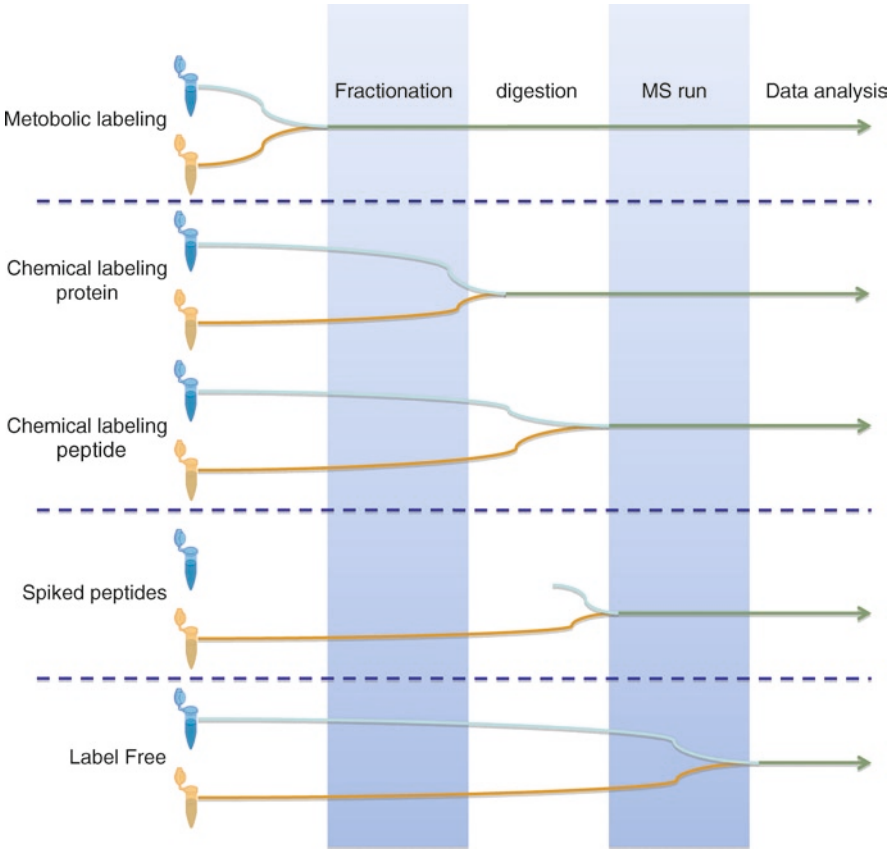

Fig. 7.3 Schematic diagram of the quantitative proteomics workflow. Tubes in blue and orange depict the two experimental conditions. The intersections of curved horizontal lines represent when samples are combined. *Shaded areas* represent different points of the experimental procedure and also where sample bias can be introduced if samples have not yet been combined

Although many cell lines can be labeled easily using SILAC, others are not. For instance, certain cell lines readily form proline from excess arginine, which can be alleviated by supplementing limited amount of arginine to the medium (Chelius et al. 2003). Some cell lines do not grow well in SILAC media and therefore cannot be labeled with this technique. Another limitation to the SILAC technology is the limited availability of useful isotopically labeled amino acids. As a consequence, in a single experiment only up to three conditions can be compared. For instance, the unlabeled sample can be compared to samples with $^{13}C_6$, $^{13}C_6$, and $^{15}N_4$ labels. One of the main advantages of SILAC specifically and metabolic labeling in general its reliable accuracy in quantitative MS-based methods due to early labeling and sample mixing. As a result, metabolic labeling is extremely useful for measurement of small variations in protein levels as well as post-translational modifications (Blagoev et al. 2004; Olsen et al. 2006; Park et al. 2006).

7.2.4.2 Post-biosynthetic Labeling

Isotopic labeling of extracted proteins and peptides can also be carried out in vitro either chemically or enzymatically. A stable isotope label can be incorporated into peptides enzymatically either during proteolytic digestion or in a separate step after proteolysis. Hence, enzymatic labeling can be very specific. For example, two ^{18}O isotope labels can be incorporated into the C-termini of peptides by either trypsin- or Glu-C during protein digestion (Yao et al. 2001; Reynolds et al. 2002). This results in a 4 Da (2 Da/^{18}O) mass shift that can be utilized for isotopomer discrimination. Other enzymes such as Lys-N introduce only one ^{18}O isotope and this mass difference cannot be detected in the spectrum (Rao et al. 2005). Since isotope labels can be lost at high pH (Schnolzer et al. 1996), electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) MS, which operate at moderate pH values, are utilized for these experiments. Another drawback of enzymatic labeling is that incorporation of isotopes is rarely complete and peptides are often differentially labeled, which can lead to tricky data interpretation (Johnson and Muddiman 2004; Ramos-Fernandez et al. 2007).

The chemical labeling approach mainly utilizes stable isotope-carrying chemical reagents to target active sites on peptides or proteins. The main targets for these reagents are the side chains of lysine and cysteine; therefore, this approach is not as specific as enzymatic labeling. The isotope-coded affinity tag (ICAT) was developed by Aebersold and co-workers (Gygi et al. 1999). The ICAT reagent targets and modifies cysteine residues and links them to a biotin tag by a polyether region, which contains either eight (heavy) or no (light) deuteriums. The biotin tag is used for affinity purification and recovery of the labeled peptides. The ICAT experiment is performed on two isolated populations of proteins that are reduced and tagged with light and heavy ICAT reagents. The proteins are then pooled and digested and the tagged peptides are recovered by affinity chromatography and quantified by MS.

ICAT-generated samples are less complex than those obtained from other chemical approaches since only cysteine, a rare amino acid, is labeled and thus analysis of complex samples is feasible. However, this also means that proteins with one or no cysteines are not detectable. In addition, the large tag effect on the fragmentation spectra and eluting time during reverse phase chromatography between lights and heavies are troublesome. These limitations have been overcome by recent technology advances, such as replacing the linker with one that is cleavable (Hansen et al. 2003; Li et al. 2003; Oda et al. 2003). A similar method makes use of a 2-thiopyridyl disulfide group to react with cysteines, a deuterium-labeled alanine, a His_6-tag for affinity purification, and a tryptic cleavage site to limit the size of the tag (Olsen et al. 2004). ICAT and similar methods are valuable tools for a host of expansive, human plasma, or targeted analyses.

Chemical labeling can also be achieved by a group of reagents that modify the N-terminus of the peptide as well as the epsilon-amino group of lysine residues. The most common and specific reagents are the *N*-hydroxysuccinimide (NHS) and other active esters and acid anhydrides. This group includes isotope tags for relative and absolute quantification (iTRAQ) (Ross et al. 2004), the isotope-coded protein

label (ICPL) (Schmidt et al. 2005), tandem mass tags (TMT) (Thompson et al. 2003), and acetic/succinic anhydride (Glocker et al. 1994; Ji et al. 2000; Che and Fricker 2002; Zhang et al. 2002). Less commonly employed reagents are isocyanates, isothiocyanates (Mason and Liebler 2003; Lee et al. 2004), and formaldehyde methylation of lysine residues followed by reduction by cyanoborohydride (Hsu et al. 2003; Ji et al. 2005; Hsu et al. 2006).

In these chemical labeling methods, labeled and unlabeled peptides are mixed and then quantification is performed by using the ratio of the MS signal intensities of the different isotopes or reporter ions. One advantage of the abovementioned methods is the use of isobaric tagging of peptides (Thompson et al. 2003). This results in peptides that co-elute during liquid chromatography, leading to reduced variability. The different tags are then distinguished by the mass spectrometer after fragmentation occurs. For instance, in single MS mode the same peptides with different labels are identical in mass. However, in tandem MS mode, where the peptides are fragmented, each tag generates a unique reporter ion. Protein quantitation is then achieved by taking the ratio the intensities of the reporter ions relative to each others in the MS spectra. This approach allows the simultaneous determination of both identity and relative abundance of peptide pairs in MS. The main advantage of the iTRAQ reagent (Ross et al. 2004) is that it allows multiple quantitation of up to eight samples at the same time thereby reducing the amount of mass spectrometry time needed for analysis.

Another type of chemical isotopic labeling targets the carboxylic acids in proteins. The C-termini of proteins as well as glutamate and aspartate are esterified by deuterated alcohols (David et al. 2001; Syka et al. 2004b). This approach has proven useful in quantitative studies of phosphorylated peptides, since it reduces the cross-reaction with ion metal chelate affinity chromatography (IMAC) mentioned earlier (Salomon et al. 2003). Also β-elimination of phosphoric acid followed by Michael addition is used for quantitation studies of phosphorylated peptides (Goshe et al. 2001, 2002; Qian et al. 2003; Tao et al. 2005). Quantitative studies of glycosylated peptides are achieved by the usage of hydrazide-based reactions, in which the carbohydrate is replaced by an isotopically labeled tag (Zhang et al. 2003).

All chemical labeling methods can be applied to proteins as well peptides. For example, labeling of the N-termini and lysine side chains of proteins has been applied using iTRAQ and ICPL. An advantage of labeling at the protein level is the minimization of bias introduced at later steps, since sample combination can be achieved early in the process. However, there is a drawback to protein labeling: Trypsin cannot cleave modified lysine residues, resulting in lower identification coverage due to presence of long peptides.

7.2.4.3 The Absolute Quantification Strategy

The absolute quantification of proteins (AQUA) strategy (Gerber et al. 2003) is accomplished by adding modified peptides to a sample as internal standards.

These peptides can contain stable isotopes and can be synthesized with covalent attachments to mimic protein posttranslational modifications such as phosphorylation, methylation, and acetylation. Data analysis is performed by comparing the signal from the synthetic peptide to the native peptide in the MS spectrum. The AQUA approach is limited to the quantitation of only a small subset of any sample, but it is still very useful if the aim of the study is focused on one or few proteins. For example, Gerber et al. (2003) measured the cell cycle-dependent modification of the human separase protein. To alleviate the limitations of AQUA, a *de novo* gene design was developed to express artificial proteins that are concatemers of tryptic Q peptides (QCAT) (Beynon et al. 2005). This strategy increases coverage, reduces bias, and provides better accuracy due to the introduction of the peptides early in the process. This strategy was applied successfully in the absolute quantification of the components of the eIF2B–eIF2 protein complex (Kito et al. 2007).

A limitation to the AQUA and similar strategies is underscored by the inherently narrow dynamic detection range of present mass spectrometry, which is compounded by the complexity of the tryptic digests of entire proteomes. The amount of labeled standard that must be spiked into the sample is rather difficult to determine since proteins of interest can be expressed differentially under diverse conditions. Also the specificity of the added standards is potentially problematic if they result in multiple isobaric peptides. The use of a very clever strategy called multiple reaction monitoring (MRM) can alleviate both of these limitations (Kirkpatrick et al. 2005). In an MRM experiment, a triple quadrupole mass spectrometer is employed to facilitate two stages of mass filtering. The intact ionized peptide is preselected and fragmented and then a small number of resultant sequence-specific fragment ions are mass analyzed. This targeted MS analysis using MRM improves specificity in peptide assignments, expands the quantitation scale, and enhances the detection limit for peptides by up to 100-fold (Wolf-Yadlin et al. 2007). Also the choice of tryptic peptides to be used can be assisted by use of a platform that predicts the most likely protein fragments to be observed, hence facilitating the choice of standard peptides (Mallick et al. 2007).

7.2.4.4 Label-Free Quantification

Proteomic quantification can also be achieved without artificially labeling parts of the sample. There exist two approaches to accomplish this label-free quantification: extraction of peptide ion intensities (Bondarenko et al. 2002; Chelius and Bondarenko 2002; Chelius et al. 2003; Wang et al. 2003; Li et al. 2005) and spectral counting (Gao et al. 2003; Liu et al. 2004). The first approach is based on comparing integrated areas under the curve of extracted peptide ion intensities (Higgs et al. 2005). The accuracy of this method is limited by the mass accuracy and reproducibility of the mass spectrometer. To achieve high accuracy, one should minimize the signal overlap by utilizing a high mass accuracy spectrometer. Also, utilization of LC alignment software can optimize the chromatographic profile of peptides (Bylund et al. 2002; Strittmatter et al. 2003; Jaitly et al. 2006; Wang et al. 2007) in

turn enhancing reproducibility. These types of experiments require an immense amount of time; therefore, a compromise has to be made between identification and quantitation. As a consequence, better quantification accuracy is achieved at the expense of coverage and vice versa.

An alternative is the spectral counting approach, which depends on wide data acquisition for both identification and quantitation. The spectral count approach is relatively new and relates the number of mass spectra identified for a protein to the protein's abundance (Gao et al. 2003; Liu et al. 2004; Gilchrist et al. 2006). Therefore, a direct comparison of two or more runs will allow the relative quantification of the protein of interest. To achieve better accuracy and reliable quantitation, an exponentially modified protein abundance index (emPAI) (Ishihama et al. 2005) is utilized, which is proportional to concentration of proteins in a sample. In addition, better quantitation is achieved through use of computational tools that select peptides in advance for detection by the mass spectrometer (Craig et al. 2005; Tang et al. 2006; Lu et al. 2007; Mallick et al. 2007). The minimum number of spectral counts required to see a significant change was determined by Old et al. (2005); they observed that the relationship is not linear, but rather exponential. They concluded that four spectra were sufficient to see threefold protein changes, but up to fifteen spectra were needed to observe a twofold change. They also showed that the spectral counting method yields reliable results as compared with extraction of peptide ion intensities, but both methods are less sensitive than isotopic labeling (Old et al. 2005). Although this approach has a great benefit highlighted by the simultaneous protein identification and quantitation, it suffers a major drawback. Quantitation is greatly dependent on the quality of MS/MS peptide identification, since errors in peptide identification can lead to inaccurate protein quantitation (Li et al. 2003; Olsen et al. 2006). Although both label-free methods have their advantages and can be applied for global quantitation studies, both require extensive platform setup. Hence only a handful of labs are able to take advantage of these methods.

7.3 Mass Spectrometric Analysis

As noted above, technological advances over the last 20 years have made mass spectrometry the tool of choice for proteomics researchers. In the late 1980s mass spectrometry of biological samples was improved by two novel ionization techniques, MALDI and ESI (Fig. 7.4). After that, the field further developed with advances in sample preparation, instrumentation, and sample analysis algorithms; all of which will be discussed in this section. In general, a mass spectrometer is used to answer two questions about a biological sample: "What is in the sample?" and "How much is in the sample?" Our goal with the following section is to provide insight into how mass spectrometers function so as to improve understanding of the resulting values of identity and quantity that are critical to modeling biological systems. As noted above, the most common proteomics paradigm uses mass

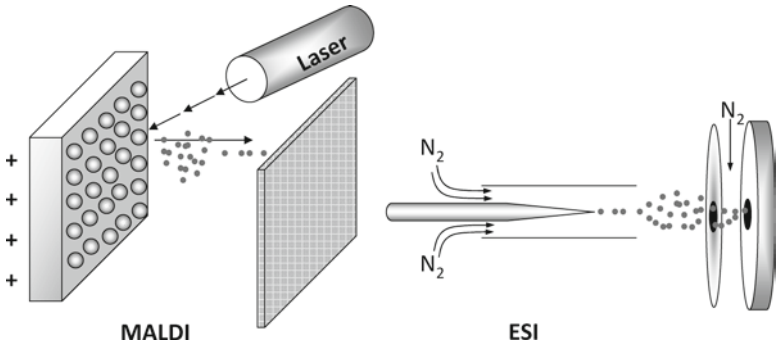

Fig. 7.4 Schematic of ionization methods. In *MALDI*, a sample is co-crystalized in a matrix solution atop a target plate. The prepared sample may then be irradiated with a laser, resulting in a vapor phase, then accelerated away from the source due to a high potential applied to the source. In *ESI*, a sample dissolved in solution passes through a highly charged needle and then passes to the inlet of the mass spectrometer, sometimes with the aid of nitrogen gas

spectrometry on polypeptide fragments of proteins as produced by digestion of a protein mixture with trypsin.

Fundamentally, mass spectrometers measure the molecular mass of a polypeptide and additional structural information, such as amino acid sequence or post-translational modifications, can be inferred. In its most basic structure, a mass spectrometer has three functions: (1) ionization, the production of gas-phase ions from the sample; (2) mass analysis, the separation of gas-phase ions according to their mass-to-charge (m/z) ratio; and (3) detection of separated ions. Initially the production of gas-phase ions proved difficult for biological samples due to "excessive" fragmentation until the advent of MALDI and ESI. These two "soft ionization" methods made it possible to generate ions from intact biomolecules.

7.3.1 Ionization

7.3.1.1 MALDI

MALDI takes advantage of lasers and matrix material to generate charged ions. Protein samples are dissolved in a matrix solution (typical compounds used are α-cyano-4-hydroxycinnamic acid or dihydrobenzoic acid). Samples in submicroliter to microliter volumes are allowed to dry on a metal substrate. After drying, samples are irradiated with nanosecond laser pulses. The matrices that dissolve the samples differ in energy necessary to desorb from surface. The matrix also affects the amount of fragmentation the samples undergo. The energy necessary to desorb is inversely correlated with ion stability. Methods have been developed to increase the stability of peptides in the ionization process, for example, the addition of

nitrocellulose increases the representation of peptides (Jensen et al. 1997). α-Cyano-4-hydroxycinnamic acid matrix solutions generally lead to highest sensitivity in MALDI for biological samples.

7.3.1.2 ESI

In ESI, samples remain in liquid form and the analytes are pumped at submicroliter to microliter per minute flow rates through a needle that is under high voltage. This voltage electrostatically disperses the sample into droplets that evaporate into charged vapor droplets. A sheath gas is used to aid in the transfer of the vapor into the mass spectrometer. This technique has the advantage of being gentler on polypeptides than MALDI. It can also be used in tandem with liquid chromatography. To be analyzed by ESI-MS, molecules must have sufficient polarity to allow attachment of a charge. The signal strength, which is essentially the peak height in the spectrum, increases linearly with the analyte concentration over a wide range until saturation occurs. There does not seem to be an upper mass limit to analysis by ESI-MS. Large ions, like proteins, are typically and are therefore in the range of mass-to-charge (m/z) ratios of typical mass spectrometers. The distribution of charges gives rise to a multiple charge envelope but spectra can be simplified by deconvolution, an algorithm that sums up the signal intensity into a single peak at the molecular weight of the analyte. Very complex mixtures can be analyzed by ESI-MS, but the spectra become increasingly difficult to interpret with increasing molecular weight and numbers of compounds.

Each ionization technique has advantages. ESI may be coupled to liquid chromatography systems. The investigator may use various gradients and separation techniques (e.g., C18 columns) to separate peptides prior to ionization to enhance the resolving power of the mass spectrometer. However, unlike the MALDI ionization set-up, the investigator may not go back to source (re-gain the sample) in order to look at it again. For ESI, the entire experiment would have to be re-run with a comparable sample. In MALDI, the samples are usually added to the matrix in spots on the target and the laser usually only irradiates a small area of that spot. Therefore, the investigator can analyze the same sample again and again. This has the advantage of increasing the mass accuracy and confidence in the data.

Electrospray ionization is the most common form of ionization coupled to Fourier transform mass spectroscopy (FTMS) and has the advantage of being configured with a nanospray source. Typical specifications for a nanospray set-up include flow rates on the order of 250 nL/min and sensitivity in the femtomolar range. Using an LC coupled to the mass spectrometer, the investigator can run a number of experiments in-line with the instrument. Complex mixtures (e.g., serum) can be separated on a column based on protein chemistry prior to ESI-MS analysis. Additionally, in some cases, samples may be fractionated into n numbers of fractions prior to MS analysis using the techniques described above. Each fraction can then further be separated on a column in-line with the instrument resulting in lower complexity fractions and therefore more proteins/peptides identified.

7.3.2 Mass Analyzers

The function of the mass analyzer is to separate ions based on their mass-to-charge ratio. A number of types are available and will be discussed in this section (Fig. 7.5).

7.3.2.1 Quadrupole Mass Analyzer

The most commonly used mass analyzer is the quadrupole mass analyzer which is often referred to as a mass filter. This analyzer has four adjacent metallic rods that

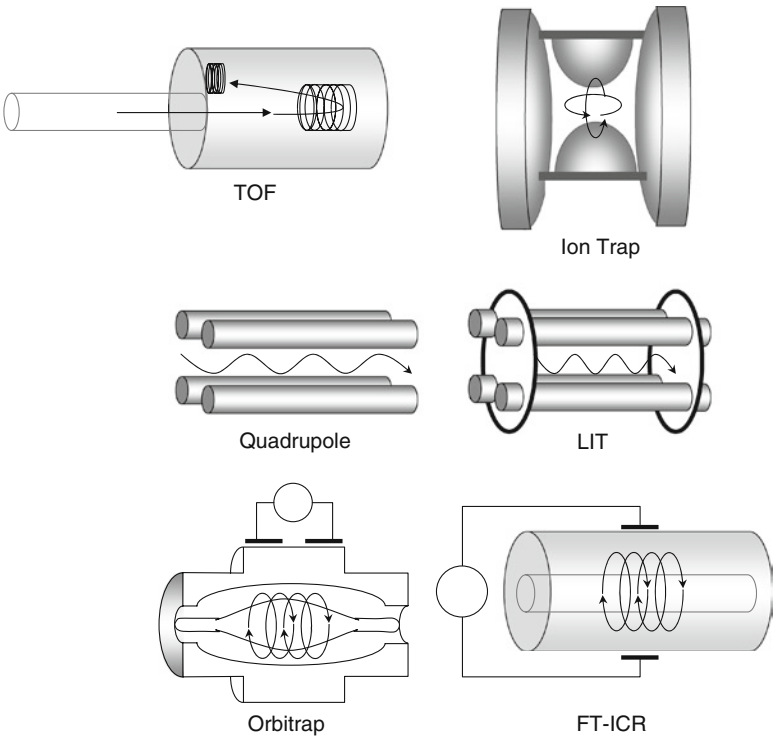

Fig. 7.5 Schematic of common mass analyzers. The *TOF* mass analyzer has a reflectron at the end to correct for shifts in flight times. The *Ion Trap* is used to perform MS/MS, can be used in tandem, and is often connected to ESI. Of the instruments available it has a low mass accuracy and resolution. *Quadrupole* mass filters can operate in tandem; in the triple quadrupole, the ion activation often occurs in the second quadrupole. *LIT* is similar to the quadrupole; it has endcaps (with DC potentials) to allow ion trapping along the long axis. The *Orbitrap* is another relative of the ion trap except with resolution and mass accuracy comparable to FT-ICR. *FT-ICR* mass spectrometers provide the highest resolution and mass accuracy of mass spectrometers available

are connected pair wise; each pair is set to a positive or a negative electrical potential. A combination of direct current (DC) and radio frequency (rf) voltages are applied between the rods in order to move the ions through the quadrupole. Depending on voltage, only ions of a given *m*/*z* value travel along the analyzer to the detector, while other ions collide with rods and are lost. By scanning the DC and RF voltages, while keeping the ratio constant, ions with different *m*/*z* ratios pass through to the analyzer successively so that a wide *m*/*z* range may be scanned (March 1997; Schuchardt and Sickmann 2007). With this design, ions may be "trapped" in a defined volume (i.e., ion trap) or drift downstream into other cells, such as other quadrupoles (tandem mass spectrometry will be discussed further below).

7.3.2.2 Time of Flight

The time of flight (TOF) mass spectrometer has a simple design: The ions are accelerated across a "field-free-drift region" of the flight tube and the velocity of the ions in the analyzer tube is dependent on their *m*/*z* values. The typical set-up has the ions traveling through a flight tube and then reflected at the end to a detector by an ion mirror called a reflectron (Karas and Hillenkamp 1988). This set-up is preferred because the alternative, the linear tube design, has relatively poor mass resolution. The reflector at the end of the flight tube is used to correct for initial energy differences; it corrects for the error in flight times by focusing the ions with the same *m*/*z* in space and time before they hit the detector. With the reflectron TOF, resolution up to 25,000 is easily achieved. MALDI ionization is often combined with time-of-flight (TOF) analyzers.

7.3.2.3 Ion Trap

In ion trap analyzers the ions are trapped in a cell for a certain time interval and then subjected to MS or MS/MS analysis. These ions are trapped using electric fields and limit for the amount of ions trapped is based on their space charge (Louris et al. 1987). The maximum number of ions is just below the number that distorts the applied field. The ions are then subjected to another electric field that ejects ions from the trap, resulting in a mass spectrum. For MS/MS, the unwanted ions are ejected first then the ions of interest are fragmented further and analyzed. Ion trap mass analyzers provide fast scanning rates, sensitivity, flexibility, and robustness and have the advantage of relatively low cost.

Ion traps have two primary designs. One is the quadrupole ion trap (QIT) and the second is the linear, or 2D, ion trap (LIT). The QIT is structurally a quadrupole analyzer that uses a combination of the rf and DC voltages to select ions of a particular *m*/*z*, but ions are trapped in the three dimensions of the cell. By ramping up the rf voltages, the QIT moves the ions out of the trap to the detector and the spectrum is scanned. These ion traps are ideal for MS/MS experiments, since the trapped ions may be excited via collision-induced dissociation (CID) to generate

the fragment (or MS/MS) spectrum. The linear ion trap is similar to the QIT, except that there are additional DC potentials to allow ions to be trapped along the long axis of the quadrupole. Ions are ejected either radially (as in the Thermo-LTQ) or axially (as in the ABI/Sciex Q-Trap) through a series of ramping protocols (Khalsa-Moyers and McDonald 2006). The primary advantage of QIT over LIT is that QIT has a greater trapping volume and therefore analyzes more ions per cycle. This improves the sensitivity and dynamic range of the ion trap.

7.3.2.4 Ion Cyclotron Resonance

One of the most powerful mass spectrometers on the market is the Fourier transform ion cyclotron resonance (ICR) mass spectrometer (FT-ICR or FTMS). The FT-ICR is a trapping mass spectrometer that captures ions under high vacuum in a high magnetic field. ICR was developed by Comisarow and Marshall in 1974 (Comisarow and Marshall 1974). In the ICR, ions travel forward and rotationally through the mass spectrometer (similar to a corkscrew motion) under high vacuum and within a magnetic field. The applied field resonates with the ions and as their "cyclotron" path is widened the rotational speed is measured in order to determine the ion size. The FT-ICR mass spectrometer was a breakthrough in resolving power and mass accuracy (Senko et al. 1997; Domon and Aebersold 2006). The FT-ICR provides high quality data and allows the detection of more signals than do instruments of lower resolving power. The development of a hybrid FT-ICR instrument with an external LIT device allows parallel full mass spectrum (MS1) and tandem mass spectrum (MS2) acquisition (not sequential); the high-quality MS1 data can be used for quantification. The system is limited by a relatively slow acquisition rate (several s per cycle).

7.3.2.5 Orbitrap

Orbitraps are similar to ICR mass spectrometers, except that rather than using a magnetic field, an electric field is used (Makarov 2000; Hardman and Makarov 2003). The Orbitrap radially traps ions about a central spindle electrode. The outer barrel-like electrode is coaxial with the inner spindle-like electrode and mass/charge values are measured from the frequency of harmonic ion oscillations of the orbitally trapped ions. Ion frequencies are measured non-destructively by acquisition of time-domain image current transients (Makarov 2000; Schuchardt and Sickmann 2007). In simpler terms, instead of measuring the ions' rotational frequencies, the translational motion along the long axis of the Orbitrap cell is measured. The Orbitrap provides high resolution (up to 150,000) and high mass accuracy (2–5 ppm) (Hardman and Makarov 2003; Hu et al. 2005). Even though, the FT-ICR provides better resolution and mass accuracy, the Orbitrap is an attractive alternative for most users and applications because there is no need for liquid helium or liquid nitrogen, as there is with the large superconducting magnet in the ICR. Similar to other mass analyzers, it may be combined with either MALDI or ESI sources.

7.3.3 *Using a Mass Spectrometer to Identify Species in a Mixture*

Given that tandem mass spectrometry is so powerful, how exactly is it used to identify ions and assign them to peptides? How is sequencing done in a mass spectrometer? Here we describe the process of sequencing using the mass spectrometer (Fig. 7.6).

7.3.3.1 Collision-Induced Dissociation

In the CID experiment, precursor ions are isolated and subjected to a neutral target gas. The gas collides with the precursor ions passing-on kinetic energy to the ions (Sleno and Volmer 2004). Multiple collisions increase the internal energy of the ions, resulting in fragmentation of the peptide backbone primarily at the amide bonds (Sleno and Volmer 2004). This results in b and y ions (Fig. 7.7). In CID, both low and high energies are used to fragment the ions. Low-energy collisions are used in ion trap and quadrupole instruments. This low-energy fragmentation results in a, b, y, and immonium ions, and ions from the neutral loss of ammonia. In the high-energy CID, the ions observed are d, v, w, and immonium ions. For mass spectrometers such as triple quadrupoles, ion traps, or TOF instruments, ions are isolated and

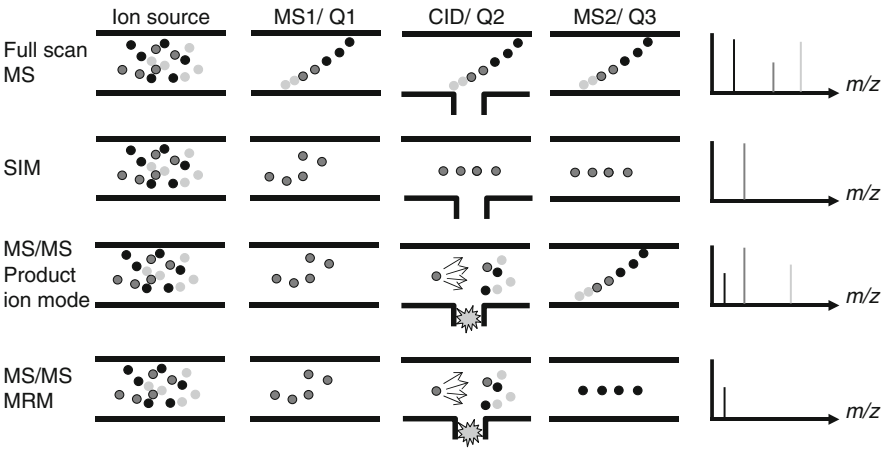

Fig. 7.6 Schematic of the different types of tandem mass spectrometry experiments, including the triple quadrupole experiments. Both *Full Scan MS* mode and *SIM* are examples of experiments using single quadrapole mode. Full Scan mode is a simple scan of all the ions from the ion source. In *SIM* mode, an ion is selected in the first quadrupole and then scanned. In *Product Ion Mode*, a precursor ion is selected in Q1, then activated in Q2 (collision induced dissociation region), and then scanned in Q3. The *MRM* mode may be a series of experiments that first select precursor ions and then filter the resulting fragment ions post Q2. The experiment may be set up to look for multiple resulting fragment ions in Q3

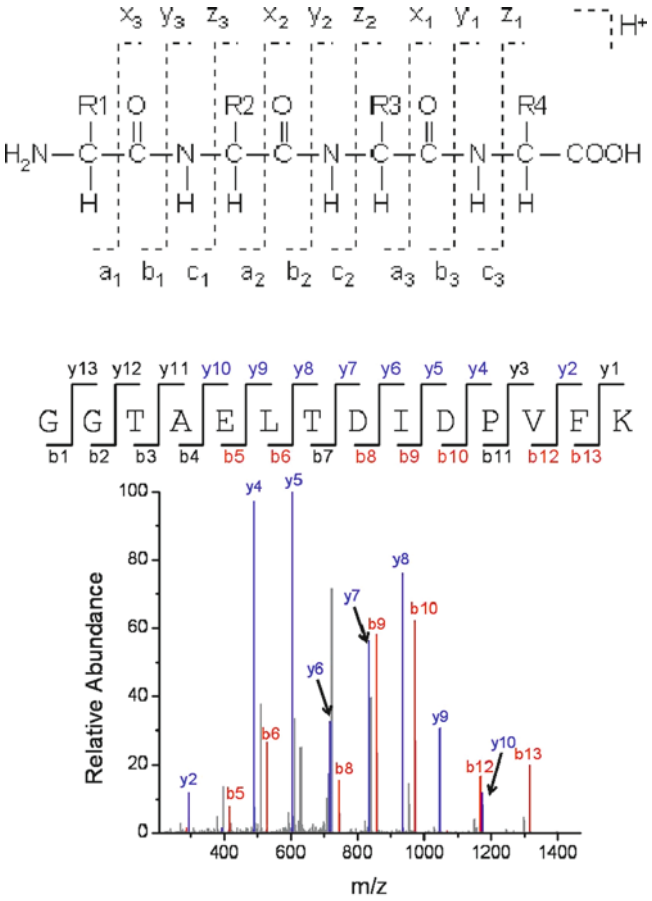

Fig. 7.7 Peptide fragment ion nomenclature. Resultant ions from cleavage of bonds along the peptide backbone. Typically b- and y-ions result from CID whereas c- and z-ions result from ECD

fragmented in a collision cell. In tandem mass spectrometers, such as QQTOF, ions enter the first quadrupole, ions of interest are then sent to the second quadrupole and the cell is filled with target gas. The resultant fragments are then sent downstream to a TOF detector.

7.3.3.2 Electron Capture Dissociation

Electron capture dissociation (ECD) was first introduced by Zubarev et al. (Zubarev et al. 1998, 2000). For ECD, a low-energy electron beam (<0.2 eV) generated by a heated filament electron gun is used for activation (Zubarev et al. 1998; Cooper et al. 2002). The positively charged precursor ions capture the electron leading to

neutralization and backbone fragmentation. Backbone cleavage usually occurs at the N–Cα bond yielding primarily c and z ions (Zubarev et al. 1998). ECD preferentially cleaves disulfide bonds but leaves other post-translational modifications intact (Zubarev 2004). ECD is an available option on most FT-ICR instruments and has been used to characterize post-translational modifications such as O-linked glycosylation, methionine oxidation, and phosphorylation (Bakhtiar and Guan 2005; Zhang et al. 2005b).

7.3.3.3 Electron Transfer Dissociation

The ECD activation method is not amenable to use in ion trap instruments. However, an analogous ion-ion method called electron transfer dissociation (ETD), where ion–ion reactions occur between singly charged anions and multiply charged peptide cations, was developed by Hunts and colleagues (Syka et al. 2004a). The electron source in ETD is a chemical ionization. The anions are introduced into the trap via an anion beam controlled by RF gating voltages. The anions interact with the multiply charged peptides resulting in a proton transfer without dissociation and in electron transfer with or without dissociation. Proton transfer results in charge reduction, whereas dissociation leads to c- and z-fragmentation, quite similar to what is observed in ECD (McLafferty et al. 2001).

7.3.3.4 Infrared Multiphoton Dissociation

Infrared multiphoton dissociation (IRMPD) is a slow heating dissociation method involving non-resonant ion activation and subsequent dissociation via photon absorption (Khalsa-Moyers and McDonald 2006). This method was historically used for small molecule analysis, but IRMPD has recently been applied to protein analysis (Shukla and Futrell 2000; Sleno and Volmer 2004). In the IRMPD experiment a low-powered CO_2 laser is used to activate ions. This laser is useful for analyzing phospho-peptides, because the phosphate preferentially absorbs at this wavelength. Since IRMPD is a low-energy ionization, it shows similar fragmentation patterns to CID (Zhang et al. 2005b).

7.4 Data Analysis

7.4.1 Identification

Peptide sequencing and identifying the peptides being fragmented in the mass spectrum is key to mass spectrometry and proteomics. How these fragmented ions are identified can be organized into three main categories: (a) database searching, similar to spectral library searching, where peptide sequences are identified based

on theoretical spectra predicted for that sequence or based on spectra from previous experiments; (b) *de novo* sequencing, where peptide sequences are read out directly from fragment ion spectra; and (c) hybrid techniques, where short stretches of the peptides are sequenced then the rest of the spectrum is searched through databases. For a comprehensive review of the publicly available tools for MS/MS-based proteomics, see (Nesvizhskii et al. 2007).

Peptide mass fingerprinting (PMF) is considered one of the fastest methods for identifying proteins recovered after gel electrophoresis or other isolation methods that provide samples containing one or two proteins. In PMF, the protein of interest is isolated from a gel and digested with a proteolytic enzyme (e.g., trypsin which selectively cleaves the protein at lysines and arginines) (James et al. 1993; Mann et al. 1993; Pappin et al. 1993). The mass spectrum obtained from a MALDI-TOFF is then searched against the masses from a known protein/peptide databases. This method has the advantage of speed, it is much faster than the labor intensive *de novo* sequencing, however it is only effective if the protein in question has actually been sequenced and is in the database.

In database searching, the spectrum of a protein is scored against theoretical fragmentation patterns constructed for peptides found in the searched databases. The peptides queried are restricted to investigator specified criteria (e.g., proteolytic enzyme and post-translational modifications allowed). Once a spectrum is matched against spectra from the database, a list of ranked peptides (scored according to the parameters set by investigator) is returned. Discerning a true match from a false match is critical in proteomic data analysis. The higher the score the more confident the investigator is that the peptide is a positive match. There are a number of scoring schemes: spectral correlation functions (e.g., SEQUEST) (Eng et al. 1994), shared fragment counts and dot products (e.g., TANDEM, OMSSA, MASCOTT) (Perkins et al. 1999; Craig and Beavis 2004; Geer et al. 2004), empirically observed rules (e.g., Spectrum Mill), and fragmentation frequencies (e.g., PHENYX) (Colinge et al. 2003). These scores can be converted into an expectation value (E value), which is the expected number of peptides with scores equal to or better than observed score under the assumption that peptides are matching the experimental spectrum by random chance (e.g., OMSAA, TANDEM and MASCOT).

Despite success of database and spectral matching searching, false peptide assignments occur for a number of reasons. Reasons for false assignments are use of simplified scoring algorithms, contaminants, low quality spectra, fragmentation of multiple peptide ions, presence of homologous peptides, incorrectly determined charge state or peptide mass, restricted/limited database search, sequence variants and new peptides (Nesvizhskii et al. 2007). The generation of “high confidence” identifications is the goal in proteomics, but scoring is software/tool-dependent. The score distribution depends on mass spectrometer performance, the quality of the sample, the instrument settings and its methods, and the size of the database. The quality of the score may be improved by approaches such as target-decoy searching and use of empirical Bayes methods (Keller et al. 2002; Storey and Tibshirani 2003; Elias and Gygi 2007). The target-decoy method is when the peptide in question is searched against the database of peptides in reverse order or with

"shuffled" sequences. Subsequently, peptides are filtered with score cut-offs. This method is useful in "weeding" out false positives but has the disadvantage of requiring twice the computing time. Programs such as PeptideProphet employ empirical Bayes approaches to validate peptide assignments made by database search programs. From each dataset, it learns distributions of search scores and peptide properties among correct and incorrect peptides and uses those distributions to compute probabilities that assignments are correct.

Identifying peptides based on their spectral match to a spectrum in a spectral library has been expanded because of the in-depth coverage of proteomes in eukaryotic species (Brunner et al. 2007; Brill et al. 2009). With extensive maps already in existence, the expansion to the proteomes of other systems by spectral matching to spectral libraries will grow at a much faster rate (Yates et al. 1998; Craig et al. 2006; Frewen et al. 2006; Lam et al. 2007).

In *de novo* sequencing, amino acids in the peptide are directly read from the fragment ion spectrum, facilitated by tools such as PepNovo and PEAKS (Ma et al. 2003; Frank and Pevzner 2005). Direct sequencing is helpful in cases where peptides are modified or there are polymorphisms. When limited genome information is known about a host, *de novo* sequencing or a hybrid of *de novo* with database spectral searching must be used. Posttranslational modifications (PTMs) are best identified through direct *de novo* sequencing; however, there are database searches and hybrid searches that can be employed to identify PTMs. A prominent inefficiency in the shotgun approach to proteomics lies in the redundancy of peptides seen. In complex samples as the human serum proteome, laboratory measures such as fractionation and immunological depletion experiments are done to compliment the analytical and software approach of the mass spectral data. Combining several search scores improves the overall confidence of the peptides identified. Programs such as TANDEM allow investigator-provided constraints, to allow for stricter and more confident identifications. Auxiliary run-condition information can also improve spectral identification. For example, the retention time (Strittmatter et al. 2004) and/or known sequence motifs such as the presence of N-linked glycosylation sites (Zhang et al. 2005a) can be useful.

7.4.2 Quantification

MS based methods for quantitative analyses of proteins/peptides, as discussed earlier, seek to compare two or more distinct proteomes in order to identify proteins with altered expression levels or post-translational forms in response to a given stimulus. In quantification analysis, regardless of whether the samples are isotopically labeled or label free, the ratio of intensities of the peptide peaks in a given mass spectrum gives a relative ratio of abundance of the two or more species. Several factors have to be considered when performing quantitative experiments. When choosing a stable isotope label, it must be determined whether the label alters the physicochemical properties of a peptide. For example, there is minimal impact

when using ^{13}C, ^{15}N, or ^{18}O labeling (Zhang and Regnier 2002), but deuterium labeling can be problematic as labeled and unlabeled peptides differ in their retention time in RP-HPLC (Zhang et al. 2001). This results in an inaccurate quantitation data analysis and requires an additional signal integration step over retention time to correct for the inaccuracy (Fig. 7.8).

As mentioned earlier, to prevent spectral overlap, the stable isotope label incorporation should result in at least 4 Da shift relative to the unlabeled peptide. Another area of great effect on the accuracy of quantification is the quality of spectrum. Data should be handled with scrutiny when the signal is very low (close to the noise level) or very high (possibly resulting in detector saturation) as both will lead to distortion of the isotope envelope intensity leading to inaccurate quantitation. It is also dependent on the ability of the instrument to discriminate between interfering signals resulting from co-eluting peptides and the peptide isotope envelopes. Even though this can be minimized by reducing the sample complexity through fractionation, it should be noted that analytes often do not elute in a narrow profile and sometimes even elute into two or more fractions in separated regions of the elution profile (Faca et al. 2007).

In MS/MS based quantitation studies, the detector saturation problem is minimal and quantitation is not dependent on the machine's mass resolution, but on the size of the sequencing window, therefore it is background contributions that may bias the results. Hence, the *m*/*z* window used for sequencing should be optimized for every run. When employing the spectral counting technique, results can be computed in any of several ways. The simplest reports the average of ratios (Saito et al. 2007) while using an intensity threshold in order to minimize the noise based bias (Wolf-Yadlin et al. 2007). More reliable results are achieved when the ratios are computed based on the intensity weighted average, on the sum of all the observed spectra (Ono et al. 2006; Saito et al. 2007), or by employing linear regression (Parish 1989).

7.5 Applications

In the previous section we described the technology underlying proteomics approaches. Here we briefly describe two primary applications of proteomics to biology. First we discuss the identification of post-translational modifications. Next, we describe how proteomics approaches can be used to characterize protein complexes.

7.5.1 Post-translational Modifications

Post-translational modifications (PTMs) are covalent processing events that change the properties of a protein by proteolytic cleavage or by addition of a modifying group to one or more amino acids. PTMs can determine a protein's activity state, localization,

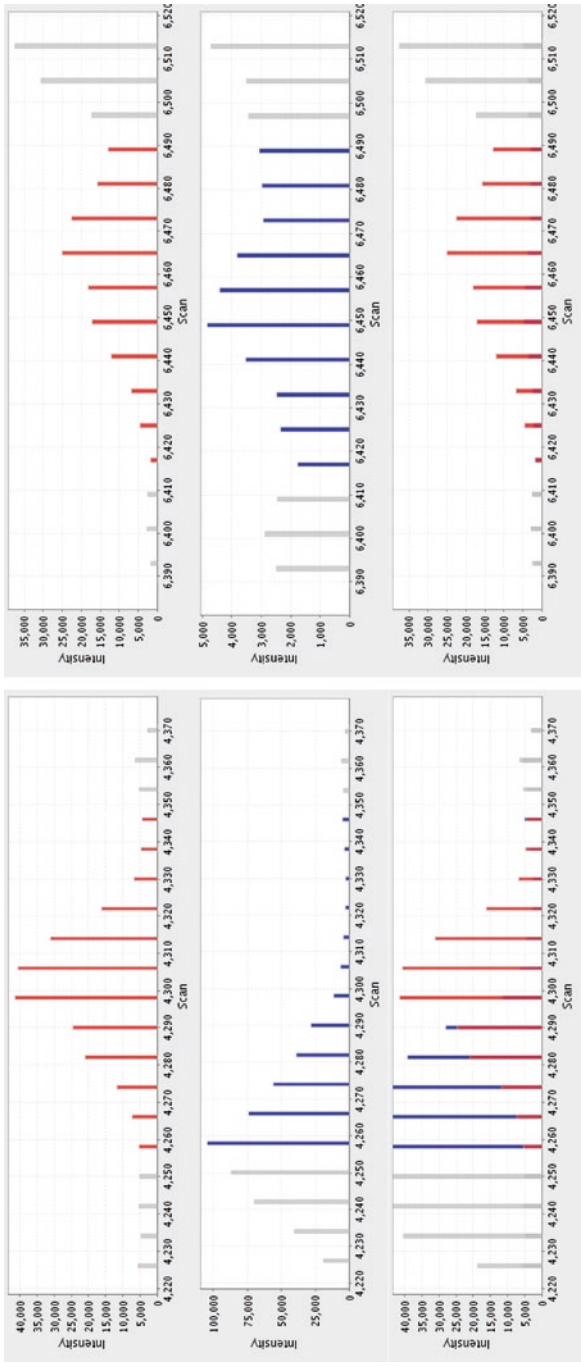

Fig. 7.8 Bias introduced by non co-eluting peptides. The elution profiles (*left*, non co-eluting; *right*, co-eluting) of peptides to be quantified and the potential of biasing the results. *Top* and *middle* represent extracted ion scans of one peptide and its isotopically labeled counterpart. *Bottom* two scans are the combination of the differentially labeled peptides

turnover, and interactions with other proteins. For example, kinase cascades critical to signaling are turned on and off by the reversible addition and removal of phosphate groups and cyclins are marked for destruction at defined time points in the cell cycle by ubiquitination. Despite the great importance of PTMs for biological function, their study on a large scale has been hampered by a lack of suitable methods and many key modifications have only been discovered late in the elucidation of various biological processes. As a result, we probably do not realize the full extent and functional importance of protein modifications in the workings of the cell.

7.5.1.1 Isolation of Modified Proteins

Modification analysis is usually done by comparison of experimental data to a known amino acid sequence. Therefore, the first step is identification of the protein to be studied, which can be done at very high sensitivity by antibody recognition (Western blotting) or by MS techniques. A central consideration in the characterization of modifications is the need for as large an amount of the protein as possible. Protein modifications are typically not homogeneous and a single gene may give rise to a bewildering number of gene products as a result of alternative splicing and the combination of different post-translational modifications. The amount of protein in a single modification state can thus be a very small fraction of the total amount of the gene product. Furthermore, as explained later, the complete characterization of the primary structure of a protein requires much more material than mere identification by MS sequencing of a few peptides.

To study the modifications of a single protein, chromatographic purifications, antibody precipitations, or both can be used to isolate sufficient amounts. Modern analysis methods tolerate contamination much better than earlier methods; therefore, the total amount of recovered protein is more important than absolute purity. Often, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) will be the final preparation step and researchers should attempt to isolate at least Coomassie-stainable amounts (several picomoles or 1 g) of protein to increase the chance of detecting and characterizing modifications.

7.5.1.2 PTM Mapping of a Purified Protein

Once a protein has been isolated, a variety of techniques can be used to determine the identities of modified amino acids. In some cases, the precise molecular weight of the intact protein can be established by MS, especially if the protein is not too heterogeneous, its mass is less than about 100 kDa, and it is in a buffer that is compatible with MS. Once the masses of the nonmodified and modified amino acid residues add up to the measured intact molecular weight, the protein is completely characterized.

Amino-terminal protein sequencing by the classical technique of Edman degradation is still the method of choice to determine proteolytic processing.

Carboxy-terminal processing can also be determined by amino acid sequencing, albeit at a much lower sensitivity. Detailed characterization of modification happens after enzymatic or chemical degradation of the protein. The resulting peptides are usually separated by high-performance liquid chromatography (HPLC) as described above. In Edman degradation, collected peptide fractions are applied to the sequencer and their amino acid sequence determined. Modified amino acids become apparent because of their absence or retention-time shift in the corresponding sequencing cycle. If the mass of the intact peptide has been determined, then the nature of the modification can be confidently assessed.

Often, the peptide mass pattern will hint at the nature of the modification, such as multiple mass differences of 162 Da for glycosylation or the presence of a "satellite mass" less 98 Da in the case of phospho-serine and phospho-threonine because of the elimination of phosphoric acid.

The mass of the modified peptide is usually not sufficient to determine the nature of the modification so peptides are fragmented by MS to localize the modification. In these "tandem mass spectrometry" (MS/MS) experiments, peptide ions are collided with inert gas, leading to fragmentation, usually at the peptide bonds. Some modified amino acid residues remain intact during this process. In this case, the fragmentation pattern is similar to the unmodified peptide with the difference that the location of the modified amino acid is revealed by its mass increment. Thus, ideally, the mass and location of the modification can be determined. In practice, the fragmentation pattern may or may not allow exact localization of the modification, depending on the completeness of the fragmentation pattern.

If the modification is labile, then it will be lost before the peptide itself fragments. In this case, the peptide can still be sequenced and identified, but only the mass increment (not the location of the modification) is determined. Examples of stable modifications are acetylation (+42 Da), which is found on the N termini of many proteins or on specific lysine residues, and arginine methylation (+14 Da). Examples of labile modifications are O-linked *N*-acetylglucosamine (GlcNAc; +203 Da) and sulfation (+80 Da). The phospho group (+80 Da) can be stable (e.g., in the case of phospho-tyrosine) or relatively labile (e.g., in the case of phospho-threonine and especially phospho-serine).

Modifying groups that are easily lost from the peptide can themselves be used as "reporter groups" to detect the presence of the modified peptide in several different ways: In "in-source fragmentation," excess energy in the ionization or ion-sampling process leads to the characteristic presence of the reporter ion in mass spectra. Subsequent sequencing of the peptide peaks can then identify the modified peptide. Conversely, in the "neutral loss" technique, mild collisions in the collision cell between the two sections of a tandem mass spectrometer lead to loss of the modifying group. The second mass analyzer is set at a mass offset corresponding to the mass-to-charge ratio (m/z) of the expected modification. A signal can only reach the detector if the peptides were modified and the mass changed by the expected amount during collision.

7.5.1.3 PTM Mapping of Protein Populations

Although the methods just described are very powerful for the characterization of individual, purified proteins and have helped elucidate numerous biological mechanisms, the real promise of proteomics is to assess systematically the modifications of large numbers of proteins. There are three commonly used approaches: analysis of affinity purified proteins using LC MS/MS, analysis of peptides using LC MS/MS, and derivatization-based methods.

The strategy of affinity-based enrichment of modified proteins combines established biochemical, genetic, and immunological methods for enrichment of modified-protein populations with recently developed MS techniques for protein mixture analysis. This strategy is particularly attractive because the enrichment step is often a single experiment (e.g., an immunoprecipitation) and the subsequent identification of the protein mixture is usually reduced to a single LC MS/MS experiment as well.

The phospho-proteome has been extensively explored with this strategy. For example, cells stimulated with EGF can be immunoprecipitated with anti-phosphotyrosine antibody. Another modification of great interest, the enzymatic attachment of ubiquitin to cellular proteins that marks them for destruction, is also under investigation. In an elegant experiment, yeast ubiquitin was replaced by a histidine-tagged version, allowing selective purification and identification of the ubiquitinated proteome. As these examples show, the combination of selective enrichment of modified proteins with MS mixture analysis can be very powerful. The critical step is the development of the enrichment protocol. Subsequently, the proteins have only to be identified, thus avoiding the difficulties of detailed modification mapping mentioned earlier.

Recent technological developments have made it increasingly feasible to directly analyze very complex peptide mixtures by LC MS/MS. A single chromatographic run can result in the identification of hundreds of modified peptides, especially as following metal or affinity enrichment strategies. Peptide mixtures derived from complex protein mixtures are very difficult to analyze comprehensively. If one is interested in specific modifications, the peptide complexity can be reduced by affinity methods. For example, phospho-peptides can be captured selectively through their negatively charged phospho group on immobilized-metal affinity (IMAC) columns. Recently, this technique has been made much more specific by esterifying, and thereby neutralizing, the negatively charged amino acid residues before the IMAC step, allowing identification of hundreds of phospho-peptides in yeast cell lysates. The method has also been used in combination with phospho-tyrosine protein affinity purification.

Chemical derivatization of the modifying group potentially allows attachment of a “hook” for affinity purification. For example, the phosphate group can be converted to an affinity tag by an elimination/Michael addition reaction or by phosphoamidate chemistry. It should be noted, however, that only very simple and extremely efficient chemical derivatization steps are compatible with proteomics. If any heterogeneity is introduced by the chemical reaction (e.g., as a result of <100% conversion efficiency or side reactions), the peptide samples become even more complex and it is then only possible to analyze modifications of the most abundant proteins.

7.5.2 Identification of Protein Complexes

Vital cellular functions such as DNA replication, transcription, and mRNA translation require the coordinated action of a large number of proteins that are assembled into an array of multiprotein complexes of distinct composition and structure. Similarly, biological processes are orchestrated and regulated by dynamic signaling networks of interacting proteins that link chemical or physical stimuli to specific effector molecules. The analysis of protein complexes and protein–protein interaction networks– and the dynamic behavior of these networks as a function of time and cell state– are therefore of central importance in biological research.

7.5.2.1 Affinity Purification

Different approaches have been used to characterize protein complexes and protein–protein interaction networks. The first interactome maps were obtained using a yeast two hybrid approach. More recently, a combination of affinity purification and mass spectrometry (AP–MS) has been used to greatly advance our understanding of protein-complex composition. With the AP–MS method, multiprotein complexes are isolated directly from cell lysates through one or more AP steps. Complex components are then identified by MS. In contrast to yeast two-hybrid and related methods, AP–MS can be performed under near physiological conditions and in the relevant organism and cell type. AP–MS does not typically perturb relevant post-translational modifications, which are often crucial for the organization and/or activity of complexes. Another advantage of AP–MS is that it can be used to probe dynamic changes in the composition of protein complexes, especially when used in combination with quantitative proteomics techniques.

Standard approaches that use affinity-tagged recombinant proteins have allowed for parallel sample preparation without the need to optimize the purification protocol for each protein complex. Proteins of interest are simply expressed in-frame with an epitope tag (at either the N or C terminus), which is then used as an affinity handle to purify the tagged protein (the bait) along with its interacting partners (the prey). Although several different tags or tag combinations have been successfully used in many low-throughput studies (see Cummings and Kornfeld 1982), high-throughput studies have primarily used either the flag or tandem affinity purification (TAP) tags.

In the flag-tag approach, C-terminally flag-tagged proteins are expressed under the control of a GAL-inducible promoter and isolated in a single step using an anti-flag antibody resin. In the TAP-tag approach genes, the proteins of interest are fused to a C-terminal dual-epitope tag via homologous recombination, such that the proteins were expressed under their own promoters. Protein purification is carried out in two steps, first via the protein A moiety in the TAP tag (which binds immunoglobulin G (IgG)–sepharose) and then via the calmodulin-binding peptide (which exhibits high affinity to calmodulin–sepharose).

The AP–MS technique generates a list of proteins detected in a given sample but does not necessarily reveal the composition of individual protein complexes. The data from a single AP–MS experiment represents an average of binding partners and protein complexes. If the bait protein is a component of multiple alternative complexes, a single AP–MS analysis cannot be used to decipher this multiplicity of associations. This is an important limitation because proteins can have dramatically different roles as components of different types of complexes. The structure of multiprotein complexes can only be revealed indirectly through high-density AP–MS approaches. However, as described below, analysis of the composition of an intact protein complex with defined biochemical properties can be used to directly reveal the composition of a given complex.

7.5.2.2 Biochemical Fractionation in Protein Complex Analysis

The fractionation approaches described above have been used widely for the separation and enrichment of protein complexes. AP of at least one of the sample components using, for example, an inhibitor or a ligand has also frequently been included in biochemical purification schemes to significantly increase enrichment. Depending on the nature of the particular protein complex, a combination of these separation methods can yield pure preparations.

Although fractionation approaches have been used successfully for the characterization of the composition of numerous biologically relevant protein assemblies, these methods are not generic and must be tailored to a particular complex of interest. This limitation prevents their application to genome-wide studies. However, combining one or more of such biochemical fractionation techniques with a generic AP protocol (such as an epitope tag) can provide a surrogate for a complete biochemical isolation of a protein complex.

Another strategy for the analysis of large multiprotein assemblies (or organelles) is to monitor co-fractionation profiles using quantitative MS and then to compare the acquired profiles with those of known components of the protein complex or organelle of interest. This can be accomplished by monitoring the number and intensity of the peptide signals for each detected protein across adjacent fractions (for example, throughout a sucrose or glycerol gradient).

7.5.2.3 Crosslinking of Protein Complexes

A problem that is encountered during the isolation of intact native protein complexes from cells or tissue is that only protein–protein interactions that are resistant to the lysis and purification conditions will survive to be detected by MS. Several different strategies have been devised to freeze transient or labile protein interactions by using chemical crosslinking reagents. Crosslinkers possess at least two reactive groups that form covalent bonds with target molecules. These reactive groups are separated by a spacer arm of a defined length (usually in the range of

5–15 Å) that determines the maximal distance between two molecules. This confers some degree of specificity to the crosslinking process: Molecules in close proximity are more likely to be crosslinked than distant species. However, protein–protein crosslinking techniques present multiple experimental and analytical challenges. The choice of crosslinker is crucial, as crosslinkers vary in cell-permeability, reactivity, and arm length. Crosslinking reaction conditions must also be closely monitored, such that bona fide protein–protein interactions are stabilized and undesired crosslinks (to contaminating proteins) are minimized.

Although many chemical crosslinkers can be used to stabilize complexes in theory, only a few have been successfully used for in vivo crosslinking followed by MS analysis. Formaldehyde and di-thiobis-succinimidyl-propionate (DSP; an amine-reactive, homobifunctional, thiol-cleavable and membrane-permeable crosslinker) have been used most often to identify novel interacting partners. Crosslinkers are also particularly attractive for revealing interactions that involve membrane proteins (or microsomes), as the detergent concentrations used to solubilize the membranes and extract the proteins typically also disrupt protein–protein interactions. Mild formaldehyde crosslinking has also been performed in animals and has allowed for the stringent immunopurification of an intact secretase complex as well as the identification of protein interactors for the cellular prion protein.

7.5.2.4 Complex Stoichiometry

Determining complex stoichiometry by mass spectrometry has thus far been challenging. However, one promising strategy to determine protein stoichiometry in a complex is to combine complex isolation with isotope-based absolute quantitative proteomics. If all of the components of a complex are known, synthetic tryptic peptides can be generated to monitor the abundance of each of the proteins in the complex. These peptides can be synthesized with heavy isotopes and then mixed with an unlabelled sample (as in the AQUA approach) or labeled in parallel to the samples (e.g., by reaction with iTRAQ).

7.6 Summary

Technologies to quantitatively interrogate proteomes are becoming a standard part of the arsenal of biologic researchers. These technologies have evolved rapidly over the past decade. Owing to the complex and dynamic nature of proteomes, it has now become clear that there is no “one-fits-all” strategy to address biological questions. While techniques for say analyzing protein complexes are already quite mature, experiments such as global protein expression profiling for biomarker discovery, are still under development. In this chapter, we attempted to provide a technical guide to the three main components of proteomics: (1) sample preparation, (2) mass spectrometric analysis, and (3) data analysis. We then highlighted two mature

applications of these components to demonstrate that a fundamental canon of approaches (e.g., fractionation, followed by LCMS) can be assembled into powerful pipelines to ask and answer important biological questions.

References

Alvarez-Manilla G, Atwood Guo Y, Warren NL, Orlando R, Pierce M (2006) Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites. J Proteome Res 5:701–708.

Andersen JS, Wilkinson CJ, Mayor T, Mortensen P, Nigg EA, Mann M (2003) Proteomic characterization of the human centrosome by protein correlation profiling. Nature 426:570–574.

Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1:845–867.

Bakhtiar R, Guan Z (2005) Electron capture dissociation mass spectrometry in characterization of post-translational modifications. Biochem Biophys Res Commun 334:1–8.

Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat Methods 2:587–589.

Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M (2003) A proteomics strategy to elucidate functional protein–protein interactions applied to EGF signaling. Nat Biotechnol 21:315–318.

Blagoev B, Ong S-E, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. Nat Biotechnol 22:1139–1145.

Bondarenko PV, Chelius D, Shaler TA (2002) Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography–tandem mass spectrometry. Anal Chem 74:4741–4749.

Brill LM, Motamedchaboki K, Wu S, Wolf DA (2009) Comprehensive proteomic analysis of Schizosaccharomyces pombe by two-dimensional HPLC-tandem mass spectrometry. Methods 48:311–319.

Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C, de Lichtenberg U, Rinner O, Lee H, Pedrioli PG, Malmstrom J, Koehler K, Schrimpf S, Krijgsveld J, Kregenow F, Heck AJ, Hafen E, Schlapbach R, Aebersold R (2007) A high-quality catalog of the *Drosophila melanogaster* proteome. Nat Biotechnol 25:576–583.

Bunkenborg J, Pilch BJ, Podtelejnikov AV, Wisniewski JR (2004) Screening for *N*-glycosylated proteins by liquid chromatography mass spectrometry. Proteomics 4:454–465.

Burke TW, Mant CT, Black JA, Hodges RS (1989) Strong cation-exchange high-performance liquid chromatography of peptides. Effect of non-specific hydrophobic interactions and linearization of peptide retention behaviour. J Chromatogr 476:377–389.

Bylund D, Danielsson R, Malmquist G, Markides KE (2002) Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography–mass spectrometry data. J Chromatogr A 961:237–244.

Che F-y, Fricker LD (2002) Quantitation of neuropeptides in Cpefat/Cpefat mice using differential isotopic tags and mass spectrometry. Anal Chem 74:3190–3198.

Chelius D, Bondarenko PV (2002) Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res 1:317–323.

Chelius D, Zhang T, Wang G, Shen R-F (2003) Global protein identification and quantification technology using two-dimensional liquid chromatography nanospray mass spectrometry. Anal Chem 75:6658–6665.

Colinge J, Masselot A, Giron M, Dessingy T, Magnin J (2003) OLAV: towards high-throughput tandem mass spectrometry data identification. Proteomics 3:1454–1463.

Comisarow MB, Marshall AG (1974) Fourier transform ion cyclotron resonance spectroscopy. Chem Phys Lett 25:282–283.
Cooper HJ, Hudgins RR, Hakansson K, Marshall AG (2002) Characterization of amino acid side chain losses in electron capture dissociation. J Am Soc Mass Spectrom 13:241–249.
Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20:1466–1467.
Craig R, Cortens JP, Beavis RC (2005) The use of proteotypic peptide libraries for protein identification. Rapid Commun Mass Spectrom 19:1844–1850.
Craig R, Cortens JC, Fenyo D, Beavis RC (2006) Using annotated peptide mass spectrum libraries for protein identification. J Proteome Res 5:1843–1849.
Cummings RD, Kornfeld S (1982) Fractionation of asparagine-linked oligosaccharides by serial lectin-Agarose affinity chromatography. A rapid, sensitive, and specific technique. J Biol Chem 257:11235–11240.
Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. Science 312:212–217.
Durr E, Yu J, Krasinska KM, Carver LA, Yates JR, Testa JE, Oh P, Schnitzer JE (2004) Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture. Nat Biotechnol 22:985–992.
Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4:207–214.
Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino-acid-sequence in a protein database. J Am Soc Mass Spectrom 5:976–989.
Faca V, Pitteri SJ, Newcomb L, Glukhova V, Phanstiel D, Krasnoselsky A, Zhang Q, Struthers J, Wang H, Eng J, Fitzgibbon M, McIntosh M, Hanash S (2007) Contribution of protein fractionation to depth of analysis of the serum and plasma proteomes. J Proteome Res 6:3558–3565.
Fan X, She YM, Bagshaw RD, Callahan JW, Schachter H, Mahuran DJ (2004) A method for proteomic identification of membrane-bound proteins containing Asn-linked oligosaccharides. Anal Biochem 332:178–186.
Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. Nat Biotechnol 20:301–305.
Frank A, Pevzner P (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal Chem 77:964–973.
Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ (2006) Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. Anal Chem 78:5678–5684.
Gao J, Opiteck GJ, Friedrichs MS, Dongre AR, Hefta SA (2003) Changes in the protein expression of yeast as a function of carbon source. J Proteome Res 2:643–649.
Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004) Open mass spectrometry search algorithm. J Proteome Res 3:958–964.
Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 100:6940–6945.
Gilchrist A, Au CE, Hiding J, Bell AW, Fernandez-Rodriguez J, Lesimple S, Nagaya H, Roy L, Gosline SJ, Hallett M, Paiement J, Kearney RE, Nilsson T, Bergeron JJ (2006) Quantitative proteomics analysis of the secretory pathway. Cell 127:1265–1281.
Glocker MO, Borchers C, Fiedler W, Suckau D, Przybylski M (1994) Molecular characterization of surface topology in protein tertiary structures by amino-acylation and mass spectrometric peptide mapping. Bioconjug Chem 5:583–590.
Goodlett DR, Keller A, Watts JD, Newitt R, Yi EC, Purvine S, Eng JK, von Haller P, Aebersold R, Kolker E (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. Rapid Commun Mass Spectrom 15:1214–1221.
Goshe MB, Conrads TP, Panisko EA, Angell NH, Veenstra TD, Smith RD (2001) Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. Anal Chem 73:2578–2586.

Goshe MB, Veenstra TD, Panisko EA, Conrads TP, Angell NH, Smith RD (2002) Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins. Anal Chem 74:607–616.

Gruhler A, Schulze WX, Matthiesen R, Mann M, Jensen ON (2005) Stable isotope labeling of *Arabidopsis thaliana* cells and quantitative proteomics by mass spectrometry. Mol Cell Proteomics 4:1697–1709.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17:994–999.

Hansen KC, Schmitt-Ulms G, Chalkley RJ, Hirsch J, Baldwin MA, Burlingame AL (2003) Mass spectrometric analysis of protein mixtures at low levels using cleavable C-13-isotope-coded affinity tag and multidimensional chromatography. Mol Cell Proteomics 2:299–314.

Hardman M, Makarov AA (2003) Interfacing the orbitrap mass analyzer to an electrospray ion source. Anal Chem 75:1699–1705.

Higgs RE, Knierman MD, Gelfanova V, Butler JP, Hale JE (2005) Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res 4:1442–1450.

Hirabayashi J (2004) Lectin-based structural glycomics: glycoproteomics and glycan profiling. Glycoconj J 21:35–40.

Hsu J-L, Huang S-Y, Chow N-H, Chen S-H (2003) Stable-isotope dimethyl labeling for quantitative proteomics. Anal Chem 75:6843–6852.

Hsu J-L, Huang S-Y, Chen S-H (2006) Dimethyl multiplexed labeling combined with microcolumn separation and MS analysis for time course study in proteomics. Electrophoresis 27: 3652–3660.

Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R (2005) The Orbitrap: a new mass spectrometer. J Mass Spectrom 40:430–443.

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4:1265–1272.

Jaitly N, Monroe ME, Petyuk VA, Clauss TRW, Adkins JN, Smith RD (2006) Robust algorithm for alignment of liquid chromatography mass spectrometry analyses in an accurate mass and time tag data analysis pipeline. Anal Chem 78:7397–7409.

James P, Quadroni M, Carafoli E, Gonnet G (1993) Protein identification by mass profile fingerprinting. Biochem Biophys Res Commun 195:58–64.

Jensen ON, Podtelejnikov AV, Mann M (1997) Identification of the components of simple protein mixtures by high-accuracy peptide mass mapping and database searching. Anal Chem 69:4741–4750.

Ji J, Chakraborty A, Geng M, Zhang X, Amini A, Bina M, Regnier F (2000) Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. J Chromatogr B Biomed Sci Appl 745:197–210.

Ji C, Guo N, Li L (2005) Differential dimethyl labeling of N-termini of peptides after guanidination for proteome analysis. J Proteome Res 4:2099–2108.

Johnson KL, Muddiman DC (2004) A method for calculating 16o/18o peptide ion ratios for the relative quantification of proteomes. J Am Soc Mass Spectrom 15:437–445.

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

Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74:5383–5392.

Khalsa-Moyers G, McDonald WH (2006) Developments in mass spectrometry for the analysis of complex protein mixtures. Brief Funct Genomic Proteomic 5:98–111.

Kirkpatrick DS, Gerber SA, Gygi SP (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. Methods 35: 265–273.

Kito K, Ota K, Fujita T, Ito T (2007) A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. J Proteome Res 6:792–800.

Krijgsveld J, Ketting RF, Mahmoudi T, Johansen J, Artal-Sanz M, Verrijzer CP, Plasterk RHA, Heck AJR (2003) Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. Nat Biotechnol 21:927–931.

Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, Aebersold R (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics 7:655–667.

Li J, Steen H, Gygi SP (2003) Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. Mol Cell Proteomics 2:1198–1204.

Li X-j, Yi EC, Kemp CJ, Zhang H, Aebersold R (2005) A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography–mass spectrometry. Mol Cell Proteomics 4:1328–1340.

Liu H, Sadygov RG, Yates JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76:4193–4201.

Lee YH, Han H, Chang SB, Lee SW (2004) Isotope-coded N-terminal sulfonation of peptides allows quantitative proteomic analysis with increased de novo peptide sequencing capability. Rapid Commun Mass Spectrom 18:3019–3027.

Louris JN, Cooks RG, Syka JEP, Kelley PE, Stafford GC, Todd JFJ (1987) Instrumentation, applications, and energy deposition in quadrupole ion-trap tandem mass spectrometry. Anal Chem 59:1677–1685.

Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 25:117–124.

Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A, Lajoie G (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17:2337–2342.

Makarov A (2000) Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem 72:1156–1162.

Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, Kuster B, Aebersold R (2007) Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol 25:125–131.

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

March RE (1997) An introduction to quadrupole ion trap mass spectrometry. J Mass Spectrom 32:351–369.

Marko-Varga G, Lindberg H, Lofdahl CG, Jonsson P, Hansson L, Dahlback M, Lindquist E, Johansson L, Foster M, Fehniger TE (2005) Discovery of biomarker candidates within disease by protein profiling: principles and concepts. J Proteome Res 4:1200–1212.

Mason DE, Liebler DC (2003) Quantitative analysis of modified proteins by LC–MS/MS of peptides labeled with phenyl isocyanate. J Proteome Res 2:265–272.

McLafferty FW, Horn DM, Breuker K, Ge Y, Lewis MA, Cerda B, Zubarev RA, Carpenter BK (2001) Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. J Am Soc Mass Spectrom 12:245–249.

Nesvizhskii AI, Vitek O, Aebersold R (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods 4:787–797.

Oda Y, Owa T, Sato T, Boucher B, Daniels S, Yamanaka H, Shinohara Y, Yokoi A, Kuromitsu J, Nagasu T (2003) Quantitative chemical proteomics for identifying candidate drug targets. Anal Chem 75:2159–2165.

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4:1487–1502.

Olsen JV, Andersen JR, Nielsen PA, Nielsen ML, Figeys D, Mann M, Wisniewski JR (2004) HysTag – a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. Mol Cell Proteomics 3:82–92.

Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127:635–648.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1:376–386.

Ong SE, Foster LJ, Mann M (2003) Mass spectrometric-based approaches in quantitative proteomics. Methods 29:124–130.

Ono M, Shitashige M, Honda K, Isobe T, Kuwabara H, Matsuzuki H, Hirohashi S, Yamada T (2006) Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. Mol Cell Proteomics 5:1338–1347.

Pappin DJ, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3:327–332.

Parish R (1989) Comparison of linear regression methods when both variables contain error: relation to clinical studies. Ann Pharmacother 23:891–898.

Park K-S, Mohapatra DP, Misonou H, Trimmer JS (2006) Graded regulation of the Kv2.1 potassium channel by variable phosphorylation. Science 313:976–979.

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567.

Porath J (1992) Immobilized metal ion affinity chromatography. Protein Expr Purif 3:263–281.

Qian W-J, Goshe MB, Camp DG, Yu L-R, Tang K, Smith RD (2003) Phosphoprotein isotope-coded solid-phase tag approach for enrichment and quantitative analysis of phosphopeptides from complex mixtures. Anal Chem 75:5441–5450.

Ramos-Fernandez A, Lopez-Ferrer D, Vazquez J (2007) Improved method for differential expression proteomics using trypsin-catalyzed 18O labeling with a correction for labeling efficiency. Mol Cell Proteomics 6:1274–1286.

Ranish JA, Hahn S, Lu Y, Yi EC, Li XJ, Eng J, Aebersold R (2004) Identification of TFB5, a new component of general transcription and DNA repair factor IIH. Nat Genet 36:707–713.

Rao KCS, Carruth RT, Miyagi M (2005) Proteolytic 18O labeling by peptidyl-Lys metalloendopeptidase for comparative proteomics. J Proteome Res 4:507–514.

Reynolds JA, Tanford C (1970) The gross conformation of protein-sodium dodecyl sulfate complexes. J Biol Chem 245:5161–5165.

Reynolds KJ, Yao X, Fenselau C (2002) Proteolytic 18O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. J Proteome Res 1:27–33.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3:1154–1169.

Saito A, Nagasaki M, Oyama M, Kozuka-Hata H, Semba K, Sugano S, Yamamoto T, Miyano S (2007) AYUMS: an algorithm for completely automatic quantitation based on LC-MS/MS proteome data and its application to the analysis of signal transduction. BMC Bioinform 8:15.

Salomon AR, Ficarro SB, Brill LM, Brinker A, Phung QT, Ericson C, Sauer K, Brock A, Horn DM, Schultz PG, Peters EC (2003) Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. Proc Natl Acad Sci U S A 100:443–448.

Schmidt A, Kellermann J, Lottspeich F (2005) A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics 5:4–15.

Schnölzer M, Jedrzejewski P, Lehmann WD (1996) Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. Electrophoresis 17:945–953.

Schuchardt S, Sickmann A (2007) Protein identification using mass spectrometry: a method overview. EXS 97:141–170.

Senko MW, Hendrickson CL, Emmett MR, Shi SD, Marshall AG (1997) External accumulation of ions for enhanced electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. J Am Soc Mass Spectrom 8:970–976.

Shukla AK, Futrell JH (2000) Tandem mass spectrometry: dissociation of ions by collisional activation. J Mass Spectrom 35:1069–1090.

Sleno L, Volmer DA (2004) Ion activation methods for tandem mass spectrometry. J Mass Spectrom 39:1091–1112.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 100:9440–9445.

Strittmatter EF, Ferguson PL, Tang K, Smith RD (2003) Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. J Am Soc Mass Spectrom 14:980–991.

Strittmatter EF, Kangas LJ, Petritis K, Mottaz HM, Anderson GA, Shen Y, Jacobs JM, Camp DG 2nd, Smith RD (2004) Application of peptide LC retention time information in a discriminant function for peptide identification by tandem mass spectrometry. J Proteome Res 3:760–769.

Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004a) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101:9528–9533.

Syka JEP, Marto JA, Bai DL, Horning S, Senko MW, Schwartz JC, Ueberheide B, Garcia B, Busby S, Muratore T, Shabanowitz J, Hunt DF (2004b) Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. J Proteome Res 3:621–626.

Tang H, Arnold RJ, Alves P, Xun Z, Clemmer DE, Novotny MV, Reilly JP, Radivojac P (2006) A computational approach toward label-free protein quantification using predicted peptide detectability. Bioinformatics 22:e481–e488.

Tao WA, Wollscheid B, O'Brien R, Eng JK, Li X-j, Bodenmiller B, Watts JD, Hood L, Aebersold R (2005) Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. Nat Methods 2:591–598.

Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75:1895–1904.

Wada Y, Tajiri M, Yoshida S (2004) Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. Anal Chem 76:6560–6565.

Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH (2003) Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem 75:4818–4826.

Wang P, Tang H, Fitzgibbon MP, Mcintosh M, Coram M, Zhang H, Yi E, Aebersold R (2007) A statistical method for chromatographic alignment of LC-MS data. Biostatistics 8:357–367.

Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proc Natl Acad Sci U S A 104:5860–5865.

Wright ME, Eng J, Sherman J, Hockenbery DM, Nelson PS, Galitski T, Aebersold R (2003) Identification of androgen-coregulated protein networks from the microsomes of human prostate cancer cells. Genome Biol 5:R4.

Wu CC, MacCoss MJ, Howell KE, Matthews DE, Yates JR (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. Anal Chem 76:4951–4959.

Yang Z, Hancock WS (2004) Approach to the comprehensive analysis of glycoproteins isolated from human serum using a multi-lectin affinity column. J Chromatogr A 1053:79–88.

Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C (2001) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. Anal Chem 73:2836–2842.

Yates JR 3rd, Morgan SF, Gatlin CL, Griffin PR, Eng JK (1998) Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. Anal Chem 70:3557–3565.

Zhang R, Regnier FE (2002) Minimizing resolution of isotopically coded peptides in comparative proteomics. J Proteome Res 1:139–147.

Zhang R, Sioma CS, Wang S, Regnier FE (2001) Fractionation of isotopically labeled peptides in quantitative proteomics. Anal Chem 73:5142–5149.

Zhang X, Jin QK, Carr SA, Annan RS (2002) N-terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. Rapid Commun Mass Spectrom 16:2325–2332.

Zhang H, Li X-j, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol 21:660–666.

Zhang H, Yi EC, Li XJ, Mallick P, Kelly-Spratt KS, Masselon CD, Camp DG 2nd, Smith RD, Kemp CJ, Aebersold R (2005a) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. Mol Cell Proteomics 4:144–155.

Zhang J, Schubothe K, Li B, Russell S, Lebrilla CB (2005b) Infrared multiphoton dissociation of O-linked mucin-type oligosaccharides. Anal Chem 77:208–214.

Zubarev RA (2004) Electron-capture dissociation tandem mass spectrometry. Curr Opin Biotechnol 15:12–16.

Zubarev RA, Kelleher NL, McLafferty FW (1998) Electron capture dissociation of multiply charged protein cations. A non-ergodic process. J Am Chem Soc 120:3265–3266.

Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. Anal Chem 72:563–573.