
1

OVERVIEW OF PROTEOMICS

1-1 WHAT IS PROTEOMICS?

The term *proteomics* is derived from a word “proteome” coined by M. Wilkins in 1995, which indicates “the entire protein complement expressed by a genome, or by a cell or tissue type” (1). The proteome is the time- and cell-specific protein complement of the genome, encompassing all proteins expressed in a cell at any given time. The study of the proteome, compared to the genome, is much more daunting for several reasons; while the genome of the cell is constant, nearly identical for all cells of an organ or organism, and consistent across a species, the proteome is extremely complex and dynamic as it continuously responds to environmental changes due to the presence of other cells, nutritional status, temperature, and drug treatment, to name only a few (2). Proteomics, which is a new field of interdisciplinary science derived from gene sequencing and classical protein chemistry, deals with the proteome; an initial goal of proteomics is the rapid identification of all the proteins expressed by a cell or tissue, although it has yet to be achieved for any species. The task of identifying and quantifying large scale proteins present in a cell or tissue or even an entire organism at a particular time is often referred to as *proteome analysis*. Because the proteome varies from time to time, any analysis of the proteome is a “snapshot.” As a result, there is no fixed proteome. Moreover, the dynamic range of protein expression within the proteome complicates study of a proteome; it may vary by as much as 7–12 orders of magnitude compared to only five orders of magnitude for DNA (3). The number of genes encoded in the genome of a species is limited

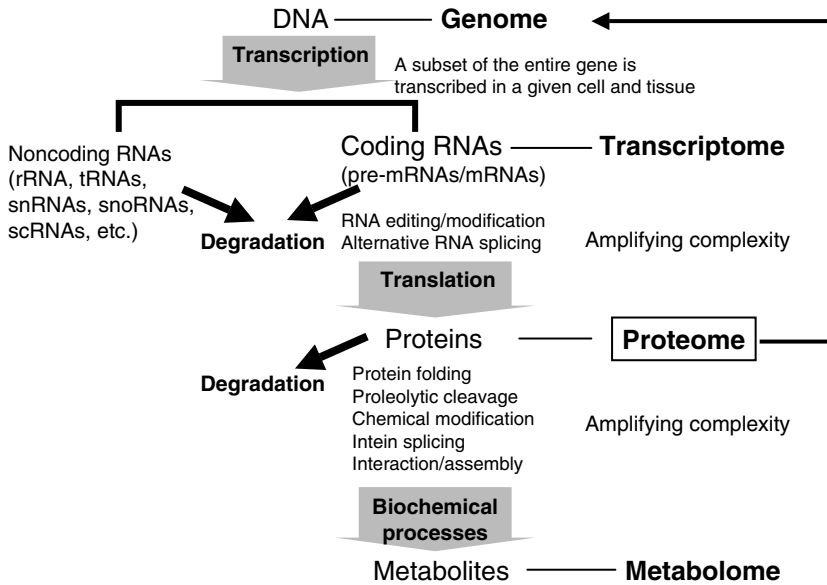


Fig. 1-1. Complexities of a proteome regulated at transcriptional and translational levels.

and ranges from a few hundred for bacteria to tens of thousands for mammalian species; however, the theoretical number of proteins defined by a genome alone is large enough that it is a struggle to entirely identify them by proteome analysis using technologies currently available. In addition, proteins are composed of 20 distinctive amino acids, each of which has its own chemical and physical characteristics including molecular weight (M_r) and isoelectric point (pI), and differ in their amino acid sequence and length; thus, they are extremely heterogeneous in chemical and physical characteristics. To make matters worse, the number of protein species produced from a genome is amplified, as the same gene can generate multiple protein products that differ as a result of RNA editing and alternative splicing of pre-mRNA, and of processing and chemical modifications of translated polypeptides (Fig. 1-1). The number of proteins, for instance, expressed by the approximately 24,000 human genes can be up to 100 times greater due to the known diversity of mRNA processing as well as to post-translational modifications of proteins (2). The main difficulty in performing proteome analysis lies in these diverse properties of proteins generated at post-transcriptional and post-translational levels in addition to the large number of proteins encoded by genes assigned in a genome. Thus, it is a formidable challenge to develop proteome analysis technologies that can handle simultaneously a huge number of proteins with extremely heterogeneous characteristics, and to identify in a comprehensive and flexible manner such extremely diversified proteins.

Proteomics should be distinguished from conventional protein chemistry, which deals with mostly “individual proteins” to determine sequence, modification state, interaction partner, activity, structure, and so on. Protein chemistry was the

mainstream approach of biological science in the 1970s but decreased in use during the flourishing era of genetic engineering in the 1980s. Proteomics, however, incorporated a number of basic ideas and technologies from conventional protein chemistry, and advanced those with new ideas incorporated into postgenomic science and emerging technologies to handle proteins on a large scale with small quantities and in a systematic way. In other words, protein chemistry has evolved and has been reincarnated as proteomics backed by technological advancements mainly in mass spectrometry and the accumulation of huge protein sequence data as a consequence of the global genomic sequencings of many species, which have catalyzed an expansion of the scope of biological studies from reductionist biochemical analysis of single proteins to proteome-wide measurements (4). Because proteomics deals with huge numbers of proteins in extremely complex mixtures of cells, tissues, and even organisms, the technologies used in proteomics have to achieve comprehensive and high throughput identification of proteins within an appropriate time.

1-2 ESTABLISHMENT OF PROTEOMICS USING MASS SPECTROMETRY

1-2-1 Protein Separations by Two-Dimensional Electrophoresis

The first step of proteome analysis is to separate proteins using high resolution and high speed techniques. Conventional protein chemistry had developed methodologies that separate proteins with high resolution, such as two-dimensional electrophoresis (2DE), and high performance, such as high performance liquid chromatography (HPLC). Two-dimensional electrophoresis, which is performed principally based on the method described in 1975 by O'Farrell, involves, in the first dimension, a separation by isoelectric focusing in gel rods containing urea and detergents, and, in the second dimension, a perpendicular separation in acrylamide (gradient) slab gels containing sodium dodecyl sulfate (SDS) (5). Because denaturing reagents are used in both dimensions, the physicochemical characteristics of polypeptides, whose secondary and tertiary structures are broken down, distinguish between the proteins; namely, proteins are separated first by isoelectric point (pI) (based on electric charge of polypeptides), and second by polypeptide length (based on molecular weight of polypeptides). Those two properties are not correlated at all; thus, proteins are spread effectively over the resulting two-dimensional pattern. The use of denaturing conditions in both dimensions is required in order to obtain high resolution and to distinguish a single charge difference, especially between two small proteins (but not always between large proteins), which may be generated by a single amino acid substitution or post-translational modification. [The single charge difference produces a much greater shift in the pI of small a protein (or subunit) than a large one (or oligomer).] In addition, 2DE can also detect some differences in molecular weight, which may be produced by proteolytic cleavage and/or post-translational modification. [The molecular weight differences are more easily detected by examination of individual polypeptides (of small proteins) rather than assembled oligomers (of large

TABLE 1-1. Typical Description of 2DE and 2D-DIGE

<i>Two-Dimensional Electrophoresis</i>
Apparatus: First dimension IPGphor/Multiphor II/HoeferDALT (Amersham Biosciences) Dry IPG gels (Immobilin, APB; Bio-Rad) Second dimension Ettan Dalt (Amersham Biosciences)
Staining method: Coomassie brilliant blue, silver ion, SYPRO ruby (Molecular Probes)
Image scanner: GS710 Imaging densitometry, Molecular Imager FX, FluorS multiimager (Bio-Rad)
Image analysis software: Melane3/PDQuest (Bio-Rad), ImageMaster 2D Elite 3.0/ Database 3.0 (Amersham Biosciences)
 <i>2D-DIGE, Two-Dimensional Differential Image Gel Electrophoresis</i>
Prelabeling of proteins: CyDye NHS (Cy3 550 nm, Cy5 649 nm)
Minimal dye labels at lysine residue (Cy2 = MW 550.59; Cy3 = MW 582.76; Cy5 = MW 580.74)
Saturation dyes at cysteine residue (Cy3 = MW 672.85; Cy5 = MW 684.86)
Fluorescence image scanner: 2D ImageMaster, Typhoon (Amersham Bioscience) FLA5000 (Fuji Film Co.)
Image analysis software: DeCyder, PDQuest (Bio-Rad)

proteins) under denatured conditions.] Thus, 2DE is a suitable method to distinguish different proteins in a chemical structure, to provide an inventory of the polypeptides in a mixture, and to create a protein catalog (5).

Although resolution of 2DE separation is dependent on the dimension of the slab gel and the pH gradient used for isoelectric focusing in the first gel rod, a typical 2DE separates 1000–3000 protein species, which can be detected by protein staining with Coomassie blue, silver ion, and fluorescent dyes, such as SYPRO Ruby (6, 7) (Table 1-1). Larger dimension slab gels achieve much better separation and can separate over 10,000 protein species on a single gel in some cases, although a large amount of sample is required. 2DE has the highest resolution of the available technologies used to separate proteins in biological samples to date. Seemingly, a staining pattern on a 2DE gel gives a whole view of a proteome (or more correctly a subset view of a proteome), but does not necessarily represent the entire set of proteins expressed in an analyzed biological sample. A comparison between the two different 2DE staining patterns obtained from two or more (related or unrelated) biological samples enables us to estimate a change or difference of the proteomes in the different biological samples, by comparing independent 2DE gels visualized either by conventional staining methods or by 2D-DIGE (two-dimensional differential image gel electrophoresis). Coomassie blue dye has traditionally been used to stain polyacrylamide gels but does not provide the sensitivity needed for visualizing low abundance proteins. Silver staining is more sensitive but is not optimal for quantitation purposes because of its narrow dynamic range (8). Consequently, fluorescent dyes that are very sensitive and have a large linear dynamic range have

been developed (9), although the cost of the dyes and the equipment used for visualization in addition to the variability between 2D gels have restricted their use (10). In 2D-DIGE, two or more pools of proteins are labeled with different fluorescent dyes (11, 12), and the labeled proteins are mixed and separated in the same 2DE gels; thus 2D-DIGE enables one to compare the differences in protein expression between the two samples or among the several samples on a single 2DE gel. Development of the immobilized pH gradient as the first-dimensional separation medium of 2DE (in which the pH gradient was fixed within the acrylamide matrix) popularized 2DE as a protein separation method for proteome analysis. In addition, the principles used for the global, quantitative analysis of gene expressions, such as the use of clustering algorithms and multivariate statistics that had been developed in the context of 2DE (13, 14), produced various image analyzing apparatus and software for quantitative comparison of the 2DE gels and further popularized 2DE as a methodology for proteome analysis (Table 1-1).

1-2-2 Development of the Technologies for Protein Identification

The second step of proteome analysis is to identify proteins separated by technologies with high resolution power such as 2DE described previously. One of the archetypal proteome analyses was that applied to human plasma, where proteins were separated by 2DE and identified by a combination of various methods, including Western blotting, coelectrophoresis of purified proteins, and Edman sequencing of a protein extracted from gels (5). Although human plasma had been profiled extensively by those analyses in the early 1980s, a large number of antibodies and purified proteins were required for Western blotting and coelectrophoresis, respectively. However, it was obvious that those approaches were not suitable for the large scale identification of proteins.

Conventional protein chemistry had made great efforts in technological development for identification of proteins separated by gels in the 1980s and early 1990s (4). Edman sequencing technology, which had been the most common technique to determine the amino acid sequence of a purified protein or enzymatically digested peptides, was applied to proteins either extracted in solution or blotted onto membrane by electroelution from gels and facilitated protein identification on gels. The gas phase protein sequencer (based on Edman sequencing technology) improved the sensitivity of protein sequencing. However, because the sequence database contained only a part of the total protein list, the chance that a particular protein and/or gene sequence was already represented in a sequence database was quite low. Therefore, during the 1980s and early 1990s, the main purpose of sequencing proteins separated on gels was either to synthesize degenerate oligonucleotide primer for cloning of its corresponding gene or to link the activity of a purified protein and amino acid sequence (4).

Genomic sequence analyses began to flourish and provided entire genomic sequences of bacteria in the mid-1990s, and subsequently established sequences for a number of other species including human beings by the early 2000s. The data of genomic sequences facilitated protein identification with even a partial sequence of proteins separated on gels and allowed one to specify the complete sequence of proteins (genes) without the need for further experimentation. Thus, the first stage

of proteomic analysis used Edman sequencing technology to identify proteins separated by 2DE. The N-terminal sequence or partial peptide sequence could be used for retrieval of the analyzed protein from the protein/gene sequence database; however, it was not sensitive enough to determine the amino acid sequence of many proteins obtained from 2DE gels, and was not applicable to proteins with amino terminus (N terminus) blocked by chemical modifiers, such as acetylate or pyroglutamate. Consequently, a number of proteins on 2DE gels remained unidentified by adopting Edman sequencing technology. In addition, Edman sequencing was a time-consuming technology; thus, it was not suitable for systematic and rapid identification of proteins. With the normalized sequence database established as a result of the completion of a number of genomic sequencing analyses, a rapid methodology for protein identification was eagerly awaited.

1-2-3 Protein Identification Based on Gel Separation and Mass Spectrometry

Mass spectrometry (MS) measured precise molecular weight, distinguished closely related molecular species, and had been a powerful method for physicochemical analysis of small molecules long before protein chemists began to apply MS to the determination of protein sequence in the late 1970s. In principle, a mass spectrometer is comprised of an ionization source (which adds charge onto a neutral analyte or forms an ion) and a mass analyzer (which measures the a mass-to-charge ratio, m/z , of a molecule). Thus, the analyte for mass spectrometry has to be ionized before being transferred into the high vacuum chamber of the mass analyzer (4). At the earliest stage, protein chemists used fast atom bombardment (FAB) as the ionization source and a magnetic sector as the mass analyzer (Fig. 1-2), but

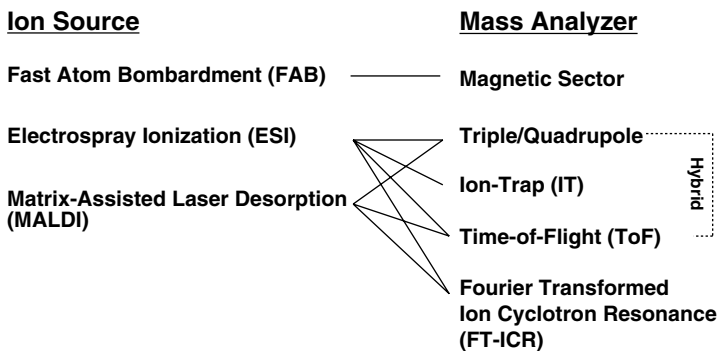


Fig. 1-2. Mass spectrometers used for proteomics research. Mass spectrometers are composed of two major compartments with different functions: an ion source that generates ions of target molecules and a mass analyzer that estimates the mass-to-charge ratio (m/z) of target ions. In mass spectrometers for proteomics research that analyze mostly proteins and peptides, an ESI or MALDI ion source is most frequently coupled with a triple-stage quadrupole (TSQ), ion-trap (IT), time-of-flight (ToF), or Fourier transformed ion cyclotron resonance (FT-ICR) mass analyzer. In some mass spectrometers, multiple mass analyzers are combined in a single instrument to perform “tandem mass spectrometry” for structural analysis of target molecules.

had only limited success for very small peptides. Proteins and even peptides were very large molecules in standard mass spectrometry and it was difficult to form ions in a vacuum or at atmospheric pressure without destroying the molecules. Therefore, protein sequence determination with a mass spectrometer was hampered by this inability for a considerable time until new methods for the ionization of large molecules were developed. In the late 1980s, two methods—matrix-assisted laser desorption ionization (MALDI) (15) and electrospray ionization (ESI) (16)—were developed that allowed the ionization of peptides and proteins at high efficiency and without excessive destruction of the molecule.

MALDI is a process that allows desorption of proteins and peptides, achieved by short laser pulses fired at those proteins and peptides that are cocrystallized with a matrix on a sample plate, which is held under vacuum (Fig. 1-3A). The matrices are chemical compounds that absorb the energy of the laser and promote

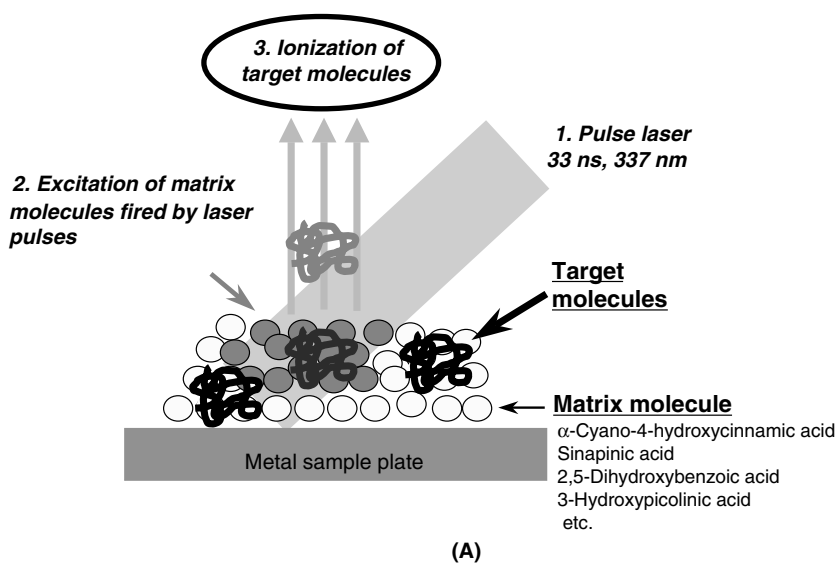


Fig. 1-3. Schematic illustration of (A) matrix-assisted laser desorption ionization (MALDI) and (B) electrospray ionization (ESI) of biological molecules for mass spectrometric analysis. MALDI is performed by embedding target molecules in an excess of a specific wavelength-absorbing matrix, such as α -cyano-4-hydroxycinnamic acid, which is dried to produce a cocrystallized mixture on a metal sample plate. Ions are produced by bombarding the sample with short-duration pulses of ultraviolet (UV) light from a nitrogen laser. The interaction of the laser pulse with the sample results in ionization, usually protonation, of both matrix and target molecule via an energy transfer mechanism from the matrix to the embedded target, rather than by direct laser ionization. ESI creates gas-phase ions by applying a potential to a flowing liquid that contains the target and solvent molecules. A fine spray of microdroplets is generated upon application of a high electric tension (typically ± 3 – 5 kV) through a needle chip. Solvent is removed as the droplets enter the mass spectrometer by heat or some other form of energy, such as energetic collisions with an inert gas.

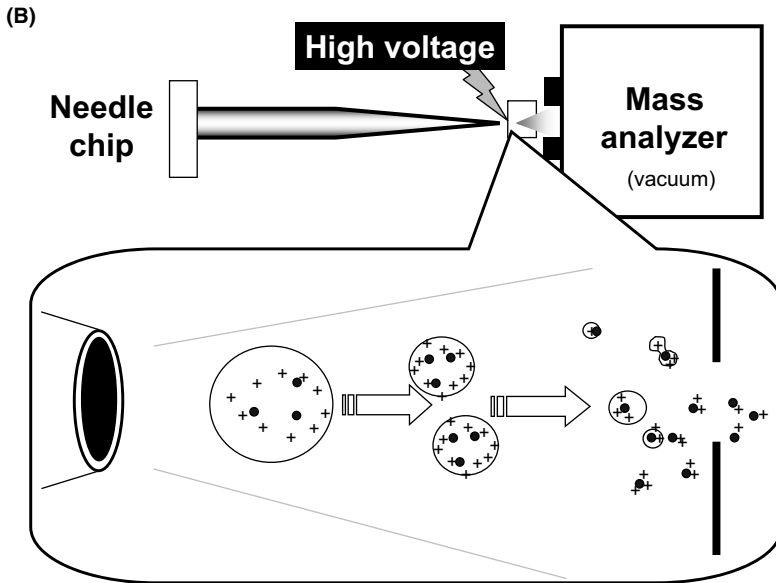


Fig. 1-3. (Continued)

ionization of the proteins and peptides. On the other hand, the ESI process is achieved by spraying a solution through a charged needle tip at atmospheric pressure toward the inlet of the mass analyzer (Fig. 1-3B). The voltage applied to the needle tip results in the formation of ions and the pressure difference results in their transfer into the mass analyzer. These ionization methods made it possible to determine molecular weights of peptides and proteins in high precision in combination with mass analyzers (Fig. 1-2).

Although MALDI ion sources are coupled with any of the time-of-flight (ToF), ToF-ToF, quadrupole, and Fourier transformed ion cyclotron resonance (FT-ICR) mass analyzers (Fig. 1-2), MALDI-ToF mass spectrometers (MSs) are often used for obtaining single-stage mass spectrometry (MS) spectra that provide mass information on all ionizable components in a sample, and determine the masses of proteins or peptides with a high degree of accuracy (4). MALDI-ToFMSs are especially suitable for determining the masses of various peptides generated by fragmentation of an isolated protein with an enzyme of known cleavage specificity, such as trypsin (cleaves at C-terminal side of lysine and arginine) or endopeptidase Lys C (cleaves at C-terminal side of only lysine). The collected list of peptide masses is matched to those masses calculated from the same proteolytic digestion of each entry in a sequence database by database search algorithms (available at [http://www/mann.emble-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html](http://www.mann.emble-heidelberg.de/GroupPages/PageLink/peptidesearchpage.html), <http://prospector.ucsf.edu/ucsfhtml4.0/msfit.html>, <http://www.matrixscience.com>, <http://prowl.rockefeller.edu/>, <http://us.expasy.org/tools/peptident.html>, etc.) that were reported originally by several independent groups in 1993 (17–20). This approach is

called the *peptide mass fingerprinting* (PMF) method (Fig. 1-4) and has greatly facilitated the identification of proteins separated by polyacrylamide gel electrophoresis in conjunction with in-gel protease digestion, in which proteolytic fragmentation of a protein was performed in a gel piece excised from an electrophorized gel (Fig. 1-4B, Protocol 1.1) (21–23). The in-gel protease digestion–PMF method using MALDI-ToF was sensitive enough to identify a protein present even in a single spot on a 2DE gel stained by silver ion or fluorescent dye, which is a highly sensitive method to detect proteins on a 2DE gel. The PMF method is far less time consuming than Edman sequencing; thus, it is now a common approach for identifying proteins separated by 2DE. The approach using MS, in which the protein is first purified and cleaved into peptides, whose relative molecular weight (M_r) values are measured by MS and used for protein identification, is often called the *bottom-up approach* (3, 24, 25); an approach where the protein mixture, without digestion, is introduced directly into the MS instrument, is separated with high resolution, and is dissociated to measure the resulting fragment masses that are matched

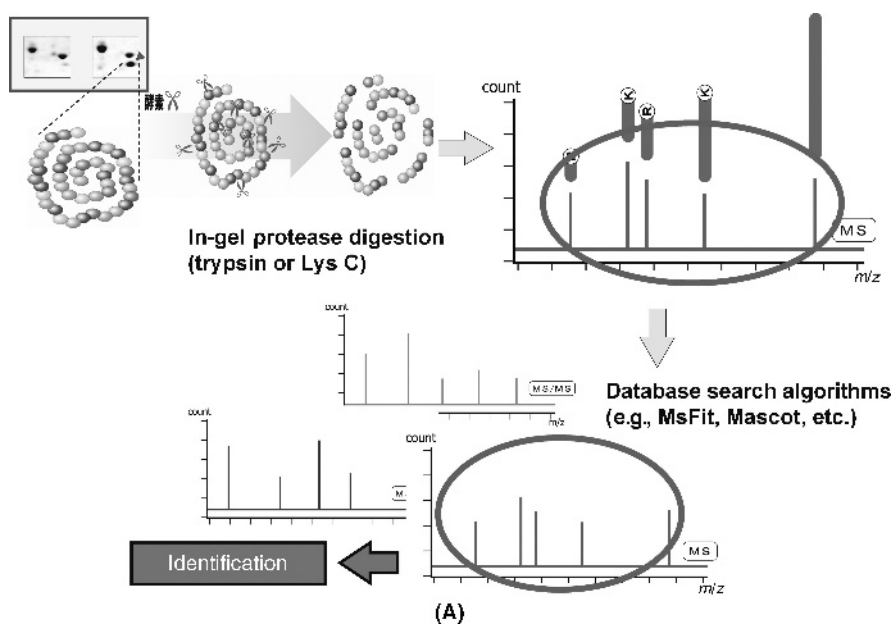


Fig. 1-4. (A) Protein identification by in-gel protease digestion and peptide mass fingerprinting. In this method, a target protein, such as that separated by two-dimensional electrophoresis, is in-gel digested with a sequence-specific protease, usually trypsin, and the resulting peptide mixture is analyzed by mass spectrometry, thus generating a peptide mass fingerprint. Experimentally determined peptide masses are compared with those obtained theoretically by a search algorithm such as MSFit or Mascot. (See insert for color representation.) (B) A typical experimental protocol for in-gel protease digestion of proteins separated by polyacrylamide gel electrophoresis. The resulting peptide mixture is analyzed directly on a mass spectrometer with a MALDI or ESI ion source.

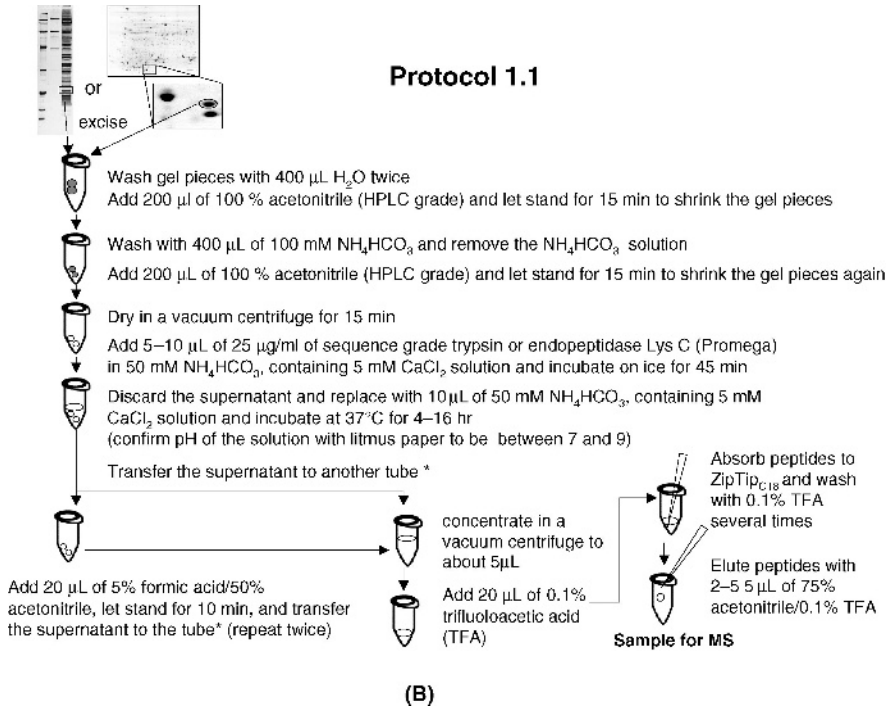


Fig. 1-4. (Continued)

against protein database is called the *top-down approach* (see Section 2-2) (26). The specificity of the protease used for in-gel digestion, the number of peptides identified from each protein species by mass spectrometry, and the mass accuracy of the mass spectrometer are key elements in successful protein identification using the PMF approach; thus, the PMF method is applicable only to the identification of a single protein or a few proteins separated mostly by gel-based techniques (1DE, 2DE). When the PMF approach is not sufficient for identification, peptides can be sequenced by tandem MS (MS/MS) with an ESI ion source and the fragmentation pattern can be used to identify the protein in a database, even on the basis of one or a few peptide sequences.

Although ESI ion sources were originally coupled with a triple-quadrupole or ion trap mass analyzer, they can be used in conjunction with most available mass analyzers including hybrid type MS/MS analyzers such as quadrupole-ToF and ToF-ToF MS/MS analyzers (Fig. 1-2). These MS/MS instruments are equipped with a mass filter that can select a peptide ion from a mixture of peptide ions (or that can isolate specific ions from a mixture on the basis of their m/z ratio), a collision cell in which peptide ions are fragmented in a sequence-dependent manner into a series of product ions through collision of the selected precursor ion with a noble gas (in a process referred to as collision-induced dissociation—CID), and a second mass analyzer that records the fragment ion mass spectrum (Fig. 1-5A) (4). Thus, these

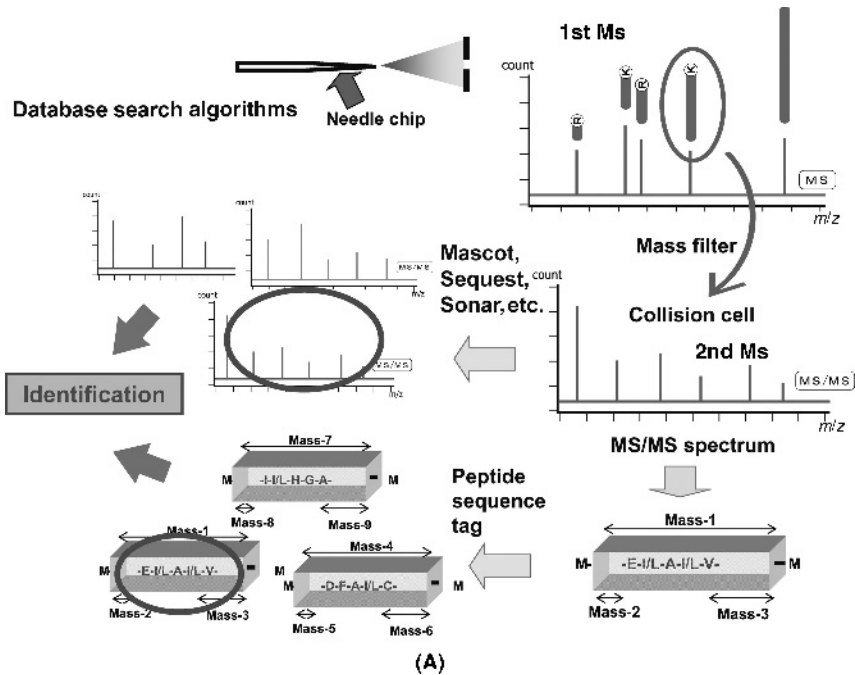


Fig. 1-5. (A) Protein identification by tandem, or MS/MS, mass spectrometry. In this method, individual peptides from a peptide mixture are isolated in the first step in the mass spectrometer and fragmented by collision-induced dissociation (CID), that is, by collision with an inert gas in a collision cell, during the second step in order to obtain the structural information of the peptide. (See pages 12–13 for more detailed information) (See insert for color representation). (B) Major fragment ions produced by tandem mass spectrometry of polypeptide chain. CID mainly induces the cleavage of peptide at peptide bonds and generates a series of fragment ions from the amino or carboxyl terminus of the peptides, which are termed “b” or “y” series ions, respectively. (C) Sequence tag. In tandem mass spectrometry, a single particular protein can be identified from the structural information of a single peptide produced by digestion with a sequence-specific protease, such as trypsin. The identification is based on the precise molecular mass of a particular peptide (mass-1), the internal amino acid sequence, which serves as a specific “tag” of the peptide, and the molecular masses of the remaining amino and carboxyl terminal portion of the peptides (mass-2 and mass-3). The computer algorithm, such as that developed by Mann and Wilm (27), selects a set of potential peptides having mass-1 from a database and then specifies one of those by the sequence tag and mass-2 and mass-3.

information on protein structure and function. In addition, the resulting datasets generally need further evaluation and reorganization, because they often include ambiguous peptide identity data and redundant peptide assignments for each protein; thus, data processing in proteomics studies is both labor intensive and time consuming (32). A number of software packages have been developed to expedite the data processing step. Of those, for example, Autoquest (33), SEQUEST SUMMARY (31), DTAslect (34), and INTERACT (35) are designed to organize and rearrange peptide/protein identification results. These software packages permit rapid data processing after identification of proteins by SEQUEST. STEM (STrategic Extractor for Mascot's results, available at <http://www.sci.metro-u.ac.jp/proteomicslab/>), on the other hand, is designed to process Mascot search data and is a stand-alone computational tool that evaluates, integrates, and compares large datasets produced by Mascot (36). Web-based software DBParser is also developed for the analysis of large scale data obtained from LC-MS/MS (37). The results obtained by STEM and DBParser actively link to the primary mass spectral data and to public online databases such as NCBI, GO, and Swiss-Prot in order to structure contextually specific reports for biologists and biochemists.

With the rapid advances in protein analytical technologies developed in the early 1990s, a huge number of protein constituents in a biological sample such as a cell or tissue extract separated by 2DE could be identified with much higher sensitivity and speed than with the technologies available before the mass-based technologies had emerged. Expansion of protein database fueled by the large scale genomic sequencing of many species made a big push to identify proteins on a large scale by those approaches. As described in Section 1-2-1, changes in a subset of the proteome could be seen on 2DE gels as different spots after the separation of proteins present in cells or tissues, similar to the appearance or disappearance of new mRNA in a differential display analysis or using DNA arrays. The PMF and/or MS/MS ion search method was immediately applied to the large scale identification of those changed proteins on 2DE gels as well as to the identification of the unchanged proteins. Because the methodologies based on 2DE and mass spectrometry were so powerful in identifying proteins in terms of sensitivity and speed when compared with methods used in conventional protein chemistry, this kind of approach dominated the first generation of proteomic studies. In fact, the proteome study involved mostly quantitative comparison and mass spectrometry-based identification of proteins separated by 2DE in the mid- to late-1990s, and resulted in a number of online 2DE gel-based protein profiling and catalog databases for various organisms, tissues, and cells, including plasma and urine, which had been used for clinical diagnosis (38–42) (<http://kr.expasy.org/ch2d/>, <http://kr.expasy.org/ch2d/2d-index.html>). Intense interest grew in applying the approach to develop new biomarkers for diagnosis and early detection of disease; this led to the identification of a number of disease-related changes in protein expression including those associated with heart disease and various cancers (43–51).

Despite the great advancement in large scale protein identification for proteomic analysis, the 2DE mass spectrometry-based identification method was realized as a rather low throughput approach in that it requires a relatively large amount of protein sample, even though all the improvements in highly sensitive detection methods,

automated spot cutting and in-gel digestion, automated mass spectrometry analysis, and so on had been made. The requirement of a large amount of protein sample was particularly problematic for clinical samples since such samples are generally procured in limited amount (51). In addition, clinical samples contain heterogeneous cells (or are mixtures of various types of cells) and thus are extremely complex in terms of protein constituents, which makes proteomic analysis much more difficult than analysis methods used in basic biology such as bacteria and cultured cells. To reduce heterogeneity of clinical samples, various tissue microdissection approaches could be used. For example, laser-capture microdissection allows one to isolate defined cell types from tissues, provides very useful samples for comparative analysis between cells of normal and disease-affected areas, and reduces tissue heterogeneity; however, it yields a small amount of proteins, making it difficult to meet the need for greater amounts for 2DE (47). In addition, 2DE has technical limitations on protein separation: that is, 2DE can resolve proteins mostly in the *pI* range of 3 to 11 and in the molecular weight (MW) range of several to 150 kDa. The other limitations facing proteomics analysis using 2DE include the inability to meet the great dynamic range of protein abundance [that spans an estimated range of five to six orders of magnitude for yeast cells (52) and more than ten orders of magnitude for human serum (e.g., from interleukin-6 at ~2 pg/mL to albumin at 50 mg/mL)] (53) and the extent of hydrophobicity and post-translational modifications, even though they are not restricted to the 2DE-based approach. Despite the fact that 2DE is the highest resolving protein separation method known, it was also realized to not always be complete; the incidence of comigrating proteins is rather high (54). Because quantification in 2DE relies on the assumption that one protein is present in each spot, comigration jeopardizes comparative quantification analysis with 2DE. Furthermore, when unfractionated cell lysates were separated by 2DE, only a subset of a cellular proteome was visualized with available protein staining methods (4). Allowing that there are a number of limitations, however, it is certain that the 2DE mass spectrometry-based methodology has revolutionized the protein identification strategy and changed ways of thinking in the fields of protein chemistry and biology.

As an estimate of all the constituents in a proteome of a single cell, a global analysis of protein expression has been done by using the yeast cell fusin library, where each open reading frame is tagged with a high affinity epitope and expressed from its natural chromosomal location (http://www.yeastgenome.org/chromosome-updates/start_changes.shtml) (55). A census of proteins expressed during log-phase growth and measurements of their absolute levels through immunodetection of the common tag suggested that about 80% of the proteome is expressed during normal growth conditions, and that the abundance of proteins ranges from fewer than 50 to more than 10^6 molecules per cell (<http://yeastgfp.ucsf.edu/>). This experiment augments efforts to view the proteome using MS-based protein identification technologies and provides a comprehensive and sensitive view of the expressed proteome in a eukaryotic cell (56). This estimate can also be used to validate the capability of the developed proteomics techniques, including MS-based protein identification methods.

1-3 STRATEGIES FOR CHARACTERIZING PROTEOMES AND UNDERSTANDING PROTEOME FUNCTION

One of the main purposes of proteomic biology is to identify particular members of proteomes that participate in specific biological processes, to assign a function to each, and to devise strategies for their selective modulation (57). However, it was realized that just a description of protein components and the abundance of a proteome did not meet the requirements for reaching that goal. Backed by the development of powerful proteomics technologies typically using gel-based mass spectrometry, new strategies for characterizing functional aspects of proteomes began to take shape. One pioneering work adopting such new strategy is the systematic identification of *in vivo* substrates of the chaperonin GroEL by using 2DE and mass spectrometry (58). GroEL has an essential role in mediating protein folding in the cytosol of *Escherichia coli* and is involved in the folding of ~10% of newly translated polypeptides *in vivo*, while it interacts *in vitro* with almost any nonnative model proteins. Thus, GroEL has an obvious preference for a subset of *E. coli* proteins *in vivo*; however, only a few proteins constituting the subset were known. In a proteomic approach to elucidate the *in vivo* substrate of GroEL, the newly synthesized substrates of GroEL were isolated by large scale immunoprecipitation with anti-GroEL antibodies in the presence of EDTA, which prevents the ATP-dependent release of protein substrates from GroEL. The study also analyzed the structurally unstable substrates of GroEL that were generated by heat shock and were also prepared by the same immunoprecipitation method. The PMF method using MALDI-ToF was applied to the analysis of the immunoprecipitated substrates after the separation by 2DE (first dimension, 13 cm Immobiline DryStrip pI 4–7L; second dimension, 11 cm SDS-PAGE gel MW 5–110 kDa) and the staining with Coomassie blue (58, 59). The analysis revealed that GroEL interacts strongly with a well-defined set of about 300 newly translated polypeptides out of 2500 polypeptides in the cytosol of *E. coli*, of which about one-third are structurally unstable and return to GroEL for conformational maintenance. Interestingly, GroEL substrates consisted preferentially of two or more domains with $\alpha\beta$ folds, which contain α helices and buried β sheets with extensive hydrophobic surfaces. Because those proteins were expected to fold slowly and be prone to aggregate, the hydrophobic binding regions of GroEL are well adapted to interact with the nonnative states of $\alpha\beta$ -domain proteins. Thus, the systematic analysis of GroEL substrate using 2DE gel mass spectrometry and database comparisons successfully highlighted the key structural features that determine the interacting proteins' need for chaperonins during protein folding *in vivo* (58).

This work provides examples of unique strategies for identifying proteins associated with a particular biological activity, thereby taking a step toward functional identification of a proteome. Thus, two main types of strategies in proteomics (that have complementary objectives to each other) have been formed so far: one is the global characterization of protein expression that is referred to as *expression proteomics*, *descriptive proteomics*, or *cataloging proteomics* (61, 62), and another is the characterization of proteome function that is referred to as *functional*

TABLE 1-2. Strategies for Quantitative Proteomics

Expression proteomics (descriptive proteomics)
Disease (Clinical) proteomics
Functional proteomics (focused proteomics)
Modification-specific proteomics
Activity-based protein profiling/enzyme substrate proteomics
Subcellular (organelle) proteomics
Machinery or complex (interaction) proteomics
Dynamic proteomics

proteomics or *focused proteomics* (Table 1-2) (3, 63). As described in the Section 1-2-3, large scale efforts to measure protein expression have typically relied on 2DE mass spectrometry-based methods and LC-MS/MS-based methods described in later sections, which are capable of simultaneously evaluating the relative abundance and permitting the identification of new proteins associated with discrete physiological and/or pathological states. By focusing on measurements of protein abundance, however, expression proteomics (descriptive or cataloging proteomics) provides only an indirect assessment of protein function and may fail to detect important post-translational forms of protein regulation, such as those mediated by enzymatic activities and protein–protein and/or protein–biomolecule interactions (64, 65).

To expedite the analysis of post-translational forms of protein regulation, focused proteomics and/or functional proteomics have formed subdivided strategies (66, 67), which analyze a limited subset of a proteome with common features including a particular post-translational modification, an enzymatic activity, a specific cellular localization, and the functional relationship among proteins in an identified protein cluster. Those strategies also intended to fill the gaps between the ideal proteome analysis that completely characterizes the entire proteome and the inability to characterize it due to the current technological limitations (60, 68). The subdivided strategies for functional or focused proteomics are modification-specific proteomics that focuses on mapping post-translational modifications (69, 70), and activity-based protein profiling (ABP), in which a chemical probe is used to label and isolate an enzyme from a complex mixture and allows searching substrates of a specific enzyme and/or each of mechanistically distinct enzyme class (64, 71–73). Those strategies also include subcellular (organelle) proteomics that focuses on mapping proteomes of subcellular structure or organelles (74–76), machinery (complex interaction) proteomics that focuses on mapping functional multiprotein complexes, cellular machinery, and interactions (interactomes) (77–83), and dynamic proteomics that deals with the need to monitor protein-redistribution events (Table 1-2) (75, 84–87). The development of various new strategies is also ongoing (88). Those subdivided strategies are well designed for attacking proteome functions from a broad point of view of post-translational forms of protein regulation as follows.

1-3-1 Modification-Specific Proteomics

Post-translational modifications (PTMs) of proteins 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, and can determine its activity state, localizations, turnover, and interactions with other proteins (69). Many vital cellular processes are governed not only by the relative abundance of proteins but also by these PTMs. However, the full extent and functional importance of protein modifications in the working cell are not well understood because of a lack of suitable methods for their large scale study. At least 200 different PTMs are known (http://en.wikipedia.org/wiki/Posttranslational_modification) (89); the best characterized PTMs in eukaryotes are phosphorylation and glycosylation, but other common PTMs include acetylation, methylation, ubiquitination, and sumoylation (Table 1-3). A single protein can be modified not only at multiple sites within the molecule but also in various combinations of those PTMs. In addition, protein modifications at given sites are typically not homogeneous; a specific modification may occur only partially at a given site of the protein. The amount of protein in a single modification state can thus be a very small fraction of the total amount of the whole population of the protein; thus, the specification of a PTM at a given site requires in general a large amount of the protein. Furthermore, a single gene can give rise to a number of gene products as a result of alternative splicing. All of those contribute to the extreme complexity of an entire proteome and cause difficulty in characterizing even a specific type of PTM on the proteomic scale.

Many techniques for mapping PTMs had been developed in classical protein chemistry and are now being examined for their applicability with new ideas and technologies on the proteomic scale. Those efforts are called *modification-specific proteomics* and have now been reported for a number of different PTMs and, especially in phosphorylation and glycosylation analyses, are beginning to yield results for proteomic-wide PTM analysis (Table 1-3) (69, 90). In an early stage of modification-specific proteomics, 2D-PAGE was used as the first choice for mapping PTMs. 2D-PAGE has the highest resolution of the known protein-separation methods and has sufficient resolution to separate modification states of a protein in some cases. For example, modifications that cause changes of protein charge, such as phosphorylation and glycosylation, result in the horizontal shift of protein spots and may be detected on 2D-PAGE gels. The same modification may also result in a change of molecular weight, depending on the molecular weights of the modified functional groups. However, the mobility changes on 2D-PAGE gels alone specify neither the protein nor the type of modification. Because MS measures mass-to-charge ratio (m/z), yielding the molecular weight and fragmentation pattern of peptide derived from proteins, it represents a general method for all modifications that change the molecular weight (69). Thus, in modification-specific proteomics, MS methods are used in conjunction with the methods for protein separation as common technologies to identify proteins, the type of PTM, and/or the specific sites of the modification in proteins (Table 1-3). Although 2D-PAGE is one choice for detecting PTM of proteins in combination with detection methods such as Western blotting, chemical labeling,

TABLE 1-3. Post-translational Modifications

Modification Type	Amino Acid Modified	Functional Group	Mass Change (Da)	References to Proteomic Scale Analysis	Principle for Experimental Specification of a Modification
Phosphorylation	Tyr, Ser/Thr, His/Asp (in prokaryote)	PO ₄ ⁻	80	131, 132	<i>In vivo</i> labeling with ³² P/ ³³ P (ATP/GTP)
Catalyzed by protein kinase	Tyr, Ser/Thr			133, 134	Enrichment of proteins with phospho-Ser/ phospho-Thr by immunoprecipitation. Immunodetection with antibody against phospho-Tyr, phospho-Thr, or phospho-Ser.
	Tyr			135	Precursor ion scanning in positive-ion mode utilizing the immonium ion of phosphotyrosine (called phosphotyrosine-specific immonium ion scanning) on a Q-ToF after immunopurification with anti-phosphotyrosine antibody
	Ser/Thr			92	Modification of phosphopeptides with free sulfhydryls that are then trapped by covalent attachment to iodoacetic acid-linked glass beads. Acid elution regenerates phosphopeptides, which are then analyzed by MS.
	Ser/Thr			93, 136	β-Elimination and addition of biotinyl-iodoacetamidyl-3,6-dioxoactanediamine containing either four alkyl hydrogen or four alkyl deuterium atoms. Biotinylated peptides are further purified in a second avidin-binding step and analyzed by MS.

Ser/Thr	137	<p>β-Elimination and addition of aminoethylcysteine convert phospho-Ser/phospho-Thr into Lys analogs aminoethylcysteine and β-methylaminoethylcysteine, respectively. Trypsin cleaves the peptide bond at the C-terminal side of the converted amino acid. Immobilized aminoethylcysteine can be used to capture proteins/peptides with Ser/Thr phosphorylation sites.</p>
Ser/Thr	138	<p>An immobilized library of partially degenerate phosphopeptides biased toward a particular protein kinase phosphorylation motif is used to isolate phospho-binding domains that bind to proteins phosphorylated by specific kinase.</p>
Tyr, Ser/Thr Tyr	139, 140 141, 142	<p>Top-down MS/bottom-up MS/MS. <i>In vivo</i> labeling with ^{12}C and ^{13}C tyrosine and immunoisolation with anti-phosphotyrosine antibody for quantitative comparison.</p>
Tyr, Ser/Thr Tyr, Ser/Thr	143 144	<p>^{32}P labeling and Edman sequencing/2DE. Immunoprecipitation of targeted phosphorylated proteins from cell extract labeled with SILAC, quantitative FTICR-MS analysis to monitor the kinetics of multiple, ordered phosphorylation events on protein players in the canonical mitogen-activated protein kinase signaling pathway.^a</p>

(continued)

TABLE 1-3. (Continued)

Modification Type	Amino Acid Modified	Functional Group	Mass Change (Da)	References to Proteomic Scale Analysis	Principle for Experimental Specification of a Modification
Phosphorylation	Tyr			145	Oligonucleotide-tagged multiplex assay. Multiple SH2 domains are labeled by domain-specific oligonucleotide tags, applied as probes to complex protein mixtures in a multiplex reaction and phosphotyrosine-specific interactions are quantified by PCR.
	Ser/Thr			146	The method involves phosphorylation of proteins using ATP- γ -S and the selective <i>in situ</i> alkylation of the resultant thiophosphorylated proteins, resulting in a stable covalent bond. The thiophosphate-specific alkylating reagent can be linked to biotin or solid support (e.g., glass or Sepharose beads) with or without a photocleavable linker to facilitate convenient, high yield isolation of phosphorylated peptide/proteins.
	Tyr, Ser/Thr			147–151	Enrichment of phosphoproteins/phosphopeptides with immobilized metal affinity chromatography (IMAC) and analysis by MS.

Glycosylation	Asn	Oligosaccharide	>800	152	<p>Isotope-coded glycosylation-site-specific tagging (IGOT) is based on the lectin column-mediated affinity capture of a set of glycopeptides generated by tryptic digestion of protein mixtures, followed by the peptide:<i>N</i>-glycosidase-mediated incorporation of a stable isotope tag, ¹⁸O, specifically into the <i>N</i>-glycosylation site. The ¹⁸O-tagged peptides were then assigned by 2D-LC-MS/MS.</p>
Catalyzed by a series of glycosyltransferases	Asn			153	<p>Proteins from two biological samples are oxidized and coupled to hydrazine resin. Nonglycosylated peptides are removed by proteolysis and extensive washes. The <i>N</i>-terminus of glycopeptides are isotope-labeled by succinic anhydride carrying either d0 or d4. The beads are then combined and the isotopically tagged peptides are released by peptide- <i>N</i>-glycosidase F (PNGase F). The recovered peptides are then identified and quantified by MS/MS.</p>

(continued)

TABLE 1-3. (Continued)

Modification Type	Amino Acid Modified	Functional Group	Mass Change (Da)	References to Proteomic Scale Analysis	Principle for Experimental Specification of a Modification
Glycosylation	Thr/Ser		>203, >800	154	Mild β -elimination followed by Michael addition of dithiothreitol (Cleland's reagent, DTT) (BEMAD) or biotin pentylamine (BAP) to tag O-GlcNAc sites (as well as phosphorylation sites). The tag allows for enrichment via affinity chromatography and is stable during collision-induced dissociation, allowing for site identification by LC-MS/MS. An immunoaffinity and enzymatic strategy is provided to discriminate between O-GlcNAc and phosphorylation sites with the use of BEMAD.
	Ser/Thr			155	The approach involves lectin-isolation, 2DE-PAGE, and in-gel protease, MS analysis.
Methylation	Arg	CH ₃	14	156	Four arginine methyl-specific antibodies (ASYM24 and ASYM25 are specific for asymmetrical DMA; SYM10 and SYM11 recognize symmetrical DMA), which were generated by using peptides with aDMA or sDMA in the context of different RG-rich sequences, are used to immunoprecipitate proteins and they are analyzed by microcapillary LC-MS/MS.

Catalyzed by methyltransferase	Arg	157	Direct LC-MS/MS identification of isolated Golgi fraction.
	Arg/Lys	158	Isotopically heavy [¹³ CD(3)]methionine is metabolically converted to the sole biological methyl donor [¹³ CD(3)]S-adenosyl methionine by SILAC method. Heavy methyl groups are fully incorporated into <i>in vivo</i> methylation sites, directly labeling the PTM. Methylated proteins are isolated by using antibodies targeted to methylated residues. Identification and relative quantitation of protein methylation is done by LC-MS/MS.
Acetylation	N terminus	140	Top-down MS/bottom-up MS/MS.
	N terminus	140	Top-down MS/bottom-up MS/MS.
Ubiquitination	Lys	159	The proteins are purified by immunoaffinity chromatography with anti-ubiquitin antibody under denaturing or native conditions. They are then digested with trypsin, and the resulting peptides were analyzed by 2DLC-MS/MS. After tryptic digestion, ubiquitination site is modified with the Gly-Gly (+114.1 Da) dipeptide.
		42	
		>1000	

TABLE 1-3. (Continued)

Modification Type	Amino Acid Modified	Functional Group	Mass Change (Da)	References to Proteomic Scale Analysis	Principle for Experimental Specification of a Modification
Ubiquitination Catalyzed by several enzymes, including an ubiquitin-activating enzyme (E1), an ubiquitin-conjugating enzyme (E2), an ubiquitin-protein ligase (E3), and a deubiquitinating enzyme (DUB)	Lys			160	The method involves 6×His-tag ubiquitin expression, affinity chromatography isolation with Ni-chelate resin column, proteolysis with trypsin, and analysis by 2D LC-MS/MS.
Sumoylation	Lys	Small ubiquitin-like modifier 1 (SUMO-1, 101 AA polypeptide)	-12 kDa	161	The method involves <i>in vitro</i> expression cloning (IVEC) screen for SUMO-1 substrates. Briefly, DNA was <i>in vitro</i> transcribed and translated (IVT) in reticulocyte lysate in the presence of ³⁵ S-methionine and subjected to <i>in vitro</i> sumoylation reactions, which contained IVT product, Aos1-Uba2, Ubc9, and SUMO-1. Sumoylation was detected SDS-PAGE followed by autoradiography. ⁶

<p>Catalyzed by a cascade of enzymes, including SUMO isopeptidases (SENFs), SUMO activating enzyme (a heterodimer of Aos1-Uba2), SUMO-conjugating enzyme (Ubc9), and SUMO ligases</p>	<p>Lys</p>	<p>SUMO-1, SUMO-2, SUMO-3</p>	<p>162–164</p>	<p>The overall strategy involves the development of a stable transfected cell line expressing a double-tagged SUMO under a tightly negatively regulated promoter, followed by the induction of the expression and conjugation of the tagged modifier to cellular proteins, the use of a tandem affinity purification (TAP) method for the specific enrichment of the modified proteins, and the identification of the enriched proteins by LC-MALDI-MS/MS.^c</p>
<p>Lys</p>	<p>SUMO-1, SUMO-2</p>	<p>165</p>	<p>His6-SUMO-1 or His6-SUMO-2 is expressed stably in HeLa cells. These cell lines and control HeLa cells are labeled with stable arginine isotopes. His6-SUMOs are enriched from lysates using immobilized metal affinity chromatography. Quantitative proteomics analyzed the target protein preferences of SUMO-1 and SUMO-2.^d</p>	

TABLE 1-3. (Continued)

Modification Type	Amino Acid Modified	Functional Group	Mass Change (Da)	References to Proteomic Scale Analysis	Principle for Experimental Specification of a Modification
Sumoylation		SUMO-2		166	A stable HeLa cell line expressing His6-tagged SUMO-2 was established and used to label and purify novel endogenous SUMO-2 target proteins. Tagged forms of SUMO-2 were functional and localized predominantly in the nucleus. His6-tagged SUMO-2 conjugates were affinity purified from nuclear fractions and identified by mass spectrometry.
	Lys	SUMO		167	The method involves expression of SUMO with both N-terminal (His)6- and FLAG-tags, two-step isolation of the sumoylated proteins by Ni-NTA chromatography and a FLAG affinity purification in tandem, and identification by LC-MS/MS analysis using an LTQ FTMS.
	Lys	SUMO		168	The method involves expression of SUMO with both N-terminal (His)8-tag, one-step isolation of the sumoylated proteins by Ni-NTA chromatography, and identification by multi-LC-MS/MS -SEQUEST analysis after digestion with Lys-C and trypsin.

S-nitrosylation

Cys

NO

30

169

The approach, termed SNOSID (SNO Site Identification), is a modification of the biotin-swap technique, comprising methylation of all Cys-thiols in a protein mixture, followed by selective reduction of S—NO bonds, thereby generating a new unmodified thiol at each former SNO-Cys site. These new thiols are then marked by introduction of a mixed-disulfide bond with a biotin-tagging reagent, captured on immobilized avidin, trypsinolysis, affinity purification of biotinylated-peptides, and amino acid sequencing by LC-MS/MS.^e

Catalyzed by NO synthase

170

The method involves the biotin-switch method. The first step is methylation of all free Cys-thiols in a protein mixture, followed by selective reduction of S—NO bonds, thereby generating a new unmodified thiol at each former SNO-Cys site. These new thiols are then marked by introduction of a mixed-disulfide bond with a biotin-tagging reagent, captured on immobilized avidin, and selectively released from avidin by reduction of the disulfide linker. The isolated proteins are separated by 1D- or 2D-PAGE and in-gel trypsinolysis for identification of SNO proteins by mass spectrometry.^f

TABLE 1-3. (Continued)

Modification Type	Amino Acid Modified	Functional Group	Mass Change (Da)	References to Proteomic Scale Analysis	Principle for Experimental Specification of a Modification
Nitration	Tyr	NO ₂	45	171	The approach involves in 2D-PAGE separation, Western blotting with anti-3-nitrotyrosine antibodies, in-gel protease digestion, and identification by MALD-ToF and MS/MS. ⁸
Carbonylation				172, 173	The method involves derivatization of the carbonyl groups in the protein side chain to 2,4-dinitrophenylhydrazine (DNPH) by reaction with 2,4-dinitrophenylhydrazine, blotted onto PVDF membrane or nitrocellulose paper after 2D-PAGE, and detected with anti-DNP antibodies.
Acylation (isoprenylation)	Cys	Farnesyl (15-carbon farnesyl isoprenoid)	204	174	The approach involves 2D-PAGE separation and LC-MS/MS analysis coupled with a hydrazide biotin-streptavidin methodology in order to identify protein carbonylation in aged mice. ⁶
				175	Tagging-via-substrate (TAS) technology involves metabolic incorporation of a synthetic azido-farnesyl analog and chemoselective derivatization of azido-farnesyl-modified proteins by a Staudinger reaction, using a biotinylated phosphine capture reagent. The resulting protein conjugates can be specifically detected and/or affinity-purified by streptavidin-linked horseradish peroxidase or agarose beads, respectively. ⁷

N terminus	Myristoyl	210	176	The N-myristoylation reaction is coupled to that of pyruvate dehydrogenase, and NADH is continuously detected spectrophotometrically.
	Palmitoyl	238		

^aDynamic analysis of phosphorylation events by quantitative proteomics.

^bSmall ubiquitin-like modifier (SUMO) regulates diverse cellular processes through its reversible, covalent attachment to target proteins. Many SUMO substrates are involved in transcription and present in chromatin structure. Sumoylation appears to regulate the functions of target proteins by changing their subcellular localization, increasing their stability, and/or mediating their binding to other proteins.

^cSumoylation consensus motifs: ΨKXXE (Ψ, a hydrophobic residue; X, any residue), mammalians.

^dFourteen of the 25 SUMO-1 conjugated proteins contain zinc fingers. Sumoylation is strongly associated with transcription since nearly one-third of the identified target proteins are putative transcriptional regulators.

^eReversible addition of NO to Cys-sulfur in proteins, a modification termed S-nitrosylation, is an ubiquitous signaling mechanism for regulating diverse cellular processes. The method is applied to rat cerebellum lysates.

^fThe method is applied to mesangial cells.

^gMany mammalian proteins are inactivated by nitration of tyrosines, among which are manganese superoxide dismutase and glutamine synthase. In addition, important enzymes or structural proteins such as manganese superoxide dismutase, neurofilament L, actin, and tyrosine hydroxylase have been indicated as the targets of tyrosine nitration in pathological conditions and animal models of disease F37. All these findings suggest an important role of protein nitration in modulating activity of key enzymes in neurodegenerative disorders.

^hProtein carbonylation content is widely used as a marker to determine the level of protein oxidation that is caused either by the direct oxidation of amino acid side chains (e.g., proline and arginine to glutamyl/semialdehyde, lysine to amino adipic semialdehyde, and threonine to aminoketobutyrate) or via indirect reactions with oxidative by-products [lipid peroxidation derivatives such as 4 hydroxynonenal (HNE), malondialdehyde (MDA), and advanced glycation end products (AGEs)]. A deleterious consequence of these oxidative impairments is protein dysfunction.

ⁱProtein farnesylation is a post-translational modification involving the covalent attachment of a 15-carbon farnesyl isoprenoid through a thioether bond to a cysteine residue near the C terminus of proteins in a conserved farnesylation motif designated the "CAAX box." Azido-farnesylated proteins maintain the properties of protein farnesylation, including promoting membrane association, Ras-dependent mitogen-activated protein kinase activation, and inhibition of lovastatin-induced apoptosis.

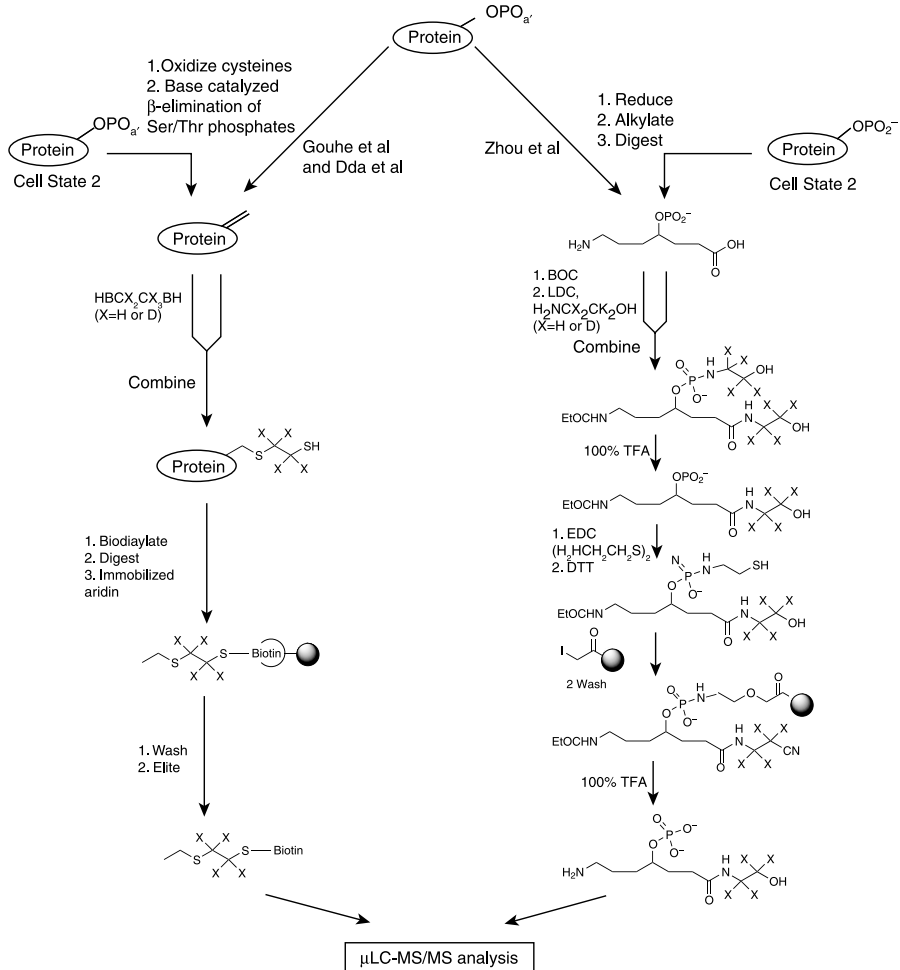


Fig. 1-6. Chemical approaches to post-translational proteomics, illustrating methods for analyzing protein phosphorylation in a complex biological mixture. The base-catalyzed phosphate elimination method (left flowchart) and the phosphoramidate modification method (right flowchart) are outlined. TFA, trifluoroacetic acid; *t*BOC, *tert*-butoxycarbonyl; DTT, dithiothreitol; EDC, 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride. [From G. C. Adam, E. J. Sorensen, and B. F. Cravatt, *Mol. Cell. Proteomics* (Ref. 64) (2003). Copyright (2003) by American Society for Biochemistry and Molecular Biology, Inc. Reproduced with permission of the ASBMB via the Copyright Clearance Center.]

or *in vivo* isotope labeling specific to a certain PTM, affinity-based isolation methods with specific antibodies or with chemical reagents that specifically react to a certain PTM site of the amino acid residues are often adopted to collect or concentrate proteins with a specific type of PTM (Fig. 1-6).

The earliest works on modification-specific proteomics using chemical reagents are those for selectively modifying phosphoproteins within complex mixtures (91).

In those approaches, modified peptides are enriched by covalent or high affinity avidin–biotin coupling to immobilized beads, allowing stringent washing to remove nonphosphorylated peptides. One method begins with a proteolytic digest, which was alkylated after reduction to eliminate reactivity from cysteine (92). After protection of N and C termini, phosphoramidate adducts at phosphorylated residues are formed by carbodiimide condensation with cystamine. The free sulfhydryl groups produced from this step are captured covalently onto glass beads coupled to iodoacetic acid, are cleaved off with trifluoroacetic acid, and are eluted from the beds. MS analyzes the regenerated phosphopeptides. On the other hand, another method starts with a protein mixture in which cysteine is oxidized with performic acid (93). β -Elimination of phosphate from phosphoserine and phosphothreonine is induced by base hydrolysis, and the resulting alkenes are modified by ethandithiol to produce free sulfhydryls that allow coupling to biotin. The biotinylated phosphoproteins are then captured with avidin-affinity beads, eluted, and digested with trypsin. The peptides are again captured with avidin beads, washed, and eluted for MS analysis.

Currently, many works adopting affinity-based isolation methods and/or isotope-labeling methods in conjunction with MS-based protein identification methods have been reported for PTMs other than phosphorylation. Each approach is well designed for adapting to each specific character of the PTM. Those are summarized in Table 1-3. Because of importance of PTMs, the details of some approaches including experimental protocols for modification-specific proteomics will be described in Section 2-2.

1-3-2 Activity-Based Protein Profiling/ Enzyme Substrate Proteomics

Activity-based profiling (ABP) provides a strategy for identifying proteins associated with a particular biological activity—typically enzymatic activity (57)—and utilizes synthetic chemistry to create tools and assays for the characterization of protein samples of high complexity (64). The ABP strategy simplifies a complex biological mixture of proteins before analysis by labeling a specific set of related proteins with an affinity or fluorescence tag (Fig. 1-7), and thus includes the development of chemical affinity tags to react to the active site of enzymes and, in certain cases, to measure the relative expression level and PTM state of proteins in cell and tissue proteomes. This strategy, in a certain sense, may share some of the approaches with those of modification-specific proteomics (described in Section 1-3-1) and quantitative proteomics (which permits the quantitative comparison of proteins and allows monitoring of dynamics in protein function in complex proteomes; this strategy, typically using ICAT reagents, is described in Chapter 2). Many of those approaches (not restricted to ABP) interface well-established approaches in molecular biology or cell biology with proteomics. As a common theme, but in contrast to classic cell biological or biochemical research, the approaches are all designed to allow systematic screening of proteins in a defined experimental paradigm (75).

In the original version of ABP, fluorophosphonates (known to be specific covalent inactivators of serine proteases) were linked to biotin. In this strategy, a fluorophosphonate irreversibly phosphonylates the active-site serine only in those proteases that are catalytically active, thereby attaching a biotin moiety that could be affinity captured with avidin. Biotin can be substituted by a fluorophore to detect

(A)



(B)

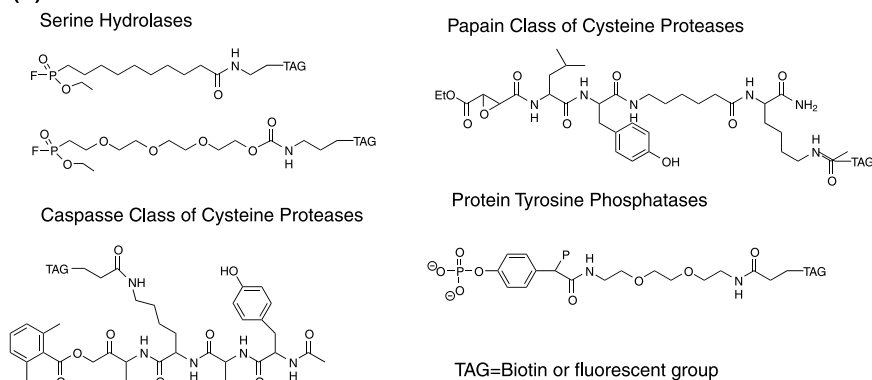


Fig. 1-7. Chemical approaches to activity-based protein profiling (ABP). (A) General structure of activity-based probes and (B) some probes directed toward specific classes of enzymes. ABP probes label the active sites of target enzymes. NU, nucleophilic amino acid residue; RG, reactive group; L, linker; TAG, detection and/or affinity tag. [From G. C. Adam, E. J. Sorensen, and B. F. Cravatt, *Mol. Cell. Proteomics* (Ref. 64) (2003). Copyright (2003) by American Society for Biochemistry and Molecular Biology, Inc. Reproduced with permission of the ASBMB via the Copyright Clearance Center.]

and increase the sensitivity of the initial screening step. Isolation of the tagged serine proteases followed by MS analysis allows the identification of active serine proteases. Thus, the approach for identifying functional proteins is expected to be specific for identifying members of the serine protease family (57). Principally, the same approach as this can be applied to the identification of many other functional families of enzymes. In fact, many ABP probes have been designed for isolating other functional families of enzymes, including cysteine proteases (94), ubiquitin-specific protease (95), threonine proteases, metalloproteases, protein phosphatases, kinases, glucosidase, exoglycosidases, and transglutaminases (Table 1-4). In some cases, well-known inhibitors were exploited to direct probe reactivity toward a specific class of enzyme and the designed probes have been shown to label selectively active enzymes, but not their inactive precursor (e.g., zymogen) or inhibitor-bound forms (71, 96, 97). Active sites of enzymes invariably contain nucleophilic groups (that participate in diverse reactions such as acids, bases, or nucleophiles) and the environments of groups (that are involved in catalysis); in some other cases, the ABP probes are also designed as nonspecific electrophiles and are directed to react over a much wider range

TABLE 1-4. Activity-Based Protein Profiling

Targeted Protein (Enzyme) Family	Members Identified	Active Site	Binding and Reactive Group	Known Inhibitor	Affinity Tag/ Reporter	Reference	Note
Serine hydrolase	Proteases, lipases, esterases, and amidases	Ser	Fluorophosphonate/ fluorophosphate (FP) derivatives	Diisopropyl fluorophosphate etc.	Biotin (5-(biotinamido) pentyl-amine) (NH ₂ -biotin, Pierce)	177, 178	The chemical probes inert to the most part of cysteine, aspartyl, and metallohydrolases and react in a catalytically active state.
Cysteine protease	Bleomycin hydrolases, calpains, caspases, cathepsins, thiol-ester motif-containing protein (TEP) 4 (fly), etc.	Cys	Epoxide electrophiles [E64 and derivatives containing a P2 leucine residue (DCG-03 and DCG-04)/DCG-04 derivatives replaced Leu with all natural amino acids]	Diazomethyl ketones, fluoromethyl ketones, acyloxymethyl ketones, <i>O</i> -acylhydroxylamines, vinyl sulfones and epoxysuccinic derivatives.	Biotin/ radio-iodine	94, 179, 180	Most cysteine proteases are synthesized with an inhibitory propeptide that must be proteolytically removed to activate the enzyme, resulting in expression profiles that do not directly correlate with activity. The largest set of papain-like cysteine proteases, the cathepsins, act in concert to digest a protein substrate. Isotope-coded activity-based probe for the quantitative profiling of cysteine proteases is also developed.

(continued)

TABLE 1-4. (Continued)

Targeted Protein (Enzyme) Family	Members Identified	Active Site	Binding and Reactive Group	Known Inhibitor	Affinity Tag/ Reporter	Reference	Note
Mechanistically distinct enzyme classes	Acetyl-CoA- acetyltransferase, aldehyde dehydrogenase, NAD/ NADP-dependent oxidoreductase, enoyl CoA hydratase, epoxide hydrolase, glutathione S-transferase, 3 β -hydroxysteroid dehydrogenase/ 5-isomerase, platelet phosphofructokinase, type II tissue transglutaminase, the endocannabinoid- degrading enzyme fatty acid amide hydrolase (FAAH), triacylglycerol hydrolase (TGH) and an uncharacterized membrane-associated hydrolase	Probably Cys/ Asp/ Glu/His	Sulfonate ester de- rivatives (phenyl-, quinoline-, octyl, nitrophenyl, naph- thyl, mesyl, pyridyl, thiophene, and azido)	—	Biotin, rhodamine/ biotin-rho- damin	71, 72, 181– 183	Libraries of candidate probes are screened against complex proteomes for activity- dependent protein reactivity in a nondirected or combinatorial strategy.

Type II tissue	Non-specific	α -chloroacetamide (α -CA) reactive group and a variable dipeptide binding group	—	Biotin, rhodamine, biotin-rhodamin	96	The α -chloroacetamide (α -CA) reactive group, consistent with the behavior of other moderately reactive carbon electrophiles, proved capable of labeling in an active site-directed manner several mechanistically unrelated enzyme classes, thereby further expanding the scope of enzymes addressable by ABPP. In total, more than 10 different classes of enzymes were identified as targets of the α -CA probe library, most of which were not labeled by previously described ABPP probes.
transglutaminase (tTG2), formiminotransferase cyclodeaminase, aldehyde dehydrogenase-9, aminolevulinatase, Δ -dehydratase, epoxide hydrolase, cathepsin Z, and the muscle and brain isoforms of creatine kinase (CK), [e.g., Asp-Asp α -CA targets two isoforms of CK, Gly-Gly α -CA target tTG2; Leu-Met (maleylacetoacetate isomerase, ATP citrate lyase), Leu-Asp (hydroxyypyruvate reductase, peroxiredoxin), Leu-Arg (malic enzyme) etc.]	Non-specific	α -chloroacetamide (α -CA) reactive group and a variable dipeptide binding group	—	Biotin, rhodamine, biotin-rhodamin	96	The α -chloroacetamide (α -CA) reactive group, consistent with the behavior of other moderately reactive carbon electrophiles, proved capable of labeling in an active site-directed manner several mechanistically unrelated enzyme classes, thereby further expanding the scope of enzymes addressable by ABPP. In total, more than 10 different classes of enzymes were identified as targets of the α -CA probe library, most of which were not labeled by previously described ABPP probes.

(continued)

TABLE 1-4. (Continued)

Targeted Protein (Enzyme) Family	Members Identified	Active Site	Binding and Reactive Group	Known Inhibitor	Affinity Tag/ Reporter	Reference	Note
Ubiquitin-specific protease (deubiquitinating enzyme)	Ub-processing proteases (UBP), Ub carboxyl-terminal hydrolases, ubiquitin-specific protease (UPS)-4, -5, -7, -8, -9X, -10, -11, -12, -13, -14, -15, -15i, -16, -19, -22, -24, -25, -28, CYLD-1, m64E, UPS flag1 (KIAA891), UCHL1, UCHL3, UCH37, and HSPC263	Cys	Ub-derived active-site-directed probes (four Michael acceptor-derived probes, vinyl methyl sulfone, vinyl methyl ester, vinyl phenyl sulfone, and vinyl cyanide, and three alkylhalide-containing inhibitors, chloroethyl, bromoethyl, and bromopropyl), which are designed to react at a position that corresponds to the C-terminal carbonyl of the Gly76 amide bond conjugating Ub to its substrate.	—	Influenza hemagglutinin (HA)	184-188	The family of ubiquitin (Ub)-specific proteases (USP) removes Ub from Ub conjugates and regulates a variety of cellular processes. Four major classes are identified. UBPs can hydrolyze both linear and branched Ub modifications whereas the activity of UCH enzymes is restricted to the hydrolysis of small Ub C-terminal extensions. A third USP family contains an ovarian tumor (OTU) domain with USP activity. Finally, RPN11/POH-1, a proteasome 19S cap subunit belonging to the Jab1/MPN domain-associated metalloisopeptidase (JAMM) family that lacks the cysteine protease signature, and cleave Ub from substrates in a Zn ²⁺ - and ATP-dependent manner.

Proteasome	Proteasomal beta subunits, HsIV subunit of the <i>Escherichia coli</i> protease complex HsIVyHsIU	Thr	Peptide vinyl sulfone, carboxybenzyl-leucyl-leucyl-leucine vinyl sulfone (Z-L3VS)	—	Biotin/nitrophenol moiety	189	Proteasomes are multicatalytic proteolytic complexes found in almost all living cells and are responsible for the degradation of the majority of cytosolic proteins in mammalian cells. The tripeptide vinyl sulfone Z-L3VS and related derivatives inhibit the trypsin-like, the chymotrypsin-like, and, unlike lactacystin, the peptidylglutamyl peptidase activity of the proteasome in vitro by covalent modification of the NH ₂ -terminal threonine of the catalytically active b subunits.
Metalloproteases	Matrix metalloproteinases (MMP)-2, 7, 9, neprilysin, aminopeptidase, and dipeptidylpeptidase	Zinc-activated water molecule	Zinc-chelating hydroxamate (that chelates the conserved zinc atom in MP-active sites in a bidentate manner -a benzophenone photocrosslinker (a photolabile diazirine group: converting tight-binding reversible MP inhibitors into active site-directed affinity labels by incorporating a photocrosslinking group into these agents.)	Hydroxamate inhibitor GM6001 (ilomastat), marimastat, peptidyl hydroxamate zinc-binding group (ZBG)	Biotin-rhodamine/rhodamine	97, 190	Metalloproteases (MPs) are a large and diverse class of enzymes implicated in numerous physiological and pathological processes, including tissue remodeling, peptide hormone processing, and cancer. In the cases of serine and cysteine proteases, ABPP probes are designed to target conserved nucleophiles in protease active sites, an approach that cannot be directly applied to MPs, which use a zinc-activated water molecule (rather than a protein-bound nucleophile) for catalysis.

(continued)

TABLE 1-4. (*Continued*)

Targeted Protein (Enzyme) Family	Members Identified	Active Site	Binding and Reactive Group	Known Inhibitor	Affinity Tag/ Reporter	Reference	Note
Phosphatase	Protein tyrosin phosphatase (PTP)- 1B, prostatic acid phosphatase, protein Ser/Thr phosphatase calcineurin	Cys	4-fluoromethylaryl phosphate [phenylphosphate (recognition head)- <i>p</i> -hydroxymandelic acid derivatives]	—	Biotin/ dansyl fluorophore	191, 192	When the designated bond between the recognition head and the latent trapping device (<i>p</i> -hydroxymandelic acid derivatives) is selectively cleaved with the assistance of the target hydrolase, it becomes activated by the release of <i>p</i> -hydroxybenzyllic fluoride intermediate. The intermediate quickly undergoes 1,6-elimination to produce highly reactive quinone methide (QM), which in turn could alkylate suitable nucleophiles on nearby hydrolases to form labeled adduct. Here, the latent trapping device serves as the core of the probe and the QM plays a critical role in the covalent labeling of target hydrolase.

	The probe is specific to PTPs including PTP1B, HePTP, SHP2, LAR, PTP, PTPH1, VHR, and Cdc14.	Cys	α -bromobenzyl-phosphonate	—	Biotin	193	The activity-based probe is targeted to the PTP active site for covalent adduct formation that involves the nucleophilic Cys.
Kinase	Cyclin G-associated kinase (GAK), casein kinase I-alpha and I-epsilon, RICK [Rip-like interacting caspase-like apoptosis-regulatory protein (CLARP) kinase/Rip2/CARDIAK], GSK3 beta, JNK,	Thr	Protein kinases are key regulators of cellular signaling and therefore represent attractive targets for therapeutic intervention in a variety of human diseases. p38 kinase inhibitor SB 203580	p38 kinase inhibitor SB 203580	Epoxy-activated sepharose 6B	194	The crystal structure of p38 in complex with SB 203580 shows exposure of the inhibitor's sulfoxide moiety at the protein surface, suggesting a suitable site for the attachment of linkers extending from solid support materials.
Glucosidase	β -glucosidase	ND	β -glucosyl unit (recognition head)- <i>p</i> -hydroxymandelic acid derivatives)	—	Dansyl fluorophore, biotin	191, 195	The active site reacts to quinone methide intermediate.

(continued)

TABLE 1-4. (Continued)

Targeted Protein (Enzyme) Family	Members Identified	Active Site	Binding and Reactive Group	Known Inhibitor	Affinity Tag/ Reporter	Reference	Note
Beta- endoglycosidases	β -1,4-glycanases (from <i>Cellulomonas fimi</i>), an endoxylanase (Bex from <i>Bacillus circulans</i>) and a mixed-function endoxylanase/ cellulase (Cex from <i>Cellulomonas fimi</i>)	Glu	2,4-dinitrophenyl 4'-amido-2,4'- dideoxy-2- fluoro-xylobioside inactivator	—	Biotin	98, 196	One enzyme molecule reacted with one DNP2FX2SSB (2,4-dinitrophenyl 4'-amino-N- {7-[N-(D-biotinoyl)-13-amino- 4,7,10-trioxatridecanyl]amino}- 4,5-dithiaheptanoyl-2,4'-dideoxy- 2-fluoro-xylobioside) molecule, forming the biotinylated fluoroglycosyl-enzyme intermediate and releasing 2,4- dinitrophenol (ate).
Transglutaminase	Substrates of transglutaminase (acyl-donor; fatty acid synthase, tumor rejection antigen-1, DNase gamma etc., acyl acceptor; myosin heavy poly peptide 9, T-complex protein Ig subunit, etc.	—	5-pentylamine (the acyl-acceptor), glutamine- containing peptide, (the acyl-donor)	—	Biotin	197, 198	Transglutaminases (TGs)1 (EC 2.3.2.13) constitute a family of enzymes that catalyze the post-translational modification of proteins. Their calcium- dependent catalytic activity is exhibited toward carboxamide groups of peptide- bound glutamine residues and amino groups of peptide-bound lysines, leading to an intrachain or interchain isopeptide bond.

of mechanistically distinct enzyme classes. The ABP probes directed for enzymes may be generalized as typically possessing three elements (Fig. 1-7): (1) a binding group that promotes interactions with the active sites of specific classes of enzymes; (2) a reactive (electrophilic) group that covalently labels those active sites (shown as “binding and reactive group” in Table 1-4); and (3) a reporter group (e.g., fluorophore or biotin) for the visualization or affinity purification of probe-labeled enzymes (shown as “affinity tag/reporter” in Table 1-4) (98). Multiple enzyme families can be attracted to the same electrophilic group; thus, some ABP probes allow the facilitated identification of large numbers of functional proteins in a particular proteome (57). By varying the electrophilic group used for the APB probe, different functional families of enzymes may be targeted. Thus, the ABP strategy analyzes proteomes based on functional properties such as enzymatic activity rather than expression level alone, and provides exceptional access to low abundance proteins in complex proteomes by concentrating specific families of enzymes with ABP probes (97).

As an extension of this strategy, the development of a new approach that utilizes rhodamine-based fluorogenic substrates encoded with PNA (protein nucleic acid) tags is a challenging attempt. The PNA tags have two arms: one is made of a chemical affinity tag similar to that used for ABP, and another is made of a defined oligonucleotide which assigns each of the substrates that interacts with the chemical affinity arm to a predefined location on an oligonucleotide microarray through hybridization with the nucleic acid arm of the PNA tag, thus allowing the deconvolution of multiple signals from a solution (99). The PNA tag approach may thus provide an additional strategy for analyzing the functional aspects of proteomes.

1-3-3 Subcellular (Organellar) Proteomics

Eukaryotic cells are compartmentalized to provide distinct and suitable environments for biochemical processes such as protein synthesis and degradation, storage of genetic materials, ribosome production, provision of energy-rich metabolites, protein glycosylation, DNA replication, and transcription. Accordingly, the compartmentalized structure of a cell is supported by subsets of proteins that are specifically targeted to particular subcellular structures. Although subcellular structures and organelles are thought to be discrete entities carrying out independent cellular functions, there are complex mechanisms of intracellular communication and contact sites between the organelles. Some proteins are associated with subcellular structures only in certain physiological states, but localized elsewhere in the cell in other states (100); among the possible mechanisms that underlie such conditional association are the protein translocation between different compartments, cycling of proteins between the cell surface and intracellular pools, or shuttling between nucleoplasm and cytoplasm. Thus, protein localization is linked to cellular function and introduces an additional strategy for proteomics at subcellular levels (75).

The strategy of cellular compartmentalization is to enrich for particular subcellular structures or organelles by subcellular fractionation with classic biochemical fractionation techniques (such as centrifugation), and to map comprehensively the proteome of these structures by an MS-based protein identification method typically

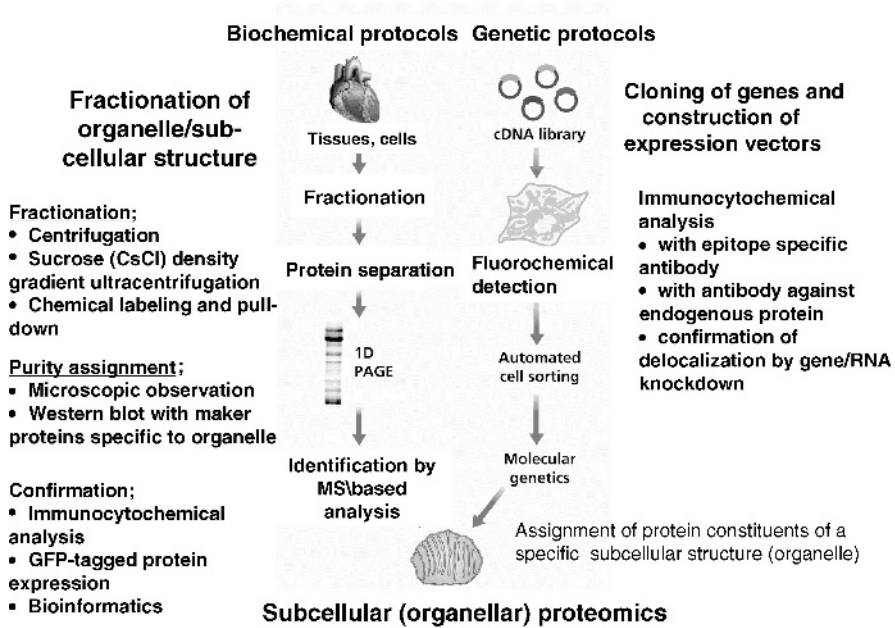


Fig. 1-8. Biochemical and genetic protocols for the isolation of subcellular structure or organelle used in subcellular (organellar) proteomics. The isolated cellular compartments are studied in terms of the protein composition and function. [Adopted by permission from Macmillan Publishers Ltd.; B. Westermann and W. Neupert, *Nat. Biotechnol.*, **21**:239–240 (2003).]

after electrophoresis-based separation or by the shotgun analysis described in Chapter 2 (see the Section 2-1) (Fig. 1-8) (75). Among the key potentials of this strategy is the capability not only to screen for previously unknown gene products but also to assign them, along with other known but poorly characterized gene products, to particular subcellular structures. This strategy is called *subcellular proteomics* (75) or *organellar proteomics* (74). Subcellular structures targeted by this strategy include not only entire organelles (such as the nucleus and mitochondria) but also nonorganelle structures (such as the postsynaptic density and raft), which can be isolated by traditional subcellular fractionation typically using sucrose density-gradient ultracentrifugation, and comprise a focused set of proteins that fulfill discrete but varied cellular functions. Subcellular proteomics or organellar proteomics ranges in scope to include cataloging studies that test the ability of a method to identify as many unique proteins as possible, in particular, unknown low abundance proteins specific to a particular organelle (74). The comprehensive identification of the proteins present in a prepared organelle by MS-based methods may reveal true components of the structure investigated at the level of the endogenous gene products, but will also yield a certain amount of false-positives, depending on the degree of impurities derived from other subcellular structures present in the preparation. Those false-positives make it hard to evaluate the biological significance of proteins that are usually associated with one organelle but are detected in the

proteome of another organelle. Although these proteins could be artifacts of subcellular fractionation procedures, they might also be biologically significant (74). Therefore, cell biological methods (such as immunocytochemical analysis and fluorescence tag analysis) as well as sequence analyses by bioinformatics tools are often required to validate the MS-based identifications (Fig. 1-8) (76). It should be noted that an innovative method called *protein correlation profiling* (PCP) was introduced to address this problem in the study of the human centrosome (101). In the PCP method, mass spectrometric intensity profiles from centrosomal marker proteins are used to define a consensus profile through a density centrifugation gradient, in direct analogy to Western blotting profiling of gradient fractions. Distribution curves generated from the intensities of tens of thousands of peptides from consecutive fractions established centrosomal proteins by their similarity to the consensus profile using mean squared deviation (χ^2 value) (102). When combined with those validation methods (103), the strategy of subcellular (organellar) proteomics allows not only assigning known proteins but also identifying previously unknown gene products in particularly defined subcellular (organellar) structures, and contributes significantly to the functional annotation of the products of a genome. Thus, subcellular structures (organelles) represent attractive targets for global proteome analysis because they represent discrete functional units; their complexity in protein composition is reduced relative to whole cells and lower abundance proteins specific to the organelle are revealed (74). In addition, the analysis at the subcellular level is a prerequisite for the detection of important regulatory events such as protein translocation in comparative studies (75). The approach has been applied to a number of subcellular structures, including plasma membrane, nucleus, nucleolus, interchromatin granule clusters, nuclear (inner, outer) membranes, centrosome, midbody, endoplasmic reticulum (microsomes), Golgi apparatus, clathrin-coated vesicles, lysosome, peroxisomes, phagosome, mitochondrion, chloroplast (plant), and to nonorganelles including synaptosome (postsynaptic density) and raft (74, 75, 104–111). The largest protein collection of subcellular localization obtained by a single study is so far that from mouse liver cells. The study localized 1404 proteins to ten cellular compartments including early endosomes, recycling endosomes, and proteasome (protein lists are available at <http://proteome.biochem.mpg.de/ormd.htm>) by using PCP introduced in the analysis of the centrosome in conjunction with other validation methods such as enzymatic assays, marker protein profiles, and confocal microscopy (102).

As a complementary approach to the biochemical protocols described earlier, a comprehensive gene expression approach in yeast has been developed (Fig. 1-8). It relies on the systematic cloning of open reading frames (ORFs) for subsequent expression or generation of genomic sets of strains expressing tagged proteins suitable for detecting cellular localization. The initial stage of proteomic scale analysis of protein localization involved a description of the cellular localization of almost half of the yeast proteins using plasmid-based overexpression of epitope-tagged proteins and genome-wide transposon mutagenesis for high throughput immunolocalization of tagged gene products (see *Saccharomyces Genome Database*, <http://www.yeastgenome.org/>) (56, 112). This approach, however, led sometimes to mislocalization of proteins, because of their overexpression, that was not responsive to normal regulatory circuitry. In the second-stage proteomic scale analysis, therefore, each yeast ORF is

fused with an affinity (TAP tag, tandem affinity purification tag) or fluorescent (GFP, green fluorescent protein) tag at the carboxyl (C) terminus, inserted on the predicted yeast ORF on the budding yeast genome by homologous recombination strategy, and is expressed from its native promoter in its endogenous chromosomal location responsive to normal regulatory circuitry (56, 113). To validate known localization, colocalization experiments are done using monomeric red fluorescent protein (RFP) fused to proteins whose cellular localization was established. This second-stage proteomic scale analysis for cellular localization has categorized about 4500 ORFs out of over 6000 strains with GFP-tagged ORFs into 22 distinguishable subcellular localization (organelle) patterns; they include cytoplasm, nucleus, mitochondrion, endoplasmic reticulum (ER), nucleolus, vacuole, cell periphery, punctate speckle, bud neck, spindle pole, vacuolar membrane, nuclear periphery, early Golgi/COPI, endosome, bud, late Golgi clathrin, Golgi, cytoskeleton, lipid particle, peroxisome, microtubules, and ER to Golgi (113). This genomic-scale information of protein localization in budding yeast enables us to make a comparison with that obtained for other species and allows us to assess functional conservation at subcellular levels among evolutionally different species. The best information available for subcellular structures other than that for yeast is that for the nucleolus of human HeLa cells. The comparison of the protein composition of yeast nucleolus with that of human HeLa cells indicates that out of the 142 yeast nucleolar proteins that have at least one human homolog, 124 are found in the human nucleolar proteome (87%). The data indicate that approximately 90% of the yeast nucleolar proteins with human homologs are also nucleolar components in HeLa cells and that the nucleolus is highly conserved throughout the eukaryotic kingdom (84). Thus, the genetic tractability of yeast allows a large fraction of yeast ORFs to be tagged for localization studies; however, such an approach is more challenging in mammalian systems due, in part, to artifacts from overexpression and to difficulty in constructing gene expression system from its native promoter in its endogenous chromosomal location (114).

All the localization information obtained for yeast is available at a public database (<http://yeastgfp.ucsf.edu>; see *Saccharomyces* Genome Database, <http://www.yeast-genome.org/>) and is useful for further analysis with yeast cells and for comparative localization analysis of the other species, although proteins with crucial C-terminal targeting signals are often mislocalized in this analysis and new fusions will have to be constructed to get an accurate view of the subcellular location of this group of proteins. Localization information is integrated into different functional genomics datasets and it will be challenging to formulate biological hypotheses, such as those regarding correlation of colocalization with transcriptional coexpression (obtained by microarray analysis) and the relationship between colocalization and physical or genetic interaction. [The relative enrichment for colocalization was assessed for the combination of protein–protein or genetic interactions in the GRID database (<http://biodata.mshri.on.ca/grid>) (56).] Thus, the challenge of global subcellular (organellar) proteomics is to provide a functional context for proteins by associating them with a distinct group of proteins in defined intracellular environments and is extremely useful for the integration of information obtained from analyses done in other proteomic strategies as well as in different functional genomics (74).

1-3-4 Machinery or Complex Interaction Proteomics

Multiprotein complexes are among the fundamental units of macromolecular organization and thus are key molecular entities that integrate multiple gene products to perform cellular functions (81). In fact, many multiprotein complexes constitute molecular machineries, such as transcription and RNA processing machinery, nuclear pore complex, preribosomal ribonucleoprotein (pre-rRNP) complex, ribosome, proteasome, and receptor–signal transduction complexes. They carry out and regulate important cell mechanisms such as transcription and RNA processing, membrane transport, ribosome synthesis, protein synthesis and degradation, and receptor–signal transduction. They require precise organization of molecules in time and space, are thought to assemble in a particular order, and often require energy-driven conformational changes, specific post-translational modifications, or chaperone assistance for proper formation. Their composition may also vary according to cellular requirements (81). Because of the difficulties of conventional protein chemical technologies in analyzing such cellular machineries or multiprotein complexes with multicomponents, most multiprotein complexes remain only partially characterized. Thus, an additional strategy for proteomics can be introduced at the levels of multiprotein complex and cellular machineries. We would like to call this strategy *machinery proteomics* or *complex (interaction) proteomics*.

Some of the multiprotein complexes (molecular machineries) can be prepared by subcellular fractionation methods using ultracentrifugation, which are often used to prepare subcellular structures or organelles as described in Section 1-3-3 [e.g., nuclear pore complex, anaphase-promoting complex, preribosomal ribonucleoprotein (pre-rRNP) complexes] (110, 111, 115, 116) or reconstituted from subcellular extracts (e.g., spliceosome) (108, 117). They can also be prepared by pull-down analyses, such as affinity purification (using epitope tag, tandem affinity purification tag, antibodies, etc.) as exemplified for spliceosome (117, 118) and pre-rRNP complexes (119–121) (see the Section 3-2-1). Once pure complexes or interacting partners are obtained, MS-based technologies allow identifying their components at subfemtomole levels on a large scale and with high throughput performance. Machinery proteomics or complex (interaction) proteomics often starts with an initial hypothesis to draft a design for particular known cellular machinery or to identify interaction partners of a particular known protein. It especially allows one to assign protein constituents with unknown function as the constituents of functionally defined cellular machinery. For instance, the constituents of the nuclear pore complex are expected to perform events related to nuclear transport, those of the preribosome perform events related to ribosome synthesis, or those of the spliceosome perform events related to mRNA splicing. Identification of unknown constituents in known machineries or of known interaction partners for a protein with unknown function may lead to the specification of function for the unknown protein. In addition, the strategy allows one to catalog data of known constituents. Thus, analysis of protein complexes or cellular machineries is one of the most useful strategies for directly assigning protein function and for annotating protein products of the genome in terms of biological activity of each protein product.

Currently, two groups, the Cellzome Corporation in Germany (82) and the University of Toronto in Canada (122), took the TAP-tag approach for genome-wide screening for complexes in an organism, the budding yeast. They are particular about endogenous expression of TAP-tagged bait proteins from their natural chromosomal locations; *Saccharomyces cerevisiae* strains are generated with in-frame insertions of TAP tags individually introduced by homologous recombination at the 3' end of each predicted open reading frame (ORF) (<http://www.yeastgenome.org/>) (55, 123). This construction of TAP-tagged protein ORFs ensures expression from its native promoter in its endogenous chromosomal location responsive to normal regulatory circuitry; thus, the formed protein complex is expected to be equal to its endogenous counterpart unless TAP tag itself has any effect on this complex formation *in vivo*. The group performed tandem affinity purification repeatedly for over 4500 different tagged proteins of the yeast; the majority of the protein complexes were purified at least several times, and the group characterized the composition and organization of the multiprotein complex/cellular machinery based on the huge dataset obtained and based on available data on expression, localization, function, evolutionary conservation, protein structure, and binary interactions. They propose that the ensemble of cellular proteins partitions into 491–547 complexes, of which about half are novel, and differentially combine with additional attachment proteins or protein modules to enable a diversification of potential functions. The detailed data is available online (the BioGRID database, <http://thebiogrid.org>; the ntAct database, <http://www.ebi.ac.uk/intact/>; the MS protein identifications, <http://yeastcomplexes.embl.de>; Euroscarf, http://web.uni-frankfurt.de/fb15/mikro/euroscarf/col_index.html) and is useful for future studies on individual proteins, biological data integration, and modeling and thus for functional genomics and systems biology. Genomic scale analysis of machinery or complex (interaction) proteomics has not been reported yet for any species other than yeast; however, at least one group in Japan has been taking an epitope-tag approach for genome-wide screening of multiprotein complexes in humans.

1-3-5 Dynamic Proteomics

Multiprotein complex modules require precise organization of molecules in time and space, are assembled in a particular order, and vary their composition according to cellular requirements as described in Section 1-3-4. The primary goal of proteomics is to describe not only the composition and connection but also the dynamics of the multiprotein modules and ideally of the entire proteome (124). The approaches using affinity purification and MS methods allow for isolation of almost any multiprotein complex formed in the cell and for the detection of many constituents of complexes in a fraction, and allow one to determine the connection between multiprotein modules on the genomic scale. The approaches enable us to probe specific states of multiprotein complexes and the network formed in some biological states by collating lists of identified proteins (profiling or cataloging analysis), or to distinguish some different states by enumerating differences in protein composition among the corresponding complexes obtained from different cell states (subtractive analysis). However, the approaches cannot tell us anything about the extent of those differences, when those

differences are caused, or how long they take to happen; thus, it remains difficult to analyze the dynamic aspect of multiprotein complexes (machineries) and even more difficult to analyze the dynamic aspect of the much higher-ordered organization of multiprotein modules (i.e., subcellular structure or organelle). To analyze the dynamic aspect, quantitative changes in protein complex abundance and composition of protein complexes/subcellular structures formed during physiological alteration in the cell have to be determined (quantitative analysis). We would like to propose a strategy for analyzing the dynamic aspect of multiprotein modules, subcellular structures, or ideally the entire proteome and call that strategy *dynamic proteomics*.

One approach used for this strategy is direct visual comparison of proteins present in the isolated protein complex or subcellular structure after protein staining on electrophoresis gels among different samples, followed by MS-based protein identification. The approach was successfully used to analyze, for example, remodeling of small nuclear ribonucleoprotein (snRNP) complexes during catalytic activation of the spliceosome, which removes introns from mRNA precursors (125). In this example, human 45S activated spliceosomes and a previously unknown 35S U5 snRNP were isolated by tobramycin-affinity selection (118) and characterized by gel-based mass spectrometry. Subtractive comparison of their protein components with those of other snRNP and spliceosomal complexes revealed dynamic changes of proteins that participated in the remodeling of splicing machinery during spliceosome activation (125). A similar approach was also applied to the analysis of reorganization of the entire human nucleolus upon transcriptional inhibition with actinomycin D (126). Proteins from nucleoli isolated from both control and actinomycin D-treated cells were separated by 1D SDS-PAGE and stained with dye. The total proteomes were similar for both control and actinomycin D-treated nucleoli on stained gels; however, there were 11 protein bands whose intensity in the actinomycin-treated nucleoli was increased relative to the control nucleoli. Those bands were excised, digested with trypsin, and analyzed by MS. All of those proteins identified by the analysis were examined by immunocytochemical analysis and were shown to be predominantly nucleoplasmic but relocated to the nucleolar periphery following actinomycin D treatment (104, 127). Those approaches are visceral and certainly very useful; however, extreme care should be taken when protein-staining bands are compared quantitatively among different samples because each compared staining may not necessarily contain a single or the same protein as that present in the corresponding staining on gels. MS-based protein identification covers the shortcomings of this comparison to some extent; however, if an excised staining band contained multiple proteins, the identified protein does not always correspond to the protein that changed staining intensity on gels among the compared samples.

The method using stable isotope tagging and MS gives an alternative approach to this gel visualization approach (128), and provides an efficient strategy for determining the specific composition, changes in the composition, and changes in the abundance of multiprotein complexes or subcellular structures. Among the reported stable isotope tagging methods, a well-known approach is based on the use of isotope-coded affinity tag (ICAT) reagents and LC-MS, which can be used to compare the relative abundances of tryptic peptides derived from suitable pairs

between different samples. Derivatization of two distinct proteomes with the light and heavy versions of the ICAT reagent provides the basis for proteome quantitation by MS analysis. Since the ICAT method can incorporate the isotopic label into only Cys-containing sites of proteins after protein extraction, it simplifies proteome analysis by isolating only Cys-polypeptides and has universal applicability. This approach, in any case, can be used to distinguish the protein components of the complex or the subcellular structure from a background of copurified proteins by comparing the relative abundances of peptides derived from a control sample and the specific complex. An example of this type of analysis is the specific identification of the components present in the RNAP II preinitiation complex that is purified from nuclear extracts by single-step promoter DNA affinity chromatography (129). This same method can certainly detect quantitative changes in the abundance and dynamic changes in the composition of protein complexes, or subcellular structures obtained from different cell states, as exemplified by an analysis of STE12 protein complexes isolated from yeast cells in different states (128).

Several other *in vitro* or *in vivo* labeling approaches in combination with mass spectrometry are introduced to quantitate relative protein levels (see Chapter 2). Although those methods are specifically directed toward quantitation of relative abundance of proteins expressed in cells or tissues, they can also be applied to describe the composition, connection, and dynamics of the multiprotein complexes and/or subcellular structure in a way similar to that described for the ICAT method. One example of *in vitro* labeling is the method using isotope-labeled *O*-methylisourea [$\text{H}_2^{15}\text{N}^{13}\text{C}(\text{OCH}_3)^5\text{NH}$] and unlabeled reagent, which allows quantitative guanidination of the N-terminus of the peptide and the epsilon amino group of lysine residues (see Chapter 2 for details). This reagent modifies all peptides generated by trypsin or Lys-C protease digestion; therefore, the peptide mixture generated by this method is very complex and requires a separation technique for higher resolution. However, the chance for quantification of multiple peptides obtained from the same protein is higher than that using ICAT reagent, which labels only cysteine-containing peptide. This guanidination method was applied to the comparative analysis of the preribosomal ribonucleoprotein (pre-rRNP) complexes associated with three typical trans-acting factors—nucleolin, fibrillarin and B23—which function at different stages of ribosome biogenesis (see Section 2-2-2, Experimental Example 2-3, and Section 3-3-1).

The most impressive work done in dynamic proteomics using MS-based organelle proteomics and *in vivo* stable isotope labeling, called amino acids in cell culture (SILAC), is the dynamic analysis of the entire nucleolus obtained from human HeLa cells (see Section 3-1-1) (84, 130). This study demonstrates the power of the quantitative approach using isotope labeling and LC-MS for the high throughput characterization of the flux of endogenous proteins through even entire subcellular structures or organelles (84). So far we have a number of methodologies on hand with high throughput enough to handle the dynamic nature of not only large cellular machineries (protein complexes) but also of an entire subcellular structure or organelle, whose protein compositions vary extensively under different environments for growth and metabolic conditions. The development of dynamic proteomics

coupled with other strategies heralds a new generation of “proteomic biology” that correlates dynamic proteome changes with cell function and thus enables us to understand biological aspects of living cells from the point of view of proteome dynamics.

REFERENCES

1. Wilkins, M. R. (1995). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotech. Gen. Eng. Rev.* **13**:19–50.
2. Neverova, I., and Van Eyk, J. E. (2005). Role of chromatographic techniques in proteomic analysis. *J. Chromatogr. B* **815**:51–63.
3. Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. *Nature* **405**:837–846.
4. Patterson, S. D., and Aebersold, R. H. (2003). Proteomics: the first decade and beyond. *Nat. Genet. Suppl.* **33**:311–323.
5. Anderson, N. L., Tracy, R. P., and Anderson, N. G. (1984). High-resolution two-dimensional electrophoretic mapping of plasma proteins. In *The Plasma Proteins IV*, F. W. Putnam (Ed.), Academic Press, New York, pp. 221–270.
6. Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., and Patton, W. F. (2000). A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21**:3673–3683.
7. Patton, W. F. (2000). A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* **21**:1123–1144.
8. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**:850–858.
9. Berggren, K., Chernokalskaya, E., Steinberg, T. H., Kemper, C., Lopez, M. F., Diwu, Z., Haugland, R. P., and Patton, W. F. (2000). Background-free, high sensitivity staining of proteins in one- and two-dimensional sodium dodecyl sulfatepolyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis* **21**:2509–2521.
10. Steen, H., and Pandey, A. (2002). Proteomics goes quantitative: measuring protein abundance. *Trends Biotechnol.* **20**:361–364.
11. Patton, W. F. (2002). Detection technologies in proteome analysis. *J. Chromatogr. B* **771**:3–31.
12. Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J. W., Hu, N., Taylor, P. R., Emmert-Buck, M. R., Liotta, L. A., Petricoin, E. F. 3rd, and Zhao, Y. (2002). 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol. Cell. Proteomics* **1**:117–124.
13. Anderson, N. L., Hofmann, J. P., Gemmell, A., and Tayler, J. (1984). Global approaches to quantitative analysis of gene-expression patterns observed by use of two-dimensional gel electrophoresis. *Clin. Chem.* **30**:2031–2036.
14. Tarroux, P., Vincens, P., and Rabilloud, T. (1987). HERMes: a second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part V: Data analysis. *Electrophoresis* **8**:187–199.

15. Tanaka, K., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1987). Detection of high mass molecules by laser desorption time-of-flight mass spectrometry. In *Proceedings of the 2nd Japan–China Joint Symposium on Mass Spectrometry*, H. Matsuda, and L. Xiao-tian (Eds.), Osaka, Japan, pp. 185–188.
16. Fenn, J. B., Mann, M., Meng, C. K., Wang, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large molecules. *Science* **246**:64–71.
17. Mann, M., Hojrup, P., and Roepstorff, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* **22**:338–345.
18. James, P., Quadroni, M., Carafoli, E., and Gonnet, G. (1993). Protein identification by mass profile fingerprinting. *Biochem. Biophys. Res. Commun.* **195**:58–64.
19. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**:327–332.
20. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* **90**:5011–5015.
21. Rosenfeld, J., Capdevielle, J., Guillemot, J. C., and Ferrara, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Anal. Biochem.* **203**:173–179.
22. Jenö, P., Mini, T., Hintermann, E., and Horst, M. (1995). Internal sequences from proteins digested in polyacrylamide gels. *Anal. Biochem.* **224**:451–455.
23. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**: 850–858.
24. Andersen, J. S., Svendsen, B., and Roepstorff, P. (1996). Electrospray ionization and matrix assisted laser desorption/ionization mass spectrometry: powerful analytical tools in recombinant protein chemistry. *Nat. Biotechnol.* **14**:449–457.
25. Qin, J., and Chait, B. T. (1997). Identification and characterization of posttranslational modifications of proteins by MALDI ion trap mass spectrometry. *Anal. Chem.* **69**:4002–4009.
26. Zabrouskov, V., Giacomelli, L., van Wijk, K. J., and McLafferty, F. W. (2003). A new approach for plant proteomics: characterization of chloroplast proteins of *Arabidopsis thaliana* by top down mass spectrometry. *Mol. Cell. Proteomics* **2**(12):1253–1260.
27. Mann, M., and Wilm, M. (1994). Error-tolerant identification of peptides in sequence database by peptide sequence tags. *Anal. Chem.* **66**:4390–4399.
28. Eng, J. K., McCormack, A. L., Yates, I., and John, R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database (Sequest). *J. Am. Soc. Mass Spectrom.* **5**:976–989.
29. Field, H. I., Fenyo, D., and Beavis, R. C. (2002). RADARS, a bioinformatics solution that automates proteome mass spectral analysis, optimises protein identification, and archives data in a relational database (Sonar). *Proteomics* **2**:36–47.
30. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence database using mass spectrometry data (Mascot). *Electrophoresis* **20**:3551–3567.
31. Ducret, A., Van Oostveen, I., Eng, J. K., Yates, J. R. 3rd, and Aebersold, R. (1998). High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci.* **7**(3):706–719.

32. MacCoss, M. J. (2005). Computational analysis of shotgun proteomics data. *Curr. Opin. Chem. Biol.* **9**:88–94.
33. Tabb, D. L., Eng, J. K., and Yates, J. R. 3rd (2001). In *Proteome Research: Mass Spectrometry*, P. James, (Ed.), Springer, New York, Vol. 1, pp. 125–142.
34. Tabb, D. L., McDonald, W. H., and Yates, J. R. 3rd (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**(1):21–26.
35. Von Haller, P. D., Yi, E., Donohoe, S., Vaughn, K., Keller, A., Nesvizhskii, A. I., Eng, J., Li, X. J., Goodlett, D. R., Aebersold, R., and Watts, J. D. (2003). The application of new software tools to quantitative protein profiling via isotope-coded affinity tag (ICAT) and tandem mass spectrometry: II. Evaluation of tandem mass spectrometry methodologies for large-scale protein analysis, and the application of statistical tools for data analysis and interpretation. *Mol. Cell. Proteomics.* **2**:428–442.
36. Shinkawa, T., Taoka, M., Yamauchi, Y., Ichimura, T., Kaji, H., Takahashi, N., and Isobe, T. (2005). STEM: a software tool for large-scale proteomic data analyses. *J. Proteome Res.* **4**(5):1826–1831.
37. Yang, X., Dondeti, V., Dezube, R., Maynard, D. M., Geer, L. Y., Epstein, J., Chen, X., Markey, S. P., and Kowalak, J. A. (2004). DBParser: Web-based software for shotgun proteomic data analyses. *J. Proteome Res.* **3**:1002–1008.
38. Evans, G., Wheeler, C. H., Corbett, J. M., and Dunn, M. J. (1997). Construction of HSC-2D PAGE: a two-dimensional gel electrophoresis database of heart proteins. *Electrophoresis* **18**:471–479.
39. Li, X. P., Pleissner, K. P., Scheler, C., Regitz-Zagrosek, V., Salnikow, J., and Jungblut, P. R. (1999). A two-dimensional gel electrophoresis database of rat heart protein. *Electrophoresis* **20**:891–897.
40. Pieper, R., Gatlin, C. L., Makusky, A. J., Russo, P. S., Schatz, C. R., Miller, S. S., Su, Q., McGrath, A. M., Estock, M. A., Parmar, P. P., Zhao, M., Huang, S.-T., Zhou, J., Wang, F., Esquer-Blasco, R., Anderson, N. L., Taylor, J., and Steiner, S. (2003). The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. *Proteomics* **3**:1345–1364.
41. Thongboonkerd, V., Mcleish, K. R., Arthur, J. M., and Klein, J. (2002). Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *Kidney Int.*, **62**:1461–1469.
42. Abbott, A. (2003). Brain protein project enlists mice in “dry run,” *Nature* **425**:110.
43. Heinke, M. Y., Wheeler, C. H., Chang, D., Einstein, R., Drake-Holland, A., Dunn, M. J., and Remedios, C. G. (1998). Protein changes observed in pacing-induced heart failure using two-dimensional electrophoresis. *Electrophoresis* **19**:2021–2030.
44. Van Eyk, J. E. (2001). Proteomics: unraveling the complexity of heart disease and striving to change cardiology. *Curr. Opin. Mol. Therapeut.* **3**:546–553.
45. Nelson, P. S., Han, D., Rochon, Y., Corthals, G. L., Lin, B., Monson, A., Nguyen, V., Franza, B. R., Plymate, S. R., Aebersold, R., and Hood, L. (2000). Comprehensive analysis of prostate gene expression: convergence of expressed sequence tag databases, transcript profiling and proteomics. *Electrophoresis* **21**:1823–1831.
46. Hanash, S. M., Madoz-Gurpide, J., and Misek, D. E. (2002). Identification of novel targets for cancer therapy using expression proteomics. *Leukemia* **16**:478–485.

47. Petricoin, E. F., Zoon, K. C., Kohn, E. C., Barrett, J. C., and Liotta, L. A. (2002). Clinical proteomics: translating bench-side promise into bedside reality. *Nat. Rev. Drug Discov.* **1**:683–695.
48. van Der Velden, J., Klein, L. J., Zaremba, R., Boontje, N. M., Huybregts, M. A., Stooker, W., Eijnsman, L., de Jong, J. W., Visser, C. A., Visser, F. C., and Stienen, G. J. (2001). Effects of calcium, inorganic phosphate, and pH on isometric force in single skinned cardiomyocytes from donor and failing human hearts. *Circulation* **104**:1140–1146.
49. Oh, P., Li, Y., Yu, J., Durr, E., Krasinska, K. M., Carver, L. A., Testa, J. E., and Schnitzer, J. E. (2004). Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. *Nature* **429**:629–635.
50. McDonough, J. L., Neverova, I., and Van Eyk, J. E. (2002). Proteomic analysis of human biopsy samples by single two-dimensional electrophoresis: Coomassie, silver, mass spectrometry, and Western blotting. *Proteomics* **2**:978–987.
51. Hanash, S. (2003). Disease proteomics. *Nature* **422**:226–232.
52. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**:1720–1730.
53. Putnam, F. W. (1984). Progress in plasma proteins. In *The Plasma Proteins*, Vol. IV, F. W. Putnam (Ed.), Academic Press, Orlando, FL, pp. 1–44.
54. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc. Natl. Acad. Sci. USA* **17**:9390–9395.
55. Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O’Shea, E. K., and Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature* **425**:737–741.
56. Andrews, B., Bader, G. D., and Boone, C. (2003). Playing tag with the yeast proteome. *Nat. Biotechnol.* **21**:1297–1299.
57. Gerlt, J. A. (2002). Fishing for the functional proteome. *Nat. Biotechnol.* **20**:786–787.
58. Houry, W. A., Frishman, D., Eckerskorn, C., Lottspeich, F., and Hartl, F. U. (1999). Identification of in vivo substrates of the chaperonin GroEL. *Nature* **402**:147–154.
59. Fountoulakis, M., and Langen, H. (1997). Identification of proteins by matrix-assisted laser desorption ionization-mass spectrometry following in-gel digestion in low-salt, nonvolatile buffer and simplified peptide recovery. *Anal. Biochem.* **250**:153–156.
60. Gygi, S. P., and Aebersold, R. (2000). Using mass spectrometry for quantitative proteomics. In *Proteomics: A Trends Guide*, Elsevier Science, London, pp. 31–36.
61. Moseley, M. A. (2001). Current trends in differential expression proteomics: Isotopically coded tags. *Trends Biotechnol.* **19**:510–516.
62. Kirkpatrick, D. S., Denison, C., and Gygi, S. P. (2005). Weighing in on ubiquitin: the expanding role of mass spectrometry-based proteomics. *Nat. Cell Biol.* **7**(8):750–757.
63. Takahashi, N., Kaji, H., Yanagida, M., Hayano, T., and Isobe, T. (2003). Proteomics: advanced technology for the analysis of cellular function. *J. Nutrition* **133**:2090–2096.
64. Adam, G. C., Sorensen, E. J., and Cravatt, B. F. (2003). Chemical strategies for functional proteomics *Mol. Cell. Proteomics* **1**(10):781–790.
65. Kobe, B., and Kemp, B. E. (1999). Active site-directed protein regulation. *Nature* **402**:373–376.
66. Blackstock, W. (2000). Trends in automation and mass spectrometry for proteomics. In *Proteomics: A Trends Guide*, Elsevier Science, London, pp. 12–16.

67. Blackstock, W. P., and Weir, M. P. (1999). Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* **3**:121–127.
68. Santoni, V., Molloy, M., and Rabilloud, T. (2000). Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **6**:1054–1070.
69. Mann, M., and Jensen, O. N. (2003). Proteomic analysis of post-translational modification. *Nat. Biotechnol.* **21**:255–261.
70. Ptacek, J., Devgan, G., Michaud, G., Zhu, H., Zhu, X., Fasolo, J., Guo, H., Jona, G., Breitkreutz, A., Sopko, R., McCartney, R. R., Schmidt, M. C., Rachidi, N., Lee, S. J., Mah, A. S., Meng, L., Stark, M. J., Stern, D. F., De Virgilio, C., Tyers, M., Andrews, B., Gerstein, M., Schweitzer, B., Predki, P. F., and Snyder, M. (2005). Global analysis of protein phosphorylation in yeast. *Nature* **438**(7068):679–684.
71. Adam, G. C., Sorensen, E. J., and Cravatt, B. F. (2002). Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat. Biotechnol.* **20**:805–809.
72. Adam, G. C., Sorensen, E. J., and Cravatt, B. F. (2002). Trifunctional chemical probes for the consolidated detection and identification of enzyme activities from complex proteomes. *Mol. Cell. Proteomics.* **1**(10):828–835.
73. Gerlt, J. A. (2002). Fishing for the functional proteome. *Nat. Biotechnol.* **20**:786–787.
74. Taylor, S. W., Fahy, E., and Ghosh, S. S. (2003). Global organellar proteomics. *Trends Biotechnol.* **21**:82–88.
75. Dreger, M. (2003). Subcellular proteomics. *Mass Spectrom. Rev.* **22**:27–56.
76. Dreger, M. (2003). Proteome analysis at the level of subcellular structures. *Eur. J. Biochem.* **270**:589–599.
77. Han, J. D., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J., Cusick, M. E., Roth, F. P., and Vidal, M. (2004). Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430**(6995):88–93.
78. Han, J. D., Dupuy, D., Bertin, N., Cusick, M. E., and Vidal, M. (2005). Effect of sampling on topology predictions of protein–protein interaction networks. *Nat. Biotechnol.* **23**(7):839–844.
79. Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., and Vidal, M. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* **303**(5657):540–543.
80. Kumar, A., and Snyder, M. (2002). Protein complexes take the bait. *Nature* **415**:123–124.
81. Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajnovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., Rudi, T., Gnau, V., Bauch, A., Bastuck, S., Huhse, B., Leutwein, C., Heurtier, M.-A., Copley, R. R., Edlmann, A., Querfurth, E., Rybin, V., Drewes, G., Raida, M., Bouwmeester, T., Bork, P., Seraphin, B., Kuster, B., Neubauer, G., and Superti-Furga, G. (2002). Functional

- organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**:141–147.
82. Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edlmann, A., Heurtier, M., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, T., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**(30): 631–636.
 83. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Soensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W. V., Figeys, D., and Tyers, M. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**: 180–183.
 84. Andersen, J. S., Lam, Y. W., Leung, A. K., Ong, S. E., Lyon, C. E., Lamond, A. I., and Mann, M. (2005). Nucleolar proteome dynamics. *Nature* **433**(7021):77–83.
 85. Patton, W. F. (1999). Proteome analysis. II. Protein subcellular redistribution: linking physiology to genomics via the proteome and separation technologies involved. *J. Chromatogr. B Biomed. Sci. Appl.* **1**(2):203–223.
 86. Andersen, J. S., and Mann, M. (2000). Functional genomics by mass spectrometry. *FEBS Lett.* **1**:25–31.
 87. Mann, M., Hendrickson, R. C., and Pandey, A. (2001). Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**:437–473.
 88. Godovac-Zimmermann, J., and Brown, L. R. (2001). Perspectives for mass spectrometry and functional proteomics. *Mass Spectrom. Rev.* **1**:1–57.
 89. Gudepu, R. G., and Wold, F. (1998). In *Proteins: Analysis and Design*, R. H. Angeletti (Ed.), Academic Press, San Diego, CA, pp. 121–207.
 90. Jensen, O. N. (2004). Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **8**(1):33–41.
 91. Ahn, N. G., and Resing, K. A. (2001). Toward the phosphoproteome. *Nat. Biotechnol.* **19**(4):317–318.
 92. Zhou, H., Watts, J., and Aebersold, R. (2001). A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.* **19**:375–378.
 93. Oda, Y., Nagasu, T., and Chait, B. T. (2001). Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**:379–382.
 94. van Swieten, P. F., Maehr, R., van den Nieuwendijk, A. M., Kessler, B. M., Reich, M., Wong, C. S., Kalbacher, H., Leeuwenburgh, M. A., Driessen, C., van der Marel, G. A., Ploegh, H. L., and Overkleeft, H. S. (2004). Development of an isotope-coded activity-based probe for the quantitative profiling of cysteine proteases. *Bioorg. Med. Chem. Lett.* **14**(12):3131–3134.
 95. Hemelaar, J., Galardy, P. J., Borodovsky, A., Kessler, B. M., Ploegh, H. L., and Ovaa, H. (2004). Chemistry-based functional proteomics: mechanism-based activity-profiling tools for ubiquitin and ubiquitin-like specific proteases. *J. Proteome Res.* **3**(2):268–276.

96. Barglow, K. T., and Cravatt, B. F. (2004). Discovering disease-associated enzymes by proteome reactivity profiling. *Chem. Biol.* **11**(11):1523–1531.
97. Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M., and Cravatt, B. F. (2004). Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Natl. Acad. Sci. USA* **101**(27):10000–10005.
98. Hekmat, O., Kim, Y. W., Williams, S. J., He, S., and Withers, S. G. (2005). Active-site peptide “fingerprinting” of glycosidases in complex mixtures by mass spectrometry. Discovery of a novel retaining beta-1,4-glycanase in *Cellulomonas fimi*. *J. Biol. Chem.* **280**(42):35126–35135.
99. Winssinger, N., Damoiseaux, R., Tully, D. C., Geierstanger, B. H., Burdick, K., and Harris, J. L. (2004). PNA-encoded protease substrate microarrays. *Chem. Biol.* **11**(10):1351-1360. Comment in: *Chem. Biol.* **11**(10):1328–1330.
100. Bryant, N. J., Govers, R., and James, D. E. (2002). Regulated transport of the glucose transporter GLUT4. *Nat. Rev. Mol. Cell Biol.* **3**:267–277.
101. Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**(6966):570–574.
102. Foster, L. J., de Hoog, C. L., Zhang, Y., Zhang, Y., Xie, X., Mootha, V. K., and Mann, M. (2006). A mammalian organelle map by protein correlation profiling. *Cell* **125**(1):187–199.
103. Donnes, P., and Høglund, A. (2004). Predicting protein subcellular localization: past, present, and future. *Geno. Prot. Bioinfo.* **2**(4):209–215.
104. Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K. L., Lam, W. W., Steen, H., Mann, M., and Lamond, A. I. (2002). Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**:1–11.
105. Scher, A., Coute, Y., De’on, C., Callé, A., Kindbeiter, K., Sanchez, J.-C., Greco, A., Hochstrasser, D., and Diaz, J.-J. (2002). Functional proteomic analysis of human nucleolus. *Mol. Biol. Cell* **13**:4100–4109.
106. Bell, A. W., Ward, M. A., Blackstock, W. P., Freeman, H. N., Choudhary, J. S., Lewis, A. P., Chotai, D., Fazel, A., Gushue, J. N., Paiement, J., Palcy, S., Chevet, E., Lafreniere-Roula, M., Solari, R., Thomas, D. Y., Rowley, A., and Bergeron, J. J. (2001). Proteomics characterization of abundant Golgi membrane proteins. *J. Biol. Chem.* **276**:5152–5165.
107. Dreger, M., Bengtsson, L., Schneberg, T., Otto, H., and Hucho, F. (2001). Nuclear envelope proteomics: novel integral membrane proteins of the inner nuclear membrane. *Proc. Natl. Acad. Sci. USA* **98**:11943–11948.
108. Neubauer, G., King, A., Rappilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* **20**: 46–50.
109. Rout, M. P., Aitchison, J. D., Suprpto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000). The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**:635–651.
110. Peters, J.-M., King, R. W., Hoog, C., and Kirschner, M. W. (1996). Identification of BIME as a subunit of the anaphase-promoting complex. *Science* **274**:1199–1201.
111. Zachariae, W., Shin, T. H., Galova, M., Obermaier, B., and Nasmyth, K. (1996). Identification of subunits of the anaphase-promoting complex of *Saccharomyces cerevisiae*. *Science* **274**:1201–1204.

112. Kumar, A., Agarwal, S., Heyman, J. A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., Cheung, K. H., Miller, P., Gerstein, M., Roeder, G. S., and Snyder, M. (2002). Subcellular localization of the yeast proteome. *Genes Dev.* **16**(6):707–719.
113. Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O’Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature* **425**(6959):686–691.
114. Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R., and Wiemann, S. (2000). Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Reports* **1**(3):287–292.
115. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, M. J. (2002). Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* **158**(5):915–927.
116. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R. (1999). Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**:676–682.
117. Zhou, Z., Licklider, L. J., Gygi, S. P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**:182–185.
118. Hartmuth, K., Urlaub, H., Vornlocher, H.-P., Will, C. L., Gentzel, M., Wilm, M., and Lührmann, R. L. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. USA* **99**(26):16719–16724.
119. Fatica, A., and Tollervey, D. (2002). Making ribosomes. *Curr. Opin. Cell Biol.* **14**:313–318.
120. Fromont-Racine, M., Senger, B., Saveanu, C., and Fasiolo, F. (2003). Ribosome assembly in eukaryotes. *Gene* **313**:17–42.
121. Takahashi, N., Yanagida, M., Fujiyama, S., Hayano, T., and Isobe, T. (2003). Proteomic snapshot analysis of preribosomal ribonucleoprotein complexes formed at various stages of ribosome biogenesis in yeast and mammalian cells. *Mass Spectrom. Rev.* **22**:287–317.
122. Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadien, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., Onge, P. S., Ghanny, S., Lam, M. H. Y., Butland, G., Altaf-Ul, A. M., Kanaya, K. S., Shilatifard, A., O’Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006). Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**(30):637–643.
123. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17**:1030–1032.
124. Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999). From molecular to modular cell biology. *Nature* **402**:C47–C52.

125. Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**(5601):2205–2208.
126. Ospina, J. K., and Matera, A. G. (2002). Proteomics: the nucleolus weighs in dispatch. *Curr. Biol.* **12**:R29–R31.
127. Fox, A. H., Lam, Y. W., Leung, A., Lyon, C. E., Andersen, J. S., Mann, M., and Lamond, A. I. (2002). Paraspeckles: a novel nuclear domain. *Curr. Biol.* **12**:13–25.
128. Ranish, J. A., Yi, E. C., Leslie, D. M., Purvine, S. O., Goodlett, D. R., Eng, J., and Aebersold, R. (2003). The study of macromolecular complexes by quantitative proteomics. *Nat. Genet.* **33**:349–355.
129. Ranish, J. A., Yudkovsky, N., and Hahn, S. (1999). Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. *Genes Dev.* **13**:49–63.
130. Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**:198–207.
131. Salih, E., Ashkar, S., Gerstenfeld, L. C., and Glimcher, M. J. (1997). Identification of the phosphorylated sites of metabolically ³²P-labeled osteopontin from cultured chicken osteoblasts. *J. Biol. Chem.* **272**:13966–13973.
132. Salih, E. (2003). In vivo and in vitro phosphorylation regions of bone sialoprotein. *Connect. Tissue Res.* **44**:223–229.
133. Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., Ohara, O., Mann, M., Jensen, O. N., and Pandey, A. (2002). A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol. Cell. Proteomics* **1**(7):517–527.
134. Yanagida, M., Miura, Y., Yagasaki, K., Taoka, M., Isobe, T., and Takahashi, N. (2000). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry analysis of proteins detected by anti-phosphotyrosine antibody on two-dimensional gels of fibroblast cell lysates after tumor necrosis factor- α stimulation. *Electrophoresis* **21**:1890–1898.
135. Steen, H., Fernandez, M., Ghaffari, S., Pandey, A., and Mann, M. (2003). Phosphotyrosine mapping in Bcr/Abl oncoprotein using phosphotyrosine-specific Immonium ion scanning. *Mol. Cell. Proteomics* **2**(3):138–145.
136. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **96**:6591–6596.
137. Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., Gibson, B. W., and Shokat, K. M. (2003). Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat. Biotechnol.* **21**:1047–1054.
138. Elia, A. E. H., Cantley, L. C., and Yaffe, M. B. (2003). Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science* **299**:1228–1231.
139. Borchers, C. H., Thapar, R., Petrotchenko, E. V., Torres, M. P., Speir, J. P., Easterling, M., Dominski, Z., and Marzluft, W. F. (2006). Combined top-down and bottom-up proteomics identifies a phosphorylation site in stem-loop-binding proteins that contributes to high-affinity RNA binding. *Proc. Natl. Acad. Sci. USA* **103**:3094–3099.

140. Meng, F., Du, Y., Miller, L. M., Patrie, S. M., Robinson, D. E., and Kelleher, N. L. (2004). Molecular-level description of proteins from *Saccharomyces cerevisiae* using quadrupole FT hybrid mass spectrometry for top-down proteomics. *Anal. Chem.* **76**(10):2852–2858.
141. Ibarrola, N., Kalume, D. E., Gronborg, M., Iwahori, A., and Pandey, A. (2003). A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. *Anal. Chem.* **75**(22):6043–6049.
142. Ibarrola, N., Molina, H., Iwahori, A., and Pandey, A. (2004). A novel proteomic approach for specific identification of tyrosine kinase substrates using [¹³C]tyrosine. *J. Biol. Chem.* **279**(16):15805–15813.
143. MacDonald, J. A., Mackey, A. J., Pearson, W. R., and Haystead, T. A. J. (2002). A strategy for the rapid identification of phosphorylation sites in the phosphoproteome. *Mol. Cell. Proteomics* **1**:314–322.
144. Ballif, B. A., Roux, P. P., Gerber, S. A., MacKeigan, J. P., Blenis, J., and Gygi, S. P. (2005). Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proc. Natl. Acad. Sci. USA* **102**(3):667–672.
145. Dierck, K., Machida, K., Voigt, A., Thimm, J., Horstmann, M., Fiedler, W., Mayer, B. J., and Nollau, P. (2006). Quantitative multiplexed profiling of cellular signaling networks using phosphotyrosine-specific DNA tagged SH2 domains. *Nat. Methods* **3**(9):737–744.
146. Kwon, S. W., Kim, S. C., Jaunbergs, J., Falck, J. R., and Zhao, Y. (2003). Selective enrichment of thiophosphorylated polypeptides as a tool for the analysis of protein phosphorylation. *Mol. Cell. Proteomics* **2**(4): 242–247.
147. Moser, K., and White, F. M. (2006). Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS. *J. Proteome Res.* **5**(1):98–104.
148. Aprilita, N. H., Huck, C. W., Bakry, R., Feuerstein, I., Stecher, G., Morandell, S., Huang, H. L., Stasyk, T., Huber, L. A., and Bonn, G. K. (2005). Poly(glycidyl methacrylate/divinylbenzene)-IDA-FeIII in phosphoproteomics. *J. Proteome Res.* **4**(6):2312–2319.
149. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., and Jorgensen, T. J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics.* **4**(7):873–886.
150. Brill, L. M., Salomon, A. R., Ficarro, S. B., Mukher, J. I. M., Stettler-Gill, M., and Peters, E. C. (2004). Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.* **76**(10):2763–2772.
151. Cantin, G. T., Venable, J. D., Cociorva, D., and Yates, J. R. 3rd. (2006). Quantitative phosphoproteomic analysis of the tumor necrosis factor pathway. *J. Proteome Res.* **5**(1):127–134.
152. Kaji, H., Saito, H., Yamauchi, Y., Sinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., and Isobe, T. (2003). Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoprotein. *Nat. Biotechnol.* **21**(6):667–672.
153. Zhang, H., Li, X. J., Martin, D. B., and Aebersold, R. (2003). Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**(6):660–666.

154. Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002). Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell. Proteomics* **1**(10): 791–804.
155. Cieniewski-Bernard, C., Bastide, B., Lefebvre, T., Lemoine, J., Mounier, Y., and Michalski, J. C. (2004). Identification of O-linked *N*-acetylglucosamine proteins in rat skeletal muscle using two-dimensional gel electrophoresis and mass spectrometry. *Mol. Cell. Proteomics* **3**(6):577–585.
156. Boisvert, F. M., Cote, J., Boulanger, M. C., and Richard, S. (2003). A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics* **2**(12):1319–1330.
157. Wu, C. C., MacCoss, M. J., Mardones, G., Finnigan, C., Mogelsvang, S., Yates, J. R. 3rd, and Howell, K. E. (2004). Organellar proteomics reveals Golgi arginine dimethylation. *Mol. Biol. Cell* **15**(6):2907–2919.
158. Ong, S. E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl SILAC. *Nat. Methods* **1**(2):119–126.
159. Matsumoto, M., Hatakeyama, S., Oyamada, K., Oda, Y., Nishimura, T., and Nakayama, K. (2005). Large-scale analysis of the human ubiquitin-related proteome. *Proteomics* **5**(16):4145–4151.
160. Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S. P. (2003). A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.* **21**(8):921–926.
161. Gocke, C., Yu, H., and Kang, J. (2004). Systematic identification and analysis of mammalian SUMO substrates. *J. Biol. Chem.* **280**(6):5004–5012.
162. Rosas-Acosta, G., Russell, W. K., Deyrieux, A., Russell, D. H., and Wilson, V. G. (2005). A universal strategy for proteomic studies of SUMO and other ubiquitin-like modifiers. *Mol. Cell. Proteomics* **4**:56–72.
163. Panse, V. G., Hardeland, U., Werner, T., Kuster, B., and Hurt, E. (2004). A proteome-wide approach identifies sumoylated substrate proteins in yeast. *J. Biol. Chem.* **279**:41346–41351.
164. Li, X. J., Pedrioli, P. G., Eng, J., Martin, D., Yi, E. C., Lee, H., and Aebersold, R. (2004). A tool to visualize and evaluate data obtained by liquid chromatography–electrospray ionization–mass spectrometry. *Anal. Chem.* **76**:3856–3860.
165. Vertegaal, A. C., Andersen, J. S., Ogg, S. C., Hay, R. T., Mann, M., and Lamond, A. I. (2006). Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol. Cell. Proteomics* **5**(12):2298–2310.
166. Vertegaal, A. C., Ogg, S. C., Jaffray, E., Rodriguez, M. S., Hay, R. T., Andersen, J. S., Mann, M., and Lamond, A. I. (2004). A proteomic study of SUMO-2 target proteins. *J. Biol. Chem.* **279**:33791–33798.
167. Denison, C., Rudner, A. D., Gerber, S. A., Bakalarski, C. E., Moazed, D., and Gygi, S. P. (2005). A proteomic strategy for gaining insights into protein sumoylation in yeast. *Mol. Cell. Proteomics* **4**(3):246–254.
168. Wohlschlege, J. A., Johnson, E. S., Reed, S. I., and Yates, J. R. 3rd. (2004). Global analysis of protein sumoylation in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **279**: 45662–45668.
169. Hao, G., Derakhshan, B., Shi, L., Campagne, F., and Gross, S. G. (2006). SNOSID, a proteomic method for identification of cysteine S-nitrosylation sites in complex protein mixtures. *Proc. Natl. Acad. Sci. USA* **103**(4):1012–1017.

170. Kunczewicz, T., Sheta, E. A., Goldknopf, I. L., and Kone, B. C. (2003). Proteomic analysis of S-nitrosylated proteins in mesangial cells. *Mol. Cell. Proteomics* **2**(3):156–163.
171. Castegna, A., Thongboonkerd, V., Klein, J. B., Lynn, B., Markesbery, W. R., and Butterfield, D. A. (2003). Proteomic identification of nitrated proteins in Alzheimer's disease brain. *J. Neurochem.* **85**:1394–1401.
172. Ballesteros, M., Fredriksson, A., Henriksson, J., and Nystrom, T. (2001). Bacterial senescence: protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosome. *EMBO J.* **20**:5280–5289.
173. Poon, H. F., Castegna, A., Farr, S. A., Thongboonkerd, V., Lynn, B. C., Banks, W. A., Morley, J. E., Klein, J. B., and Butterfield, D. A. (2004). Quantitative proteomics analysis of specific protein expression and oxidative modification in aged senescence-accelerated-prone mice brain. *Neuroscience* **126**:915–926.
174. Soreghan, B. A., Yang, F., Thomas, S. N., Hsu, J., and Yang, A. J. (2003). High-throughput proteomic-based identification of oxidatively induced protein carbonylation in mouse brain. *Pharm. Res.* **20**(11):1713–1720.
175. Kho, Y., Kim, S. C., Jiang, C., Barma, D., Kwon, S. W., Cheng, J., Jaunbergs, J., Weinbaum, C., Tamanoi, F., Falck, J., and Zhao, Y. (2004). A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *Proc. Natl. Acad. Sci. USA* **101**(34):12479–12484.
176. Boisson, B., and Meinnel, T. (2003). A continuous assay of myristoyl-CoA:protein *N*-myristoyltransferase for proteomic analysis. *Anal. Biochem.* **322**(1):116–123.
177. Liu, Y., Patricelli, M. P., and Cravatt, B. F. (1999). Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. USA* **96**(26):14694–14699.
178. Kidd, D., Liu, Y., and Cravatt, B. F. (2001). Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40**(13):4005–4015.
179. Greenbaum, D., Medzihiradzky, K. F., Burlingame, A., and Bogyo, M. (2000). Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem. Biol.* **7**(8):569–581.
180. Kocks, C., Maehr, R., Overkleeft, H. S., Wang, E. W., Iyer, L. K., Lennon-Dumenil, A. M., Ploegh, H. L., and Kessler, B. M. (2003). Functional proteomics of the active cysteine protease content in *Drosophila* S2 cells. *Mol. Cell. Proteomics.* **2**(11):1188–1197.
181. Adam, G. C., Cravatt, B. F., and Sorensen, E. J. (2001). Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chem. Biol.* **8**(1):81–95.
182. Speers, A. E., Adam, G. C., and Cravatt, B. F. (2003). Activity-based protein profiling in vivo using a copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**(16):4686–4687.
183. Leung, D., Hardouin, C., Boger, D. L., and Cravatt, B. F. (2003). Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat. Biotechnol.* **21**(6):687–691.
184. Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Ploegh, H. L., and Kessler, B. M. (2002). Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**:1149–1159.
185. Ovaa, H., Kessler, B. M., Rolen, U., Galardy, P. J., Ploegh, H. L., and Masucci, M. G. (2004). Activity-based ubiquitin-specific protease (USP) profiling of virus-infected and malignant human cells. *Proc. Natl. Acad. Sci. USA* **101**(8):2253–2258.

186. Ovaa, H., Galaray, P. J., and Ploegh, H. L. (2005). Mechanism-based proteomics tools based on ubiquitin and ubiquitin-like proteins: synthesis of active site-directed probes. *Methods Enzymol.* **399**:468–478.
187. Galaray, P., Ploegh, H. L., and Ovaa, H. (2005). Mechanism-based proteomics tools based on ubiquitin and ubiquitin-like proteins: crystallography, activity profiling, and protease identification. *Methods Enzymol.* **399**:120–131.
188. Rolen, U., Kobzeva, V., Gasparjan, N., Ovaa, H., Winberg, G., Kisseljov, F., and Masucci, M. G. (2006). Activity profiling of deubiquitinating enzymes in cervical carcinoma biopsies and cell lines. *Mol. Carcinog.* **45**(4):260–269.
189. Bogoy, M., McMaster, J. S., Gaczynska, M., Tortorella, D., Goldberg, A. L., and Ploegh, H. (1997). Covalent modification of the active site threonine of proteasomal subunits and the *Escherichia coli* homolog HslV by a new class of inhibitors. *Proc. Natl. Acad. Sci. USA* **94**:6629–6634.
190. Chan, E. W., Chattopadhyaya, S., Panicker, R. C., Huang, X., and Yao, S. Q. (2004). Developing photoactive affinity probes for proteomic profiling: hydroxamate-based probes for metalloproteases. *J. Am. Chem. Soc.* **126**(44):14435–14446.
191. Lo, L. C., Chiang, Y. L., Kuo, C. H., Liao, H. K., Chen, Y. J., and Lin, J. J. (2005). Study of the preferred modification sites of the quinone methide intermediate resulting from the latent trapping device of the activity probes for hydrolases. *Biochem. Biophys. Res. Commun.* **326**(1):30–35.
192. Lo, L.-C., Pang, T.-L., Kuo, C.-H., Chiang, Y.-L., Wang, H.-Y., and Lin, J.-J. (2002). Design and synthesis of class-selective activity probes for protein tyrosine phosphatases. *J. Proteome Res.* **1**:35–40.
193. Kumar, S., Zhou, B., Liang, F., Wang, W.-Q., Huang, Z., and Zhang, Z.-Y. (2004). Activity-based probes for protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. USA* **101**:7943–7948.
194. Godl, K., Wissing, J., Kurtenbach, A., Habenberger, P., Blencke, S., Gutbrod, H., Salassidis, K., Stein-Gerlach, M., Missio, A., Cotten, M., and Daub, H. (2003). An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc. Natl. Acad. Sci. USA* **100**:15434–15439.
195. Tsai, C.-S., Li, Y.-K., and Lo, L.-C. (2002). Design and synthesis of activity probes for glycosidases. *Org. Lett.* **4**:3607–3610.
196. Williams, S. J., Hekmat, O., and Withers, S. G. (2006). Synthesis and testing of mechanism-based protein-profiling probes for retaining endo-glycosidases. *Chem. Biochem.* **7**(1):116–124.
197. Orru, S., Caputo, I., D'Amato, A., Ruoppolo, M., and Esposito, C. (2003). Proteomics identification of acyl-acceptor and acyl-donor substrates for transglutaminase in a human intestinal epithelial cell line: implication for celiac disease. *J. Biol. Chem.* **273**:31766–31773.
198. Ruoppolo, M., Orru, S., D'Amato, A., Francese, S., Rovero, P., Marino, G., and Esposito, C. (2003). Analysis of transglutaminase protein substrates by functional proteomics. *Protein Sci.* **12**:1290–1297.

