4.1 INTRODUCTION

The objective of some proteomics experiments is simply to catalog the proteins that are found in a given sample, and this form of analysis can be described as predominantly descriptive. The comparison of related samples may reveal proteins that are present in one sample but not in another, due to the samples representing different cell types, developmental stages, and cell states (for example, stage of the cell cycle or in response to changes in the environment) so there may be numerous proteins uniquely expressed under particular circumstances. There are also proteomic changes associated with disease; for example, a comparison of normal skin and squamous cell carcinoma might reveal a set of protein spots unique to the disease. Once identified, these proteins could be useful as disease markers and might even represent potential new therapeutic targets (Chapter 10). However, there are very few proteins that show such unambiguous on/off changes. More often, the difference between samples is one of degree. Therefore, the accurate quantitation of proteins is now a vital aspect of proteomics.

There are several well-established methods for the quantitation of individual proteins, either in solution or using a solid-phase assay, which are based on the use of labeled antibodies (Box 4.1). The adaptation of such assays for proteomic analysis is difficult because even if antibodies could be found to bind to every protein in the proteome, the signal intensity for each antigen–antibody interaction would depend not only on the abundance of the target protein but also on the strength of the antigen–antibody binding (that is, the affinity of the antibody). Despite these technical hurdles, some analytical protein microarrays have been manufactured that are arrayed with thousands of antibodies, and these are described in Chapter 9. Generally, the most successful microarrays contain a small number of well-characterized antibodies. The more complex the device, the greater the problems with sensitivity and specificity, and the resulting quantitative data are less reliable.

Large-scale protein quantitation in proteomics relies primarily on the use of general labeling or staining, or on the selective labeling or staining of particular classes of proteins. There are various methods for measuring the total amount of protein in a solution (Box 4.2). However, it is necessary to compare the abundances of thousands of proteins in parallel across multiple samples in typical proteomics experiments. The chosen strategy depends largely on how the protein samples are prepared and fractionated, and can be divided into two broad categories: those based on the analysis of two-dimensional gel images and those based on comparing the abundance of ions across samples by mass spectrometry.
CHAPTER 4: STRATEGIES FOR PROTEIN QUANTITATION

4.2 QUANTITATIVE PROTEOMICS BASED ON 2DGE

The quantitation of proteins in two-dimensional gels involves the creation of digital data from analog images. The abundance of different proteins on a two-dimensional gel is reflected by the shape, size, and intensity of the corresponding spots. Assuming that spots are well resolved, protein quantitation requires the conversion of an analog gel image into digital data, resulting in a catalog of individual spots listed as x, y positions, shape parameters, and quantitative values (integrated spot intensities). It is then possible to carry out objective comparisons of equivalent spots on different gels and thus to determine whether a particular protein is more or less abundant in one sample compared with another. It can be difficult to reproduce the exact conditions for protein separation in 2DGE, so the identification of corresponding spots even on two-dimensional gels
containing the same original sample can be a challenge. Robust methods are therefore required for the analysis of gels representing different samples if many of the spots differ in abundance and some spots are present on one gel and absent on another.

The first stage in protein quantitation is **image acquisition**, and the method used depends on how the proteins were labeled or stained. Radioactively labeled proteins can be detected on an X-ray film or by phosphorimaging. The X-ray film may then be scanned by a CCD camera or a densitometer, whereas phosphorimagers come with their own scanning devices. A **charge-coupled device (CCD)** is simply a solid-state electrical component that is divided into a series of light-sensitive areas or photosites composed of a material that emits electrons when struck by a photon of light. The image from a CCD camera is generated by a microprocessor that counts the electrons at each photosite. A **densitometer** is a scanning device that works on a similar principle, that is, light reflected from or transmitted through the surface of a film is detected by a photodiode, which thus records the density of the light and dark areas on the image. Coomassie-stained and silver-stained gels may also be scanned with a CCD camera or densitometer, whereas gels stained with the fluorescent reagents or gels containing fluorescently labeled proteins may be scanned using a CCD camera or a fluorescence imager.

The quality of the digital data depends critically on the resolution of the scanned image, which can be considered in terms of both spatial resolution (expressed as pixels per unit length or area) and **densitometric resolution** (that is, the range of gray values that can be interpreted). However, the densitometric resolution also depends on the labeling or staining method. **Silver staining** was the major non-radioactive detection method used for separated proteins because it is 10–100 times more sensitive than Coomassie Brilliant Blue. However, silver stains do not detect glycoproteins very efficiently and the most sensitive detection methods lead to chemical modification of cysteine residues, thereby interfering with downstream analysis by mass spectrometry (this reflects the use of formaldehyde for stain development, and its replacement has helped to increase the compatibility between silver staining and MS). In terms of comparative protein quantitation, the major disadvantage of silver staining is its narrow linear range (about one order of magnitude). This means that it is possible to accurately determine whether one protein is twice as abundant as another (or more importantly, if one protein is twice as abundant in one sample compared with another), but it is not possible to accurately compare protein abundance if there is a tenfold or greater difference. Fluorescent stains such as SYPRO Ruby, Deep Purple, and Flamingo are now strongly preferred as these are at least as sensitive as silver staining but share none of its disadvantages. That is, they detect glycoproteins efficiently, they do not cause any covalent protein modifications, and they have an extensive linear range (over three orders of magnitude), which means they can be used to compare protein abundances very effectively. The different ways for detecting proteins in two-dimensional gels are summarized in **Box 4.3**.

**Spot detection, quantitation, and comparison can be challenging without human intervention**

Spots on protein gels are not uniform in shape, size, or density. Some spots appear as discrete entities while others overlap to a greater or lesser degree. The edges of some spots are clearly defined while those of others may be blurred. Small spots may appear as shoulders on larger ones, or several spots may be joined together in a line. The densitometric landscape within different spots (that is, the distribution of gray values) is not always consistent. These variations may be compounded by nonspecific changes in the gel background.
Pre-labeling with organic fluorophores
A number of different organic molecules can be covalently attached to proteins prior to electrophoretic separation, allowing the direct detection and quantitation of labeled proteins within two-dimensional gels. Methods utilizing well-characterized fluorophores such as fluorescein and fluorescein isothiocyanate have been available since the 1970s, but in the context of proteomics these methods have a number of drawbacks, including the altered solubility and/or mobility of labeled proteins and the variable sensitivity of labeling depending on the number of functional groups available for modification. However, the use of two or more different fluorophores, for example propyl-Cy3 and methyl-Cy5, to label different protein samples allows the abundance of proteins in the samples to be compared on the same gel (difference in-gel electrophoresis, see p. 75).

Silver staining
Silver staining is one of the most popular techniques for staining proteins in polyacrylamide gels. There are many variations on the staining protocol but staining is generally carried out using the dye in a mixture of concentrated acid with ethanol or methanol. This produces a colloidal suspension that stains proteins strongly with low background. Depending on the exact make-up of the stain, the dye can also modify glutamic acid side chains, which can complicate the interpretation of mass spectrometry data (although adjustments to the search criteria can accommodate this).

Coomassie Brilliant Blue
Coomassie Brilliant Blue is an organic dye that is commonly used to stain proteins in polyacrylamide gels. There are many variations on the staining protocol but staining is generally carried out using the dye in a mixture of concentrated acid with ethanol or methanol. This produces a colloidal suspension that stains proteins strongly with low background. Although widely used for general protein analysis, Coomassie Brilliant Blue and related organic dyes lack the sensitivity for proteomic analysis, having a detection limit of 10–30 ng. Depending on the exact make-up of the stain, the dye can also modify glutamic acid side chains, which can complicate the interpretation of mass spectrometry data (although adjustments to the search criteria can accommodate this).

Fluorescent stains
A number of fluorophores are known to bind noncovalently to proteins, which makes them particularly compatible with downstream mass spectrometry analysis. These stains generally demonstrate little fluorescence in aqueous solution but fluoresce strongly when associated with SDS–protein complexes, and therefore produce a very low background in stained gels. The most versatile of these molecules include SYPRO Ruby (Invitrogen), Lucy (Sigma-Aldrich), Deep Purple (GE Healthcare), Krypton (Pierce), and Flamingo (Bio-Rad). These agents are very sensitive and show a broad linear dynamic range. SYPRO Ruby is one of the most widely used, and matches the sensitivity of the best silver staining techniques but has a superior linear dynamic range (extending over three orders of magnitude) and stains proteins that do not show up well with silver stains, for example many glycoproteins. The staining protocol is also simple and rapid, unlike Coomassie and silver staining techniques, which each require a lengthy de-staining step. The proprietary fluorescent dyes can be expensive but compounds with similar properties can be synthesized in the laboratory.

The human eye can generally tell the difference between a spot and background artifacts on a two-dimensional gel but humans are too subjective in their judgment to define spots rigorously. Machines can apply a fixed set of rules and parameters to the definition of individual spots and therefore interpret spot patterns more objectively. However, getting machines to see the spots in the same way that humans do can be challenging. Normally, the first stage in automated spot detection is digital image enhancement, which helps to clear the background and improve the contrast of the image to make the spot boundaries easier to delineate. Smoothing is used to eliminate variable background noise and the background is then subtracted from the rest of the image. The contrast in the subtracted image is enhanced by reassigning gray values from the mid-range to make the pixels either darker...
or lighter. In many cases, edge detection filters are used that aim to identify regions of the image in which there is a sharp change in pixel intensity. 

Once a processed image is available, a number of different algorithms can be applied to detect and quantitate individual spots. These must take all the possible variations in spot morphology into account and calculate the integrated spot intensities, which are essentially absolute values that represent protein abundances. The algorithms generally use either Gaussian fitting (which assumes the gray values in the spot have a normal distribution along both the x and y axes) or Laplacian of Gaussian (LOG) spot detection methods. Other algorithms are based on the watershed transformation method in which a grayscale image is converted into a topographic surface with darker sections representing peaks and lighter sections representing troughs. The idea is then to “flood” the image from the minima, which divides the image into catchment basins representing individual spots and watershed lines representing divisions (Figure 4.1a and color plates). In practice, the indiscriminate flooding of gel images in this manner leads to over-segmentation due to background variation in pixel intensity (Figure 4.1a and color plates). To avoid this outcome, flooding can be initiated from a previously defined set of markers, which avoids any over-segmentation (Figure 4.1b and color plates). Another useful method is line analysis in which the computer focuses on individual vertical scan lines to identify density peaks. The density peaks in adjacent scan lines can be assembled into chains and these represent the centers of spots.

Once the two-dimensional gel has been reduced to a series of digital data representing spot intensities, the comparison of different gels is a simple process of comparing data values and determining whether the abundance of a given protein differs significantly, according to some predefined threshold, among two or more samples. A prerequisite for this type of analysis is the identification of equivalent spots on different gels, which may be challenging because gel-running conditions cannot be reproduced exactly. This may be due to several factors:

- Differences in sample preparation.
- Differences in gel composition. This can be minimized by preparing several gels from the same mixture at the same time, or by using commercially available pre-cast gels.
- Variations in running conditions. As discussed in Chapter 2, this is a significant problem for isoelectric focusing (IEF) gels with carrier ampholytes, particularly non-equilibrium gels, but it also applies to a lesser degree to immobilized pH gradient gels. The problems can be addressed to some extent by running several gels in parallel, but this is not always possible.

FIGURE 4.1 The watershed method for contour finding on two-dimensional gel images. (a) Any grayscale image can be considered as a topographic surface. If flooded from its minima without allowing water from different sources to merge, the image is partitioned into catchment basins and watershed lines, but in practice this leads to over-segmentation. (b) Therefore, markers (bright red shapes) are used to initiate flooding, and this reduces over-segmentation considerably. See also color plates. (Adapted from images by Serge Beucher, CMM École Nationale Supérieure des Mines de Paris.)
CHAPTER 4: STRATEGIES FOR PROTEIN QUANTITATION

- Minor variations within each gel that lead to regional differences in protein mobility. Again this is a major problem with carrier ampholyte IEF gels, but also applies to others.

In the absence of gels, or images thereof, showing perfect spot-to-spot correspondence, it becomes necessary to force equivalent gels into register, a process known as gel matching. This process makes use of landmarks, that is, spots that are present on all gels in the comparison and can be used as a common frame of reference. Gel matching algorithms then apply image transformation procedures such as stretching, skewing, and rotating, at both local and global levels, to bring multiple gel images into register and make them comparable. This can be thought of as a procedure in which several equivalent gels are stacked above each other and a pin is used to pierce the center of the first landmark spot through all the stacked gels. Further pins are inserted through other landmarks. When the gels are held in position by a number of pins, flexible wires can be inserted to link equivalent spots that are not perfectly in register (Figure 4.2). In some gels, a given spot may be absent, but with a number of matched landmarks surrounding the space, the algorithm can assign a zero value to the spot with reasonable confidence. As an alternative to matching gels at the spot level, other algorithms perform essentially the same task at the pixel level. An extension to the use of landmarks is a gel matching method known as propagation. In this approach, the algorithm begins at a known landmark and then maps the nearby spots and returns a list of x, y displacement values. Other gels are scrutinized for spots at the same displacements relative to the landmark and matches are identified. These matches can then be used as new landmarks for recursive searching.

The end result of spot detection, quantitation, and gel matching should be a table of spot values (x, y coordinates, shape parameters, and integrated spot intensities) arranged as an $N \times M$ matrix where $N$ represents all the different spots that have been identified and $M$ represents all the gels (Figure 4.3). $M$ should be divided into groups based on the experimental conditions. For example, $M_1, ..., M_{15}$ might represent five control gels, five from...
experimental condition 1, and five from experimental condition 2 (perhaps different stages of a disease or different time points after drug administration). The quantitative values must be normalized for any differences in the overall signal intensities on the gels (for example, due to different exposure times) and then various statistical methods can be used to identify protein spots whose abundance varies over the experimental conditions. Recent developments in proteomic gel imaging technology allow matched gels to be overlain in false color so that protein spots with differential abundance over two or more gels can be visually identified. This is essentially an artificial method for generating difference gel electrophoresis data (see below) from samples separated on different gels.

### 4.3  MULTIPLEXED IN-GEL PROTEOMICS

Multiplexed proteomics in the context of 2DGE is the use of fluorescent stains or probes with different excitation and emission spectra to detect different groups of proteins simultaneously on the same gel. This helps to reduce the number of duplicate gels that are required to compare different proteins and, at least in theory, obviates the need for gel matching to identify corresponding proteins. Gel matching is necessary because the staining methods discussed above are intrinsically limited to a single-color display.

**Difference in-gel electrophoresis involves the simultaneous separation of comparative protein samples labeled with different fluorophores**

In Chapter 1, we discussed the comparative analysis of mRNA levels in different samples by labeling each population with a different fluorophore and hybridizing both populations simultaneously to the same DNA microarray. By scanning the microarray twice, at the emission wavelengths of each fluorophore, it is possible to determine the relative abundance of different mRNAs within each sample and the relative abundance of the same mRNA between samples. The signals can be rendered in false color and combined to provide a composite image that immediately identifies differentially expressed genes.
CHAPTER 4: STRATEGIES FOR PROTEIN QUANTITATION

**Difference in-gel electrophoresis (DIGE)** is an analogous method in proteomics based on 2DGE (see Chapter 2). The protein extracts from related samples (for example, healthy versus diseased tissue or stimulated versus unstimulated cells) are labeled on lysine side chains with succinimidyl esters of propyl-Cy3 and methyl-Cy5, two cyanine family fluorophores with the same mass and charge but different emission wavelengths. The protein samples are mixed prior to separation and loaded onto the IEF gel for separation in the first dimension, then transferred to an SDS–polyacrylamide gel for orthogonal separation. After electrophoresis, the gel is scanned using a CCD camera or fluorescence reader fitted with two different filters and two sets of data are obtained. The images from each filter can be pseudocolored and combined, immediately revealing the spots representing proteins whose abundance differs across the sample (Figure 4.4 and color plates). The use of further labels, for example Cy2, can allow even more samples to be run concurrently. Because the samples run together, all potential differences in gel preparation, running conditions, and local gel structure are eliminated.

DIGE has many advantages in terms of simplified data analysis, but the technique also has several drawbacks. The fluorescent labels are less sensitive than both SYPRO dyes and silver staining. This primarily reflects the fact that only a small proportion of the proteins in each sample can be labeled before solubility is compromised such that the proteins precipitate during electrophoresis. A further consequence of partial labeling is that the bulky fluorescent conjugate reduces the mobility of the proteins during SDS-PAGE so the gels must be post-stained, for example with Coomassie Brilliant Blue, to identify the “true” protein spot to be excised for downstream analysis by mass spectrometry. Such registration errors between the labeled and unlabeled protein populations are minimized during isoelectric focusing because the dyes carry a single positive charge that replaces the positive charge on the lysine side chain to which they bind and thus the labeled and unlabeled proteins have the same pI. Alternative cyanine reagents with a maleimide reactive group are designed to bind covalently to the thiol group of cysteine residues via a thioether linkage.

Accurate protein quantitation may be difficult because proteins differ in their labeling efficiency, solubility when conjugated to the label, and the extent to which they might exhibit quenching (a phenomenon in which there is energy transfer between two fluorophores that are close together on the same molecule, thus preventing the emission of light). Therefore, bright spots and dim spots may represent abundant and scarce proteins, or may represent proteins that are present at approximately the same level but show differential labeling efficiency or quenching effects.

**Parallel analysis with multiple dyes can also be used to identify particular structural or functional groups of proteins**

The sensitivity of standard gels can be combined with the convenience of multiplex fluorescence by using fluorescent reagents such as SYPRO Ruby to stain and compare protein spots on different gels plus more selective reagents that identify specific classes of proteins. These proteins can be used as landmarks for gel matching or to identify subsets of proteins that share specific structural or functional attributes. A number of stains have been developed that recognize various structurally or functionally related proteins: for example, glycoproteins and phosphoproteins (these are discussed in more detail in Chapter 8), oligo-histidine tagged proteins, calcium-binding proteins, and even proteins that have the capability to bind or metabolize particular drugs (see Chapter 9). For example, penicillin analogs have been produced carrying BODIPY dyes, which are relatively nonpolar and have a neutral chromophore and therefore do not interfere with the structure or...
The chemical behavior of the antibiotic. These so-called BOCILLIN reagents can efficiently identify penicillin-binding proteins on a two-dimensional gel with SYPRO Ruby used as a general counterstain. Similarly, BODIPY dyes have been used to generate analogs of the cysteine protease inhibitor trans-epoxysuccinyl-L-leucylamido(4-guanidino)butane, thus allowing cysteine proteases to be identified and changes in their expression levels following different types of cell treatment to be investigated.

4.4 QUANTITATIVE MASS SPECTROMETRY

Label-free quantitation may be based on spectral counting or the comparison of signal intensities across samples in a narrow m/z range

Because sample processing, separation, and transfer to the mass spectrometer are generally automated, quantitative data can only be obtained from LC-MS and LC-MS/MS experiments by determining the abundance of different proteins from their mass spectra. Quantitative mass spectrometry is often achieved by comparing samples that have been labeled with alternative discriminatory mass tags, but it is also possible to achieve an accurate quantitative comparison between unlabeled samples. The principles of the available strategies are summarized in Table 4.1 and their relative merits and drawbacks are compared in Table 4.2. However, it is important to emphasize that the vast majority of quantitative proteomics experiments provide relative rather than absolute quantitative data, which makes proteomics as a research approach distinct from the use of similar methods to measure protein levels in body fluids, for example (see Chapter 10).

The first label-free approach is spectral counting, which is conceptually similar to the census sequencing approach discussed in Chapter 1. The basis of this approach is that the number of recorded spectra corresponding to a particular peptide correlates with the abundance of that peptide in the original sample. This is a sensitive method for detecting differentially expressed proteins, although precise quantitation is affected by peptide ionization and fragmentation characteristics and becomes less accurate in the case of scarce peptides. The other major label-free strategy is based on the measurement of precursor ion signal intensity, which can be achieved in standard MS experiments by isolating m/z values representing one or more analytes of interest from a standard chromatogram. This is known as the extracted ion chromatogram (XIC) method. More sensitive quantitation is possible by preselecting the ions for analysis. This can be achieved by selected ion monitoring (SIM) in MS instruments, in which the m/z values for analysis are selected and only this restricted m/z range appears in the dataset. The preferred method is selected reaction monitoring (SRM) in triple quadrupole instruments using transition pairs (precursor and product ions) because the latter does not require full mass spectra to be recorded. By carrying out multiple SRM experiments (multiple reaction monitoring, MRM) and spiking the sample with isotopically labeled peptides as concentration standards, MRM can be used to construct a calibration curve that achieves absolute rather than relative quantitation.

Label-based quantitation involves the incorporation of labels that allow corresponding peptides in different samples to be identified by a specific change in mass

Quantitative proteomics is often based on the incorporation of stable isotopes or mass tags into different samples, allowing equivalent peptides (or peptide fragments) to be identified by a specific increase in mass. The general approach is to label alternative protein or peptide samples with equivalent
## TABLE 4.1 PRINCIPLES OF QUANTITATIVE MASS SPECTROMETRY METHODS

<table>
<thead>
<tr>
<th>Method</th>
<th>Principle</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Label-free methods</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectral counting</td>
<td>Counting the frequency of particular mass spectra</td>
<td>Unreliable for rare peptides</td>
</tr>
<tr>
<td>Precursor ion peak intensity (preferably SRM)</td>
<td>Direct comparison of peaks between spectra</td>
<td>Sensitive to instrument accuracy. More reliable in FT-ICR and Orbitrap mass analyzers</td>
</tr>
<tr>
<td><strong>Selective labeling of proteins and peptides in vitro</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICAT</td>
<td>Cysteine residues labeled with isotopic/normal mass tags containing reactive iodoacetamide (or acrylamide) groups. Carries biotin tag for affinity purification</td>
<td>Purification of cysteine-peptides simplifies analysis but presence of biotin complicates it (addressed by the development of a cleavable derivative). Restricts analysis to cysteine-containing proteins</td>
</tr>
<tr>
<td><strong>Nonselective labeling of proteins and peptides in vitro</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICPL</td>
<td>Derivatization of amines with isotopic/normal chemical groups</td>
<td>Labeling can be inefficient</td>
</tr>
<tr>
<td>MCAT</td>
<td>Derivatization of amines with O-methylisourea in one peptide population only</td>
<td>Inexpensive because stable isotopes are not used, but lacks accuracy</td>
</tr>
<tr>
<td>Proteolysis</td>
<td>Proteolysis in presence of $^{18}$O incorporates isotope into peptide carboxyl groups</td>
<td>Theoretically labels all peptides except C-terminal one, but incorporation often incomplete. Post-digestion labeling with immobilized trypsin is more efficient</td>
</tr>
<tr>
<td>TMT, iTRAQ</td>
<td>Derivatization of amines with isobaric mass tags</td>
<td>Eliminates quantitation problems caused by peak overlaps in first mass spectrum</td>
</tr>
<tr>
<td><strong>Nonselective labeling of proteins in vivo</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SILAC</td>
<td>Incorporation of isotopically labeled amino acids during metabolic activity</td>
<td>Corrects for preparation artifacts but only applicable to microbes and cultured cells</td>
</tr>
</tbody>
</table>

## TABLE 4.2 COMPARISON OF QUANTITATIVE MASS SPECTROMETRY METHODS

<table>
<thead>
<tr>
<th>Method</th>
<th>Application</th>
<th>Accuracy</th>
<th>Quantitative coverage</th>
<th>Linear dynamic rangea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metabolic protein labeling</td>
<td>Complex biochemical workflows&lt;br&gt;Comparison of 2–3 states&lt;br&gt;Cell culture systems only</td>
<td>+++</td>
<td>++</td>
<td>1–2 logs</td>
</tr>
<tr>
<td>Chemical protein labeling (MS)</td>
<td>Medium to complex biochemical workflows&lt;br&gt;Comparison of 2–3 states</td>
<td>+++</td>
<td>++</td>
<td>1–2 logs</td>
</tr>
<tr>
<td>Chemical peptide labeling (MS)</td>
<td>Medium-complexity biochemical workflows&lt;br&gt;Comparison of 2–3 states</td>
<td>++</td>
<td>++</td>
<td>2 logs</td>
</tr>
<tr>
<td>Chemical peptide labeling (MS/MS)</td>
<td>Medium-complexity biochemical workflows&lt;br&gt;Comparison of 2–8 states</td>
<td>++</td>
<td>++</td>
<td>2 logs</td>
</tr>
<tr>
<td>Enzymatic labeling (MS)</td>
<td>Medium-complexity biochemical workflows&lt;br&gt;Comparison of 2 states</td>
<td>++</td>
<td>++</td>
<td>1–2 logs</td>
</tr>
<tr>
<td>Spiked peptides</td>
<td>Medium-complexity biochemical workflows&lt;br&gt;Targeted analysis of few proteins</td>
<td>++</td>
<td>+</td>
<td>2 logs</td>
</tr>
<tr>
<td>Label-free (ion intensity)</td>
<td>Simple biochemical workflows&lt;br&gt;Whole-proteome analysis&lt;br&gt;Comparison of multiple states</td>
<td>+</td>
<td>+++</td>
<td>2–3 logs</td>
</tr>
<tr>
<td>Label-free (spectrum counting)</td>
<td>Simple biochemical workflows&lt;br&gt;Whole-proteome analysis&lt;br&gt;Comparison of multiple states</td>
<td>+</td>
<td>+++</td>
<td>2–3 logs</td>
</tr>
</tbody>
</table>

---

*In MRM mode, dynamic range may be extended to 4–5 logs.

reagents, one of which contains a heavy isotope and one of which contains a light isotope, or one of which contains a heavy mass tag and the other a light mass tag (or no tag at all). The samples are mixed, fractionated, and analyzed by mass spectrometry. The ratio of the two isotopic or mass tag variants can be determined from the peaks in the mass spectra and used to identify proteins that differ in abundance. Several variants of the approach can be used, which are discussed below and summarized in Figure 4.5.

ICAT reagents are used for the selective labeling of proteins or peptides

One of the first developments in quantitative mass spectrometry was a class of reagents known as isotope-coded affinity tags (ICATs). These are biotinylated derivatives of iodoacetamide (and later acrylamide) both of which react with the cysteine side chains of denatured proteins. Originally, the reactive group and biotin were joined by a linker that was available in two versions, one normal or light form and one heavy or deuterated form in which hydrogen atoms were replaced by deuterium. The heavy and light forms were used to label different protein samples and then the proteins were combined and digested with trypsin. The biotin allowed cysteine-containing peptides to be isolated from the complex peptide mixture through affinity to streptavidin, thereby considerably simplifying the peptide mixture and reducing the number of different peptides introduced into the mass spectrometer (Figure 4.6).

The original deuterated ICAT reagents were prone to partial peak separation during chromatography and the presence of the biotin group interfered with database searching. Therefore, a new cleavable ICAT reagent was introduced in which the biotin could be removed by acid treatment before mass spectrometry, and the heavy version incorporated $^{13}$C rather than deuterium. A solid-phase cleavable ICAT reagent has also been developed containing a photolabile linker arm so that cysteine-containing peptides from a complex mixture can be captured onto small plastic beads and then released by exposure to light. However, the main drawback of ICAT reagents that bind cysteine residues is that approximately 10% of proteins do not contain cysteine and are excluded from subsequent analysis.

**FIGURE 4.5 Overview of MS-based strategies for quantitative proteomics.**

Depending on the point at which the label is introduced, most procedures are classified as (a) in vivo labeling, (b) pre-digestion labeling in vitro, or (c) post-digestion labeling in vitro. (From Sechi S & Oda Y (2003) Curr. Opin. Chem. Biol. 7, 70. With permission from Elsevier.)
FIGURE 4.6 The ICAT reagent strategy for protein quantitation. Two protein mixtures representing two different cell states are treated with the isotopically light (pink) or heavy (red) ICAT reagents, respectively. The labeled protein mixtures are then combined and proteolyzed; tagged peptides are selectively isolated and analyzed by MS. The relative abundance is determined by the ratio of signal intensities of the tagged peptide pairs. The CID spectra are recorded and searched against large protein sequence databases to identify the protein. Therefore, in a single operation, the relative abundance and sequence of a peptide are determined. (From Tao WA & Aebersold R (2003) Curr. Opin. Biotechnol. 14, 110. With permission from Elsevier.)

Proteins and peptides can also be labeled nonselectively

More versatile systems have been introduced that allow nonselective protein and peptide labeling. For example, isotope-coded protein labeling (ICPL) is similar in principle to ICAT labeling but in this case the reagent labels lysine side chains by taking advantage of the ability of N-hydroxysuccinimide (NHS) ester derivatives to fully derivatize primary amino groups in intact proteins. ICPL reagents also have variants with different numbers of deuterium atoms to allow multiplex quantitative analysis. Similar approaches include the specific labeling of the exposed N-termini of peptides with alternative light and heavy versions of chemicals such as N-acetoxysuccinimide and 1-nicotinoyloxysuccinimide. There are also methods that enrich for the N-termini of proteins, allowing global analysis of the N-terminome (Box 4.4).

Nonselective labeling can also be achieved by taking advantage of the catalytic properties of proteases such as trypsin. As discussed in Chapter 3, when trypsin cleaves a protein and generates peptides, it uses oxygen atoms derived from water to create the new carboxyl group of each peptide C-terminus. This reaction can be exploited to identify y-series ions in fragment ion spectra (see p. 62), but it can also be used to differentially label peptides derived from alternative protein samples if normal water is used in one buffer and water substituted with heavy oxygen ($^{18}$O) is used in the other (Figure 4.7). The abundance of the peptides can then be compared, since they will appear as doublets separated by four mass units (although the C-terminal peptide of each protein is not labeled and no discrimination is possible). Whereas labeling is concurrent with digestion in the above method, it is also possible to uncouple the reactions and label the peptides after digestion by incubating the already digested peptides with immobilized trypsin and H$_2^{18}$O. The advantages of post-digestion labeling include the lower requirement for isotopic substrate and the ability to optimize the reaction to reduce incomplete oxygen exchange (that is, where only one $^{18}$O is incorporated instead of two, generating a more complex spectrum).

The use of isotopes is avoided with the mass-coded abundance tag (MCAT) system in which the primary amine groups of one population of peptides are derivatized with O-methylisourea and the other population is left without a label. This is inexpensive but not as accurate as isotope-based methods.

Isobaric tagging allows protein quantitation by the detection of reporter ions

All the selective and nonselective labeling methods above generate two versions of each protein or peptide, differing in mass by a specific amount.
This produces two peaks on the first mass spectrum, and, depending on the resolution of the mass analyzer, it may be difficult to achieve accurate quantitation because the peaks may overlap to a greater or lesser degree. **Isobaric tagging** means the labeling of proteins or peptides with chemical groups that are the same in mass, so that proteins from both samples behave in the same manner during fractionation and mass spectrometry, generating a single peak in the first mass spectrum. However, the reagents

**BOX 4.4 ALTERNATIVE APPLICATIONS.**

**Terminal amine isotopic labeling of substrates (TAILS)**

**TAILS** is a high-throughput proteomics approach that is useful for the quantitative analysis of N-terminal peptides. There are several variants of the method, but all involve the uniform labeling of exposed amines (N-terminal amines and lysine side chains) followed by the negative selection of blocked N-terminal peptides. The use of differential labeling allows quantitative comparison between samples, and is highly useful for the analysis of protease targets by comparing the N-termini before digestion and the neo-N-termini afterwards. This method is superior to those relying on chemical modification and/or biotinylation (which do not provide reliable quantitative data) and to combined factional diagonal chromatography (COFRADIC), which requires multiple chemical processing steps before separation and analysis. The dimethylation-TAILS method only allows pairwise comparisons, but the method can be combined with SILAC or iTRAQ labeling (see p. 82) to increase the number of different tags that can be used simultaneously (**Figure 1**).

**FIGURE 1** Dimethylation-TAILS.
Exposed amine groups in two protein samples are isotopically labeled in vitro with $^{2}$CH$_{2}$-formaldehyde in one case and $^{13}$CH$_{2}$-formaldehyde in the other, simultaneously blocking the groups. The samples are then pooled, digested with trypsin, and mixed with a hyperbranched polymer that captures the exposed (unblocked) terminal amine groups of the resulting peptides, leaving a peptide mixture highly enriched for the original (blocked) N-terminal peptides.
are designed so that fragmentation during MS/MS releases reporter ions with different masses, allowing the abundance of the corresponding peptide to be determined. This is achieved by using mass tags comprising three regions: a reporter region, a mass balancing region, and a linker region connected to the reactive group. The mass of the reporter region plus the mass balancing region is the same in all forms of the reagent but the individual masses differ so that when the reporter ion is released it can be resolved to a particular source (Figure 4.8).

Two major isobaric tag platforms are available. The first is known as the tandem mass tag (TMT) system and comprises a mass reporter region separated from a mass normalization region via a linker that is vulnerable to fragmentation. The different forms of the label are generated by differential isotopic substitutions in the mass reporter and normalization regions, and currently there are duplex and 6-plex versions available. The other system is known as iTRAQ (isobaric tags for absolute and relative quantification) and it works on similar principles, with 4-plex and 8-plex versions available. In the 4-plex version, the four reporting groups have masses of 114, 115, 116, and 117 Da with balancing groups of 31, 30, 29, and 28 Da ensuring that all four tags have a mass of 145 Da.

FIGURE 4.7 Enzymatic stable isotope coding of proteomes. For enzymatic labeling, proteins from two distinct proteomes are proteolytically digested in aqueous buffer containing either normal water (H$_2$)$_{16}$O; white squares) or isotopically labeled water (H$_2$)$_{18}$O; red squares). This encoding strategy effectively labels every C terminus produced during digestion. The samples are combined at the peptide level and then analyzed by microcapillary LC-MS. (From Goshe MB & Smith RD (2003) Curr. Opin. Biotechnol. 14, 101. With permission from Elsevier.)
Metabolic labeling introduces the label before sample preparation but is limited to simple organisms and cultured cells

Metabolic labeling involves the incorporation of an isotopic label into proteins while the sample is still metabolically active, for example, by growing cells in a medium containing heavy isotopes and comparing them with controls growing in normal medium. For example, Washburn and colleagues (see Further Reading) grew yeast cells in $^{15}$N-minimal medium or $^{15}$N-enriched medium, then pooled the cells for protein extraction, digestion, fractionation, and analysis by MS/MS. They identified more than 800 differentially expressed proteins as doublets differing in mass by one unit. The advantage of this approach is that the label is introduced early in the experiment, thereby eliminating variation arising from sample preparation and purification losses (Figure 4.9). One widely used variant of this approach is stable-isotope labeling with amino acids in cell culture (SILAC), which involves the inclusion of isotypically labeled amino acids (for example, $[^{15}$N]lysine) in the medium for one population of cells, which are then compared with controls fed with normal lysine. In more ambitious strategies, it has been possible to compare cultures fed with up to five different isotopic forms of arginine. The drawback of SILAC and other metabolic labeling methods is that they are restricted to the analysis of simple biological systems that can be maintained in a controlled environment. It is not possible to use this method with tissue explants, biopsies, body fluids, or cells that are difficult to maintain in culture.

The relative merits of different quantitative mass spectrometry methods based on SILAC have been explored in yeast, as discussed in Box 4.5.
CHAPTER 4: STRATEGIES FOR PROTEIN QUANTITATION

**FIGURE 4.9** Metabolic stable isotope labeling. Cells from two distinct cultures are grown on media supplemented with normal amino acids (1H/12C/14N) or 14N-minimal media (white spheres) or stable-isotope amino acids (2H/13C15N) or 15N-enriched media (red spheres). These mass tags are incorporated into proteins during translation, thus providing complete proteome coverage. An equivalent number of cells for each sample is combined and processed for microcapillary LC-MS analysis. (From Goshe MB & Smith RD (2003) Curr. Opin. Biotechnol. 14, 101. With permission from Elsevier.)

---

**FURTHER READING**


In 2006, Matthias Mann and colleagues showed that the SILAC method could be used to detect more than half of the proteins in the proteome of log-phase yeast cells (based on previous experiments showing that approximately 4500 proteins were expressed in such cells). In 2008, they published another ground-breaking study in which they used SILAC to differentially label the proteins in haploid and diploid yeast cells, and then carried out a comprehensive quantitative analysis to compare protein abundance in the different cell states.

Three different strategies were used as shown in Figure 1, resulting in more than 32% protein coverage by peptides and hence the unambiguous identification of 4399 proteins. The second strategy, which involved the digestion of proteins in solution followed by separation by IEF, was both the simplest and the most successful, yielding 3987 proteins. There was an 89% overlap between the proteome dataset produced in this experiment and previous large-scale studies based on protein tagging by homologous recombination (Chapter 7). The data did not appear to select against low-abundance proteins (indeed, several of the identified proteins are thought to be present at fewer than 50 molecules per cell) and the representation of membrane proteins was higher within the experimental dataset than within the yeast genome.

The quantitative data were based on the analysis of 1,788,451 SILAC peptide pairs, which represents more than 30 peptides per protein, and this analysis revealed 196 proteins whose abundance differed significantly between haploid and diploid cells (Figure 2 and color plates). The affected proteins included key members of the pheromone signaling pathway that is responsible for mating in yeast as well as transposon-associated proteins and proteins associated with the cell wall. There was little agreement between the proteomic data and previous transcriptomic studies, although once low-confidence microarray results were filtered out there was better correlation, at least among the genes involved in the pheromone response pathway.

**FIGURE 1** Three different strategies based on the differential incorporation of labeled amino acids to compare the proteomes of haploid and diploid yeast cells. (From de Godoy LMF, Olsen JV, Cox J et al. (2008) *Nature* 455, 1251–1254. With permission from Macmillan Publishers Ltd.)

**Continued on next page**
Box 4.5 Case Study.
Comparative quantitative proteomics of haploid and diploid yeast cells.

**Figure 2:** Quantitative difference between the haploid and diploid yeast proteome (overall fold change). Proteins to the left (becoming deeper gray) are more strongly represented in haploid cells. Proteins to the right (becoming deeper red) are more strongly represented in diploid cells. (From de Godoy LMF, Olsen JV, Cox J et al. (2008) *Nature* 455, 1251–1254. With permission from Macmillan Publishers Ltd.) See also color plates.


