BASIC PROTEOMIC PROTOCOLS

Casado-Vela, J. (jcasado@cnb.csic.es)
Centro Nacional de Biotecnología, Spanish National Research Council. (CSIC). Darwin 3, Campus de Cantoblanco. 28049 Madrid, Spain.

Acknowledgements. J. Casado-Vela is a JAE-DOC (CSIC) holder supported by Ministerio de Economía y Competitividad, Spain and co-funded by the European Social Fund.

Contents

1. INTRODUCTION

The objective of Basic Proteomic Protocols is to get a first glance to routine proteomic practice and protocols to recently initiated researchers willing to introduce these protocols in their daily work or in collaboration with proteomic platforms sharing their expertise and instrumentation. We do believe that this may help as a quick starting guide before wet-lab work. Readers are invited to contribute and to include corrections to this guide.

The term proteomics was first coined in the mid-1990s and has become a key research area during the last decade. It is dedicated to study the biological implication of proteins in disease, pathologies, or alterations of living organisms. The proteome was first defined as the protein complement of the genome. The proteome is dynamic and its composition changes as a consequence of alterations in an effort to adapt to environmental changes and keep the homeostasis of living systems. There are at least four main questions that proteomic studies aim to find a response for: How many proteins change as a consequence of an alteration? (2) How much do these proteins change? (3) Did the changes observed involve the regulation of any biochemical pathways? (4) What is the biological relevance or can we take advantage of that information to solve the alteration? To answer these questions, a number of techniques are routinely used to comprehensively and systematically separate, visualize, identify, and quantify the plethora of proteins coexisting in tissues and cells. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), high performance liquid chromatography (HPLC), and mass spectrometry (MS) are the most important tools in proteomics. The combination of all these techniques allows the identification and quantification of up to thousands of proteins, the identification of proteins differentially expressed in normal versus altered organisms/tissues, and the characterization of posttranslational modification of proteins.

2. PROTEIN ARRAYS

Protein arrays represent a step forward and innovative approach in proteomics, as they enable to address biological and functional questions by large-scale approaches (i.e., up to several thousand proteins may be monitored in a single experiment). Protein array technology enables the identification and characterization of proteins, including their interactions. Protein arrays open the door for addressing a plethora of biological questions through sensitive (i.e., low abundant proteins/molecules should be detected) and high-throughput (i.e., several thousand proteins/molecules can be screened in a single experiment) approaches. Albeit the apparent simplicity of the concept, it is difficult to build a general classification including all kinds of protein arrays currently available. In this sense, three main factors contribute to add some complexity including: i) the analytes of interest or their substrates could be either immobilized on the surface of arrays or layered on the array surface as soluble chemicals, ii) the number of different applications of protein arrays is wide and is continuously expanding and evolving, iii) the proteins/peptides/antibodies attached on arrays can be obtained from natural sources, purified after their expression using heterologous cellular systems or synthesized directly onto the array surface. The molecular characteristics of proteins make construction of protein arrays much more challenging than building DNA microarrays. The major obstacle to the development of this field is the

content of the microarrays. In contrast to the PCR technique for nucleic acids that can generate large quantities of virtually any DNA, there is no counterpart for protein amplification. Protein expression or production systems are difficult to automate and always unpredictable. Many proteins are unstable, which complicates microarray shelf life. Finally, in contrast to the simple hybridization procedures for nucleic acids, proteins have shown a wide range of chemistries and specificities that complicate substantially any interaction study at the multiplex level. Two major approaches have been followed for the production of protein microarrays: abundance- based microarrays and function-based microarrays.

2.1. Abundance-Based Microarrays

There are two types of abundance-based microarrays: (1) antibody or antibody-like microarrays and (2) reverse phase protein microarrays. Antibody microarrays are the most popular option given the stability, selectivity, and high affinity of the antibodies as well as the high number of antibodies commercially available. Therefore, this option has been the fastest growing in the field. In any case, some precautions should be taken into account: not all the antibodies are actually functional in the microarray format, and even monospecific antibodies can show cross-reaction on the microarray when confronted with complex protein mixtures. The most practical approach for the use of capture microarrays is the direct labeling of the analytes with one or two fluorochroms. This approach can be used for the analysis of differential expression of proteins in different conditions (e.g., for the analysis of tumour tissues versus normal tissues in cancer). This approach has recently been reported by a number of authors in bot h prostate and breast cancer. Reverse phase protein microarrays rely entirely on the existence of analyte specific reagents. This is particularly important as the concentration of analytes in the spots is locally high and may facilitate spurious interactions. Other authors used reverse phase microarrays to classify different types of cancer using the NCI-60 cell lines. The cell lines were spotted as lysates and then probed with antibodies. More than 200 antibodies were tested to finally select 52 that gave a reliable signal. Also, some interference problems may appear from the locally high concentration of analytes. This might be alleviated by prefractionation of the samples before spotting.

2.2. Function-Based Protein Microarrays

Function-based microarrays consist of panels of proteins spotted at defined positions. They are used to study the biochemical properties and activities of the target proteins spotted on the microarray. Thus, protein microarrays are mainly used to examine protein interactions with other proteins or other molecules. There are also other arrays used for studying the enzyme activity and substrate specificity. Proteins have to be printed in the right conditions to maintain conformation, integrity, and activity. In some cases, the integrity of the protein requires it to be printed as a complex of several proteins to remain active and functional. The major challenge about this type of microarray continues to be the production of the proteins. It is time consuming and costly to produce proteins of good purity and yield. Generally, these arrays are made by using three different approaches: (1) chemical linkage, (2) peptide fusion tag, and (3) self-assembling protein arrays. All three methods have advantages and disadvantages. Alternatively, it may be simpler to study protein domains rather than the full length proteins, because small domains are usually simple to express and purify. Self-assembling microarrays, also called nucleic acid programmable protein arrays (NAPPAs), are a new and interesting alternative based on the printing of cDNAs on the microarray surface and expressed in situ using a mammalian cell-free expression system. The nascent molecule, labeled with a tag, is captured by an adjacent antibody against the tag printed on the chip surface. This approach is very promising because it excludes the necessity of expressing, purifying, and storing the proteins. Also, it opens new avenues to

prepare custom-made protein microarrays. There is great enthusiasm about this new area of protein microarrays. Still, major efforts need to be made in different areas of production before these tools become commonly used in laboratories, as happens with DNA chips. Few articles have reported original findings of proteomic interest and most of the communications to date have been based more on technology optimization than in solving real problems. Therefore, the utility of protein microarrays remains to be fully determined.

3. TWO-DIMENSIONAL POLYACRYLAMIDE GEL ELECTROPHORESIS

Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) allows the separation of hundreds (up to thousands) of proteins in a single experiment. The result is a gel matrix where proteins appear as single spots that can be visualized after a protein staining process. After digitalization of the gel to obtain an image with the help of transmission light scanners, the subsequent images can be processed for spot analysis. It can be used to find differential expression profiles between normal versus altered protein patterns and to quantify these changes by densitometric measurement of the intensity of the protein spots observed in the gel. The quality and biological significance of the results are compromised by the protein sample preparation and the reproducibility of the experiments. The main steps involved in 2D-PAGE analysis and some of the potential problems are described next.

3.1. Protein Sample Preparation

Proteins can be found soluble in the cytosol, embedded in cell membranes, and bound to nucleic acids or to other proteins. For 2D-PAGE, it is necessary that the pool of proteins submitted be completely soluble. Different treatments can be used to solubilize different types of proteins. The choice of cell/tissue disruption method, composition of lysis buffer, and the use of detergents directly affect the solubilisation of proteins. Since proteases are normally present in biological samples and proteolysis may cause undesired results in the 2D gels, cocktails of protease inhibitors should be added during the protein extraction steps. Some of the protease inhibitors frequently used typically include 4-(2-aminoethyl) benzenesulfonyl fluoride (AEBSF), ethylenediaminetetraacetic acid (EDTA), bestatin, E-64, leupeptin, aprotinin, and phenylmethylsulfonyl fluoride (PMSF), among others. If only a subset of proteins from the biological sample is of interest (i.e., subcellular proteins), prefractionation methods need to be employed. This is normally achieved by differential solubilization or ultracentrifugation in sucrose gradients. Once the proteins of interest have been extracted, it is common to include some steps to remove contaminants. Dialysis for salt removal, lyophilization, precipitation of proteins with trichloroacetic acid–cold acetone, or DNA removal with the use of endonucleases are among the current procedures.

Protocol A: Protein Cleaning and Preparation for Isoelectrofocusing Sample volumes corresponding to 100 μg of protein, previously quantified, are pelleted by centrifugation at 13,000 rpm after addition of 6% trichloroacetic acid in chilled acetone and incubation at 4 °C for 30 min. The addition of 0.2% deoxycholate usually facilitates the precipitation of proteins. The precipitation can be repeated several times if there is an excess of salts in the sample of interest. After removal of the supernatant, the protein pellet is dried under a nitrogen stream at room temperature. The appearance of several spots corresponding to the same protein is frequently observed in 2D gels. This can be attributed most times to the oxidation state of thiol groups in the protein of interest. This problem is solved by oxidizing the thiol groups, for instance, adding chemicals, such as DeStreak (Amersham Biosciences), immediately before or during the rehydration of immobilized pH gradient (IPG) strips.

3.2. First Dimension: Isoelectrofocusing

Prior to the isoelectrofocusing (IEF), dry and clean precipitates need to be solubilized in a rehydration solution (typically 100–300 μL) containing 8 M urea, 2% 3[(3- cholamidopropyl)dimethylamonio]-1-propanesulfonate (CHAPS), 2 mg/mL dithiothreitol (DTT), 0.0125 g/mL iodoacetamide (IAM), a trace of bromophenol blue, and carrier ampholytes IPG buffer (commercially available depending on the pH range of the IPG strip). Proteins are embedded in the IPG strip at low voltages (50 V per IPG strip) and are ready for IEF. It consists of different steps: a first step of 500 V for 1 h, a second step of 1000 V for 1 h. The final step requires 8,000 V for 4 h, until a total voltage of about 40,000 (volt × hour)V · h. After completion of the IEF, each strip is deposited in a vial with 5 mL of equilibrating solution and left with agitation for 15 min. 3.3. Second Dimension: Polyacrylamide Gel Electrophoresis—Choice of Gel Size, Concentration, and Protein Molecular Weight Range After the IEF, the IPG strip must be equilibrated by covering it with a solution containing 50 mM tris(hydroxymethyl)aminometane (Tris-base) pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.01 g bromophenol blue, and 10 mg/mL DTT. Proteins are separated according to their molecular weight under denaturing conditions in the presence of SDS.

Protocol B: Standard SDS-PAGE Gel Solution for 100 mL Solution for 12.5% Acrylmide-Bisacrylamide Gels Volume Stock solution 40% acrylamide/bisacrylamide (37.5 : 1, 2.6% C) 41.7 mL 1.5 M Tris-HCl pH 8.8 25 mL 10% Sodium dodecyl sulfate (SDS) 1 mL Water 31.8 mL 10% Ammonium persulfate 0.5 mL N,N,N',N'-tetramethylethylenediamine (TEMED) 33 μL Once the gel is polymerized, the IPG strips, containing the proteins focused at their isoelectric point (pI) and equilibrated, are placed on top of the gel and the proteins are separated according to their molecular weight in a chamber containing running buffer (25 mM Tris-HCl, 192 mM glycine, and 0.1% SDS).

4. GEL STAINING AND IMAGE PROCESSING

The staining of 2D gels is a crucial step in proteomic analysis. It allows the visualization of protein spots resolved in the 2D gel and enables further analysis, including the identification of differential proteins among samples, densitometric quantification, and recovery of proteins of interest.

4.1. Coomassie Brilliant Blue and Silver Staining

An improved method for Coomassie Brilliant Blue (CBB) stained gel is the colloidal Coomassie Brilliant Blue. Silver staining is used to visualize proteins that are present at a low concentration in the gel (2 ng/spot) (Fig. 1a).

Protocol C: Colloidal Coomassie Brilliant Blue G-250 Staining Fix the gels with 40% methanol and 10% acetic acid for 30 min. Prepare a stock solution of colloidal Coomassie Brilliant Blue G-250 staining as follows: mix 4 g of Brilliant Blue G-205 with 230 mL of 85% phosphoric acid and stir. Add 640 mL of saturated ammonium sulfate solution and stir. Add distilled water to 4 L. Prepare 100 mL of staining solution by adding 20 mL methanol to 80 mL of the stock solution. The staining solution should be prepared immediately before use, as the colloid only lasts for a few hours. Stain gels overnight (if maximum sensitivity is required) using a rocking platform to avoid colloidal stain sticking to the surface of the gel. Finally, destain the gels with 3–5 changes of 25% methanol until the spots appear clearly defined in the gel. The use of rocking platforms during the staining is crucial, because the staining of gels without agitation can cause the appearance of noise and background during software-

based image densitometry of the gels. The duration and temperature of staining and destaining steps are also crucial in order to obtain reproducible results.

Protocol D: Silver Staining/Destaining of Proteins in Polyacrylamide Gels Cover the gel with fixing solution (50% methanol, 10% acetic acid), remove the fixing solution, cover gel with 5% methanol for 15 min, and wash 3 times with distilled water. Cover the gel with sensitizing solution 0.2 g/L of Na 2 S 2 O of silver nitrate (AgNO 3 freshly prepared for 2 min and wash with distilled water. Cover gel with 0.2 g/L 3) for 25 min and wash 3 times with distilled water. Cover gel with developing solution containing 3 g/100 mL Na 2 CO 3 , 50 µL/100 mL of 37% HCOH, and 2 mL of sensitizing solution. Allow staining until the protein spots are visible over the background of the gel (10 min maximum). Remove the developing solution and stop with 14 g/L Na 2 -ethylenediaminetetraacetic acid (EDTA) for 10 min. Finally, wash with distilled water. Once stained, the gels can be destained by covering the gels with 0.2 g potassium ferricyanide in 100 mL of sensitizing solution, leaving a yellow background in the gel. Restaining is also possible starting the protocol again by adding the silver nitrate solution.

4.2. Fluorescent Dyes

Silver staining, while sensitive, exhibits a low dynamic range. Handling may significantly change the results and the background of the gels frequently interferes with further characterization of silver-stained proteins using mass spectrometry. Fluorescent staining is superior in sensitivity and linear response as compared to Coomassie Brilliant Blue and silver staining, respectively. A range of fluorescent dyes are commercially available for different purposes, including Sypro (r) Ruby, Deep purple, 5-hexadecanoylamino-fluorescein. Sypro Ruby is a sensitive fluorescent stain for detecting proteins separated by polyacrylamide gel electrophoresis. This fluorescent stain does not interfere with subsequent analysis of proteins by mass spectrometry and is quantitative over three orders of magnitude. Apart from its high sensitivity, Sypro Ruby will not stain nucleic acids.

Protocol E: Sypro Ruby Staining of Proteins in Polyacrylamide Gels It is highly recommended to bind the gel to one of the glass plates prior to running the electrophoresis as follows: prepare a Bind-Silane working solution by mixing 8 mL of ethanol, 200 µL of acetic acid, 10 µL of Bind- Silane, and 1.8 mL of distilled water. Cover the inner surface of one of the glass plates used to polymerize the acrylamide gel and dry to room temperature for 1–2 h. Perform the assembly and polymerization of the gels as usual. The gels will stay attached to the glass during electrophoresis, staining procedures, scanning, and storage. Incubate the gel in fixing solution (30% methanol, 7.5% acetic acid) for at least 2 h. Remove fixing solution and cover the gel with Sypro Ruby stain. Incubate 3 h or overnight with gentle shaking protected from direct light. Wash extensively with distilled water (at least 4 changes of 20 min). Finally, scan the gels and select the spots of interest. Sypro Ruby protein gel stain can be imaged using laser-based imaging systems equipped with 450, 473, 488, or 532 nm laser lines.

4.3. Differential In-Gel Electrophoresis Staining

An outstanding application of fluorescent dyes to proteomics using 2D gels is the differential in-gel electrophoresis (DIGE) technology, which is based on the use of three CyDye™ fluorochroms, Cy2, Cy3, and Cy5, for protein labeling (Fig. 1b). It allows one to label up to three samples that can be separated in a 2D gel in order to search for differential protein profiles, thus avoiding experimental variations from gel to gel and increasing the accuracy of the results. A pooled internal standard, containing equal amounts of protein from each sample, should also be created as a reference for normalization in these

experiments. Fluorescent two-dimensional DIGE constitutes the proteomic equivalent to gene expression analysis by DNA microarrays. Even minor differences of protein expression can be detected across multiple samples simultaneously with statistical confidence by using the DeCyder software. The comparison of spot intensities using the DIGE approach and the DeCyder software is more objective than the conventional approach based on the comparison of the brightness of gel images obtained by conventional staining. The quantification of abundance changes are obtained over a linear dynamic range of almost four orders of magnitude for the Cy dyes. This novel technology is being applied to the analysis of differential protein expression in different neoplasia for the search of cancer markers or to study drug resistance.

Protocol F: DIGE Protein Labeling Reconstitute each vial of commercial CyDye in high quality N,N-dimethylformamide (DMF) to a final concentration of 400 pmol/ μL. After complete resuspension, each dye gives a visible color, Cy2—yellow, Cy3— red, and Cy5— blue. Add the proper amount of dye to the different protein (up to three) extracts. Normally, 400 pmol of dye is used to label 50 μg of protein, although higher concentration of protein can be efficiently labeled by adding proportional amounts of each dye. Let dyes react with the sample during 30 min at room temperature in the dark and quench the labeling reaction by adding 10 mM lysine. Combine the labeled samples into a single microfuge tube and mix. Centrifuge for 10 min at 13,000 rpm to pellet any particles not dissolved. Add to each labeled sample and pooled samples 100–300 μL (depending on drystrip length) of rehydration solution, containing 8 M urea, 2% CHAPS, 2 mg/mL DTT, 0.0125 g/mL iodoacetamide, a trace of bromphenol blue, and carrier ampholytes IPG buffer (commercially available depending on the pH range of the IPG strip). The following steps of the 2D electrophoresis are performed as described in Section 3.2. Other specific applications of fluorescent dyes for the detection of subsets of proteins have also been developed; such as the case of ProQ diamond for phosphoproteins.

4.4. Software-Based Image Analysis for Quantification and Comparison of Protein Patterns

Software-based analysis of conventional (non-DIGE) 2D gels includes the detection of spots (individual proteins) present inside the gel matrix. Although the reproducibility from gel to gel is relatively high, some degree of experimental variability can be observed from gel to gel (even if the same protein mixture is analysed). For this reason, triplicates (at least) of 2D gels should be run for every individual sample in order to compare the pattern of spots that appear in each of the gels. In this way, the software is able to measure the average "volume" of each spot, measured as the area (in dots per square area of the gel surface) by the intensity of the same spot (the value of intensity assigned to each spot) assigned as a percentage of the total intensity of the surface of the 2D gel. The scanner-densitometer used has been calibrated by assigning an algorithm able to transform the transmission light that reaches the detectors through the gel. Therefore, the software can reveal the differences in protein profiles and measure changes in the intensity of spots from each sample.

4.5. Drawbacks and Limitations of 2D-PAGE

2D-PAGE gels can resolve a limited range of proteins, in terms of molecular weight and isoelectric point. In general, high range (above 150 kDa) or low range (below 12 kDa) molecular weight proteins are difficult to detect with this technique. The same limitations apply to hydrophobic proteins (e.g., membrane proteins), which normally migrate very poorly in these gels. Finally, it should be taken into account that only relatively abundant proteins can be visualized. Commercial pI strips including immobilized ampholites range from 3 (most acidic) to 11 (most basic). Although these pI values cover a

wide range of the proteins present in living organism, it is not possible to include proteins with higher or lower pI points in a 2D gel. The size of the gel needs to be considered. Larger 2D gels allow a better resolution of proteins that could appear as overlapped spots in small 2D gels, but they require longer pI strips and considerable consumption of acrylamide and other reagents (especially when replicates of the experiments are necessary), and the handling for staining and scanning is a hurdle. On the other hand, conventional staining approaches have a limited dynamic range. The use of sensible staining using fluorescent dyes requires the use of equipment such as laser scanners and robotized spot-pickers to pick up the spots of interest for further analysis. This instrumentation is expensive and not available in every lab.

5. IN-GEL IN-SOLUTIN ENDOPROTEASE DIGESTION OF PROTEINS

In most proteomic studies, the start-up material used for high throughput protein identification is a proteolytic mixture of the proteins. Here we describe some of the current methods for protein digestion.

5.1. Reduction and Alkylation of Proteins

Reduction with 10 mM DTT is necessary to break disulfide bonds and to unfold the protein. Alkylation with 50 mM iodoacetamide is an efficient reaction that yields carboxyamidomethylated residues and keeps the protein unfolded. This allows the endoprotease to access all the cleavage sites efficiently.

5.2. Trypsin Digestion

Trypsin is the most frequently used endoprotease for proteomic analyses for several reasons. It possesses high cleavage specificity after Lys or Arg residues (provided Pro is not the following amino acid in the protein sequence). The frequency of appearance of Lys and Arg residues inside proteins is relatively high compared to other amino acids. Modified trypsins are readily available in the market that reduce the autoproteolysis and increase the efficiency of the reaction. Moreover, trypsin cleaves after basic residues, which favors the protonation of the peptides under low acidic conditions (which is crucial for mass spectrometry in positive ionization mode). The amount of trypsin added to samples must be proportional to the amount of protein, normally 1:20 or even 1:50 (w/w) trypsin:protein ratio.

5.3. Alternative Proteases

In some cases, tryptic digestion is not sufficient to characterize the protein of interest because the resulting peptides are too small or too large for their separation with C18 reverse phase columns and subsequent analysis with mass spectrometry. In these cases, the combination of two (or more) endopeptidase digestions with mass spectrometric analysis of the resulting peptides enables the comprehensive identification or proteins present in complex mixtures. It is advisable to use combinations of proteases with different specificities, such as trypsin combined with chymotrypsin, Asp-N, or Glu-C.

5.4. In-Gel Digestion of Proteins

This method is based on the behavior of gel plugs to act as a "sponge," shrinking when acetonitrile is added or swelling when aqueous buffer or other reagents are added to the piece of gel.

Protocol G: In-Gel Digestion Proteins Cover the gel plug containing the protein of interest with acetonitrile for 10 min; remove the acetonitrile. Cover the gel with 10 mM DTT in 50 mM ammonium bicarbonate and store at 56 °C for 30 min; remove the DTT. Cool to room temperature. Alkylate the

sample by covering the gel with 50 mM iodoacetamide in 50 mM ammonium bicarbonate and keep at room temperature in the dark for 15 min; remove the iodoacetamide. Add trypsin at a 1 : 20 ratio and store at 37 °C for 8 hours. Stop the endoprotease reaction by adding 1 µL of concentrated acetic acid. Remove the supernatant and place it in a new clean tube. Cover the plug with acetonitrile for 10 min and recover the supernatant. Mix both supernatants and lyophilize the peptide mixture. The peptide mixture will be ready for mass spectrometry after resuspension in appropriate buffer.

5.5. In-Solution Digestion of Proteins

Proteins in solution in different buffers can be lyophilized to dryness and redissolved in 50 mM ammonium bicarbonate containing relatively high concentrations of urea (4–8 M) to help the solubilization of the proteins at basic pH, necessary for tryptic or chymotryptic endoprotease digestion. When 8 M urea is used to solubilize proteins, the concentration of urea should be diluted to 2 M prior to endoprotease digestion in order to maintain the protease activity.

Protocol H: In-Solution Digestion of Proteins Lyophilize the sample to dryness and redissolve in a small volume (10 µL of 50 mM ammonium bicarbonate). Vortex and/or bath sonicate to ensure the solubilization of your protein(s) of interest. Add 10 µL of DTT in 50 mM ammonium bicarbonate and store at 56 °C for 30 min. Cool to room temperature. Alkylate by adding 10 µL of iodoacetamide in ammonium bicarbonate and keep at room temperature in the dark for 15 min. Dilute in one volume of 50 mM ammonium bicarbonate to reduce the urea concentration to 2 M. Add trypsin at a 1:20 ratio and store at 37 °C for 8 h. Your peptide mixture will be ready for mass spectrometry analysis after resuspension in the appropriate buffer.

6. PRINCIPLES OF MASS SPECTROMETRY: IONIZATION SOURCES AND DIFFERENT MASS ANALYSERS

Mass spectrometry has become a key area in proteomic studies. Several sources of soft ionization for peptides and proteins have been developed during the last decades to allow high throughput analysis of peptides and routine identification and characterization of proteins (or their proteolytically derived peptides) present at very low concentrations in biological samples (even attomoles / nanograms of protein). Moreover, tandem mass spectrometry (MS/MS) approaches enable the confident identification of mixtures of proteins based on amino acid sequence information. Mass spectrometers are composed of three parts: the ionization source, the mass analyser, and the detector. Several mass analysers can be found in mass spectrometers, including triple quadrupoles (Q), time-of-flight (TOF) analysers, ion traps (IT), Fourier cyclotron resonance (FCR), and combinations such as quadrupole time-of-flight (Q-TOF). All these analysers are able to accurately measure the mass of compounds (proteins or peptides) that enter the analyser in gas phase as a function of their charge state (mass-to-charge (m/z) ratio).

6.1. Matrix Assisted Laser Desorption Ionization–Time of Flight (MALDI-TOF)

Spotting on MALDI Plates and Use of Different Matrices The MALDI technique has primarily been used in conjunction with TOF analysers for molecular mass determination. For MALDI, peptides are mixed with a matrix that transfers the energy of a laser beam to the peptides embedded in the crystallized matrix. The resulting spectra, characterized by singly charged precursor ions, present high energy collision fragments, which are readily interpretable.

Protocol I: Dried Droplet Method for Spotting of Samples on MALDI Plates After tryptic digestion and lyophilization, the resulting peptides are resuspended in 5 µL water with 0.1% trifluoroacetic acid. It is

advisable to introduce desalting steps by using Zip-tips ® containing C18 beads (commercially available). The solution containing the peptides is mixed with 5 µL of MALDI-TOF matrix. Normally, 1% acyano-4-hydroxy-cinnamic acid dissolved in 90% acetonitrile and 10% trifluoroacetic acid is used as matrix for peptide mass fingerprinting of proteins. Normally, 1 µL of the mixture (peptides and matrix) is spotted on the MALDI plate and allowed to dry for 5 min before MALDI analysis.

6.2. Acquisition of Peptide Mass Fingerprints (PMFs) for Protein Identification

Peptide mass fingerprinting is the most common strategy for identifying proteins proceeding from SDS-PAGE separations, either 1D or 2D. The proteins are usually in-gel digested with trypsin or other endoproteases (e.g., Lys C or Glu C). The resulting proteolytic peptides are extracted from the gel piece and analysed by MALDITOF MS. The peptide masses contained in the resulting spectra are the peptide mass fingerprints. This mass profile is matched against the theoretical masses obtained from the in silico digestion of all proteins contained in the database. This interrogation is made with algorithms such as MASCOT (http://www.matrixscience. co.uk), Phenyx (http:/www.phenyx-ms.com), or Profound (http://prowl.rockefeller. edu/profound_bin) among others. The choice of the algorithm may influence the final identification. A combination of several search algorithms is advisable. The results of the matching are ranked according to the number of peptide masses matching their sequence within a given mass error tolerance. Successful protein identification requires several factors: (1) mass accuracy of the instrument, (2) ratio of assigned versus non-assigned peaks in the spectrum, and (3) quality and size of the database used. Common databases used for the analysis are SwissProt or NCBI. The whole process is now automated and hundreds of samples can be analysed per day. As a general rule, when MASCOT is used for identification, protein identification is considered as significant when the MASCOT score is higher than 66 and the sequence coverage is higher than 15%. 6.3. Time of Flight–Time of Flight (TOF-TOF) Fragmentation of Precursor Ions TOF-TOF instruments combine the high throughput of the PMF analysis with the increased confidence given by peptide fragmentation that extends protein identification coverage. This approach is particularly useful when complex mixtures of peptides are analysed. These instruments have the ability to acquire MALDI-TOF mass spectra followed by high energy collision-induced dissociation (CID) and to perform subsequent MS/MS analysis of fragment ions for more definitive protein identification. The peptide sequence information gained in MS/MS experiments leads to higher confidence in the database search results and identification of more proteins. The acquisition queue is automated in such a way that the five or ten most prominent ions of every spectrum are subjected to further MS/MS fragmentation. The accuracy of these instruments has improved considerably in the last few years; 2.5 parts per million (ppm) are now easily achieved. This accuracy is very useful for the study of proteomes from unknown genomes and the *de-novo* sequencing approaches (Lopez-Llorca et al., 2010). Although MALDI instruments are very useful for classical gel-based proteomics and protein identification, they are not the best option for the analysis of complex proteomes, since they lack multidimensionality. On the other hand, they are not particularly appropriate for the study of posttranslational modifications and more specifically phosphoproteome analysis. For these cases, other mass spectrometry techniques based on electrospray ionization coupled to liquid chromatography are more efficient.

7. ELECTROSPRAY IONIZATION OF PROTEINS AND TANDEM MASS SPECTROMETRY

In electrospray ionization, proteins present in aqueous buffers are desolvated into a gas phase when flowing through a narrow capillary and are subjected to high voltage. Contrary to other ionization techniques, such as fast atom bombardment (FAB), electrospray ionization (ESI) is more efficient and

robust for the ionization of large organic molecules, including peptides. Therefore, it has become widely used in proteomics. Basically, the mechanism underlying ESI is that the application of voltage to a capillary (made of metal or silica), at atmospheric pressure and containing proteins or peptides in solution, produces charged tiny droplets that form a spray. As droplets evaporate, charged peptides pass from the aqueous phase to the gas phase. 7.1. Basic Concepts of Electrospray-Based Mass Analysers The basic principle of electrospray-based mass spectrometers is that ions, produced in an ESI source, can be isolated as a function of voltage applied to opposite electrodes and radiofrequency. By controlling these variables, ions can be stored, fragmented, or forced to hit a detector in order to measure their mass-to-charge ratio (m/z). Once peptides are charged and in the gas phase, they enter the mass spectrometer with the aid of rotary vacuum pumps. Basic amino acid residues become protonated when solved in low acidic buffers. Therefore, proteins and peptides in low acidic buffers become multiply charged and can be analysed as positive ions. ESI sources are usually coupled to ion traps, triple quadrupoles, and Q-TOF instruments. Ion traps and quadrupoles are able to isolate ions of interest and perform tandem mass spectrometry (MS/MS or MS) at high speed and resolution rates. In order to choose what kind of mass analyser is more n appropriate for a certain experimental approach, three issues need to be considered: mass resolution, scan speed, and accuracy of mass measurement. Ion traps are composed of a three dimensional (3D) chamber (3D traps or linear ion, where ions can be stored, isolated, and fragmented in order to achieve structural information on those. Ion traps are capable of performing multiple stages of precursor ion fragmentation (MS^n, where n = number of sequential n fragmentation) in a single experiment. Linear ion traps are characterized by higher ion capacity and scan rates, compared to 3D ion traps. Triple quadrupoles can be used to detect, isolate, and fragment specific precursor ions that produce a selected product ion of interest, such as a neutral loss. During the last few years, ultrahigh resolution equipment has been developed, including Orbitrap or Fourier transform (FT) ion cyclotron resonance technologies, which are able to reach up to sub-ppm mass accuracy. 7.2. Mass Determination of Intact Proteins A direct measurement of the molecular weight of proteins can be performed with ESI-equipped mass spectrometers. In this approach, proteins of interest are resuspended in low acidic buffers (0.1% acetic or formic acid) and placed inside small capillaries with metallic tips to facilitate the appearance of a spray with the aid of voltage. Protein ions spraying from the ESI source are normally detected as a mixture of different charge states, which is called a charge envelope. The mass determination is performed by deconvoluting the set of mass-to-charge ratios, corresponding to consecutive charged species, detected for a specific protein, with specific algorithms such as MagTran (free software available at http://www.geocities.com/SiliconValley/Hills/2679/magtran.html) or Maxent software (http://www.cs.princeton.edu/~schapire/maxent/). Mass determination of intact proteinscan routinely be performed on purified proteins in solution. Protein purifications using chromatographic procedures usually contain certain amounts of salt or other impurities that can dramatically affect the mass measurement of the protein of interest. Desalting steps prior to mass determination are carried out by dialysis or by reverse phase chromatography with hydrophobic columns (C2, C4, or C8 depending on the size of the protein). Using Fourier transform ion cyclotron resonance (FTICR), the mass determination of whole intact proteins ranging from 10,000 to 100,000 Da may be provided with an accuracy near 1 ppm. Protein masses of up to 229 kDa have also been reported. In some cases, this data can lead to unambiguous protein identification. However, the mass of the intact protein is compromised by posttranslational modifications or partial proteolysis of the purified protein. In that case, ions corresponding to the intact protein are subsequently fragmented in the mass spectrometer, yielding the molecular masses of both the protein and the fragment ions. If a sufficient number of informative

fragment ions are observed, this analysis can provide a complete description of the primary structure of the protein and reveal all of its primary structure and its modifications. Many proteomic studies, including the ability to fragment and analyse large intact proteins for protein characterization, protein–protein interaction, and chemical cross-link location inside proteins, are called top–down approaches. A second widely used alternative is the bottom–up approach, in which complex mixtures of peptides are analysed by mass spectrometry in two stages. In the first stage, the masses of peptides are determined; in a second stage, peptides are fragmented to produce information on the sequence of the peptides to reconstruct the primary structure of the original protein. 7.3. Peptide Bond Fragmentation Using Collision Energy Peptide structure determination using mass spectrometry is based on ionization of peptides and subsequent fragmentation of the peptide amide bonds after collision induced dissociation (CID) by controlled collision with an inert gas (such as He or Ar). The collision of ionized peptide molecules against the inert gas inside a collision chamber is sufficient to overcome the energetic threshold of the amide bonds. After fragmentation, the series of fragment ions generated from the N and C termini, respectively, are scanned, which leads to the elucidation of the primary structure backbone of the peptide. The series of fragment ions derived from the N terminus are designated as a, b, or c ions, whereas the ions derived from the C terminus are x, y, or z ions. This is a nomenclature widely accepted for the fragmentation of peptides, first proposed in 1984 and modified four years later. 8. MULTIDIMENSIONAL LIQUID CHROMATOGRAPHY COUPLED TO MASS SPECTROMETERS 8.1. Multidimensional LC-ESI-Q-TRAP Here we focus on the analysis of complex mixtures of peptides by coupling an ESI source at the end of a separation step using high performance liquid chromatography (HPLC). Although ESI was originally designed to be performed at a few microliters per minute flow rate (micro HPLC), separation techniques have rapidly evolved to run at nano-flow rates below 100 nL per minute using nanocolumns with diameters ranging from 1.5 to 3 μm. The major advantage of using nano-LC flow rates is the significant increase in the sensitivity of the identification. Peptide mixtures can routinely be separated using reverse phase (RP) chromatographic columns packed with, normally, C18. Peptides elute as independent chromatographic peaks as long as an increasing concentration of acetonitrile is pumped through the reverse phase column. As soon as the peptides elute from the RP column and reach the ESI source, ions are produced and can readily be scanned and fragmented. If the mixture of peptides of interest is more complex, a pre-fractionation can be achieved by coupling a strong cation exchange (SCX) column before the RP. This is called two-dimensional liquid chromatography (2D-LC) or multidimensional protein identification (MudPit). In this approach, after injecting the peptide mixture, increasing concentrations of salt (normally from 0.05 mM up to 2 M NaCl or ammonium formate) are pumped first through the SCX column. The eluting peptides from each salt concentration go through a second separation through the RP column. Chromatographic runs from 30 min up to 6 hours can be performed, depending on the complexity of the peptide mixture. Modern HPLC instruments allow one to configure this experimental approach online, which permits a comprehensive identification of proteins derived from tryptic digestions with a high degree of automation. Nevertheless, if high concentrations of peptides need to be analysed, the first fractionation using SCX may be performed offline using capillary or even preparative columns and salt gradients, instead of salt plugs. The development of C18 beads smaller than 2 μm, together with ultrahigh pressure liquid chromatography (UPLC) has led to better separation and resolution during chromatographic runs and has been shown to be very advantageous for both peptide and metabolite screening. 8.2. LC-MALDI-TOF(-TOF) The same principle of "divide and conquer" can be applied for MALDI-TOF analysis. Complex mixtures of peptides can be separated through RP columns and directly spotted as independent spots onto the surface of a MALDI plate, where an appropriate matrix is added

for subsequent analysis. This process can be automated using specific robots such as Probot (LC-Packings). Contrary to what happens during peptide mass fingerprint analysis, where all the peptides derived from a unique protein are spotted in a single position, in this new approach the peptides belonging to the same protein are usually spotted in consecutive positions of the MALDI plate. Hence, a correction in the acquisition and search algorithm needs to be introduced in order to take into account this possibility. TOF instruments are characterized by a high sensitivity and accuracy to measure the m/z ratios of peptides quickly and reliably, but when coupled to peptide fractionation using HPLC, these capabilities significantly improve.

9. DATABASE SEARCH AND USE OF BIOINFORMATIC TOOLS FOR PROTEIN IDENTIFICATION

9.1. Search Engines Against Nucleotide and Amino Acid Databases

The analysis of even a single protein tryptic digest yields a vast amount of fragmentation data after analysis with a mass spectrometer. Manual sequencing of peptides and subsequent identification of the original protein would be a major constraint. Several search engines have been developed to help in this task, including MS-Fit, MASCOT, OMSSA, PHENYX, SEQUEST, and X! Tandem, among others. Basically, search engines compare the experimental mass-to-charge ratios measured by mass spectrometers with in silico mass-to-charge ratios derived from the fragmentation of amino acid sequence databases. Nucleotide databases can also be used for protein identification as long as protein sequence can be derived from nucleotides. Most proteomic studies are therefore biased toward the identification of peptides and proteins contained in databases (either as nucleic acid or as amino acid sequences). Hence, most "shotgun" protein identification experiments, aimed at the successful identification of as many proteins present in the sample as possible, are constrained to organisms whose genomes have been sequenced, or at least partially sequenced. Annotated databases, including functional information of proteins as well as their modifications, are rapidly improving and new nucleotide databases from several organisms are also emerging. The Human Genome Project and other genome sequencing projects are turning out in rapid succession the complete genome sequences of specific species and thus, in principle, the amino acid sequence of every protein potentially encoded.

9.2. Use of Customized Databases for Tagged Recombinant Proteins and De Novo Sequencing

Mass spectrometry is also useful to characterize the quality and sequence content of recombinant proteins that contain peptide tags (His-tag, GST-tag, etc.) for their purification after expression in E. coli or other systems. In these cases, where the exact amino acid sequence of the recombinant protein is known, customized databases can be prepared in order to identify all the proteolytic peptides of the protein of interest. The goal is to detect the possible occurrence of deletions or partial degradation from the N or C terminus or the loss of the tags. An alternative to protein identification using search algorithms is the identification of unknown proteins, from unknown genomes, by *de-novo* sequencing. There are two things that need to be considered in *de-novo* sequencing experiments. First, better accuracy of the measured mass leads to better sequence information from the fragmented peptide. Second, appropriate software (i.e., PEAKS, http://www. bioinfor.com) needs to be used to gather all the information detected after fragmentation of the peptides of interest (Casado-Vela et al., 2010; Lopez-Llorca et al., 2010; Selles-Marchart et al., 2008).

10. IDENTIFICATION OF POSTTRANSLATIONAL MODIFICATION OF PROTEINS

Many critical events involved in cellular responses (control of enzymatic activity, protein–protein interaction, cellular localization, signal transduction) are mediated by changes in posttranslational protein modifications rather than transcriptional changes. More than 300 modifications have been reported on proteins and peptides both in prokaryotic and eukaryotic cells (http://www.abrf.org/index.cfm/dm.home). Currently, mass spectrometry is considered as the most appropriate approach for the identification and characterization of posttranslational modification of proteins. It can unambiguously identify the occurrence of posttranslational modifications at the residue level. Almost 2% of the human genome encodes protein kinases, and it is estimated that protein phosphorylation can affect up to one-third of all proteins, due to the relative abundance of phosphorylated residues near 1% pTyr, 12% pThr, and87% pSer. Phosphorylation analysis of proteins is made by cleavage into their constituent peptides using endopeptidase digestion, generally trypsin, separation, and analysis of the resulting peptides by LC-MS/MS. The high sensitivity and the ability to isolate and fragment peptides of ion traps and triple quadrupoles serves as a quick and reliable method for the detection of phosphorylations. The use of nano-HPLC coupled to mass spectrometers with high ion capacity and fast scan rates enhances the detection of phosphorylated peptides and enables the determination of the phosphorylation at the residue level. In order to detect phosphorylation sites on protein digests, data-dependent mass analysis is frequently used to trigger MS 3 scans, where specific neutral losses (98.0,49.0, 32.7 Da) are observed. This is called data-dependent neutral loss (DDNL) analysis. Nevertheless, phosphopeptide analysis with mass spectrometry has to face the inherent problem of low concentrations inside the peptide mixture. Therefore, several enrichment strategies have been developed to favour their analysis using mass spectrometry. These methods include the use of immobilized metal ion (Fe 3+ or Ga 3+) affinity chromatography (IMAC), zirconium dioxide, or titanium dioxide. Phosphopeptides can also be collected in a few fractions by strong cation exchange (SCX) chromatography using a salt gradient (e.g., 1 mM to 1 M NaCl or ammonium formate). Another strategy for tyrosine kinases is the use of immunological methods based on anti- phosphotyrosine antibodies, which are sufficiently specific for the enrichment of phosphotyrosine-containing peptides by affinity or immunoprecipitation.

11. PROTEIN QUANTITATION

Until recently, the quantitative study of the pool of proteins expressed in biological samples using differential display was restricted to 2D gel analyses. Protein quantitation using the 2D gel-based approach consists of the measurement of the volume (area in pixels by intensity of the spots) of spots identified as a single protein in different gels. A significant statistical variation of the amount of protein measured in gels corresponding to drug treatments or specific biological alterations of interest with respect to a control can therefore be expressed in terms of quantitative variation. An alternative non-gel-based approach has been the use of isotope coded affinity tags (ICAT) for the quantitative study of protein expression at the proteome level. Recently, an improved approach, analogous to ICAT, has been developed. It is called iTRAQ (Applied Biosystems) and consists of a range of up to eight different amine-specific, stable isotope tags that label all peptides in up to eight different biological samples, enabling simultaneous identification and quantitation of the whole set of proteins. A more recent and powerful alternative is the stable isotope labeling with amino acids in cell culture (SILAC), which can be used mainly in cell cultures because it is based on the metabolic incorporation of labeled amino acids containing heavy isotopes (i.e., 12 C, or 15 N instead of 14 2 H instead of 1 H, N) in living cells. After at least five cell population turnovers, complete labeling of all the proteins is achieved. The cell cultures to be compared are mixed and the changes in the expression of proteins versus the control can be

quantified because the peptides containing the different isotopes co-elute simultaneously from the C18 column, leading to a mass shift that is observed in the mass spectrometer. A major limitation of the iTRAQ and SILAC approaches is the cost of the reagents necessary to perform these experiments. A more economical approach for isotopic peptide labeling and quantification is based on the enzymatic 16 O/ 18 O labeling. This methodology is based on the trypsin catalysed oxygen exchange at the free terminus from a peptide.

CONCLUSION

The measurement of changes in gene expression as a response to the appearance of an alteration/disease is not sufficient in many cases. It is estimated that only 50% of the measured increments of mRNA are directly correlated to an increase in the amount of the corresponding expressed protein, as observed in human, yeast, bacteria, and other cell line. Therefore, changes detected at the protein level using proteomic techniques could be more indicative of the existence of alterations and modifications. The classical approach to study proteins and biological processes, which was based on purification to homogeneity followed by biochemical assays of the specific activity of the purified proteins, has been replaced today by a tremendous increase in the application of proteomic technologies to the study and characterization en masse of proteins and complete proteomes. Although the analysis of full proteomes remains a formidable task, the new generation of multidimensional chromatographic steps coupled to the unprecedented resolution of the new mass spectrometers allows optimism for the future. Technological advances have

Casado-Vela, J., Muries, B., Carvajal, M., Iloro, I., Elortza, F., & Martinez-Ballesta, M. C. (2010). Analysis of root plasma membrane aquaporins from Brassica oleracea: post-translational modifications, de novo sequencing and detection of isoforms by high resolution mass spectrometry. *J Proteome Res, 9*(7), 3479-3494. doi: 10.1021/pr901150g.

Lopez-Llorca, L. V., Gomez-Vidal, S., Monfort, E., Larriba, E., Casado-Vela, J., Elortza, F., et al. (2010). Expression of serine proteases in egg-parasitic nematophagous fungi during barley root colonization. *Fungal Genet Biol, 47*(4), 342-351. doi: 10.1016/j.fgb.2010.01.004.

Selles-Marchart, S., Luque, I., Casado-Vela, J., Martinez-Esteso, M. J., & Bru-Martinez, R. (2008). Proteomics of multigenic families from species underrepresented in databases: the case of loquat (Eriobotrya japonica Lindl.) polyphenol oxidases. *J Proteome Res, 7*(9), 4095-4106. doi: 10.1021/pr700687c.