# PROTEOME ANALYSIS OF NON-MODEL PLANTS: A CHALLENGING BUT POWERFUL APPROACH

**Sebastien Christian Carpentier,[1]* Bart Panis,[1] Annelies Vertommen,[1] Rony Swennen,[1] Kjell Sergeant,[2] Jenny Renaut,[2] Kris Laukens,[3] Erwin Witters,[4,5] Bart Samyn,[6] and Bart Devreese[6]**

[1]*Division of Crop Biotechnics, K.U.Leuven, Leuven, Belgium*
[2]*Department of Environment and Agrobiotechnologies, CRP Gabriel Lippmann, Belvaux, Luxembourg*
[3]*Intelligent Systems Laboratory, University of Antwerp, Antwerpen, Belgium*
[4]*Centre for Proteome Analysis and Mass Spectrometry, University of Antwerp, Antwerp, Belgium*
[5]*Flemish Institute for Technological Research, VITO, Mol, Belgium*
[6]*Laboratory of Protein Biochemistry and Biomolecular Engineering, Ghent University, Gent, Belgium*

*Received 11 December 2007; received (revised) 25 February 2008; accepted 25 February 2008*

*Published online 1 April 2008 in Wiley InterScience (www.interscience.wiley.com) DOI 10.1002/mas.20170*

*Biological research has focused in the past on model organisms and most of the functional genomics studies in the field of plant sciences are still performed on model species or species that are characterized to a great extent. However, numerous non-model plants are essential as food, feed, or energy resource. Some features and processes are unique to these plant species or families and cannot be approached via a model plant. The power of all proteomic and transcriptomic methods, that is, high-throughput identification of candidate gene products, tends to be lost in non-model species due to the lack of genomic information or due to the sequence divergence to a related model organism. Nevertheless, a proteomics approach has a great potential to study non-model species. This work reviews non-model plants from a proteomic angle and provides an outline of the problems encountered when initiating the proteome analysis of a non-model organism. The review tackles problems associated with (i) sample preparation, (ii) the analysis and interpretation of a complex data set, (iii) the protein identification via MS, and (iv) data management and integration. We will illustrate the power of 2DE for non-model plants in combination with multivariate data analysis and MS/MS identification and will evaluate possible alternatives.*
 Mass Spec Rev 27:354–377, 2008

**Keywords:** *2DE; cross-species identification; data analysis; data integration; de novo peptide sequencing; plant proteomics; methodology; MS/MS; non-model organism; sample preparation; sequence similarity search*

## I. INTRODUCTION: GENE EXPRESSION PROFILING AND UNDERSTANDING GENE FUNCTION IN NON-MODEL PLANT SYSTEMS

### A. Non-Model Plant Systems and Proteomics

In past decades, the term ''model organism'' has been narrowly applied to species that facilitate experimental laboratory research because of their particular size and generation time. Research communities have focused on model organisms to gain insight into some general principles that underlie various disciplines. The first and only classical plant model, *Arabidopsis thaliana*, is ideal for laboratory studies. It has a short life cycle, a small size, an important production of seeds and a relatively small genome that is completely sequenced (Arabidopsis Genome Initiative, 2000). To date 67% of the Arabidopsis genome is covered by annotated genes (TAIR7 genome statistics). However, with the recent increase in the number of genome-sequencing projects, the definition of model organism has broadened (Hedges, 2002). Advances in high-throughput and computational technologies have resulted nowadays in the genome sequencing of hundreds of organisms across the three domains of life. Currently, 645 genomes are considered to be completely sequenced and publicly available and the number is still growing (www.genomesonline.org). Those species fall under the new and broad definition of ''model organism.'' In most cases, economics has had an important impact on the choice of organism to study. The green plants or *Viriplantae* are largely under-represented with only two plant genomes completed, publicly available and reasonably well annotated: *A. thaliana* (thale cress, ~120 Mb, 5 chromosomes) and *Oryza sativa* (japonica cultivar-group) (rice, ~450 Mb, 12 chromosomes). The manageable size of the rice genome (though almost 4 times the size of Arabidopsis) and the fact that rice is a staple food for half of the world's population led to the effort of unraveling its sequence (Barry, 2001; Goff et al., 2002; Yu et al., 2002). The NCBI Plant Genomes Central considers also *Medicago truncatula* (barrel medic, ~500 Mb and 8 chromosomes) and *Populus trichocarpa* (black cottonwood, ~550 Mb, 19 chromosomes) (Tuskan et al., 2006) as completed

Contract grant sponsor: Belgian National Fund for Scientific Research (FWO-Flanders); K.U. Leuven.
*Correspondence to: Sebastien Christian Carpentier, Faculty of Bioscience Engineering, Division of Crop Biotechnics, K.U. Leuven, Kasteelpark Arenberg 13, box 2455, 3001 Leuven, Belgium.
E-mail: sebastien.carpentier@biw.kuleuven.be

large scale sequencing projects. All those new model plants have a relatively small genome size but are not necessary ideal as a laboratory model. The NCBI Plant Genomes Central expects that the genome sequencing of *Lotus japonicus* (lotus), *Manihot esulenta* (cassava), *Solanum lycopersicum* (tomato), *Solanum tuberosum* (potato), *Sorghum bicolor* (sorghum), and *Zea mays* (corn) will be completed and the data will be publicly available in the near future. However, there are approximately 300,000 known species of land plants but the model plants represent only a handful of species and families. Even the arrival of new model plants cannot reflect the biodiversity of the plant kingdom and all the economic or agricultural interests. The genome size of non-model plants is in general large and complex. Numerous crops and plant species have different levels of ploidy. Cultivated sugar cane (*S. officinarum*) is a good example. It is a hybrid of different species and it has a complex octoploid genome with chromosome number ranging from $2n = 70$ to 140 (Asano et al., 2004).

Gene expression profiling and understanding gene function can be approached via several techniques. RNA based system biology approaches have largely been applied to the classical model organisms. These so-called transcriptomics approaches are extremely powerful and highly automated, allowing massive screening of hundreds of genes simultaneously. However, the success of those approaches depends greatly on the genomic progress. Successful techniques like cDNA microarrays, cDNA amplified fragment length polymorphism (AFLP) and serial analysis of gene expression (SAGE) are in practice restricted to model organisms or species that are already characterized to great extent. The power of those transcript-based techniques is lost in non-model organisms due to the lack of genomic information or due to the sequence divergence from a related model organism. Gene sequences are rarely identical from one species to another and orthologous genes are usually riddled with nucleotide substitutions.

An alternative for examining gene expression is studying its end products, the proteins. Protein sequences are more conserved making the high-throughput identification of non-model gene products by comparison to well known orthologous proteins quite efficient (Liska & Shevchenko, 2003). Moreover, it is important to recognize that there is a possible discrepancy between the messenger (transcript) and its final effector (mature protein). As most biological functions in a cell are executed by proteins rather than by mRNA, transcript expression profiling does not always provide pertinent information for the description of a biological system. Several post-transcriptional and post-translational control mechanisms such as the translation rate, the half-lives of mRNAs and proteins, protein modifications and intercellular protein trafficking, have an important influence on the phenotype (Mata, Marguerat, & Bahler, 2005; Higashi et al., 2006).

As a matter of fact, also most of the proteome studies are still performed on model plants (Jorrín, Maldonado, & Castillejo, 2007). Insights from the model Arabidopsis will undoubtedly boost crop science but Arabidopsis is not a crop and will never feed the world (Adam, 2000). Also rice, as a new model, has become a cornerstone for crop proteomics (Agrawal & Rakwal, 2006). Nonetheless, there is a great need for proteomics of non-model plants and crops. Some features and processes are unique and cannot be approached via a model plant. Woody plants for example, are perennials with a quite long life cycle and have special features to offer like survival to harsh winters and wood formation. Renaut et al. (in press) have applied proteomics to peach trees to understand the mechanisms triggered by low temperatures and a short photoperiod. A proteomic study of wood formation has recently been performed (Gion et al., 2005; Celedon et al., 2007).

The interests of crop proteomics in general are (i) to get insight into the different varieties and their performance toward yield, specific pathogens, abiotic constraints, fruit set, low input systems, etc., (ii) to develop safe and high quality food, (iii) to reduce the impact of agriculture on the environment, and (iv) to fulfill needs for food, feed, and industry. Our intention in this is not to give a full review of all the proteomic studies carried out on non-model plant species. A review of the applications of proteomics to crop species has been published recently (Salekdeh & Komatsu, 2007). Our goal is rather to discuss the specific technical challenges for studying non-model plants via proteomic approaches.

## B. Proteomics and Technology

A normal proteomics workflow consists of (i) protein extraction, (ii) protein (peptide) separation and quantification, (iii) protein identification, and (iv) data integration. An array of approaches has been developed to address proteomics. There are two main complementary approaches: the so-called ''gel-based'' approach and the ''gel-free'' approach. Both approaches differ in the way (poly)peptides are isolated (extracted), separated, and detected and consequently, each of them covers a typical subset of proteins. Indeed, the proteome of a cell or tissue at any specific time point is extremely complex and diverse. Any available technique is only able to focus on a sub-fraction of the protein set due to the complex chemical nature of proteins and to the large dynamic range.

The gel-based approach is the cornerstone of proteome analysis and has an unequalled resolving power for separation of complex protein mixtures. Two dimensional gel electrophoresis (2DE) is a complete methodology, resulting in a qualitative and quantitative high resolution image of intact proteins that can provide a good overview of different isoforms and post-translational modifications. The classical 2DE protocol separates denatured proteins according to two independent properties: isoelectric point (pI) (IEF: iso-electric focusing) and molecular size [more often referred to as molecular weight (MW)]. In order to separate proteins under denaturing conditions in the first dimension, proteins are solubilized in the presence of high concentrations of chaotropes, a reductant and a neutral detergent. The use of a detergent in conjunction with chaotropes is of paramount importance and is decisive for the subset of proteins that can be analyzed.

Unfortunately, 2DE is difficult to automate and the control of technical variation greatly depends on a scientist's skills. Since total automation is the ultimate objective for every high-throughput method, gel-free approaches were developed. Yates and colleagues were among the pioneers to explore the use of liquid chromatography coupled to electrospray ionization tandem mass spectrometry (LC/MS/MS) in an attempt to realize

automated high-throughput proteomics (McCormack et al., 1997; Ducret et al., 1998; Link et al., 1999). Most gel-free approaches use a bottom-up strategy where proteins are first digested with a proteolytic enzyme and the obtained complex peptide mixture is then separated via reversed-phase (RP) chromatography coupled to a tandem mass spectrometer. The whole dataset of acquired tandem mass spectra is subsequently used to search protein databases and to link the individual peptides to the original proteins. This concept is only successful, however, when identifying proteins in relatively simple mixtures. The problem of resolution was anticipated by introducing multidimensional chromatography. Different concepts such as Direct Analysis of Large Protein Complexes (DALPC) (Link et al., 1999) and MUltiDimensional Protein Identification Technology (MudPit) (Washburn, Wolters, & Yates, 2001) have been described. Although a great improvement, the resolving power was still limited, the eluate complexity exceeded the analytical capacity of most mass spectrometers and the method was not quantitative (Regnier & Julka, 2006). Aebersold and colleagues tackled these problems and described an approach for accurate quantification and identification of individual proteins using Isotope-Coded Affinity Tags (ICAT) (Gygi et al., 1999). Further improvements were introduced (Gygi et al., 2002; Zhou et al., 2002) and many groups adapted the principle of labeling with stable isotopes, generating different approaches with their own strengths and weaknesses (Moritz & Meyer, 2003; Regnier & Julka, 2006). In general, such peptide centred bottom-up approaches have the disadvantage that both qualitative and quantitative information on protein isoforms and differential post-translational modifications are lost.

Cross-species identification is the only option for protein identification whenever a genome is poorly characterized (Wilkins & Williams, 1997; Lester & Hubbard, 2002; Mathesius et al., 2002; Liska & Shevchenko, 2003; Witters et al., 2003; Samyn et al., 2007). In this approach, proteins are identified by comparing peptides of the proteins of interest to orthologous proteins of species that are well characterized. All bottom-up gel-free strategies are peptide-based separation techniques and have the disadvantage to lose the connectivity between peptides derived from the same protein. Haynes and Roberts (2007) reviewed the possibilities of using a shotgun approach in plants. They acknowledge that a completely sequenced genome is essential for a peptide-based separation and that shotgun proteomics is currently only applicable in model plants. 2DE-based proteome analysis is at present the most powerful option for non-model organisms: it is a protein-based separation and quantification technique where the connectivity between protein derived peptides is preserved with high confidence allowing to compare multiple peptides per protein as a diagnostic assembly.

## II. PROTEIN SEPARATION AND ANALYSIS

Plant proteome research groups are often confronted with sample preparation issues and are forced to explore the limits. Many important technical sample preparation improvements have been developed by plant research groups (Westermeier, 2006) (see Table 1). The major limitations of proteome analysis in general are associated with the heterogeneity of proteins in terms of physicochemical properties and the huge differences in abundance (Wilkins et al., 1998b). Depending on the origin, an actual proteome can have a dynamic range of 7–12 orders of magnitude and only a few orders can be analyzed simultaneously with the current proteomics platforms (Corthals et al., 2000). Although classical 2DE is up till now unequalled for resolution and one of the few general methods able to separate protein isoforms, it is still not appropriate to analyze hydrophobic proteins (Gorg, Weiss, & Dunn, 2004). Further limits are associated with the size of the proteins and the extreme pI of certain proteins. Streaking and the presence of artifactual spots in the basic region of a 2DE gel is a well-known problem and has been addressed to some extent (Herbert et al., 2001; Hoving et al., 2002; Olsson et al., 2002). The analysis of extreme basic proteins is a problem mainly due to hydrolysis of acrylamide under extreme basic conditions ($pH > 10$). Different optimization steps with respect to pH engineering and gel composition were introduced to obtain reproducible basic IPG strips (Gorg et al., 1997). High MW proteins (>150 kDa) are poorly transferred from the first to the second dimension and the classical one-dimensional Laemmli SDS–PAGE system has not enough resolving power below 10 kDa (Schägger & von Jagow, 1987). Schägger and von Jagow presented in 1987 a new method. The superiority of this method, especially for the separation of proteins ranging from 5 to 30 kDa, is mainly based on the introduction of tricine as trailing ion and the introduction of an additional spacer gel. We focus here on approaches applied or applicable in plants.

### A. Adaptations to the Iso-Electric Focusing-Based Two-Dimensional Gel Electrophoresis to Push the Limits

#### 1. Protein Extraction

Most plant tissues are not a ready source for protein extraction and need specific precautions. The cell wall and the vacuole make up the majority of the cell mass, with the cytosol representing only 1–2% of the total cell volume. Subsequently, plant tissues have a relatively low protein content compared to bacterial or animal tissues. The cell wall and the vacuole are associated with numerous substances responsible for irreproducible results such as proteolytic breakdown, streaking and charge heterogeneity. Most common interfering substances are phenolic compounds, proteolytic and oxidative enzymes, terpenes, pigments, organic acids, ionic species, and carbohydrates.

Most of the agricultural interesting species contain high levels of interfering compounds and several specific protocols have been developed (e.g., see Table 1). Banana (*Musa* spp.) for example contains extremely high levels of oxidative enzymes (e.g., polyphenol oxidase) (Gooding, Bird, & Robinson, 2001; Wuyts, De Waele, & Swennen, 2006) and phenolic compounds (simple phenols, flavonoids, condensed tannins, lignin), and high levels of latex and carbohydrates. Phenolic compounds reversibly combine with proteins by hydrogen bonding and irreversibly by oxidation followed by covalent condensations (Loomis & Bataille, 1966), leading to charge heterogeneity and

**TABLE 1.** List of articles focusing on the sample preparation for 2DE analysis of some important agricultural species and their major outcome

| Reference | Aim | Major outcome | Recalcitrant species/tissue |
|---|---|---|---|
| Hari 1981 | Optimization of protocol | Acetone precipitation | Tobacco |
| Wu & Wang 1984 | Optimization of protocol | TCA precipitation | Bean/anther<br>Petunia/anther |
| Damerval et al. 1986 | Optimization of protocol | TCA/acetone precipitation | Wheat |
| Hurkman & Tanaka 1986 | Comparison of different methods | Phenol extraction /ammonium acetate precipitation | Barley/root |
| Granier 1988 | Comparison of different methods | TCA/acetone precipitation | Maize/leaf |
| Wang et al. 2003 | Optimization of protocol | Combination of TCA/acetone precipitation and phenol extraction | Olive/leaf |
| Saravanan & Rose 2004 | Comparison of different methods | Phenol extraction /ammonium acetate precipitation | Avocado/fruit<br>Banana/fruit<br>Tomato/fruit<br>Orange/fruit |
| Carpentier et al. 2005 | Downscaling for small amounts of starting material and comparison of different methods | Phenol extraction /ammonium acetate precipitation | Banana/leaf + meristems<br>Potato/ shoots<br>Strawberry/ shoots |
| Delaplace et al. 2006 | Comparison of two methods | Hot SDS extraction | Potato/tuber |
| Song et al. 2006 | Comparison of different methods | TCA/acetone precipitation and hot SDS buffer | Apple/fruit |
| Valcu & Schlink 2006 | Testing of several chaotrope/detergent combinations | Species and tissue dependent combinations of detergents | Norway spruce/needles<br>Beech/leaf and root |
| Vincent, Wheatley, & Cramer 2006 | Comparison of different methods | Phenol extraction /ammonium acetate precipitation | Grape/fruit |

(*Continued*)

**TABLE 1.** (*Continued*)

| | | | |
|---|---|---|---|
| Wang et al. 2006 | Developing a universal rapid method | Combination of TCA/acetone precipitation and phenol extraction | Bamboo/leaf<br>Grape/leaf<br>Iris/leaf<br>Olive/leaf<br>Lemon/leaf<br>Pine/leaf<br>Redwood/leaf<br>Sugar cane/leaf<br>Tobacco/leaf<br>Possidonia grass /leaf<br>Apple/fruit<br>banana/fruit<br>grape/fruit<br>kiwi/fruit<br>orange/fruit<br>pear/fruit<br>tomato/fruit |
| Yao, Yang, & Liu 2006 | Optimization of protocol | Phenol extraction /ammonium acetate precipitation and PVP and SDS in first extraction buffer | Cotton/fiber |
| Carpentier et al. 2007c | Safe storage and transport at room temperature of lyophilized tissue | Lyophilized tissue samples can perfectly be analyzed after phenol extraction /ammonium acetate precipitation | Banana/leaf |
| Maserti et al. 2007 | Optimizing a method | TCA/acetone precipitation | Citrus/leaf |
| Pedreschi et al. 2007 | Optimizing a method | Phenol extraction /ammonium acetate precipitation | Pear/fruit |
| Xie et al. 2007 | Optimizing a method | Chloroform isoamylalcohol fractionation combinated with phenol extraction /ammonium acetate precipitation | Bupleurum/root |
| Zheng et al. 2007 | Comparison of different methods | Phenol extraction /ammonium acetate precipitation | Apple/fruit<br>Strawberry/fruit |

streaks in the gels. Carbohydrates and latex interfere with the electrophoresis and can block gel pores causing precipitation and extended focusing times, resulting in streaking and loss.

The original 2DE protocol contains only a single sample preparation step, that is, denaturing extraction of cellular proteins in a lysis buffer (O'Farrell, 1975). This one step protocol is restricted to "clean" samples and is rarely useful for plant material. In the 1980s, much effort has been invested in the establishment of two dimensional gel electrophoresis sample preparation methods for plant tissue (Damerval et al., 1988; Granier, 1988; Meyer et al., 1988). The majority of the plant protocols introduce a precipitation step to concentrate the proteins and to separate them from the interfering compounds. Proteins are usually precipitated by the addition of high concentrations of salts (Cremer & Vandewalle, 1985), extreme pH (Wu & Wang, 1984), organic solvents (Hari, 1981; Wessel & Flugge, 1984; Flengsrud & Kobro, 1989; Schroder & Hasilik, 2006) or a combination of organic solvents and ions (Van Etten, Freer, & McCune, 1979; Damerval et al., 1986). The protein precipitation step can be preceded by a denaturing or non-denaturing extraction step and goes along with one or two washing steps to remove introduced salt ions and other remaining interfering substances.

The most commonly used method for extraction of plant proteins is the trichloroacetic acid (TCA)/acetone precipitation method (Damerval et al., 1986). TCA is a strong acid ($pK_a$ 0.7) that is soluble in organic solvents. The extreme pH and negative charge of TCA and the addition of acetone realizes an immediate denaturation of the protein, along with precipitation, thereby arresting instantly the activity of proteolytic enzymes (Wu & Wang, 1984). The main disadvantage of TCA precipitated proteins is that they are difficult to redissolve (Nandakumar et al., 2003). Moreover, the extreme low pH might create problems with alkaline chemical labeling methods used in 2DE (such as Cy dyes). We evaluated different plant protocols for the extraction of banana leaf and meristem proteins which resulted in an optimized phenol-based extraction procedure for small amounts of fresh weight (e.g., Fig. 1) (Carpentier et al., 2005) and for lyophilized tissues (Carpentier et al., 2007c). The extraction buffer has been designed to minimize enzymatic reactions and to remove as much interfering compounds as possible. It forms the aqueous phase containing carbohydrates, nucleic acids and cell debris. The extremely abundant (poly)-phenols (quinones) are eliminated by DTT to form thio-ethers. The phenol rich phase contains the proteins and some remaining interfering compounds such as lipids and pigments. The proteins are separated from the remaining interfering compounds via ammonium acetate/methanol induced precipitation. This "micro" phenol protocol is applicable to a lot of (plant) species. We applied it already successfully to other eukaryotic non-model species such as apple, pear, potato (Carpentier et al., 2005), stevia, and Trypanosoma (unpublished results).

Apart from the optimization of the extraction protocol, also protein solubilization is a critical factor. The introduction of

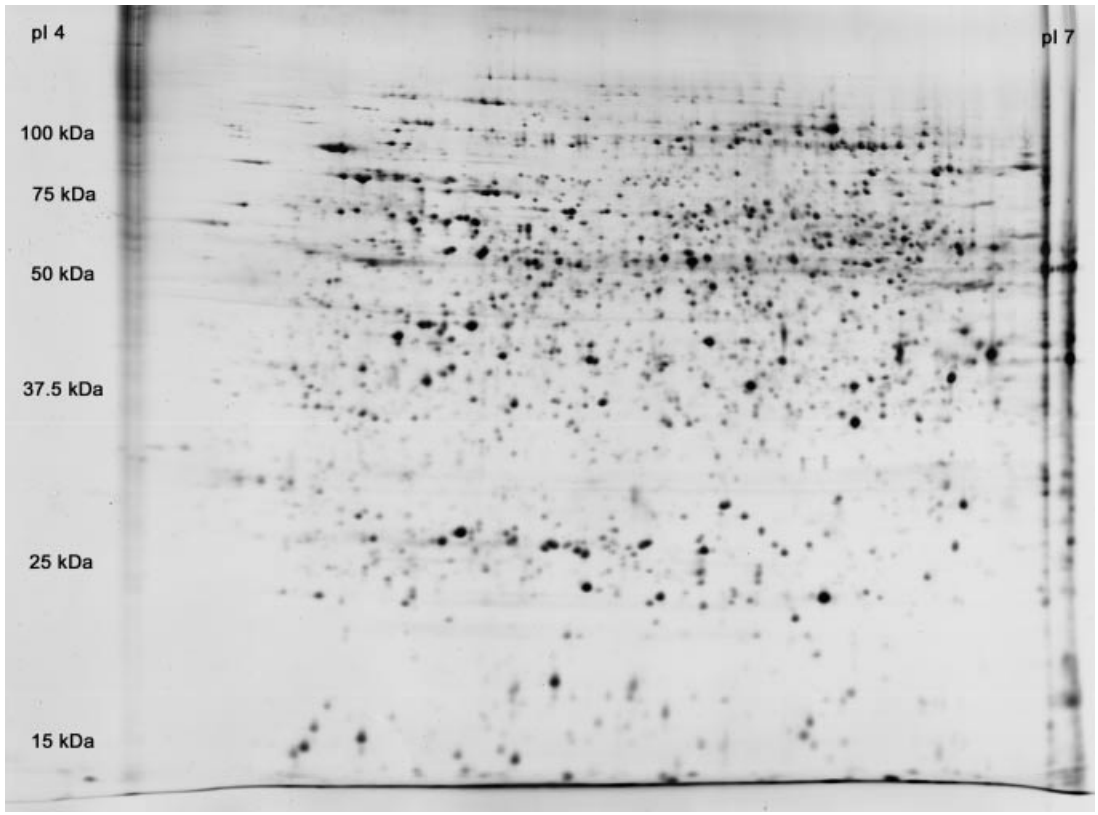


**FIGURE 1.** A representative gel of banana meristem proteins separated via 2DE and visualized via silver staining (24 cm gel, pI 4–7, 12.5% acrylamide).

thiourea by Rabilloud et al. (1997), in combination with urea and the neutral sulfobetaine detergent CHAPS was a noticeable improvement. Thiourea clearly enhances the chaotropic power but has to be used in combination with urea due to its poor solubility. Méchin et al. (2003) also explored the possibilities to increase the resolving power of 2DE gels for plant proteins. Apart from the successful chaotrope combination of 2 M thiourea and 7 M urea, their buffer ''R2D2'' is a combination of two reducing agents (DTT and TCEP) and two detergents (CHAPS and SB3).

### 2. *Fractionation and Enrichment Tools*

A wide variety of fractionation tools are available to cope with the issue of dynamic range. They are based on electrophoretic or chromatographic separation or physico-chemical properties to zoom in at a specific subset of proteins (Canut, Bauer, & Weber, 1999; Lopez, 2000; Righetti et al., 2005a,b).

Proteomes containing extremely abundant proteins like the ''green proteome'' (Rubisco) and the ''seed proteome'' (storage proteins) suffer from this problem considerably. Specific attempts for removing Rubisco from the proteome have been made by PEG fractionation (Kim et al., 2001; Xi et al., 2006) and commercial kits are available using specific antibodies. So far, publications of this commercial system are only available for plasma samples (Huang et al., 2005). Espagne et al. (2007) report that the electrophoretic behavior of the large subunit of Rubisco can be influenced by changing the composition of the extraction buffer. This has no direct impact on the dynamic range, but it enables the characterization of proteins that were previously masked by Rubisco. Similarly, Hurkman and Tanaka (2004) exploited the solubility properties of storage proteins, the dominant proteins in seeds, in different buffers to separate the abundant proteins from the less abundant ones.

The introduction of an extra dimension prior to isoelectric focusing allows loading of higher amounts of protein and thus facilitates the detection of less abundant proteins. A technical improvement came from the laboratory of A. Görg, who picked up an old idea of introducing a pre-fractionation step using neutral beads of the dextran Sephadex (Delincee & Radola, 1970; Radola, 1975; Gorg et al., 2002). A nice illustration of this preparative native isoelectric focusing, in combination with column chromatography and 2D PAGE, is the purification of glyoxysomal processing protease from watermelon (Helm et al., 2007).

Plant cells have characteristic subcellular components. Some components are relatively easy to isolate in a pure form, whereas others are easily susceptible to contamination from other compartments (e.g., endomembrane organelles like Golgi fractions). Lilley and Dupree (2006, 2007) give a recent overview on plant organelle proteomics. Physico-chemical pre-fractionation of different subcellular components is a powerful approach to focus on the subproteomes of the cell (e.g., see Table 2). Differential and isopycnic centrifugation using sucrose or Percoll gradients have been traditionally used to separate and purify plant organelles. Canut, Bauer, and Weber (1999) review the possibilities to separate plant membranes and organelles by electromigration techniques. Eubel et al. (2007) recently optimized the purification of mitochondria and combined differential centrifugation and Percoll density centrifugation with free-flow electrophoresis prior to 2DE.

Rolland et al. reviewed techniques to fractionate plant membrane proteins (Ephritikhine, Ferro, & Rolland, 2004; Rolland et al., 2006). A technique which is capable to fractionate selectively hydrophobic proteins is the chloroform/methanol fractionation. This simple and efficient strategy was developed to extract the most hydrophobic proteins from spinach (*Spinacia oleracea* L.) chloroplast envelope membranes (Joyard et al., 1982; Seigneurin-Berny et al., 1999; Ferro et al., 2000, 2002). It enriches hydrophobic proteins that, unfortunately, cannot easily be resolved by IEF-based 2DE. Santoni, Molloy, and Rabilloud (2000) review this problem very thoroughly and begin with the rhetorical question: ''Membrane proteins and proteomics: un amour impossible?'' Chloroform/methanol extracted proteins are frequently analyzed by one-dimensional SDS–PAGE (Fig. 2). This one-dimensional approach coupled to tandem MS proved to be most successful to separate and identify proteins from Arabidopsis (Ferro et al., 2003; Friso et al., 2004; Marmagne et al., 2004). However, non-model organisms are dependent on cross-species identification and the resolution of chloroform/methanol-1DE might not be sufficient. Indeed, after 1DE analysis of chloroform/methanol extracted proteins from spinach, Ferro et al. (2002) could not assign more than 40% of the tryptic peptides to a protein, leaving a significant amount of orphan peptides and, consequently a significant amount of unidentified proteins. Moreover, the limited resolution puts severe restrictions on the protein quantification. Therefore alternative electrophoresis tools have to be evaluated.

## B. Non-Model Plants and Membrane Proteins: Un Amour Impossible?

Membranes play a pivotal role in cell biology and are involved in signal transduction and stress monitoring, cell–cell communication, cellular and organellar trafficking and transport, and the formation of mitochondrial and plastidial electron transfer chains. From the studies performed on model organisms, it was estimated that transmembrane proteins represent 20–30% of the total proteome (Santoni, Molloy, & Rabilloud, 2000). As mentioned, classical 2D PAGE fails to resolve those proteins.

Some alternative two-dimensional electrophoresis techniques apply different ionic detergents (Hartinger et al., 1996; Buxbaum, 2003) or different acrylamide concentrations in the two dimensions (Rais, Karas, & Schagger, 2004). The anomalous migration of hydrophobic proteins in function of the acrylamide concentration was the basis for the dSDS-based two-dimensional gel electrophoresis (Akiyama & Ito, 1985). Proteins with an anomalous migration are dispersed around a diagonal of proteins with a normal migration. Rais, Karas, and Schagger (2004) improved resolution, especially of low MW proteins, by using Tricine-SDS–PAGE and addition of urea to the first dimension gel. Further optimization comprised the incorporation of glycerol and increased Tris concentrations into the gel solutions and the use of Bicine in stead of Tricine as trailing ion (Williams et al., 2006). Although dSDS was successfully applied to study highly hydrophobic proteins in mammalian (Rais, Karas, & Schagger, 2004; Burre et al., 2006) and bacterial (Williams et al., 2006)

**TABLE 2.** List of articles focusing on a plant subproteome

| Subproteome | Reference | Species |
|---|---|---|
| plasma membrane proteome and detergent resistant membrane regions | Marmagne et al. 2004 | arabidopsis |
| | Borner et al. 2005 | arabidopsis |
| | Morel et al. 2006 | tobacco |
| mitochondrial proteome | Brugiere et al. 2004 | arabidopsis |
| | Millar et al. 2005 | arabidopsis |
| | Bardel et al. 2002 | pea |
| chloroplast proteome | Seigneurin-Berny et al. 1999 | spinach |
| | Gomez et al. 2002 | pea, tobacco, spinach |
| | Ferro et al. 2002 | arabidopsis |
| | Ferro et al. 2003 | arabidopsis |
| | Schubert et al. 2002 | arabidopsis |
| | Zabrouskov et al. 2003 | arabidopsis |
| | Sun, Emanuelsson, & Van Wijk 2004 | arabidopsis |
| | Friso et al. 2004 | arabidopsis |
| | Whitelegge 2004 | arabidopsis, spinach |
| | Peltier et al. 2004 | arabidopsis |
| | Van Wijk 2004 | *Chlamydomonas-Reinhardtii*, arabidopsis |
| cell wall proteome | Feiz et al. 2006 | arabidopsis |
| vacuolar proteome | Shimaoka et al. 2004 | arabidopsis |
| | Carter et al. 2004 | arabidopsis |
| | Endler et al. 2006 | arabidopsis |
| | Jaquinod et al. 2007 | arabidopsis |
| | Schmidt et al. 2007 | cauliflower |
| nuclear proteome | Pendle et al. 2005 | arabidopsis |

samples, to our knowledge this promising technique was not yet applied to non-model plants. Penin, Godinot, and Gautheron (1984) suggested to use the cationic detergent cetyltrimethylammoniumbromide (CTAB) in the first dimensional separation and also benzyldimethyl-n-hexadecylammonium chloride (16-BAC) proved to be promising to establish an alternative 2DE technique (Macfarlane, 1989; Hartinger et al., 1996). Once their critical micellar concentration is achieved, the cationic detergents CTAB (Eley et al., 1979) and 16-BAC (Macfarlane, 1983) bind at a constant ratio to proteins, analogous to the well-known anionic detergent SDS. Similar, those cationic detergents will mask the intrinsic charge of proteins and a separation based on their molecular size is possible. Further improvements to the CTAB electrophoresis were introduced by Buxbaum (2003), though since their first publications the cationic detergent methods remained largely unnoticed. Only the last 4 years they started to gain popularity for the analysis of membrane proteins (Braun et al., 2007). Plant research is still lagging behind. 16-BAC has been used to identify sucrose-binding-protein homologs via western blot in isolated Golgi fractions of pea (*Pisum sativum* L.) seeds (Wenzel et al., 2005).

The separation range of dSDS and cationic-based 2DE is still limited because these two-dimensional techniques lack true orthogonality as both dimensions discriminate on the basis of molecular size. Though, depending on the complexity of the protein mixture and on the genome status of the plant species, these techniques are promising for non-model plants to characterize hydrophobic proteins.

A technique that did already prove its value to analyze non-model plant membrane proteins is native electrophoresis. Moreover, as a result of the use of non-denaturing agents, information about the organization of protein complexes or protein–protein interactions is obtained. Clear-Native (or colorless-native) electrophoresis (CN-PAGE) uses the inherent negative charge of proteins (with an acidic pI) for separation. Alternatively, negative charges are provided by adding the negatively charged protein-binding dye Coomassie Brilliant Blue G-250 in Blue-Native electrophoresis (BN–PAGE). Schägger and von Jagow developed these techniques to separate mitochondrial membrane proteins and complexes from muscle tissue (Schägger & von Jagow, 1991; Schägger, Cramer, & von Jagow, 1994). Krause (2006) summarizes the most important

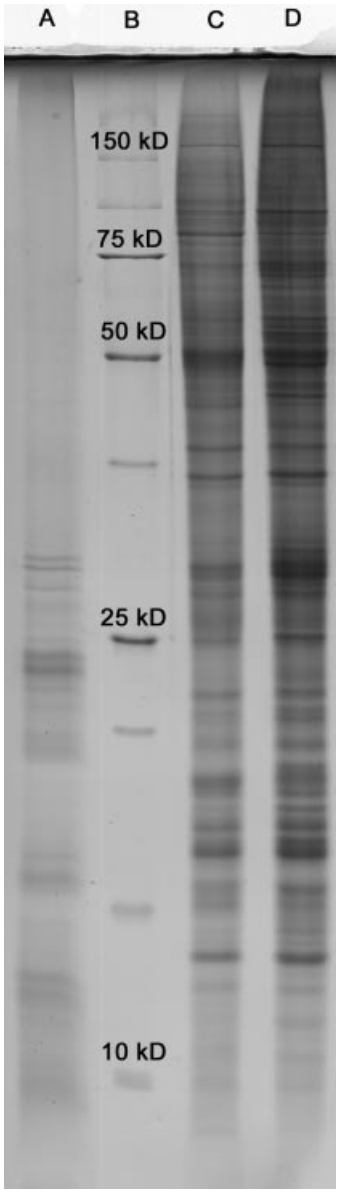


**FIGURE 2.** A representative gel of banana meristem hydrophobic proteins fractionated via chloroform methanol and visualized via silver staining. A: Chloroform/methanol soluble proteins, (B) Protein standards, (C) Total amount of proteins, (D) Chloroform/methanol insoluble proteins. There is a visible fractionation of the proteins and a bias toward small (hydrophobic) proteins.

achievements of both techniques to elucidate protein–protein interactions. BN–PAGE proved to be an elegant tool for studying the large multienzyme complexes in specific organelles, that is, mitochondria (oxidative phosphorylation) and chloroplasts (photosynthetic apparatus). The latter is also the reason why BN–PAGE became quite popular in plant studies (Eubel, Braun, & Millar, 2005; Granvogl, Reisinger, & Eichacker, 2006; Reisinger & Eichacker, 2006, 2007). Likewise, protein complexes of the plasma membrane of *Spinacia oleracea* leaves (Kjell et al., 2004) and *Dunaliella salina* (Katz et al., 2007) have been successfully resolved by BN–PAGE as well as the complexes in the peribacteroid membrane from *Lotus japanicus* root nodules (Wienkoop & Saalbach, 2003). Remmerie et al. (submitted) presented protein–protein interactions starting from whole cell lysates from *Nicotiana tabacum* BY 2 cells. This application of BN–PAGE to whole plant cell lysates increases its possibility as an analytical tool for functional proteomics. A general overview of stable protein complexes from virtually all organelles, including plastids, mitochondrion, nucleus, endoplasmetic reticulum, and plasma membrane, are provided in one single gel (Fig. 3). With this strategy, not only known protein complexes could be separated and detected, also possible protein–protein interactions that had not been observed in plants before were visualized. Nevertheless, as it is the case with dSDS and cationic-based 2DE, the separation range of BN/SDS–PAGE is still limited. Reducing sample complexity of the total cell extract by prefractionation methods anticipates this problem. The advantage of this total cell BN/SDS–PAGE is that it may reveal many complexes in a cellular context and that it can be used to monitor the kinetics of complexes during perturbation or time-course experiments. In order to eliminate false positives, the authenticity of novel interaction partners (not described in model plants) has to be confirmed by an alternative protein interaction identification technique such as co-immunoprecipitation (Co-IP), tagged affinity purification (TAP), two hybrid strategies or bimolecular fluorescence complementation approaches (Hink, Bisseling, & Visser, 2002; Figeys, 2003). The latter technique also enables to localize the subcellular interaction site and thereby can rule out false positive interactions following organelle disruption (Hu, Chinenov, & Kerppola, 2002).

### C. Explorative Multivariate Analysis: A Useful Tool for Both Non-Model and Model Organisms

After separation through 2DE several hundreds of individual protein abundances can be quantified in a cell population or sample tissue. Data from 2DE analysis are generated through image analysis software that detects and quantifies the protein abundances and matches the proteins across the different gels. Though the matching quality is dependent on the software algorithm, it is above all determined by the quality and reproducibility of the gels. As discussed above, non-model plants are generally not ideal laboratory models and it is therefore essential to first get insight into the reproducibility of the data. The typical tests that are applied in proteome research (univariate statistical tests like the *T*-test, the Kolmogorov–Smirnov test, ANOVA or the Kruskal–Wallis test) analyze the individual variables (i.e., protein spots) one by one and have not been designed to analyze complex datasets containing multiple correlated variables. Consequently, they do not give an overview of the data. Exploratory data analysis approaches a biological problem from a different perspective and tries to describe patterns, relationships, trends and outlying data. In contrast to a univariate approach, it can be used to explore the general inter- and intra-group variability of the biological samples. It gives insight into the experimental groups, it displays the interrelationships between the large number of variables and it is helpful to improve the image analysis and to detect protein mismatches (Fig. 4).

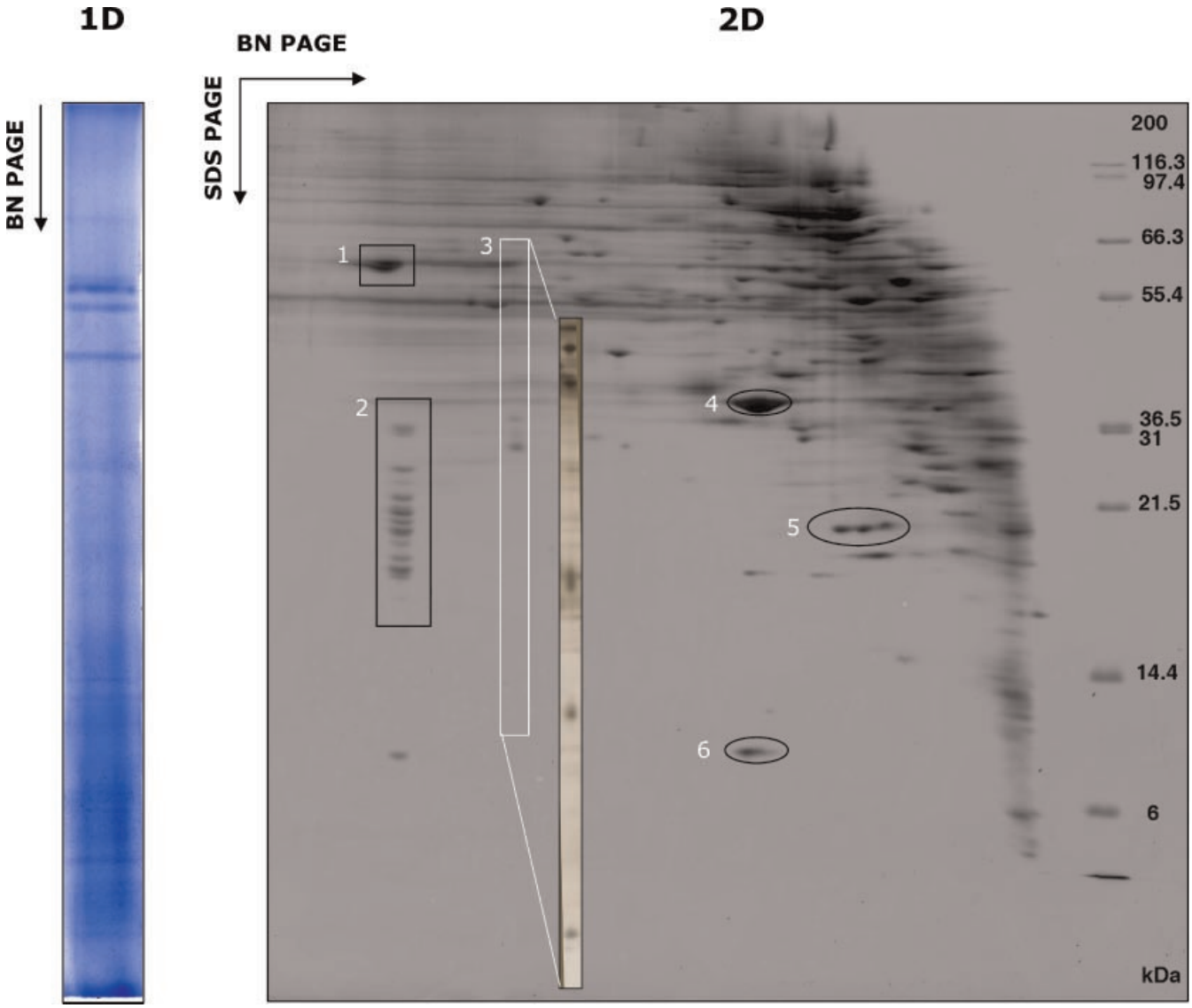


**FIGURE 3.** A representative gel of tobacco BY2 proteins separated via Blue Native electrophoresis. Both the results are presented: a 1DE gel where protein complexes are separated according to their molecular size under native conditions and a 2DE gel where the denatured proteins of the individual complexes are also separated according to their molecular size in the second dimension. A selected number of proteins and complexes are annotated: (1) both units of the Rubisco binding protein complex, (2) the individual units of the 20S proteasome, (3) enlarged in an inset the individual silver stained proteins of the F0F1-ATPase complex, (4) glyceraldehydes phosphate dehydrogenase, (5) nucleoside diphosphate kinase, (6) different expression forms of triosephosphate isomerase. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Principal component analysis (PCA) is one of the multivariate methods available to perform explorative data analysis. A comprehensive overview of the use of PCA in statistics is given by Sharma (1996). Explorative PCA does not put strict requirements to the data. The only requirement is that the data set has to be complete, meaning that missing spot values among the different samples are not allowed. A missing value in 2DE proteomics is undoubtedly correlated to gel quality. Apart from real absent proteins, the causes of missing values might be (i) faint spots at the detection limit and detected in one gel but not detected in another; (ii) mismatches probably caused by distortions in the protein pattern or (iii) spots being absent due to bad transfer from the first to the second dimension. The concept of difference gel electrophoresis (DIGE), with its common internal standard and the co-running approach, anticipates the missing value problem to some extent but we have recently shown it remains an issue that must be addressed (Pedreschi et al., in press). Different studies have described this

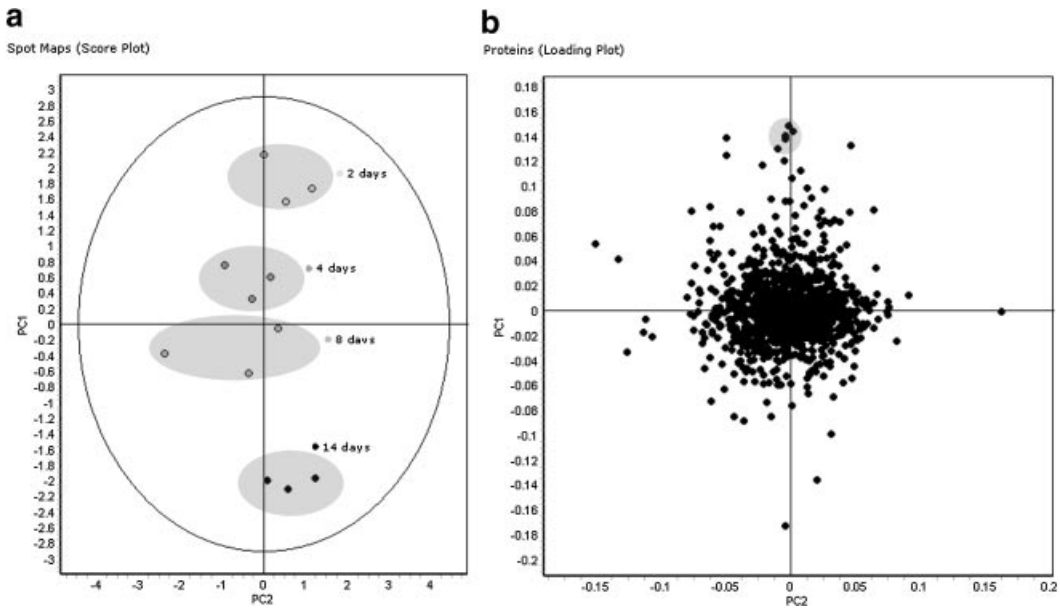


**FIGURE 4.** PCA analysis. **a**: Score plot. The big circle is based on the Hotellings $T^2$-test statistic and is used to detect outlying observables ($\alpha = 0.95$). The three biological replicates of the same experimental group cluster together, indicating an acceptable intra-group variability (gray ellipse). The different experimental groups are also separated, indicating a certain inter-group variability. There is a clear difference between 2 and 14 days of treatment. **b**: The loading plot indicates the correlation between the original variables. A protein with a high loading score for a specific PC explains an important part of the sample variance. As an example, we focus on five proteins that, from the loading plot, seem highly correlated (highlighted in B). Confirmatory differential expression analysis via ANOVA confirms that all five proteins have a very similar expression pattern over time. Four of them have been identified as isoforms. Reproduced from Carpentier et al. (2007b), with permission from Humana Press, copyright 2007.

missing value issue in detail (Troyanskaya et al., 2001; Oba et al., 2003; Jung et al., 2006; Krogh et al., 2007).

The combination of explorative multivariate analysis and confirmative univariate analysis is a powerful approach. For an overview of design and analysis issues in proteomics see Carpentier et al. (2007b). For a recent overview on univariate proteomics see Karp and Lilley (2007).

## III. MASS SPECTROMETRY

### A. Mass Spectrometric Analysis of Proteins

The standard approach for the analysis of 2DE separated proteins involves an enzymatic digestion of the protein in the spot of interest and extraction of the peptides followed by mass spectral analysis. The traditional way of analysis involves peptide mass fingerprint (PMF) analysis, typically performed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS since it provides a simple profile by producing a single peak per peptide. The concept behind PMF analysis was independently implemented by several groups at approximately the same time (Henzel et al., 1993; James et al., 1993; Mann, Hojrup, & Roepstorff, 1993; Pappin, Hojrup, & Bleasby, 1993; Yates et al., 1993). State-of-the-art equipment combines excellent sensitivity (subfemtomole region) and high accuracy (typically better than 10 ppm) with high-throughput capacity (typically less than 5 sec to generate a PMF). Unfortunately, PMF data have very little power to identify proteins from species with a fragmentary genome and protein repository (non-model organism). Hence, the chance of finding significant and conserved peptides decreases and PMF fails or results in false positive hits (Mathesius et al., 2002).

Tandem mass spectrometry (MS/MS) has been used for decades to obtain structural information of (bio)molecules. For peptides, MS/MS generates sequence specific information and the information content of such spectra is thus much higher than for PMF. Since the introduction of nano-electrospray ionization, routine use of electronspray ionization (ESI)-MS/MS for the analysis of 2D-PAGE spots has become feasible (Wilm & Mann, 1994). The formation of multiple charged species (a characteristic of ESI) generates complex spectra for peptide mixtures. This is usually circumvented by prior separation of the peptides by capillary or nanoscale liquid chromatography (nano-LC), which is, moreover, amenable for automated 2D spot analysis (Gatlin et al., 1998). Unfortunately, separation of peptides prior to MS/MS is expensive and time consuming. For these reasons, MALDI is often preferred because of ease of use, speed and the ability to include MALDI spotting in automated digestion protocols on liquid handling systems (Shevchenko et al., 1997). In addition, a MALDI approach has the advantage that it has the

potential to store temporarily the targets for re-analysis when certain data are not yet fully explored. Nevertheless both techniques are useful and complementary.

MALDI instruments were for a long time equipped with time-of-flight (TOF) analyzers that were poorly amenable to MS/MS type of experiments (Arnott, Henzel, & Stults, 1998). The development of post-source-decay (PSD) analysis (Kaufmann et al., 1996; Gevaert et al., 1997) allowed to visualize spontaneous fragmentation of peptides. Nevertheless, it was only with the development of hybrid instruments, first Q-TOF instruments (Shevchenko et al., 2001; Wattenberg et al., 2002), later TOF/TOF (Medzihradszky et al., 2000; Yergey et al., 2002) and quadruple ion trap (QIT)/TOF instruments (Martin & Brancia, 2003) that also MALDI became routinely used for MS/MS. Recently, the PSD phenomenon was rediscovered as a tool for peptide sequencing with the introduction of the ''lift'' concept (Suckau et al., 2003).

The main advantage of protein identification based on peptide structural information for species with a partially sequenced genome is that it can be successfully used for the high-throughput identification of protein orthologs (see further). Unfortunately, the software tools are developed to search against a database of known proteins, which might produce low scores when working with non-model organisms. Recent examples of proteomic research on non-model plants confirm this statement: 2D-PAGE-based proteomic on strawberry proteins resulted only in 40% protein identification (Alm et al., 2007), for sunflower a success rate of 51% was reported (Hajduch et al., 2007). Taking into account the abundance of the proteins, we could clearly show a bias toward abundance and found an average success rate of 36% in banana (Carpentier et al., in press). For comparison: rice proteomics studies report a near 100% identification (Yang et al., 2007). Depending on the genome status of the species under investigation and on the degree of homology to a model organism, *de novo* sequencing might be essential to obtain significant sequence information. Grossmann et al. (2007) report in a case study with spinach, bell pepper and cassava that on average 11% of their identified proteins could be identified only via *de novo* peptide sequencing and sequence similarity searching.

### B. *De Novo* Sequencing

Sequence reconstruction of an unknown peptide based solely on the acquired mass data is referred to as *de novo* peptide sequencing. Early *de novo* sequencing involved the use of chemical microsequencing using Edman chemistry and still quite recently, it was used complementary to mass spectrometry to identify proteins from plant origin (Beyer et al., 2002; Kao et al., 2004; Khan & Komatsu, 2004). This method, however, is quite expensive in terms of reagent cost and suffers from a low throughput and sensitivity. Nowadays *de novo* sequencing is almost exclusively realized via a MS-based approach. The challenge of MS-based *de novo* sequence determination resides in: (i) finding fragment ion peaks that correspond to the progressively shorter peptide and (ii) determining the ion-type of an observed series of a partial peptide (directionality).

ESI-MS/MS routinely provides more informative MS/MS spectra of unmodified tryptic peptides than MALDI-TOF MS/MS. Doubly charged peptide ions tend to fragment more equally across a given sequence than do singly charged ion species (Cramer & Corless, 2001; Tabb et al., 2003). Apart from requiring higher activation energies, singly charged ions, as preferentially produced during MALDI-ionization, often yield relatively low quality collision-induced dissociation (CID) mass spectra which result from a small number of preferred fragmentation pathways (Qin & Chait, 1995). This leads to poorer scoring in classical search algorithms, in particular for cross-species identification. An elegant solution to the problem of directionality is offered by proteolytic differential isotopic labeling (Shevchenko et al., 1997). Other chemistries that result in facilitated *de novo* sequence analysis by differentiation of *N*- and *C*-terminal fragments have been proposed, involving either the introduction of a label during cell culturing (Gu et al., 2002; Shui et al., 2005) or derivatization of peptides after proteolytic digestion (Brancia et al., 2004; Beardsley, Sharon, & Reilly, 2005). Although *de novo* sequence determination is simplified, these methods do not improve the fragmentation reaction or the detection of fragments.

Based on the emerging knowledge on reactions resulting in peptide bond cleavage during MS/MS (Wysocki et al., 2000; Paizs & Suhai, 2005), another set of methods was developed to improve *de novo* sequencing. Derivatization of peptides with a charged moiety influences the fragmentation mechanism by sequestration of protons (Jones et al., 1994; Dongre et al., 1996). The research group of Keough, Youngquist, and Lacey (1999) developed a general procedure for high-sensitivity *de novo* peptide sequencing using PSD-MALDI. By adding a permanent negative charge to the *N*-terminus of tryptic peptides (*in casu* a sulfonic acid group), the *C*-terminal positive charge is counter-balanced. As the most basic residue is already protonated, excess protons, so-called ionizing protons, will be more or less free to randomly ionize backbone amide groups favoring charge directed fragmentation of the weakened peptide bonds. Because fragments that contain the negatively charged *N*-terminus are not detected in positive ion mode, only *C*-terminal *y*-ions will be visible and *de novo* sequence determination simply requires the calculation of mass differences between consecutive peaks (Fig. 5). Nonetheless, the preferential fragmentation at specific residues as described for non-derivatized peptides using ESI (Tabb et al., 2003, 2004) is preserved (Samyn et al., 2004).

Different reagents have been used for sulfonation and sulfonated peptides have been sequenced using different types of mass spectrometers (Bauer et al., 2000; Keough, Lacey, & Youngquist, 2000; Keough et al., 2000; Keough, Lacey, & Strife, 2001). We published in 2005 an efficient protocol that allows the extracted tryptic peptides to be *N*-terminally sulfonated without any further sample purification (Sergeant et al., 2005). We have validated this application on proteins isolated from banana (*Musa* spp.) (Samyn et al., 2007). Furthermore, the possibility to discriminate several isoforms showed that apart from cross-species identification, sulfonation can also be used to identify biologically important modifications such as single nucleotide polymorphisms or the differentiation between paralogs and between variety specific homologs. The characterization of isoforms benefits undoubtedly from *de novo* analysis but it is not

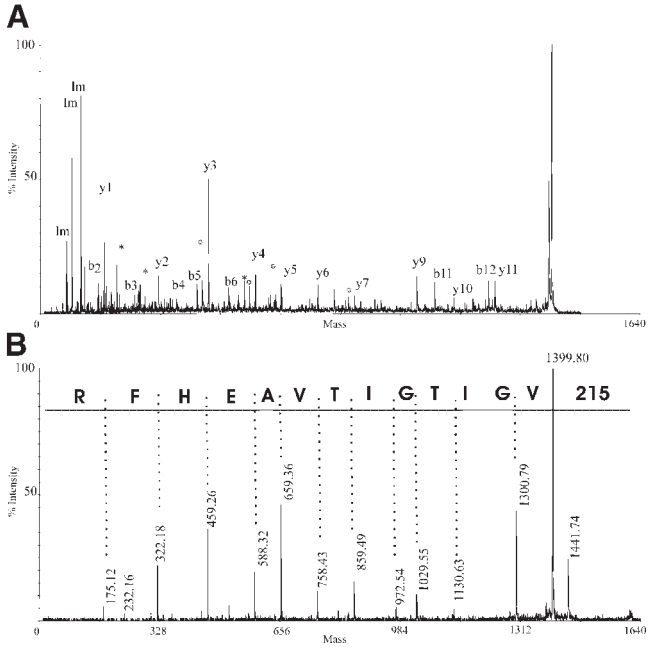


**FIGURE 5.** MALDI MS/MS spectra from respectively unsulphonylated (**A**) and sulphonylated (**B**) peptide from Oak. The protein has been identified as F1-ATP synthase, beta subunit. The closest homolog was ''gi:4388533'' from *Sorghum bicolor*. A: The contributing y and b ions are indicated as well as the immonium ions (Im), the internal fragments (*) and neutral loss ions (y-17°). B: Only *y*-ions are detected.

a prerequisite. Laugesen et al. (2007) managed to identify several tissue and cultivar specific isoforms of peroxidase in barley. The nearly complete sequence coverage that was necessary to realize this, was obtained through the analysis of different complementary mass spectra datasets generated by MALDI-TOF MS and Q-TOF MS/MS and by searching these spectra against several databases including specific expressed sequence tag (EST) libraries.

Novel precision mass spectrometric approaches using Fourrier Tranform (FT)-MS dramatically improve the performance for *de novo* sequencing without prior modifications. A recent article by Frank et al. (2007) clearly demonstrates the potential of these systems in off-line ESI-MS and MS/MS identification. Moreover, the high precision eliminates drastically the number of candidate peptide sequences that fit a tandem mass spectrum. The ongoing explosion in the availability of hybrid quadrupole or linear ion trap FTMS (both Fourrier Transform-Ion Cyclotron Resonance FT-ICR and Orbitrap) likely paves the way for wider use to obtain *de novo* sequences. Currently, ESI is the main ionization technique in such configurations. It seems that the high resolution allows both the separation of the complex multiple charged ions and the easy determination of the charged state which potentially eliminates some of the arguments pro MALDI for *de novo* sequencing.

### C. Tools for Terminal Sequencing

For non-model organisms, the identification of the *N*- or *C*-terminal sequences of proteins might give crucial information enabling the identification of a protein and the discrimination between isoforms. Furthermore, it has been demonstrated very recently that this information can be used to study proteolytic processing events and to verify the correctness of genome annotations, both in prokaryotic and eukaryotic organisms (Dormeyer et al., 2007; Gupta et al., 2007).

Using MS, the characterization of the *N*- or *C*-terminal sequence relies first on the selection of the particular terminal peptide. The use of diagonal chromatography solved this problem in a gel-free approach for *N*-terminal peptides (Gevaert et al., 2003). For *C*-terminal peptide selection, no such method is described. Since *C*-terminal peptides tend to be more selective for protein identification purposes (Wilkins et al., 1998a), a number of chemical/isotopic labeling techniques have been described to isolate them. Those techniques led rarely to the identification of peptides (Zhou et al., 2004), most likely due to problems associated with their recovery and the need for larger sample amounts (Kosaka, Takazawa, & Nakamura, 2000). Recently, we reported a new approach directed at selective characterization of the *C*-terminal sequence (Fig. 6). The MS-based enzymatic ladder sequencing approach is applied on the unseparated peptide mixture after chemical cleavage of the protein by CNBr in gel or in solution (Samyn et al., 2005). Upon cleavage, Met residues are converted to homoserine lactone (hsl). During subsequent incubation with CarboxyPeptidases (CPX) only the original *C*-terminal fragment is accessible to enzymatic degradation and forms a ladder (Fig. 6B). Ladder read-out is performed using a MALDI-TOF/TOF instrument as this ionization technique produces predominantly ladders of singly charged ions. In experiments where insufficient *C*-terminal residues were removed by CPX, the peptide is further subjected to MALDI MS/MS fragmentation (Samyn, Sergeant, & Van Beeumen, 2006). As a proof of principle, the method was used to investigate the proteolytic processing of procardosin A, an aspartic proteinase isolated from the artichoke thistle *Cynara cardunculus* (Castanheira et al., 2005).

The isolation of the specific terminal fragments can be avoided by performing the analysis on the intact protein. The fragmentation of intact proteins with MS, the so-called ''top-down'' approach, has been demonstrated on a variety of instruments (Aebersold & Mann, 2003). The improvements in FTMS technology and fragmentation technology such as electron capture dissociation (ECD) or combined InfraRed MultiPhoton Dissociation (IRMPD) and ECD now allow to obtain sequence information including information on the terminal sequences. The method is not yet applicable in high-throughput proteomics. The recently described matrix compositions for improved in-source decay of intact proteins in MALDI may lead to similar applications (Demeure et al., 2007). However, a major bottleneck for top-down proteomics is the requirement of separation of intact proteins by liquid chromatography and so it is restricted to relatively simple protein mixtures.

### D. Cross-Species Protein Identification

Since protein identification usually refers to determining from which gene a certain protein originates, proteins derived from

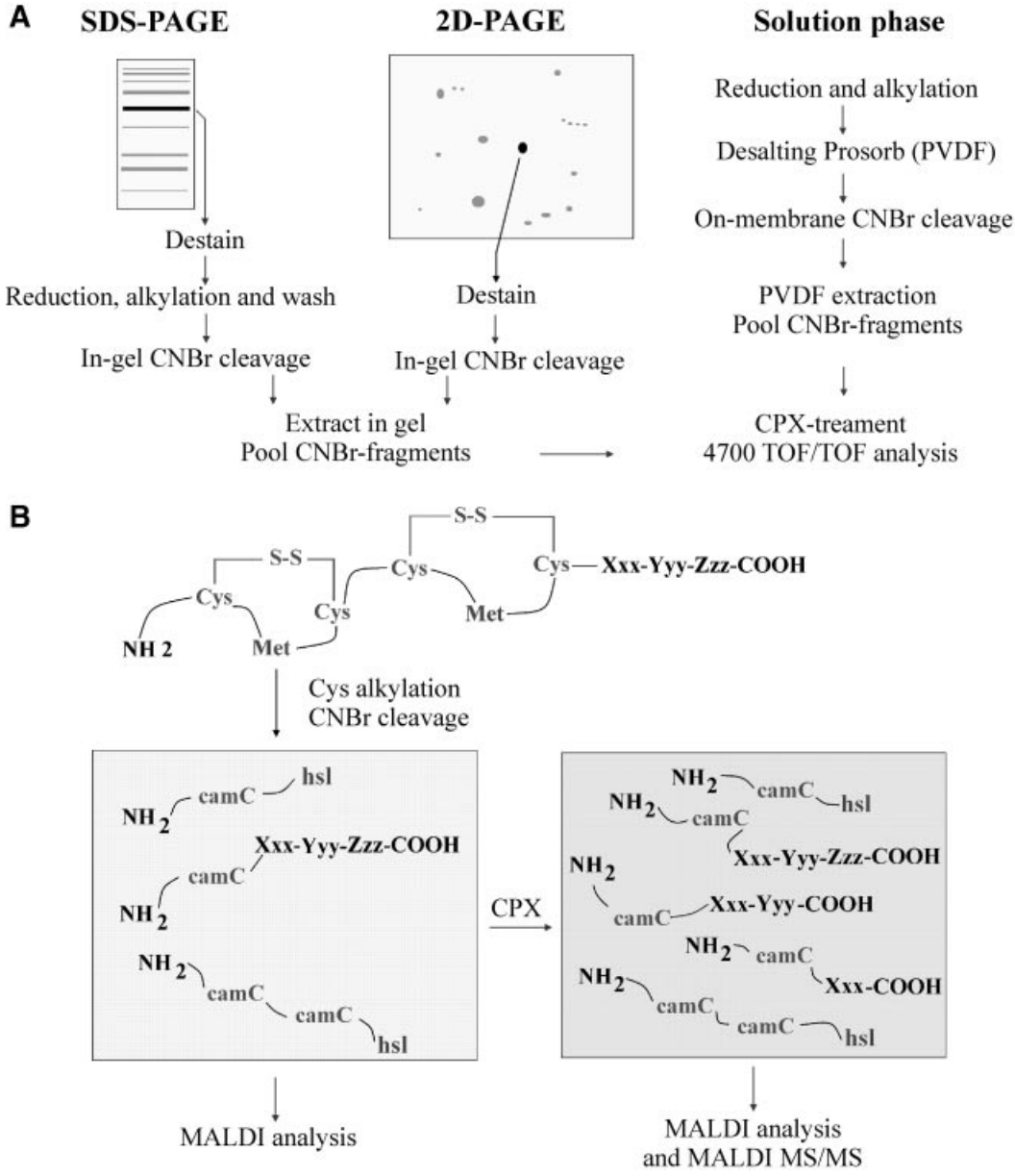


**FIGURE 6.** **A**: Schematic representation of the steps involved in the *C*-terminal sequencing method. **B**: Conversion of Met residues to homoserine lactone (hsl) upon cysteine alkylation and cleavage, and MALDI-TOF/TOF-based ladder read-out of the original *C*-terminal fragment after incubation with carboxypeptidases (CPX).

non-sequenced genes can in a strict sense not be ''identified.'' If the corresponding gene of the investigated organism is unknown, the aim is rather ''finding the most similar gene in a closely related organism.'' Coming from cross-species protein identification, protein ''identity'' information cannot be led back to one single gene locus in a single non-redundant genomic database, but is dynamically rather associated with a number of ''most similar'' database entries. The typical ''identity'' field should be evaluated together with a statistical parameter describing the significance and completeness of the homology.

Conventional protein identification strategies are generally based on two complementary approaches: one based on the comparison of the experimentally acquired masses, one based on the comparison of derived sequences (based on *de novo* sequencing), and one that combines both. The search for a homologous hit instead of an identical hit imposes some specific requirements on these spectral data analysis approaches. Not all approaches are equally tolerant toward variations between sequences.

Tandem MS-based identification is more tolerant to variation. A single mutation completely abolishes the information content of a peptide in PMF by shifting its mass. MS/MS

retains also sequence information of the unmodified residues of the peptide. Information from multiple peptides derived from a single protein can be combined to gain additional confidence. Herein lies the advantage of the two-dimensional gel-based approach, which allows to conserve the connectivity between all information derived from multiple peptides originating from one observed protein. Compared to the more ''simple'' mass correlation applied in PMF database searching, identification from tandem mass spectra requires that some sort of weight needs to be assigned to peaks, since not all ion types are equally diagnostic. Additional valuable information lies in the continuity of mass peaks and their intensity. Widely employed algorithms take this extra information into account and apply a certain information prioritization when calculating scores. The correlation between theoretical and experimental data gives an indication about the degree of homology between the given protein and the hit sequence. A probability based score then typically indicates the significance of the hit. This database dependent strategy is probably the most scalable approach that allows for a relatively high-throughput analysis of the proteome of a non-sequenced organism. The approach becomes even more powerful when additional sequence data sources are taken into account, such as the growing EST sequence repositories, which are usually queried in six-frame translations (Mooney & Thelen, 2004; Carpentier et al., 2007a). A number of spectrum-based identification engines are currently available. Well known is Mascot (Perkins et al., 1999), which can, with moderate restrictions, be freely used at www.matrixscience.com. Alternative commercial products are SEQUEST (Thermo Fisher Corp.) and Phenyx (Genebio) and free engines such as X!Tandem (Craig & Beavis, 2004), VEMS (Matthiesen et al., 2004) Protein prospector (Clauser, Baker, & Burlingame, 1999), and profound (Zhang & Chait, 2000).

The success rate of protein identification via conventional database searching based on the MS/MS information from multiple peptides derived from a single protein is, due to its restricted error tolerance, still limited for non-model organisms. Early software, specifically the Peptide Sequence Tag algorithm, involved partial user interpretation of mass spectra in order to derive a minimum of sequence information to be incorporated in the database searching (Mann & Wilm, 1994; Wilm, Neubauer, & Mann, 1996). Basically, a peptide sequence tag consists of the mass of the precursor ion (the peptide mass), a short sequence fragment directly derived from the spectrum, accompanied by the masses of its flanking ions. This may yield perfect matches, but in order to cope with the typical inter-species sequence variations, more error-tolerant methods were developed. In the MultiTag method (Sunyaev et al., 2003; Liska et al., 2005) multiple error-tolerant searches are carried out with the peptide sequence tags in a loosened query specificity, for example, by allowing for one or both mass mismatches or an amino acid substitution. Error-tolerant peptide tag searches tend to produce enormous lists of potential hits, from which the MultiTag attempts to extract the most probable hit. This is done by combining the results from multiple error-tolerant searches and assigning an *E*-value to each hit to judge its confidence level. The recently introduced Paragon algorithm differs from others in that it models modifications and substitutions with probabilities, rather than implementing user controlled settings whether to consider or not to consider them (Shilov et al., 2007).

Considering the cross-species principle, it is worthwhile to mention the importance of results validation: false positive and false negative hit rates should ideally be reduced to zero. The rate of incorrect identifications can be estimated in a target-decoy approach, in which the proportion of hits against a decoy database containing ''nonsense'' sequences gives a reliable false positive estimation. When analyzing large datasets, it is common practice to repeat the homology-based identification search algorithm using the same database containing the reversed or randomized sequences. Since no real matches are expected from this decoy database, the number of positive hits in this search strategy gives an excellent estimate of the number of potential false positives. Instead of repeating the search, good arguments are at hand to perform only a single search using a concatenated database containing both the target and the decoy database. Obvious advantages include reduction of the total processing time and a direct simultaneous competition of decoy and target sequences for the highest ranking peptide. A significant additional advantage is that the search report analysis allows interpreting the relative score and ranking of especially high scoring false positives and low scoring true positives (Elias & Gygi, 2007). While these empirical approaches are especially effective for large datasets, estimation of false positives on restricted datasets is less accurate and additional statistical analysis is needed. A means to assess the error of false positive estimation associated with the size of the database has been described and a method has been developed to calculate this uncertainty (Huttlin et al., 2007).

In *de novo* sequencing, the mass spectrum is interpreted directly and a corresponding amino acid sequence is deduced independent of any database. This is often done manually, usually assisted by a computer program that calculates and displays differences between spectral peaks that exactly correspond to the mass of one (or more) amino acid(s). Obviously this approach is not scalable to high-throughput situations. The derived sequences are subsequently compared to a database of known sequences via sequence similarity search engines. There exist different dedicated sequence similarity search engines such as CIDentify (Taylor & Johnson, 1997), an MS tailored version of gapped BLAST (Huang et al., 2001), a MS driven BLAST (Shevchenko et al., 2001), FASTS (Mackey, Haystead, & Pearson, 2002), MS-homology (Clauser et al., 1999), and Open Sea (Searle et al., 2004). Some have already successfully been applied on different non-model plants (Castro et al., 2005; Jorge et al., 2005; Grossmann et al., 2007; Samyn et al., 2007; Waridel et al., 2007).

The automatic *de novo* sequencing of not-simplified spectra constitutes a major computational challenge, although consistent progress has been made over the years. The first attempts towards automated *de novo* sequencing considered all theoretically possible sequences that correspond to the parent ion mass, and compared their theoretical spectra with the experimental spectrum (Sakurai et al., 1984). Due to the number of possible combinations this approach is computationally too demanding. A different approach starts with the finding of short matching sequences, which are then gradually extended as long as the match is conserved (Johnson & Biemann, 1989). Though, this method may lead to false negatives when certain fragment ions

are missing. Most current approaches are based on the graph theory, a common data representation method in computer sciences, for example, Fast (Bartels, 1990), SeqMS (Fernandez-De-Cossio et al., 1998), Lutefisk (Taylor & Johnson, 2001), SHERENGA (Dancik et al., 1999), PepNovo (Frank & Pevzner, 2005). Some non-graph theoretic approaches have been introduced more recently, for example, Peaks (Ma et al., 2003) and ByOnic (Bern, Cai, & Goldberg, 2007). None of the current techniques readily converts tandem mass spectra into sequences, but these algorithms greatly assist in reducing the number of verifiable options. For more background and comparisons of *de novo* sequencing algorithms the reader is referred to several reviews (Pevtsov et al., 2006; Xu & Ma, 2006; Colinge & Bennett, 2007).

## E. Data Integration

### 1. Reference Databases and Spectral Libraries

It was recognized early that identifications obtained in previous proteome studies remain of great value. A first effort to make previous identifications useful for future studies was the development of the federated 2D databases http://www.expasy.ch/ch2d/ (Appel et al., 1996). This offered researchers a rapid way to locate a certain protein on an experimental ''reference'' 2D map and made it possible to derive the identity of an unknown spot by matching it with a corresponding spot on a reference map. Proticdb is a web-based application of the UMR of Génétique Végétale du Moulon to store, track, query, and compare plant proteome data (Ferry-Dumazet et al., 2005) and is exclusively dedicated to plants. We have previously generated reference maps for tobacco BY-2 cells and banana meristems (Laukens et al., 2004; Carpentier et al., 2007a). Such a reference database is most useful in serving as rapid ''pre-analysis'' information, but another potential application, in particular for non-sequenced organisms, lies in the reverse approach. Given the fact that pI and molecular size of the primary structure is reasonably well conserved between isoforms and orthologs these federated repositories can be used in guiding experiments. When a researcher is interested in a certain protein for PTM analysis or other purposes, a reference database of the particular or related species could point to the spots on a matched map which could then be picked for further analyses.

Gel images and identification results are typical features included in proteome-databases, but even more potential lies in the inclusion of spectral data. Incomplete or insufficiently informative spectra can lead to positive identification when first matched with previously acquired spectra in a database (Frewen et al., 2006). During the last few years several initiatives were launched to develop spectrum databases or spectral libraries. Preprocessed experimental spectra are usually combined into synthetic reference spectra using some kind of averaging technique (Craig, Cortens, & Beavis, 2005; Craig et al., 2006; Liu et al., 2007). Parallel, the need has arisen to have methods to compare spectra with each other, or to search a spectral library with a freshly acquired un-interpreted spectrum.

The integration of proteome data in the case of a non-model species involves some specific concerns. Regular re-annotation of proteins by re-analysis of their digest MS-spectra against new database versions can lead to further improvement of the existing annotation database. Spectral libraries are potentially valuable to assist with the identification of incomplete spectra. The connectivity between spectra from multiple peptides originating from a single gel spot may also facilitate future LC-based MudPIT approaches, even for incompletely sequenced organisms.

An important function of integration efforts is maintaining, promoting and developing the internal relationships between different types of data. In particular for organisms dependent on cross-species identification, the connectivity between multiple peptides originating from a single protein spot is an important characteristic. Often the knowledge of a previous analysis of a certain sample or spot can be of great value for repeated analysis and results interpretation. The pProRep tool enables such a gel-centric data integration into a relational database. It offers a web-interface to which users can import, analyze, visualize, and export experimental data sets (Laukens et al., 2006). Some advanced query functions allow for novel ways to search the database, for example with experimental peak lists. Query results and their internal relationships can be visualized for example on the spot level, and labeled. The application can be used as an ''analytical workbench'' for experimental proteome data, and will be further developed to offer more advanced data mining functions. Experimental data sets are growing and can be of great value for the future interpretation of new experiments if the right tools are available. In addition these analytical and integrative functions, pProRep also intends to serve the purpose of online data sharing.

It can be expected that this need for proper integration and electronic sharing of experimental data will receive momentum, as minimal reporting requirements are established under the guidance of the leading proteome journals and the Hupo Standards Initiative (http://psidev.info/).

### 2. Cross-Species and Functional Annotation

The current possibilities to acquire ''omics''-data imposes new challenges to the interpretation, clustering, comparison, and functional annotation of biological data. This includes integration with biological knowledge databases such as the Gene Ontology Database, protein interaction databases, pathway databases, as well as mining the biological literature. A number of automated tools to accomplish this task are now becoming available.

The annotation of genes from non-model organisms depends totally on what is known from the model organisms. The attributes known from the retrieved homologous hit, such as functions, molecular interactions and localization, can then, though carefully, be projected onto the organism and process under investigation and be interpreted in the context of the study. The lack of annotated data complicates this task, but does not preclude. To a certain extent annotation data can be successfully transferred over the species-border if the (sequence) similarity criteria are carefully selected (Yu et al., 2004). Further maturation of this field can be expected in the near future.

## IV. BANANAS: EPILOGUE

This closing section focuses shortly on bananas and plantains as an example of a non-model crop with a booming proteome analysis. With an annual production of approximately 100 million tons, banana and plantain are one of the most important food commodities after rice, wheat and maize (FAOstat, 2007). Bananas and plantains are cultivated in more than 120 countries and are a staple food source of 400 million people, as only 13% are exported. The internationally-traded banana varieties belong to the ''Cavendish'' group. These very similar varieties are grown in monocultures which favor pest and disease development. A disease outbreak can wipe out the entire crop. The dessert banana ''Gros Michel'' was such a case when it was destroyed by the Panama disease (or Fusarium wilt) and as such the export industry completely collapsed in the 1950–1960s. ''Gros Michel'' was then replaced by the current ''Cavendish'' varieties which are nowadays threatened again by virulent diseases, abiotic stresses, etc. (Heslop-Harrison & Schwarzacher, 2007).

Despite its importance, both as a food source and as a commercial crop, the genome of banana is poorly characterized. The A genome is estimated at 638 Mb (11 chromosomes) and the B genome at 529 Mb (11 chromosomes) (Lysak et al., 1999). The Global Musa Genome Consortium (http://www.musagenomics.org/) currently reports that only 0.4% of the A genome and 0.1% of the B-genome is sequenced. Large scale gene expression profiling based on transcripts is feasible but remains quite challenging (Coemans et al., 2005) and thus a proteomics approach is more appropriate (Carpentier et al., in press). In order to start a proteome approach on banana, we focussed first on the specific problems of protein extraction of this extremely recalcitrant plant and established a powerful protein extraction for 2DE separation (Carpentier et al., 2005). Subsequently we focussed on data analysis and the possibilities of statistical analysis (Carpentier et al., 2007b) and on the specific problems associated with protein identification. In order to maximize the identification rate, we combined different ways of database searching: high-throughput database dependent searching (cross species and EST-based) (Carpentier et al., 2007a) and database independent *de novo* sequencing combined with error tolerant BLAST searching (Samyn et al., 2007). The added value of such a strategy is that the ease and high-throughput of database dependent searching is complemented by the more powerful *de novo* sequencing and error tolerant BLAST searching: proteins that have not been identified successfully via database dependent searching are subsequently subjected to *de novo* sequencing. Currently we are exploring alternative electrophoresis systems to tackle the membrane proteome and are creating species and tissue specific EST libraries. Results are expected to improve the exploitation of the International Musa Germplasm Collection which is currently stored at the Laboratory of Tropical Crop Improvement (K.U.Leuven, Belgium), under the auspices of Bioversity International.

The optimized workflow for a non-model organism comprises (i) the investment in a powerful protein extraction method capable to deal with the interfering compounds, (ii) the combination of different complementary protein fractionation, separation and quantification techniques to maximize the resolution and to cover the proteome as good as possible, and (iii) the usage of different complementary MS techniques and error tolerant database searches.

From a prospective view the ideal workflow for a non-model organism should bundle the spectral data from 2DE experiments into libraries. The connectivity between spectra from multiple peptides originating from a single gel spot may facilitate future LC-based approaches.

## ACKNOWLEDGMENTS

The authors thank Prof. Josef Van Beeumen for critical reading and Noor Remmerie for the sharing of data. Financial support from the Belgian National Fund for Scientific Research (FWO-Flanders) is gratefully acknowledged. Dr. S.C. Carpentier is supported by a postdoctoral fellowship of the K.U.Leuven.

## REFERENCES

Adam D. 2000. Now for the hard ones. Nature 408:792–793.

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

Agrawal GK, Rakwal R. 2006. Rice proteomics: A cornerstone for cereal food crop proteomes. Mass Spectrom Rev 25:1–53.

Akiyama Y, Ito K. 1985. The SecY membrane component of the bacterial protein export machinery—Analysis by new electrophoretic methods for integral membrane-proteins. EMBO J 4:3351–3356.

Alm R, Ekefjard A, Krogh M, Hakkinen J, Emanuelsson C. 2007. Proteomic variation is as large within as between strawberry varieties. J Proteome Res 6:3011–3020.

Appel RD, Sanchez JC, Bairoch A, Golaz O, Ravier F, Pasquali C, Hughes GJ, Hochstrasser DF. 1996. The SWISS-2DPAGE database of two-dimensional polyacrylamide gel electrophoresis, its status in 1995. Nucleic Acids Res 24:180–181.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis. Nature 408:796–815.

Arnott D, Henzel WJ, Stults JT. 1998. Rapid identification of comigrating gel-isolated proteins by ion trap mass spectrometry. Electrophoresis 19:968–980.

Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K. 2004. Complete nucleotide sequence of the sugarcane (Saccharum officinarum) chloroplast genome: A comparative analysis of four monocot chloroplast genomes. DNA Res 11:93–99.

Bardel J, Louwagie M, Jaquinod M, Jourdain A, Luche S, Rabilloud T, Macherel D, Garin J, Bourguignon J. 2002. A survey of the plant mitochondrial proteome in relation to development. Proteomics 2:880–898.

Barry GF. 2001. The use of the Monsanto draft rice genome sequence in research. Plant Physiol 125:1164–1165.

Bartels C. 1990. Fast algorithm for peptide sequencing by mass-spectroscopy. Biomed Environ Mass Spectrom 19:363–368.

Bauer MD, Sun YP, Keough T, Lacey MP. 2000. Sequencing of sulfonic acid derivatized peptides by electrospray mass spectrometry. Rapid Commun Mass Spectrom 14:924–929.

Beardsley RL, Sharon LA, Reilly JP. 2005. Peptide *de novo* sequencing facilitated by a dual-labeling strategy. Anal Chem 77:6300–6309.

Bern M, Cai YH, Goldberg D. 2007. Lookup peaks: A hybrid of *de novo* sequencing and database search for protein identification by tandem mass spectrometry. Anal Chem 79:1393–1400.

Beyer K, Bardina L, Grishina G, Sampson HA. 2002. Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: Seed storage proteins as common food allergens. J Allergy Clin Immun 110:154–159.

Borner GHH, Sherrier DJ, Weimar T, Michaelson LV, Hawkins ND, Macaskill A, Napier JA, Beale MH, Lilley KS, Dupree P. 2005. Analysis of detergent-resistant membranes in Arabidopsis. Evidence for plasma membrane lipid rafts. Plant Physiol 137:104–116.

Brancia FL, Montgomery H, Tanaka K, Kumashiro S. 2004. Guanidino labeling derivatization strategy for global characterization of peptide mixtures by liquid chromatography matrix-assisted laser desorption/ ionization mass spectrometry. Anal Chem 76:2748–2755.

Braun RJ, Kinkl N, Beer M, Ueffing M. 2007. Two-dimensional electrophoresis of membrane proteins. Anal Bioanal Chem 389:1033–1045.

Brugiere S, Kowalski S, Ferro M, Seigneurin-Berny D, Miras S, Salvi D, Ravanel S, D'herin P, Garin J, Bourguignon J, Joyard J, Rolland N. 2004. The hydrophobic proteome of mitochondrial membranes from Arabidopsis cell suspensions. Phytochemistry 65:1693–1707.

Burre J, Beckhaus T, Schagger H, Corvey C, Hofmann S, Karas M, Zimmermann H, Volknandt W. 2006. Analysis of the synaptic vesicle proteome using three gel-based protein separation techniques. Proteomics 6:6250–6262.

Buxbaum E. 2003. Cationic electrophoresis and electrotransfer of membrane glycoproteins. Anal Biochem 314:70–76.

Canut H, Bauer J, Weber G. 1999. Separation of plant membranes by electromigration techniques. J Chromatogr B 722:121–139.

Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B. 2005. Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. Proteomics 5:2497–2507.

Carpentier SC, Witters E, Laukens K, Van Onckelen H, Swennen R, Panis B. 2007a. Banana (*Musa* spp.) as a model to study the meristem proteome: Acclimation to osmotic stress. Proteomics 7:92–105.

Carpentier SC, Panis B, Swennen R, Lammertyn J. 2007b. Finding the significant markers: Statistical analysis of proteomic data. In: Vlahou A, editor. Methods in molecular biology. Totowa, NJ: Humana Press, Inc. pp 327–347.

Carpentier SC, Dens K, Van den houwe I, Swennen R, Panis B. 2007c. Lyophilization, a practical way to store and transport tissues prior to protein extraction for 2DE analysis? Proteomics 7:S64–S69.

Carpentier SC, Coemans B, Podevin N, Witters E, Laukens K, Matsumura H, Terauchi R, Swennen R, Panis B. in press. Functional genomics in a non-model crop: Transcriptomics or proteomics? Physiol Plant DOI: 10.1111/j.1399-3054.2008.01069.x.

Carter C, Pan SQ, Jan ZH, Avila EL, Girke T, Raikhel NV. 2004. The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. Plant Cell 16:3285–3303.

Castanheira P, Samyn B, Sergeant K, Clemente JC, Dunn BM, Pires E, Van Beeumen J, Faro C. 2005. Activation, proteolytic processing, and peptide specificity of recombinant cardosin A. J Biol Chem 280:13047–13054.

Castro AJ, Carapito C, Zorn N, Magne C, Leize E, Van Dorsselaer A, Clement C. 2005. Proteomic analysis of grapevine (*Vitis vinifera* L.) tissues subjected to herbicide stress. J Exp Bot 56:2783–2795.

Celedon PAF, De Andrade A, Xavier Meireles KG, De Carvalho Mcdcg, Gomes Caldas DG, Moon DH, Carneiro RT, Franceschini LM, Oda S, Labate CA. 2007. Proteomic analysis of the cambial region in juvenile *Eucalyptus grandis* at three ages. Proteomics 7:2258–2274.

Clauser KR, Baker P, Burlingame AL. 1999. Role of accurate mass measurement (+/− 10 ppm) in protein identification strategies employing MS or MS MS and database searching. Anal Chem 71:2871–2882.

Coemans B, Matsumura H, Terauchi R, Remy S, Swennen R, Sagi L. 2005. SuperSAGE combined with PCR walking allows global gene expression profiling of banana (*Musa acuminata*), a non-model organism. Theor Appl Genet 111:1118–1126.

Colinge J, Bennett KL. 2007. Introduction to computational proteomics. Plos Comput Biol 3:1151–1160.

Corthals GL, Wasinger VC, Hochstrasser DF, Sanchez JC. 2000. The dynamic range of protein expression: A challenge for proteomic research. Electrophoresis 21:1104–1115.

Craig R, Beavis RC. 2004. TANDEM: Matching proteins with tandem mass spectra. Bioinformatics 20:1466–1467.

Craig R, Cortens JP, Beavis RC. 2005. The use of proteotypic peptide libraries for protein identification. Rapid Commun Mass Spectrom 19:1844–1850.

Craig R, Cortens JC, Fenyo D, Beavis RC. 2006. Using annotated peptide mass spectrum libraries for protein identification. J Proteome Res 5: 1843–1849.

Cramer R, Corless S. 2001. The nature of collision-induced dissociation processes of doubly protonated peptides: Comparative study for the future use of matrix-assisted laser desorption/ionization on a hybrid quadrupole time-of-flight mass spectrometer in proteomics. Rapid Commun Mass Spectrom 15:2058–2066.

Cremer F, Vandewalle C. 1985. Method for extraction of proteins from green plant-tissues for two-dimensional polyacrylamide-gel electrophoresis. Anal Biochem 147:22–26.

Damerval C, Devienne D, Zivy M, Thiellement H. 1986. Technical improvements in two-dimensional electrophoresis increase the level of genetic-variation detected in wheat-seedling proteins. Electrophoresis 7:52–54.

Damerval C, Zivy M, Granier F, de Vienne D. 1988. Two-dimensional electrophoresis in plant biology. In: Chrambach A, Dunn MJ, Radola BJ, editors. Advances in electrophoresis. Weinheim: VCH. pp 265–280.

Dancik V, Addona TA, Clauser KR, Vath JE, Pevzner PA. 1999. *De novo* peptide sequencing via tandem mass spectrometry. J Comput Biol 6:327–342.

Delaplace P, Van Der Wal F, Dierick JF, Cordewener JHG, Fauconnier ML, Du Jardin P, America AHR. 2006. Potato tuber proteomics: Comparison of two complementary extraction methods designed for 2-DE of acidic proteins. Proteomics 6:6494–6497.

Delincee H, Radola BJ. 1970. Thin-layer isoelectric focusing on Sephadex layers of horseradish peroxidase. Biochim Biophys Acta 200:404–407.

Demeure K, Quinton L, Gabelica V, DePauw E. 2007. Rational selection of the optimum MALDI matrix for top-down proteomics by in-source decay. Anal Chem 79:8678–8685.

Dongre AR, Jones JL, Somogyi A, Wysocki VH. 1996. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. J Am Chem Soc 118:8365–8374.

Dormeyer W, Mohammed S, Breukelen BV, Krijgsveld J, Heck AJ. 2007. Targeted analysis of protein termini. J Proteome Res 6:4634–4645.

Ducret A, Van Oostveen I, Eng JK, Yates JR, Aebersold R. 1998. High throughput protein characterization by automated reverse-phase chromatography electrospray tandem mass spectrometry. Protein Sci 7:706–719.

Eley MH, Burns PC, Kannapell CC, Campbell PS. 1979. Cetyltrimethylammonium bromide polyacrylamide-gel electrophoresis—Estimation of protein subunit molecular-weights using cationic detergents. Anal Biochem 92:411–419.

Elias JE, Gygi SP. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4:207–214.

Endler A, Meyer S, Schelbert S, Schneider T, Weschke W, Peters SW, Keller F, Baginsky S, Martinoia E, Schmidt UG. 2006. Identification of a

vacuolar sucrose transporter in barley and Arabidopsis mesophyll cells by a tonoplast proteomic approach. Plant Physiol 141:196–207.

Ephritikhine G, Ferro M, Rolland N. 2004. Review: Plant membrane proteomics. Plant Physiol Biochem 42:943–962.

Espagne C, Martinez A, Valot B, Meinnel T, Giglione C. 2007. Alternative and effective proteomic analysis in Arabidopsis. Proteomics 7:3788–3799.

Eubel H, Braun H-P, Millar AH. 2005. Blue-native PAGE in plants: A tool in analysis of protein–protein interactions. Plant Methods 1:11.

Eubel H, Lee C, Kuo J, Meyer EH, Taylor NL, Millar AH. 2007. Free-flow electrophoresis for purification of plant mitochondria by surface charge. Plant J 52:583–594.

Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E. 2006. Evaluation of cell wall preparations for proteomics: A new procedure for purifying cell walls from Arabidopsis hypocotyls. Plant Methods 2:10.

Fernandez-De-Cossio J, Gonzalez J, Betancourt L, Besada V, Padron G, Shimonishi Y, Takao T. 1998. Automated interpretation of high-energy collision-induced dissociation spectra of singly protonated peptides by 'SeqMS', a software aid for *de novo* sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 12:1867–1878.

Ferro M, Seigneurin-Berny D, Rolland N, Chapel A, Salvi D, Garin J, Joyard J. 2000. Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. Electrophoresis 21: 3517–3526.

Ferro M, Salvi Dl, Riviere-Rolland H, Vermat T, Seigneurin-Berny D, Grunwald Di, Garin J, Joyard J, Rolland N. 2002. Integral membrane proteins of the chloroplast envelope: Identification and subcellular localization of new transporters. Proc Natl Acad Sci USA 99:11487–11492.

Ferro M, Salvi D, Brugiere S, Miras S, Kowalski S, Louwagie M, Garin J, Joyard J, Rolland N. 2003. Proteomics of the chloroplast envelope membranes from *Arabidopsis thaliana*. Mol Cell Proteomics 2:325–345.

Ferry-Dumazet H, Houel G, Montalent P, Moreau L, Langella O, Negroni L, Vincent D, Lalanne C, De Daruvar A, Plomion C, Zivy M, Joets J. 2005. PROTICdb: A web-based application to store, track, query, and compare plant proteome data. Proteomics 5:2069–2081.

Figeys D. 2003. Novel approaches to map protein interactions. Curr Opin Biotechnol 14:119–125.

Flengsrud R, Kobro G. 1989. A method for two-dimensional electrophoresis of proteins from green plant-tissues. Anal Biochem 177:33–36.

Frank A, Pevzner P. 2005. PepNovo: *De novo* peptide sequencing via probabilistic network modeling. Anal Chem 77:964–973.

Frank AM, Savitski MM, Nielsen ML, Zubarev RA, Pevzner PA. 2007. *De novo* peptide sequencing and identification with precision mass spectrometry. J Proteome Res 6:114–123.

Frewen BE, Merrihew GE, Wu CC, Noble WS, Maccoss MJ. 2006. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. Anal Chem 78:5678–5684.

Friso G, Giacomelli L, Ytterberg AJ, Peltier J-B, Rudella A, Sun Q, van Wijk KJ. 2004. In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: New proteins, new functions, and a plastid proteome database. Plant Cell 16:478–499.

Gatlin CL, Kleemann GR, Hays LG, Link AJ, Yates JR. 1998. Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography microspray and nanospray mass spectrometry. Anal Biochem 263:93–101.

Gevaert K, Demol H, Verschelde JL, Vandamme J, Deboeck S, Vandekerckhove J. 1997. Novel techniques for identification and characterization of proteins loaded on gels in femtomole amounts. J Protein Chem 16:335–342.

Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, Vandekerckhove J. 2003. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nat Biotechnol 21:566–569.

Gion JM, Lalanne C, Le Provost G, Ferry-Dumazet H, Paiva J, Chaumeil P, Frigerio JM, Brach J, Barre A, De Daruvar A, Claverol S, Bonneu M, Sommerer N, Negroni L, Plomion C. 2005. The proteome of maritime pine wood forming tissue. Proteomics 5:3731–3751.

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchinson D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong JP, Miguel T, Paszkowski U, Zhang SP, Colbert M, Sun WL, Chen LL, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu YS, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp japonica). Science 296:92–100.

Gomez SM, Nishio JN, Faull KF, Whitelegge JP. 2002. The chloroplast grana proteome defined by intact mass measurements from liquid chromatography mass spectrometry. Mol Cell Proteomics 1:46–59.

Gooding PS, Bird C, Robinson SP. 2001. Molecular cloning and characterisation of banana fruit polyphenol oxidase. Planta 213:748–757.

Gorg A, Obermaier C, Boguth G, Csordas A, Diaz JJ, Madjar JJ. 1997. Very alkaline immobilized pH gradients for two-dimensional electrophoresis of ribosomal and nuclear proteins. Electrophoresis 18:328–337.

Gorg A, Boguth G, Kopf A, Reil G, Parlar H, Weiss W. 2002. Sample prefractionation with sephadex isoelectric focusing prior to narrow pH range two-dimensional gels. Proteomics 2:1652–1657.

Gorg A, Weiss W, Dunn MJ. 2004. Current two-dimensional electrophoresis technology for proteomics. Proteomics 4:3665–3685.

Granier F. 1988. Extraction of plant-proteins for two-dimensional electrophoresis. Electrophoresis 9:712–718.

Granvogl B, Reisinger V, Eichacker LA. 2006. Mapping the proteome of thylakoid membranes by *de novo* sequencing of intermembrane peptide domains. Proteomics 6:3681–3695.

Grossmann J, Fisher B, Baerenfaller K, Owiti J, Buhmann J, Gruissem W, Baginsky S. 2007. A workflow to increase the deection rate of proteins from unsequenced organisms in high-throughput proteomics experiments. Proteomics 7:4245–4254.

Gu S, Pan SQ, Bradbury EM, Chen X. 2002. Use of deuterium-labeled lysine for efficient protein identification and peptide *de novo* sequencing. Anal Chem 74:5774–5785.

Gupta N, Tanner S, Jaitly N, Adkins JN, Lipton M, Edwards R, Romine M, Osterman A, Bafna V, Smith RD, Pevzner PA. 2007. Whole proteome analysis of post-translational modifications: Applications of mass-spectrometry for proteogenomic annotation. Genome Res 17:1362–1377.

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

Gygi SP, Rist B, Griffin TJ, Eng J, Aebersold R. 2002. Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. J Proteome Res 1:47–54.

Hajduch M, Casteel JE, Tang SX, Hearne LB, Knapp S, Thelen JJ. 2007. Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. J Proteome Res 6:3232–3241.

Hari V. 1981. A method for the two-dimensional electrophoresis of leaf proteins. Anal Biochem 113:332–335.

Hartinger J, Stenius K, Hogemann D, Jahn R. 1996. 16-BAC/SDS-PAGE: A two-dimensional gel electrophoresis system suitable for the separation of integral membrane proteins. Anal Biochem 240:126–133.

Haynes PA, Roberts TH. 2007. Subcellular shotgun proteomics in plants: Looking beyond the usual suspects. Proteomics 7:2963–2975.

Hedges SB. 2002. The origin and evolution of model organisms. Nat Rev Genet 3:838–849.

Helm M, Luck C, Prestele J, Hierl G, Huesgen PF, Froehlich T, Arnold GJ, Adamska I, Gorg A, Lottspeich F, Gietl C. 2007. Dual specificities of the glyoxysomal/peroxisomal processing protease Deg15 in higher plants. Proc Natl Acad Sci USA 104:11501–11506.

Henzel W, Billeci T, Stults J, Wong S, Grimley C, Watanabe C. 1993. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci USA 90:5011–5015.

Herbert B, Galvani M, Hamdan M, Olivieri E, Maccarthy J, Pedersen S, Righetti PG. 2001. Reduction and alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how? Electrophoresis 22:2046–2057.

Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. Ann Bot 100:1073–1084.

Higashi Y, Hirai MY, Fujiwara T, Naito S, Noji M, Saito K. 2006. Proteomic and transcriptomic analysis of Arabidopsis seeds: Molecular evidence for successive processing of seed proteins and its implication in the stress response to sulfur nutrition. Plant J 48:557–571.

Hink MA, Bisseling T, Visser A. 2002. Imaging protein-protein interactions in living cells. Plant Mol Biol 50:871–883.

Hoving S, Gerrits B, Voshol H, Muller D, Roberts RC, Van Oostrum J. 2002. Preparative two-dimensional gel electrophoresis at alkaline pH using narrow range immobilized pH gradients. Proteomics 2:127–134.

Hu CD, Chinenov Y, Kerppola TK. 2002. Visualization of interactions among bZip and Rel family proteins in living cells using bimolecular fluorescence complementation. Mol Cell 9:789–798.

Huang L, Jacob RJ, Pegg SCH, Baldwin MA, Wang CC, Burlingame AL, Babbitt PC. 2001. Functional assignment of the 20 S proteasome from *Trypanosoma brucei* using mass spectrometry and new bioinformatics approaches. J Biol Chem 276:28327–28339.

Huang L, Harvie G, Feitelson JS, Gramatikoff K, Herold DA, Allen DL, Amunngama R, Hagler RA, Pisano MR, Zhang WW, Fang XM. 2005. Immunoaffinity separation of plasma proteins by IgY microbeads: Meeting the needs of proteomic sample preparation and analysis. Proteomics 5:3314–3328.

Hurkman WJ, Tanaka CK. 1986. Solubilization of plant membrane-proteins for analysis by two-dimensional gel-electrophoresis. Plant Physiol 81:802–806.

Hurkman WJ, Tanaka CK. 2004. Improved methods for separation of wheat endosperm proteins and analysis by two-dimensional gel electrophoresis. J Cereal Sci 40:295–299.

Huttlin EL, Hegeman AD, Harms AC, Sussman MR. 2007. Prediction of error associated with false-positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. J Proteome Res 6:392–398.

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

Jaquinod M, Villiers F, Kieffer-Jaquinod S, Hugouvieu V, Bruley C, Garin J, Bourguignon J. 2007. A proteomics dissection of *Arabidopsis thaliana* vacuoles isolated from cell culture. Mol Cell Proteomics 6:394–412.

Johnson RS, Biemann K. 1989. Computer-program (seqpep) to aid in the interpretation of high-energy collision tandem mass-spectra of peptides. Biomed Environ Mass Spectrom 18:945–957.

Jones JL, Dongre AR, Somogyi A, Wysocki VH. 1994. Sequence dependence of peptide fragmentation efficiency curves determined by electrospray-ionization surface-induced dissociation mass-spectrometry. J Am Chem Soc 116:8368–8369.

Jorge I, Navarro RM, Lenz C, Ariza D, Porras C, Jorrin J. 2005. The holm oak leaf proteome: Analytical and biological variability in the protein expression level assessed by 2-DE and protein identification tandem mass spectrometry *de novo* sequencing and sequence similarity searching. Proteomics 5:222–234.

Jorrín JV, Maldonado AM, Castillejo MA. 2007. Plant proteome analysis: A 2006 update. Proteomics 7:2947–2962.

Joyard J, Grossman A, Bartlett SG, Douce R, Chua NH. 1982. Characterization of envelope membrane polypeptides from spinach chloroplasts. J Biol Chem 257:1095–1101.

Jung K, Ganooun A, Sitek B, Apostolov O, Schramm A, Meyer H, Stühler K, Urfer W. 2006. Statistical evaluation of methods for the analysis of dynamic protein expression data from a tumor study. Stat J 4:67–80.

Kao SH, Hsu CH, Su SN, Hor WT, Chang WHT, Chow LP. 2004. Identification and immunologic characterization of an allergen, alliin lyase, from garlic (*Allium sativum*). J Allergy Clin Immun 113:161–168.

Karp NA, Lilley KS. 2007. Design and analysis issues in quantitative proteomics studies. Proteomics 7:42–50.

Katz A, Waridel P, Shevchenko A, Pick U. 2007. Salt-induced changes in the plasma membrane proteome of the halotolerant alga Dunaliella salina as revealed by blue native gel electrophoresis and nano-LC-MS/MS analysis. Mol Cell Proteomics 6:1459–1472.

Kaufmann R, Chaurand P, Kirsch D, Spengler B. 1996. Post-source decay and delayed extraction in matrix-assisted laser desorption/ionization/reflectron time-of-flight mass spectrometry. Are there trade-offs? Rapid Commun Mass Spectrom 10:1199–1208.

Keough T, Youngquist RS, Lacey MP. 1999. A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry. Proc Natl Acad Sci USA 96:7131–7136.

Keough T, Lacey MP, Youngquist RS. 2000. Derivatization procedures to facilitate *de novo* sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. Rapid Commun Mass Spectrom 14:2348–2356.

Keough T, Lacey MP, Fieno AM, Grant RA, Sun YP, Bauer MD, Begley KB. 2000. Tandem mass spectrometry methods for definitive protein identification in proteomics research. Electrophoresis 21:2252–2265.

Keough T, Lacey MP, Strife RJ. 2001. Atmospheric pressure matrix-assisted laser desorption/ionization ion trap mass spectrometry of sulfonic acid derivatized tryptic peptides. Rapid Commun Mass Spectrom 15:2227–2239.

Khan MMK, Komatsu S. 2004. Rice proteomics: Recent developments and analysis of nuclear proteins. Phytochemistry 65:1671–1681.

Kim ST, Cho KS, Jang YS, Kang KY. 2001. Two-dimensional electrophoretic analysis of rice proteins by polyethylene glycol fractionation for protein arrays. Electrophoresis 22:2103–2109.

Kjell J, Rasmusson AG, Larsson H, Widell S. 2004. Protein complexes of the plant plasma membrane resolved by blue native page. Physiol Plant 121:546–555.

Kosaka T, Takazawa T, Nakamura T. 2000. Identification and C-terminal characterization of proteins from two-dimensional polyacrylamide gels by a combination of isotopic labeling and nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry. Anal Chem 72:1179–1185.

Krause F. 2006. Detection and analysis of protein–protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (membrane) protein complexes and supercomplexes. Electrophoresis 27:2759–2781.

Krogh M, Fernandez C, Teilum M, Bengtsson S, James P. 2007. A probabilistic treatment of the missing spot problem in 2D gel electrophoresis experiments. J Proteome Res 6:3335–3343.

Laugesen S, Bak-Jensen KS, Hagglund P, Henriksen A, Finnie C, Svensson B, Roepstorff P. 2007. Barley peroxidase isozymes—Expression and post-translational modification in mature seeds as identified by two-dimensional gel electrophoresis and mass spectrometry. Int J Mass Spectrom 268:244–253.

Laukens K, Deckers P, Esmans E, Van Onckelen H, Witters E. 2004. Construction of a two-dimensional gel electrophoresis protein database

for the *Nicotiana tabacum* Cv. Bright Yellow-2 cell suspension culture. Proteomics 4:720–727.

Laukens K, Matthiesen R, Lemiere F, Esmans E, Onckelen HV, Jensen ON, Witters E. 2006. Integration of gel-based proteome data with pProRep. Bioinformatics 22:2838–2840.

Lester PJ, Hubbard SJ. 2002. Comparative bioinformatic analysis of complete proteomes and protein parameters for cross-species identification in proteomics. Proteomics 2:1392–1405.

Lilley KS, Dupree P. 2006. Methods of quantitative proteomics and their application to plant organelle characterization. J Exp Bot 57:1493–1499.

Lilley KS, Dupree P. 2007. Plant organelle proteomics: Cell biology. Curr Opin Plant Biol 10:594–599.

Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR. 1999. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol 17:676–682.

Liska AJ, Shevchenko A. 2003. Expanding the organismal scope of proteomics: Cross-species protein identification by mass spectrometry and its implications. Proteomics 3:19–28.

Liska AJ, Sunyaev S, Shilov IN, Schaeffer DA, Shevchenko A. 2005. Error-tolerant EST database searches by tandem mass spectrometry and MultiTag software. Proteomics 5:4118–4122.

Liu J, Bell AW, Bergeron JJM, Yanofsky CM, Carrillo B, Beaudrie CEH, Kearney RE. 2007. Methods for peptide identification by spectral comparison. Proteome Sci vol. 5 article 3.

Loomis WD, Bataille J. 1966. Plant phenolic compounds and the isolation of plant enzymes. Phytochemistry 5:423–438.

Lopez MF. 2000. Better approaches to finding the needle in a haystack: Optimizing proteome analysis through automation. Electrophoresis 21:1082–1093.

Lysak MA, Dolezelova M, Horry JP, Swennen R, Dolezel J. 1999. Flow cytometric analysis of nuclear DNA content in *Musa*. Theor Appl Genet 98:1344–1350.

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

Macfarlane DE. 1983. Use of benzyldimethyl-normal-hexadecylammonium chloride (16-BAC), a cationic detergent, in an acidic polyacrylamide-gel electrophoresis system to detect base labile protein methylation in intact-cells. Anal Biochem 132:231–235.

Macfarlane DE. 1989. Two-dimensional benzyldimethyl-*N*-hexadecylammonium chloride-sodium dodecyl-sulfate preparative polyacrylamide-gel electrophoresis—A high-capacity high-resolution technique for the purification of proteins from complex-mixtures. Anal Biochem 176: 457–463.

Mackey AJ, Haystead TAJ, Pearson WR. 2002. Getting more from less—Algorithms for rapid protein identification with multiple short peptide sequences. Mol Cell Proteomics 1:139–147.

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

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

Marmagne A, Rouet M-A, Ferro M, Rolland N, Alcon C, Joyard J, Garin J, Barbier-Brygoo H, Ephritikhine G. 2004. Identification of new intrinsic proteins in Arabidopsis plasma membrane proteome. Mol Cell Proteomics 3:675–691.

Martin RL, Brancia FL. 2003. Analysis of high mass peptides using a novel matrix-assisted laser desorption/ionisation quadrupole ion trap time-of-flight mass spectrometer. Rapid Commun Mass Spectrom 17:1358–1365.

Maserti BE, Della Croce CM, Luro F, Morillon R, Cini M, Caltavuturo L. 2007. A general method for the extraction of citrus leaf proteins and separation by 2D electrophoresis: A follow up. J Chromatogr B 849:351–356.

Mata J, Marguerat S, Bahler J. 2005. Post-transcriptional control of gene expression: A genome-wide perspective. Trends Biochem Sci 30:506–514.

Mathesius U, Imin N, Chen HC, Djordjevic MA, Weinman JJ, Natera SHA, Morris AC, Kerim T, Paul S, Menzel C, Weiller GR, Rolfe BG. 2002. Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. Proteomics 2:1288–1303.

Matthiesen R, Bunkenborg J, Stensballe A, Jensen ON, Welinder KG, Bauw G. 2004. Database-independent, database-dependent, and extended interpretation of peptide mass spectra in VEMS V2.0. Proteomics 4:2583–2593.

McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR. 1997. Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. Anal Chem 69:767–776.

Méchin V, Consoli L, Le Guilloux M, Damerval C. 2003. An efficient solubilization buffer for plant proteins focused in immobilized pH gradients. Proteomics 3:1299–1302.

Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL. 2000. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. Anal Chem 72:552–558.

Meyer Y, Grosset J, Chartier Y, Cleyetmarel JC. 1988. Preparation by two-dimensional electrophoresis of proteins for antibody-production—Antibodies against proteins whose synthesis is reduced by auxin in tobacco mesophyll protoplasts. Electrophoresis 9:704–712.

Millar AH, Heazlewood JL, Kristensen BK, Braun HP, Moller IM. 2005. The plant mitochondrial proteome. Trends Plant Sci 10:36–43.

Mooney BP, Thelen JJ. 2004. High-throughput peptide mass fingerprinting of soybean seed proteins: Automated workflow and utility of UniGene expressed sequence tag databases for protein identification. Phytochemistry 65:1733–1744.

Morel J, Claverol S, Mongrand S, Furt F, Fromentin J, Bessoule JJ, Blein JP, Simon-Plas F. 2006. Proteomics of plant detergent-resistant membranes. Mol Cell Proteomics 5:1396–1411.

Moritz B, Meyer HE. 2003. Approaches for the quantification of protein concentration ratios. Proteomics 3:2208–2220.

Nandakumar MP, Shen J, Raman B, Marten MR. 2003. Solubilization of trichloroacetic acid (TCA) precipitated microbial proteins via NaOH for two-dimensional electrophoresis. J Proteome Res 2:89–93.

O'Farrell PH. 1975. High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021.

Oba S, Sato M, Takemasa I, Monden M, Matsubara K, Ishii S. 2003. A Bayesian missing value estimation method for gene expression profile data. Bioinformatics 19:2088–2096.

Olsson I, Larsson K, Palmgren R, Bjellqvist B. 2002. Organic disulfides as a means to generate streak-free two-dimensional maps with narrow range basic immobilized pH gradient strips as first product dimension. Proteomics 2:1630–1632.

Paizs B, Suhai S. 2005. Fragmentation pathways of protonated peptides. Mass Spectrom Rev 24:508–548.

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

Pedreschi R, Vanstreels E, Carpentier S, Hertog M, Lammertyn J, Robben J, Noben JP, Swennen R, Vanderleyden J, Nicolai BM. 2007. Proteomic analysis of core breakdown disorder in Conference pears (*Pyrus communis* L.). Proteomics 7:2083–2099.

Pedreschi R, Hertog M, Carpentier SC, Lammertyn J, Robben J, Noben J-P, Panis B, Swenen R, Nicolai BM. in press. Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. Proteomics 8(7):1371–1383.

Peltier JB, Ytterberg AJ, Sun Q, Van Wijk KJ. 2004. New functions of the thylakoid membrane proteome of *Arabidopsis thaliana* revealed by a simple, fast, and versatile fractionation strategy. J Biol Chem 279: 49367–49383.

Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JWS, Shaw PJ. 2005. Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol Biol Cell 16:260–269.

Penin F, Godinot C, Gautheron DC. 1984. Two-dimensional gel-electrophoresis of membrane-proteins using anionic and cationic detergents—Application to the study of mitochondrial f0-f1-ATPase. Biochim Biophys Acta 775:239–245.

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

Pevtsov S, Fedulova I, Mirzaei H, Buck C, Zhang X. 2006. Performance evaluation of existing *de novo* sequencing algorithms. J Proteome Res 5:3018–3028.

Qin J, Chait BT. 1995. Preferential fragmentation of protonated gas-phase peptide ions adjacent to acidic amino-acid-residues. J Am Chem Soc 117:5411–5412.

Rabilloud T, Adessi C, Giraudel A, Lunardi J. 1997. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 18:307–316.

Radola BJ. 1975. Isoelectric focusing in layers of granulated gels II. Preparative isoelectric focusing. Biochim Biophys Acta 386:181–185.

Rais I, Karas M, Schagger H. 2004. Two-dimensional electrophoresis for the isolation of integral membrane proteins and mass spectrometric identification. Proteomics 4:2567–2571.

Regnier FE, Julka S. 2006. Primary amine coding as a path to comparative proteomics. Proteomics 6:3968–3979.

Reisinger V, Eichacker LA. 2006. Analysis of membrane protein complexes by blue native PAGE. Proteomics 6:S6–S15.

Reisinger V, Eichacker LA. 2007. How to analyze protein complexes by 2D blue native SDS-PAGE. Proteomics 7:S6–S16.

Renaut J, Hausman J-F, Bassett C, Artlip T, Cauchie H-M, Witteres E, Wisniewski M. Quantitative proteomic analysis of short photoperiod and low temperature responses in bark tissues of peach (*Prunus persica* L. Batsch). Tree Genet Genom (in press).

Righetti PG, Castagna A, Antonioli P, Boschetti E. 2005a. Prefractionation techniques in proteome analysis: The mining tools of the third millennium. Electrophoresis 26:297–319.

Righetti PG, Castagna A, Herbert B, Candiano G. 2005b. How to bring the ''unseen'' proteome to the limelight via electrophoretic pre-fractionation techniques. Biosci Rep 25:3–17.

Rolland N, Ferro M, Ephritikhine G, Marmagne A, Ramus C, Brugiere S, Salvi D, Seigneurin-Berny D, Bourguignon J, Barbier-Brygoo H, Joyard J, Garin J. 2006. A versatile method for deciphering plant membrane proteomes. J Exp Bot 57:1579–1589.

Sakurai T, Matsuo T, Matsuda H, Katakuse I. 1984. Paas-3—A computer-program to determine probable sequence of peptides from mass-spectrometric data. Biomed Mass Spectrom 11:396–399.

Salekdeh GH, Komatsu S. 2007. Crop proteomics: Aim at sustainable agriculture of tomorrow. Proteomics 7:2976–2996.

Samyn B, Debyser G, Sergeant K, Devreese B, Van Beeumen J. 2004. A case study of *de novo* sequence analysis of N-sulfonated peptides by MALDI TOF/TOF mass spectrometry. J Am Soc Mass Spectrom 15:1838–1852.

Samyn B, Sergeant K, Castanheira P, Faro C, Van Beeumen J. 2005. A new method for C-terminal sequence analysis in the proteomic era. Nat Methods 2:193–200.

Samyn B, Sergeant K, Van Beeumen J. 2006. A method for C-terminal sequence analysis in the proteomic era (proteins cleaved with cyanogen bromide). Nat Protocols 1:317–321.

Samyn B, Sergeant K, Carpentier S, Debyser G, Panis B, Swennen R, Van Beeumen J. 2007. Functional proteome analysis of the banana plant (Musa spp.) using *de novo* sequence analysis of derivatized peptides. J Proteome Res 6:70–80.

Santoni V, Molloy M, Rabilloud T. 2000. Membrane proteins and proteomics: Un amour impossible? Electrophoresis 21:1054–1070.

Saravanan RS, Rose JKC. 2004. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. Proteomics 4:2522–2532.

Schmidt UG, Endler A, Schelbert S, Brunner A, Schnell M, Neuhaus HE, Marty-Mazars D, Marty F, Baginsky S, Martinoia E. 2007. Novel tonoplast transporters identified using a proteomic approach with vacuoles isolated from cauliflower buds. Plant Physiol 145:216–229.

Schroder B, Hasilik A. 2006. A protocol for combined delipidation and subfractionation of membrane proteins using organic solvents. Anal Biochem 357:144–146.

Schubert M, Petersson UA, Haas BJ, Funk C, Schroder WP, Kieselbach T. 2002. Proteome map of the chloroplast lumen of *Arabidopsis thaliana*. J Biol Chem 277:8354–8365.

Schägger H, von Jagow G. 1987. Tricine sodium dodecyl-sulfate polyacrylamide-gel electrophoresis for the separation of proteins in the range from 1-kDa to 100-kDa. Anal Biochem 166:368–379.

Schägger H, von Jagow G. 1991. Blue native electrophoresis for isolation of membrane-protein complexes in enzymatically active form. Anal Biochem 199:223–231.

Schägger H, Cramer WA, von Jagow G. 1994. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane-protein complexes by 2-dimensional native electrophoresis. Anal Biochem 217:220–230.

Searle BC, Dasari S, Turner M, Reddy AP, Choi DS, Wilmarth PA, McCormack AL, David LL, Nagalla SR. 2004. High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS *de novo* sequencing results. Anal Chem 76:2220–2230.

Seigneurin-Berny D, Rolland N, Garin J, Joyard J. 1999. Differential extraction of hydrophobic proteins from chloroplast envelope membranes: A subcellular-specific proteomic approach to identify rare intrinsic membrane proteins. Plant J 19:217–228.

Sergeant K, Samyn B, Debyser G, Van Beeumen J. 2005. *De novo* sequence analysis of N-terminal sulfonated peptides after in-gel guanidination. Proteomics 5:2369–2380.

Sharma S. 1996. Applied multivariate techniques. USA: Wiley.

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

Shevchenko A, Sunyaev S, Loboda A, Shevehenko A, Bork P, Ens W, Standing KG. 2001. Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and BLAST homology searching. Anal Chem 73:1917–1926.

Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, Hunter CL, Nuwaysir LM, Schaeffer DA. 2007. The Paragon algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. Mol Cell Proteomics 6:1638–1655.

Shimaoka T, Ohnishi M, Sazuka T, Mitsuhashi N, Hara-Nishimura I, Shimazaki KI, Maeshima M, Yokota A, Tomizawa KI, Mimura T. 2004.

Isolation of intact vacuoles and proteomic analysis of tonoplast from suspension-cultured cells of *Arabidopsis thaliana*. Plant Cell Physiol 45:672–683.

Shui WQ, Liu YK, Fan HZ, Bao HM, Liang SF, Yang PY, Chen X. 2005. Enhancing TOF/TOF-based *de novo* sequencing capability for high throughput protein identification with amino acid-coded mass tagging. J Proteome Res 4:83–90.

Song J, Braun G, Bevis E, Doncaster K. 2006. A simple protocol for protein extraction of recalcitrant fruit tissues suitable for 2-DE and MS analysis. Electrophoresis 27:3144–3151.

Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A. 2003. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal Bioanal Chem 376:952–965.

Sun Q, Emanuelsson O, Van Wijk KJ. 2004. Analysis of curated and predicted plastid subproteomes of Arabidopsis. Subcellular compartmentalization leads to distinctive proteome properties. Plant Physiol 135:723–734.

Sunyaev S, Liska AJ, Golod A, Shevchenko A, Shevchenko A. 2003. MultiTag: Multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry. Anal Chem 75:1307–1315.

Tabb DL, Smith LL, Breci LA, Wysocki VH, Lin D, Yates JR. 2003. Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. Anal Chem 75:1155–1163.

Tabb DL, Huang YY, Wysocki VH, Yates JR. 2004. Influence of basic residue content on fragment ion peak intensities in low-energy—Collision-induced dissociation spectra of peptides. Anal Chem 76:1243–1248.

Taylor JA, Johnson RS. 1997. Sequence database searches via *de novo* peptide sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 11:1067–1075.

Taylor JA, Johnson RS. 2001. Implementation and uses of automated *de novo* peptide sequencing by tandem mass spectrometry. Anal Chem 73: 2594–2604.

Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB. 2001. Missing value estimation methods for DNA microarrays. Bioinformatics 17:520–525.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van De Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604.

Valcu CM, Schlink K. 2006. Efficient extraction of proteins from woody plant samples for two-dimensional electrophoresis. Proteomics 6:4166–4175.

Van Etten JL, Freer SN, McCune BK. 1979. Presence of a major (storage?) protein in dormant spores of the fungus *Botryodiplodia theobromae*. J Bacteriol 139:650–652.

Van Wijk KJ. 2004. Plastid proteomics. Plant Physiol Biochem 42:963–977.

Vincent D, Wheatley MD, Cramer GR. 2006. Optimization of protein extraction and solubilization for mature grape berry clusters. Electrophoresis 27:1853–1865.

Wang W, Scali M, Vignani R, Spadafora A, Sensi E, Mazzuca S, Cresti M. 2003. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. Electrophoresis 24:2369–2375.

Wang W, Vignani R, Scali M, Cresti M. 2006. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. Electrophoresis 27:2782–2786.

Waridel P, Frank A, Thomas H, Surendranath V, Sunyaev S, Pevzner P, Shevchenko A. 2007. Sequence similarity-driven proteomics in organisms with unknown genomes by LC-MS/MS and automated *de novo* sequencing. Proteomics 7:2318–2329.

Washburn MP, Wolters D, Yates JR. 2001. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19:242–247.

Wattenberg A, Organ AJ, Schneider K, Tyldesley R, Bordoli R, Bateman RH. 2002. Sequence dependent fragmentation of peptides generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification. J Am Soc Mass Spectrom 13:772–783.

Wenzel D, Schauermann G, Von Lupke A, Hinz G. 2005. The cargo in vacuolar storage protein transport vesicles is stratified. Traffic 6:45–55.

Wessel D, Flugge UI. 1984. A method for the quantitative recovery of protein in dilute-solution in the presence of detergents and lipids. Anal Biochem 138:141–143.

Westermeier R. 2006. Preparation of plant samples for 2-D electrophoresis. Proteomics 6:S56–S60.

Whitelegge JP. 2004. Mass spectrometry for high throughput quantitative proteomics in plant research: Lessons from thylakoid membranes. Plant Physiol Biochem 42:919–927.

Wienkoop S, Saalbach G. 2003. Novel proteins identified at the peribacteroid membrane from *Lotus japonicus* root nodules. Plant Physiol 131:1080–1090.

Wilkins MR, Williams KL. 1997. Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: A theoretical evaluation. J Theor Biol 186:7–15.

Wilkins MR, Gasteiger E, Tonella L, Ou K, Tyler M, Sanchez JC, Gooley AA, Walsh BJ, Bairoch A, Appel RD, Williams KL, Hochstrasser DF. 1998a. Protein identification with N and C-terminal sequence tags in proteome projects. J Mol Biol 278:599–608.

Wilkins MR, Gasteiger E, Sanchez JC, Bairoch A, Hochstrasser DF. 1998b. Two-dimensional gel electrophoresis for proteome projects: The effects of protein hydrophobicity and copy number. Electrophoresis 19:1501–1505.

Williams TI, Combs JC, Thakur AR, Strobel HJ, Lynn BC. 2006. A novel bicine running buffer system for doubled sodium dodecyl sulfate—Polyacrylamide gel electrophoresis of 2 membrane proteins. Electrophoresis 27:2984–2995.

Wilm M, Mann M. 1994. Electrospray and taylor-cone theory, doles beam of macromolecules at last. Int J Mass Spectrom 136:167–180.

Wilm M, Neubauer G, Mann M. 1996. Parent ion scans of unseparated peptide mixtures. Anal Chem 68:527–533.

Witters E, Laukens K, Deckers P, Van Dongen W, Esmans E, Van Onckelen H. 2003. Fast liquid chromatography coupled to electrospray tandem mass spectrometry peptide sequencing for cross-species protein identification. Rapid Commun Mass Spectrom 17:2188–2194.

Wu FS, Wang MY. 1984. Extraction of proteins for sodium dodecyl-sulfate polyacrylamide-gel electrophoresis from protease-rich plant-tissues. Anal Biochem 139:100–103.

Wuyts N, De Waele D, Swennen R. 2006. Extraction and partial characterization of polyphenol oxidase from banana (Musa acuminata Grande naine) roots. Plant Physiol Biochem 44:308–314.

Wysocki VH, Tsaprailis G, Smith LL, Breci LA. 2000. Mobile and localized protons: A framework for understanding peptide dissociation. J Mass Spectrom 35:1399–1406.

Xi JH, Wang X, Li SY, Zhou X, Yue L, Fan J, Hao DY. 2006. Polyethylene glycol fractionation improved detection of low-abundant proteins by two-dimensional electrophoresis analysis of plant proteome. Phytochemistry 67:2341–2348.

Xie H, Pan SL, Liu SF, Ye K, Huo KK. 2007. A novel method of protein extraction from perennial Bupleurum root for 2-DE. Electrophoresis 28:871–875.

Xu CX, Ma B. 2006. Software for computational peptide identification from MS-MS data. Drug Discov Today 11:595–600.

Yang PF, Chen H, Liang Y, Shen SH. 2007. Proteomic analysis of de-etiolated rice seedlings upon exposure to light. Proteomics 7:2459–2468.

Yao Y, Yang YW, Liu JY. 2006. An efficient protein preparation for proteomic analysis of developing cotton fibers by 2-DE. Electrophoresis 27:4559–4569.

Yates JR, Speicher S, Griffin PR, Hunkapiller T. 1993. Peptide mass maps: A highly informative approach to protein identification. Anal Biochem 214:397–408.

Yergey AL, Coorssen JR, Backlund PS, Blank PS, Humphrey GA, Zimmerberg J, Campbell JM, Vestal ML. 2002. *De novo* sequencing of peptides using MALDI/TOF-TOF. J Am Soc Mass Spectrom 13:784–791.

Yu HY, Luscombe NM, Lu HX, Zhu XW, Xia Y, Han JDJ, Bertin N, Chung S, Vidal M, Gerstein M. 2004. Annotation transfer between genomes: Protein–protein interologs and protein-DNA regulogs. Genome Res 14:1107–1118.

Yu J, Hu SN, Wang J, Wong GKS, Li SG, Liu B, Deng YJ, Dai L, Zhou Y, Zhang XQ, Cao ML, Liu J, Sun JD, Tang JB, Chen YJ, Huang XB, Lin W, Ye C, Tong W, Cong LJ, Geng JN, Han YJ, Li L, Li W, Hu GQ, Huang XG, Li WJ, Li J, Liu ZW, Li L, Liu JP, Qi QH, Liu JS, Li L, Li T, Wang XG, Lu H, Wu TT, Zhu M, Ni PX, Han H, Dong W, Ren XY, Feng XL, Cui P, Li XR, Wang H, Xu X, Zhai WX, Xu Z, Zhang JS, He SJ, Zhang JG, Xu JC, Zhang KL, Zheng XW, Dong JH, Zeng WY, Tao L, Ye J, Tan J, Ren XD, Chen XW, He J, Liu DF, Tian W, Tian CG, Xia HG, Bao QY, Li G, Gao H, Cao T, Wang J, Zhao WM, Li P, Chen W, Wang XD, Zhang Y, Hu JF, Wang J, Liu S, Yang J, Zhang GY, Xiong YQ, Li ZJ, Mao L, Zhou CS, Zhu Z, Chen RS, Hao BL, Zheng WM, Chen SY, Guo W, Li GJ, Liu SQ, Tao M, Wang J, Zhu LH, Yuan LP, Yang HM. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp indica). Science 296:79–92.

Zabrouskov V, Giacomelli L, Van Wijk KJ, Mclafferty FW. 2003. New approach for plant proteomics—Characterization of chloroplast proteins of *Arabidopsis thaliana* by top-down mass spectrometry. Mol Cell Proteomics 2:1253–1260.

Zhang WZ, Chait BT. 2000. Profound: An expert system for protein identification using mass spectrometric peptide mapping information. Anal Chem 72:2482–2489.

Zheng Q, Song J, Doncaster K, Rowland E, Byers DM. 2007. Qualitative and quantitative evaluation of protein extraction protocols for apple and strawberry fruit suitable for two-dimensional electrophoresis and mass spectrometry analysis. J Agric Food Chem 55:1663–1673.

Zhou HL, Ranish JA, Watts JD, Aebersold R. 2002. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. Nat Biotechnol 20:512–515.

Zhou XW, Blackman MJ, Howell SA, Carruthers VB. 2004. Proteomic analysis of cleavage events reveals a dynamic two-step mechanism for proteolysis of a key parasite adhesive complex. Mol Cell Proteomics 3:565–576.

---

**Sebastien Carpentier** studied bioscience engineering at the KULeuven. He then joined the Centre of Human Genetics where he worked for Prof. Marynen developing a chromosomal vector for controlled trangene dosage and gene therapy. Sebastien joined the team of Prof. Swennen in 2003 and started his Ph.D on the development of proteomic methods to unravel biochemical processes in bananas. He obtained his Ph.D in bioscience engineering from the KULeuven in 2007. As a post-doctoral fellow, his current research focus is proteomics methods for non-model plants, applied statistics, and abiotic stress physiology.