

REGULAR ARTICLE

Proteome analysis of *Arabidopsis thaliana* by two-dimensional gel electrophoresis and matrix-assisted laser desorption/ionisation-time of flight mass spectrometry

Patrick Giavalisco^{1,2*}, Eckhard Nordhoff¹, Thomas Kreitler¹, Klaus-Dieter Klöppel¹, Hans Lehrach¹, Joachim Klose^{1,2} and Johan Gobom¹

¹ Max-Planck-Institute for Molecular Genetics, Berlin, Germany

² Institute for Human Genetics, Humboldt University Berlin, Berlin, Germany

In the present study we show results of a large-scale proteome analysis of the recently sequenced plant *Arabidopsis thaliana*. On the basis of a previously published sequential protein extraction protocol, we prepared protein extracts from eight different *A. thaliana* tissues (primary leaf, leaf, stem, silique, seedling, seed, root, and inflorescence) and analysed these by two-dimensional gel electrophoresis. A total of 6000 protein spots, from three of these tissues, namely primary leaf, silique and seedling, were excised and the contained proteins were analysed by matrix assisted laser desorption/ionisation time of flight mass spectrometry peptide mass fingerprinting. This resulted in the identification of the proteins contained in 2943 spots, which were found to be products of 663 different genes. In this report we present and discuss the methodological and biological results of our plant proteome analysis.

Received: June 23, 2004

Revised: August 27, 2004

Accepted: October 1, 2004

**Keywords:**

Arabidopsis thaliana / Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry / Two-dimensional gel electrophoresis

1 Introduction

After having sequenced completely several pro- and eukaryotic genomes (<http://www.genomesonline.org/>), including those of two higher land plants, *Arabidopsis thaliana* [1] and rice [2, 3], it became obvious that to understand the function of the contained genes, large technical and methodological efforts are required to identify and analyse all their gene products, including basic properties such as tissue distribution, relative quantity, or detection of different isoforms. For the purpose of large-scale mRNA expression analysis, differ-

ent microarray techniques were developed [4–6]. However, comparisons of mRNA expression and protein expression alone is no substitute for detailed protein analysis, since, in general, no linear correlation between RNA transcription and protein abundance exists [7–9] and no information on post-translational protein modifications can be deduced. Currently, the techniques most often employed for analysing large protein populations, are 2-DE [10, 11] and HPLC [12, 13]. Both techniques can be coupled, either with ESI MS or MALDI-TOF MS, to identify proteins.

In preceding plant proteome studies, different combinations of these techniques have been successfully employed for the elucidation of different plant physiological processes like germination [14] or senescence [15]. A number of studies have focused on specific organelles, such as the chloroplast [16, 17], mitochondria [18, 19] and the nucleus [20, 21], and

Correspondence: Dr. Johan Gobom, Max-Planck-Institute for Molecular Genetics, Ihnestraße 73, D-14195 Berlin, Germany

Email: gobom@molgen.mpg.de

Fax: +49 (0) 30 8413-1690

Abbreviation: LSU, large subunit of riblose 1, 5 bisphosphate carboxylase/oxygenase.

* Current address: Boyce Thompson Institute for Plant Research, Ithaca, New York, USA

plant subcellular compartments, such as the cell wall [22, 23], the phloem [24] and the plasma membrane [25, 26]. Apart from an extensive analysis of the rice proteome employing HPLC and 2-DE based protein separation [27], no global proteomic studies have been published thus far.

The aim of the current work was therefore to fill this gap and reveal a large part of the expressed proteins from the widely studied model plant *A. thaliana*. For this purpose we selected the large-gel 2-DE protein separation technique [28, 29], which offers a high resolution for protein separation, in combination with MALDI-TOF MS-based protein identification. To get access to a large number of functional and morphological divergent proteins we analysed different tissues of the model plant (primary leaf, leaf, stem, silique, seedling, seed, root, and inflorescence). The large number of selected tissues was chosen to achieve a representative overview of the different functional organs and tissues of the plant.

On the basis of a functional work flow, which included a newly developed tissue protein fractionation protocol [30], different protein separation and staining techniques (Gialvalisco *et al.*, manuscript in preparation) and robotic techniques for spot excision and sample handling, including new software tools for improved MS-based protein identification [31–33], we constructed a set of tissue specific 2-DE protein reference maps, and analysed the proteins in more than 6000 spots by MALDI-TOF MS [21]. In the current paper we will provide data that illustrates the large-scale analysis of the proteome of *A. thaliana*, including the preparation of reference maps, comprehensive image analysis and comparisons of the properties and functions of the identified proteins.

2 Materials and methods

2.1 Plant material

The plant material used in these experiments was sampled from *A. thaliana* ecotype Columbia 0, provided by Dr. Christina Walz from the Max Planck Institute for Molecular Plant Physiology in Golm, Germany. Seedlings were harvested from seeds placed on wetted tissue, 7 d after exposure. Primary leaves were harvested from plants 14 d after sowing. Leaf, root, inflorescence and stem tissue were harvested 50 d after the seeds were sown, while silique tissue was harvested 58 d after sowing. Seed tissue was harvested 64 d after the initial seeds were sown. The growing conditions for the plants were long day (16 h light and 8 h dark) and cold night (day temperature 20°C, night temperature 6°C), for the first 7 d after sowing. Then the plants were grown for 8 d in short day (8 h light and 16 h dark) with warm night conditions (20°C day and 16°C night), before they were finally placed in the greenhouse with long day and warm night conditions. The whole plant material, except the seeds and inflorescence tissue, was washed once with cold water and then stored at –80°C prior to protein extraction.

2.2 Protein extraction

The extraction of proteins from *A. thaliana* tissues was performed according to a previously published protocol [30] with some modifications, as described below. Frozen tissue (300 mg) was ground to a fine powder in a mortar placed in a liquid nitrogen bath in the presence of a frozen droplet (30 µL) of a protease inhibitor mixture (one Complete tablet (Roche Molecular Biochemicals, Mannheim, Germany)) dissolved in 2 mL 100 mM KCl (Merck, Darmstadt, Germany), 20% v/v glycerol (Merck) and 50 mM Tris pH 7.1 (Sigma, Steinheim, Germany) and 15 µL of a second protease inhibitor mixture consisting of 1 mM pepstatin A (Sigma) and 1.4 µM PMSF (Sigma) dissolved in ethanol. The homogenate was centrifuged for 30 min at 226 000 × *g* at 4°C. The resulting supernatant constituted fraction 1 for the tissue, while the resulting pellet was submitted to further protein extraction. From this supernatant a 50 µL aliquot was taken out and mixed with 54 mg urea (Bio-Rad, Hercules, CA, USA), 5 µL 700 mM DTT (Sigma) and 5 µL of the carrier ampholyte mixture Servalyte 2–4 (Serva, Heidelberg, Germany). Unless analysed immediately by 2-DE, the samples were stored at –80°C. To the pellet from the first ultracentrifugation, an aliquot of 12.5% w/w protease inhibitor mixture (complete) and 100% w/w 100 mM phosphate buffer, pH 7.1 containing 0.2 M KCl, 20% glycerol, 2 mM MgSO₄ and 2% w/w amido sulfobetaine 14, were added. The frozen components were mixed with 2.5% w/w benzonase (Merck) and stirred at 4°C for 45 min. An aliquot of 23% w/w of 700 mM DTT solution, 50% w/w of urea and 30% w/w of thiourea were added to the homogenate under continued stirring and left for incubation for an additional 45 min at room temperature. When the urea had been dissolved completely, the sample was centrifuged for 30 min at 30 000 × *g* at 17°C. The supernatant obtained was considered as fraction 2. After addition of 5% w/w carrier ampholyte pH 2–4, fraction 2 was used for 2-DE analysis or stored at –80°C. The remaining pellet was then incubated for 15 min in SDS sample buffer [35] and finally centrifuged at 10 000 × *g* for 15 min. The resulting supernatant was considered fraction 3 and was run immediately on SDS-PAGE gels [35] or stored at –80°C until used.

2.3 2-DE

IEF was carried out, according to a previously described protocol [29], in 20 cm tube gels (diameter 1.5 mm) containing 3.5% w/v acrylamide (Serva), 0.3% w/v bisacrylamide (Serva) and 2% v/v of an ampholyte mixture consisting of one part of ampholine pH 3.5–10.0 (Pharmacia, Uppsala, Sweden), one part of Servalyt pH 2.0–11.0 (Serva, Heidelberg, Germany), three parts of pharmalyte pH 4.0–6.5 (Pharmacia), two parts of pharmalyte pH 5.0–8.0 and one part of pharmalyte pH 6.5–9.0. The anode buffer contained 90 g urea and 25 mL of 85% v/v phosphoric acid (Merck) in a final volume of 500 mL water. The cathode buffer contained 20 mL ethanoldiamine (Merck), 216 g of urea and 20 g glycerol in a final volume of

400 mL water. Millipore water (Millipore, Bedford, MA, USA) with a minimum electric resistance of 18.2 M Ω was used for all solutions. Typically, 50 μ L protein sample, containing 600 μ g of protein, was loaded on the preparative IEF gel (Coomassie G-250 staining). Analytical IEF gels were loaded with 10 μ L, 120 μ g total protein (silver nitrate staining). The proteins were focused for 1 h at 100 V, 1 h at 200 V, 17.5 h at 400 V, 1 h at 650 V min, 30 min at 1000 V, 10 min at 1500 V, and finally 5 min at 2000 V. After IEF the gels were ejected from the glass tube and, either used immediately for SDS-PAGE, or stored at -80°C . SDS-PAGE [35] was performed in $23.2 \times 30 \times 0.1$ cm 15% w/v acrylamide gels containing 0.2% w/v bisacrylamide. The electrophoresis conditions were 30 mA/gel for 15 min followed by 50 mA/gel for 5–6 h. After completing the protein separation, the gels were fixed or stored in the adequate staining buffer.

2.4 Protein staining

Coomassie G-250 staining was performed according to the protocol by Doherty *et al.* [36]. Protein fixation was performed by incubating the gels in a 50% v/v methanol (J.T. Baker, Deventer, Netherlands), 2% v/v phosphoric acid solution for a minimum of 2 h. The gels were then rinsed for 20 min in water after which they were equilibrated for 1 h in a solution containing 30% v/v methanol, 2% v/v phosphoric acid and 17% w/v ammonium sulfate (Merck). A portion of 660 mg crystalline Coomassie G-250 (Carl Roth, Karlsruhe, Germany) *per* litre staining solution was added. The staining proceeded for up to 5 d under constant shaking. The staining process was stopped by replacing the staining solution with water. Silver-staining was performed according to Heukeshofen *et al.* [37]. The gels were incubated for 2 h in a solution of 41 g/L sodium acetate, 1.25 g/L sodium thiosulfate, 30% ethanol and 20 mL/L glutaraldehyde, then washed twice for 10 min in water, and incubated in 0.1% AgNO₃, 0.01% formaldehyde for a minimum of 30 min. After this step the gels were washed once with water for 20 s and once with a solution of 2.5% Na₂CO₃ for 1 min. The stain was developed in a 2.5% Na₂CO₃, 0.02% formaldehyde solution, within a time period of 3 to 15 min. The development of the staining was stopped by placing the gel in 1 L of a solution of 18.5 g/L Titriplex (Merck).

2.5 Image analysis

The 2-DE gels were scanned with an UMAX Mirage II A3 scanner (Willich, Germany) and then analysed with Z3 imaging software, version 3.0 (Compugen, Tel Aviv, Israel). Spot detection was performed with the parameters minimum spot area and a minimum spot contrast set to 1. The parameter confidence level, which is a measure of the probability that a detected feature on the gel image truly is a protein spot, was set to 0.99. For a more detailed description of the program we refer to the internet address of the software manufacturer (<http://www.2dgels.com>) and to a recent publication by Smilansky *et al.* [38]. The protein *pI* scale on the

gels was calibrated by a linear correlation, using the *pI* values calculated for the amino acid sequences of all identified proteins. The *M_r* scale was calibrated using *M_r* standards run alongside the second dimension.

2.6 Spot excision and tryptic in-gel digestion

Gel sample handling and protein digestion were performed as previously described [31, 39] with some minor modifications. Protein spots were sampled from the 2-DE gel using an automatic spot excision workstation (Proteineer; Bruker Daltonics, Bremen, Germany). The excision head was equipped with a single circular excision head with a diameter of 2 mm. The excised gel spots were delivered into 96-well polypropylene microtiter plates (Costar Thermowell, Cornis, NY, USA). The microtiter plates were pretreated by punching two 0.5 mm holes in the bottom of every well, using an in-house modified robot (LND-1; Luigs und Neumann, Ratingen, Germany). This preparation allowed liquid to be removed by flow-through centrifugation [31, 39]. The protocol for tryptic in-gel digestion was based on the protocol of Shevchenko *et al.* [40] with some modifications. Briefly, the excised gel pieces were washed three times for 20 min in 100 μ L washing buffer (50 mM NH₄CO₃, pH 8, 50% ethanol v/v). The gel pieces were then dehydrated for 5 min by addition of 100 μ L ethanol, followed by a 30 min incubation step at 37°C in 10 μ L 10 mM DTT. The DTT solution was removed and the gel pieces were dehydrated. As a next step the reduced cysteine residues were carbamidomethylated by incubation of the gel pieces in 50 mM iodoacetamide in the dark for 30 min at room temperature. After removal of the iodoacetamide solution, a 10 min washing step in 100 μ L of the washing buffer was performed, after which the gel pieces were dehydrated for 10 min in 100 μ L of ethanol. To completely dehydrate the gel pieces they were centrifuged for 15 min in a Speed Vac concentrator (SVC100H; Savant, Holbrook, NY, USA). The digest was then performed with a trypsin solution (Roche) containing 15 ng trypsin/ μ L in 50 mM NH₄CO₃, pH 8.4 (60 ng trypsin) of this solution, stored on ice, and left for incubation for 30 min at 4°C. Following the incubation, 8 μ L 50 mM NH₄CO₃, pH 8 was added and the digestion was performed at 37°C for 4 h. Finally the digestion was stopped by adding 10 μ L of extraction buffer (0.3% TFA, 4 mM *n*-octylglycopyranoside). The resulting supernatant was then centrifuged in a fresh microtiter plate and 1 μ L of it was used for the subsequent MALDI-TOF MS analysis.

2.7 MALDI-TOF MS

The sample preparation on MALDI AnchorChip targets (Bruker Daltonics) was performed as previously described [31, 39]. The acquisition of the MALDI spectra was performed on a Bruker Autoflex (Bruker Daltonics) MALDI-TOF mass spectrometer operated in reflector mode and using delayed ion extraction. Positively charged ions in the *m/z* range 700–3500 were analysed in automatic mode.

2.8 Data analysis

The spectra were calibrated using a recently described procedure relying on external calibration followed by internal mass correction [33]. Assignment of the first monoisotopic signals in the spectra was performed automatically using the signal detection algorithm SNAP (Bruker Daltonics). Protein identification was performed using the ProFound database search engine [41] with the following search settings: 50 parts *per* million mass error tolerance, one permitted missed cleavage site and carbamidomethylation on cysteine residues as complete- and methionine oxidations as variable modifications. The *A. thaliana* MIPS database (<http://mips.gsf.de>) was employed for protein identification. The certainty of the identifications results is characterised by a probability score, *i.e.*, the probability that the retrieved protein is a random match. Proteins are reported identified if their probability score is below 0.02, corresponding, for the MIPS database, to a 95% confidence.

3 Results and discussion

3.1 Construction of the *A. thaliana* proteomic reference map

A previously described protein extraction protocol, optimised for plant tissues [30], was employed for the preparation of protein extracts from eight different *A. thaliana* tissues: primary leaf, leaf, stem, silique, seedling, seed, root, and inflorescence. From each tissue, three fractions were produced, fractions 1, 2 and 3, by the successive use of different protein solubilising agents. Fractions 1 and 2 were analysed by 2-DE, while fraction 3, not suitable for IEF due to a high concentration of the anionic detergent SDS, was analysed by 1-D SDS-PAGE [35]. The separated proteins were visualised by silver nitrate staining [37], which allowed sensitive and reproducible protein detection. The resulting 2-DE images, which were reproduced from a minimum of three independent experiments, constituted the reference gel images in this study. The 1-D PAGE and the 2-DE protein patterns of the corresponding tissues are shown in a compressed format in Fig. 1, while downloadable full-size images of the tissues can be viewed on the internet pages of the GABI Primary Database of the Resource Centre in Berlin at the following URL: https://gabi.rzpd.de/projects/Arabidopsis_Proteomics.

3.2 Evaluation of the 2-DE protein spot patterns of different tissues

Automatic spot detection (Z3; Compugen) resulted in a total of 39 000 protein spots in the 2-DE images of the analysed plant tissues. On the basis of this large number of detected protein spots from the different 2-DE gels, spot comparisons of the different tissues and the different fractions were performed. As can be seen in Fig. 1, the 2-DE spot patterns of

fraction 1 for all tissues contain protein spots predominantly in the acidic and neutral region of the gels, whereas the spot patterns of fraction 2 show a large number of additional spots in the basic part of the gel. According to automatic, pairwise, spot pattern comparisons, implemented in the image analysis software, the overlap between the spot patterns of the different fractions for each tissue ranged between 20 and 30%, demonstrating the complementary character of the differentially prepared protein extracts.

Comparing the 2-DE gels from the same fractions but different tissues showed that some tissues and fractions had a higher degree of similarity, while others were more divergent, indicating closer or more distant functional and compositional relations between the contained tissue proteomes. To illustrate these similarities/differences, we selected a representative region from the analysed 2-DE gels and magnified them (Fig. 2). As can be seen from this figure a large variability, not only in the spot density (number of displayed protein spots *per* region), ranging from a minimum of 30 spots for the seed fraction 1 gel to a maximum of 492 spots for the stem fraction 1 gel, but also in their general spot pattern, was observed. According to their similarities we grouped the eight tissues into two major classes, which lined up with two physiological characteristics of the different tissues. Specifically, the nongreen, chlorophyll-free, tissues (seed and root) and the green, photosynthetically active tissues (seedling, stem, primary leaf, leaf, inflorescence and silique). While the green tissues showed a higher extent of homogeneity within their group, the nongreen tissues resulted in 2-DE spot patterns that were highly divergent to each other (Fig. 2).

Comparing the 2-DE spot patterns of the green tissues in more detail, revealed varying degrees of similarity between the different tissues. The highest degree of similarity was found between leaf and primary leaf tissue. The 2-DE images of fraction 1 from these tissues share 220 protein spots, which corresponds to 53% of the spots detected in this section of primary leaf and 41% of the spots detected in the same section of the leaf 2-DE gel, respectively. An even higher spot pattern similarity was observed between the 2-DE images of fraction 2 of these tissues. Here 160 protein spots showed overlap between the two tissues, corresponding to 61% of the protein spots detected in the primary leaf gel and 60% of the leaf 2-DE gel section, respectively. The other green tissues (inflorescence, silique and stem) showed a significantly lower degree of similarity, which varied between 20 and 45% overlap. Nevertheless, the observation from leaf and primary leaf tissue, that the 2-DE images of fraction 2 show higher similarities than their corresponding images of fraction 1, holds also true for these tissues. Since most of the overlapping protein spots are identical to all of the 2-DE gels from these green tissues, while most of them are missing from the nongreen tissues, it is reasonable to assume that most of these proteins are associated with the chloroplasts. To what extent the chloroplasts stay intact after fraction 1 extraction is still not clear, but it became evident,

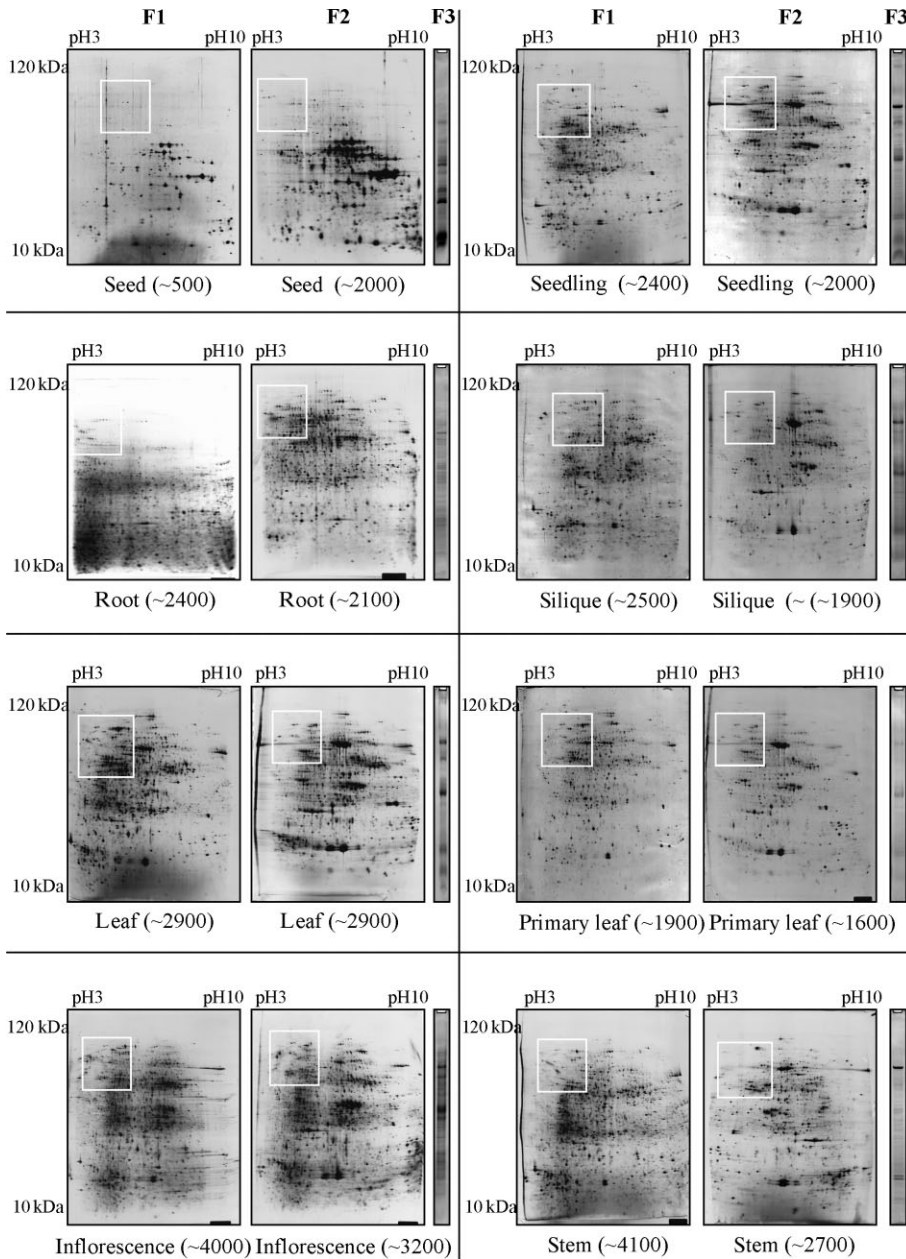


Figure 1. Overview of 2-DE and 1-D SDS PAGE gels analysed from the eight different *A. thaliana* tissues. F1, F2 and F3 refer to protein fractions 1, 2 and 3, respectively. The numbers under the gel indicate the number of detected protein spots within each 2-DE gel. Boxed regions correspond to the magnified regions displayed in Fig. 2.

that the major part of the large subunit of the chloroplast protein, ribulose 1, 5 biphosphate carboxylase/oxygenase (LSU), could be detected in fraction 2 gels. This very abundant protein was therefore also mainly responsible for the clearly increased similarity of the 2-DE images of fraction 2 compared to their counterparts from fraction 1.

The low resolution of the SDS-PAGE images resulting from fraction 3 protein extracts (Fig. 1), prevents detailed image comparisons. Many of the observed bands comigrate and superimpose each other, making a reliable distinction of individual proteins therefore difficult. Nevertheless, much more important than a detailed image analysis of these 1-D

gels is, that these SDS-PAGE gels allow the separation and MS-based protein identification of highly hydrophobic proteins, which otherwise might escape the analysis, if only 2-DE gel based separation would have been employed.

3.3 Protein identification

As a starting point for the MS-based protein analysis, we selected three tissue types; primary leaf, silique, and seedling. Instead of using silver nitrate protein staining [40], which was used for the reference 2-DE gels, a colloidal Coomassie G-250 protein stain was employed, which is our pre-

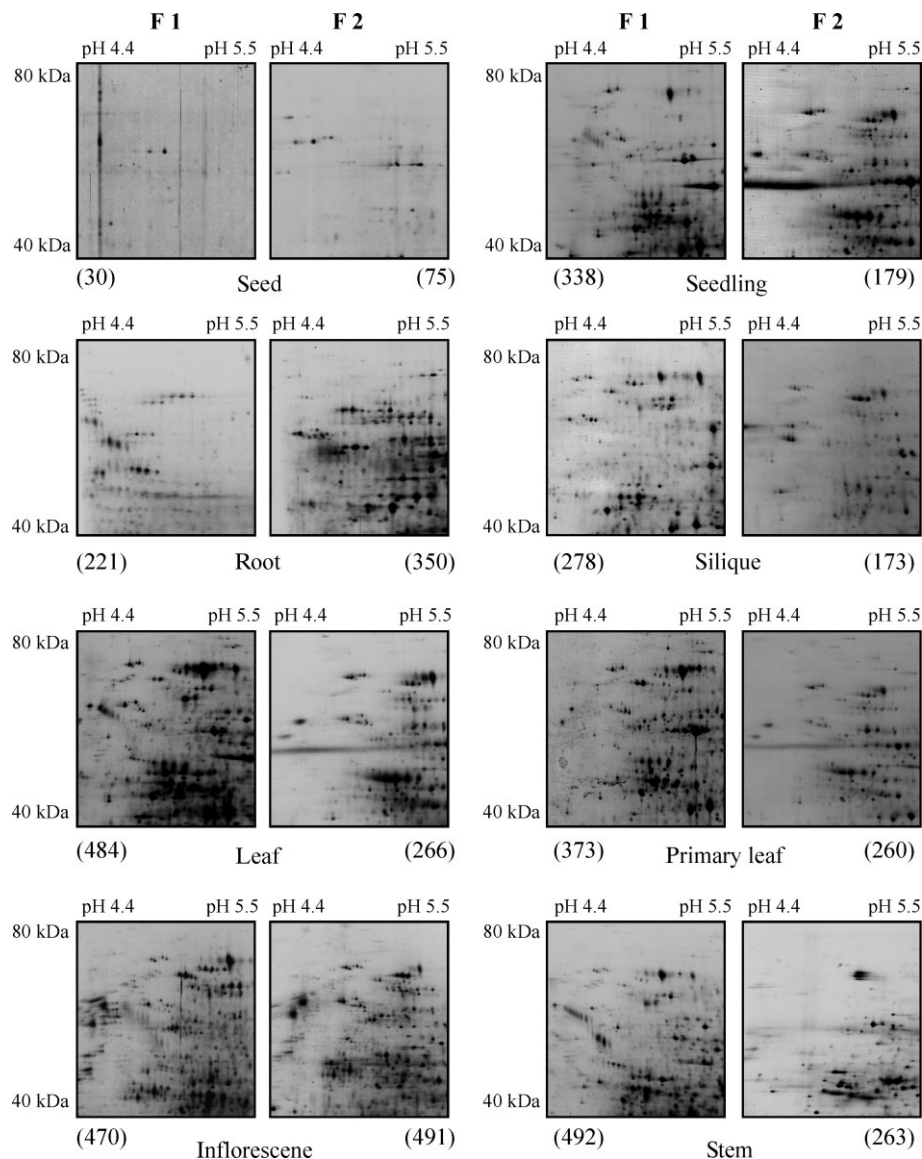


Figure 2. Comparison of a 2-DE gel region, indicating the numeric and spatial diversity between the eight different analysed 2-DE gels. The numbers under each gel image indicate the number of proteins detected within each section.

ferred choice for automated mass spectrometric analysis of proteins separated by 2-DE [36]. Supplementary Fig. 1, shows the 2-DE images of leaf fraction 1. As can be seen, the detection sensitivity of the employed Coomassie G-250 staining is close to that of the silver nitrate staining. Starting from six different 2-DE gels (two fractions *per* tissue) 6000 protein spots were excised and analysed by MALDI-TOF MS PMF, resulting in a total of 2943 identifications. The spots of the identified proteins from the Coomassie 2-DE gels were subsequently matched to their silver nitrate stained counterparts, revealing the annotated 2-DE reference maps. An interactive version of these annotated reference gels and additional tables containing all identified proteins can be viewed and downloaded at the GABI primary database internet page under: https://gabi.rzpd.de/projects/Arabidopsis_Proteomics.

3.4 Mass and *pI* distribution of the identified and predicted proteins

To evaluate potential bias in the analytical techniques used, the distribution of the identified proteins' mass and *pI*, as determined by their migration on the 2-DE gels, were compared to the calculated mass and *pI* distributions of all predicted proteins in the *A. thaliana* protein database [42]. A clear correlation between the relative occurrence of identified proteins from the 2-DE gel and the calculated occurrence from the *A. thaliana* protein database can be seen for proteins in the mass ranges 20–50 kDa and 70–90 kDa. (Fig. 3a). For proteins of lower and higher masses, the relative occurrence of predicted proteins exceeds the occurrence of the identified proteins, while within the mass range 50–70 kDa a higher percentage of identified *versus* predicted proteins was observed.

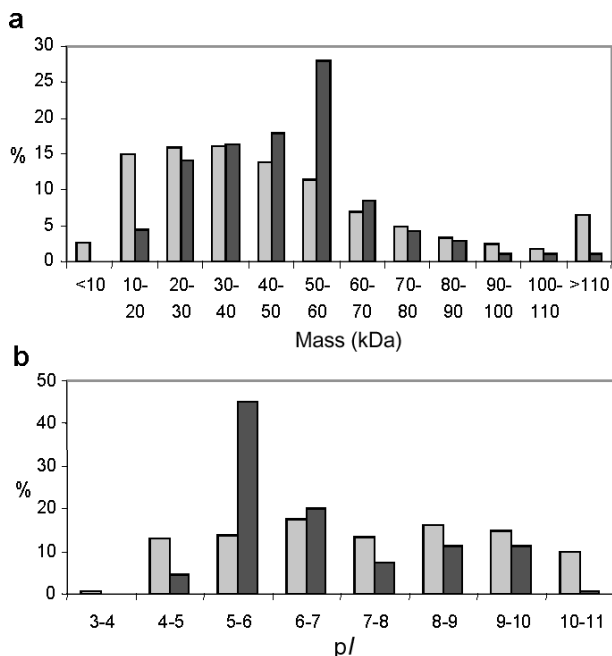


Figure 3. (a) Diagram of the comparison of the relative mass distribution of the predicted proteins from the MIPS *A. thaliana* protein database (MatDB) [28] to the relative distribution of the identified proteins from the selected 2-DE gels. The grey bars indicate the values for the predicted proteins, while the black bars indicate the values for the identified proteins. (b) Same as 3a, but concerning the relative distribution of pI values.

For example, the mass of 17.7% of the proteins contained in the *A. thaliana* protein database falls into the 5–20 kDa range, while only 4.5% of the identified proteins fall within this mass range. The most likely explanation for this underrepresentation of small proteins in our data set is, that although a relative large number of small mass proteins are present on the 2-DE gels, few of these proteins were readily identified. Namely, these small mass proteins typically yield only a few peptides upon proteolytic digestion which makes an unambiguous identification by PMF less likely [43]. Protein identification techniques based on peptide fragment ion mass spectra aid in the identification of low mass proteins [44]. Ten point nine percent of the proteins from the Arabidopsis protein database were expected to have masses above 90 kDa, while only 3.4% of the proteins identified from the 2-DE gels fell into this category. A significant difference which was expected for the employed 2-DE gels. Since we used 15% polyacrylamide gels, the optimal resolving capacities of these gels covers the mass range 10–120 kDa. To get access to larger proteins, lower percentage polyacrylamide, or even better, agarose gels [45] should be employed.

In contrast to these underrepresented proteins, proteins with a mass of 50–70 kDa were more frequently identified than expected. Here the experimentally determined number (36.6%) was twice as high as that expected (18.3%), indicating that proteins in this mass range are best suited for iden-

tification by 2-DE and MALDI-TOF MS PMF. A similar comparison was made for the distribution of the pIs of all predicted versus all identified proteins. The result of this comparison is displayed in Fig. 3b. As can be seen from the figure, identified proteins in the pI ranges below pH 5 and above pH 7 are underrepresented, whereas the span between these values is clearly overrepresented. Nearly 65.1% of the identified proteins fall into this range, whereas only 31.4% were expected therein. Both observations concerning the mass and the pI distribution indicate that resolving all expressed protein species from an organism like *A. thaliana* requires more than one analytical protein separation and identification method.

3.5 Detection of protein isoforms

The migration of proteins on a 2-DE gel is sensitive to small structural differences. Therefore, for instance, PTM [46, 47] and alternatively spliced [48, 49] forms of a protein often appear as additional spots on the 2-DE image. An estimation of the overall degree of protein modification in a tissue can thus be made by calculating the average number of spots to which the same protein sequence was assigned. This calculation was performed for the 2-DE gels of each of the three tissues, for the combination of both gels of each tissue (fraction 1 and fraction 2), and for the proteins identified from all tissues. Figure 4 summarises the results of these comparisons. A redundancy of approximately two spots per identified protein sequence was found at the individual 2-DE gel level for all analysed fraction 1 and fraction 2 gels, whereas the combined redundancy of the fraction 1 and fraction 2 gels yielded 2.5 protein spots per protein. This increase indicates an overlap of ~20% between the different fractions of one tissue. Interestingly, this value matches the observed overlap between the spot patterns of the fraction 1 and fraction 2 gels from the different tissues, calculated by comparative image analysis (see Section 3.4).

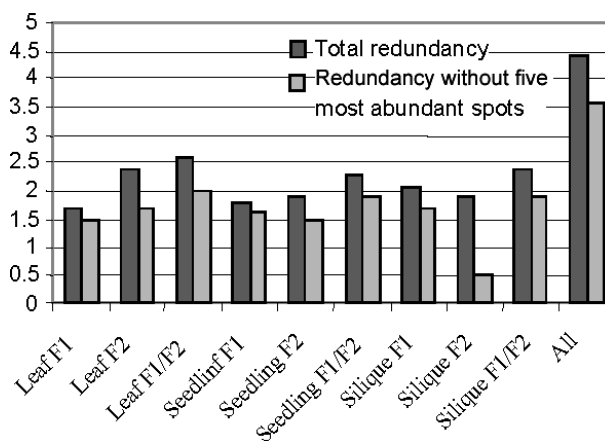


Figure 4. Diagram indicating the number of protein spots yielding the same identified proteins, in different tissues.

Comparing the protein identities from different tissues on their protein fraction level revealed that the 2-DE spot patterns from primary leaf and seedling fraction 1 were, with 212 identical protein identities (33%), more similar to each other, than they were to the protein identities from the silique fraction 1 gel (Table 1). Only 107 proteins (22%) from primary leaf fraction 1 found identical assignments to the identified proteins from silique fraction 1, while seedling fraction 1 showed 124 overlapping protein identities (23%) to silique fraction 1.

The protein identification results from fraction 2 clearly supports the result already mentioned in the image analysis section (Section 3.2), namely that the protein compositions of fraction 2 were more similar to each other (31–35% overlap) than their fraction 1 counterparts (Table 1). This observation can not be solely based, as mentioned in the image analysis section (Section 3.2), on the presence of the LSU in these gels, since only the total number of different identified proteins was compared in this analysis. Therefore it seems obvious that a further number of different proteins, possibly associated with chloroplasts or other intracellular organelles, might be responsible for this increased degree of similarity on the protein identity level for the fraction 2 gels.

Another interesting finding, concerning the expression pattern of different protein isoforms revealed that many of these proteins, which were expressed from different genes, show high tissue specificity. For instance quinone oxidoreductase, which was found to be expressed from gene At5g16990 in fraction 1 of silique tissue, was expressed in the primary leaf tissue by gene At5g17000. Another isoform of this protein was found in primary leaf and seedling fraction 1, expressed from gene At5g16970. In described cases no protein spot could be detected in the other tissues that matches to the tissue-specific isoforms. Interestingly, there were several of these examples within the sum of the identified proteins, confirming previously reported results, which employed HPLC- and 2-DE-based protein separation in combination with MS-based protein identification in the model plant rice [27].

The average redundancy for the entire data set of the identified protein spots was 4.4 protein spots *per* protein (Fig. 4), indicating the accumulating redundancy between the different plant tissues. Although, this value of 4.4 protein spots *per* protein might appear relatively high, it is significantly lower than a redundancy of 10 and more protein spots *per* assigned protein, reported for animal tissues [50–52]. The reasons for this observed discrepancy between plant and animal tissue are not clear. A detailed comparison of the relatedness of the different 2-DE gel spot identities is provided in Table 1.

3.6 Abundant and redundant proteins

The most abundant protein in plants [54], LSU [54, 55], was also the protein identified the most in each of the analysed 2-DE gels. This protein alone, represented by a total of 366 identified protein spots, makes up 12.5% of all assigned spots. The highest number of protein spots (123 spots) in a single 2-DE gel that were assigned to LSU, was obtained from fraction 2 of primary leaf. In many cases, their location on the gel was not close to the expected mass and *pI* (53 kDa and *pH* 5.88), but was distributed over a large area across the entire gel, ranging from approximately 5–70 kDa and from *pH* 5–9. The majority, 264 (72%), of the protein spots that were assigned to LSU were obtained from fraction 2, which is part of the explanation for the higher similarity of the fraction 2 protein spot patterns, compared to their counterparts from fraction 1. This is indeed a surprisingly large number. How many of these protein spots represent physiologically relevant protein species is not known. Also, without further analyses, it cannot be ruled out that some of the detected protein forms are artefacts introduced during sample preparation for 2-DE. The second most abundant protein in our 2-DE gels was the beta subunit of the ATPase multi-protein complex (*atpB*) [56]. This protein was identified in 84 spots, corresponding to 1.8% of all identified spots. Interestingly, *atpB* was not detected in fraction 1 from silique tissue,

Table 1. Overlap of identified proteins among different plant tissues

Overlap of protein identities	Primary leaf F1	Primary leaf F2	Seedling F1	Seedling F2	Silique F1	Silique F2
Primary leaf F1 (448)	286	133	212	99	107	112
Primary leaf F2 (669)	<i>23%</i>	283	146	141	82	140
Seedling F1 (621)	<i>33%</i>	<i>23%</i>	344	108	124	124
Seedling F2 (320)	<i>21%</i>	<i>31%</i>	<i>21%</i>	168	60	111
Silique F1 (383)	<i>22%</i>	<i>17%</i>	<i>23%</i>	<i>17%</i>	182	84
Silique F2 (466)	<i>25%</i>	<i>32%</i>	<i>25%</i>	<i>35%</i>	<i>26%</i>	150

The numbers in brackets after the tissue and the fraction indicate the total number of identified proteins spots from that 2-DE gel. Bold numbers in the diagonal of the table indicate the number of different protein sequences identified *per* 2-DE gel. The italic percentage indicates the overlap between the different 2-DE gels, while the other Arabic numbers indicate the absolute numbers of overlapping protein species from the different 2-DE gels.

while it was in 25 spots of the gels run with fraction 1 from primary leaf and seedling. The same observation was made for the alpha subunit of the ATPase multiprotein complex [57], which was not detected in fraction 1 from silique, whereas it was found in seven and four spots of the gels of leaf and the seedling, respectively.

A third example underlining the diversity of the protein composition of silique fraction 1, and also underlining the overall heterogeneity of all analysed tissues is illustrated in Fig. 5 by the spot pattern of the rubisco activase protein, which is the catalytic chaperone of the LSU complex [57]. In the gels of fraction 1 of primary leaf and seedling, rubisco activase was found in 12 and 15 spots, respectively, while it was not found in the fraction 1 gel from silique (Fig. 5). These results are in contrast to the result obtained by the comparisons of the rubisco activase pattern for the fraction 2 gels. Here a higher similarity between the silique and the seedling spot patterns was observed, namely for the two fraction 2 gels obtained from these tissues, rubisco activase was assigned to only five and four spots, respectively, whereas it was identified in 16 spots in the fraction 2 gel of the primary leaf (Fig. 5). The position of all spots assigned to rubisco activase show, similarly to the observed tissue specific expression of protein isoforms (see above), distinct, tissue specific patterns. Except for a few spots present in every 2-DE gel, each fraction from each tissue shows a specific rubisco activase pattern in the corresponding 2-DE gel images. This kind of observation was made for many other proteins that were abundant and found in sev-

eral spots, supporting the notion that many of the abundant proteins vary significantly in their modification status between the different tissues.

3.7 Membrane proteins

According to previous reports, based on bioinformatic protein sequence analysis, proteins with at least one transmembrane-spanning domain, constitute approximately 20% of all proteins in eukaryotic genomes [58, 59]. It is a well-known problem that these candidates for transmembrane proteins are often under-represented in 2-DE-based proteome analyses because of the tendency of the hydrophobic transmembrane regions to cause the proteins to precipitate, mainly during IEF [60]. To compare the relative frequency of potential transmembrane proteins, identified in our 2-DE gels, to the relative frequency of transmembrane proteins predicted from the Arabidopsis protein database, we made use of the web-based membrane protein prediction tool SOSUI (<http://sosui.proteome.bio.tuat.ac.jp/sosui/frame0.html>). According to SOSUI 27% of all *A. thaliana* proteins were likely candidates containing transmembrane domains, while only 8.1% of the identified proteins were predicted to contain at least one transmembrane domain. These comparisons render likely that our experimental approach, as previously predicted for other analyses of this type [60], selectively excludes hydrophobic transmembrane proteins from the analysis. This assumption is further supported by the fact that 97% of the identified transmembrane protein candidates contain only one or two transmembrane domains.

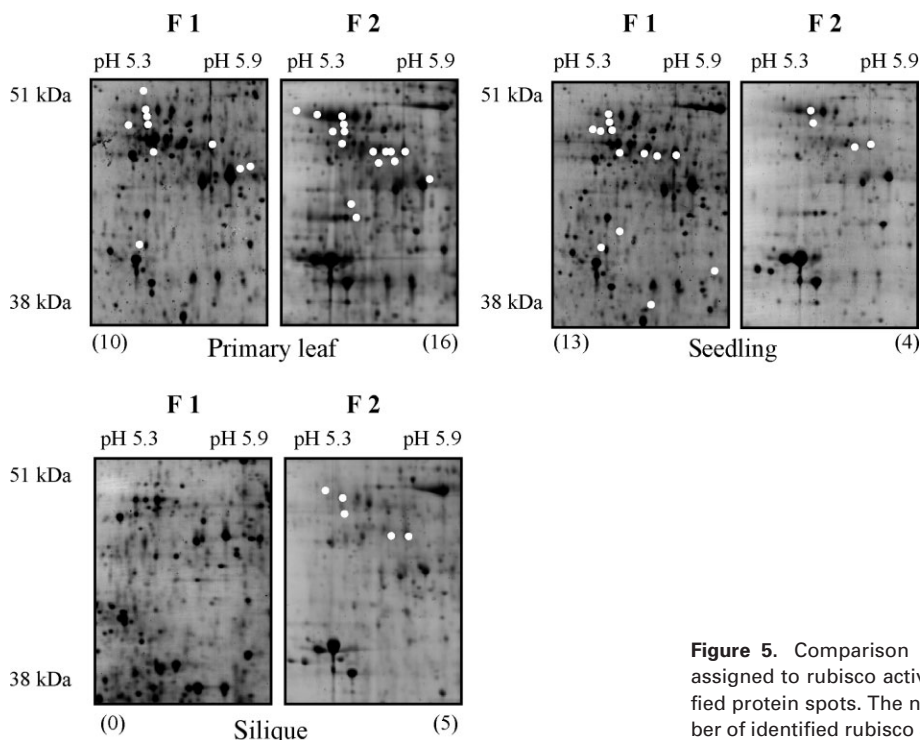


Figure 5. Comparison of different protein spots which were assigned to rubisco activase. The white dots indicate the identified protein spots. The numbers under the gel indicate the number of identified rubisco activase spots within each 2-DE gel.

The sequential protein extraction protocol used in this study utilises the successive application of protein-solubilising agents of increasing strength to fractionate the extracted proteins. Fraction 1, obtained by extraction using a solution containing only buffer and salt, was expected to contain highly soluble, cytosolic proteins, whereas fraction 2, which was obtained by extraction in chaotropes and detergents, was expected to additionally contain hydrophobic proteins. However, no significant difference was observed for the occurrences of membrane proteins in fraction 1 and fraction 2. Therefore although a high number of hydrophobic proteins were expected to be present in the fraction 2 sample, these are likely to be selectively lost during 2-DE. This loss might take place during the IEF step, due to protein precipitation, or during the transfer of the proteins from the first to the second dimension. To check if at least an increase of transmembrane domain proteins can be observed in fraction 3, we excised the six most intense bands of the 1-D SDS PAGE gel from primary leaf (Fig. 6) and analysed them by MALDI-TOF PMF. Three of the samples were found to be transmembrane proteins, containing six respectively three transmembrane spanning regions, indicating that the above-mentioned restrictions do not account for the SDS-PAGE separated proteins of fraction 3.

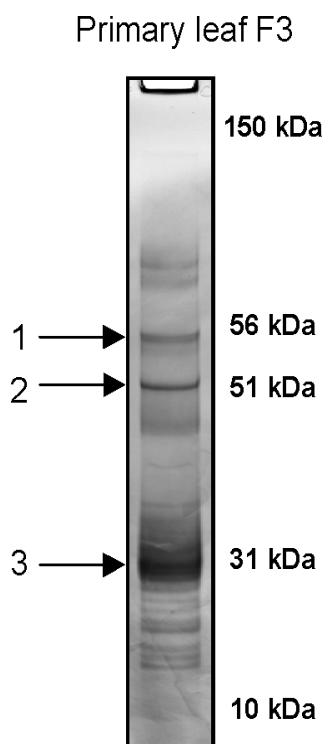


Figure 6. 1-D SDS PAGE gel from primary leaf fraction 3. The numbered arrows indicate bands of identified transmembrane spanning proteins from this gel. 1. photosystem II P680 chlorophyll a apoprotein (3 transmembrane domains); 2. photosystem II 44 kDa reaction center protein (6 transmembrane domains); 3. chlorophyll a/b-binding protein CP29 (3 transmembrane domains).

3.8 Functional classification of the identified proteins

The identified proteins were sorted into functional categories, using the MAPMAN program [48]. This classification, provides a general overview of the different types of protein detected in different tissues. Thirty-eight percent of the different proteins (238 proteins), corresponding to 52% of all excised protein spots (1541 protein spots), constitute enzymes or proteins with predicted enzymatic activity (Fig. 7a). This large group is quite heterogeneous and was further subdivided in more functionally oriented entities, displayed in Fig. 7b. A portion of the enzymes (15.9%) had functions in amino acid metabolism (41 proteins). Others (23 proteins, 8.9%) were involved in the citric acid cycle, 8.5% (22 proteins) were involved in the glycolysis pathway, and 8.1% (21 proteins) participate in cell wall synthesis and cell wall modification. Interestingly, only 3.1% (12 proteins) were assigned to secondary metabolism. This result clearly underlines that most of the proteins detected on our 2-DE gels are associated with central metabolic processes involved in the housekeeping of the cell.

The members of the second largest group, comprising 24% (157 proteins detected from 311 protein spots) of all identified proteins, are unknown or hypothetical proteins (Fig. 7a). This group consists of predicted open reading frames for which functions in some cases can be assigned based on sequence homology to proteins characterised in other organisms. Their tissue-specific detection by 2-DE and identification by MALDI-TOF MS PMF confirms the expression of these genes. Another category in Fig. 7a containing 8% (50 proteins detected from 112 protein spots) of the

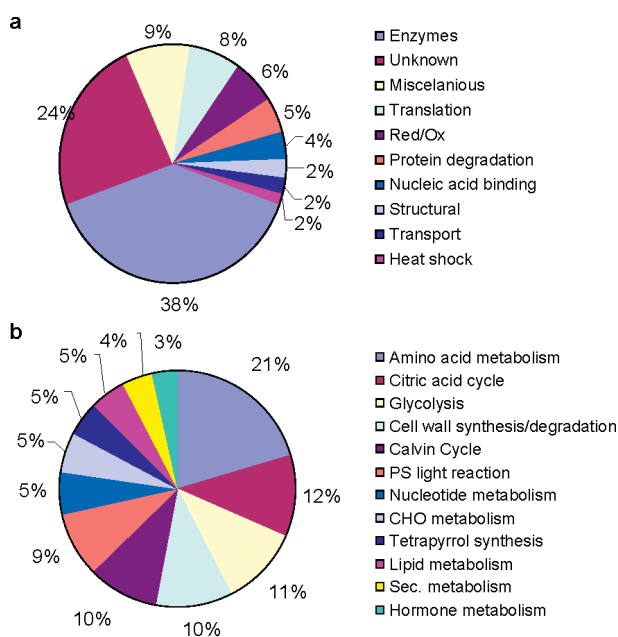


Figure 7. (a) Pie chart grouping the identified proteins from the 2-DE gels into 10 functional categories according to their occurrence. (b) Pie chart subgrouping the proteins from the enzymes group from (a) according to metabolic pathways they can be associated with.

identified proteins, is involved in translational processes. These proteins are either associated directly to the cytosolic 80S or the organellar 70S ribosome as core components [62] or indirectly as associated factors like initiation or elongation factors [63] of the translational machinery. Other proteins, like thioredoxins [64] or peroxyredoxins [65] are associated with the group of proteins regulating the redox status of the cell which constitutes 6% of the identified proteins (35 proteins from 164 protein spots). A further 2.1% of the identified proteins are involved in intercellular transport processes (13 proteins derived from 32 spots) [66] or are components of the cytoskeleton (15 proteins derived from 33 spots) [67, 68].

An interesting group, constituting approximately 5% of the identified proteins (28 proteins derived from 59 protein spots) is the group associated with the protein degradation machinery, mainly the 26S proteasome [69, 70]. This potentially large protein complex has been reported to employ up to 1300 proteins, representing ~5% of the total *A. thaliana* genome [71]. Approximately 2% of the identified proteins, namely 10 protein spots derived from the relatively large number of 72 protein spots, were classified as chaperones from the heat shock protein type [71, 72], which have their primary function in assisting or recovering correct protein folding. The high degree of different protein spots detected for this class of proteins might further indicate the post-translational regulation these proteins are submitted to [71, 72]. The remaining 9.0% of the proteins (83 protein spots from 390 protein spots) do not fit into the previous nine categories. These proteins have different cellular functions and were therefore collected in the category miscellaneous. The identified proteins and assigned functional classification are listed in supplementary Table 1, which can also be accessed online at https://gabi.rzpd.de/projects/Arabidopsis_Proteomics.

4 Concluding remarks

In the presented proteomic study, proteins in eight different *A. thaliana* tissues were analysed by 2-DE. We found spot patterns with less than 10% similarity to any of the other 2-DE spot patterns (seed and root), whereas the patterns of tissues like seedling and primary leaf showed up to 60% overlap, indicating the functional related- or unrelatedness of the different plant tissues. We selected three from the eight tissues, *i.e.* primary leaf, silique and seedling, for mass spectrometric analysis and identified 2943 out of 6000 excised proteins, representing 663 different genes. These proteins were further grouped according to their metabolic or cellular functions and as a result we found that the majority were so called housekeeping proteins, which are associated to basic functions like primary metabolism, the translational machinery [63], the protein degradation machinery [73] or regulating the redox status of the cell. Interestingly, a large number of the identified proteins from the different tissues

were expressed as tissue-specific isoforms, derived from different genes. These proteins, which have similar biochemical functions and high sequence similarities, were selectively expressed in the different tissues. The same observation of high tissue specificity could be made for protein spot families that were assigned to the same gene. Many of these different protein spots are believed to represent different PTMs or degradation events [46, 47] and/or alternative splicing processes [48, 49], which to a large degree contribute to the heterogeneity and complexity of the analysed tissue specific proteomes.

We would like to thank Marion Herrmann, Beata Lukaszewka, Yvonne Kläre and Janine Stuwe for their excellent technical assistance. We would also like to thank Dr. Svenja Meyer and Axel Nagel from the Resource Centre Berlin for providing the proteomics Web page. Furthermore we would like to thank Dr. Oliver Thimm from the Max Planck Institute for Molecular Plant Physiology in Golm, for sharing data from the MAPMAN program. This project was funded by the German plant genome project: Genome Analysis in the Biological System Plant (GABI) support by the German Ministry of Education and Research (BMBF/GABI Project 32P2244) and the Max Planck Society.

5 References

- [1] Arabidopsis Genome Initiative, *Nature* 2000, 408, 796–815.
- [2] Yu, J., Hu, S. N., Wang, J., Wong, G. K. S. *et al.*, *Science* 2002, 296, 79–92.
- [3] Goff, S. A., Ricke, D., Lan, T. H., Presting, G. *et al.*, *Science* 2002, 296, 92–100.
- [4] Schena, M., Shalon, D., Davis, R. W., Brown, P. O., *Science* 1995, 270, 467–470.
- [5] Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T. *et al.*, *Nat. Biotechnol.* 1996, 14, 1675–1680.
- [6] Lashkari, D. A., DeRisi, J. L., McCusker, J. H., Namath, A. F. *et al.*, *Proc. Natl. Acad. Sci. USA* 1997, 94, 13057–13062.
- [7] Anderson, L., Seilhamer, J., *Electrophoresis* 1997, 18, 533–577.
- [8] Haynes, P. A., Gygi, S. P., Figeys, D., Aebersold, R., *Electrophoresis* 1998, 19, 1862–1871.
- [9] Gygi, S. P., Rochon, Y., Franza, B. R., Aebersold, R., *Mol. Cell. Biol.* 1999, 19, 1720–1730.
- [10] Klose, J., *Humangenetik* 1975, 26, 231–243.
- [11] O'Farrell, P. H., *J. Biol. Chem.* 1975, 250, 4007–4021.
- [12] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. *et al.*, *Nat. Biotechnol.* 1999, 17, 676–682.
- [13] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, *Nat. Biotechnol.* 1999, 17, 994–999.
- [14] Gallardo, K., Job, C., Groot, S. P. C., Puype, M. *et al.*, *Plant Physiol.* 2001, 126, 835–848.
- [15] Wilson, K. A., McManus, M. T., Gordon, M. E., Jordan, T. W., *Proteomics* 2002, 2, 1114–1122.

- [16] Herranen, M., Battchikova, N., Zhang, P., Graf, A. *et al.*, *Plant Physiol.* 2004, *134*, 470–481.
- [17] Friso, G., Giacomelli, L., Ytterberg, A. J., Peltier, J. B. *et al.*, *Plant Cell* 2004, *16*, 478–499.
- [18] Werhahn, W., Braun, H. P., *Electrophoresis* 2002, *23*, 640–646.
- [19] Krufft, V., Eubel, H., Jansch, L., Werhahn, W., Braun, H. P., *Plant Physiol.* 2001, *127*, 1694–1710.
- [20] Calikowski, T. T., Meulia, T., Meier, I., *J. Cell. Biochem.* 2003, *90*, 361–378.
- [21] Bae, M. S., Cho, E. J., Choi, E. Y., Park, O. K., *Plant J.* 2003, *36*, 652–663.
- [22] Chivasa, S., Ndimba, B. K., Simon, W. J., Robertson, D. *et al.*, *Electrophoresis* 2002, *23*, 1754–1765.
- [23] Ndimba, B. K., Chivasa, S., Hamilton, J. M., Simon, W. J., Slabas, A. R., *Proteomics* 2003, *3*, 1047–1059.
- [24] Walz, C., Giavalisco, P., Schad, M., Juenger, M. *et al.*, *Phytochemistry* 2004, *65*, 1795–1804.
- [25] Santoni, V., Dumas, P., Rouquie, D., Mansion, M. *et al.*, *Biochimie* 1999, *81*, 655–661.
- [26] Santoni, V., Vinh, J., Pflieger, D., Sommerer, N., Maurel, C., *Biochem. J.* 2003, *373*, 289–296.
- [27] Koller, A., Washburn, M. P., Lange, B. M., Andon, N. L. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 11969–11974.
- [28] Klose, J., Kobalz, U., *Electrophoresis* 1995, *16*, 1034–1059.
- [29] Klose, J., in: Link, A. J. (Ed.), *2-D Proteome Analysis Protocols*, Humana Press, Totowa, New Jersey, USA 1999, pp. 147–172.
- [30] Giavalisco, P., Nordhoff, E., Lehrach, H., Gobom, J., Klose, J., *Electrophoresis* 2003, *24*, 207–216.
- [31] Nordhoff, E., Egelhofer, V., Giavalisco, P., Eickhoff, H. *et al.*, *Electrophoresis* 2001, *22*, 2844–2855.
- [32] Egelhofer, V., Gobom, J., Seitz, H., Giavalisco, P. *et al.*, *Anal. Chem.* 2002, *74*, 1760–1771.
- [33] Gobom, J., Mueller, M., Egelhofer, V., Theiss, D. *et al.*, *Anal. Chem.* 2002, *74*, 3915–3923.
- [34] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C. *et al.*, *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5011–5015.
- [35] Laemmli, U. K., *Nature* 1970, *227*, 680–655.
- [36] Doherty, N. S., Littman, B. H., Reilly, K., Swindell, A. C. *et al.*, *Electrophoresis* 1998, *19*, 355–363.
- [37] Heukeshoven, J., Dernick, R., *Electrophoresis* 1988, *9*, 28–32.
- [38] Smilansky, Z., *Electrophoresis* 2001, *22*, 1616–1626.
- [39] Gobom, J., Schuerenberg, M., Mueller, M., Theiss, D. *et al.*, *Anal. Chem.* 2001, *73*, 434–438.
- [40] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.
- [41] Zhang, W. Z., Chait, B. T., *Anal. Chem.* 2000, *72*, 2482–2489.
- [42] Schoof, H., Zaccaria, P., Gundlach, H., Lemcke, K. *et al.*, *Nucleic Acids Res.* 2002, *30*, 91–93.
- [43] Jensen, O. N., Podtelejnikov, A., Mann, M., *Rapid Commun. Mass Spectrom.* 1996, *10*, 1371–1378.
- [44] Rappsilber, J., Moniatte, M., Nielsen, M., Podtelejnikov, A., Mann, M., *Int. J. Mass Spectrom.* 2003, *226*, 223–237.
- [45] Oh-Ishi, M., Satoh, M., Maeda, T., *Electrophoresis* 2000, *21*, 1653–1669.
- [46] Krishna, R. G., Wold, F., *Adv. Enzymol. Relat. Mol. Biol.* 1993, *67*, 265–298.
- [47] Gooley, A. A., Packer, N. H., in: Wilkins, M. R., Williams, K., Appel, R. D., Hochstrasser, D. F. (Eds.), *Proteome Research: New Frontiers in Functional Genomics.*, Springer, Berlin, Germany 1997, pp. 65–92.
- [48] Smith, C. W. J., Patton, J. G., Nadal-Ginard, B., *Annu. Rev. Genet.* 1989, *23*, 527–577.
- [49] Brett, D., Pospisil, H., Valcárcel, J., Reich, J., Bork, P., *Nat. Genet.* 2002, *30*, 29–30.
- [50] Fountoulakis, M., Berndt, P., Langen, H., Suter, L., *Electrophoresis* 2002, *23*, 311–328.
- [51] Fountoulakis, M., Juranville, J.-F., Dierssen, M., Lubec, G., *Proteomics* 2002, *2*, 1547–1576.
- [52] Fountoulakis, M., Suter, L., *J. Chromatogr. B* 2002, *782*, 197–218.
- [53] Kellogg, E. A., Juliano, N. D., *Am. J. Bot.* 1997, *84*, 413–428.
- [54] Spreitzer, R. J., *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 1993, *44*, 411–434.
- [55] Spreitzer, R. J., Salvucci, M. E., *Annu. Rev. Plant Biol.* 2002, *53*, 449–475.
- [56] Junge, W., *Proc. Natl. Acad. Sci. USA* 1999, *96*, 4735–4737.
- [57] Portis, A. R., *Photosynth. Res.* 2003, *75*, 11–27.
- [58] Wallin, E., von Heijne, G., *Protein Sci.* 1998, *7*, 1029–1038.
- [59] Stevens, T. J., Arkin, I. T., *Proteins* 2000, *39*, 417–420.
- [60] Santoni, V., Molloy, M., Rabilloud, T., *Electrophoresis* 2000, *21*, 1054–1070.
- [61] Thimm, O., Blasing, O., Gibon, Y., Nagel, A. *et al.*, *Plant J.* 2004, *37*, 914–939.
- [62] Barakat, A., Szick-Miranda, K., Chang, I.-F., Guyot, R. *et al.*, *Plant Physiol.* 2001, *127*, 398–415.
- [63] Wilson, D. N., Nierhaus, K. H., *Angewandte Chemie-International Edition* 2003, *42*, 3464–3486.
- [64] Baumann, U., Juttner, J., *Cell. Mol. Life Sci.* 2002, *59*, 1042–1057.
- [65] Dietz, K. J., *Annu. Rev. Plant Biol.* 2003, *54*, 93–107.
- [66] Ohno, M., Fornerod, M., Mattaj, I. W., *Cell* 1998, *92*, 327–336.
- [67] Hussey, P. J., Hawkins, T. J., Igarashi, H., Kaloriti, D., Smerthenko, A., *Plant Mol. Biol.* 2002, *50*, 915–924.
- [68] Mayer, U., Jurgens, G., *Curr. Opin. Plant Biol.* 2002, *5*, 494–501.
- [69] Fu, H. Y., Girod, P. A., Doelling, J. H., van Nocker, S. *et al.*, *Mol. Biol. Rep.* 1999, *26*, 137–146.
- [70] Vierstra, R. D., *Trends Plant Sci.* 2003, *8*, 135–142.
- [71] Hartl, D. L., Ochman, H., *Methods Mol. Biol.* 1996, *58*, 293–301.
- [72] Bukau, B., Horwich, A. L., *Cell* 1998, *92*, 351–366.
- [73] Estelle, M., *Curr. Opin. Plant Biol.* 2001, *4*, 254–260.
- [74] Nielsen, H., Engelbrecht, J., Brunak, S., vonHeijne, G., *Protein Eng.* 1997, *10*, 1–6.