Available online at www.sciencedirect.com
SCIENCE DIRECT®
Phytochemistry 65 (2004) 1609–1618
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

A two-dimensional proteome map of maize endosperm ☆

Valérie Méchin [a,b], Thierry Balliau [a,c], Sophie Château-Joubert [d], Marlène Davanture [c], Olivier Langella [a,c], Luc Négroni [a,c], Jean-Louis Prioul [d], Claudine Thévenot [d], Michel Zivy [a], Catherine Damerval [a,*]

[a] *INRA/INA-PG/UPS/CNRS UMR8120, UMR de Génétique Végétale, Ferme du Moulon, 91190 Gif-sur-Yvette, France*
[b] *Unité de Chimie Biologique UMR 206, INRA/INA-PG, Institut National Agronomique, 78850 Thiverval-Grignon, France*
[c] *IFR 87 La Plante et son Environnement, Plate-forme de Protéomique du Moulon, UMR de Génétique Végétale, Ferme du Moulon, 91190 Gif-sur-Yvette, France*
[d] *Laboratoire Structure et Métabolisme des Plantes, Institut de Biotechnologie des Plantes (CNRS UMR 8618), Bâtiment 630, Université Paris-Sud, 91405 Orsay Cedex, France*

Received 12 January 2004; received in revised form 27 April 2004
Available online 25 June 2004

Abstract

We have established a proteome reference map for maize (*Zea mays* L.) endosperm by means of two-dimensional gel electrophoresis and protein identification with LC–MS/MS analysis. This investigation focussed on proteins in major spots in a 4–7 p*I* range and 10–100 kDa M_r range. Among the 632 protein spots processed, 496 were identified by matching against the NCBInr and ZMtuc-tus databases (using the SEQUEST software). Forty-two per cent of the proteins were identified against maize sequences, 23% against rice sequences and 21% against *Arabidopsis* sequences. Identified proteins were not only cytoplasmic but also nuclear, mitochondrial or amyloplastic. Metabolic processes, protein destination, protein synthesis, cell rescue, defense, cell death and ageing are the most abundant functional categories, comprising almost half of the 632 proteins analyzed in our study. This proteome map constitutes a powerful tool for physiological studies and is the first step for investigating the maize endosperm development.

Keywords: *Zea mays* L.; Poaceae; Maize; 2D PAGE; LC–MS/MS; Endosperm proteins; 2D map

1. Introduction

During maize seed development, several physiological steps involving complex and inter-related processes can be recognized. For the first 3 days after pollination, free nuclear divisions followed by cellularisation of the coenocyte occur. The formation of cell walls begins at 4 days after pollination when the endosperm contains several thousand cells. From 4 to 10 days after pollination, cell division and differentiation take place, then storage compounds are synthesized from 12 days after pollination to the maturity. During this period, the endosperm accumulates large amounts of starch and storage proteins. At around 25 days after pollination, the relative water content of the endosperm begins to decrease and the dry down process initiates. Whilst morphological steps of this development are well known, until now the underlying physiological and molecular mechanisms are largely unknown. One way to appraise these complex processes is to study the proteome of the endosperm at several key stages along development, which would give insight to the functional gene products and how their expression is modulated.

Most proteomic studies still rely on two-dimensional (2D) electrophoresis, which uses two independent

☆ Supplementary data associated with this article can be found at doi:10.1016/j.phytochem.2004.04.035.

Abbreviations: ACN, acetonitrile; DTT, dithiothreitol; EST, expressed sequence tag; HPLC, high-pressure liquid chromatography; IT, ion trap; MS, mass spectrometry; PDA, piperazine diacrylyl; SB3-10, *N*-decyl-*N*,*N*-dimethyl-3-ammonio-1-propane-sulfonate; SDS, sodium dodecyl sulfate; TCA, tricarboxylic acid cycle; TCEP, tris(2-carboxyethyl)phosphine; TFA, trifluoroacetic acid.

* Corresponding author. Tel.: +33-1-69-33-2366; fax: +33-1-69-33-2340.
E-mail address: damerval@moulon.inra.fr (C. Damerval).

0031-9422/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2004.04.035

biochemical properties of proteins, the isoelectric point (p*I*) and the apparent molecular mass (M_r), to achieve good resolution of complex mixtures. The instrumental developments in 2D gel electrophoresis have offered the context for proteomics. A similar break-through in the methods for protein identification was essential. Although Edman sequencing is still useful for certain aspects of proteomics, it has generally been replaced with mass spectrometry (MS) for routine protein identification. More recently, Link et al. (1999) have initiated a novel approach whereby protein extracts are cleaved and the peptide mixture immediately separated by MS to generate peptide sequence data. This technology is termed MudPIT (Multidimensional Protein Identification Technology) and allows very efficient and high throughput protein identification studies. However, this identification approach does not enable quantitative analysis of individual proteins as allowed by 2D gel electrophoresis separation.

Proteomics in plants is still in its infancy and lags significantly behind proteomics in other eukaryotes and unicellular prokaryotes. Many eukaryotes outside of the plant kingdom have relatively complete genome sequence data available (Goffeau et al., 1996), thus facilitating proteomics approaches. In contrast, the first complete genome sequence of the model plant *Arabidopsis thaliana* has only recently been made available (Arabidopsis Genome Initiative, 2000). The sequencing of the larger and more complex genome of cereal crops is still at an early stage of progress, except for rice (*Oryza sativa*) where a quite complete sequence, still not fully annotated, is available (Goff et al., 2002; Yu et al., 2002; Niiler, 2000). A wealth of information (description of the function, domain structure, subcellular location, post-translational modifications, variants, similarities to other proteins, etc.) based on knowledge from sequenced prokaryotes and eukaryotes is now available in various public databases, as well as expressed sequence tags (ESTs) from various organisms, and provides the opportunity to develop systematic plant proteomic analysis and to study metabolic pathways. Thus, despite a slow start, there has been a noticeable increase in the number of plant proteomic studies in recent years.

A well-annotated reference map of endosperm proteome is essential for maize endosperm development studies. Here, we present the achievement of such a map based on 2D gel electrophoresis of 14 days after pollination endosperm proteins. This stage was chosen because it takes place after cell differentiation and at the beginning of storage compound synthesis. Thus, a large spectrum of functions is expected to be present among the expressed genes. Several hundreds of proteins were picked from the 2D gel, then in-gel digested prior to LC–MS/MS analysis and identified by bioinformatics (Steven et al., 2000; Dutt and Lee, 2000).

2. Results and discussion

2.1. Overview of the reference map

Fourteen days after pollination endosperm proteins were separated by 2D PAGE and stained with colloidal Coomassie blue. One thousand seven hundred and one protein spots were detected in a pH range from 4 to 7 and a M_r range from 10 to100 kDa. Six hundred and thirty-two well-resolved spots corresponding to abundant proteins, covering the whole M_r and p*I* range were manually excised and in-gel digested with trypsin (Fig. 1). The extracted peptides were then separated and analyzed by LC–MS/MS.

The identification of the peptides was performed using public databases (ZMtuc-tus and NCBInr) searching with the SEQUEST software. Four hundred and ninety-six protein spots were thus identified, 60 of which matched to two or more proteins with different functions (78% of successful identification) and 136 had no homology to any protein in the NCBInr or ZMtuc-tus databases. The characterized proteins are listed in supplementary data file. This file table contains the ProteinId, the functional category, the function description, the corresponding accession code (NCBI gi identifier), the p*I*, the M_r and the peptide sequence identified.

In comparison to proteome maps of other plants, including *A. thaliana* (Kamo et al., 1995; Tsugita et al., 1996), clover (Morris and Djordjevic, 2001), wheat (Skylas et al., 2001; Andon et al., 2002), *Medicago truncatula* (Mathesius et al., 2001) and rice (Tsugita et al., 1994, 1996; Komatsu et al., 1993, 1999; Koller et al., 2002; Heazlewood et al., 2003) our map is one of the most extensive, in that it comprises the highest number of spots processed and identified from a 2D gel obtained from a single tissue. Studies using MudPIT, as performed by Koller et al. (2002) on rice, obviously allow most extensive proteome coverage. However, considering only their seed data obtained using both 2D electrophoresis associated with tandem mass spectrometry and MudPIT, their rice map contains 877 processed protein spots, which is in a same magnitude order as our maize endosperm map.

In our survey, 42% of the proteins were identified according to similarity with maize sequences (EST or NCBInr), 23% and 21% with rice and *Arabidopsis* sequences, respectively. The quite high contribution of rice was expected due to its almost sequenced genome added to the synteny and high homology encountered in cereal genomes. The model eu-dicot plant *Arabidopsis*, whose genome is fully sequenced and annotated also provided a lot of information, in spite of the divergence between monocots and eu-dicots. Various species (monocots or eu-dicots) contributed to the remaining identifications.

The identified proteins are located within the cytoplasmic, mitochondrial, endoplasmic, amyloplastic and

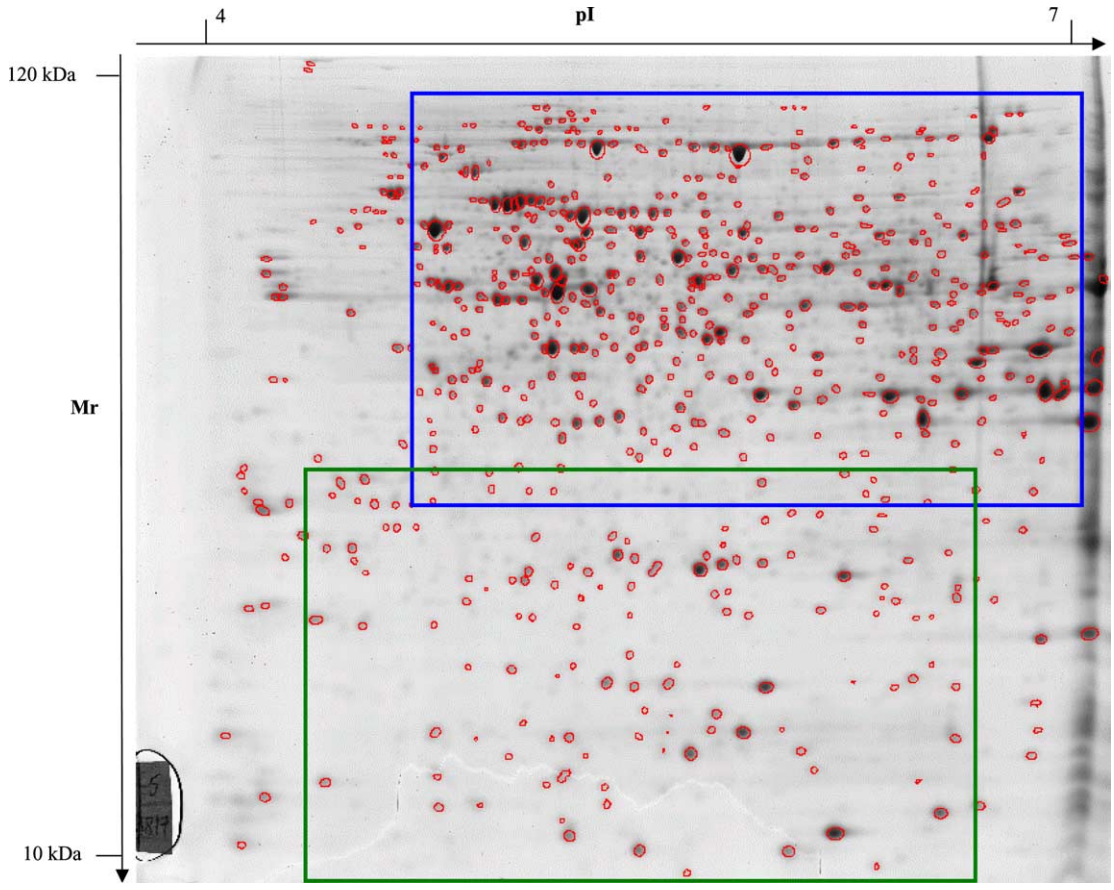

Fig. 1. Two-dimensional gel of 14 days after pollination endosperm proteins. Proteins (250 μg) were separated on a 4–7 pH gradient IPG strip and subsequently on an 11% SDS–PAGE gel. Gel was stained with colloidal Coomassie blue. 1701 spots were detected using the Melanie software. The red circles indicate the 632 spots that were excised prior to in-gel digestion and LC-MS/MS analysis. Blue and green frames correspond to areas of the 2D gel presented in detail in Figs. 4 and 6.

nuclear compartments. It is encouraging to note that our protein extraction and solubilization protocol allowed visualization of nuclear and organelle proteins, and not only cytoplasmic ones. As a matter of fact, the sample preparation should break organelles but the important dilution of organelle and nuclear proteins in comparison with cytoplasmic ones might have prevented visualization.

Among the identified proteins, two categories could not be assigned a function. First, 45 proteins matched with sequences with not yet clear cut function in the databases. Second, 60 spots matched with 2–4 different proteins in databases according to the SEQUEST search results. It is already known that a spot can sometimes correspond to a mix of several proteins (Parker et al., 1998). These overlapping proteins may be separated to discrete spots by using narrower pH gradient and/or using wider SDS–PAGE gels for second dimensional electrophoresis.

We found 136 non-identified (NI) protein spots and 45 proteins with not yet clear cut function (29%), which could be compared to the results of Koller et al. (2002) on rice seed, reporting about 30% of proteins with unidentified function or with very low or no homology. Both percentages were similar, in spite of additional identification using the MudPIT technology and the genome being completely sequenced in the rice study. On a whole the protein spots with no identification or no clear function (not yet clear cut or multiple homology) represented 38% of the proteins processed (Fig. 2).

The 391 protein spots unambiguously identified corresponded to 180 different functions, which means that 71% of the identified protein spots were actually isoforms (Fig. 3). At this stage of the analysis, it is not possible to know whether these multiple isoforms corresponded to products of different yet related genes or to post-translational modifications of a same gene product. For example, three different protein spots (90, 99 and 129, Fig. 4) were identified as cytosolic triose phosphate isomerase. All had similar M_r (27,650, 26,250 and 27,000 Da, respectively), close to the theoretical value, but different p*I* (5.4, 5.05 and 5.7, respectively). Triose phosphate isomerase plays a central role in glycolysis and as

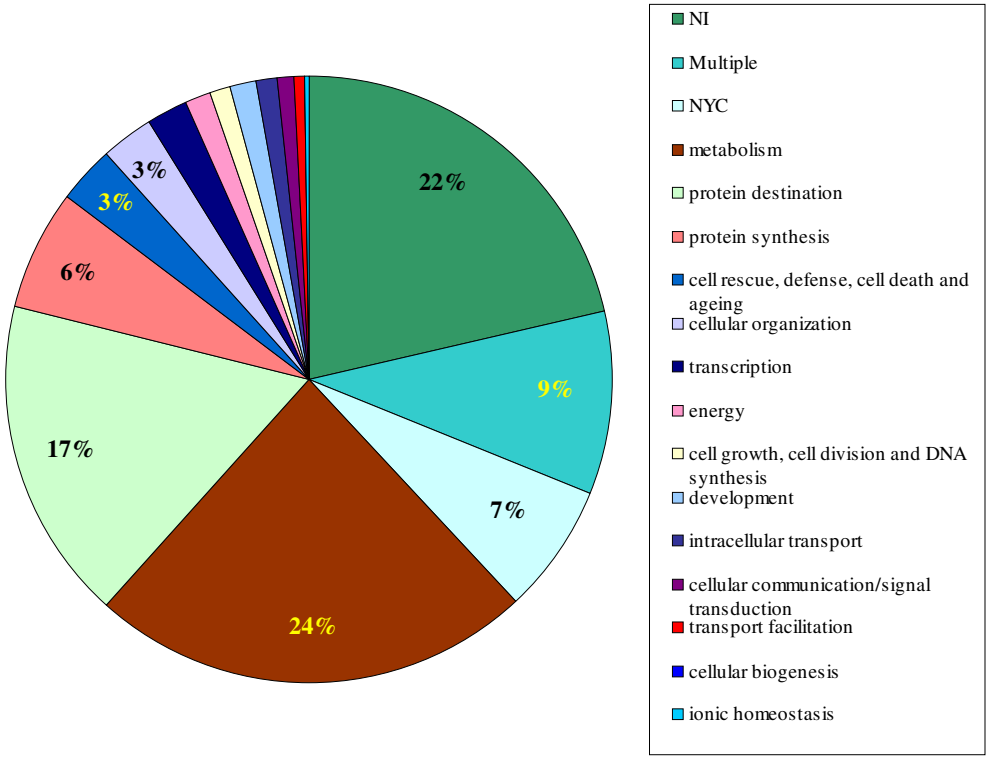

Fig. 2. Proportion of the categories (15 functional categories and proteins either not identified or without a clear function) among the 632 processed protein spots.

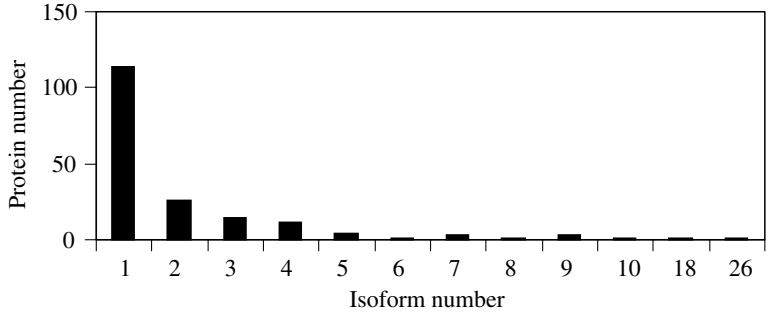

Fig. 3. Distribution of isoform number among the identified proteins.

such, has been identified in numerous plant studies (Skylas et al., 2001; Gallardo et al., 2001; Ostergaard et al., 2002; Koller et al., 2002; Finnie et al., 2002). Ostergaard et al. (2002) on barley, Koller et al. (2002) on rice and Finnie et al. (2002) on barley detected, respectively, 4, 2 and 7 spots for triose phosphate isomerase. Another example concerned ascorbate peroxidase, a major enzyme involved in detoxication of hydrogen peroxide, which occurred as four protein spots (88, 131, 135 and 136, supplementary data file) with similar M_r (about 27 kDa) but with different p*I*. Finnie et al. (2002) showed that a cytosolic form of this enzyme was only detectable in early developmental stage of barley seed. This result is consistent with our study performed on 14 days after pollination maize endosperm.

2.2. Functional categories of identified proteins

A comprehensive knowledge of the function of each protein can help us in clarifying there contribution in cellular functioning. Fifteen functional categories as defined by Schoof et al. (2002) were found among the 391 unambiguously identified protein spots (Fig. 2). No additional category was found among the 60 protein spots with multiple homologies and the 10 most abundant functional categories were found in similar relative abundance compared to the unambiguously identified spots. However, few protein types were uniquely found among this subset (e.g., enzymes of sulfur assimilation, or small subunit of ADP glucose pyrophosphorylase, see below).

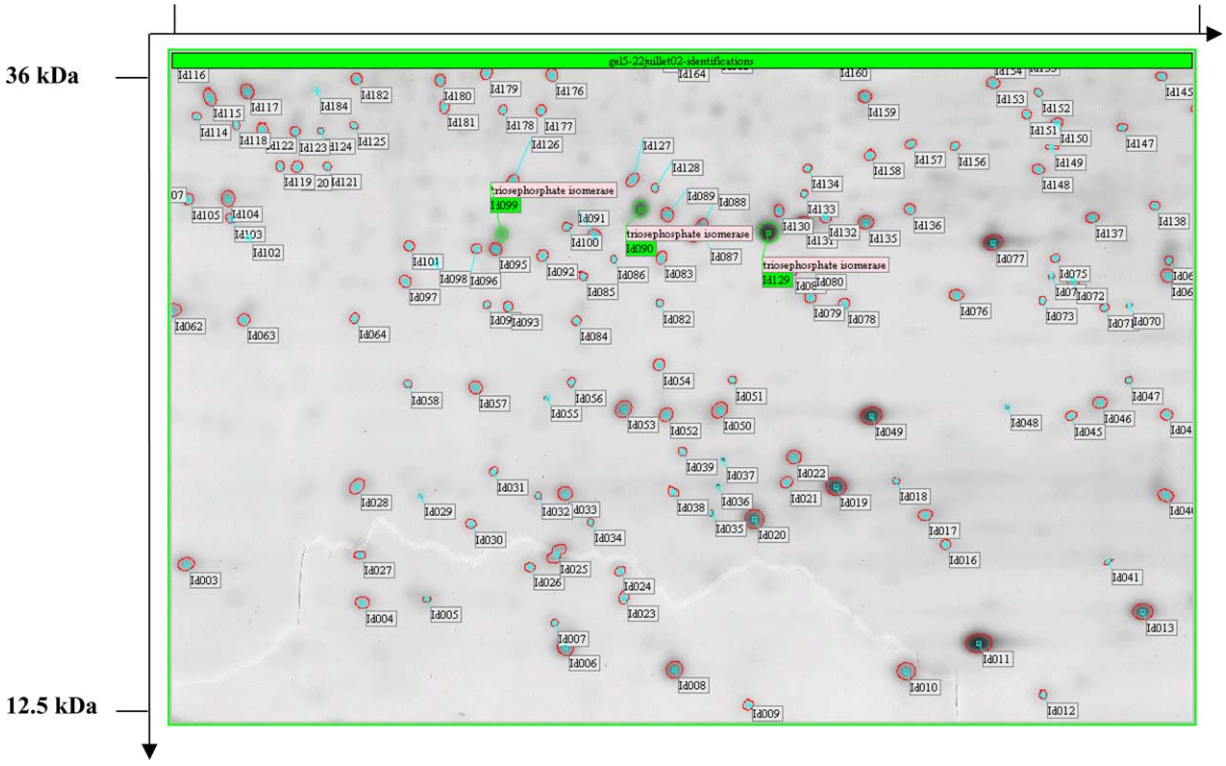

Fig. 4. Portion of the 2D reference map showing, in green, the three isoforms of triose phosphate isomerase.

The most abundant functional category related to metabolic processes (24%, Fig. 2), with numerous pathways represented. The predominance of this very global metabolic category is in agreement with other data from rice and *Arabidopsis*. Even if in silico gene prediction is an imperfect process (Hogenesch et al., 2001) the classification performed on translated genes by Goff et al. (2002) on rice genome and on *A. thaliana* by the Arabidopsis Genome Initiative (2000) also showed a predominance of the metabolism function category. The same result was obtained in an extensive rice proteome study involving 2528 proteins (Koller et al., 2002).

The second most abundant category corresponded to proteins involved in protein destination (17%, Fig. 2). Actually, the predominance of this category came largely from the heat-shock proteins (Hsp or Hsc) that represented 57% of the proteins identified in this class. Proteins involved in other cellular activities were also detected, covering protein synthesis (6%), cell rescue and defense (3%), cellular organisation (3%), energy, transport, transcription, DNA synthesis and cellular communication (Fig. 2). Very few storage proteins were identified and classified in the development functional category. In maize, zeins are the core of protein endosperm reserves and are the equivalent of prolamins in other cereals. They begin to be synthesized in the endosperm 12–15 days after pollination. Zeins, which are very basic proteins, were not allowed to focus on a 4–7 pH gradient. Not surprisingly, we did not find any zein proteins in our survey. On another hand, three legumin-like proteins homologous to globulins, and two germins were identified.

A few functional categories deserve a more precise analysis of the pathways represented.

2.2.1. Metabolism

Maize is an autotrophic organism that only needs minerals, light, water and air to synthesize organic compounds and to grow, however, endosperm is a heterotrophic organ. A large proportion of its proteins supports primary metabolic processes and synthesis of more or less complex molecules such as nucleotides, amino acids, carbohydrates, lipids and secondary compounds. Accordingly, many enzymes involved in these processes were identified in this study.

The primary metabolism, including glycolysis, TCA cycle, fermentation and biochemical reactions of photosynthesis, sorted 44% of the identified proteins within the metabolic class (Fig. 5). Among the 10 enzymatic steps involved in the glycolytic pathway, nine were identified as 32 protein spots: triose phosphate isomerase (three spots), glyceraldehyde 3-phosphate dehydrogenase (nine spots), aldolase (one spot), enolase (seven spots), pyruvate kinase (one spot), phosphoglycerate kinase (four spots), phosphofructokinase (one spot), phosphoglycerate mutase (five spots) and hexokinase (one spot). No polypeptide corresponding to phosphohexose isomerase involved in the second step of the

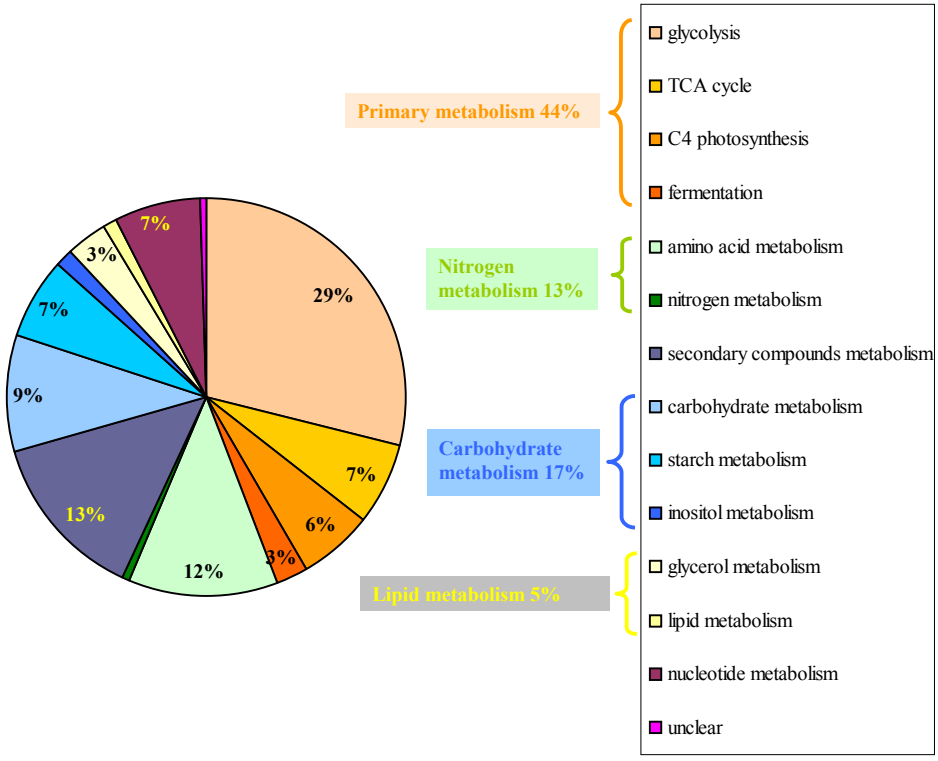

Fig. 5. Proportion of the various pathways identified within the metabolism functional category.

glycolytic pathway was found among the excised protein spots.

Eleven protein spots were also identified as elements of the pyruvate dehydrogenase complex that catalyzes the conversion of pyruvate to acetyl-CoA, a step linking the glycolytic pathway to the TCA cycle. Half of the enzymes involved in the TCA cycle were identified in our study: aconitase (three spots), malate dehydrogenase (three spots), succinate deyhdrogenase (one spot) and succinyl-coA ligase (two spots). This small proportion of identified enzymatic functions, in comparison to the percentage identified for glycolysis, is most probably explained by the mitochondrial localization of the reactions involved in the TCA cycle. Remarkably, most proteins involved in these primary metabolic pathways were identified as multiple isoforms, rising the question of whether these various isoforms may be located in different cell compartments, and/or differentially regulated, thus possibly assuming slightly different physiological roles. Plants have frequently many genes encoding one enzyme. For example, the 10 enzymes involved in the glycolytic process may be encoded by up to 51 genes that are present in 1–8 copies (Arabidopsis Genome Initiative, 2000).

Synthesis of amino acids is critical for cell survival. They not only serve as the building blocks for proteins, but also as starting points for the synthesis of many important cellular molecules including vitamins, phytohormones and nucleotides. The amino acid synthesis pathways can be grouped into several categories: simple reactions, branched chain amino acids, aromatic amino acids, threonine/lysine and serine/glycine. Eighteen proteins participating in amino acid metabolism were identified in our map (12%, Fig. 5), including 12 proteins specifically involved in the synthesis of branched amino acids (3), aromatic amino acids (1), tryptophan (2), cystein (1), threonine (1), methionine (2), glycine (1) and serine (1).

Starch, the most common storage polysaccharide in plants (Smith et al., 1997) is synthesized in endosperm amyloplasts. The first step in starch biosynthetic pathway is the conversion of glucose-1-phosphate into ADP-glucose, through the action of a tetrameric enzyme, the ADP-glucose pyrophosphorylase (AGPase). The ADP-glucose next serves as a glucosyl donor for α-glucan synthesis by starch synthases, starch branching and debranching enzymes (Beckles et al., 2001; Slattery et al., 2000). AGPase is a homotetramer in bacteria while in plants, it is a heterotetramer of two different, yet evolutionary related subunits (reviewed in Sivak and Preiss, 1994). The two small subunits are responsible for the catalytic properties and the two large subunits for the regulatory properties (Smith-White and Preiss, 1992; Fu et al., 1998). In maize, the large and small subunits encoded by the *Shrunken2* and *Brittle2* loci, respectively, are considered as specific of the endosperm (Sivak and Preiss, 1994). The large subunit of AGPase was detected in our analysis, with three protein

spots (271, 283 and 298, Fig. 6) matching to the endosperm specific subunit and one protein spot (626) matching to the large subunit mainly present in embryo and in lesser extent in endosperm. The number of large subunit isoforms is a minimum estimate since spot 272 matched both the large subunit and a component of the pyruvate dehydrogenase complex. Similarly, the endosperm specific small subunit cannot unambiguously be identified, since it was found in spot 576 that also matched an enolase (Fig. 6). Relative apparent molecular masses of the two subunits on the 2D gel are not consistent with the theoretical values (57–58 kDa for the large subunit and about 52 kDa for the small subunit of maize), which raises the question of the actual identity of spot 576. The use of subunit-specific antibodies should help us to confirm the presence and relatedness of the endosperm subunit isoforms, and the presence of the large subunit usually considered to be expressed mainly in the embryo. Among starch synthesizing enzymes, one starch synthase was also observed.

Enzymes for the starch degradation pathway include debranching enzymes, disproportionating enzymes, isoamylases, α-amylases, β-amylases, α-glucosidases and starch phosphorylases (Kossmann and Lloyd, 2000). Only one starch phosphorylase enzyme was identified among the excised spots. The presence of several starch biosynthetic enzymes and the absence of most catabolic enzymes among the identified protein spots might simply be explained by the fact that starch is just beginning to be synthesized at 14 days after pollination stage.

One α-amylase inhibitor was also identified. According to Finnie et al. (2002), this protein is expected to accumulate during grain filling and is involved in the defense of starch reserves of the grain against invading insect pathogens by inhibiting the α-amylase activity of extracts from insect pests (Gutierrez et al., 1990).

Among the secondary compound category, seven protein spots involved in cell wall lignification or cell wall polysaccharides synthesis were identified. In contrast, in rice, Koller et al. (2002) identified enzymes involved in the phenylpropanoid metabolism in a quite important proportion in leaves and in a lesser extent in roots, but not in seeds where only the phenylalanine ammonia lyase was found.

2.2.2. Protein synthesis and protein destination

In accordance with previous studies (Koller et al., 2002; Goff et al., 2002; Arabidopsis Genome Initiative, 2000), protein synthesis and protein destination were two major classes in our functional map, since they encompassed 40 and 109 protein spots, respectively. Thirty three proteins belonging to the protein destination family were involved in protein degradation, mostly the ubiquitin-proteasome pathway. It is noteworthy that

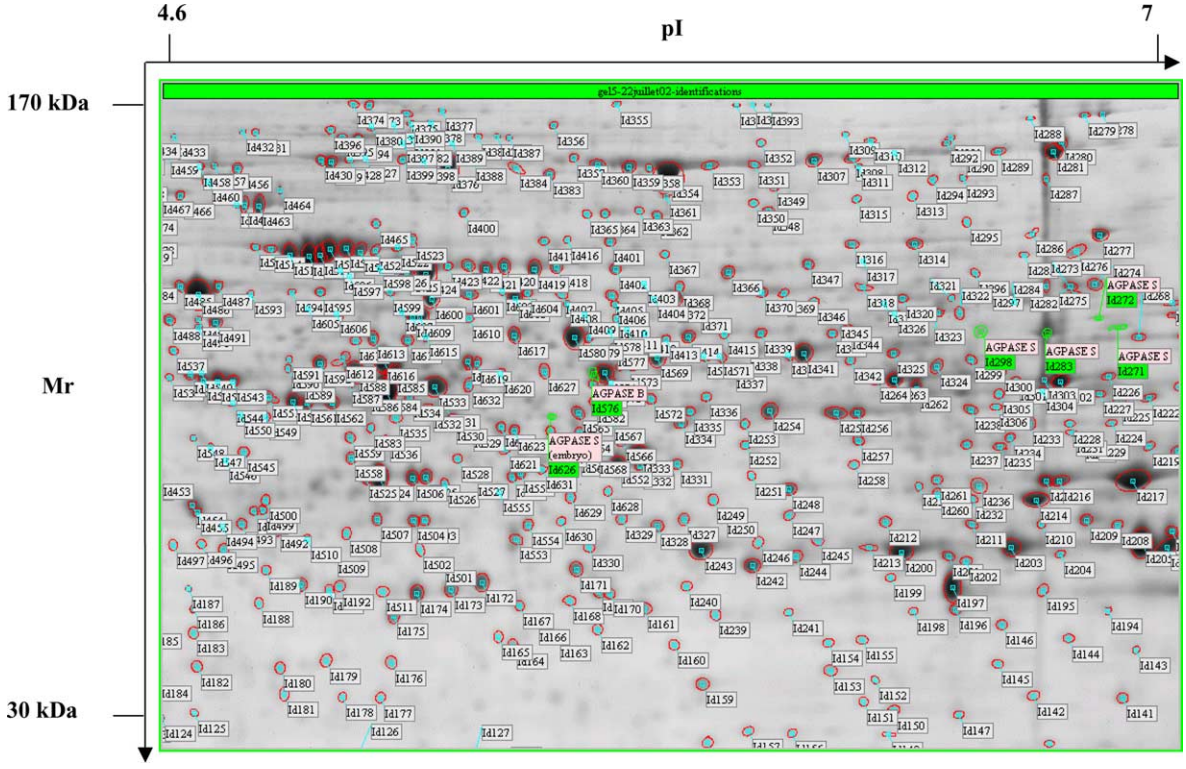

Fig. 6. Portion of the 2D reference map showing the isoforms of AGPase. Id271, Id272, Id283 and Id298 are for the endosperm specific large subunit while Id626 matched the embryo large subunit. Id576 corresponds to the endosperm specific small subunit.

the numbers of proteins participating in protein degradation and protein synthesis are in the same range. In the protein synthesis category, mainly 60S and 40S ribosomal proteins, translation initiation and elongation factors were identified. In the protein destination category, 62 proteins were identified as Hsp and/or chaperone/chaperonins. All Hsp categories were represented (Hsp100, 90, 70, 60 and Hsp20), and all appeared as multiple isoforms (up to 26 isoforms for Hsp70, Fig. 3). Actually Hsp proteins represented about 10% of the proteins processed.

3. Experimental

3.1. Preparation of the protein sample

The INRA maize inbred line F2 is becoming a reference genotype for many studies in Europe. F2 plants were grown and selfed in the field at Le Moulon (INRA, 91190 Gif-sur-Yvette, France) in 2001, and ears were collected 14 days after pollination. Kernels from the middle part of the ear were sampled and dissected to remove embryo and pericarp. Endosperms were stored in liquid nitrogen.

Protein extraction was performed according to Damerval et al. (1986). Briefly, 30 endosperms were ground together in a mortar with a pestle in liquid nitrogen. The powder was resuspended in acetone with 0.07% v/v β-mercaptoethanol and 10% w/v trichloroacetic acid. Proteins were allowed to precipitate for 1 h at −20 °C. Then the pellet was washed overnight with acetone containing 0.07% v/v β-mercaptoethanol. The supernatant was discarded and the pellet dried under vacuum. The resulting pellet was resuspended in R2D2 buffer (Urea 5M, Thiourea 2M, CHAPS 2%, SB3-10 2%, DTT 20 mM, TCEP 5 mM carrier and ampholytes) according to Mechin et al. (2003).

3.2. Two-dimensional gel electrophoresis

3.2.1. Isoelectric focusing (first dimension)

Isoelectric focusing (IEF) was performed using 24 cm immobilized pH gradient (IPG) strips (Amersham Biosciences, Uppsala, Sweden) with a linear pH gradient from 4 to 7. Solubilized protein concentration was estimated using the PlusOne 2D Quant Kit (Amersham Biosciences, Uppsala, Sweden) and 250 μg of the solubilized proteins was applied on an IPG strip for in-gel rehydration. The focusing was achieved using a Protean IEF Cell (Bio-Rad, Hercules, CA, USA). An active rehydration was performed at 22 °C during 12 h at 50 V; then the focusing itself was achieved. For improved sample entry the voltage was increased step by step from 50 to 10,000 V (0.5 h at 200 V, 0.5 h at 500 V, 1 h at 1000 V then 10,000 V for a total of 94,000 V h). After IEF, strips were equilibrated according to Görg et al. (1987) to improve protein transfer to the second dimension gel.

3.2.2. SDS–PAGE (second dimension)

The equilibrated strips were sealed at the top of the 1 mm thick second dimensional gel (24 × 24 cm) with the help of 1% low-melting agarose in SDS–electrophoresis buffer (Tris 25 mM, glycine 0.2 M and SDS 0.1%). Continuous gels (11% T, 2.67% C gels with PDA as cross-linking agent) were used. Separation was carried out at 20 V for 1 h and subsequently at a maximum of 30 mA/gel, 120 V overnight, until the bromophenol blue front had reached the end of the gel.

3.2.3. Protein staining

Following SDS–PAGE, the gels were stained with colloidal Coomassie blue G250. Scanning was carried out at 300 dpi with an Image Scanner (Amersham Biosciences, Uppsala, Sweden) and image analysis was performed with Melanie 3.08 software (Swiss Institute of Bioinformatics, Geneva, Switzerland).

3.3. Mass spectrometry analyses

3.3.1. In-gel trypsin digestion

The Coomassie stained protein spots were manually excised. In-gel digestion was performed with a Progest system (Genomic Solution, Huntingdon, UK) according to a standard trypsin protocol. Briefly, after a washing step, gel particles were digested during 5 h with 125 ng of modified trypsin (Promega, Madison, WI). The resulting peptides were extracted with 30 μl of 5% TFA, 10% ACN then 30 μl of 0.2% TFA, 83% ACN. After drying in a vacuum centrifuge, peptide extracts were resuspended in 20 μl of 0.1% TFA, 3% ACN.

3.3.2. NanoHPLC-MS/MS and IT-MS analysis

HPLC was performed with Ultimate LC system combined with Famos autosample and Switchos II microcolumn switching for pre-concentration (LC Packings, Amsterdam, The Netherlands). The sample was loaded on the column (PEPMAP C18, 5 μm, 75 μm internal diameter, 15 cm, LC Packing) using a preconcentration step on a micro precolumn cartridge (300 μm internal diameter, 5 mm). Five microlitres of sample was loaded on precolumn at 5 μl/min. After 3 min, the precolumn was connected with the separating column and the gradient was started at 200 nl/min. Buffers were 0.1% HCOOH, 3% ACN (A) and 0.1% HCOOH, 95% ACN (B). A linear gradient from 5% to 30% B for 25 min was applied. Including the regeneration step, one run was 60 min length. The LCQ deca xp+ (Thermofinnigan, les Ulis, France) was used with a nano electrospray interface. Ionization (1.2–1.4 kV ionization potential) was

performed with liquid junction and non coated capillary probe (New Objective, Cambridge, USA). Peptides ions were analysed by the *N*th-dependent method as follows: (i) full MS scan (m/z 500–1500), (ii) ZoomScan (scan of the two major ions with higher resolution) and (iii) MS/MS of this two ions.

3.4. Database analysis

The SEQUEST software (Thermofinnigan, les Ulis, France) was used to interpret MS/MS. Identification was performed with Sequest^{TM} using protein sequences databases downloaded from the NCBI (http://www.ncbi.nlm.nih.gov/) and maize EST databases from Plant GDB (http://www.plantgdb.org/). Peptides identified by SEQUEST were filtered according to their charge state, cross-correlation score (Xcorr, >1.7 for $n+1$ and >2.2 for $n+2$), normalized difference in correlation score ($\Delta C_n \geqslant 0.2$) and the tryptic nature of each peptide.

4. Conclusion

This study shows that large-scale proteomic analyses in maize have become realistic provided that comprehensive genome databases are now available in plants. Actually, the 2D technique followed by MS analysis is more limited than the high throughput MudPIT methodology for protein identification, however it settles the ground for quantitative developmental studies of individual proteins. The reference map of endosperm described here is a first step towards understanding the coordination of physiological and molecular events during maize endosperm development. It could of course be complemented. Firstly, it is possible to adapt the sample processing in order to characterize other specific proteins like membrane proteins or low abundance nuclear proteins, basic proteins, etc. Secondly, the identification of additional proteins, for example showing a remarkable development scheme, will also be possible using 2D gels of other developmental stages. Such analysis of other endosperm developmental stages is currently under way in our laboratory. The obtained dynamic reference map will serve as an important tool for maize endosperm functional genomics, i.e., for linking proteomic, transcriptomic and metabolic analyses.

Acknowledgements

This work was funded by the European Community in the context of the Zeastar European program (QLRT-2000-00020), and V.M. was supported by a grant within the same program. The Plate-forme de Protéomique du Moulon and its equipment was supported by grants from Région Ile-de-France, IFR 87, Génoplante, INRA and CNRS.

References

Andon, N.L., Hollingworth, S., Koller, A., Greenland, A.J., Yates 3rd, J.R., Haynes, P.A., 2002. Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. Proteomics 2, 1156–1168.

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

Beckles, D.M., Smith, A.M., ap Rees, T., 2001. A cytosolic ADP-glucose pyrophosphorylase is a feature of graminaceous endosperms, but not of other starch-storing organs. Plant Physiology 125, 818–827.

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

Dutt, M.J., Lee, K.H., 2000. Proteomic analysis. Current Opinion in Biotechnology 11, 176–179.

Finnie, C., Melchior, S., Roepstorff, P., Svensson, B., 2002. Proteome analysis of grain filling and seed maturation in barley. Plant Physiology 129, 1308–1319.

Fu, Y., Ballicora, M.A., Preiss, J., 1998. Mutagenesis of the glucose-1-phosphate-binding site of potato tuber ADP-glucose pyrophosphorylase. Plant Physiology 117, 989–996.

Gallardo, K., Job, C., Groot, S.P., Puype, M., Demol, H., Vandekerckhove, J., Job, D., 2001. Proteomic analysis of arabidopsis seed germination and priming. Plant Physiology 126, 835–848.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., 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, R.M., 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.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G., 1996. Life with 6000 genes. Science 274 (5287), 546–567.

Görg, A., Postel, W., Weser, J., Günther, S., Strahler, J.R., Hanash, S., Somerlot, L., 1987. Elimination of point streaking on silver stained two-dimensional gels by addition of iodoacetamide to the equilibration buffer. Electrophoresis 8, 122–124.

Gutierrez, C., Sanchez-Monge, R., Gomez, L., Ruiz-Tapiador, M., Castañera, P., Salcedo, G., 1990. Amylase activities of agricultural insect pests are specifically affected by different inhibitor preparations from wheat and barley endosperms. Plant Sciences 72, 37–44.

Heazlewood, J.L., Howell, K.A., Whelan, J., Millar, A.H., 2003. Towards an analysis of the rice mitochondrial proteome. Plant Physiology 132, 230–242.

Hogenesch, J.B., Ching, K.A., Batalov, S., Su, A.I., Walker, J.R., Zhou, Y., Kay, S.A., Schultz, P.G., Cooke, M.P., 2001. A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. Cell 106, 413–415.

Kamo, M., Kawakami, T., Miyatake, N., Tsugita, A., 1995. Separation and characterization of *Arabidopsis thaliana* proteins by two-dimensional gel electrophoresis. Electrophoresis 16, 423–430.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., Yates 3rd, J.R., 2002. Proteomic survey of metabolic pathways in rice. Proceedings of the National Academy of Sciences of the United States of America 99, 11969–11974.

Komatsu, S., Kajiwara, H., Hirano, H., 1993. A rice protein library — a data-file of rice proteins separated by 2-dimensional electrophoresis. Theoretical and Applied Genetics 86, 935–942.

Komatsu, S., Muhammad, A., Rakwal, R., 1999. Separation and characterization of proteins from green and etiolated shoots of rice (*Oryza sativa* L.): towards a rice proteome. Electrophoresis 20, 630–636.

Kossmann, J., Lloyd, J., 2000. Understanding and influencing starch biochemistry. Critical Review in Biochemical Molecular Biology 35, 141–196.

Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., Yates 3rd, J.R., 1999. Direct analysis of protein complexes using mass spectrometry. Nature Biotechnology 17, 676–682.

Mathesius, U., Keijzers, G., Natera, S.H., Weinman, J.J., Djordjevic, M.A., Rolfe, B.G., 2001. Establishment of a root proteome reference map for the model legume *Medicago truncatula* using the expressed sequence tag database for peptide mass fingerprinting. Proteomics 1, 1424–1440.

Mechin, 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.

Morris, A.C., Djordjevic, M.A., 2001. Proteome analysis of cultivar-specific interactions between *Rhizobium leguminosarum* biovar trifolii and subterranean clover cultivar Woogenellup. Electrophoresis 22, 586–598.

Niiler, E., 2000. Monsanto releases rice data to academia. Nature Biotechnology 18, 484.

Ostergaard, O., Melchior, S., Roepstorff, P., Svensson, B., 2002. Initial proteome analysis of mature barley seeds and malt. Proteomics 2, 733–739.

Parker, K.C., Garrels, J.I., Hines, W., Butler, E.M., McKee, A.H., Patterson, D., Martin, S., 1998. Identification of yeast proteins from two-dimensional gels: working out spot cross-contamination. Electrophoresis 19, 1920–1932.

Schoof, H., Zaccaria, P., Gundlach, H., Lemcke, K., Rudd, S., Kolesov, G., Arnold, R., Mewes, H.W., Mayer, K.F., 2002. MIPS Arabidopsis thaliana Database (MAtDB): an integrated biological knowledge resource based on the first complete plant genome. Nucleic Acids Research 30, 91–93.

Sivak, M.N., Preiss, J., 1994. Synthesis of starch in seeds. In: Negbi, M., Kigel, J., Galili, G. (Eds.), Seed Development and Germination. Marcel Dekker, New York, pp. 139–168.

Skylas, D.J., Copeland, L., Rathmell, W.G., Wrigley, C.W., 2001. The wheat-grain proteome as a basis for more efficient cultivar identification. Proteomics 1, 1542–1546.

Slattery, C.J., Kavakli, I.H., Okita, T.W., 2000. Engineering starch for increased quantity and quality. Trends in Plant Science 5, 291–298.

Smith, A.M., Denyer, K., Martin, C., 1997. The synthesis of the starch granule. Annual Review in Plant Physiology Plant Molecular Biology 48, 67–87.

Smith-White, B.J., Preiss, J., 1992. Comparison of proteins of ADP-glucose pyrophosphorylase from diverse sources. Journal of Molecular Evolution 34, 449–464.

Steven, P., Gygi, S.P., Aebersold, R., 2000. Mass spectrometry and proteomics. Current Opinion in Chemical Biology 4, 489–494.

Tsugita, A., Kamo, M., Kawakami, T., Ohki, Y., 1996. Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. Electrophoresis 17, 855–865.

Tsugita, A., Kawakami, T., Uchiyama, Y., Kamo, M., Miyatake, N., Nozu, Y., 1994. Separation and characterization of rice proteins. Electrophoresis 15, 708–720.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L., Yang, H., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296, 79–92.