Proteomics analysis of albumins and globulins from either endosperm or the aleurone layer

Gerard G. Branlard, M. M. Merlino, Isabelle Nadaud, Emmanuelle Bancel, Bastien Laubin

PROTEOMICS ANALYSIS OF ALBUMINS AND GLOBULINS FROM EITHER ENDOPSERM OR THE ALEURON LAYER

Branlard G., Merlino M., Nadaud I., Bancel E., Laubin B.

Institut National de la Recherche Agronomique, UMR 1095, 234, av. du Brezet, 63 100 Clermont Ferrand, France, E-mail: branlard@clermont.inra.fr

Key words: wheat endosperm, developing kernel, chromosome mapping, proteomics

Abstract

To illustrate the usefulness of the proteomics tool, we report on two experiments whose objectives were to (1) analyse albumins and globulins in developing kernel, or (2) map albumins and globulins using segregating progeny, or reveal these proteins in the aleurone layer of two different cultivar.

Proteomic analysis every 50°C day (average cumulative day temperature) during kernel formation from the ovule stage to 18 days after anthesis allowed many albumins and globulins to be tracked during cell and tissue formation. Quantitative analysis of spot volume and qualitative analysis (appearance/disappearance of a spot) evidenced five types of protein expression.

Two-dimensional electrophoresis pH 4-7 x SDS Page of the albumins-globulins extracted from wholemeal kernels of 'Synthetic' or 'Opata' revealed 2125 and 2163 spots respectively. Only 10.6% of the spots differed between these two parents of the mapped progeny. The unbiased segregating spots were successfully mapped on 21 chromosomes.

Wheat AL was isolated from mature seeds of cv. Chinese Spring and cv. Recital and more than 700 Coomassie stained proteins were revealed on 24 cm IPGE pH 3-10 x SDS-Page.

Nearly 20% of these spots differed between the two cultivars.

Proteomics analysis of seed is an increasingly available tool that enables biochemists and geneticists to investigate the diversity and function of the proteins expressed in different cell organites such as protein bodies, starch granules or kernel compartments.

Introduction

The term proteome covers all proteins present in a sample at a given point in time and under given conditions. Although plant proteomics is still in its infancy, many different approaches have already been used for genetic and physiological studies (for review see Thiellement et al., 2006).

Wheat endosperm proteins, which were first separated using two-dimensional electrophoresis (2DE), were partly identified after Edman amino-acid sequencing (Skylas et al., 2000, Nakamura 2001). The numerous spots resolved on 2DE can be visualized through image analysis and their variation – i.e. the presence / absence and quantitative fluctuations- can be detected. Subsequently, spots of interest are subjected to mass spectrometry for identification. Several proteomic studies of the wheat kernel focused on the impacts of aneuploidy (Islam et al., 2002, Dumur et al., 2004), heat stress (Skylas et al., 2002, Majoul et al., 2003, 2004, Dupont et al. 2006), nitrogen fertilization (Bahrman et al., 2004) and during kernel development (Vensel et al., 2005) or on the results of the effect of the 1BL/1RS translocation (Gobaa et al., 2007a, b).

Although wheat storage proteins have received a great deal of attention in the two last decades, the proteomics approach would be very useful for deciphering the complex regulation mechanism involved in their synthesis. The storage proteins whose genetic determinism, allelic diversity, protein sequences and functional properties have already been investigated are in fact, the result of numerous seed enzymes that remain to be identified. Most of these enzymes belong to the soluble group of endosperm proteins: albumins-globulins (AG). The diversity of AG present in the developing endosperm, or still present in the mature kernel, is largely unknown. To illustrate the usefulness of the proteomics tool, we report on experiments carried out at INRA on (1) developing wheat kernel where the AG were analysed during the first 18 days after anthesis (DAA), (2) mature wheat kernel in which AG were identified and mapped in a segregating progeny, and (3) proteins present in the isolated aleurone layer (AL). The usefulness of these proteomics approaches for geneticists and technologists involved in improving wheat kernel components is discussed.

Plant Materials

Hexaploid wheat cultivars were grown at the INRA Research Centre in Clermont Ferrand, France. For developmental proteomics, the Recital cultivar was grown in controlled conditions and the day of flowering was recorded for each main spike. Developing kernels located in the medium part of the main spike of each wheat plant were collected every 50°C cumulative day temperature from 0 to 18 DAA. From 4 to 20 developing seeds were used for protein extraction.

The two parents, Synthetic and Opata cultivars, and their ITMI progeny, were obtained from a field nursery in 2000, where they had been grown in normal conditions with full fungicide protection. Fifteen mature seeds of each of the 112 F9 lines were ground and the wholemeal used for protein extraction.

Proteomics of the aleurone layer was performed on mature seeds of Recital and Chinese Spring cvs. The procedure used for isolating the aleurone layer is described in Laubin et al. 2007.

Methods

The following were the main steps in proteomic analysis:

1- Protein extraction: Extraction of AG from mature kernel was performed in a sodium phosphate buffer as previously described (Branlard, Bancel, 2006). For developmental proteomics, the young kernels were crushed under liquid nitrogen using a pestle and mortar. Proteins were extracted using glacial acetone, 0.07% β-mercaptoethanol, inhibitor proteases at 3.4 µl/100 mg. The proteins extracted were then precipitated with glacial acetone and dried before storing at -80°C, or used for isofocussing. Proteins were extracted from the aleurone layer in the following buffer: Urea 7 M, 4% CHAPS, Thiourea 2M, 1.2% Destreak reagent (Amersham Biosciences) and 1% IPG buffer (Amersham Biosciences).

2- 2DE. Although 2DE is not perfect and must be technically correctly performed, it is currently the most efficient method for protein separation (Rabilloud, 2002). In our case, Immobiline pH gradient gel electrophoresis (IPGE) was used as the first dimension and SDS poly-acrylamid gel electrophoresis as the second dimension. The IPG gradients were 3-10, 4-7 and 3-10 for developing kernel, AG mapping in ITMI progeny and AL analysis, respectively. Three to six replicates were performed to identify the common spots that characterise each proteic sample. Coomassie blue (CB) or silver staining procedures were used before gel scanning in transmission mode.

3- **Image analysis**. Melanie-3 software (GeneBio, Geneva, Switzerland) or 2D Platinum (GE Healthcare) was used to compare images and spots. The SAS GLM procedure (SAS, 1985) was used to test spot volume and percentage of spot volume between images of consecutive sampling in developing kernel. Spots in the ITMI progeny were identified by comparing the parental images to the images of each ITMI line.

4- Mass spectrometry (MS). Spots of interest were excised and subjected to trypsic digestion as described by Gobaa et al. (2007a). Two spectrometers were used, (1) matrix-assisted laser desorption – the time of flight (MALDI-TOF) spectrometer giving peptide mass fingerprints and/or (2) Electrospray ion trap MS/MS spectrometer giving partial sequences of peptides.

5-. Protein identification. The monoisotopic peptide masses resulting from the trypsic digest were compared with peptide masses in databases such as SwissProt, (http://www.expasy.org/sprot/), NCBI (http://www.ncbi.nlm.nih.gov/) or EST databases using Mascot and Profound software (http://www.matrixscience.com and http://prowl.rockefeller.edu). Matches to protein sequences from the *Viridiplantae* taxon were considered acceptable if at least four peptide masses from the PMF matched, and a Z score of 1.60 or higher was obtained using ProFound, or a significant score was obtained using MASCOT based on the program's algorithm.

Mass data collected during the LC-MS/MS analysis were processed with Sequest software. The protein was identified by searching for the peptide masses and the MS/MS sequences in the SwissProt (http://www.expasy.org/sprot/) and NCBI (http://www.ncbi.nlm.nih.gov/) non-

redundant sequence databanks. Proteins were considered to have been identified if at least two non-redundant peptides matched a single reference in the databases.

Results and discussion

1- Developing wheat kernel.

2DEs of proteins extracted from young kernels of the Recital cv. revealed up to 1200 CB stained proteins. The number of spots detected varied from 800 to 1200 over the whole 18-day period.

There was an increase of 30% in the number of spots between 9 and 12 DAA (Fig. 1). This rapid increase in new proteins was associated with the synthesis of storage proteins. Together with the appearance of new spots, many proteins that were present in the previous sampling stages may disappear. For example at 12 DAA, out of the 1200 spots detected, 483 were new spots (not revealed in the previous sampling during kernel formation) and 433 disappeared at 15 DAA (see Fig. 2)

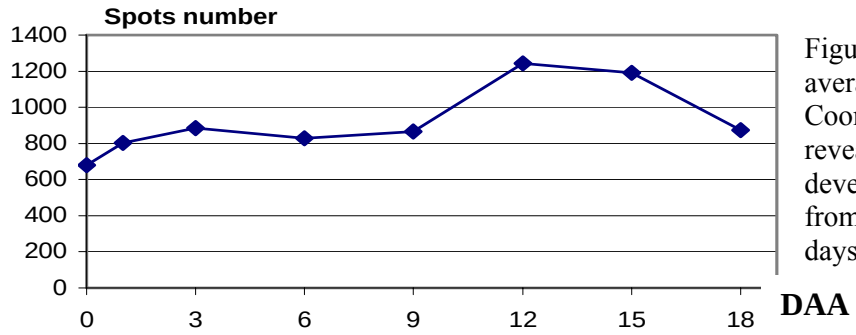


Figure 1 : Variations in the average number of Coomassie stained proteins revealed in 2DE of the developing wheat kernel from the ovule stage to 18 days after anthesis

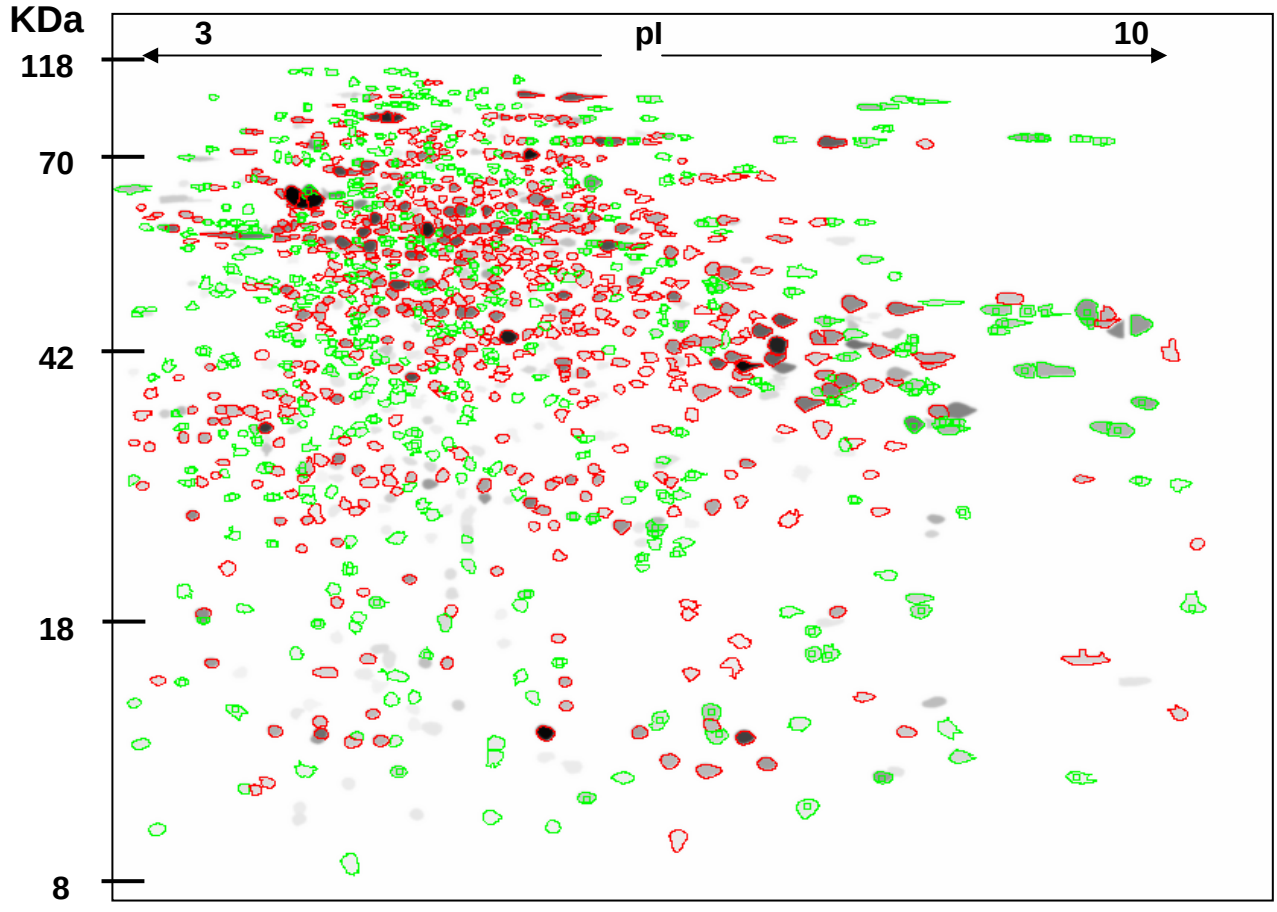


Figure 2 : 2-dimensional electrophoresis (IPGE pH 3-10 x SDS Page) of the kernel proteins at 12 DAA. 1200 spots were Coomassie stained. Red spots appeared at 12 DAA. Green spots are those that were absent at 15 DAA.

Quantitative analysis of spot volume and qualitative analysis (appearance/disappearance of spot) evidenced five types of protein expression. Four were found among the spots that were present at all sampling stages: (1) 78.2 % showed a regular increase in protein content from 1 to 9 DAA followed by a decrease, (2) 11.2 % showed a regular increase in protein content from 1 to 18 DAA, (3) 8.8 % showed a regular increase in protein content from 1 to 9 DAA which then remained constant from 9 to 18 DAA, and (4) in 1.8 %, protein content remained constant from 1 to 9 DAA then increased until 18 DAA.

A fifth type was composed of proteins expressed for a short period of time that were, in our case, new proteins present only in the analysed sample stage and disappeared in the following stage. The major part of this type is probably composed of proteins involved in the regulation of metabolism such as transcription factors but also proteins with a short life in the developing endosperm.

Many major spots were excised from gels then trypsic digested and subjected to MALDI-TOF mass-spectrometry for protein identification. Several spots whose pI and MW were identical to storage proteins were revealed between 9 and 12 DAA. More than 230 other spots were identified either in the SwissProt, TrEMBL and NCBI databases or using the EST database. The successful identification of the majority of the proteins expressed in the developing endosperm will make several approaches possible:

- For the identification of the cascade of protein expression to be linked with the morphological changes that occur in the developing tissues of the wheat kernel.
- For the association between the proteome and the transcriptome, which needs to be investigated in depth to enable progress in the design of micro arrays.
- For the identification of the key kernel enzymes involved in the quantitative accumulation of starch and storage proteins, although several transcription factors have already been found (Xian-Jun Song et al. 2007).

2 Mapping the albumins globulins from mature wheat kernel.

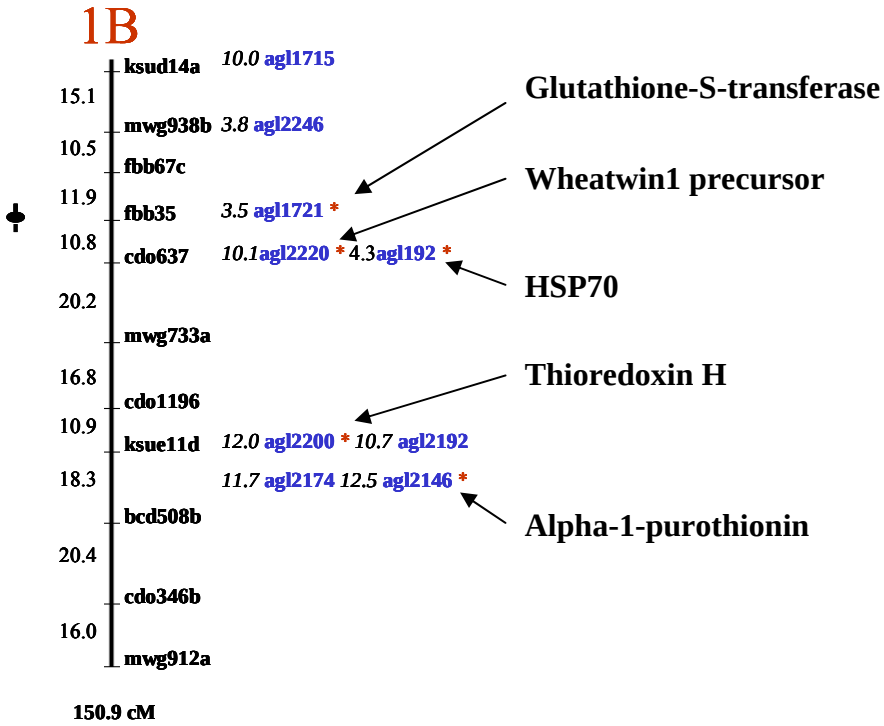
Silver staining of 2DE (IPGE pH 4-7 x SDS Page) of the AG extracted from wholemeal kernels of 'Synthetic' or 'Opata', parents of the ITMI progeny, revealed 2125 and 2163 spots respectively (data not shown). Comparison of the images of the AG of these two cultivars evidenced 226 significantly different spots. These 226 spots were segregating in the ITMI progeny composed of 112 F9 lines, each of which was analysed by running 3 replicates of 2DEs. Among the 226 spots, 120 showed an unbiased segregation and were successfully mapped on 21 chromosomes. The chromosome assignment of each AG spot (between markers anchors) was carried out using the software Mapmaker/exp/version 3.0b at Lod 3.0.

Figure 3 shows the nine AGs assigned on chromosome 1A. The chromosome with the lowest (highest) number of assigned AGs was 1D (5B). All the chromosomes encoded AGs; however, this was not the case of the kernel amphiphilic proteins which were also analysed using 2DE and mapped using the same ITMI progeny (Amiour et al. 2002, 2003). The kernel amphiphilic proteins were mapped on 15 chromosomes; the majority being located on the short arm of group 1 chromosomes, whereas the AGs were located on both the short and the long arm of the different chromosomes. The segregating spots were also subjected to mass spectrometry and their peptide mass fingerprint used for databases interrogation. The proteins mapped on chromosome 1B that were identified are also indicated in Figure 3. Several other proteins mapped on the other chromosomes were also identified.

Mapping the AGs involved more than 350 2DEs being individually analysed using the image analysis software Melanie-3. The proteomic approach is innovative and important for several reasons. It offers the opportunity to:

- Map genes encoding proteins that differ between the two parents;
- Study the association between these AGs and any compositional and morphological characters of the wheat kernel. This progeny has already been used for QTLs analysis of some quality characters (Nelson et al. 2006) but many flour components remain to be analysed;

- Enrich this important international wheat map (the ITMI map) with genes that are biologically active in the kernels of the progeny and may represent new candidate genes to understand differences that have not been investigated up till now.



- Figure 3 : Nine albumins/globulins of the wheat kernel that were segregating in the ITMI progeny were mapped on chromosome 1B. These proteins were identified through MALDI-TOF mass spectrometry and data mining.

3- Wheat Aleurone Layer

Wheat AL is known to contain several vitamins (B1, B2, B3, B5, B6, C) and minerals (iron, manganese, phosphorus) (McMasters et al., 1971, Pomeranz, 1988). A large proportion of the nutritional value of the wheat kernel nutritional value -but also its health benefit- comes from the AL. The different layers of wheat bran have already been characterized using biochemical approaches (Antoine et al. 2004), but the proteins specific to the AL remain to be analysed. Proteomics would be a valuable tool to study the proteins specific to the aleurone layer, which is an important seed tissue. The first proteomics analysis of the AL of barley was performed by Finnie et al., 2003 showing more numerous silver stained proteins than in the endosperm. Our team developed a procedure for hand isolation under the binocular of the AL of mature wheat kernel (Laubin et al. 2007). This procedure enabled us to investigate proteins specific to the AL using proteomics analysis. Wheat AL was isolated from mature seeds of Chinese Spring and Recital cvs. To obtain sufficient extracted proteins to run three replicate 2DEs, a minimum of ten seeds per cultivar were used to isolate the AL. The proteins were separated on 24 cm IPGE pH 3-10 x SDS Page and revealed using CB staining. This procedure enabled more than 700 proteins to be revealed (Figure 4). These numerous proteins are currently being identified in the frame of the European HELTHGRAIN project and an attempt to map the majority of them on wheat chromosomes is underway.

Besides these studies, it was important to see if AL proteomes could differ between cultivars of unrelated pedigree. Surprisingly, near 20% of the AL spots differed between Chinese Spring and Recital. In contrast to the 10.6% of seed albumins–globulins (mostly from

endosperm) that differed between Synthetic and Opata (226 spots out of 2125), it appears that ALs can differ much more between cultivars than previously expected. This finding may be the result of the absence of selection pressure on this tissue composed of living cells, which is not the case of enzymes associated with kernel components such as starch and storage proteins. This high level of genetic difference opens a new door to study the genetic variability of AL components, which could then be further improved by breeding.

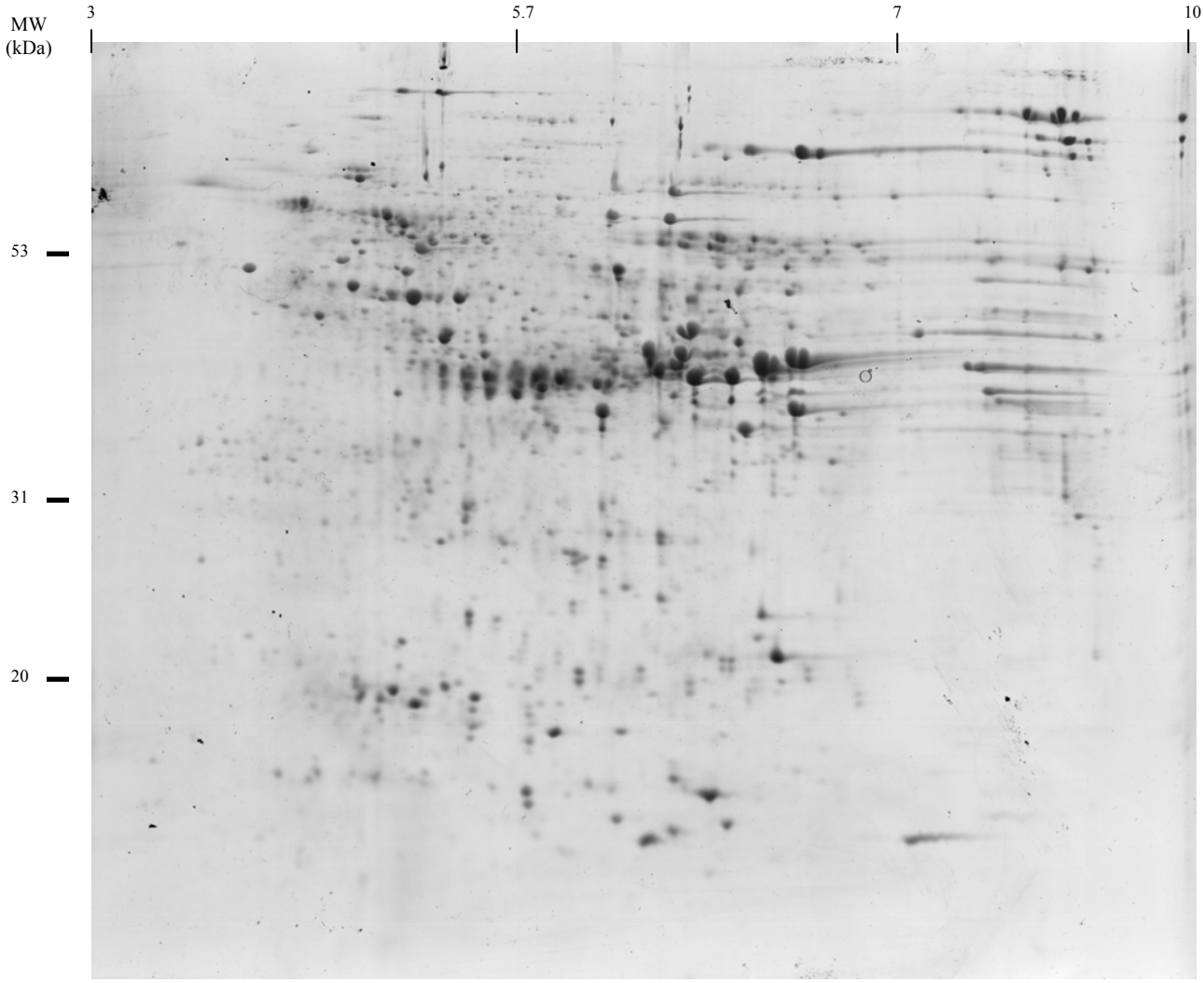


Figure 4 : IPGE 24cm pH 3-10 x SDS Page of the Coomassie blue stained proteins present in the aleurone layer of cv. Chinese Spring. The procedure for AL isolation and 2DE was reported by Laubin et al., 2007

For several obvious reasons the proteomics of the AL is of prime importance for future wheat improvements:

- This part of the kernel is the reservoir of numerous vitamins and minerals whose genetic variability needs to be analysed
- Tools have to be developed to explore the components of AL. Proteomics will provide significant insights into the genetic variability of the enzymes present in the AL and will greatly help us to locate the gene and to design dedicated micro-arrays.
- Such tools will be useful for breeding some proteins in AL as well as its numerous nutritional components. Other components present in the AL such as phytates, or outside the AL, such as arabinoxylanes, lignans, alkylresorcynols, also have to be investigated to understand their

environmental and genetic variability and finally to develop wheat breeding programmes for health value. The proteomics of AL and of the seed peripheral layers will be of great help in this respect.

The proteomics study of the wheat AL is still in its infancy. However, the approach will rapidly provide important data on the proteins specific to this cell layer. Besides the identification of the many proteins and their coding sequences that are specific to the AL, geneticists are interested in understanding the mechanisms involved in the formation of a unique cell aleurone layer in wheat as opposed to barley, which may have two or three aleurone layers. The transcriptomic and proteomic approach of the early stages of kernel formation will provide insights in that point.

Using proteomics for the analysis of the seed will enable biochemists and geneticists to investigate the diversity and function of the proteins expressed in different cell organites such as protein bodies, starch granules or kernel compartments. The analysis of the isolated protein bodies will provide information about the biochemical pathway and genetic factors involved in the accumulation of storage proteins and their polymerisation. The resulting size of the gluten polymers was shown to be the main criterion involved in the stability of bread loaf volume measured in cultivars grown in several locations (Lemelin et al., 2005). The analysis of the gene and of transcription factors leading to the quantitative regulation of the oxidative molecules that are active in the ER, as well as of protein bodies, will provide breeders with the data they need. Beside the protein bodies, starch granules are the other major cell organites that are currently under investigation using proteomics. Several important questions related to the size of starch granules but also to their composition and variation through genetic and environmental influences could be addressed by using proteomics and transcriptomics on developing starch granules. The characterisation of the albumins-globulins, which are located inside the granules and amount to less than 5% of the weight of the starch granules, will be a decisive step in the analysis of this important component. Many other topics will benefit from the proteomics of kernel albumins–globulins. One of the most immediate results of proteomics analysis of albumins-globulins in the endosperm will probably be the identification of allergenic proteins. Many wheat allergens have already been identified (Mills et al., 2003) such as a-amylase/trypsin inhibitor, non-specific lipid transfer protein, seed storage prolamins, but concerning Bakers' asthma, the majority of the allergens fall in the albumins-globulins fraction which remain to be accurately identified. Our proteomics approach could be very useful in achieving this goal.

Acknowledgements
Christophe Chambon who performed MALDI-Tof and MSMS analyses and Clément Debiton for 2 dimensional electrophoresis of proteins and image analysis in developing kernel are gratefully acknowledged.

References

Amiour N., Merlino M., Leroy P., Branlard G.. 2002. Proteomic analysis of amphiphilic proteins of hexaploid wheat kernels. Proteomics, 2: 632-641.

Amiour N., Merlino M., Leroy P., Branlard G. 2003 Chromosome mapping and identification of amphiphilic proteins of hexaploid wheat kernels. Theoretical and Applied Genetics, 108: 62-72.

Antoine C., Peyron S., Lullien-Pellerin V., Abecassis J., Rouau X., 2004. Wheat bran tissue fractionation using biochemical markers. J. Cereal Sci, 39, 387-393.

Bahrman N., Le Gouis J., Negroni L., Amilhat L., Leroy P., Laine A.L., Jaminon O., 2004. Differential protein expression assessed by two-dimensional gel electrophoresis for two wheat varieties grown at four nitrogen levels. Proteomics, 4, 709-719.

Branlard G., Bancel E. 2006. Protein extraction from cereal seeds. In: Thiellement H, Zivy M, Damerval C., Méchin V.E., eds. Plant Proteomics: Methods and Protocols, *Humana Press* Inc, Totowa, NJ, 15-25.

Dumur J., Jahier J., Bancel E., Lauriere M., Bernard M., Branlard G., 2004. Proteomic analysis of aneuploid lines in the homeologous group 1 of the hexaploid wheat cultivar Courtot. Proteomics, 4, 2685-2695.

Dupont F. M., Hurkman W. J., Vensel W. H., Tanaka C., Kothari K. M., Chung O. K., Altenbach S. B., 2006. Protein accumulation and composition in wheat grain: Effects on mineral nutrients and high temperature. Europ. J. Agronomy, 25:96-107.

Finnie C., Svensson B., 2003. Feasability of a tissue specific approach to barley proteome analysis: aleurone layer, endosperm, embryo and single seeds. J. Cereal Sci, 38, 217-227.

Gobaa S., Bancel E., Branlard G., Kleijer G., Stamp P., 2007a. Proteomic analysis of wheat recombinant inbred lines: variations in prolamin and and dough rheology. J. Ceral Sci, in press

Gobaa S., Bancel E., Kleijer G., Stamp P., Branlard G., 2007b Effect of the 1BL.1RS translocation on the wheat grain proteome as revealed by proteomic analysis. Proteomics, Accepted.

Islam N., Woo S.-H., Tsujimoto H., Kawasaki H., Hirano H. 2002. Proteome approaches to characterize seed storage proteins related to ditelocentric chromosomes in common wheat (*Triticum aestivum* L.). Proteomics, 2:1146-1155.

Laubin B., Lullien-Pellerin V., Nadaud I., Branlard G., 2007 2D Electrophoresis of the wheat aleurone layer. J. Cereal Sci, submitted.

Lemelin E., Branlard G., Salvo L., Lein V., Aussenac T., Daydé J. 2005. Breadmaking stability of wheat flours: relation between mixing properties and molecular weight distribution of polymeric glutenins. J.Cereal Sci, 42: 317-326.

Majoul T, Bancel ., Triboi E, Ben Hamida J, Branlard G, 2003. Proteomic analysis of the effect of heat stress on hexaploid wheat grain: Characterization of heat-responsive proteins from total endosperm. Proteomics, 3, 175-183.

Majoul T., Bancel E., Triboï E., Ben Hamida J., Branlard G. 2004. Proteomic analysis of the effect of heat stress on hexaploid wheat grain: Characterization of heat-responsive proteins from non-prolamins fraction. Proteomics, 4: 505-513.

McMasters M. M., Hinton J. J. C., Bradbury D., 1971. Microscopic structure and composition of the wheat kernel. Wheat: chemistry and technology 2nd edition. 51-113.

Mills E. N. C., Jenkins J. A., Griffiths-Jones S., Shewry P. R., 2003 The structural and biological relationships of cereal proteins involved in type I allergy. In “the gluten proteins” (Lafiandra D., Masci S., R D’Ovidio Eds) Proceedings of the 8th Gluten Workshop,Viterbo, Italy, September 8-10, 2003, RSC publisher, 367-370.

Nakamura H., 2001 N-terminal amino acid sequence analysis of endosperm proteins in Japanese hexaploïd wheat. Cereal Chem, 78 : 79-83.

Nelson J. C., Andreescu C., Breseghello F., Finney P. L., Gualberto D. G., Bergman C. J., Peňa R. J., Perretant M. R., Leroy P., Qualset C. O., Sorrells M. E., 2006. Quantitative trait locus analysis of wheat quality traits. Euphytica, 149, 145-159.

Pomeranz Y., 1988 Chemical composition of kernel structures. Wheat: chemistry and technology 3rd edition. p. 108.

Rabilloud T., 2002. Two-dimensional gel electrophoresis in proteomics: Old, old fashioned, but it still climbs up the mountains Proteomics, 2, 3-10.

SAS Institute Inc. SAS/STAT User's Guide: Statistics, Version 5 Edition (1985), Cary, NC, U.S.A.

Skylas D. J., Mackintosh J., Cordwell S. J., Basseal D. J., Walsh B. J., Harry J, Blumenthal C., Copeland L., Wrigley C. W., Rathmell W., 2002. Proteome approach to the characterization of protein composition in the developing and mature wheat grain endosperm. J. Cereal Sci, 32: 169-188.

Skylas D. J., Cordwell S. J., Hains P. G., Larsen M. R., Basseal D. J., Walsh B. J., Blumenthal C., Rathmell W., Copeland L., Wrigley C. W., 2002. Heat Shock of Wheat During Grain Filling: Proteins Associated with Heat-tolerance. J. Cereal Sci, 35: 175-188.

Thiellement H., Zivy M., Damerval C., Méchin V. E. 2006. Plant Proteomics: Methods and Protocols, Humana Press Inc, Totowa, NJ, 1- 399.

Vensel W. H., Tanaka C. K., Cai N., Wong J. H., Buchanan B. B., Hurkman W. J., 2005 Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics, 5, 1594-1611.

Xian-Jun Song, Wei Huang, Min Shi, Mei-Zhen Zhu, Hong-Xuan Lin. 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase Nature Genetics, 39: 623-230.