2008

A PROTEOMIC STUDY OF SOYBEAN SEED COATS (Glycine max)

M Carmen Romero
Western University

A PROTEOMIC STUDY OF SOYBEAN SEED COATS (*Glycine max*)

(Spine title: Soybean seed coat proteome)

(Thesis format: Integrated-Article)

by

M. Carmen Romero

Graduate Program
In
Biology

2

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

School of Graduate and Post Doctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Supervisor

Dr. Daniel Brown

Co-Supervisor

Dr. Mark Gijzen

Supervisory Committee

Dr. Mark Bernards
Dr. Susanne Kohalmi
Dr. Gilles Lajoie

Examiners

Dr. Mark Bernards

Dr. David Edgell

Dr. Brian Miki

Dr. Yolanda Morbey

The thesis by

M. Carmen Romero

entitled:

Proteomic Study of Soybean Seed Coats *(Glycine max)*

is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date September 26, 2008

Chair of the Thesis Examination Board

ABSTRACT

The seed coat arises from maternal integuments and its main role is the protection of the developing embryo. At maturity the soybean seed coat is composed mainly of dead cells that impart protection, enable germination and enhance dispersal. Soybean seed coat is an agricultural-by product that is currently under utilized, although some reports have found diverse applications that are currently explored.

Using gel-based pre fractionation of proteins followed by electro spray ionization tandem mass spectrometry, a comprehensive proteomic database of physiologically mature soybean seed coats was created. Around 150,000 spectral was acquired and used to challenge current protein databases. The gene ontology assignment of over 1,000 seed coat proteins allowed a correlation with important seed metabolic pathways such as cell wall biosynthesis, proteolytic pathway, synthesis of amino acids, carbohydrates and nucleotides thorough the C_1 metabolic pathway, fatty acids and isoflavonoids.

The most abundant protein in the soybean seed coat is methionine synthase. Besides the synthesis of methionine, it could be associated with the production of ethylene in the seed coat, promoting fruit ripening. There is an apparent increase in the relative amount of metabolic proteins at the onset of seed maturation. This finding suggests that the seed coat remains metabolically active for a longer time than the embryo. Proteases are an important protein group in the seed coat proteome and are most likely involved in tissue remodelling. They could also be further studied as potential candidates for catalysis of industrial reactions.

A comprehensive protein database is reported along with protein ontology assignments. The correlation of proteomic and transcriptomic data for specific proteins allowed the identification of control mechanisms of protein expression in the seed coat. This will be very valuable in future molecular-based approaches to

modify the seed coat proteome in order to control aspects of seed development, and also as a target organ for the heterologous expression of proteins.

Key words: BLAST, cell wall, electrospray ionization, fatty acid synthesis, isoflavonoids, proteins, proteome, proteases, seed coat, seed development, tandem mass spectrometry, two-dimensional electrophoresis.

Para Camila

ACKNOWLEDGEMENTS

This work was possible thanks to the support and trust of my supervisor Dr. Daniel Brown and the guidance of Dr. Mark Gijzen, for which I am indebt. I would like to thank my advisors, Dr. Mark Bernards and Dr. Susanne Kohalmi for their constructive comments and suggestions during my graduate studies at the Biology Department, at UWO. I would like to thank Dr. Gilles Lajoie for taking me under his wing to teach me proteomics.

I need to thank Kuflom Kuflu for his patience to teach the protein basics. Tara Rintoul for reading my thesis and helpful comments. Huaiyu Wang for assistance with the statistics. Marysia Latoszek-Green for helping in the logistics in the lab. In the Lajoie lab, I am indebt to Paula Pittock, Sean Bendall and Cunjie Zhang for endless help, cheerful attitude and friendship.

To my sister and brother, my inspiration. To my dear husband Ruben, for his unconditional support and patience, lots of patience. I love you Ruben.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABREVIATIONS

2 DE	two-dimensional electrophoresis
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
ACN	acetonitrile
Ah	*Arachis hypogea* (peanut)
At	*Arabidopsis thaliana* (thale cress)
BLAST	basic local alignment search tool
BPB	bromophenol blue
CHAPS	3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate
DPA	days post anthesis
DTT	dithiothreitol
ESI	electrospray ionization
EST	expressed sequence tag
FA	formic acid
Gm	*Gycine max* (soybean)
Hv	*Hordeum vulgare* (barley)
IAA	indoleacetic acid
IEF	isoelectrofocusing
IPG	immobiline dry strip gel
LC	liquid chromatography
Le	*Lycopersicum esculentum* (tomato)
MALDI	laser desoprtion/ionization
MS/MS	tandem mass spectrometry
Nt	*Nicotiana tobacum* (tobacco)
Os	*Oryza sativa* (rice)
Ps	*Pisum sativum* (pea)
SDS	sodium dodecyl sulfate
SPI	spectral peak intensity
St	*Soluanum tuberosum* (potato)
Ta	*Triticum aestivum* (wheat)

TCA	trichloroacetic acid
Tris HCl	Tris (hydroxymethyl) aminomethane hydrochloride
Zm	*Zea mays* (maize)

Chapter 1

INTRODUCTION

The seed is essential to flowering plant reproduction because it protects and nourishes the developing embryo that represents the next generation. Seed development is triggered by a novel double fertilization process that leads to the differentiation of the embryo, endosperm, and the seed coat, which are the major compartments of the seed (Miller et al., 1999; Haughn and Chaudhury, 2005; Moise et al., 2005). Each compartment has different origins and specialized roles. The maternally-derived seed coat differentiates from the ovule integument that surrounds the embryo sac in what is regarded as the most dramatic cellular changes observed during seed development. It is the main protective structure for the embryo, transferring nutrients from the maternal plant to the developing embryo (Murray, 1987; Schuurmans *et al.*, 2003; Borisjuk et al., 2004; Zhang et al., 2007). At maturity, the seed coat is composed mainly of dead cells, and even in death, the specialized cell types impart protection, enable germination and enhance seed dispersal (Haughn and Chaudhury, 2005).

The embryo and endosperm, on the other hand, are derived from the fertilized egg and central cell, respectively. The endosperm proliferates to occupy most of the post-fertilization embryo sac and nourishes the embryo during early development (Lopes and Larkins, 1993). In legumes, the endosperm is absorbed by the embryo during seed development and some remnants are present in the seed (Le et al., 2007). Soybean is virtually devoid of endosperm with some traces in the aleurone layer (Yaklich et al., 1984); whereas, *Medicago truncatula* possesses a more structured endospermic layer at maturity (Lei et al., 2007). The embryo represents the new sporophytic (diploid) generation and contains the shoot and root meristems that are responsible for generating organ systems of the mature plant after seed germination.

In legumes, food reserves stored in the embryonic cotyledons make seeds an important food source for both human and animal consumption. Soybean is one of the most important seed crops in the world (Wilcox, 2004) and constitutes

about one third of the world supply of vegetable oil, most of which is used for food and cooking (Kinney, 1998). It provides an inexpensive source of protein as the main ingredient of animal diet, with as much as 90% of the soybean production used to feed livestock worldwide (Steinfeld and Wassenaar, 2007).

In recent years, researchers have focused on the use of crop plants for the production of oil to power engines, in an effort to reduce pressure on the use of fossil energy (Doll et al., 2008). Consequently, the consumption of canola, soybean, palm and other oil crops for biodiesel has increased. The 'net energy balance ratio' for biodiesel from soybean is three to four times more favorable than for ethanol from maize (Hill, 2007). In the USA, it is estimated that approximately 22% of domestic soybean oil production by 2016 will be devoted to biodiesel (Durrett *et al.*, 2008). This production process normally leaves the hulls or seed coats as unutilized by-products, which represent up to 10% of the total seed mass.

1.1 Seed coat anatomy

In legume seeds the seed coat is a complex organ comprised of a palisade or epidermis layer, hypodermis or hourglass cell layer, parenchyma of maternal origin and aleurone with some endosperm debris of filial origin (Yaklich et al., 1998) (Figure 1.1). Several extensive studies have described the anatomy of soybean seed coats (Thorne, 1981; Yaklich et al., 1984; Ma et al., 2004) and, in this section, major aspects will be brought into consideration as we will use them as a guide to the dissection of the proteome of this organ.

The epidermis (palisade) is a layer of tightly packed, elongated cells that have pitted walls in the upper part of the cell. By maturity these cells are fully cutinized providing a strong, gas-impermeable surface (Thorne, 1981; Ma et al., 2004). The hypodermis is a layer of hourglass-shaped cells that have unevenly thickened cell walls, thin at the ends of the cell and very thick in the central, constricted portion, thus forming a strong supporting layer with considerable intercellular space (Figure 1.1B). The prevailing characteristic of both layers, as of any sclereid-type cell, is the presence of extremely thickened cell walls most

likely to impart physical resistance to these tissues and provide protection to the embryo (Thorne, 1981).

Two types of tissue compose the parenchyma in the seed coat. Adjacent and below the hourglass layer, lays the articulated parenchyma with large and irregularly shaped cells with thick cell walls. Abundant plasmodesmata span the thin walls at points of cellular interconnection and there are extensive intercellular spaces. These cells have been reported to possess a dense cytoplasm enriched in constituents characteristic of cells engaged in active carbohydrate transport and excretion (Thorne, 1981). The reticulate venation that originates in two large vascular bundles in the pod placenta is imbedded within the narrow zone separating the two distinctly different parenchyma tissues in the seed coat. It is composed of small, thick-walled sieve tubes surrounded by a bundle sheath of small vascular parenchyma cells which are abundant in plasmodesmata. Xylem tissue is absent from the seed coat.

Below this level of seed coat vascularization, the parenchyma is composed of 10 to 15 layers of thin-walled aerenchyma cells, all interconnected to form a three-dimensional lattice. The cytoplasm of these aerenchyma cells is almost devoid of organelles. This lattice appears to form a continuous apoplastic route from the vascular plane to the inner seed coat surface. These aerenchyma cells are easily observed because their interconnections are delicate, and this is where a natural fault line facilitating dissection occurs (Figure 1.1C) at the junction of the aerenchyma and the endothelium or aleurone. No pores, plasmodesmata or vascular tissue exist to carry photosynthates to the embryo from the surrounding endothelium and there is no vascular communication between the seed coat and the embryo (Thorne, 1981).

The aleurone cells represent a special tissue of endospermic origin at the interphace between seed coat and embryo. In the *Glycine* genus some of the endosperm remains unabsorbed during seed development and occurs as the antipit adhering to aleurone (Yaklich et al., 1992). The aleurone is formed by a single layer of small, thick-walled cells that are distinguishable from parenchyma

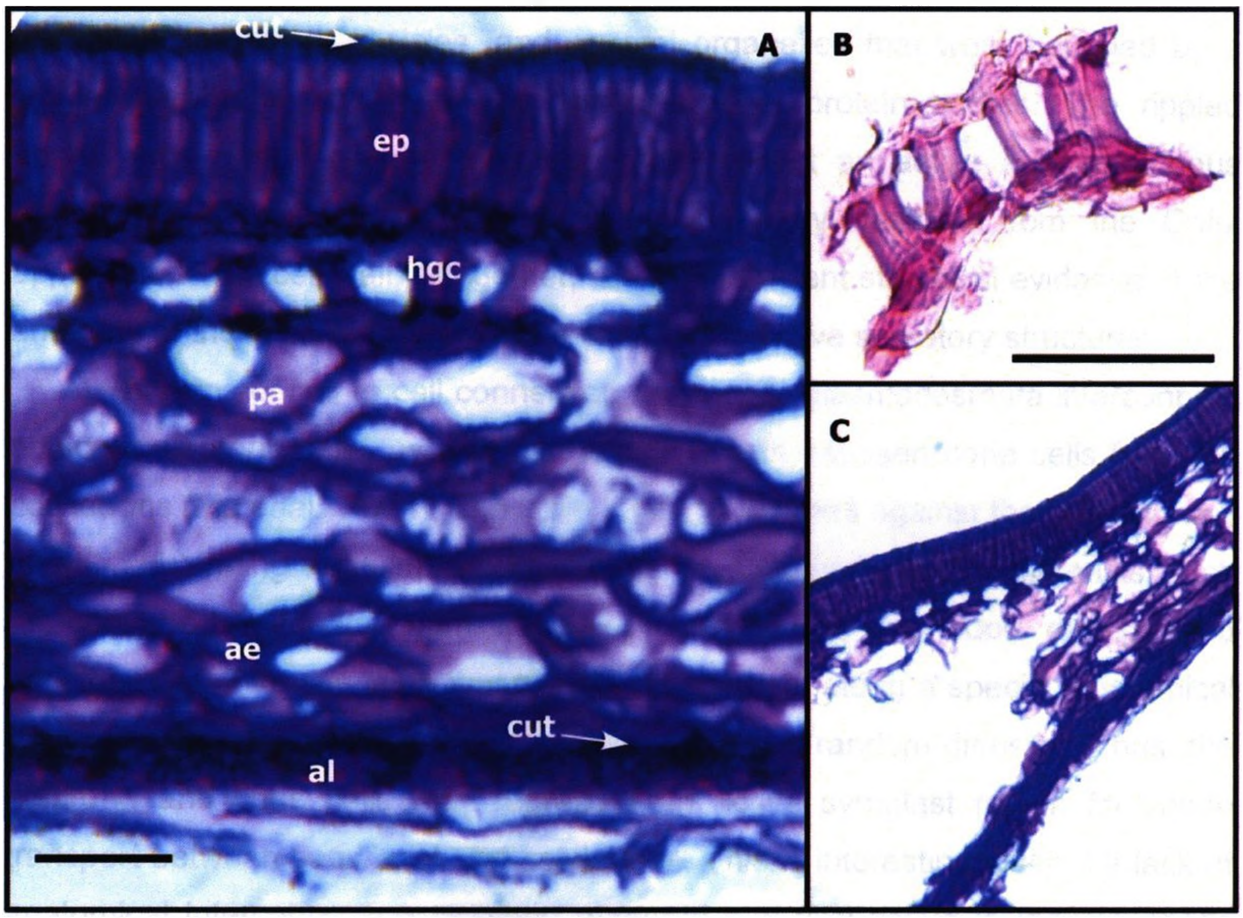


Fig. 1.1 Microphotographs of soybean seed coats. A) Agarose 40 µ inverted section of 70 DPA soybean seed coat (40X). B) Hourglass cells (hgc) isolated from epidermal layer (40X). C) Tissue separation along a natural plane of weakness between the palisade-hourglass and the parenchyma-aleurone layers (20X). ae, aerenchyma; al, aleurone; cut, cuticle; ep, epidermis (palisade); pa, parenchyma. Scale bar = 100 µm.

by their dense cytoplasm and small cuboidal shape. The cytoplasm contains numerous dilated cisternae of rough endoplasmic reticulum, Golgi apparatus with associated secretory vesicles, and several organelles that were bounded by a single membrane and have the appearance of protein bodies. The rippled appearance of the plasma membrane may reflect an active and continuous transfer of membrane and contents by secretory vesicles from the Golgi apparatus to the cell wall. Altogether, there is sufficient structural evidence in the anatomy of these cells for them to be considered active secretory structures.

In terms of cell to cell connection, numerous plasmodesmata interconnect adjacent aleurone cells, as well as the cross walls between cone cells from the cotyledons (Ma et al., 2004). The cone cells tightly press against the cotyledonary pit and subtending vascular bundle, suggesting that movement of materials, such as converted nutrients, from the seed coat to the cotyledon during seed development. This suggests that transport can occur along a specific anatomical route and that this material is not subject entirely to random diffusion. Thus, the aleurone and cone cells may represent an active symplast region for solute transport between seed coat and cotyledons. This is interesting given the lack of anatomical interconnection between maternal and filial tissue in legume seeds providing access to plasma membrane transport events (Patrick and Offler, 1995; Weber et al., 1997a).

1.2 Seed coat function

The mature seed coat has been classically viewed as a protective structure with limited physiological functions. However, investigations have concentrated on the transport and metabolism of carbohydrate and amino acid compounds from the parent plant to the developing seed (Thorne, 1985; Tegeder et al., 1999; Tegeder et al., 2000; Tilsner et al., 2005). These studies have indicated that the pod wall and the seed coat were involved in converting carbohydrate and amino acid compounds into forms readily usable by the embryo. It was found that in legume seeds, some storage precursors undergo

processing in the seed coat before they are transported to the embryo. For example, ureides are converted in the seed coat into transportable amino acids (Hsu et al., 1984); whereas, the regulation of sucrose transport in the seed coat and subtending cotyledons occurs through the apoplast of both organs (Wang et al., 1995; Weber et al., 1997b; Ritchie et al., 2003; Zhang et al., 2007).

During legume seed development the seed coat and endosperm develop first, followed by the development of the embryo, maturation of the seed coat and maturation of the embryo (Weber et al., 2005). The coordination of these events is governed by communication among tissues of the seed organs. Early embryo development and differentiation is controlled by maternal tissues; therefore, signals must be transmitted through the seed coat and endosperm before they can reach the embryo. Specialized transfer cells facilitate the transport of nutrients within the seed (Offler and Patrick, 1993). Sufficient published literature exists to support the notion that maternal control of embryo development is exerted through sugar metabolism and metabolic gradients in legume seeds (Weber et al., 1996; Weber et al., 1997b; Weber et al., 1997a; Weber et al., 1998; Wobus and Weber, 1999a, 1999b). The effect of phytohormones (Bleecker and Kende, 2000), hypoxia (Rolletschek *et al.*, 2005) and carbon dioxide recycling (Furbank et al., 2004) has been well documented. Cross talk among various pathways must play a major role in the control of seed development and most likely there are differences among species due in part to varying morphology and structure. In *Arabidopsis* it was determined that the maternal control of seed coat elongation and the zygotic control of endosperm growth are coordinated to determine seed size (Garcia et al., 2005).

The maternal seed coat, filial endosperm and embryo interact physically. In pea seeds, the general pattern of seed development appears largely determined by the maternal parent; that is, the seed and final seed size is positively correlated with the maximum volume of the endosperm (Wang and Hedley, 1993). Large-seeded genotypes of *Vicia faba* develop a larger seed resulting in a longer cell division period of the embryo (Weber et al., 1996). This

is in accordance with the observation that cell number in cotyledons is correlated with seed size, which is predominantly maternally determined (Davies, 1975). Because the cell number is determined by cell division cycles during early growth, control of cell division is crucial. However, seed size control is complex. Studies in small- and large-seeded genotypes of *Vicia faba* suggests that the seed coat derived metabolic signals are critical (Weber et al., 1996).

Genetic approaches have been traditionally utilized when studying legume seed development. More recently, the use of large-scale proteomic studies has been useful to elucidate general trends of protein expression during plant and seed development (Ruuska et al., 2002; Hajduch *et al.,* 2005; Hajduch et al., 2006a; Gallardo et al., 2007; Hajduch et al., 2007). An in depth study of the soybean seed coat proteome has not yet been reported. Given its biological relevance and the economic importance of soybean, a proteomic analysis of soybean seed coats would be very useful to determine biotechnological approaches to increase crop value and improve agronomic traits. From the basic research point of view, such a study would provide a comprehensive understanding of global protein expression and metabolic pathways prevalent in this organ.

1.3 Proteomic method approach

The proteome is the full complement of proteins expressed by a genome (Wasinger et al., 1995) at a specific point in time. Proteomics is the systematic analysis of proteins and peptides that are encoded by a genetic code, and provides a link between cell physiology and the genetic code. The objectives of proteomics include large-scale identification and quantification of all protein types in a cell or tissue, analysis of post-translational modification and association with other protein, and characterization of protein activities and structures (Rhee et al., 2006). Application of proteomics in plants is still in its initial phase, mostly in protein identification (Canovas et al., 2004; Newton et al., 2004). Other aspects of proteomics (reviewed in Zhu et al., 2003), such as identification and prediction of protein-protein interactions, protein activity

profiling, protein subcellular localization and protein structure, have not been widely used in plant science. However, recent efforts such as the structural genomic initiative that includes *Arabidopsis thaliana* are certainly encouraging (http://www.uwstructuralgenomics.org).

The genome of model-plants such as *A. thaliana* (Arabidopsis Genome Initiative, 2000) and rice (*Oryza sativa*) (Goff et al., 2002) have now been elucidated. The NCBI Plant Genomes Central considers also *Medicago truncatula* (barrel medic) and *Populus trichocarpa* (black cottonwood) as completed (Carpentier et al., 2008) and several others are in the process of being sequenced (http://www.genomesonline.org). In the case of soybean, the entire genome sequencing was completed and made available to the scientific community on January 18, 2008 (http://phytozome.net/soybean); however, the information is still preliminary and unsuitable for protein searches, which is an important aspect of our study.

While identification of genomes has been, and continues to be, a technically and intellectually demanding process, the identification of the proteome contains inherently greater difficulties. The first major difference between genome and proteome analysis is that genome is static, while the proteome of each living cell is dynamic, altering in response to the individual's cell metabolic state and reception of intracellular and extracellular signal molecules. Thus while the genome enables a prediction of the proteome simply as the gene products, this cannot be described as the proteome, since it remains unknown which genes are expressed at any specific moment in time, and many of the proteins which are expressed will be post-translationally altered, by one or more of approximately 200 modifications (Mann and Jensen, 2003; Newton et al., 2004). Despite the differences in the nature of genome and proteome, genetic characterization of an organism is required for the massive identification of proteins from its complement.

Once the genomic data is acquired, it is stored in large databases containing the nucleotide sequence code and gene annotations and provides basic foundations for studying biological systems (Ivakhno and Kornelyuk, 2006). However, sequence information alone is insufficient for understanding the biology of a given organism. Data on mRNA expression, protein interaction, protein localization, and dynamics of signaling pathways is needed before a full appreciation of complexity of living cells is grasped. The power of high-throughput approaches in functional genomics is exemplified by DNA microarray technologies. However, because proteins are the predominant functional macromolecules, the identity of potentially expressed proteins at a given time defines the functional state of the cell. Since significant molecular control is exercised at the level of translation initiation, post translational modifications, and mRNA turnover, the investigation of proteome dynamics is a vital requirement for understanding of the cell's regulatory mechanism.

When only model organisms are used, the power of transcript-based techniques is lost in non-model organisms due to the lack of genomic information or due to the sequence divergence from a related model organism. Gene sequences are rarely identical from one species to another and orthologous genes are usually riddled with nucleotide substitutions. An alternative for examining gene expression is studying its end products, the proteins. Protein sequences are more conserved making the high-throughput identification of non-model gene products by comparison to well known orthologous proteins quite efficient (Liska and Shevchenko, 2003).

Seeds of several species have been investigated at the proteome level, such is the case of seeds of the model species *Arabidopsis thaliana* (Ruuska et al., 2002), *Medicago* truncatula (Gallardo et al., 2003; Gallardo et al., 2007), Brasica napus (Hajduch et al., 2006b), soybeans (Herman et al., 2003; Hajduch et al., 2005; Agrawal and Thelen, 2006). In every case, complicated networks have been reported to orchestrate seed development, such as the onset of cell division followed by storage product accumulation and desiccation. However,

many of the fundamental questions remain unanswered and the extent to which comparison can be established between species becomes limited, given the observations that these networks can differ considerably.

The standard approach for proteomic analysis is the in-gel separation of proteins followed by mass spectrometry. Gel-based pre fractionation of proteins allows the separation of complex protein mixtures (Granvogl et al., 2007). Two-dimensional electrophoresis (2 DE), although developed over 50 years ago, continues to be relevant and useful in the separation of protein complexes (Aebersold and Mann, 2003; Gorg et al., 2004). Also, the use of one-dimensional gel electrophoresis of proteins offers a number of important advantages compared to gel-free approaches (Shevchenko et al., 2006). For instance, the sequencing of sharp, molecular weight-separated protein bands increases the dynamic range of analysis of protein mixtures (ratio of lowest to highest abundance protein detectable) as peptides produced by in-gel tryptic cleavage of each band are sequenced in separate experiments. For complex mixture analysis, spreading the proteome over 10-20 gel slices dramatically increases the depth of analysis, and hence the number of identified proteins. The same idea of pre-fractionation holds valid for 2D gels, only that the number of samples are increased several times.

After protein separation using gel electrophoresis and protein digestion using an enzyme (e.g., trypsin, pepsin), proteins are identified by typically using mass spectrometry (MS). Mass spectrometric measurements are carried out in the gas phase on ionized analytes. By definition, a mass spectrometer consists of an ion source, a mass analyzer that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that registers the number of ions at each m/z value. Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are the two techniques most commonly used to volatize and ionize the proteins or peptides for mass spectrometric analysis (Aebersold and Mann, 2003). ESI ionized the analytes out of a solution and is therefore readily coupled to liquid-based (chromatographic) separation tools. ESI-MS systems (LC-MS) are the methods of choice to analyze complex samples.

The high-throughput data generated from mass spectrometers are often complicated and computational analyses are critical in interpreting the data for protein identification (Gorg et al., 2005). Tandem mass spectrometry (MS/MS) breaks each digested peptide into smaller fragments, whose spectra provide effective signatures of individual amino acids in the peptide for protein identification. Many tools have been developed for MS/MS-based peptide/protein identification that rely on the comparison between theoretical peptides derived from database and experimental mass spectrometric tandem spectra. In general terms, protein identification from mass spectra (Veljanovski et al., 2006) can be considered as being straightforward for plant species whose genome have been sequenced or with a considerable number of ESTs available in either general (UniProt, Swiss-Prot, NCBI) or plant specific databases (Basu et al., 2006; Xu et al., 2006). In parallel, the development of bioinformatic tools and specific algorithms permits data integration, modeling and prediction (Rhee et al., 2006). The opposite situation is encountered when dealing with proteomic analysis from non-model plants or with poorly characterized genomes, such as oak (Navarro et al., 2006) and banana (Samyn et al., 2007). In such cases, sequence databases from closely related species are interrogated by de novo sequencing and/or basic local alignment search tool (BLAST) (Altschul et al., 1997) similarity searching. BLAST results are difficult to score and require a large amount of manual validation. However, BLAST remains as a suitable option for cases such as soybean, with predicted proteome data not yet available.

The de novo sequencing approach based on MS/MS spectra is an active research area (Carpentier et al., 2008). Typically the algorithms match the separations of peaks by the mass of one or several amino acids and infer the probable peptide sequences that are consistent with the matched amino acids (Chen et al., 2001). There are some popular software packages for peptide de novo sequencing using MS/MS data such as PEAKS (Ma et al., 2003)(http://www.bioinformaticssolutions.com/products/peaks). One limitation of current de novo methods is that they often cannot provide the exact sequence of

peptide. Instead, several top candidate sequences are suggested (Rhee et al., 2006).

1.4 Research goal and objectives

The objectives of the present research center on the identification and relative quantification of the most abundant proteins present in physiologically mature soybean seed coats. The long-term goal is to identify the processes involved in seed coat development that could potentially be modified in order to improve agronomic traits of soybean seeds, as well as the comprehensively assess the effects of introduced genetic modifications on the seed coat proteome. This latter objective is expected to be of critical importance in the future as new compliance regulations demand information on intended and non-target gene expression of new cultivars being considered for commercial release; making the basic proteomic information of a target organ a requirement.

The first objective of my research is the identification of seed coat proteins in order to create the proteomic database that permits the connection of functional classes with mainstream biosynthetic and physiological pathways known to occur at cellular level.

The second objective is to determine protein expression trends and relative amounts of the most abundant proteins expressed during seed coat development, in order to understand the nature of the changes and potentially determine strategies for seed coat proteome manipulation.

These objectives will be targeted with the use of resources and scientific methodology presented in Chapters 2 and 3 of this document.

1.5 References

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature **422,** 198-207.

Agrawal, G.K., and Thelen, J.J. (2006). Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. Mol Cell Proteomics **5,** 2044-2059.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25,** 3389-3402.

Basu, U., Francis, J.L., Whittal, R.M., Stephens, J.L., Wang, Y., Zaiane, O.R., Goebel, R., Muench, D.G., Good, A.G., and Taylor, G.J. (2006). Extracellular proteomes of Arabidopsis thaliana and Brassica napus roots: analysis and comparison by MudPIT and LC-MS/MS. Plant And Soil **286,** 357-376.

Bleecker, A.B., and Kende, H. (2000). Ethylene: A gaseous signal molecule in plants. Annu Rev Cell Dev Bi **16,** 1-18.

Borisjuk, L., Rolletschek, H., Radchuk, R., Weschke, W., Wobus, U., and Weber, H. (2004). Seed development and differentiation: A role for metabolic regulation. Plant Biol. **6,** 375-386.

Canovas, F.M., Dumas-Gaudot, E., Recorbet, G., Jorrin, J., Mock, H.P., and Rossignol, M. (2004). Plant proteome analysis. Proteomics **4,** 285-298.

Carpentier, S.C., Panis, B., Vertommen, A., Swennen, R., Sergeant, K., Renaut, J., Laukens, K., Witters, E., Samyn, B., and Devreese, B. (2008). Proteome analysis of non-model plants: A challenging but powerful approach. Mass Spectrometry Reviews **27,** 354-377.

Chen, T., Kao, M.Y., Tepel, M., Rush, J., and Church, G.M. (2001). A dynamic programming approach to de novo peptide sequencing via tandem mass spectrometry. Journal Of Computational Biology **8,** 325-337.

Davies, D.R. (1975). Studies of seed development in Pisum sativum I. Seed size in reciprocal crosses. Planta **124,** 297-301.

Doll, K.A., Sharma, B.K., Suarez, P.A.Z., and Erhan, S.Z. (2008). Comparing biofuels obtained from pyrolysis, of soybean oil or soapstock, with traditional soybean biodiesel: Density, kinematic viscosity, and surface tensions. Energy & Fuels **22,** 2061-2066.

Durrett, T.P., Benning, C., and Ohlrogge, J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. Plant J **54,** 593-607.

Furbank, R.T., White, R., Palta, J.A., and Turner, N.C. (2004). Internal recycling of respiratory CO2 in pods of chickpea (Cicer arietinum L.): the role of pod wall, seed coat, and embryo. J Exp Bot **55,** 1687-1696.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., and Burstin, J. (2003). Proteomics of Medicago truncatula seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol **133,** 664-682.

Gallardo, K., Firnhaber, C., Zuber, H., Hericher, D., Belghazi, M., Henry, C., Kuster, H., and Thompson, R. (2007). A combined proteome and

transcriptome analysis of developing Medicago truncatula seeds. Mol Cell Proteomics **6**, 2165-2179.

Garcia, D., Fitz Gerald, J., and Berger, F. (2005). Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in Arabidopsis. Plant Cell **17**, 52-60.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R.L., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchinson, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J.P., Miguel, T., Paszkowski, U., Zhang, S.P., Colbert, M., Sun, W.L., Chen, L.L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y.S., 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., and Briggs, S. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp japonica). Science **296**, 92-100.

Gorg, A., Weiss, W., and Dunn, M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. Proteomics **4**, 3665-3685.

Gorg, A., Weiss, W., and Dunn, M.J. (2005). Current two dimensional electrophoresis technology for proteomics (vol 4, pg 3665, 2004). Proteomics **5**, 826-827.

Granvogl, B., Ploscher, M., and Eichacker, L.A. (2007). Sample preparation by in-gel digestion for mass spectrometry-based proteomics. Analytical And Bioanalytical Chemistry **389**, 991-1002.

Hajduch, M., Ganapathy, A., Stein, J.W., and Thelen, J.J. (2005). A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol **137**, 1397-1419.

Hajduch, M., Casteel, J.E., Hurrelmeyer, K.E., Song, Z., Agrawal, G.K., and Thelen, J.J. (2006a). Proteomic analysis of seed filling in Brassica napus. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. (vol 141, pg 32, 2006). Plant Physiol **141**, 1159-1159.

Hajduch, M., Casteel, J.E., Hurrelmeyer, K.E., Song, Z., Agrawal, G.K., and Thelen, J.J. (2006b). Proteomic analysis of seed filling in Brassica napus. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol **141**, 32-46.

Hajduch, M., Casteel, J.E., Tang, S.X., Hearne, L.B., Knapp, S., and Thelen, J.J. (2007). Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. Journal Of Proteome Research **6**, 3232-3241.

Haughn, G., and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. Trends in Plant Science **10**, 472-477.

Herman, E.M., Helm, R.M., Jung, R., and Kinney, A.J. (2003). Genetic modification removes an immunodominant allergen from soybean. Plant Physiol **132**, 36-43.

Hill, J. (2007). Environmental costs and benefits of transportation biofuel production from food- and lignocellulose-based energy crops. A review. Agronomy for Sustainable Development **27,** 1-12.

Hsu, F.C., Bennett, A.B., and Spanswick, R.M. (1984). Concentrations of sucrose and nitrogenous compounds in the apoplast of developing soybean seed coats and embryos. Plant Physiol **75,** 181-186.

Ivakhno, S., and Kornelyuk, A. (2006). Quantitative proteomics and its applications for systems biology. Biochemistry-Moscow **71,** 1060-1072.

Kinney, A.J. (1998). Plants as industrial chemical factories - new oils from genetically engineered soybeans. Fett-Lipid **100,** 173-176.

Le, B.H., Wagmaister, J.A., Kawashima, T., Bui, A.Q., Harada, J.J., and Goldberg, R.B. (2007). Using genomics to study legume seed development. Plant Physiol **144,** 562-574.

Lei, Z., Nagaraj, S., Watson, B.S., and Sumner, L.W. (2007). Proteomics of Medicago truncatula. Plant Proteomics, 121-136.

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

Lopes, M.A., and Larkins, B.A. (1993). Endosperm origin, development and function. Plant Cell **5,** 1383-1309.

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

Ma, F.S., Cholewa, E., Mohamed, T., Peterson, C.A., and Gijzen, M. (2004). Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. Ann Bot **94,** 213-228.

Mann, M., and Jensen, O.N. (2003). Proteomic analysis of post-translational modifications. Nat Biotechnol **21,** 255-261.

Miller, S.S., Bowman, L.A.A., Gijzen, M., and Miki, B.L.A. (1999). Early development of the seed coat of soybean (Glycine max). Ann Bot **84,** 297-304.

Moise, J.A., Han, S., Gudynaite-Savitch, L., Johnson, D.A., and Miki, B.L.A. (2005). Seed coats: Structure, development, composition, and biotechnology. In Vitro Cellular & Developmental Biology-Plant **41,** 620-644.

Murray, D.R. (1987). Nutritive Role of Seedcoats in Developing Legume Seeds. Am J Bot **74,** 1122-1137.

Navarro, J.I., Lenz, R.M., Ariza, D., and Jorrin, J.V. (2006). Variation in the holm oak leaf proteome at different plant developmenta stages, between provenances and in response to drought stress. Proteomics **6,** S206-S214.

Newton, R.P., Brenton, A.G., Smith, C.J., and Dudley, E. (2004). Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. Phytochemistry **65,** 1449-1485.

Offler, C.E., and Patrick, J.W. (1993). Pathway of photosynthate transfer in the developing seed of Vicia faba L: a structural assessment of the role of transfer cells in unloading from the seed coat. J Exp Bot **44,** 711-724.

Patrick, J.W., and Offler, C.E. (1995). Post-sieve element transport of sucrose in developing seeds. Australian Journal of Plant Physiology **22,** 681-702.

Rhee, S.Y., Dickerson, J., and Xu, D. (2006). Bioinformatics and its applications in plant biology. Annu Rev Plant Biol **57,** 335-360.

Ritchie, R.J., Fieuw-Makaroff, S., and Patrick, J.W. (2003). Sugar retrieval by coats of developing seeds of Phaseolus vulgaris L. and Vicia faba L. Plant Cell Physiol **44,** 163-172.

Rolletschek, H., Radchuk, R., Klukas, C., Schreiber, F., Wobus, U., and Borisjuk, L. (2005). Evidence of a key role for photosynthetic oxygen release in oil storage in developing soybean seeds. New Phytol **167,** 777-786.

Ruuska, S.A., Girke, T., Benning, C., and Ohlrogge, J.B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell **14,** 1191-1206.

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

Schuurmans, J., van Dongen, J.T., Rutjens, B.P.W., Boonman, A., Pieterse, C.M.J., and Borstlap, A.C. (2003). Members of the aquaporin family in the developing pea seed coat include representatives of the PIP, TIP, and NIP subfamilies. Plant Mol Biol **53,** 655-667.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc **1,** 2856-2860.

Steinfeld, H., and Wassenaar, T. (2007). The role of livestock production in carbon and nitrogen cycles. Annual Review Of Environment And Resources **32,** 271-294.

Tegeder, M., Offler, C.E., Frommer, W.B., and Patrick, J.W. (2000). Amino acid transporters are localized to transfer cells of developing pea seeds. Plant Physiol **122,** 319-325.

Tegeder, M., Wang, X.D., Frommer, W.B., Offler, C.E., and Patrick, J.W. (1999). Sucrose transport into developing seeds of Pisum sativum L. Plant J **18,** 151-161.

Thorne, J.H. (1981). Morphology and ultrastructure of maternal seed tissues of soybean in relation to the import of photosynthate. Plant Physiol **67,** 1016-1025.

Thorne, J.H. (1985). Phloem unloading of C and N assimilates in developing seeds. Annual Review of Plant Physiology **36,** 317-343.

Tilsner, J., Kassner, N., Struck, C., and Lohaus, G. (2005). Amino acid contents and transport in oilseed rape (Brassica napus L.) under different nitrogen conditions. Planta **221,** 328-338.

Veljanovski, V., Vanderbeld, B., Knowles, V.L., Snedden, W.A., and Plaxton, W.C. (2006). Biochemical and molecular characterization of AtPAP26, a vacuolar purple acid phosphatase up-regulated in phosphate deprived Arabidopsis suspension cells and seedlings. Plant Physiol **3,** 1282-1293.

Wang, T.L., and Hedley, C.L. (1993). Genetic and developmental analysis of the seed. In Peas: genetics, molecular biology and biochemistry, ed R. Casey, D.R. Davies, 83-120.

Wang, X.D., Harrington, G., Patrick, J.W., Offler, C.E., and Fieuw, S. (1995). Cellular pathway of photosynthate transport in coats of developing seed of Vicia faba L. and Phaseolus vulgaris L. II. Principal cellular site(s) of efflux. J Exp Bot **46,** 49-63.

Wasinger, V.C., Cordwell, S.J., Cerpapoljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., and Humpherysmith, I. (1995). Progress With Gene-Product Mapping Of The Mollicutes - Mycoplasma-Genitalium. Electrophoresis **16,** 1090-1094.

Weber, H., Borisjuk, L., and Wobus, U. (1996). Controlling seed development and seed size in Vicia faba: a role for seed coat-associated invertases and carbohydrate state. Plant J. **10,** 823-834.

Weber, H., Borisjuk, L., and Wobus, U. (1997a). Sugar import and metabolism during seed development. Trends in Plant Science **2,** 169-174.

Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. Annu. Rev. Plant Biol. **56,** 253-279.

Weber, H., Borisjuk, L., Heim, U., Sauer, N., and Wobus, U. (1997b). A role for sugar transporters during seed development: molecular characterization of a hexose and a sucrose carrier in fava bean seeds. Plant Cell **9,** 895-908.

Weber, H., Heim, U., Golombek, S., Borisjuk, L., Manteuffel, R., and Wobus, U. (1998). Expression of a yeast-derived invertase in developing cotyledons of Vicia narbonensis alters the carbohydrate state and affects storage functions. Plant J **16,** 163-172.

Wilcox, J.R. (2004). World distribution and trade of soybean. In RH Boerma, JE Specht, eds, Soybean: Improvement, Production and Uses, Ed 3., 1-14.

Wobus, U., and Weber, H. (1999a). Sugars as signal molecules in plant seed development. Biol Chem **380,** 937-944.

Wobus, U., and Weber, H. (1999b). Seed maturation: genetic programmes and control signals. Curr Opin Plant Biol **2,** 33-38.

Xu, X.Y., Zheng, R., Li, C.M., Gai, J.Y., and Yu, D.Y. (2006). Differential proteomic analysis of seed germination in soybean. Progress In Biochemistry and Biophysics **33,** 1106-1112.

Yaklich, R.W., Vigil, E.L., and Wergin, W.P. (1984). Scanning Electron-Microscopy Of Soybean Seed Coat. Scanning Electron Microscopy, 991-1000.

Yaklich, R.W., Vigil, E.L., Erbe, E.F., and Wergin, W.P. (1992). The fine-structure of aleurone cells in the soybean seed coat. Protoplasma **167,** 108-119.

Yaklich, R.W., Vigil, E.L., Erbe, E.E., and Wergin, W.P. (1998). Developmental Changes of Antipit Cell Ultrastructure in the Soybean Seed Coat During Seed Maturation. Scanning Microscopy International **12,** 523-532.

Zhang, W.H., Zhou, Y.C., Dibley, K.E., Tyerman, S.D., Furbank, R.T., and Patrick, J.W. (2007). Nutrient loading of developing seeds. Funct Plant Biol **34,** 314-331.

Zhu, H., Bilgin, M., and Snyder, M. (2003). Proteomics. Annual Review of Biochemistry **72,** 783-812.

Chapter 2

CHARACTERIZATION OF THE SOYBEAN SEED COAT PROTEOME (*Glycine max*)

2.1 Introduction

During the 2006-07 growing season, 3500 tons of soybeans were produced in Canada, generating around $900 million in annual crop value (Statistics Canada, 2007). Soybeans are a significant source of fatty acids and proteins for human and animal nutrition as well as for non-edible uses. These uses include industrial feedstock and combustible fuel (Thelen and Ohlrogge 2002). The major source of these commodities is the seed, from which the seed coats represent 6-8% of the total weight (Yoshida et al., 2006b). In recent years, several studies have focused on the development of seeds, both model organisms such as *Arabidopsis thaliana* and *Medicago truncatula*, but also in crop species such as soybean and sunflower. Although the general understanding of seed development and seed biology has substantially increased, there are still major questions to be answered in terms of how the different processes governing development are controlled (Le, 2007). Undoubtedly, a better understanding of the seed structure will be important for future biotechnology efforts, given its potential economic importance.

The function of the seed coat in seed biology has received attention for several decades, generating information on individual proteins extracted from seed coats mainly involved in defence and storage. Seed proteins have been studied in seed development (Weber et al., 2005, Haughn and Chaudhury 2005, Miranda et al., 2003), germination (Kirmizi et al., 2006; Ferreira et al.,1995; Xu et al., 2006b), food allergens (Herman et al., 2003, Xu et al., 2007), seed quality (Blackman et al., 1992), and the understanding of the involvement of the seed coat in such processes has been assessed.

Several studies have reported on the expression of proteins in the seed coat and have drawn some attention to the potential industrial use of this agricultural by-product. Proteomic approaches have been undertaken to study seed filling process in soybean (Hajduch et al., 2005; Mooney and Thelen 2004), and some model systems; e.g., *Arabidopsis* (Ruuska et al., 2002), *Medicago truncatula* (Gallardo et al., 2003), *Brassica napus* (Hajduch et al., 2006). These studies have laid the foundation for an understanding of the complex regulation of seed filling process.

Seed structure is quite diverse among species, even within the legume family (Le et al., 2007). To date, a comprehensive proteomic study of soybean seed coats has not been produced. This study aims at elucidating the soybean seed coat proteome at 35-50 days post anthesis (DPA), a developmental stage in which seeds are fully developed, but the desiccation process has not yet started. From previous studies we understand that the main role of the seed coat is protection of the embryo and its nurture in development (Haughn and Chaudhury (2005). We predict, based on the diversity of cell types in this organ, that the seed coat has diverse functions that may or may not change during overall seed development. A comprehensive proteomic description of the seed coat should enable us to gain knowledge on the metabolic processes that take place in this organ, as well as provide an opportunity to understand the areas that could be further utilized in efforts to enhance the soybean crop value, an important commodity in agriculture.

2.2 Experimental procedures

The workflow of the methods used for the analysis of soybean seed coat proteome is presented in Figure 2.1.

2.2.1 Sample collection and preparation

a) Plant materials and growth conditions

Soybean seeds (*Glycine max*) L. Merr. cv Harosoy 63 were planted at the Agriculture and Agri-Food Canada Research Centre in London, Ontario, in 2006 and 2007. Regular agronomic practices and planting dates were followed. Flowers at anthesis at nodes 3 and 4 were tagged and harvested weekly between 35-50 days post anthesis (DPA). The pods were collected randomly from 20-30 plants, and seed coats were excised from seeds, frozen in liquid nitrogen, and stored at -80 °C.

b) Trichloroacetic acid precipitation of proteins

Total protein was isolated from soybean seed coats and subjected to trichloroacetic acid (TCA) precipitation according to Gorg *et al.* (1997) with modifications from Natarajan *et al.* (2005). Dissected seed coats were pulverized in liquid nitrogen in a mortar with a pestle. Soybean seed coat powder (1 g) was homogenized with ~3 ml of a solution containing 10% (v/w) TCA in acetone (-18 °C) with 0.07% (v/v) 2-mercaptoethanol.

Homogenate was vortexed for 1h at 4 °C. Total protein was precipitated overnight at -20 °C. Following centrifugation at 10,500 x g for 20 min at 4 °C, the supernatant was discarded and the pellet was washed three times with a solution containing acetone (-18 °C) and 0.07% (v/v) 2-mercaptoethanol, after which, it was dried under vacuum for 30 min, resuspended for immediate gel fractionation or stored at -20 °C.

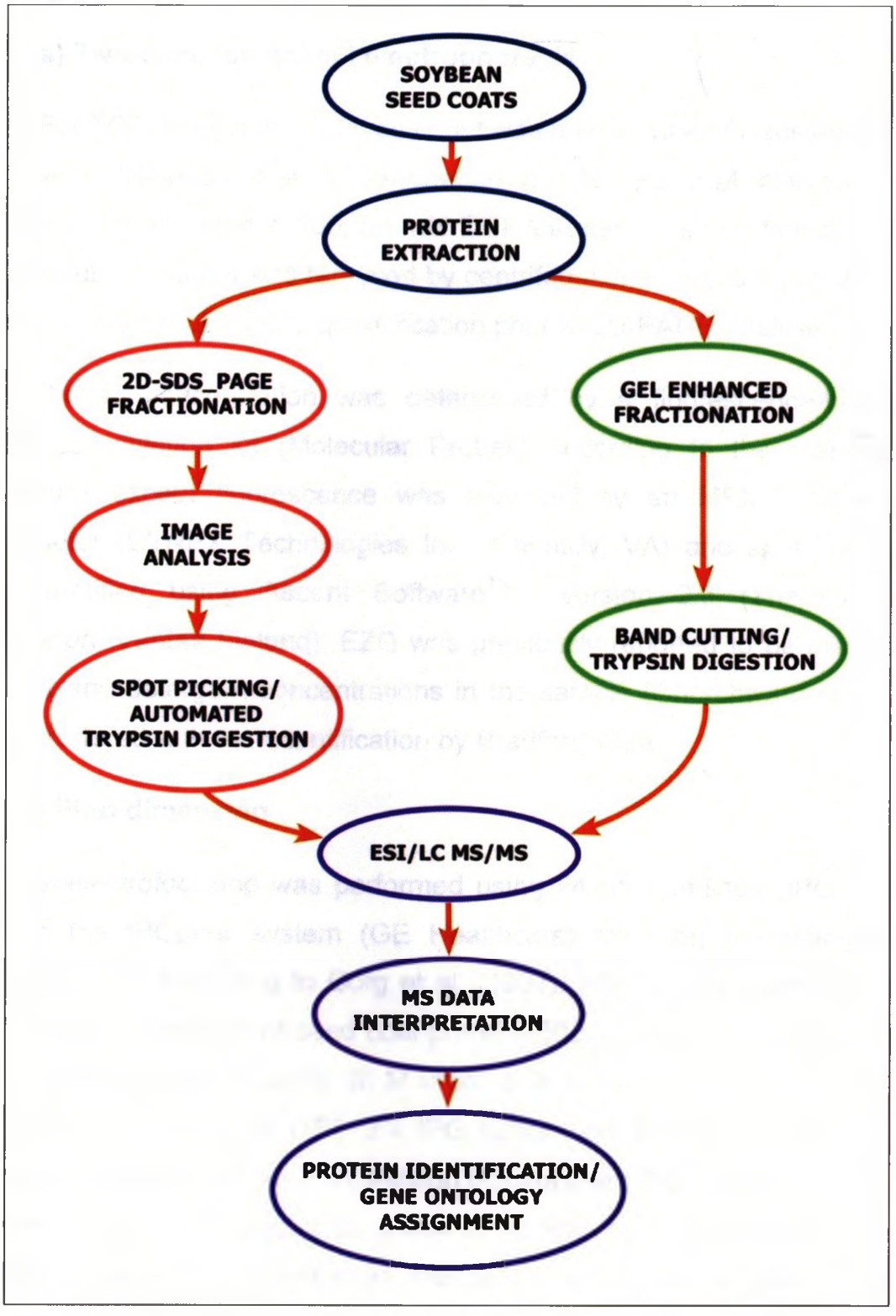


Figure 2.1. Flow chart of methods employed to determine the soybean seed coat proteome.

2.2.2 Sample fractionation

a) Two-dimensional gel electrophoresis

For TCA extraction, the dried pellet was resuspended in isoelectrofocusing (IEF) media (Hajduch et al., 2005) containing 8 M urea, 2 M thiourea, 2% [w/v] CHAPS, 2% [v/v] Triton X-100, 50 mM DTT followed by sonication on ice for 30 min. Insoluble material was removed by centrifugation at 10,500 g for 20 min at 4 °C and extracts subjected to quantification prior to 2D-PAGE analysis.

Protein concentration was determined by a fluorescence-based EZQ Protein Quantitation kit (Molecular Probes) according to the manufacturer's instructions except fluorescence was recorded by an MFX Microtiter Plate Fluorometer (DYNEX Technologies Inc., Chantilly, VA) and spot fluorescence was quantified using Ascent Software™ version 2.6 (Thermo Electron Corporation, Vantaa Finland). EZQ was previously reported to be insensitive to high salt and detergent concentrations in the sample (Churchward et al., 2005) which interfered with the quantification by Bradford assay.

I) First dimension

Isoelectrofocusing was performed using 24-cm non-linear IPG strips (pH 3-11) in the IPGphor system (GE Healthcare) following the manufacturer's instructions and according to Görg et al. (1997). All IPG strips were rehydrated with the desired amount of seed coat protein (500 mg) which was brought up to 450 µL with rehydration buffer (8 M urea, 2 M thiourea, 2% [w/v] CHAPS, 2% [v/v] Triton X-100, 50 mM DTT, 2% IPG buffer [v/v], 0.002% BPB [w/v], 12 µL DeStreak reagent/ml rehydration solution [Amehrsham Biosciences]. The mixture was vortexed and centrifuged for 5 min at 10,500 x g to remove the remaining insoluble matter before rehydration. The protein mix was then transferred to an IEF tray, and a 24-cm non-linear Immobiline Dry Strip gel (IPG) (pH 3-11) (Amersham Biosciences, Upssala) was carefully placed onto the protein sample, covered with mineral oil and allowed to rehydrate for 15-18 hrs. Isoelectrofocusing and SDS electrophoresis were carried out in a flatbed

Multiphor II Elctrophoresis system (GE Healthcare). For IEF, the following voltage settings were used: 100 V for 2 h, 500 V for 1 min, 2990 for 1 h 45 min, 2990 V for 16 h 51 min to a total of 55.88 kVh. The focused strips were removed from the focusing tray and either run immediately on a 2D electrophoresis or stored at - 80 °C.

II) Second dimension

For the 2D electrophoresis, the focused strips were incubated with the equilibration buffer 1 (50 mM Tris-HCl [pH 8.8], 6 M urea, 30% [v/v] glycerol, 2% [w/v] SDS, 0.002% [w/v] BPB, 1% [w/v] DTT) followed with the equilibration buffer 2 (50 mM Tris-HCl [pH 8.8], 6 M urea, 30% [v/v] glycerol, 2% [w/v] SDS, 0.002% [w/v] BPB, 2.5% [w/v] IAA) for 15 min each. Second dimension SDS-PAGE was performed using precast ExcelGel 12.5% polyacrylamide homogeneous gels (11x 24 cm) under constant current 20 mA for 40 min followed by 50 mA until the front dye has reached the anodic buffer strip (~1 h 10 min).

Following SDS-PAGE, gels were visualized by staining with colloidal Coomasie Brilliant Blue G-250 (Pierce) following manufacturer's instructions and according to Syrovy and coworkers (Syrovy and Hodny, 1991). The gels were fixed overnight in 40% ethanol and 10% acetic acid followed by 3 x 30 min washes in distilled water. Then the gels were stained for at least 24 hr with a solution containing 20% [v/v] methanol, 0.8% [v/v] phosphoric acid, 8% [w/v] ammonium sulfate and 0.08% [w/v] Coomasie Brilliant Blue G-250. The gels were stored in 20% glycerol at 4 °C until further analysis.

b) Analysis of 2-D Gels

Image acquisition was performed using a PowerLook 1120 scanner (UMAX Technologies Inc., Taiwan) with a resolution of 300 dpi and 16-bit grayscale pixel depth. Image analysis was carried out with Progenesis PG220 v2006 and Progenesis SameSpots TT900 SDS^{TM} software (Nonlinear Dynamics, Newcastle upon Tyne, UK).

Gel digital images were analyzed in four technical replicates (four different protein extractions from plants grown during the same growing season) following the instructions on the software user's manual. The aligned images were corrected for positional variations using manual and automatic applied vectors between spots on two images followed by an image alignment. Protein spots were detected using automatic spot detection and the background was subtracted using the mode of non-spot method. The digital images were also subjected to spot filtering in order to remove artifacts, for contrast enhancement, background subtracting, etc. The total intensity of pixels within each spot (the integrated intensity) was determined by the software. The integrated intensity of each spot (normalized volume) was expressed as percentual fractions of the total integrated intensity of all spots within the region of analysis of the gel. This normalized the amount of any given spot and gave relative protein abundance values for each sample. Protein spots in the two different gels (Rep 2 against Rep 1, 3 and 4) were then matched for qualitative and/or quantitative differences between the 2D patterns. For the particular task of protein identification, gels were compared for the same proteins present on all the replicates using Pearson linear regression.

After gel image analysis, spots were selected based upon the following criteria: a) were present in all 4 replicates and b) their expression was above a normalized volume of 10.7 (provided that they were big and resolved enough to be picked), were manually excised from the gels using a OneTouch manual spot picker (The Gel Company) (3.0 mm). Excised spots were subjected to automated in-gel trypsin digestion using a MassPREP Automated Digestor (Waters) following the manufacturer's instructions. Excised spots (discs) were individually placed on wells in a 96 well microplate. Gel discs were distained by washing with a solution of 100 mM NH_4CO_3 and 20% ACN. For cysteine reduction, the discs were incubated for 30 min in 20mM DTT in 100 mM NH_4CO_3. DTT solution was discarded followed by 20-min incubation in a solution of 55 mM IAA in 100 mM NH_4 CO_3 for alkylation followed by a washing step. Discs were dehydrated with 100% ACN and rehydrated with 100 mM NH_4CO_3. For digestion, discs were

dehydrated with 100% ACN and rehydrated with trypsin (Promega-Porcine modified)(6 ng/mL) in 100 mM NH_4CO_3 for 5 hrs. To extract the peptides, gel discs were washed three times with 10% FA and once with 100% ACN. Samples were evaporated to dryness in a Speedvac and resuspended in 10% FA for IE-LC-MS/MS analysis.

c) Gel enhanced fractionation (Gel LC MS)

For gel-enhanced fractionation, a composite seed coat extract was obtained by mixing aliquots from each of the 4 replicates after TCA precipitation. The seed coat extracts were reconstituted in 1X Laemmli loading buffer, and resolved on a 1.5 mm, 8-15 % gradient SDS-PAGE precast mini-gel (BIO-RAD). Two replicate gels were stained with Coomassie Brilliant Blue G-250 and the entire lane representing the concentrated sample was divided into ~10 sections (fractions). A total of 45 µg of protein was loaded in each replicate.

Each gel section was digested manually (Shevchenko et al., 2006). Briefly, gel bands were cubed into smaller pieces (~2 mm^2) and destained by washing in 1M (NH_4CO_3), 20% ACN. For cysteine reduction, the gel pieces were dehydrated with 100% ACN and rehydrated with 10 mM DTT in 100 mM NH_4CO_3 for 30 min. The DTT solution was removed and the gel pieces were alkylated by adding 100mM IAA in 100 mM NH_4CO_3 for 30 min. The gel pieces were washed and dehydrated with 100% ACN, then rehydrated with 50 mM NH4CO3. For digestion, the gel pieces were first dehydrated with 100% ACN, then rehydrated with trypsin (Promega – Porcine modified) (20 µg/mL) in 50 mM NH_4CO_3 on ice for 15 min. Excess trypsin solution was removed. The gel pieces were covered with 50 mM NH_4CO_3 and digested for 18 hrs at 37°C. To extract the resulting peptides, the supernatant was collected and gel pieces were extracted three times with 10% FA and followed once with 100 % ACN. Samples were evaporated to dryness with a Speedvac and resuspended in 10% FA for LC-MS/MS analysis.

2.2.3 Mass Spectrometry and protein identification

a) IE-LC-MS/MS analysis

For ion exclusion liquid chromatography tandem mass spectrometry analysis, all dried fractions, including 2D spots and gel enhanced gel bands were reconstituted in 10% FA prior to injection. For analysis, spots and band samples were kept separated. The complexity of each sample was estimated based on the apparent clear, light or dark intensity of the Coomasie stain and samples were subjected to MS/MS in this ascending order. 2D samples were analyzed using a 60 min LC method; whereas, band samples were analyzed in three steps of 60 min each, first run and two exclusion steps (exclusions lists) (Bendall, 2008). Liquid chromatography (5-40% ACN, 0.1% FA gradient) was performed on a NanoAcquity UPLC (Waters, Milford, MA) with a 25 cm x 75 µm C18 reverse phase column. Peptide ions were detected in data-dependent acquisition (DDA) mode by tandem MS (Q-ToF Ultima - Waters) using the following parameters: survey scan (MS only) range *m/z* 400-1800, 1 s scan time, 1-4 precursor ions selected based on charge state (+2, +3, and +4). For each MS/MS scan, the *m/z* range was extended to *m/z* 50 – 2000, scan times used ranged from 1.5 - 6 s (signal dependent), and a charge state-dependent collision energy profile was used.

I) Iterative exclusion list method

For the analysis of each gel band fraction, the *m/z* and RT values were manually extracted from the ".RAW" data folder ('*auto.txt*' file) for all ions selected in the previous MS/MS analysis. All previously selected ions were excluded, not just those identified as peptides. This approach ensures that ions with high spectral intensity are not analyzed more than once, even if they were not identified as peptides via conventional MS/MS analysis. To create an exclusion window centered on the major isotopes and avoid excluding masses below the monoisotopic peak, a *m/z* shift of 0.7 was added to each *m/z* value selected for MS/MS. These ions were excluded (Waters-MassLynx DDA exclude functionality) from all analyses performed after that fraction using a *m/z* tolerance

window of ±0.8 and RT window of ±45 s. This process was repeated twice for each fraction. The creation of iterative exclusion lists was successfully used for the determination of low abundance proteins in the seed coat, increasing the positive by 40 and 15% in each successive round. It allowed the identification of new peptides in every round or increased the number of peptides for previously identified proteins, augmenting the confidence on the identification.

II) MS Data Interpretation and Gene Ontology Assignment

The acquired MS/MS spectra were processed by using the ProteinLynx Global SERVER 2.2.5 (Waters) and searched against extracted subsets for Plants or *Glycine max* (forward and reverse) of NCBInr protein databases (www.ncbi.nlm.nih.gov) using Spectrum Mill (Agilent Technologies, Santa Clara, CA). The following settings were employed: a mass tolerance of 100 ppm for MS spectra and 100 ppm for MS/MS spectra, a spectral peak intensity (SPI) limit of 60%, minimum peptide score of 6, and minimum protein score of 13. To minimize false positives to a rate of 0.0001%, peptides with reverse database scores higher than forward scores were removed from the summaries.

Gene ontology was assigned to all identified proteins in all samples according to a classification for yeast adapted for the *Arabidopsis* genome (Bevan et al., 1998) with modifications that make it more suitable for a seed study (Hajduch et al., 2006b).

2.2.4 Seed coat dissection

Seed coat tissue preparation was previously reported (Dhaubhadel et al., 2005). Briefly, seed coats from mature soybean seeds (80 DPA) were soaked in distilled water for 2-3 h, then cut in halves and the embryos were removed. Seed coats were immersed in luke warm 3% agarose solution. Once cooled off, cubes of the gel containing tissue were cut and sectioned into 40 µ slides using a vibrating blade microtome (Leica VT 1000S), stained with 0.05% (w/v) toluidine blue and observed under an inverted microscope. Digital pictures were taken with a DXM 1200 Nikon camera.

2.3 Results

2.3.1 Protein extracts fractionation by 1D and 2D-SDS-PAGE

In this study, the seed coat proteome of physiologically mature soybean seeds (35-50 DPA) was pre-fractionated by one and two-dimensional SDS PAGE. A flow chart showing the steps in 2D gel analysis is presented in Figure 2.2., the 2D-SDS-PAGE gel from which spots were excised is shown at the bottom of the flow chart. The four technical replicates utilized for the 2D-gel image analysis are reported in Appendix I. The Pearson product-moment correlation coefficient R^2 was calculated for the relationships between normalized spot volume of replicate 2 (which was chosen to excise the spots) and the same values in replicates 1, 3 and 4 and the regression is shown in Figure 2.3. The R^2 value of 0.92 indicates that there is an almost perfect positive linear relationship between the replicates, and therefore the level of reproducibility of the gels is very high. It is important to consider this aspect prior the selection and excision of spots, to be confident in the consistent expression of the spots for further analysis.

The SDS-PAGE pre-fractionation includes two technical replicates shown in Figure 2.4 along with the molecular weights at which the 10 bands were cut. The bands from each replicate were submitted for trypsin digestion and mass spectrometry individually. To better analyze the complex sample from each gel fraction, we used the iterative exclusion approach (Bendall, 2008) in order to maximize the protein identification per fraction. Figure 2.5 shows the characteristics of the identification of unique peptide obtained following this methodology.

Once the all spectral data was acquired, regardless of pre-fractionation strategy, that is, from 2D spots and gel fractions (with the iterative exclusion results), the files were utilized to create a composite database of proteins found in the seed coat of fully developed seed coats.

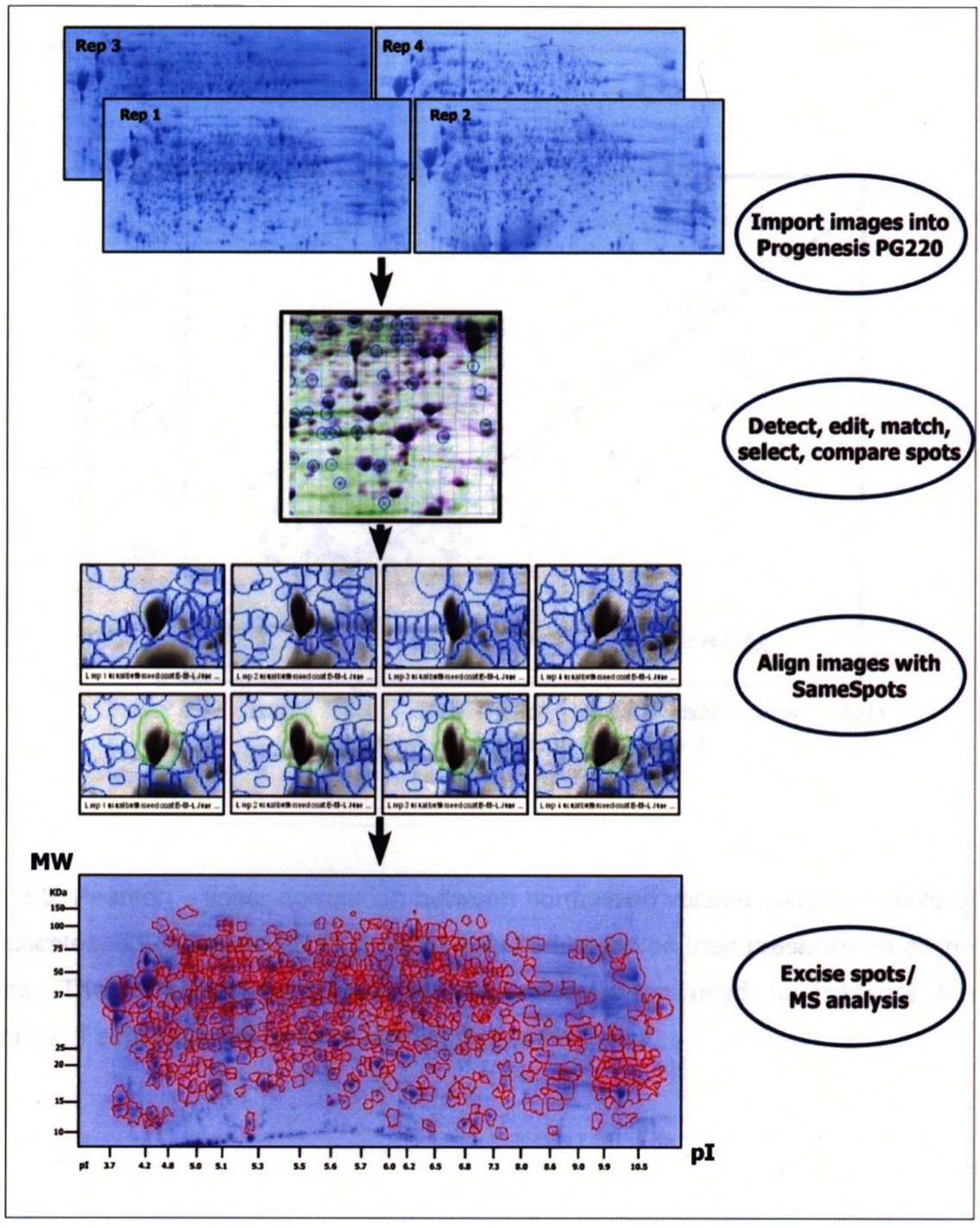


Figure 2.2. Flow chart of analysis of 2D gel image analysis using Progenesis220 with SameSpots. Four technical replicates of 500 µg of seed coat proteins separated by 2D-SDS-PAGE and images analyzed for MS studies. Outlined in red are the spots that were found in all four technical replicates of 2D gels after image analysis.

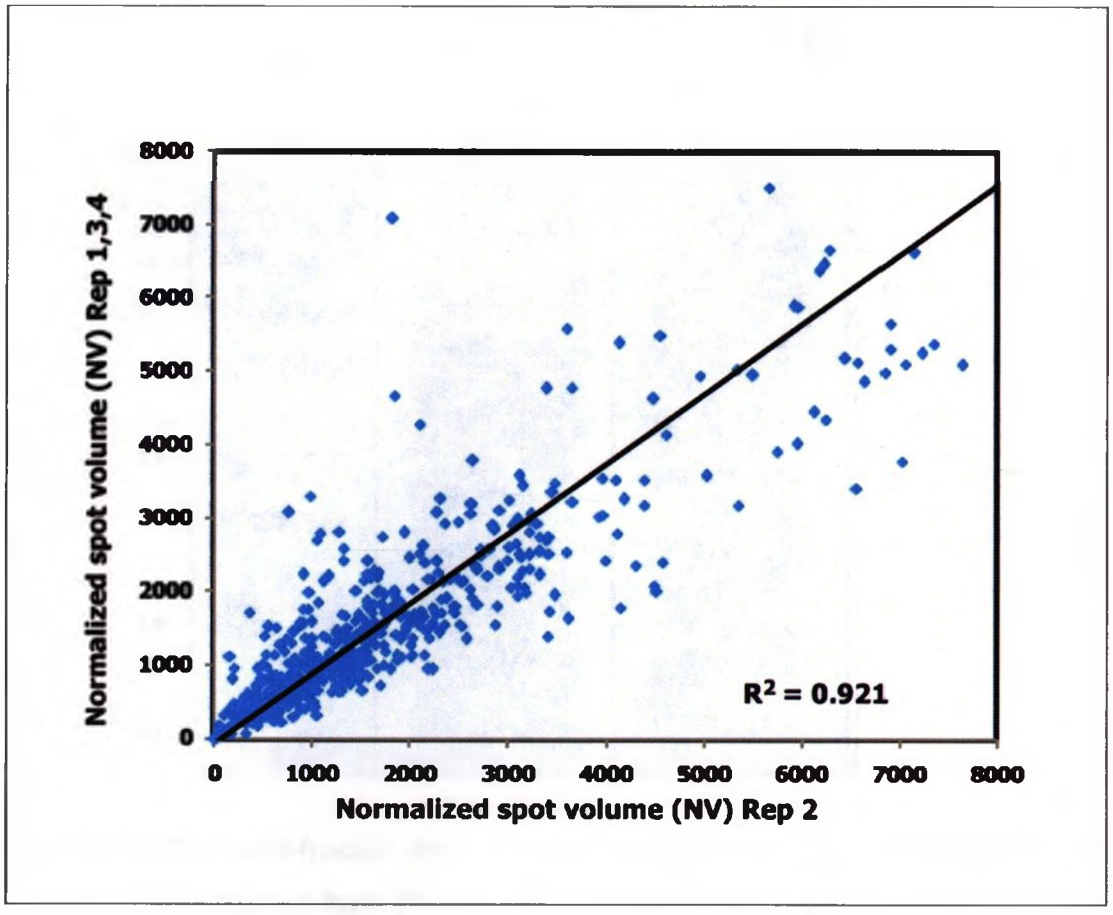


Figure 2.3. Pearson – linear correlation between normalized volume values of spots of 4 gel replicates. 532 selected 2D spots were selected based on their presence on all the replicates. The acquired normalized volume values were used to measure the consistency of expression of proteins among replicates.

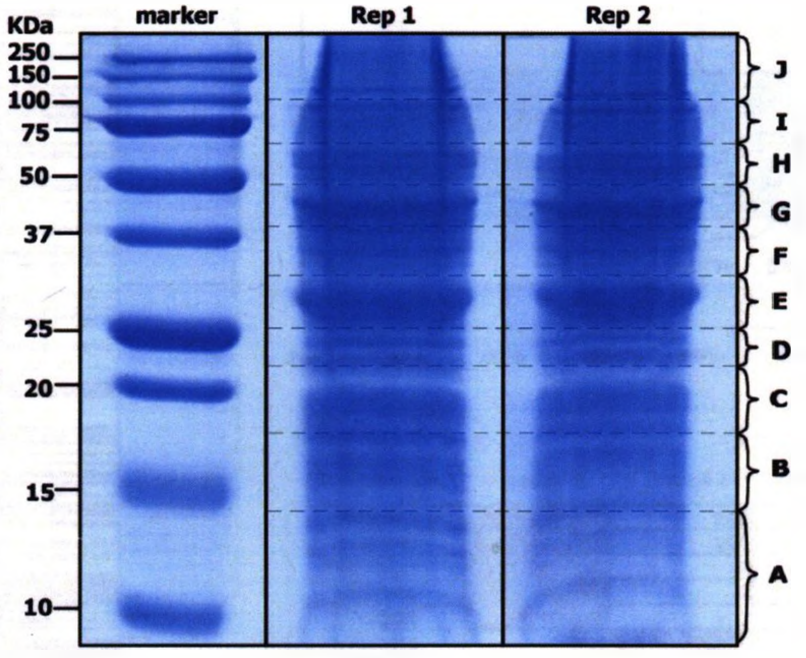


Figure 2.4. SDS-PAGE pre-fractionation of seed coat proteins. Two technical replicates of 45 µg of protein extracted from 35-50 DPA soybean seed coats separated by SDS-PAGE. Letters A - J represent excised bands that were further analyzed by LC-MS/MS

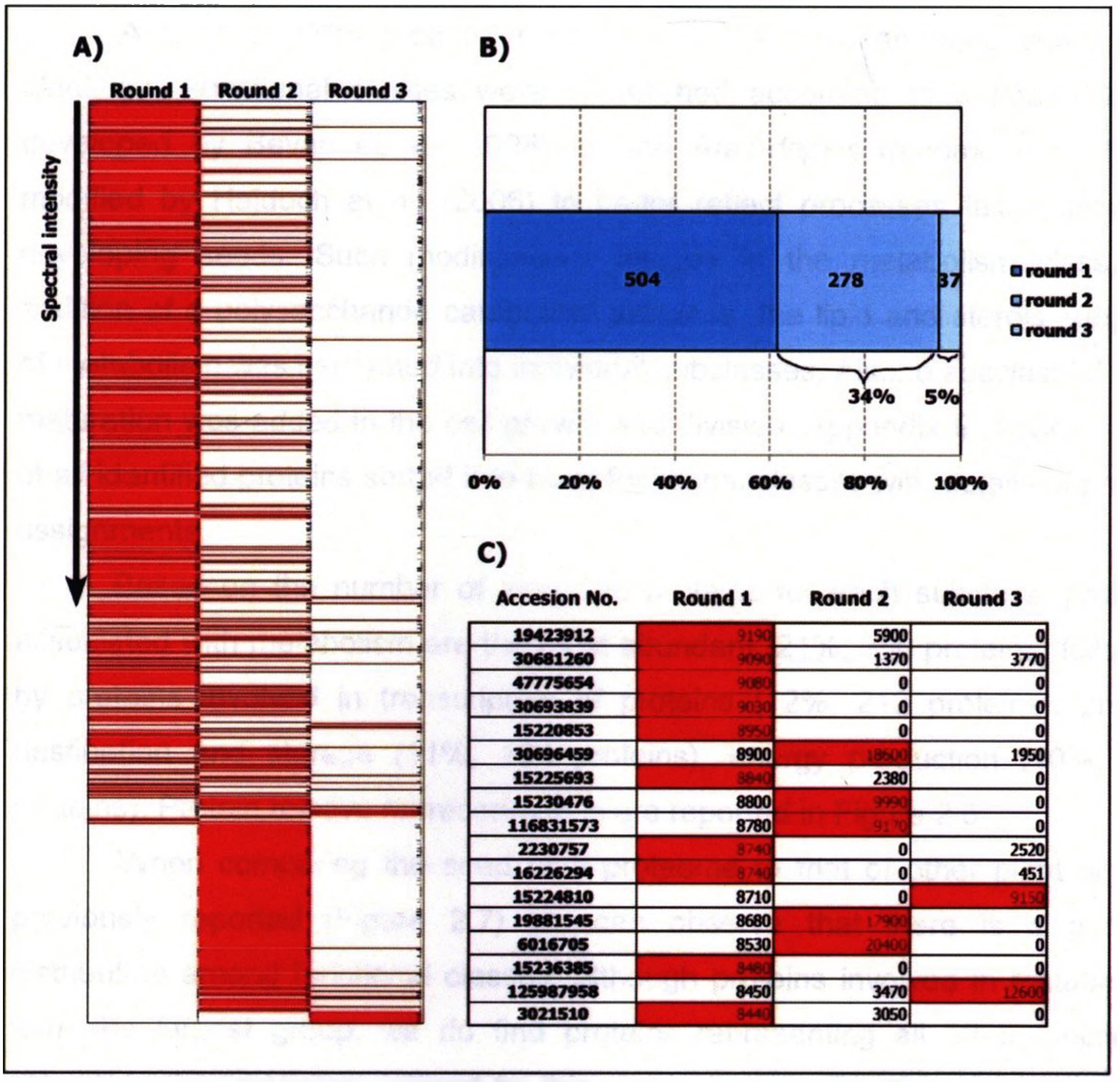


Figure 2.5. Protein identification from seed coat extracts with iterative spectral exclusion lists. 10 different gel separated fractions were analyzed using iterative exclusion lists. A) Heat maps for iterative exclusion where each row corresponds to a protein identification sorted by total spectral intensity show which round of analysis in which 1 or more unique peptides were identified. B) The relative increments in protein identification in successive iterative exclusions. C) A table outlining high confidence (> 1 unique peptide, score > 13) seed coat proteins in which entries are listed according to total spectral intensity of unique peptides identified for each analysis round. Highlighted in red are spectral intensities that led to the identification of unique peptides.

2.3.2 Identification of proteins in soybean seed coat extracts

A total of 1705 proteins from 35-50 DPA soybean seed coats were identified. Functional classes were established according to a nomenclature developed by Bevan et al. (1998) for the *Arabidopsis* genome project and modified by Hajduch et al. (2006) to better reflect processes taking place in developing seeds. Such modifications include in the metabolism class, the addition of a polysaccharide catabolism subclass, the lipid and sterols subclass of metabolism was separated into individual subclasses. Also, a subclass of seed maturation was added in the cell growth and division. Appendix II provides a list of all identified proteins sorted into plant functional classes with details of protein assignments.

Based on the number of identified proteins for each subclass, proteins associated with metabolism are the most abundant (21%, 350 proteins) followed by proteins involved in transcription of proteins (12%, 212 proteins), protein destination and storage (11%, 184 proteins), energy production (10%, 165 proteins). Protein relative representations are reported in Figure 2.6.

When comparing the seed coat proteome to that of other plant organs previously reported (Figure 2.7) we can observe that there is a general distribution among functional classes, although proteins involved in metabolism form the largest group, we do find proteins representing all other functional groups. In the proteomes utilized for this comparison we find that there is a trend to have a very well represented group of metabolic proteins, regardless the organ. In the case of leaves, there is however, a very strong specialization of the proteins, with about 50% of them devoted to energy production and related activities. From this comparison we can infer that the seed coat is a multifunctional organ with proteins representing most of the functional classes.

Our results show an unprecedented and unexpected wealth of proteins present in the seed coat, both in terms of number as well as in diversity of functions. From a preliminary inspection of the data, several enzymes were identified that are involved in the biosynthesis of the cell wall, fatty acids, cutin,

and isoflavonoids and also in C_1 metabolism and the proteolytic pathway. Such pathways will be considered in greater detail in the following sections.

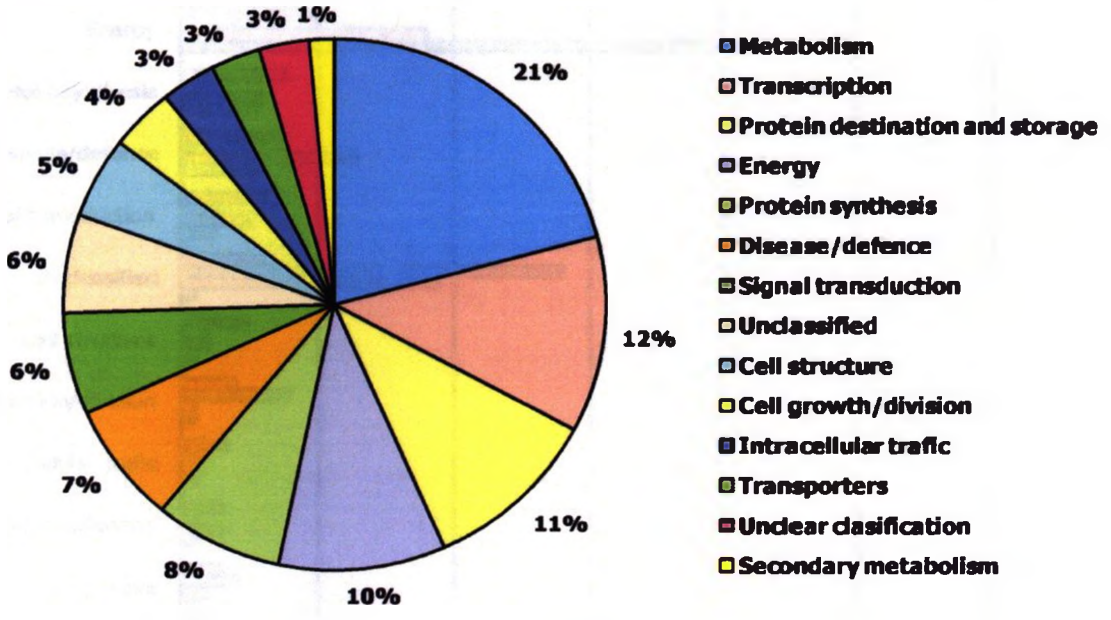


Figure 2.6. The functional distribution of 1705 non-redundant proteins identified from fully developed soybean seed coats (35-50 DPA). Classification was based upon nomenclature by Hajduch et al. (2006).

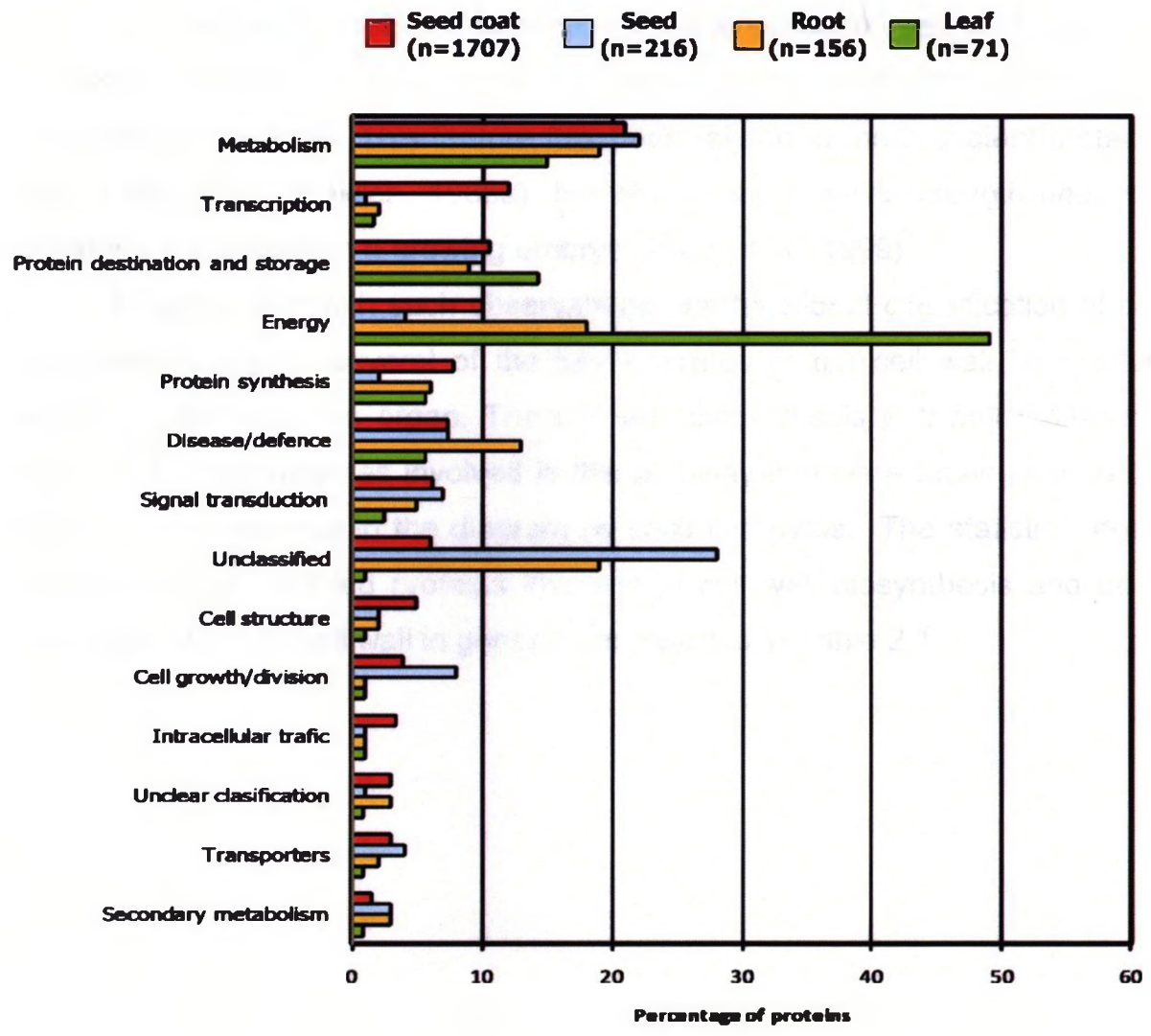


Figure 2.7. Comparison of functional categories of proteins identified from different plant tissues. The data were summarized based on data of studies of soybean seed coats, seeds (Hajduch et al., 2005), leaves (Xu et al., 2006) and *Arabidopsis* roots (Mooney et al., 2006). The plant tissue is shown in colored squares and n is the number of non-redundant proteins identified in each study. The classification was based on nomenclature by Bevan et al., (1998) and the categories are shown in the y-axis, and the x-axis shows the percentage of identified proteins in each study.

2.3.3 Cell wall biosynthesis in the seed coat

As previously presented in section 1.1, the soybean seed coat possesses different cell types that have thickened cell walls: the epidermal and hourglass and parenchyma cells. This feature has been related to the prevalent protective role of this organ (Yaklich, 1986a), but also to the relative strength needed to withstand the extension a growing embryo (Miller et al., 1999).

In agreement with such observations, the functional classification of seed coat proteins shows several of the key enzymes of the cell wall biosynthesis pathway present in this organ. The cell wall biosynthesis pathway is shown in Figure 2.8. The enzymes involved in this pathway that were found in seed coat extracts are presented in the diagram as solid line ovals. The statistical details for each of the reported proteins involved in cell wall biosynthesis and others associated with the cell wall in general are reported in Table 2.1.

Figure 2.8. Seed coat enzymatic sequences involved in cell wall biosynthesis adapted from Seifert *et al.* (2004). Most enzymes are localized in the cytosol, where they interact metabolically with glycolysis and gluconeogenesis through the reversible actions of phosphomannose isomerase (PMI), phosphoglucose isomerase (PGI), phosphomannomutase (PMM), and phosphoglucomutase (PGM). Nucleotide sugars are generated in vivo by UDP-$_D$-glucose pyrophosphorylase (UGP), GDP-$_D$-mannose pyrophosphorylase (GMP), and UDP-$_D$-glucuronic acid pyrophosphorylase (UAP). UDP-$_D$-glucose is also generated by sucrose synthase (SUS). GDP-$_D$-mannose is converted either to GDP-$_L$-fucose by the sequential action of the directly interacting GMD (GDP-$_D$-mannose-4,6 dehydratase) and GER (GDP-4-keto-6-deoxy-$_D$-mannose-3,5-epimerase-4-reductase) or into GDP-$_L$-galactose and GDP-$_L$-gulose by GME (GDP-$_D$-mannose 3,5-epimerase). UDP-$_D$-glucose is converted into UDP-$_D$-galactose by UGE (UDP-glucose 4-epimerase), into UDP-$_D$-glucuronic acid by UGD (UDP-$_D$-glucose dehydrogenase) or into UDP-$_L$-rhamnose by RHM (rhamnose synthase), which hypothetically consists of sequentially acting UDP-$_D$-glucose 4,6-dehydratase and UDP-4-keto-6-deoxy-$_D$-glucose 3,5-epimerase 4-reductase. The sequential action of inositol oxygenase (INO), $_D$-glucuronokinase (GAK) and UAP represents an alternative pathway of UDP-$_D$-glucuronic acid biosynthesis. Two different cytosolic UDP-$_D$-glucuronic acid decarboxylase: UXS (UDP-$_D$-xylose synthase) and AXS (UDP-$_D$-apiose/UDP-$_D$-xylose synthase) give rise to UDP-$_D$-xylose or to a mixture of UDP-$_D$-xylose and UDP-$_D$-apiose, respectively. UDP-$_D$-glucose, UDP-$_D$-galactose, UDP-$_D$-glucuronic acid, UDP-$_L$-rhamnose, UDP-$_D$-apiose, GDP-$_D$-mannose, GDP-$_L$-fucose and GDP-$_L$-galactose are transported into the endomembrane system, where specific glycosyltransferases are localized. UDP-$_D$-glucose is also channeled to cellulose synthase (CeS) and callose synthase (CaS), which are localized at the plasma membrane. In the lumen of the endomembrane system, UDP-$_D$-glucuronic acid is either converted into UDP-$_D$-xylose by membrane-bound UXS or into UDP-$_D$-galacturonic acid by membrane-bound GAE (UDP-$_D$-glucuronic acid 4-epimerase). UDP-$_D$-xylose is converted into UDP-$_L$-arabinose by membrane-bound UXE (UDP-$_D$-xylose 4-epimerase). Api, apiose; Ara, arabinose; Frc, fructose; Gal, galactose; GalA, galacturonic acid; Glc, glucose; Gul, gulose; GlcA, glucoronic acid; Man, mannose; PPi, inorganic pyrophosphate; UTP, uridine triphosphate; Xyl, xylose. Enzymes represented by a solid line oval were identified in the seed coat proteome. Those presented by a dashed line oval were not.

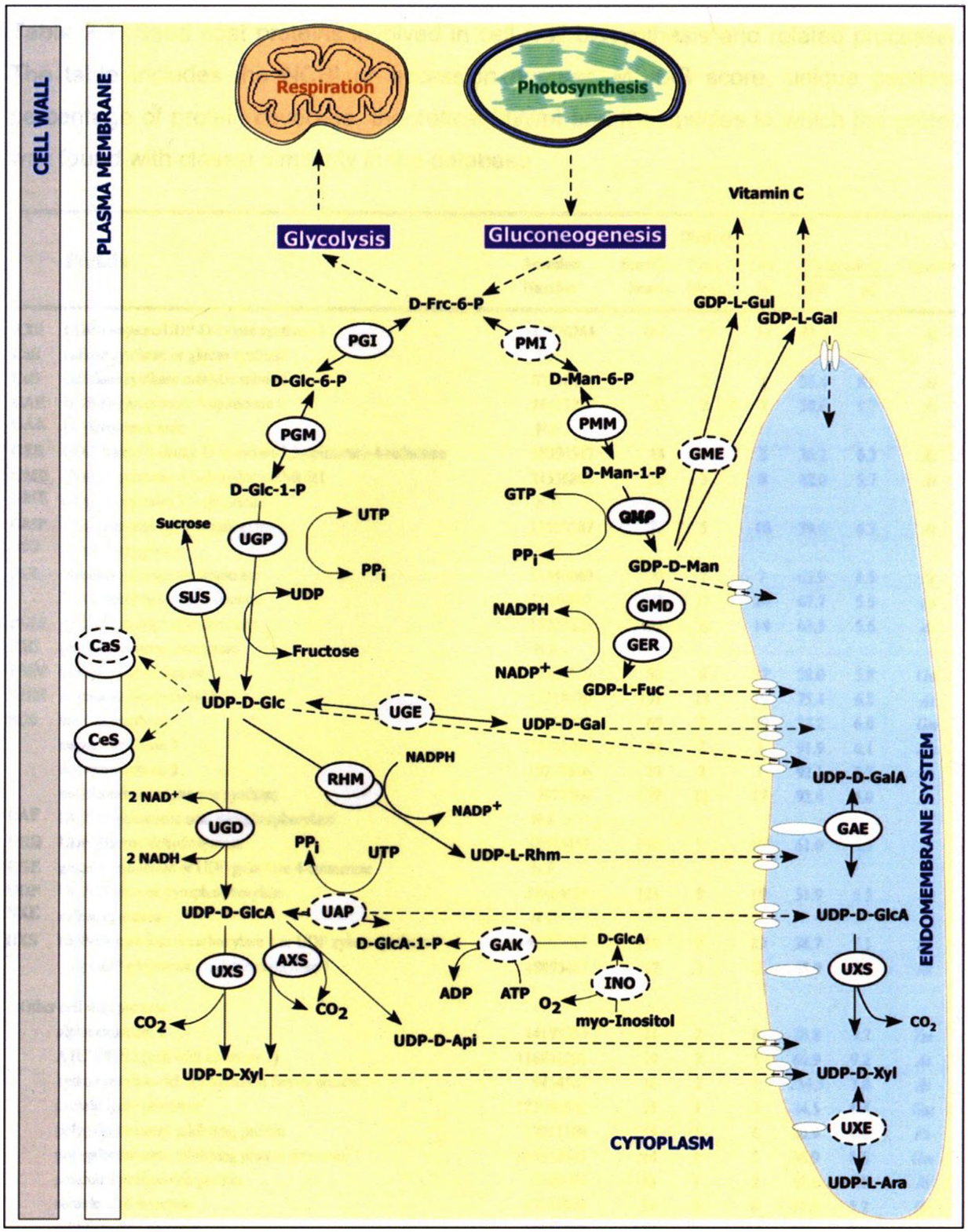

CELL WALL
PLASMA MEMBRANE
Respiration
Photosynthesis
Glycolysis
Gluconeogenesis
Vitamin C
D-Frc-6-P
PGI
PMI
D-Glc-6-P
D-Man-6-P
PGM
PMM
GME
GDP-L-Gul
GDP-L-Gal
D-Glc-1-P
D-Man-1-P
GTP
GMP
PPi
GDP-D-Man
Sucrose
UGP
UTP
PPi
SUS
UDP
Fructose
NADPH
GMD
GER
NADP+
GDP-L-Fuc
CaS
CeS
UDP-D-Glc
UGE
UDP-D-Gal
UDP-D-GalA
NADPH
RHM
NADP+
2 NAD+
UGD
2 NADH
PPi
UTP
UDP-L-Rhm
GAE
UDP-D-GlcA
UAP
UDP-D-GlcA
D-GlcA-1-P
GAK
D-GlcA
UXS
AXS
ADP
ATP
INO
O2
myo-Inositol
UXS
CO2
CO2
CO2
UDP-D-Api
UDP-D-Xyl
UDP-D-Xyl
UXE
UDP-L-Ara
ENDOMEMBRANE SYSTEM
CYTOPLASM

Table 2.1. Seed coat proteins involved in cell wall biosynthesis and related processes. The table includes the NCBI nr accession number, MS/MS score, unique peptides, percentage of protein coverage, theoretical MW/pI and the species in which the protein was found with closest similarity in the database.

	Protein	NCBI Accesion Number	MS/MS Search Score	Distinct Pept. Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
AXS	UDP-D-apiose/UDP-D-xylose synthase 1	15226264	167	12	31	43.6	5.5	*At*
CaS	callose synthase or glucan synthase	N.F.						
CeS	cellulose synthase catalytic subunit	62318989	19	2	3	28.4	8.9	*At*
GAE	UDP-D-glucuronate 4-epimerase 6	24417280	20	2	7	50.6	9.7	*At*
GAK	D-glucuronokinase	N.F.						
GER	GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase	18394547	18	1	3	36.2	6.3	*At*
GMD	GDP-D-mannose-4,6-dehydratase MUR1	21536808	28	3	8	42.0	5.7	*At*
GME	GDP-D-mannose 3,5-epimerase	N.F.						
GMP	GDP-mannose pyrophosphorylase	13509287	77	5	16	39.6	6.3	*At*
INO	inositol oxygenase	N.F.						
PGI	glucose-6-phosphate isomerase	51340060	20	2	7	62.9	6.5	*St*
	or phosphoglucose isomerase	51340062	143	11	24	67.7	5.5	*Ls*
PGM	cytosolic phosphoglucomutase	15223226	123	8	14	63.5	5.6	*At*
PMI	phosphomannose isomerase	N.F.						
PMM	phosphomannomutase	90762150	90	6	27	28.0	5.8	*Gm*
RHM	rhamnose biosynthesis 1	15218420	191	13	19	75.4	6.8	*At*
SUS	sucrose synthase	63852202	69	5	39	23.2	6.0	*Gm*
	sucrose synthase 3	22121990	24	2	1	91.9	6.1	*Zm*
	sucrose synthase 2	15239816	19	2	1	92.1	5.7	*At*
	nodule-enhanced sucrose synthase	3377764	179	12	17	92.4	6.0	*Ps*
UAP	UDP-D-glucuronic acid pyrophosphorylase	N.F.						
UGD	UDP-glucose dehydrogenase	48093457	118	7	17	61.0	6.5	*Nt*
UGE	glucose epimerase or UDP-galactose 4-epimerase	N.F.						
UGP	UDP-D-glucose pyrophosphorylase	28863909	123	9	19	51.9	5.5	*St*
UXE	xylose epimerase	N.F.						
UXS	UDP-glucuronate decarboxylase 1 or UDP xylose synthase	48093461	119	8	23	38.7	7.1	*Nt*
	or UDP-glucuronic acid decarboxylase 1	48093467	32	3	8	45.9	9.0	*Nt*
Other cell wall proteins								
	alpha-expansin 2	14193753	21	2	6	28.8	9.7	*Zm*
	ATCWINV2 (cell wall invertase 2)	116831291	19	2	5	66.9	9.2	*At*
	hydroxyproline-rich glycoprotein family protein	9454580	16	2	1	234.3	7.8	*At*
	pectate lyase precursor	127464581	15	1	3	34.5	6.4	*Gm*
	polygalacturonase inhibiting protein	37051109	14	1	6	20.9	9.4	*Ps*
	polygalacturonase inhibiting protein precursor	110836643	14	1	5	36.9	8.6	*Gm*
	predicted proline-rich protein	7269684	13	1	2	55.6	5.8	*At*
	soluble acid invertase	47969540	14	2	4	71.7	5.2	*Hv*
	subtilisin-like protease	33621210	118	8	14	83.2	9.0	*Gm*
	subtilisin-like protease	86439745	15	2	3	76.5	9.3	*Ta*
	xyloglucan endotransglycosylase precursor	89145876	15	1	7	19.5	6.6	*Gm*

2.3.4 Lipid metabolism in the seed coat

Yoshida *et al.* (2000) reported the composition of lipids of soybean coats to be of 70% of TAGs, 17% phospholipids and 12% of other types of lipids, when studying the effect of microwave roasting on the lipid composition of seed coats. A later chemical analysis study revealed that the inner seed coat and outer seed coat layers may differ in their fatty acid composition. This study related to cutin deposition in the seed coat with around 30 ng mm^{-2} seed surface area in each layer for the particular Harosoy 63 cultivar that was also analyzed in the present study (Shao *et al.,* 2007). It is therefore not a surprise that our results confirm the existence of several enzymes involved in lipid metabolism and fatty acid biosynthesis in the proteome of 35-50 DPA soybean seed coats. A general biosynthetic pathway for fatty acids is presented in Figure 2.9.

Other enzymes involved in lipid metabolism were also found and are summarized in Table 2.2 with statistical details for each of the proteins involved in fatty acid metabolism identified in the seed coat proteome.

Figure 2.9. The fatty acid (FA) biosynthesis pathway in soybean seed coats (adapted from Lung and Weselake, 2006). In plastids, FA are synthesized from acetyl-CoA in a three-step process: (a) irreversible carboxylation of acetyl-CoA by the action of acetyl CoA carboxylase (ACCase) to form malonyl-CoA; (b) repeated condensations of malonyl-CoA with a growing acyl carrier protein (ACP)-bound acyl chain by the action of FA synthase complex (FAS), consecutively adding two carbon units to form 16:0-ACP; and (c) elongation and desaturation of 16:0-ACP to form 18:0-ACP and 18:1-ACP, respectively. The *de novo* synthesized FA enter the cytosolic pool in an esterified form known as acyl-CoA, which are synthesized by an ATP dependent esterification of FA and CoA thorugh the action of acyl-CoA synthetase (ACS). In the ER, the sequential incorporation of FA onto the glycerol backbone (Kennedy pathway). This pathway starts with the acyl-CoA-dependent acylation of *sn*-glycerol-3-phosphate to form lysophosphatidic acid through the action of *sn*-glycerol-3-phosphate acyltransferase (GPAT). The second acyl-CoA-dependent acylation is catalyzed by lysophosphatidic acid acyltransferase (LPAAT), leading to the formation of PA, which is dephosphorylated through the action of phosphatidate phosphatase (PAP) to form *sn*-1,2-DAG. The third acyl-CoA-dependent acylation catalyzed by acyl-CoA:diacylglycerol acyltransferase (DGAT) leads to the production of TAG. DAG is situated at the branch point of the pathway between TAG and membrane phospholipid formation. Cytidine diphosphate (CDP)-choline:1,2-DAG cholinephosphotransferase (CPT) catalyzes the transfer of a phosphocholine from CDP-choline into DAG and leads to the formation of PC and cytidine monophosphate. The acyl moiety at the *sn*-2 position of PC may undergo acyl exchange with the acyl-CoA pool by the reversible reaction catalyzed by lysophosphatidylcholine acyltransferase (LPCAT). Lysophosphatidylcholine can also be formed through hydrolysis of PC catalyzed by phospholipase A_2. PDAT catalyzes the acyl transfer from PC to DAG leading to the formation of lysophosphatidylcholine and TAG, whereas DGTA catalyzes the transfer of an acyl moiety between two DAG molecules to form TAG and monoacylglycerol (MAG). Δ^9-DES, Δ^9-desaturase; ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; ACS, acyl CoA synthetase; CPT, CDP-choline: 1,2-diacylglycerol cholinephosphotransferase; DGAT, diacylglycerol acyltransferase; DGTA, diacylglycerol transacylase; FAS, fatty acid synthase; GPAT, sn-glycerol-3-phosphate acyltransferase; LPA, lysophosphatidic acid; LPCAT, lysophosphatidylcholine acyltransferase; LPAAT, lysophosphatidic acid acyltransferase; LPC, lysophosphatidylcholine; PAP, phosphatidate phosphatase; PDAT, phospholipid:diacylglycerol acyltransferase; PLA_2, phospholipase A_2; TS, acyl-ACP thioesterase. Enzymes represented by a solid line oval were identified in the seed coat proteome. Those presented by a dashed line oval were not.

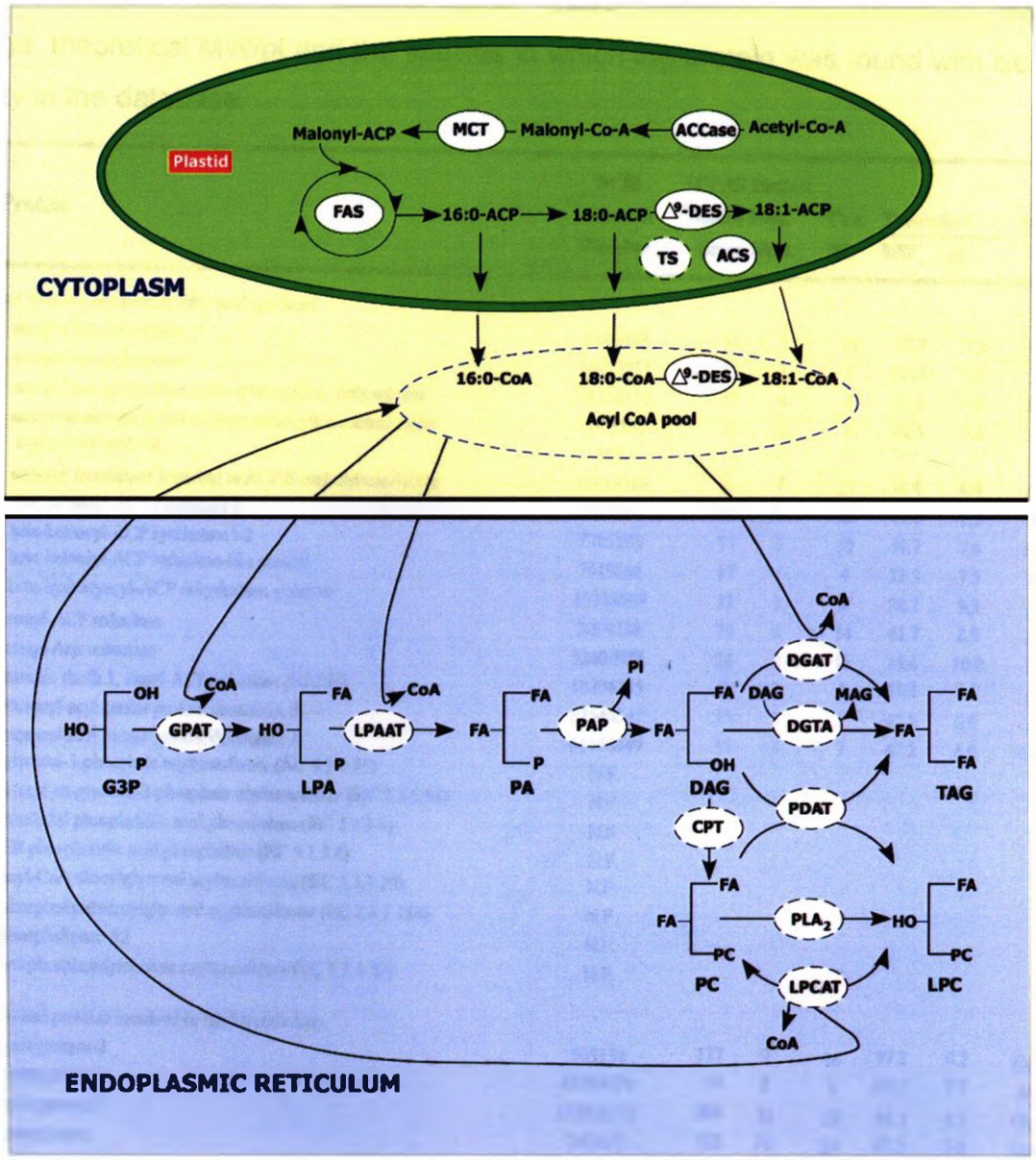

Plastid
Malonyl-ACP
MCT
Malonyl-Co-A
ACCase
Acetyl-Co-A
FAS
16:0-ACP
18:0-ACP
Δ9-DES
18:1-ACP
TS
ACS
CYTOPLASM
16:0-CoA
18:0-CoA
Δ9-DES
18:1-CoA
Acyl CoA pool
CoA
DGAT
PI
OH
CoA
FA
CoA
FA
FA
FA
DAG
MAG
FA
HO
GPAT
HO
LPAAT
FA
PAP
FA
DGTA
FA
P
P
P
OH
FA
G3P
LPA
PA
DAG
TAG
CPT
PDAT
FA
FA
PLA2
HO
FA
PC
PC
LPCAT
LPC
CoA
ENDOPLASMIC RETICULUM

Table 2.2. Seed coat enzymes involved in fatty acid metabolism. The table includes the NCBI nr accession number, MS/MS score, unique peptides, percentage of protein coverage, theoretical MW/pI and the species in which the protein was found with closest similarity in the database.

	Protein	NCBI Accession Number	MS/MS Search Score	Distinct Pept. Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
Enzymes and proteins involved in fatty acid synthesis								
ACCase	acetyl-CoA carboxylase	8886469	54	5	11	58.8	7.2	*Gm*
	acetyl-CoA carboxylase	14423251	15	2	3	120.6	6.0	*Zm*
	acetyl-CoA carboxylase carboxyltransferase beta subunit	91214152	57	4	9	49.0	4.8	*Gm*
	acetyl co-enzyme A carboxylase carboxyltransferase alpha	4895181	13	2	4	88.5	5.8	*At*
ACP	acyl carrier proteins	N.F.						
MCT	malonyl transferase homolog to ACP-S-malonyltransferase	82618886	92	7	27	36.4	6.4	*Gm*
KAS	beta-ketoacyl-ACP synthetase I	7385201	132	9	30	49.7	7.2	*Gm*
	beta-ketoacyl-ACP synthetase I-2	7385203	57	4	19	49.7	7.6	*Gm*
	beta-ketoacyl-ACP reductase-like protein	7019664	17	1	4	32.3	7.5	*At*
	beta-hydroxyacyl-ACP dehydratase, putative	15238069	17	1	3	24.1	9.3	*At*
	enoyl-ACP reductase	2204236	79	6	14	41.7	8.9	*Nt*
	enoyl-Acp reductase	32400828	28	2	16	15.6	10.0	*Ta*
	mosaic death 1; enoyl-ACP reductase (NADH)	18396215	41	3	8	41.2	9.1	*At*
	stearoyl-acyl carrier protein desaturase B	62546347	35	3	7	47.2	6.0	*Gm*
	stearoyl-acyl carrier protein desaturase A	62546349	35	3	7	47.2	6.0	*Gm*
GPAT	glycerol-3-phosphate acyltransferase (EC 2.3.1.15)	N.F.						
LPAAT	1-acyl-sn-glycerol-3-phosphate acyltransferase (EC 2.3.1.51)	N.F.						
PAP	plastidial phosphatidic acid phosphatase (EC 3.1.3.4)	N.F.						
DGAT	ER phosphatidic acid phosphatase (EC 3.1.3.4)	N.F.						
DGTA	acyl-CoA:diacylglycerol acyltransferase (EC 2.3.1.20)	N.F.						
PDAT	phospholipid:diacylglycerol acyltransferase (EC 2.3.1.158)	N.F.						
PLA_2	phospholipase A2	N.F.						
LPCAT	lysophosphatidylcholine acyltransferase (EC 2.3.1.23)	N.F.						
Other enzymes and proteins involved in lipid metabolism								
	lipoxygenase-2	505138	127	9	16	97.2	6.2	*Gm*
	lipoxygenase 3	18394479	19	2	3	103.7	7.7	*At*
	lipoxygenase-9	152926332	304	22	32	96.3	6.5	*Gm*
	lipoxygenase	242462	182	11	24	67.3	5.6	*Gm*
	lipoxygenase	2598612	50	5	6	97.7	6.0	*Ps*
	lipoxygenase	2459611	37	3	4	97.4	6.1	*Ps*
	lipoxygenase	541746	25	2	3	97.1	6.3	*Ps*
	lipoxygenase	493730	24	2	5	97.0	6.1	*Ps*
	lipoxygenase	6002055	19	2	3	103.7	7.7	*At*
	lipoxygenase	9665131	19	2	3	102.9	7.7	*At*
	lipoxygenase	12620877	18	2	3	96.4	5.7	*Zm*
	lipoxygenase	1407703	16	2	3	96.9	5.5	*St*
	lipoxygenase	1495806	15	2	9	51.2	5.9	*St*
	lipoxygenase, putative	15218506	15	2	3	104.8	7.1	*At*
	lipoxygenase, putative	12323766	15	2	4	79.5	5.8	*At*
	phosphatidylserine decarboxylase	29465780	15	2	5	50.2	9.3	*Ls*
	ataxia telangiectasia-mutated and RAD3-related	18422029	14	2	1	30.2	6.6	*At*
	phosphatidylinositol 3- and 4-kinase family protein	18407090	14	2	4	62.6	5.8	*At*
	phosphatidylinositol 3/4-kinase family protein	22329206	43	3	1	43.3	6.8	*At*

Table continues from previous page

Table continued next page

Protein	NCBI Accession Number	MS/MS Search Score	Distinct Pept. Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
phosphatidylinositol-4-phosphate 5-kinase family protein	15230176	24	3	6	89.2	6.8	*At*
contains a phosphatidylinositol kinase domain	8569097	24	3	1	28.2	6.4	*At*
allene oxide cylase	14423351	51	3	12	26.5	8.7	*Ls*
allene oxide synthase	82795997	51	4	9	58.8	9.2	*Gm*
24 kDa oleosin isoform	18720	59	5	33	15.8	8.2	*Gm*
3-ketoacyl-CoA thiolase	62321535	36	3	12	35.7	8.8	*At*
3-ketoacyl-CoA thiolase	37549269	128	8	26	47.0	8.8	*Gm*
acyl-CoA oxidase	15553478	38	4	6	74.3	7.3	*Gm*
Other enzymes and proteins involved in lipid metabolism							
3-hydroxybutyryl-CoA dehydrogenase, putative	15232545	20	1	5	31.7	6.6	*At*
long-chain-fatty-acid-ACP ligase	22328609	17	2	3	81.5	8.9	*At*
AMP-dependent synthetase and ligase family protein	15218839	16	2	3	64.9	7.6	*At*
esterase, putative	15227376	36	3	5	31.7	5.9	*At*
fatty acyl coA reductase	22003082	16	2	4	57.5	8.8	*Ta*
inorganic pyrophosphatase-like protein	21593570	99	7	27	24.6	5.3	*At*
lipase class 3 family protein	2244965	17	2	3	75.7	5.5	*At*
lipase, putative	13569989	26	3	4	56.8	8.7	*Os*
peroxisome defective 1, acetyl-CoA C-acyltransferase	15225798	36	3	9	48.6	8.6	*At*
acetyl-CoA synthetase, putative	12323178	17	2	1	290.1	5.8	*At*
Other enzymes and proteins involved in lipid catabolism							
GDSL-motif lipase/hydrolase family protein	15241404	21	2	5	43.6	8.7	*At*
GLIP7 (GDSL-motif lipase 7)	9755617	24	3	5	40.5	8.7	*At*
phospholipase D alpha 1	2499708	14	1	1	92.2	5.4	*Zm*
phospholipase D alpha 1	15232671	28	2	2	91.8	5.5	*At*
phospholipase D alpha	6573119	17	1	1	92.2	5.4	*Ls*
abnormal inflorescence meristem, enoyl-CoA hydratase	15235527	22	2	2	77.9	9.4	*At*
enoyl-CoA hydratase/isomerase	79473201	19	2	7	46.3	6.2	*At*
enoyl-CoA hydratase/isomerase family protein	42565158	17	1	2	45.7	6.1	*At*
enoyl-CoA hydratase/isomerase family protein	30683577	15	1	4	28.8	9.1	*At*
MFP2 (multifunctional protein), enoyl-CoA hydratase	15231317	23	2	3	78.8	9.2	*At*

2.3.5 Isoflavonoids synthesis in the seed coat

A proposed phenylpropanoid pathway and several of the branch groups based on the related enzymes found in soybean seed coats is presented in Figure 2.10 based on previous studies by (Winkel-Shirley, 2001; Dhaubhadel et al., 2008; Yu and McGonigle, 2005). The phenylpropanoid and related isoflavonoid pathways are well represented by protein matches. Most steps appear to operate despite that pathway flux is restricted by lack of chalcone synthase (CHS) which is known to control seed coat pigmentation (yellow or black) (Tuteja et al., 2004). The list of proteins and enzymes related to phenylpropanoid metabolism found in the seed coat proteome along with their statistical information is presented in Table 2.3.

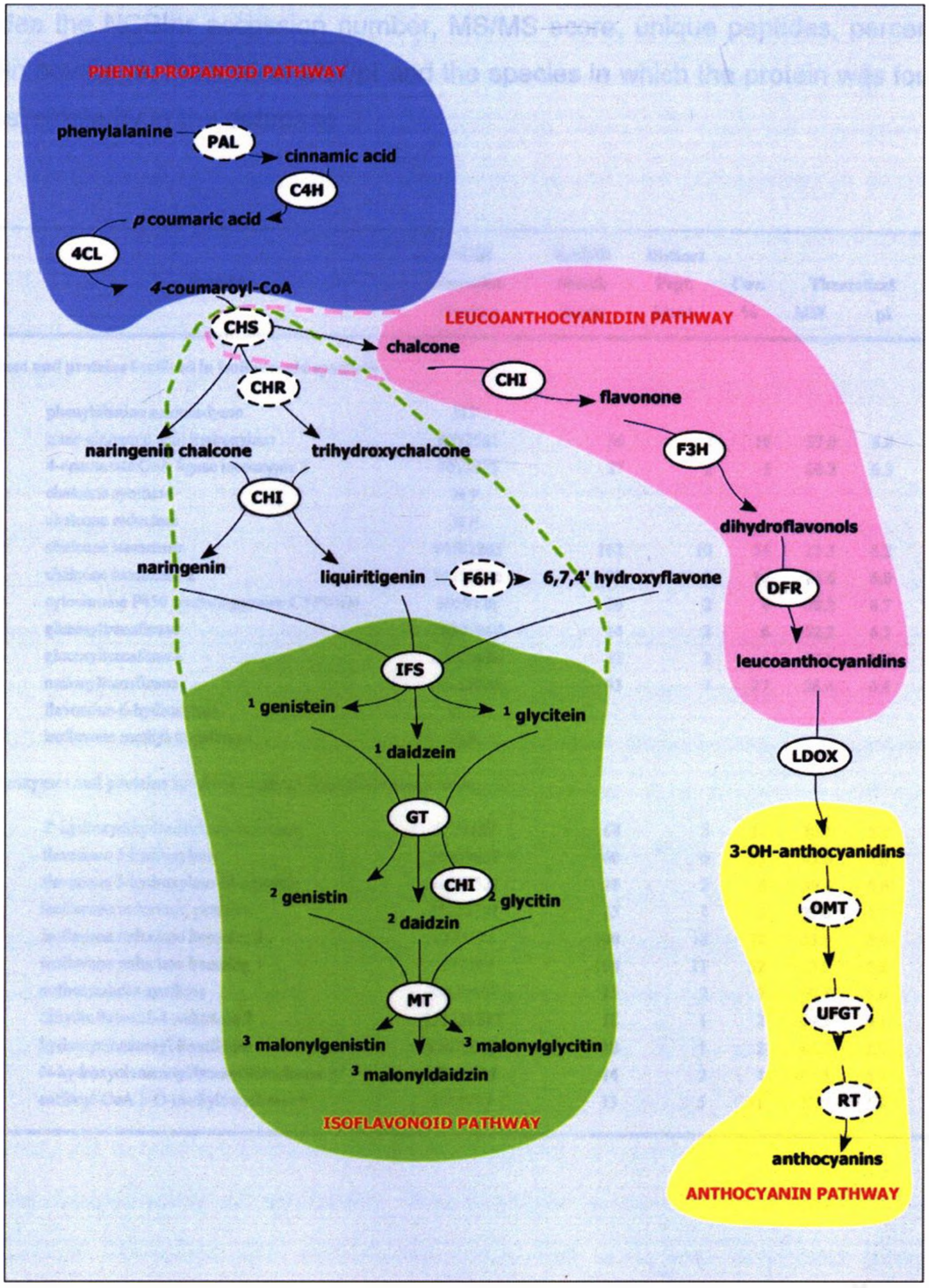


Figure 2.10. A scheme of the major branch pathways of the phenylpropanoid biosynthesis in soybean seed coats modified from (Winkel-Shirley 2001) and (Dhaubhadel *et al.*, 2003). The pathways for the isoflavonoid and leucoanthocyanidin groups are shown in green and pink backgrounds. The anthocyanin group is absent from soybean. Enzymes represented by a solid line oval were identified in the seed coat proteome. Those presented by a dashed line oval were not.

Table 2.3. Seed coat enzymes involved in phenylpropanoid metabolism. The table includes the NCBInr accession number, MS/MS score, unique peptides, percentage of protein coverage, theoretical MW/pI and the species in which the protein was found with closest similarity in the database.

	Protein	NCBI Accesion Number	MS/MS Search Score	Distinct Pept. Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
Enzymes and proteins involved in isoflavonoid synthesis								
PAL	phenylalanine amonia-lyase	N.F.						
C4H	trans-cinnamic acid hydroxylase	9957081	36	3	10	57.0	8.8	*Ps*
4CL	4-coumarate:CoA ligase isoenzyme 2	4038975	27	2	5	60.2	6.3	*Gm*
CHS	chalcone synthase	N.F.						
CHR	chalcone reductase	N.F.						
CHI	chalcone isomerase	14582263	162	10	56	23.3	6.2	*Gm*
	chalcone isomerase 2	51039626	33	2	15	24.6	6.0	*Gm*
IFS	cytochrome P450 monooxygenase CYP93D1	5059126	20	2	4	58.2	8.7	*Gm*
GT	glucosyltransferase	28302068	14	2	6	52.2	6.3	*Gm*
	glucosyltransferase	82618888	32	2	4	53.1	6.7	*Gm*
MT	malonyltransferase	82618886	92	7	27	36.4	6.4	*Gm*
F6H	flavonone-6-hydroxylase	N.F.						
IMT	isoflavone methyl-transferase	N.F.						
Other enzymes and proteins involved in phenylpropanoid metabolism								
	2'-hydroxydihydrodaidzein reductase	6573167	68	5	17	36.1	5.7	*Gm*
F3H	flavanone 3-hydroxylase	51039637	60	6	21	42.6	5.6	*Gm*
	flavanone 3-hydroxylase-like protein	21553527	18	2	6	39.4	5.6	*At*
IFR	isoflavone reductase, putative	15222191	15	1	3	35.6	6.1	*At*
	isoflavone reductase homolog 2	6573171	300	18	72	33.9	5.6	*Gm*
	isoflavone reductase homolog 1	6573169	198	11	52	33.9	5.8	*Gm*
	anthocyanidin synthase	38679407	15	2	7	40.1	5.6	*Gm*
DFR	dihydroflavonol-4-reductase 2	121755811	18	1	2	39.5	6.1	*Gm*
	hydroxycinnamoyl transferase	27475616	18	1	2	48.2	5.9	*Nt*
	N-hydroxycinnamoyl/benzoyltransferase 4	83853813	14	2	4	52.5	6.5	*Gm*
	caffeoyl-CoA 3-O-methyltransferase 5	2511737	13	5	32	27.2	5.4	*Nt*

2.3.6 Proteolysis in the seed coat

Many proteolytic processes take place in both the embryo and seed coat (see Chapter 1). It is therefore expected that the seed coat be equipped with a wide complement of hydrolyzing enzymes as well as proteases. In accordance with this expectation, we find a large group of proteases in the seed coat which are reported in Table 2.4. What is surprising is the diversity of proteolytic enzymes in the seed coat. The emerging picture is that plant proteases are key regulators of a striking variety of biological processes, including meiosis, gametophyte survival, embryogenesis, seed coat formation, cuticle deposition, epidermal cell fate, stomata development, chloroplast biogenesis, and local and systemic defense responses (van der Hoorn, 2008).

The classification of 98 seed coat proteases in the five major catalytic classes threonine, cysteine, serine, aspartic and metallopeptidases proteolytic groups is reported in Figure 2.11. This classification was based on the MEROPS, a peptidase database, in which they have been subdivided into families and clans on the basis of evolutionary relationships (http://merops.sanger.ac.uk) (Rawlings et al., 2006). The largest group of seed coat proteases falls in the cysteine type, which are generally known to play a role in programmed cell death in response to both developmental cues and pathogens, although, they can also regulate epidermal cell fate, flowering time and pollen or embryo development. The genomes of rice and *Arabidopsis* encode 678 and 826 proteases respectively, with serine proteases as the dominant group (van der Hoorn, 2008), whereas, 23 soybean proteases were reported, as of May 2008, prior to whole genome sequencing of soybean. This number is obviously under represented and further searches are necessary once the complete soybean genome is available in an adequate format for proteomic data mining.

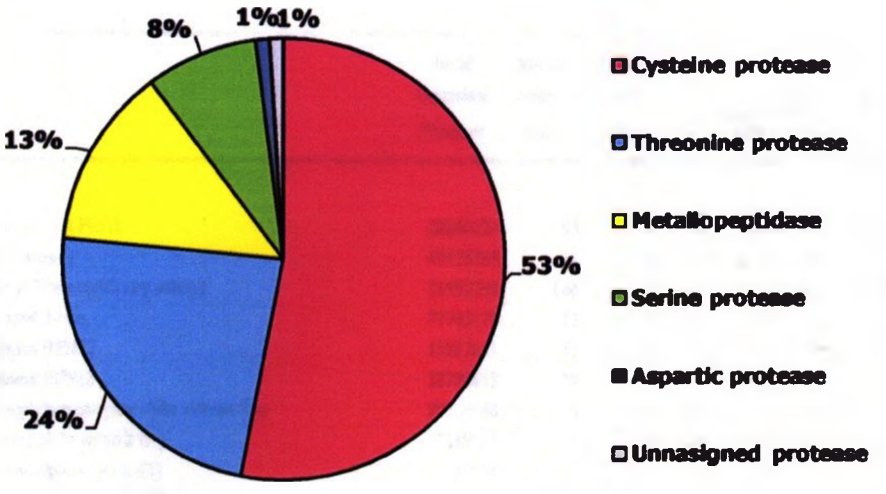


Figure 2.11. Distribution of 98 seed coat proteases over the different catalytic classes. Classification was based on that reported in the peptidase database MEROPS (Rawlings et al., 2006).

Table 2.4. Soybean seed coat proteases present at 35-50 DPA. Table includes the NCBI nr accession number, MS/MS score, unique peptides, percentage of protein coverage, theoretical MW/pI and the sp. in which the protein was found with closest similarity in the database. Classification was done according to the catalytic activity as reported on MEROP protease database (http://merops.sanger.ac.uk) (Rawlings et al., 2006).

Protein	NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
Threonine protease							
20S proteasome beta subunit PBB2	20260224	95	6	11	29.6	6.7	*At*
26S proteasome beta subunit	49175785	22	2	29	6.0	10.2	*Ps*
26S proteasome non-ATPase regulatory subunit	21592398	146	9	42	34.4	6.4	*At*
26S proteasome subunit 4-like	77745479	72	5	14	49.6	6.1	*St*
26S proteasome subunit RPN12	15217661	35	2	10	30.7	4.8	*At*
26S proteasome subunit RPN1b	32700812	38	3	5	98.0	5.1	*At*
multicatalytic endopeptidase complex alpha subunit-like	20260140	96	6	25	27.3	5.4	*At*
PAB1 (20S proteasome alpha subunit B1)	15219317	113	6	26	25.7	5.5	*At*
PAE1 (20S proteasome alpha subunit E1)	15220961	125	8	46	25.9	4.7	*At*
PAG1 (20S proteasome alpha subunit G1)	15225839	37	3	10	27.4	5.9	*At*
PBA1 (20S proteasome beta subunit A 1)	79325892	81	5	18	25.3	5.3	*At*
PBC1 (20S proteasome beta subunit C1)	21553663	58	4	20	22.8	5.3	*At*
PBD1 (proteasome subunit PBGB)	15228805	60	4	20	22.5	6.0	*At*
PBE1 (20S proteasome beta subunit E1)	14594931	96	6	45	18.6	9.2	*Nt*
PBG1 (20S proteasome beta subunit G1)	15223537	21	1	7	27.7	6.1	*At*
proteasome alpha subunit-like protein	76160982	111	7	33	28.1	5.4	*St*
proteasome-like protein alpha subunit	77999287	125	8	36	27.1	7.0	*St*
proteosome subunit	608387	32	2	10	25.3	7.8	*At*
putative 26S proteasome ATPase subunit	6056389	14	2	5	50.3	5.2	*At*
putative alpha7 proteasome subunit	14594925	98	7	29	27.2	6.1	*Nt*
putative beta7 proteasome subunit	14594935	41	3	23	14.7	8.2	*Nt*
RPT5B (26S proteasome AAA-ATPase subunit RPT5B)	15217431	293	18	47	47.0	4.9	*At*
RPT1A (regulatory particle triple-A 1A)	15220930	131	9	26	47.8	6.3	*At*
Cysteine protease							
AESP (aspraae)	79482708	18	2	1	244.8	6.8	*At*
ATP-dependent Clp protease ClpB protein-related	145323770	17	2	3	107.8	8.1	*At*
cathepsin B-like cysteine protease, putative	18378947	15	1	3	40.0	6.5	*At*
ClpC (Clp protease ATP binding subunit)	2921158	38	3	4	103.5	6.3	*At*
CLPP5 (nuclear encoded CLP protease 1)	18378982	27	2	9	32.4	8.4	*At*
CLPX (Clp protease regulatory subunit X)	18423503	22	2	3	62.0	7.6	*At*
CUL2 (cullin 2)	22329305	18	2	2	86.0	7.3	*At*
cysteine protease TDI-65	5726641	14	1	3	51.1	5.9	*La*
cysteine proteinase	479060	27	2	8	41.6	6.0	*Gm*
cysteine proteinase	31559530	25	2	9	40.1	6.1	*Gm*
cysteine proteinase inhibitor	1944319	82	5	29	27.6	7.3	*Gm*
cysteine proteinase inhibitor	1277164	56	4	36	10.3	5.9	*Gm*
cysteine proteinase inhibitor	1277168	25	2	27	11.1	5.8	*Gm*
cysteine-type peptidase	15241982	17	2	6	56.1	4.7	*At*
DJ-1 family protein / protease-related	15232958	14	1	2	41.6	5.2	*At*
DNA-damage inducible protein DDI1-like	21537297	38	3	7	45.4	4.8	*At*
ubiquitin	1762935	13	1	23	8.7	8.1	*Nt*
ubiquitin activating enzyme E1	1808656	56	4	5	120.3	5.4	*Nt*
ubiquitin carboxyl-terminal hydrolase	42566353	14	2	6	46.6	9.5	*At*

Table continues from previous page

Table continued next page

	Protein	NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	ubiquitin carboxyl-terminal hydrolase-related	15242114	15	2	2	132.2	5.9	*At*
	ubiquitin carboxyl-terminal hydrolase-related	42572001	15	2	1	130.9	5.5	*At*
	ubiquitin conjugating enzyme-like	82623381	25	2	11	21.4	5.0	*Si*
	ubiquitin family protein	15232924	25	2	5	44.2	4.8	*At*
	ubiquitin fusion-degradation protein-like	76160972	63	5	16	35.5	6.2	*Si*
	ubiquitin isopeptidase T	11994150	14	1	1	88.4	5.0	*At*
	ubiquitin-conjugating enzyme family protein-like protein	76160962	92	6	34	16.6	6.2	*Si*
	ubiquitin-conjugation enzyme	22597164	46	3	25	16.4	7.7	*Gm*
	ubiquitin-specific protease 6	11993465	14	1	1	53.7	5.8	*At*
	ECR1 (E1 C-terminal related 1)	18419850	19	1	2	50.5	5.5	*At*
	F1E22.3 (ubiquitin carboxyl-terminal hydrolase 1)	6686411	17	1	6	41.4	5.2	*At*
	UBC12 (ubiquitin-conjugating enzyme 12)	18398288	14	1	6	16.7	7.7	*At*
	UBC30 (ubiquitin-protein ligase)	18423829	16	1	7	16.5	6.8	*At*
	UBC36, ubiquitin-protein ligase	18394416	142	9	75	17.2	6.7	*At*
	UBC9 (ubiquitin conjugating enzyme 9)	18417097	30	2	19	20.2	7.0	*At*
	UPL2 (ubiquitin-protein ligase 2)	15223117	14	2	0	403.6	4.8	*At*
	Ulp1 protease family protein	15232756	23	3	4	94.1	5.5	*At*
	Ulp1 protease family protein	15242433	18	2	2	105.8	8.6	*At*
	Ulp1 protease family protein	15229144	14	2	2	146.5	5.3	*At*
	Ulp1 protease family protein	15234224	14	2	2	82.1	5.4	*At*
	polyubiquitin	3452083	101	6	41	12.8	9.7	*Gm*
	MMZ4 (MMS zwei homologue 4)	18409633	44	3	22	16.5	6.2	*At*
	putative ubiquitin protein ligase	13174246	16	2	3	84.5	6.4	*Os*
	SUM2 (small ubiquitin-like modifier 2)	15240471	43	3	25	11.7	5.4	*At*
	thiol protease isoform B	1619903	40	3	11	35.0	7.6	*Gm*
	thiolprotease	3980198	18	1	2	51.3	6.1	*Ps*
	RUB1 (related to ubiquitin 1)	30692436	43	3	17	17.4	5.8	*At*
	RUB1-conjugating enzyme-like protein	76573335	26	2	8	21.0	8.3	*Si*
	PEX4 (peroxin 4)	18420949	24	2	17	17.7	8.4	*At*
	unknown protein, contains peptidase C12 domain	15238875	21	2	5	49.7	8.2	*At*
	unknown protein, with PPPDE putative peptidase domain	15231383	17	1	5	28.8	5.6	*At*
	phosphatidylinositol 3- and 4-kinase family protein	18407090	14	2	4	62.6	5.8	*At*
	ICS1 (isochorismate synthase 1) (ubiquinone biosynthesis)	42572105	20	2	4	69.0	6.1	*At*
Serine protease								
	leucine-rich repeat family protein	15220080	18	2	3	110.2	6.0	*At*
	SFC (scarface, vascular network defective 3)	3914085	17	2	2	97.7	7.7	*Zm*
	subtilase family protein	42567017	25	3	5	78.5	9.4	*At*
	subtilase family protein	18416719	25	3	3	82.9	6.2	*At*
	subtilase family protein	18423316	14	1	1	85.0	9.4	*At*
	subtilisin-like protease	33621210	118	8	14	83.2	9.0	*Gm*
	subtilisin-like protease	86439745	15	2	3	76.5	9.3	*Ta*
	subtilisin-type protease precursor	11611651	397	25	40	82.7	6.9	*Gm*

Table continued next page

Table continues from previous page

Protein	NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
Metallopeptidase							
ATPREP1/ATZNMP (presequence protease 1)	22331173	35	3	4	121.0	5.5	*At*
cytosol aminopeptidase family protein	15235763	84	5	9	61.3	6.6	*At*
FTSH3 (FtsH protease 3)	30684118	33	3	3	89.4	6.8	*At*
Ftsh-like protease	50892959	22	2	2	86.9	6.2	*Ps*
FtsH-like protein Pftf precursor	4325041	39	3	6	74.4	6.0	*Nt*
mitochondrial processing peptidase	587564	14	1	2	59.3	6.2	*St*
mitochondrial processing peptidase alpha subunit, putative	21594084	23	2	3	54.5	5.9	*At*
MPPALPHA (mitochondrial processing peptidase alpha subunit)	14334534	30	3	5	54.1	6.2	*At*
peptidase	12324166	18	2	8	34.9	9.0	*At*
peptidase M20/M25/M40 family protein	42566909	25	2	9	48.2	5.2	*At*
TPP2 (tripeptidyl peptidase II)	5262775	20	2	1	154.2	6.1	*At*
urease	14599161	20	2	2	90.7	5.8	*Gm*
urease	14599413	16	2	2	89.8	5.6	*St*
Aspartic protease							
pepsin A	79507883	16	1	2	48.7	9.7	*At*
Unassigned protease							
ATHMOV34 (asymmetric leaves enhancer 3)	77745499	84	5	18	34.8	5.9	*St*

2.3.7. C_1 metabolism-related enzymes in the seed coat

In the comprehensive proteomic list presented in Appendix II and the relative gene ontology assignment (Figure 2.1) it is shown that an important fraction of the whole protein complement of seed coats is committed to amino acid, nucleotide and protein synthesis. In fact, from a complement of 1705 proteins, 136 enzymes were found to be involved in amino acid metabolism (8%), 51 in nucleotide metabolism (3%) and 130 in protein synthesis (8%), rising to as high as 19% of the identified proteome committed to C_1 related pathways. This is evidence of the high metabolic activity of the seed coat at maturity. Figure 2.12 presents an overview of C_1 metabolism in soybean seed coats. This is an adaptation from (Hanson, 2001), who presented biochemical and DNA evidence, to which we add protein evidence of C1 metabolism found in seed coats. The enzymes involved in the proposed pathway are summarized, along with statistical information, in Table 2.5. The enzyme methionine synthase, with 39 unique peptides matching the database entry, was found to be the most abundant protein in the seed coat proteome at 35-50 DPA.

Figure 2.12. Reactions of plant C_1 metabolism in soybean seed coats. The principal sources of C_1 units are boxed and highlighted in yellow. For simplicity, enzymes are numbered and listed in Table 2.5. The question mark shows the reaction catalyzed by 5-forminino-THF cyclodeaminase, for which there is as yet no evidence in plants. Enzymes represented by a solid line oval were identified in the seed coat proteome. Those presented by a dashed line oval were not. The double question mark shows the reduction of formaldehyde to methanol, which occurs in vivo but the enzymatic basis is uncertain. The double asterisks mark adenosine salvage reactions. Substrate abbreviations: 10-CHO-THF, 10-formyl-THF; 5-CHO-THF, 5-formyl-THF; 5-10=CH-THF, 5,10-methenyl-THF; 5,10-CH_2-THF, 5,10-methylene-THF; 5-CH_3-THF, 5-methyl-THF; 5-CH=NH-THF, 5-formino-THF; DHF, dihydrofolate; Hcy, homocysteine; Met, methionine; GSH, glutathione; HM-GSH, *S*-hydroxymethylglutathione; FGAR, formylglycinamide ribonucleotide; FAICAR, formamidoimidazolecarboxamide ribonucleotide; CH=NH-Glu, formiminoglutamate. Pathways based on (Hanson and Roje, 2001).

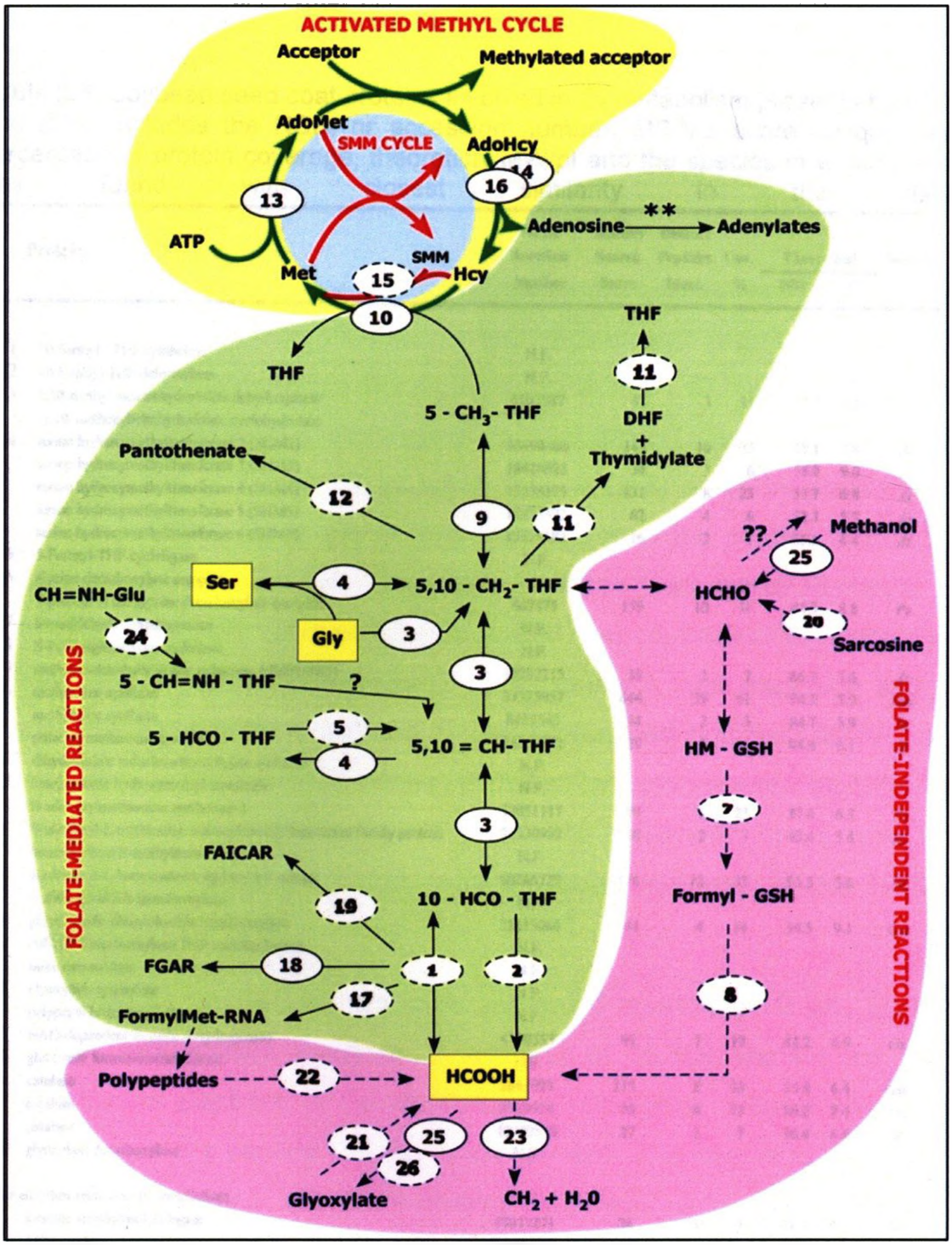

ACTIVATED METHYL CYCLE
Acceptor
Methylated acceptor
AdoMet
SMM CYCLE
AdoHcy
16
13
ATP
Adenosine
**
Adenylates
SMM
Met
15
Hcy
10
THF
THF
11
5 - CH3- THF
DHF
+
Thymidylate
Pantothenate
12
9
11
??
Methanol
25
Ser
4
5,10 - CH2- THF
HCHO
CH=NH-Glu
Gly
3
20
Sarcosine
24
3
5 - CH=NH - THF
?
FOLATE-MEDIATED REACTIONS
FOLATE-INDEPENDENT REACTIONS
5 - HCO - THF
5
4
5,10 = CH- THF
HM - GSH
3
7
FAICAR
Formyl - GSH
19
10 - HCO - THF
FGAR
18
1
2
17
8
FormylMet-RNA
Polypeptides
22
HCOOH
21
25
23
26
Glyoxylate
CH2 + H20

Table 2.5. Soybean seed coat proteins involved in C_1 metabolism present at 35-50 DPA. The table includes the NCBI nr accession number, MS/MS score, unique peptides, percentage of protein coverage, theoretical MW/pI and the species in which the protein was found with closest similarity in the database.

	Protein	NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1	10-formyl - THF synthetase	N.F.						
2	10-Formyl-THF deformylase	N.F.						
3	5,10-methylenetetrahydrofolate dehydrogenase -5,10-methenyltetrahydrofolate cyclohydrolase	4103987	48	3	11	31.3	8.3	*Ps*
4	serine hydroxymethyltransferase 2 (SHM2)	30690400	143	10	23	59.1	8.8	*At*
	serine hydroxymethyltransferase 3 (SHM3)	18418028	34	3	6	58.0	9.0	*At*
	serine hydroxymethyltransferase 4 (SHM4)	15236375	131	8	23	51.7	6.8	*At*
	serine hydroxymethyltransferase 5 (SHM5)	15236371	62	4	6	52.3	5.7	*At*
	serine hydroxymethyltransferase 6 (SHM6)	15219182	16	2	1	66.6	6.4	*At*
5	5-Formyl-THF cycloligase	N.F.						
6	glycine decarboxylase complex							
	T-protein of the glycine decarboxylase complex	407475	159	10	31	44.3	8.8	*Ps*
7	formaldehyde dehydrogenase	N.F.						
8	S-Formylglutathione hydrolase	N.F.						
9	methylenetetrahydrofolate reductase 1 (MTHFR1)	15232215	38	3	7	66.3	5.6	*At*
10	methionine synthase	33325957	644	39	61	84.3	5.9	*Gm*
	methionine synthase	8439545	34	2	3	84.7	5.9	*St*
	putative methionine synthase	14532772	29	3	6	84.6	6.1	*At*
11	dihydrofolate reductase/thymidylate synthase	N.F.						
12	ketopantoate hydroxymethyltransferase	N.F.						
13	S-adenosylmethionine synthetase-2	37051117	89	6	23	37.6	6.3	*Ps*
14	S-adenosyl-L-methionine:carboxyl methyltransferase family protein	22330992	16	2	3	43.4	5.4	*At*
15	homocysteine S-methyltransferase	N.F.						
16	S-adenosyl-L-homocysteine hydrolase 1 mutant	60266729	176	13	32	53.5	5.6	*At*
17	methionyl-tRNA transformylase	N.F.						
18	glycinamide ribonucleotide transformylase	32815066	44	4	14	34.5	9.1	*Gm*
19	AICAR Transformylase/ IMP cyclohydrolase	N.F.						
20	sarcosine oxidase	N.F.						
21	glyoxylate synthetase	N.F.						
22	polypeptide deformylase	N.F.						
23	NAD-dependent formate dehydrogenase	4760553	91	7	19	41.2	6.9	*Os*
24	glutamate formiminotransferase	N.F.						
25	catalase	2661023	115	8	33	35.8	6.4	*Gm*
	catalase	3929924	96	6	13	56.2	7.4	*Os*
	catalase	40950550	27	2	7	56.4	6.6	*St*
26	glyoxylate decarboxylase	N.F.						
Other enzymes related to 1C metabolism								
	formate tetrahyrofolate ligase	17017271	36	3	9	31.6	7.2	*Zm*
	ACC oxidase	25989506	15	2	6	35.6	6.1	*St*
	ethylene response sensor 1	15226788	17	2	10	68.3	6.1	*At*
	ethylene responsive protein	33331083	24	3	8	42.2	5.1	*Gm*
	ethylene-responsive RNA helicase	15231074	57	4	7	69.2	7.7	*At*
	LEA14 (late embryogenesis abundant 14)	15223413	24	2	10	16.5	4.7	*At*
	S-adenosyl-L-homocysteine hydrolase 1 mutant	60266729	176	13	32	53.5	5.6	*At*
	S-adenosylmethionine synthetase-2	37051117	89	6	23	37.6	6.3	*Ps*
	seed maturation protein PM22	4585271	37	3	19	16.7	5.2	*Gm*
	seed maturation protein PM24	6648964	13	1	6	26.8	5.1	*Gm*
	seed maturation protein PM31	4838149	32	2	17	17.7	6.1	*Gm*
	seed maturation protein PM34	9622153	88	6	23	31.8	6.6	*Gm*
	seed maturation protein PM37	5802244	71	5	15	46.3	5.9	*Gm*

2.4. Discussion

2.4.1 Protein identification of seed coat proteins using a combined pre-fractionation and iterative exclusion lists

In this study, the seed coat proteome of physiologically mature soybean seeds (35-50 DPA) was pre-fractionated by one and two-dimensional SDS PAGE. The tryptic digests were submitted for ESI-LC MS/MS. The use of trypsin allowed the specific cleavage of proteins at the C-terminals of the amino acids lysine and arginine, except when they were followed by proline, which could have implications in the overall composition of the peptide mixture. We used iterative exclusion lists to identify peptides from the SDS-PAGE pre-fractionations, with a total of 10 bands and two replicates. In round 1 (no exclusion) 504 proteins were identified, in round 2 (first exclusion list) 278 proteins more (34 % increase) and in round 3 (second exclusion list) 37 other proteins were identified (5%). Protein identification was increase by 39% using two exclusion rounds. We postulate that these increments could have been greater had the peptides found more matching proteins in the NCBI nr database.

A total of 149,346 spectra were generated combining 1D and 2D-SDS-PAGE pre fractionation methods. A sub set of 15,368 spectra led to the identification of peptides in the database, which leaves around 90% of the spectra without any peptide or protein assignment. This seems to be a normal scenario when dealing with plant protein databases, only about 10% of the spectral data acquired by MS/MS methodology gets peptide identification (Dr. Martina Stromvik, personal communication). Considering the amount of spectral data acquired, it would be necessary to search for further identities once the soybean genome and predicted proteome becomes available.

A total of 1705 proteins from 35-50 DPA soybean seed coats were identified and functionally classified (Appendix II, Figure 2.6). To classify the

proteins the assumption was that proteins sharing functional domains have the same activity. It should be stressed that the function of a relatively small portion of the identified proteins has ever been experimentally demonstrated and the assumption that proteins sharing functional domains have the same biological function can become invalid in cases where the protein has more than one functional domain. Of course, this classification has to evolve to take into account new results obtained by other experimental approaches such as biochemistry or genetics. At present, such a classification can be proposed for 1705 proteins from soybean seed coats. To our knowledge, this is the most extensive report on the protein complement of a plant organ. Microarray studies have revealed the presence of at least 1,382 up regulated genes in seed coats of *Medicago truncatula* (Gallardo et al., 2007) and 15,683 genes expressed in the soybean seed coat during development, from which 4,860 were expressed uniquely in this organ (M. Gijzen, unpublished).

From the comparison of published proteomic reports on different plant organs (Figure 2.7), we observe that there is a general distribution among functional classes, although metabolic proteins is the most heavily represented, in correspondence with the general trend presented in other organs such as *Arabidopsis* roots (Mooney et al., 2006) and soybean seeds (Hajduch et al., 2005). No single particular specialization was noted for seed coats, as it was for the energy class of proteins in the case of soybean leaves (Xu et al., 2006a).

Based on the functional categories of proteins, the classes of transcriptional control and metabolism are particularly well represented (Figure 2.6). In a comparison of the transcriptome and proteome of *Medicago truncatula* mature seeds, it was found that genes related to transcription and RNA processing was unregulated (Gallardo et al., 2007). These are believed to contribute to the stored mRNA pool used for protein synthesis during germination (Rajjou et al., 2004) and are an indication of the potential of germination performance in the case of cotyledons. Our results suggest that the transcriptional control taking place in the seed coat is related to cell differentiation

and fate, given the remarkable cell diversity present in this organ (Gijzen, 1999; Miller, 1999; Yaklich, 1990).

Seed development is highly anabolic and it is not surprising to find the metabolic class as the most represented. One indicator of metabolic activity is methione synthase (#77 in Appendix II, NCBI accession number 33325957), which is the most abundant protein in the seed coat proteome (Appendix II) and suggests its commitment to anabolism. The same enzyme was found during high metabolic activity of germinating *Arabidopsis* seeds (Gallardo *et al.*, 2002) once metabolic activity was resumed after imbibition. The metabolic activity in soybean seed is about to decline when the seed enters quiescence, as previously found in seeds of *Brassica napus* (Hajduch et al., 2006b), soybean (Hajduch et al., 2005), *Arabidopsis* (Ruuska et al., 2002) and *Medicago truncatula* (Gallardo et al., 2003) seeds. The commitment of the seed coat proteins to C_1 and Met metabolism will be discussed later.

Other functional groups such as protein storage, energy related and defense related proteins are important and will be discussed in the light of mainstream pathways. Our data demonstrates that the proteins in the seed coat carry out a diversity of roles as per their functional classification. Our results show an unprecedented wealth of proteins present in the seed coat, both in terms of number as well as in diversity of functions. From a preliminary inspection of the data, several enzymes were identified as involved in the biosynthesis of the cell wall, fatty acids, cutin, isoflavonoids and also in C_1 metabolism and the proteolytic pathway. Such pathways will be considered in greater detail in the following sections.

2.4.2 Cell wall related proteins in the seed coat

Most of the carbon fixed by plant photosynthesis is incorporated into cell wall carbohydrates; the remainder forms glycoproteins, glycolipids, storage polysaccharides and small molecules such as glycosides and oligosaccharides. The monosaccharide building blocks of plant carbohydrates are highly diverse, and carbohydrate biosynthesis requires subtle quantitative control throughout growth and development, putting formidable pressure on the evolution of versatile regulatory mechanisms. The glycosyltransferases involved in carbohydrate biosynthesis typically depend on nucleotide sugars as substrates.

The hourglass cells are the most prominent anatomical feature in mature soybean seed coats. The thickened cell walls provide structural support for the seed and allow them to withstand the tensile pressure of the growing embryo (Thorne, 1981). Thickened cell walls are also observed in the vascular parenchyma and aerenchyma, where they may enhance the apoplastic transport of nutrients to the embryo during seed filling (Miller et al., 1999; Yaklich et al., 1995). In *Vicia faba* and *Pisum sativum*, cell wall invertases play a role in creating sink strength for sucrose (Weschke et al., 2003; Weber et al., 1996). A cell wall invertase (116831291) and an acid invertase (47969540) were found at maturity in the seed coat, in agreement with several reports on the role of the seed coat in nutrient uploading (Zhang et al., 2007; Harrington et al., 2005; Van Dongen et al., 2003).

In the analysis of the seed coat proteome we found several key proteins involved in the synthesis of the cell wall (Table 2.1). Detailed analysis of the data allowed the identification of several nucleotide sugar interconverting enzymes previously reported to be involved in cell wall synthesis (Figure 2.8). For example, rhamnose synthase (RHM) was reported as a key enzyme in pectin production in *Arabidospsis* seed coats, affecting directly the amount of mucilage produced (Usadel et al., 2004; Western et al., 2004). Although we were not able to identify peptides from UDP-glucose epimerase (UGE), the presence of up and downstream enzymes could be taken as an indication of their presence. Mutant

analysis of UGE encoding genes demonstrated that its up regulation does not have noticeable effects in the cell wall composition (Doermann et al., 1998) but that its down regulation causes a 25% reduction of cell-wall bound galactose (Seifert et al., 2002) affecting the cell wall integrity.

A soybean cell wall serine protease encoded by *SCS1* was reported to express specifically in the parenchyma (Batchelor et al., 2000). We found two other subtilisin-like serine proteases in the seed coat proteome (#1060, 33621210; #1061, 86439745) that could be involved in the tissue remodeling of this layer.

The epidermis and hourglass cells from the seed coat form a rigid outer shell around the seed coats. This rigidity limits seed size and results in the crushing of some of the inner seed coat layers as the embryo grows (Weber *et al.*, 2005; Murray *et al.*, 1979). The presence of hydroxyproline-rich glycoproteins (extensins) (#1207, 14193753) and proline-rich proteins (#1608, 7269684; #1099, 9454580) confirms the postulate that their accumulation during development and cross-link to the extracellular matrix helps to solidify the cell walls of the epidermis, hourglass and vascular cells.

The current view of the cell wall is of a highly dynamic, responsive structure which not only is associated with a variety of developmental events but is also important in processing information from external stimuli (Humphrey *et al.*, 2007; Pilling *et al.*, 2003). The cell wall continuum is extended to the plasma membrane and underlying cytoskeleton so that the external and the internal environments are linked. Cell expansion is initiated by increase in turgor pressure and followed by a controlled loosening of the cell wall and simultaneous deposition of new wall material (Cosgrove *et al.*, 2005). Expansins are low abundance cell wall proteins that are important agents in the control of cell wall loosening, as they disrupt the non covalent bonds between cellulose and matrix polysaccharides (McQueenmason et al., 1995). It has been reported that xyloglucan endotransglycosylases cleave and re-graft xyloglucans and have a role in cell wall loosening and strengthening. They have also been implicated in

the physiological response to mechanical stimuli and auxin-mediated growth (Antosiewicz et al., 1997; Fry et al., 1992; Talbott et al., 1992). The presence of both expansin and xyloglucan endotransglycosylase in the seed coat proteome is not surprising (#1207, 14193753; #298, 89145876), but it brings some detail to the actual mechanism of cell wall loosening, a process that is necessary for elongation.

Enzymes such as polygalacturonase and pectate lysases cause enzymatic degradation of pectin in the plant cell walls and they are known to be pathogen secreted, releasing oligogalacturonides, which can act as a signal to trigger defense responses (Cote and Hahn, 1994). The presence of these defense-related enzymes in the seed coat proteome (#1210, 37051109; #1211, 110836643; #275, 127464581) is most likely associated with the remodeling of cell layers during development in order to affect the cell wall pectin. Other cell wall proteins, such as proteases, polysaccharide hydrolytic enzymes, and lipases were reported to contribute to the generation of defense signals and response to the environment and many still unknown proteins may fall in this category.

Our results establish a baseline for further scientific investigation and discovery of key players in the mechanism seed coat response to the environment. The spectral data should be further analyzed against a soybean cell wall proteome once the resources become available.

2.4.3 Lipid metabolism in the seed coat

In terms of soybean oil production, the cotyledons contribute 98% of the oil and the seed coats 0.5% (Liu et al., 1995). Although its contribution is relatively small, the seed coat potential in the oil industry could become relevant when considering the production volume. At maturity, soybean seed oil content is approximately 20%, from which 90% is TGA stored in oil bodies in the cotyledon (Weber et al., 2005; Wilson et al., 1986).

Our results demonstrate at the protein level that the seed coats are capable of *de novo* lipid synthesis. Although, ER enzymes from the Kennedy pathway were elusive, we assume that the elongation of fatty acids take place in

the seed coat, as there are studies reporting on the lipid content of the seed coat. Also, transcripts from GPAT were detected in a seed coat microarray study (M. Gijzen, personal communication), which suggests that ER enzymes involved in FA elongation were in fact present, but probably in low abundance at the present developmental stage. Alternatively, the FA production had already ceased at the developmental stage in our study. The lipid composition of seed coats is as follows: 17-20% is phospholipids, 67% TAG, 15% others (steryl esters, 1,3- and 1,2-DAG, free FAs, and glycolipids (Yoshida et al., 2006a).

It is also noteworthy that seed coat seems to have the capability to synthesize tocopherols, from a branching of the phenylpropanoid pathway and FA synthesis. Soybean and their products are relatively good sources of vitamin E (tocopherols). Tocopherols are important biological and nutritive components of foods and are regarded as neutraceuticals for their positive impact on human health. Tocopherols belong to the group of antioxidant vitamins and prevent formation of free radicals (Bramley et al., 2000). They protect the chloroplasts from photo oxidative damage (Havaux *et al.,* 2005). In *Arabidopsis* it was demonstrated that tocopherols in the seed inhibit the oxidation of polyunsaturated fatty acids during dormancy and germination, increasing germination fitness (Sattler et al., 2004a). This report provides additional support to the postulated function of the seed coat in germination enhancement.

In soybean oil extracted from different seed components, it was found that seed coats contain as much as 30 mg of tocopherols per 100 g of oil, compared to 100 mg in cotyledons (Yoshida et al., 2006b). In a metabolic engineering approach taken to increase the vitamin E content in soybean utilizing the over expression of 2-methyl-6-phytylbenzoquinol methyltransferase (VTE3) from *Arabidopsis*, the levels of α-tocopherol, which is the active vitamin E, was increased 7-fold (Van Eenennaam et al., 2003).

Fatty acids are the building blocks for the production of plant cuticles (Pollard et al., 2008). We should not forget that the soybean seed coat synthesizes two distinct cuticles with the outer one having a direct impact on water intake of the seed (Shao et al., 2007; Ma et al., 2004). The functional

importance of the cuticle to the whole plant is evidenced by the significant commitment of epidermal cells to cuticle production. For example, over one-half of the fatty acids made by epidermal cells of the rapidly expanding *Arabidopsis* stem are estimated to be channeled into cuticular lipids, more than intracellular membrane and storage lipids combined (Suh et al., 2005). Although we know that cuticles from stem and seed coats differ in their composition, it is valid to assume that the commitment of the soybean epidermal cells to the cuticle formation is substantial.

Shao et al. (2007) demonstrated that the inner and outer cuticles of soybean seed coats are unique in chemical composition and that they are different from cuticles of other organs even in the same plant. The fine tuning of the FA composition of the cuticle has been shown to have direct impact on the quality of the seed. Some seeds are unable to imbibe water, and are termed "stone seeds". This hardness is due to a continuous impermeable cuticle that prevents the water from entering the seed. Stone seeds are a problem in the seed processing industry. It was demonstrated that the lack of mid-chain hydroxylated FA and the elevated amount of other hydroxylated FA cause the cuticle to be continuous (without cracks) preventing the flow of water into the seed. In seed coat cutin, there is a predominance of 2-hydroxy- and ω-hydroxy-fatty acids and the absence of mid-chain hydroxylated fatty acids, and a high proportion of long chain (non-wax) monomers in the monomer profile of the seed cuticle. It was postulated that these FA are critical in modifying either the degree of cross-linking of the components of the cuticle or its integration with the underlying cell wall components, thereby preventing the cuticle from cracking during development.

Other proteins that are plausibly involved in cutin biosynthesis or related processes that control the strength and integrity of the cuticle and its adherence to the cell wall, such as fatty acid desaturases, lipoxygenases (#327 - #334), thioredoxins (e.g., #964, #1502, #1512), dehydrogenases (#1508, #1509), glycosyl transferase (#257) and expansin (#1207) are also present in the seed coat proteome.

Recently there has been some interest in the mechanism of deposition of plant cuticles (Pollard et al., 2008). Although still in its infancy, the elucidation of this mechanism promises to be crucial for the improvement of crop plants. The role of a plasma membrane ABC transporter as main vehicle has been demonstrated in the accumulation of stem wax in Arabidopsis (CER5) (Pighin et al., 2004). It was suggested however, that other mechanisms would be necessary for the apoplastic transport of larger molecules. Our data could be further analyzed once more understanding is gained on this process, since ABC transporters (#1152-#1162) and oxidoreductases (e.g., #464, #1425, #1426), the two main candidates for cutin and wax deposition, are present in the soybean seed coat proteome.

2.4.4 Isoflavonoids synthesis in the seed coat

In soybean, isoflavonoids accumulate mostly in developing seeds and leaves. In the seed coat, the amount of isoflavonoids present was reported to be in the range of 10-90 nmol g/FW (Dhaubhadel et al., 2003). It is known that soybean embryos have the capability to synthesize isoflavonoids *de novo* from simple precursors and it was proposed that the isoflavonoids from the seed coat are transported to the embryo, helping to increase the total amount of these metabolites in the seed.

It was also noted that the inheritance of isoflavonoids in soybean seeds presents a maternal effect; that is, it is transmitted from plant to progeny in the maternal integuments, from which the seed coat arises. Our results confirm the proposed notion that the seed coat is programmed for *de novo* synthesis of isoflavonoids.

The absence of pigmentation in the seed coat results in a yellow color at maturity. Yellow seed coat color the dominant phenotype brought about by the dominant *I* allele (inhibitor); whereas, the homozygous recessive *i* allele gives rise to a fully pigmented seed coat. The locus *I* locus corresponds to a region of chalcone synthase (CHS) (Todd et al., 1996). It comprises six genes, from which CHS7/CHS8 controls the increased pigmentation of the recessive allele *i* (Tuteja

et al., 2004). We were unable to detect CHS (Table 2.3), which was expected, since the cultivar used in our study, Harosoy 63, carries the *I* gene and has yellow seed coat at maturity. The capability of *de novo* synthesis of isoflavonoids may be restricted by CHS silencing via the *I* locus.

The transport mechanism of isoflavonoid has received some attention in the last few years. An ABC-binding cassette-type has been found to be involved in the exudation of isoflavonoid ginstein from soybean roots to the rizhosphere, during the establishment of symbiosis (Sugiyama et al., 2007). In barley, the uptake of endogenous flavonoid glucosides into the vacuoles was found to be mediated by a proton antiporter (Klein et al., 1996), whereas the vacuolar transport of the same compounds was performed by an ABC-transporter in *Arabidopsis* (Frangne et al., 2002). Also, the role of gluthatione S-transferase (GST) and the gluthatione pump from the ABC family of transporters that have been suggested to be important in the transport of pigments in maize, petunia and soybean (Winkel-Shirley, 1999) and could be common to the transport of isoflavonoids.

Isoflavonoids are small molecules and the glycosylated derivatives are reasonably soluble, so phloem mobility is certainly feasible. The detection of relatively high concentrations of isoflavonoid glucosides in pod exudates provides evidence that transport between different organs may occur within the plant (Dhaubhadel et al., 2008).

Our results confirm that maternal tissues are programmed for *de novo* synthesis of isoflavonoids and support the notion that seed coat isoflavonoid biosynthesis contributes to the overall content of the soybean seed in black-seeded varieties. We also demonstrate the presence of several types of transporters that have been previously reported in the transport of isoflavonoids. The detailed complement of biosynthetic enzymes and potential transporters provide a baseline for the closer inspection of the specific mechanisms of isoflavonoid accumulation and transport in the seed coats and will certainly be helpful in the undertaking of customizing the isoflavonoid contents of soybean.

2.4.5 Proteolysis in the seed coat

In Chapter 1 we presented an overview of the concomitant proteolytic processes taking place both in the seed and the seed coat. The importance of proteolysis in the seed coat and endosperm of *Medicago truncatula* has been highlighted by the presence of several proteases at the protein and RNA level (Gallardo *et al.*, 2007). This seems to be the case also for soybean seed coats, as our data shows a substantial set of proteases present at physiological maturity (Table 2.4).

Gallardo et al. (2007) suggested the potential role of a subtilisin-type seed coat protease in endogenous nitrogen remobilization. In agreement with this report, we found six of these proteases (#1057, 42567017; #1058, 18416719; #1059, 18423316; #1060, 33621210; #1061, 86439745 and #1062, 11611651), the latter one was the most abundant with 25 unique peptides. Clp-proteases were also reported as amino acid recyclers (Cahoon et al., 2003), 4 of such a class were present in the seed coat (#1001, 145323770; # 1004, 2921158; #1005, 18378982 and #1006, 18423503). Other seed coat candidate proteins involved in nitrogen recycling are 20S proteasome (#991, 20260224; #1031, 15219317; #1032, 15220961; #1033, 15225839; #1034, 79325892; #1035, 21553663; #1036, 15228805; #1037, 14594931 and #1038, 15223537). In *M. truncatula* these proteins were expressed at the onset of seed filling and reached a maximum at physiological maturity, which is also the case for the ones expressed in the soybean seed coat. We present here evidence of the involvement of the seed coat in amino acid recycling and protein degradation.

Soybean seed coats undergo "yellowing" at maturity, which is the loss of photosynthetic activity due to the disruption of chloroplasts. Cysteine proteases and carboxypeptidases are present in lytic vacuoles, especially during senescence (Guo et al., 2002). Some cellular evidence points to a possible role of vacuolar proteases in the degradation of plastidial proteins after vacuolar autophagy of chloroplasts, but evidence for such a process is at

present not clear-cut (Hortensteiner et al., 2002). In any case, the seed coat proteome is equipped with at least 54 cysteine proteases (half of the protease component), which gives a direct indication of active chloroplast degradation taking place at physiological maturity.

The ubiquitin/proteasome pathway has proven to be one connected to most biological processes (Schaller et al., 2004). However, there are many other proteases involved with protein turnover and the control of protein half-life by degradation, protein trafficking, processing and limiting the activity by proteolysis of specific active sites.

It should not be forgotten that glycoside hydrolases are involved in the hydrolysis of cell wall polyssacharides and signaling (Minic et al., 2007). We found 3 of these enzymes in the seed coat (#247, 42561840; #248, 15224879; #249, 30689724). Their presence is an indication of the cell wall degradation and remodeling that takes place in the seed coat.

There is a wealth of proteases in the soybean seed coat that could be related to the regulation of several processes such as chloroplast biogenesis and local systemic defense responses (van der Hoorn, 2008). This data provides a detailed prospecting of the proteolytic complement in soybean seed coats and will be helpful in future molecular-based efforts to modify the protein composition of seed coats.

2.4.6. C_1 metabolism-related enzymes in the seed coat

Gallardo et al. (2007) demonstrated that there is a remarkable compartmentalization of enzymes involved in methione synthesis in the different organs of *M. truncatula* seeds. They suggested that this could regulate the availability of sulfur-containing amino acids for embryo protein synthesis during seed filling. The seed coat supports the synthesis of storage compounds during seed development. It transmits organic nutrients from the phloem, mainly sugars, glutamine and asparagines (Weber et al., 2005). Sulfur

containing compounds and molecules, such as sulfate, S- methylmethionine (SMM) are supplied to the embryo via the phloem and can substantially affect seed composition (Tabe et al., 2002; 2001).

Based on relative volume, the most abundant protein in the seed coat at 35-50 DPA is methionine synthase (#77, 33325957 and #78, 8439545) (Table 2.5). Together with *S*-adenosylmethionine synthase (AdoMet) (#107, 37051117), these two enzymes were previously associated with the status of metabolic activity in seeds (Gallardo et al., 2003; 2002, Rajjou et al., 2004). It is noteworthy to find these metabolic enzymes as the most abundant in the seed coat, as an indication of the importance of this process. Their decreased levels could be considered an indication of the switch from active metabolism to a quiescent state. In our study, the high abundance of this protein indicates that quiescence is not yet a reality. It will be interesting to follow up on the expression of this protein in the development of the seed coat.

It is also known that the overall content on sulfur-containing amino acids cysteine and methionine in legumes is low (<1.5%). Although the amino acid composition of seed proteins is genetically programmed, it can be influenced by the rate of accumulation and nutrient availability during seed filling (Weber et al., 2005).

We propose that the high levels of methionine synthase in the seed coat could also be related to its participation in genesis of ethylene. Ethylene is known to promote fruit ripening (Barry and Giovannoni, 2007). *S*-adenosyl-L methionine (SAM)(#107, 37051117), amonocyclopropane 1- carboxylic acid (ACC) oxidase (ACO) (#1445, 25989506) and various forms of ACC synthases (late embryogenesis and maturation proteins) (#584 - #589; 15223413, 4585271, 6648964, 4838149, 9622153, 5802244) were found in the seed coat, supporting the active production of ethylene.

A general pathway for C_1 metabolism is presented in Figure 2.12. The presence of catalase (# 25 in the pathway)(#1465 - #1467; 2661023, 3929924,

40950550, also reported in Table 2.5) becomes important, as its involvement in the glyoxylate cycle will aid in the lipid hydrolysis during the mobilization of storage products necessary in the germination and seedling establishment .

Altogether, the data herein presented constitutes evidence of the multi-functionality of the seed coat at physiological maturity (35-50 DPA). The seed coat proteome's role in the synthesis of cell wall formation and tissue remodeling was analyzed, as well as the *de novo* synthesis of isoflavonoids. The C_1 metabolism and its potential implications in amino acid synthesis and fruit ripening were discussed. The data was presented in the format of structured pathways that will be useful for further applied investigation in soybean seed coats. The regulation of some of these proteins during development will be explored in the next chapter.

2.5. References

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature **422,** 198-207.

Antosiewicz, D.M., Purugganan, M.M., Polisensky, D.H., and Braam, J. (1997). Cellular localization of Arabidopsis xyloglucan endotransglycosylase-related proteins during development and after wind stimulation. Plant Physiol **115,** 1319-1328.

Barry, C.S., and Giovannoni, J.J. (2007). Ethylene and fruit ripening. Journal of Plant Growth Regulation **26,** 143-159.

Batchelor, A.K., Boutilier, K., Miller, S.S., Labbe, H., Bowman, L., Hu, M., Johnson, D.A., Gijzen, M., and Miki, B.L.A. (2000). The seed coat-specific expression of a subtilisin-like gene, SCS1, from soybean. Planta **211,** 484-492.

Bendall, S. (2008). Proteomic characterization of human embryonic stem cell culture. Ph. D. dissertation, The University of western Ontario, 236 pages.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., Van Staveren, M., Stiekema, W., Drost, L., Ridley, P., Hudson, S.A., Patel, K., Murphy, G., Piffanelli, P., Wedler, H., Wedler, E., Wambutt, R., Weitzenegger, T., Pohl, T.M., Terryn, N., Gielen, J., Villarroel, R., De Clerck, R., Van Montagu, M., Lecharny, A., Auborg, S., Gy, I., Kreis, M., Lao, N., Kavanagh, T., Hempel, S., Kotter, P., Entian, K.D., Rieger, M., Schaeffer, M., Funk, B., Mueller-Auer, S., Silvey, M., James, R., Montfort, A., Pons, A., Puigdomenech, P., Douka, A., Voukelatou, E., Milioni, D., Hatzopoulos, P., Piravandi, E., Obermaier, B., Hilbert, H., Dusterhoft, A., Moores, T., Jones, J.D.G., Eneva, T., Palme, K., Benes, V., Rechman, S., Ansorge, W., Cooke, R., Berger, C., Delseny, M., Voet, M., Volckaert, G., Mewes, H.W., Klosterman, S., Schueller, C., Chalwatzis, N., and Project, E.A.G. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. Nature **391,** 485-488.

Blackman, S.A., Obendorf, R.L., and Leopold, A.C. (1992). Maturation Proteins and Sugars in Desiccation Tolerance of Developing Soybean Seeds. Plant Physiol **100,** 225-230.

Bramley, P.M., Elmadfa, I., Kafatos, A., Kelly, F.J., Manios, Y., Roxborough, H.E., Schuch, W., Sheehy, P.J.A., and Wagner, K.H. (2000). Vitamin E. Journal Of The Science Of Food And Agriculture **80,** 913-938.

Cahoon, A.B., Cunningham, K.A., and Stern, D.B. (2003). The plastid clpP gene may not be essential for plant cell viability. Plant Cell Physiol **44,** 93-

95.

Churchward, M., Lang, J.C., Hsu, K., and Coorssen, J.R. (2005). Enhanced detergent extraction for analysis of membrane proteomes by two-dimensional gel electrophoresis. Proteome Science 5.

Cosgrove, D.J. (2005). Growth of the plant cell wall. Nature Reviews Molecular Cell Biology **6,** 850-861.

Cote, F., and Hahn, M.G. (1994). Oligosaccharins - Structures and Signal-Transduction. Plant Mol.Biol **26,** 1379-1411.

Dhaubhadel, S., Farhangkhoee, M., and Chapman, R. (2008). Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. J Exp Bot **59,** 981-994.

Dhaubhadel, S., McGarvey, B.D., Williams, R., and Gijzen, M. (2003). Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol.Biol **53,** 733-743.

Dhaubhadel, S., Kuflu, K., Romero, M.C., and Gijzen, M. (2005). A soybean seed protein with carboxylate-binding activity. J Exp Bot **56,** 2335-2344.

Doermann, P., and Benning, C. (1998). The role of UDP-glucose epimerase in carbohydrate metabolism of Arabidopsis. Plant J **13,** 641-652.

Ferreira, R.M.B., Ramos, P.C.R., Franco, E., Ricardo, C.P.P., and Teixeira, A.R.N. (1995). Changes In Ubiquitin And Ubiquitin-Protein Conjugates During Seed Formation And Germination. J Exp Bot **46,** 211-219.

Frangne N, E.T., Koblischke C, Weissenbock G, Martinoia E, Klein M, Weissenbock G, Dufaud A, Gaillard C, Kreuz K, Martinoia E. (2002). Flavone glucoside uptake into barley mesophyll and Arabidopsis cell culture vacuoles: energization occurs by H(1)-antiport and ATP-binding cassette-type mechanisms. Plant Physiol **128,** 726–733.

Fry, S.C., Smith, R.C., Renwick, K.F., Martin, D.J., Hodge, S.K., and Matthews, K.J. (1992). Xyloglucan Endotransglycosylase, A New Wall-Loosening Enzyme-Activity From Plants. Biochemical Journal **282,** 821-828.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., and Burstin, J. (2003). Proteomics of Medicago truncatula seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol **133,** 664-682.

Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J., and Job, D. (2002). Importance of methionine biosynthesis for

Arabidopsis seed germination and seedling growth. Physiologia Plantarum **116**, 238-247.

Gallardo, K., Firnhaber, C., Zuber, H., Hericher, D., Belghazi, M., Henry, C., Kuster, H., and Thompson, R. (2007). A combined proteome and transcriptome analysis of developing Medicago truncatula seeds. Mol Cell Proteomics **6**, 2165-2179.

Gijzen, M, Miller, S.S., Bowman, L.A., Batchelor, A.K., Boutilier, K., and Miki, B.L.A. (1999). Localization of peroxidase mRNAs in soybean seeds by in situ hybridization. Plant Mol Biol **41**, 57-63.

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

Guo, Y.M., Shen, S.H., Jing, Y.X., and Kuang, T.Y. (2002). Plant proteomics in the post-genomic era. Acta Botanica Sinica **44**, 631-641.

Hajduch, M., Ganapathy, A., Stein, J.W., and Thelen, J.J. (2005). A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol **137**, 1397-1419.

Hajduch, M., Casteel, J.E., Hurrelmeyer, K.E., Song, Z., Agrawal, G.K., and Thelen, J.J. (2006b). Proteomic analysis of seed filling in Brassica napus. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol **141**, 32-46.

Hanson, A.D., and Roje, S. (2001). One-carbon metabolism in higher plants. Annu. Rev. Plant Physiol. Plant Molec Biol **52**, 119-137.

Harrington, G.N., Dibley, K.E., Ritchie, R.J., Offler, C.E., and Patrick, J.W. (2005). Hexose uptake by developing cotyledons of Vicia faba: physiological evidence for transporters of differing affinities and specificities. Funct Plant Biol **32**, 987-995.

Haughn, G., and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. Trends in Plant Science **10**, 472-477.

Havaux, M., Eymery, F., Porfirova, S., Rey, P., and Dormann, P. (2005). Vitamin E protects against photoinhibition and photooxidative stress in Arabidopsis thaliana. Plant Cell **17**, 3451-3469.

Herman, E.M., Helm, R.M., Jung, R., and Kinney, A.J. (2003). Genetic modification removes an immunodominant allergen from soybean. Plant Physiol **132**, 36-43.

Hortensteiner, S., and Feller, U. (2002). Nitrogen metabolism and remobilization during senescence. J Exp Bot **53,** 927-937.

Humphrey, T.V., Bonetta, D.T., and Goring, D.R. (2007). Sentinels at the wall: cell wall receptors and sensors. New Phytol. **176,** 7-21.

Kirmizi, S., and Guleryuz, G. (2006). Protein mobilization and proteolytic enzyme activities during seed germination of broad bean (Vicia faba L.). Zeitschrift Fur Naturforschung C-A Journal of Biosciences **61,** 222-226.

Klein M, W.G., Dufaud A, Gaillard C, Kreuz K, Martinoia E. (1996). Different energization mechanisms drive the vacuolar uptake of a flavonoid glucoside and a herbicide glucoside. J Biol Chem **271,** 29666–29671.

Le, B.H., Wagmaister, J.A., Kawashima, T., Bui, A.Q., Harada, J.J., and Goldberg, R.B. (2007). Using genomics to study legume seed development. Plant Physiol **144,** 562-574.

Liu, K.S., Brown, E.A., and Orthoefer, F. (1995). Fatty-acid composition within each structural part and section of a soybean seed. J Agric Food Chem **43,** 381-383.

Lung, S.C., and Weselake, R.J. (2006). Diacylglycerol acyltransferase: A key mediator of plant triacylglycerol synthesis. Lipids **41,** 1073-1088.

Ma, F.S., Cholewa, E., Mohamed, T., Peterson, C.A., and Gijzen, M. (2004). Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. Ann Bot **94,** 213-228.

McQueenmason, S.J., and Cosgrove, D.J. (1995). Expansin Mode Of Action On Cell-Walls - Analysis Of Wall Hydrolysis, Stress-Relaxation, and Binding. Plant Physiol **107,** 87-100.

Miller, S.S., Bowman, L.A.A., Gijzen, M., and Miki, B.L.A. (1999). Early development of the seed coat of soybean (Glycine max). Ann Bot **84,** 297-304.

Minic, Z., Jamet, E., Negroni, L., der Garabedian, P.A., Zivy, M., and Jouanin, L. (2007). A sub-proteome of Arabidopsis thaliana mature stems trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. J Exp Bot **58,** 2503-2512.

Miranda, M., Borisjuk, L., Tewes, A., Dietrich, D., Rentsch, D., Weber, H., and Wobus, U. (2003). Peptide and amino acid transporters are differentially regulated during seed development and germination in faba bean. Plant Physiol **132,** 1950-1960.

Mooney, B.P., and Thelen, J.J. (2004). High-throughput peptide mass

fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. Phytochemistry **65,** 1733-1744.

Mooney, B.P., Miernyk, J.A., Greenlief, C.M., and Thelen, J.J. (2006). Using quantitative proteomics of Arabidopsis roots and leaves to predict metabolic activity. Physiologia Plantarum **128,** 237-250.

Murray, D.R. (1979). Nutritive Role of the Seed Coats During Embryo Development in Pisum-Sativum. Plant Physiol **64,** 763-769.

Natarajan, S., Xu, C.P., Caperna, T.J., and Garrett, W.A. (2005). Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. Anal Biochem **342,** 214-220.

Pighin, J.A., Zheng, H.Q., Balakshin, L.J., Goodman, I.P., Western, T.L., Jetter, R., Kunst, L., and Samuels, A.L. (2004). Plant cuticular lipid export requires an ABC transporter. Science **306,** 702-704.

Pilling, E., and Hofte, H. (2003). Feedback from the wall. Curr Opin Plant Biol **6,** 611-616.

Pollard, M., Beisson F., Li, Y., Ohlrogge, J. (2008). Building lipid barriers: biosynthesis of cutin and suberin. Trends in Plant Science **available online April 24, 2008**.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of alpha-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiol **134,** 1598-1613.

Rawlings, N.D., Morton, F.R., and Barrett, A.J. (2006). MEROPS: the peptidase database. Nucleic Acids Research **34,** D270-D272.

Ruuska, S.A., Girke, T., Benning, C., and Ohlrogge, J.B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell **14,** 1191-1206.

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M., and DellaPenna, D. (2004b). Vitamin E is essential for seed longevity, and for preventing lipid peroxidation during germination. Plant Cell **16,** 1419-1432.

Schaller, A. (2004). A cut above the rest: the regulatory function of plant proteases. Planta **220,** 183-197.

Seifert, G.J. (2004). Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. Curr Opin Plant Biol **7,** 277-284.

Seifert, G.J., Barber, C., Wells, B., Dolan, L., and Roberts, K. (2002). Galactose biosynthesis in Arabidopsis: Genetic evidence for substrate channeling from UDP-D-galactose into cell wall polymers. Curr Biol **12,** 1840-1845.

Shao, S.Q., Meyer, C.J., Ma, F.S., Peterson, C.A., and Bernards, M.A. (2007b). The outermost cuticle of soybean seeds: chemical composition and function during imbibition. J Exp Bot **58,** 1071-1082.

Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc **1,** 2856-2860.

Statistics Canada, C.a.O.R.S. (2007). Cat. No. 22-077.

Sugiyama, A., Shitan, N., and Yazaki, K. (2007). Involvement of a soybean ATP-binding cassette - Type transporter in the secretion of genistein, a signal flavonoid in legume-Rhizobium Symbiosis(1). Plant Physiol **144,** 2000-2008.

Suh, M.C., Samuels, A.L., Jetter, R., Kunst, L., Pollard, M., Ohlrogge, J., and Beisson, F. (2005). Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. Plant Physiol **139,** 1649-1665.

Syrovy, I., and Hodny, Z. (1991). Staining and Quantification of Proteins Separated by Polyacrylamide-Gel Electrophoresis. J Chromatogr-Biomed **569,** 175-196.

Tabe, L., Hagan, N., and Higgins, T.J.V. (2002). Plasticity of seed protein composition in response to nitrogen and sulfur availability. Curr. Opin. Plant Biol **5,** 212-217.

Tabe, L.M., and Droux, M. (2001). Sulfur assimilation in developing lupin cotyledons could contribute significantly to the accumulation of organic sulfur reserves in the seed. Plant Physiol **126,** 176–187.

Talbott, L.D., and Ray, P.M. (1992). Molecular-Size And Separability Features Of Pea Cell-Wall Polysaccharides - Implications For Models Of Primary Wall Structure. Plant Physiol **98,** 357-368.

Thelen, J.J., and Ohlrogge, J.B. (2002). Metabolic engineering of fatty acid biosynthesis in plants. Metab Eng **4,** 12-21.

Thorne, J.H. (1981). Morphology and ultrastructure of maternal seed tissues of soybean in relation to the import of photosynthate. Plant Physiol **67,** 1016-1025.

Todd, J.J., and Vodkin, L.O. (1996). Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell **8,** 687-699.

Tuteja, J.H., Clough, S.J., Chan, W.C., and Vodkin, L.O. (2004). Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in Glycine max. Plant Cell **16,** 819-835.

Usadel, B., Kuschinsky, A.M., Rosso, M.G., Eckermann, N., and Pauly, M. (2004). RHM2 is involved in mucilage pectin synthesis and is required for the development of the seed coat in Arabidopsis. Plant Physiol **134,** 286-295.

van der Hoorn, R. (2008). Plant Proteases: from phenotypes to molecular mechanisms. Annu. Rev Plant Biol **59,** 191-223.

Van Dongen, J.T., Ammerlaan, A.M.H., Wouterlood, M., Van Aelst, A.C., and Borstlap, A.C. (2003). Structure of the developing pea seed coat and the post-phloem transport pathway of nutrients. Ann Bot **91,** 729-737.

Van Eenennaam, A.L., Lincoln, K., Durrett, T.P., Valentin, H.E., Shewmaker, C.K., Thorne, G.M., Jiang, J., Baszis, S.R., Levering, C.K., Aasen, E.D., Hao, M., Stein, J.C., Norris, S.R., and Last, R.L. (2003). Engineering vitamin E content: From Arabidopsis mutant to soy oil. Plant Cell **15,** 3007-3019.

Weber, H., Borisjuk, L., and Wobus, U. (1996). Controlling seed development and seed size in Vicia faba: a role for seed coat-associated invertases and carbohydrate state. Plant J **10,** 823-834.

Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. Annu. Rev. Plant Biol. **56,** 253-279.

Weschke, W., Panitz, R., Gubatz, S., Wang, Q., Radchuk, R., Weber, H., and Wobus, U. (2003). The role of invertases and hexose transporters in controlling sugar ratios in maternal and filial tissues of barley caryopses during early development. Plant J **33,** 395-411.

Western, T.L., Young, D.S., Dean, G.H., Tan, W.L., Samuels, A.L., and Haughn, G.W. (2004). MUCILAGE-MODIFIED4 encodes a putative pectin biosynthetic enzyme developmentally regulated by APETALA2, TRANSPARENT TESTA GLABRA1, and GLABRA2 in the Arabidopsis seed coat. Plant Physiol **134,** 296-306.

Wilson, R.F., and Kwanyuen, P. (1986). Triacylglycerol Synthesis And Metabolism In Germinating Soybean Cotyledons. Biochimica Et Biophysica Acta **877,** 231-237.

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol **126,** 485-493.

Xu, C.P., Garrett, W.M., Sullivan, J., Caperna, T.J., and Natarajan, S. (2006a). Separation and identification of soybean leaf proteins by two-dimensional gel electrophoresis and mass spectrometry. Phytochemistry **67,** 2431-2440.

Xu, C.P., Caperna, T.J., Garrett, W.M., Cregan, P., Bae, H.H., Luthria, D.L., and Natarajan, S. (2007). Proteomic analysis of the distribution of the major seed allergens in wild, landrace, ancestral, and modern soybean genotypes. Journal of The Science of Food and Agriculture **87,** 2511-2518.

Xu, X.Y., Zheng, R., Li, C.M., Gai, J.Y., and Yu, D.Y. (2006b). Differential proteomic analysis of seed germination in soybean. Progress In Biochemistry and Biophysics **33,** 1106-1112.

Yaklich, R.W. (1990). Freeze fracture study of soybean (Glycine max L. Merr.) seed coat cells. Physiologia Plantarum **79,** A17.

Yaklich, R.W., Vigil, E.L., and Wergin, W.P. (1995). Morphological and fine-structural characteristics of aerenchyma cells in the soybean seed coat. Seed Science and Technology **23,** 321-330.

Yoshida, H., Takagi, S., and Hirakawa, Y. (2000). Molecular species of triacylglycerols in the seed coats of soybeans (glycine max L.) following microwave treatment. Food Chemistry **70,** 63-69.

Yoshida, H., Kanei, S., Tomiyama, Y., and Mizushina, Y. (2006a). Regional distribution in the fatty acids of triacylglycerols and phospholipids within soybean seeds (Glycine max L.). European Journal of Lipid Science and Technology **108,** 149-158.

Yoshida, H., Kanrei, S., Tomiyama, Y., and Mizushina, Y. (2006b). Regional characterization of tocopherols and distribution of fatty acids within soybean seeds (Glycine max L.). Journal of Food Lipids **13,** 12-26.

Yu, O., and McGonigle, B. (2005). Metabolic engineering of isoflavone biosynthesis. Advances In Agronomy, Volume 86 **86,** 147-190.

Zhang, W.H., Zhou, Y.C., Dibley, K.E., Tyerman, S.D., Furbank, R.T., and Patrick, J.W. (2007b). Nutrient loading of developing seeds. Funct Plant Biol **34,** 314-331.

Chapter 3

DEVELOPMENTAL ANALYSIS OF SOYBEAN SEED COAT PROTEOME

3.1 Introduction

In Chapter 2 we reported on the soybean seed coat proteome at physiological maturity (35-50 DPA). The protein complement shows a diversity of processes taking place in the seed coat, supporting its role as a protective, defensive and metabolically engaged organ. To gain insight on the regulation of protein expression in this organ, a developmental study is necessary to address the major changes during the different seed developmental stages.

The complex process of seed development can be divided into three sequential phases: embryogenesis, seed filling and seed maturation. The longest phase is the seed filling stage, characterized by cell division, cell expansion and accumulation of storage product synthesis (Mienke, 1981). The metabolic events that occur during seed filling will eventually determine the overall composition of seed, and targeted transgenic alteration of these pathways have the potential to greatly impact seed quality traits (Thelen and Ohlrogge, 2002). Given this biotechnological importance, the soybean seed filling has been studied in depth by Hajduch and coworkers (2005), and at least 216 non-redundant proteins from soybean embryos have been reported.

Relative to seed embryos, the molecular and cellular events underlying seed coat differentiation have received less attention. A lot of the research in seed coats has been done in *Arabidopsis* (Haughn and Chaudhury, 2005; McFarlane et al., 2008; Rautengarten et al., 2008; Truernit and Haseloff, 2008), and it as greatly contributed to the understanding of many aspects of seed coat biology. There is sufficient evidence to assume that the fundamental regulatory mechanisms are similar in legume seeds. However, too much generalization could be misleading because, unlike *Arabidopsis*, grain legumes are crop plants selected for high yield and characterized by high metabolic activity and fluxes in seeds.

In legume seeds, seed coat studies have focused on understanding transport events (Murray, 1979b; Thorne, 1981; Grusak and Minchin, 1988; Offler and Patrick, 1993; Walker et al., 1995; Wang et al., 1995; Rolletschek et al., 2005), seed filling and morphology (Murray, 1979a; Miller et al., 1999). Other aspects of legumes seed coat were also studied, such as seed hardness and water uptake (Shao et al., 2007; Shackel and Turner, 2000; Mullin and Xu, 2001; Ma et al., 2004; Qutob et al., 2008), morphology and structure (Yaklich et al., 1992; Van Dongen et al., 2003), and individual proteins (Weber et al., 1995; Gijzen, 1997; Schuurmans et al., 2003; Zhou et al., 2007).

More comprehensive approaches were used to study the proteomics of seed filling of *M. truncatula* (Gallardo et al., 2003) and the correlation between proteome and transcriptome was established at the seed coat endosperm and embryo tissues (Gallardo et al., 2007). Surprisingly, only a few reported proteins in all these studies were common among species, pointing to marked proteome differences between different legume seeds. The difference was also significant when comparing different seed compartments, the seed coat, endosperm and embryo. The authors concluded in fact that there is a clear specialization in the different compartments, at least in amino acid metabolism.

Our study aims at elucidating the proteomic changes that occur during soybean seed coat development, a matter that has been overlooked. The expectation is that the information generated from this study will help in the understanding of general mechanisms taking place in soybean seed development, an important topic in biology and biotechnology.

3.2. Methods

3.2.1 Sample collection and preparation

a) Plant materials and growth conditions

Soybean seeds (Glycine max) L. Merr. cv Harosoy 63 were planted at the Agriculture and Agri-Food Canada Research Centre in London, Ontario, in 2006 and 2007. Regular agronomic practices and planting dates were followed.

Flowers at anthesis from nodes 3 and 4 were tagged and harvested at 20, 35, 50 and 80 days post anthesis (DPA). Pods were collected randomly from 20-30 plants, and seed coats were excised from seeds, frozen in liquid nitrogen, and stored at -80 °C.

b) Protein extraction

Total protein was isolated from soybean whole seeds (20 DPA) and seed coats only (35, 50, 80 DPA) and subjected to trichloroacetic acid (TCA) precipitation according to Gorg et al. (1997) with modifications from Natarajan et al. (2005). The procedure was described in section 2.2 in the previous Chapter.

3.2.2 Two-dimensional gel electrophoresis

The first and second dimensions of the 2D-SDS-PAGE along with the gel staining procedures were performed as previously described in section 2.2. For the developmental analysis of seed coat proteome, 4 technical replicates were analyzed at each stage (E, M, L, Mat). Image acquisition was performed using a PowerLook 1120 scanner (UMAX Technologies Inc., Taiwan) with a resolution of 300 dpi and 16-bit grayscale pixel depth.

3.2.3 Mass Spectrometry and protein identification

After gel image analysis, 342 selected spots that met specific criteria, i.e.: a) were present in all 4 replicates in at least 3 developmental stages and b) their expression was above a normalized volume of 10.7 (provided that they were big and resolved enough to be excised), were chosen. Spots were manually excised from the reference gel of Late stage (35-50 DPA) using a OneTouch manual spot picker (The Gel Company) (3.0 mm). The selected spots were excised and subjected to automated in-gel trypsin digestion using a MassPREP Automated Digestor (Waters) followed by ESI-LC MS/MS as described in the previous Chapter.

For electrospray ionization tandem mass spectrometry analysis (ESI LC-MS/MS), all 2D spot-dried fractions were reconstituted in 10% FA prior to

injection. For analysis, spot samples were kept separated. Samples were analyzed using a 60 min LC method. Liquid chromatography (5-40% ACN, 0.1% FA gradient) was performed on a NanoAcquity UPLC (Waters, Milford, MA) with a 25 cm x 75 µm C18 reverse phase column. Peptide ions were detected in data-dependent acquisition (DDA) mode by tandem MS (Q-ToF Ultima - Waters) using the following parameters: survey scan (MS only) range *m/z* 400-1800, 1 s scan time, 1-4 precursor ions selected based on charge state (+2, +3, and +4). For each MS/MS scan, the *m/z* range was extended to *m/z* 50 – 2000, scan times used ranged from 1.5 - 6 s (signal dependent), and a charge state-dependent collision energy profile was used.

The acquired MS/MS spectra were processed by using ProteinLynx Global SERVER 2.2.5 (Waters) and searched against extracted subsets for "Plants" or "*Glycine max*" (forward and reverse) of NCBI nr protein databases (www.ncbi.nlm.nih.gov) using Spectrum Mill (Agilent Technologies, Santa Clara, CA). The following settings were employed: a mass tolerance of 100 ppm for MS spectra and 100 ppm for MS/MS spectra, a spectral peak intensity (SPI) limit of 60 %, minimum peptide score of 6, and minimum protein score of 13. To minimize false positives to a rate of 0.0001%, peptides with reverse database scores higher than forward scores were removed from the summaries.

Gene ontology was assigned to all identified proteins in 2D samples according to a classification for yeast adapted for the Arabidopsis genome (Bevan *et al.*, 1998) with modifications that make it more suitable for a seed study (Hajduch *et al.*, 2006).

3.2.4 Relative protein quantification and expression patterns

a) Quantification of protein abundance

Gel digital images were analyzed with Progenesis PG220 v2006 and Progenesis SameSpots TT900 SDSTM software (Nonlinear Dynamics, Newcastle upon Tyne, UK). Image analysis was carried out using the default analysis wizard, which combines spot detection, warping, and matching on each

set of gels (E, M, L, Mat). Background subtraction was carried out by "Mode of non-Spot" with a margin of 45. Normalization was carried out by "Total Spot Volume" in which total intensity of pixels of each of the software-delineated polypeptide spots was expressed as a percentage of the total intensity of pixels of all software-delineated polypeptide spots. This normalized the amount of any given polypeptide spot to the total polypeptides on each gel. At each stage, the best gel replicate was chosen as a reference (see Appendix III).

In this manner, 565 spots were matched between different stages allowing their comparison. Spots of interest were chosen due to sheer abundance or changes along development by examining the comparison window in the Progenesis220 software package and then highlighted for further study. Only consistent spots present in all 4 stages and 4 replicates were finally selected for mass spectrometry identification. These spots were then checked for accurate detection, and where required manual corrections to spot detection were carried out, and spot volumes were re-determined. The total intensity of pixels within each spot (the integrated intensity) was determined by the software. The relative abundance of each spot per reference gel was then compared with the "mid" reference gel and these values were used for statistical analysis.

b) Protein expression profiles

The integrated intensity of each spot or normalized volume was determined by Progenesis220 based on the area (number of pixels) and intensity of staining (height) and was expressed as percentual fractions of the total integrated intensity of all spots within the region of analysis of the gel. This normalizes the amount of any given spot and gives relative protein abundance values for each sample. The reference gel in each stage allowed for detection of qualitative and/or quantitative differences between replicates. SAS statistical package (SAS Institute) was used to perform a one way ANOVA analysis followed by Dunett's Multiple Comparison test to determine the overall change of protein expression along development. Differences were considered significant when $P < \alpha$.

c) Cluster analysis of 565 differentially expressed protein spots

The normalized relative volumes of each spot were imported into GeneSpring v7.3 (Agilent Technologies). The relative protein expression for seed coat development as a function of time was calculated using a single channel input for each developmental stage imported into GeneSpringGX. To validate this approach, *k*-clustering was performed with the choice of 5 distinct expression groups. Within each cluster, proteins were cross referred to the functional classification presented in Chapter 2 (Appendix II) to determine the correlation between expression pattern and functional class of proteins.

d) Protein and transcript levels of proteins of interest

The expression patterns of individual proteins of interest were examined in-depth. For selected proteins, a comparison was established between the levels of protein and transcript levels. The transcript hybridization intensities were those of a soybean seed coat microarray analysis using a platform of 18,462 low redundancy cDNAs spotted to glass slides (Vodkin et al., 2004). The values were obtained from seed coat cDNA from developing seed coats corresponding to the exact developmental stages in our study. This set of data was provided from an independent study (M. Gijzen, unpublished).

3.3. Results

3.3.1. Staging and characterization of developing soybean seed coats

The objective of this study was to characterize the seed coat global protein expression during soybean seed development. For the best coverage of this period, we analyzed whole seeds (20 DPA), corresponding to the late morphogenetic phase. The seeds were too small to be dissected at this point; hence, the use of whole seeds. Next, the seed coats of 35 DPA–seeds were analyzed, which were undergoing cell division; followed by 50 DPA, with seeds at the end of cell enlargement period. The last stage was at 80 DPA, corresponding to maturity and desiccation, which is the stage at which soybean seeds are harvested.

Figure 3.1 shows the characteristics of the developing seeds used in this study. The whole seed data trend is in agreement with measurements published previously (Hajduch et al., 2005); (Hill, 1974). However, fresh weight and protein values are lower, which can be attributed to differences between varieties and growth conditions.

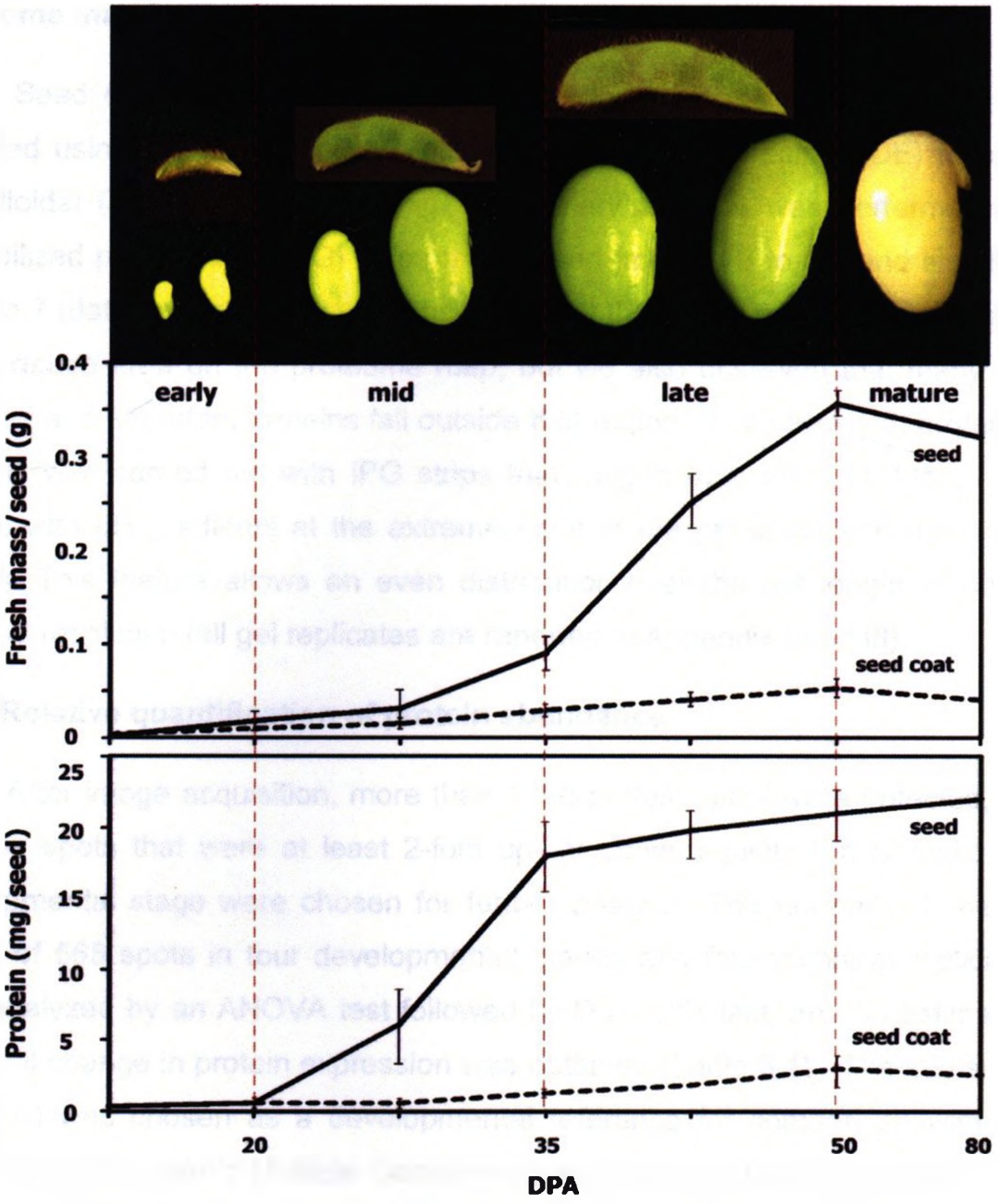


Figure 3.1 Development of soybean seeds during the experimental period. A, Whole seeds and pods at four stages of seed development. Experimental sampling began at 10 DPA and continued at precisely 7-d intervals until 50 DPA; last sampling was performed at 80 DPA. B, Individual seed and seed coat fresh mass during the experimental period expressed as mass per seed. Values are the average of 25 determinations; SD is shown. C, Total protein content per seed and seed coat during the investigated period of seed development. Values are the average of 20 determinations for mass and five for protein; SD is shown.

3.3.2. Broad-range isoelectric focusing is appropriate for high-resolution proteome maps

Seed coat proteins from developing soybean seeds were resolved and detected using high-resolution two-dimensional electrophoresis (2-DE) followed by colloidal Coomasie Blue staining. Preliminary analysis was performed with immobilized pH gradient (IPG) strips that ranged from pH 3 to 10, and also from pH 4 to 7 (data not shown). It was observed that the region from pH 4 to 7 was a highly dense area on the proteome map; but we also observed that many well resolved and abundant proteins fall outside that region. The seed coat proteome analysis was carried out with IPG strips that ranged from pH 3 to 11NL (non-linear), with pH gradients at the extreme ends of the pH scale with non linear scaling. This feature allows an even distribution over the gel length to obtain maximal resolution (all gel replicates are reported in Appendix I and III).

3.3.3. Relative quantification of protein abundance

After image acquisition, more than 1700 protein spots were detected, but only the spots that were at least 2-fold up- or down-regulated in at least one developmental stage were chosen for further analysis. The normalized volume values of 565 spots in four developmental stages and four technical replicates were analyzed by an ANOVA test followed by Dunnett's test, and an estimation of overall change in protein expression was obtained (Table 3.1). The mid-stage (35 DPA) was chosen as a developmental reference for comparison with the other stages (Dunnett's Multiple Comparison test) (SAS statistical package). No differences were found between the replicates and very significant differences were found between the developmental stages at both α levels. The most noteworthy changes in protein expression took place between late and mature stages.

Table 3.1 Analysis of normalized spot volume. Normalized spot volume was measured from 2D gel images using Progenesis220. The software calculates the volume based on the intensity of the stain and the area assigned to each spot and then normalizes it to the total relative volume of all the spots in the gel. The table shows results from 565 spots at each developmental stage and 4 technical replicates. An ANOVA test was performed with α value of 5% and 1%, followed by Dunnett's Multiple Comparison test for the developmental stage parameter.

Parameter	α	ANOVA p - value	Difference between means	Confidence intervals limits	
				lower	upper
Replication	0.05	0.626			
R2-R1			-42.92	-221.75	135.91
R3-R1			-37.75	-216.58	141.08
R4-R1			-99.65	-278.48	79.18
Developmental stage	0.05	< 0.0001			
Early - Mid			155.53	-23.3	334.36
Mid - Late			29.74	-149	208.57
Late - Mature			437.86	259.03	616.69 ***
	0.01	< 0.0001			
Early - Mid			155.53	-66.42	377.47
Mid - Late			-29.74	-251.68	192.21
Late - Mature			437.86	215.91	659.8 ***

*** Based on Dunnett's test the comparison presents significant differences at the given α level.

3.3.4. LC-MS/MS using the NCBI nr database yielded 304 protein assignments

Each of the 342 spots with confirmed expression profiles were excised from reference gels for identification by ESI LC MS/MS as described previously (refer to section 3.2.3 of this Chapter). After mass spectrometry analysis of tryptic peptides, MS/MS spectral data was used determine protein identities. A total of 304 protein assignments were obtained using Spectrum Mill search engine (Agilent technologies, Santa Clara, CA) against the NCBI nr database. The cut off thresholds were established at score 6 for peptides and 13 for proteins. Using this approach 304 proteins out of 342 were identified (89%) (Table 3.2). One unique protein was often represented by more than one spot on the 2D gel, most likely due to post translational modifications, genetic isoforms and proteolysis. Taking into account this redundancy, 185 unique proteins were identified (39% redundancy). Previously, higher levels of redundancy (~49%) were reported in soybean (Hajduch et al., 2005) and 55% in Brassica napus filling studies (Hajduch et al., 2006).

Table 3.2 Proteins identified by LC-MS/MS from 2D SDS-PAGE gels of 35-50 DPA soybean seed coats. Proteins were classified according to protein functional categories described by Bevan *et al.* (1998). Proteins were identified by ESI-LC-MS/MS analysis of tryptic peptides following searching against NCBI nr database. The putative protein identifications with score ≥ 13 were considered as positive. The table includes spot number, *k*-cluster, NCBI nr accession number, MS/MS score, unique peptides, percentage of protein coverage, theoretical MW /pI and the species in which the protein was found with closest similarity in the database.

Expression cluster	Protein	Spot No. PG220	NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
	01 Metabolism										
4	acetylornithine transaminase, putative	223	21554043	29	3	8	40.0	5.3	48.8	6.3	*At*
4	aspartate aminotransferase glyoxysomal isozyme	2506	2654094	58	6	14	42.0	7.8	49.7	8.7	*Gm*
3	chorismate synthase	513	77547031	38	3	5	44.0	7.0	47.3	6.3	*Gm*
2	cysteine synthase	195	126508784	32	2	11	32.0	5.1	34.7	5.3	*Gm*
4	cytosolic glutamine synthetase GSbeta1	2632	10946357	77	5	22	35.5	5.3	39.0	5.5	*Gm*
2	glutamate--ammonia ligase	2534	547508	55	4	12	37.0	5.8	39.2	5.9	*Gm*
2	isovaleryl-CoA Dehydrogenase	382	5869967	23	2	5	41.0	5.6	44.7	6.3	*Ps*
1	methionine synthase	1896	33325957	140	12	18	0.0	5.4	84.3	5.9	*Gm*
2	methionine synthase	42	33325957	40	4	6	39.0	3.9	84.3	5.9	*Gm*
2	methionine synthase	67	33325957	40	4	6	32.0	3.9	84.3	5.9	*Gm*
4	methionine synthase	557	33325957	270	21	39	85.0	6.2	84.3	5.9	*Gm*
4	methionine synthase	2140	33325957	94	10	15	80.0	5.1	84.3	5.9	*Gm*
4	pyridoxine biosynthesis protein	2293	72256519	82	7	24	31.0	5.6	33.2	5.6	*Gm*
3	serine hydroxymethyltransferase 4 (SHM4)	466	11762130	25	2	5	50.0	6.8	51.8	7.1	*At*
4	serine hydroxymethyltransferase 4 (SHM4)	2616	11762130	52	4	10	52.0	7.1	51.8	7.1	*At*
3	serine hydroxymethyltransferase 2 (SHM2)	441	30690400	56	5	13	90.0	6.3	59.1	8.8	*At*
	01 Metabolism										
	01.02 Nitrogen and sulphur										
3	auxin amidohydrolase	436	51538213	16	1	3	44.0	5.5	47.3	5.5	*Ta*
	01 Metabolism										
	01.03 Nucleotides										
3	ferric leghemoglobin reductase	2715	546360	32	3	6	53.0	6.3	55.8	6.9	*Gm*
1	nucleoside diphosphate kinase	2494	6435320	41	3	16	16.0	7.8	25.3	9.4	*Ps*
4	nucleoside diphosphate kinase	319	26245395	44	4	32	15.8	5.6	16.4	6.9	*Gm*
4	nucleoside diphosphate kinase	489	26245395	66	5	39	50.0	10.5	16.4	6.9	*Gm*
5	nucleoside diphosphate kinase	1651	26245395	38	3	24	15.0	6.7	16.4	6.9	*Gm*
	01 Metabolism										
4	AXS2 (UDP-D-APIOSE/UDP-D-XYLOSE SYNTHASE 2)	531	18390863	13	2	5	43.0	6.1	43.8	5.6	*At*
4	catalytic/ coenzyme binding	1991	18404496	17	2	7	28.5	5.6	34.9	8.4	*At*
4	chloroplast NAD-MDH	2689	3256066	75	5	16	33.0	5.5	42.4	8.5	*At*
1	fruit ripening protein	3008	7580480	13	2	5	24.0	5.0	24.0	6.1	*Ls*
4	gamma-aminobutyrate transaminase subunit isozyme 1	3221	29837282	16	2	5	49.0	6.9	56.7	7.7	*Ls*
4	gamma-aminobutyrate transaminase subunit isozyme 2	3221	29837284	18	1	2	49.0	6.9	50.5	6.6	*Ls*
3	gamma-aminobutyrate transaminase subunit isozyme 3	211	29837286	17	1	2	49.0	6.7	57.2	6.7	*Ls*
4	glutamate dehydrogenase 1	2993	59668638	106	8	24	38.0	6.2	44.5	6.0	*Gm*
1	lactoylglutathione lyase, putative / glyoxalase I, putative	1647	15220397	14	3	11	32.0	5.0	29.4	5.0	*At*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
5	succinate dehydrogenase iron-sulphur subunit	603	21555840	28	2	7	30.0	8.3	31.3	8.8	*At*
4	phosphomannomutase	3010	90762150	64	6	27	28.0	5.8	28.0	5.8	*Gm*
1	RHM1/ROL1 (RHAMNOSE BIOSYNTHESIS1)	5555	15218420	46	5	7	72.0	6.5	75.4	6.8	*At*
3	RHM1/ROL1 (RHAMNOSE BIOSYNTHESIS1)	535	15218420	63	6	8	75.0	7.0	75.4	6.8	*At*
5	succinyl-CoA ligase alpha 2 subunit	672	49617539	33	3	13	32.0	8.5	35.4	9.0	*Ls*
2	THFS (10-FORMYLTETRAHYDROFOLATE SYNTHETASE)	6666	18403095	17	2	3	62.0	5.9	67.8	6.3	*At*
3	UDP-glucose 6-dehydrogenase	716	48093457	45	3	6	53.0	5.6	61.0	6.5	*Nt*
3	UDP-glucose:protein transglucosylase-like	675	77416931	84	6	17	39.0	5.6	41.2	5.6	*St*
	01 Metabolism										
4	allene oxide cyclase	574	40644130	25	2	8	25.0	9.3	26.5	9.1	*Nt*
2	epoxide hydrolase	2152	2764806	15	2	7	33.0	5.5	39.2	5.4	*Gm*
2	epoxide hydrolase	2603	2764806	43	4	12	33.0	5.4	39.2	5.4	*Gm*
4	inorganic pyrophosphatase-like protein	664	21593570	35	3	14	30.0	5.5	24.6	5.3	*At*
	02 Energy										
	02.01 Glycolysis										
1	cytosolic phosphoglycerate kinase	1410	9230771	68	5	18	35.0	6.2	42.3	5.7	*Ps*
3	cytosolic phosphoglycerate kinase	2652	9230771	144	11	37	24.0	5.1	42.3	5.7	*Ps*
4	cytosolic phosphoglycerate kinase	734	9230771	95	7	19	37.0	6.2	42.3	5.7	*Ps*
3	enolase	1849	42521309	136	9	27	50.0	5.2	47.7	5.3	*Gm*
4	enolase	445	42521309	180	12	45	47.0	5.3	47.7	5.3	*Gm*
4	enolase	567	42521309	144	10	31	47.0	5.4	47.7	5.3	*Gm*
4	enolase	2492	42521309	155	11	36	45.0	4.2	47.7	5.3	*Gm*
4	fructose-bisphosphate aldolase-like protein	443	15231715	39	3	8	40.0	6.7	38.5	6.1	*At*
3	glyceraldehyde-3-phosphate dehydrogenase C subunit	3096	15229231	59	5	17	37.0	6.3	36.9	6.6	*At*
4	glyceraldehyde-3-phosphate dehydrogenase C subunit	2962	15229231	101	8	26	37.0	7.0	36.9	6.6	*At*
4	glyceraldehyde-3-phosphate dehydrogenase	455	85720768	86	7	25	37.0	7.5	36.8	6.7	*Gm*
4	glyceraldehyde-3-phosphate dehydrogenase	459	85720768	25	3	7	38.0	7.0	36.8	6.7	*Gm*
4	glyceraldehyde-3-phosphate dehydrogenase	1829	85720768	54	5	14	38.0	7.0	36.8	6.7	*Gm*
4	glyceraldehyde-3-phosphate dehydrogenase A subunit	2962	77540210	94	6	19	39.0	6.7	43.2	8.4	*Gm*
4	glycine hydroxymethyltransferase	275	7433553	16	2	3	50.0	7.0	59.3	9.0	*At*
5	pfkB-type carbohydrate kinase family protein	1712	15221364	24	3	10	17.5	5.1	37.6	5.5	*At*
2	PGK1 (PHOSPHOGLYCERATE KINASE 1)	376	15230595	36	3	9	17.0	5.6	50.1	5.9	*At*
4	phosphoglycerate kinase-like	418	15223484	14	2	5	40.0	6.3	49.9	8.3	*At*
1	phosphoglycerate mutase	3185	551288	47	4	6	70.0	5.6	60.6	5.3	*Zm*
3	phosphoglycerate mutase	2015	551288	33	3	4	70.0	5.6	60.6	5.3	*Zm*
4	T-protein of the glycine decarboxylase complex	447	407475	38	3	9	39.0	8.3	44.3	8.8	*Ps*
4	T-protein of the glycine decarboxylase complex	2962	407475	39	4	12	41.0	7.3	44.3	8.8	*Ps*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accesion Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
1	triosephosphate isomerase	2371	77540216	33	4	18	17.5	5.1	27.2	5.9	*Gm*
2	triosephosphate isomerase	2276	77540216	79	7	38	31.0	6.1	27.2	5.9	*Gm*
4	triosephosphate isomerase	646	48773765	49	4	21	29.0	6.1	27.2	5.9	*Gm*
	02 Energy										
	02.02 Gluconeogenesis										
3	cytosolic malate dehydrogenase	2886	15241923	16	1	4	33.0	5.9	36.9	5.8	*At*
4	cytosolic malate dehydrogenase	3040	42521311	159	9	46	34.0	5.8	35.5	6.3	*Gm*
4	cytosolic malate dehydrogenase	2819	42521311	74	6	26	35.0	5.8	35.5	6.3	*Gm*
	02 Energy										
	02.07 Pentose phosphate										
3	6-phosphogluconate dehydrogenase	485	2529229	82	7	15	46.0	7.0	56.4	5.6	*Gm*
	02 Energy										
	02.10 TCA pathway										
4	cytosolic aconitase	683	11066033	58	6	6	100.0	6.0	98.1	5.9	*Nt*
4	isocitrate dehydrogenase (NADP+)	660	3021512	13	1	2	44.0	5.8	53.9	8.3	*Nt*
1	pyruvate dehydrogenase E1 beta subunit isoform 2	3140	162458637	13	2	8	36.0	5.1	35.8	4.8	*Zm*
	02 Energy										
	02.20 Electron-transport										
4	cinnamyl alcohol dehydrogenase 1	385	60265616	18	1	3	35.0	7.3	35.5	6.5	*Nt*
1	F22C12.4 (similar to vacuolar ATPase)	3161	6692094	34	3	8	31.0	5.9	35.8	6.1	*At*
4	FQR1 (FLAVODOXIN-LIKE QUINONE REDUCTASE 1)	2862	3269288	15	2	9	25.0	6.8	22.3	6.3	*At*
4	malate dehydrogenase	2683	5929964	131	10	57	35.0	6.1	36.1	8.2	*Gm*
2	malate dehydrogenase Glycine max	2534	5929964	13	4	17	37.0	5.9	36.1	8.2	*Gm*
4	malate dehydrogenase, cytoplasmic	1785	18202485	33	3	8	36.0	6.5	35.6	5.8	*Zm*
3	NADH-cytochrome b5 reductase, putative	681	18420117	29	2	5	32.0	7.4	36.0	8.8	*At*
4	NADPH dependent mannose 6-phosphate reductase	2802	21554266	18	2	5	34.0	5.9	35.1	6.1	*At*
4	vacuolar H+-ATPase A2 subunit isoform	2515	27884018	110	9	17	67.0	5.1	68.7	5.3	*Ls*
4	VHA-A	297	15219234	39	4	8	67.0	4.9	68.8	5.1	*At*
	02 Energy										
	02.30 Photosynthesis										
1	33kDa precursor protein of oxygen-evolving complex	2407	809113	34	3	5	32.0	5.1	35.3	5.9	*St*
1	33kDa precursor protein of oxygen-evolving complex	519	809113	37	3	5	31.0	5.0	35.3	5.9	*St*
4	ATP synthase CF1 beta subunit	445	91214126	159	14	37	47.0	5.3	53.8	5.3	*Gm*
2	ribulose 1 5 bisphosphate carboxylase small subunit	1743	1055368	17	4	23	15.8	5.3	20.0	8.9	*Gm*
2	ribulose-1,5-bisphosphate carboxylase small subunit	2850	10946375	47	3	18	16.0	7.8	20.0	8.9	*Gm*
4	ribulose-1,5-bisphosphate carboxylase small subunit	2008	10946375	24	3	18	13.0	6.8	20.0	8.9	*Gm*
3	ribulose-1,5-bisphosphate carboxylase large subunit	2550	91214125	71	6	13	50.0	5.9	52.6	6.0	*Gm*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accesion Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
	03 Cell growth/division										
	03.19 Recombination/repair										
3	proliferating cell nuclear antigen Glycine max	2481	18726	13	4	14	33.0	4.7	0.0	12-3400	*Gm*
	03 Cell growth/division										
	03.22 Cell cycle										
1	transitional endoplasmic reticulum ATPase	3279	98962497	140	12	16	70.0	4.2	89.9	5.1	*Nt*
2	transitional endoplasmic reticulum ATPase	2421	11265361	106	10	12	100.0	5.1	93.6	5.4	*At*
4	transitional endoplasmic reticulum ATPase	2021	11265361	104	9	10	100.0	5.1	93.6	5.4	*At*
	03 Cell growth/division										
	03.30 Seed maturation										
2	desiccation protectant protein homolog of Lea14	369	472850	13	2	16	19.0	4.8	19.0	4.5	*Gm*
4	desiccation protectant protein homolog of Lea14	652	472850	54	6	45	24.9	4.6	16.5	4.7	*Gm*
2	seed maturation protein PM34	435	9622153	16	2	9	34.0	6.4	31.8	6.6	*Gm*
3	seed maturation protein PM34	2151	9622153	41	4	15	30.0	6.3	31.8	6.6	*Gm*
	04.10 tRNA synthesis										
5	glycyl-tRNA synthetase / glycine--tRNA ligase	429	15292923	50	4	5	78.0	6.4	82.0	6.6	*At*
	04.19 mRNA synthesis										
3	elongation factor EF-2	468	6056373	139	12	18	92.0	6.1	94.2	5.9	*At*
	04.1901 General TFs										
3	F28C11.12 (Arf1-Arf5-like subfamily)	597	8778579	33	3	14	19.5	6.5	28.4	9.0	*At*
4	F28C11.12 (Arf1-Arf5-like subfamily)	1858	8778579	45	4	17	26.0	8.5	28.4	9.0	*At*
4	glycine rich RNA binding protein Glycine max	408	5726567	13	1	9	15.8	5.3	15.8	7.2	*Gm*
2	glycine-rich RNA-binding protein	2027	5726567	32	3	23	15.5	5.1	15.8	6.6	*Gm*
3	glycine-rich RNA-binding protein	1633	5726567	75	6	50	20.0	6.4	15.8	6.6	*Gm*
4	glycine-rich RNA-binding protein	375	5726567	74	6	54	16.5	6.3	15.8	6.6	*Gm*
1	novel calmodulin-like protein	3074	1235664	14	1	6	16.0	4.0	21.0	4.8	*Os*
	05 Protein synthesis										
	05.01 Ribosomal proteins										
2	60S ribosomal protein L10 (RPL10C)	477	18408550	31	3	14	25.0	3.9	24.9	10.6	*At*
	05 Protein synthesis										
	05.04 Translation factors										
4	eukaryotic initiation factor 3I1 subunit	708	12407664	21	2	6	35.0	6.4	36.5	6.9	*At*
4	translation initiation factor	230	2286151	50	5	12	45.0	4.6	47.0	5.4	*Zm*
3	translational elongation factor EF-TuM	528	11181616	99	7	16	40.0	6.0	48.5	6.0	*Zm*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accesion Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
	05 Protein synthesis										
	05.99 Others										
3	calreticulin-1	520	117165712	93	7	18	58.0	4.4	48.2	4.4	*Gm*
4	calreticulin-1	258	117165712	53	4	13	34.8	4.3	48.2	4.4	*Gm*
	06 Protein destination and storage										
	06.01 Folding and stability										
4	BiP	2059	62433284	304	24	36	72.0	5.0	73.6	5.1	*Gm*
4	CCH (COPPER CHAPERONE)	132	6525011	14	1	9	41.0	4.4	13.6	4.7	*Gm*
3	chaperonin 21 precursor	537	7331143	21	2	8	24.0	5.1	26.6	6.9	*Ls*
1	cyclophilin	2193	17981611	62	5	44	18.0	10.3	18.2	8.7	*Gm*
3	cyclophilin	607	17981611	69	5	29	19.0	7.6	18.2	8.7	*Gm*
1	endoplasmic reticulum HSC70-cognate binding precursor	2722	2642238	122	10	18	65.0	4.9	73.6	5.2	*Gm*
2	endoplasmic reticulum HSC70-cognate binding precursor	422	2642238	156	11	20	65.0	5.0	73.6	5.2	*Gm*
2	endoplasmic reticulum HSC70-cognate binding precursor	2636	2642238	224	17	25	72.0	5.1	73.6	5.2	*Gm*
4	endoplasmic reticulum HSC70-cognate binding precursor	3160	2642238	75	8	12	71.0	5.0	73.6	5.2	*Gm*
5	heat shock protein 70	2986	6746592	95	7	9	71.0	4.7	77.1	5.1	*At*
3	MnSOD (superoxide dismutase)	2238	147945633	41	3	12	25.4	7.9	26.7	8.6	*Gm*
1	MTHSC70-1 (mitochondrial heat shock protein 70-1)	1875	30691626	68	5	8	72.0	4.8	73.1	5.5	*At*
3	MTHSC70-1 (mitochondrial heat shock protein 70-1)	2386	30691626	82	6	11	56.0	4.3	73.1	5.5	*At*
	06 Protein destination and storage										
	06.07 Modification										
4	protein disulfide isomerase-like protein	699	49257111	112	8	26	37.0	5.5	40.4	5.7	*Gm*
4	protein disulfide isomerase-like protein	3103	49257115	84	6	18	37.0	5.8	39.8	6.0	*Gm*
	06 Protein destination and storage										
	06.10 Complex assembly										
3	CPN60A (chloroplast / 60 kDa chaperonin alpha subunit)	288	21554572	46	5	8	52.5	5.0	62.1	5.0	*At*
	06 Protein destination and storage										
	06.13 Proteolysis										
4	20S proteasome alpha 4 subunit	3190	125662835	45	4	16	25.6	8.0	27.2	8.3	*Zm*
1	26S proteasome AAA-ATPase subunit RPT5a	728	23197864	32	3	7	43.0	5.0	47.5	5.0	*At*
1	26S proteasome non-ATPase regulatory subunit	2454	21592398	20	2	6	34.0	6.3	34.4	6.4	*At*
3	26S proteasome non-ATPase regulatory subunit	537	21592398	57	4	25	24.0	5.1	34.4	6.4	*At*
3	cytosol aminopeptidase family protein	2222	15235763	46	3	6	59.0	5.6	61.3	6.6	*At*
3	PAB1 (20S proteasome alpha subunit B1)	2751	15219317	54	4	25	24.0	5.3	25.7	5.5	*At*
4	PAB1 (20S proteasome alpha subunit B1)	2319	15219317	27	3	16	24.5	5.4	25.7	5.5	*At*
4	PBA1 (20S proteasome beta subunit A 1)	648	79325892	41	3	12	22.0	5.4	25.3	5.3	*At*
5	PBE1 (20S proteasome beta subunit E1)	1626	14594931	42	4	26	17.0	8.6	18.6	9.2	*Nt*
4	putative alpha7 proteasome subunit	475	14594925	70	5	22	28.0	5.5	27.2	6.1	*Nt*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accesion Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
3	putative beta4 proteasome subunit	561	14594929	26	2	11	13.0	5.9	14.5	6.4	*Nt*
1	RPT5B (26S PROTEASOME AAA-ATPASE SUBUNIT RPT5B)	2234	15217431	115	9	31	44.0	5.0	47.0	4.9	*At*
2	subtilisin-type protease precursor	2207	11611651	23	2	2	18.0	5.4	82.7	6.9	*Gm*
3	subtilisin-type protease precursor	482	11611651	102	8	14	49.8	5.7	82.7	6.9	*Gm*
3	subtilisin-type protease precursor	2042	11611651	163	10	20	60.0	6.4	82.7	6.9	*Gm*
4	subtilisin-type protease precursor	246	11611651	14	1	0	20.0	5.6	82.7	6.9	*Gm*
4	subtilisin-type protease precursor	491	11611651	68	6	9	52.0	6.4	82.7	6.9	*Gm*
3	UBC35	1774	3834310	47	4	30	17.0	5.3	18.3	5.9	*At*
4	UBC36 (UBIQUITIN CONJUGATING ENZYME 36)	2000	18394416	59	5	39	16.0	6.0	17.2	6.7	*At*
4	UBC9 (UBIQUITIN CONJUGATING ENZYME 9)	1954	18417097	16	1	6	30.0	8.2	20.2	7.0	*At*
	06 Protein destination and storage										
	06.20 Storage proteins										
5	beta-conglycinin alpha subunit	2986	14245736	137	11	20	65.0	5.0	70.3	5.1	*Gm*
3	beta-conglycinin alpha-subunit	155	15425633	79	7	11	60.0	5.0	72.5	5.3	*Gm*
4	beta-ketoacyl-ACP synthetase I-2	491	7385203	51	4	8	47.0	6.4	49.8	7.6	*Gm*
1	glycinin	615	18641	68	5	12	21.0	10.6	63.9	5.2	*Gm*
1	glycinin	2497	4249568	47	3	8	12.5	5.6	63.8	5.2	*Gm*
2	glycinin	477	18641	50	4	9	22.0	10.3	63.9	5.2	*Gm*
2	glycinin	233	18641	70	5	12	22.0	10.2	63.9	5.2	*Gm*
4	glycinin	458	18641	31	3	6	31.0	7.9	63.9	5.2	*Gm*
4	glycinin	2768	18641	17	2	4	32.0	7.9	63.9	5.2	*Gm*
1	glycinin A1bB2-784	1935	18609	58	4	12	21.0	7.7	54.3	5.6	*Gm*
1	glycinin A1bB2-784	108	18615	63	4	9	51.5	6.0	55.5	6.2	*Gm*
1	glycinin A1bB2-784	250	18609	57	5	14	52.0	5.1	54.3	5.6	*Gm*
1	glycinin A1bB2-784	2247	18615	58	4	9	21.0	8.6	55.5	6.2	*Gm*
2	glycinin A1bB2-784	290	18609	80	6	20	32.0	5.1	54.3	5.6	*Gm*
3	glycinin A1bB2-784	2040	18609	25	3	7	52.0	5.3	54.3	5.6	*Gm*
3	glycinin A1bB2-784	347	18615	42	4	8	23.0	6.3	55.5	6.2	*Gm*
4	glycinin A1bB2-784	2264	18609	93	6	20	32.0	5.1	54.3	5.6	*Gm*
1	prepro beta-conglycinin alpha prime subunit	1731	32328882	97	11	20	67.0	5.3	72.2	5.5	*Gm*
2	prepro beta-conglycinin alpha prime subunit	2722	32328882	117	11	21	68.0	5.3	72.2	5.5	*Gm*
3	prepro beta-conglycinin alpha prime subunit	1675	32328882	117	10	21	70.0	5.6	72.2	5.5	*Gm*
5	prepro beta-conglycinin alpha prime subunit	2982	32328882	120	10	21	67.0	5.3	72.2	5.5	*Gm*
	07 Transporters										
	07.01 Ions										
2	chloroplast ferritin	3144	117650780	49	5	20	25.0	5.1	28.0	5.6	*Gm*
3	chloroplast ferritin	497	117650780	66	6	32	75.0	6.7	28.0	5.6	*Gm*
4	manganese-superoxide dismutase	3276	27526758	67	5	33	25.5	6.3	15.4	6.1	*Gm*
2	EDA9 (embryo sac development arrest 9)	2460	15235282	27	3	5	54.0	5.6	63.3	6.2	*At*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
	08 Extracellular matrix										
	08.01 Nuclear										
3	Ran1	718	123192431	38	3	12	29.0	6.3	25.3	6.4	*Ps*
4	Ran1	3084	123192431	32	3	15	25.0	6.4	25.3	6.4	*Ps*
	09 Cell structure										
	09.01 Cell wall										
1	3-deoxy-D-manno-2-octulosonic acid-8-phosphate	3199	32169731	31	3	20	15.0	7.1	19.5	7.6	*Nt*
4	UXS3 (UDP-GLUCURONIC ACID DECARBOXYLASE)	1829	145334845	42	3	9	38.0	7.0	40.2	8.5	*At*
	09 Cell structure										
	09.04 Cytoskeleton										
1	actin	3275	50058115	147	10	35	40.0	5.2	41.6	5.3	*Nt*
1	actin	3275	1666234	76	7	20	41.0	5.1	42.7	5.3	*Ps*
4	actin	555	1498346	57	5	14	32.0	4.6	37.2	5.3	*Gm*
4	alpha tubulin-4A	3236	90289610	110	8	24	49.0	5.1	49.8	4.8	*Ta*
4	beta tubulin	498	15451226	124	11	27	50.0	4.5	50.6	4.8	*At*
1	profilin	1934	156938901	27	3	24	13.2	4.6	14.1	4.7	*Gm*
1	profilin	982	156938901	18	2	16	12.0	4.5	14.1	4.7	*Gm*
3	TUB2 (Tubulin beta-2)	622	18424620	21	2	4	55.0	4.9	50.7	4.7	*At*
4	TUB2 (Tubulin beta-2)	331	18424620	157	12	34	50.0	5.0	50.7	4.7	*At*
5	TUB2 (Tubulin beta-2)	331	18424620	106	11	26	50.0	5.0	50.7	4.7	*At*
3	tubulin A	1432	62546341	53	5	14	47.0	4.3	49.7	5.0	*Gm*
4	tubulin A	1724	62546341	143	10	31	47.0	5.1	49.7	5.0	*Gm*
4	tubulin B4	492	62546343	171	14	34	48.0	4.9	50.4	4.7	*Gm*
	09 Cell structure										
	09.16 Mitochondria										
1	34 kDa outer mitochondrial membrane porin	2425	83283993	29	2	10	31.0	9.3	29.4	7.7	*St*
	10 Signal transduction										
	10.0404 Kinases										
2	adenosine kinase isoform 2S	1877	51949802	25	2	9	36.0	4.9	37.6	5.2	*Nt*
	10 Signal transduction										
	10.99 Others										
4	14-3-3-like protein	514	4775555	62	6	18	27.0	4.3	29.4	4.7	*Ps*
	11 Disease/defence										
	11.01 Resistance genes										
5	universal stress protein (USP) family protein	1142	30693971	18	2	10	16.0	5.4	17.8	5.7	*At*
	11 Disease/defence										
	11.02 Defence-related										
1	24 kDa protein SC24	1052	18448973	26	2	10	20.0	10.3	24.6	9.1	*Gm*
1	24 kDa protein SC24	1028	18448973	57	4	18	23.0	9.0	24.6	9.1	*Gm*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accesion Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
	11 Disease/defence										
	11.05 Stress responses										
2	caffeoyl-CoA 3-O-methyltransferase 5	544	2511737	26	2	6	29.0	5.4	27.2	5.4	*Nt*
2	caffeoyl-CoA 3-O-methyltransferase 5	636	2511737	26	2	6	29.0	5.4	27.2	5.4	*Nt*
	11 Disease/defence										
	11.06 Detoxification										
2	alcohol dehydrogenase 1	515	22597178	88	7	17	37.0	6.8	40.0	6.2	*Gm*
3	alcohol dehydrogenase 1	2066	22597178	52	4	10	37.0	7.0	40.0	6.2	*Gm*
4	alcohol dehydrogenase 1	2962	22597178	90	7	17	39.0	6.7	40.0	6.2	*Gm*
4	alcohol dehydrogenase 1	2903	22597178	77	6	15	39.0	6.7	40.0	6.2	*Gm*
4	alcohol dehydrogenase 1	661	22597178	14	2	4	16.8	5.6	40.0	6.2	*Gm*
4	alcohol dehydrogenase 1	453	22597178	19	2	5	31.0	6.6	40.0	6.2	*Gm*
4	alcohol dehydrogenase 1	2478	22597178	60	6	14	42.0	6.8	40.0	6.2	*Gm*
2	alcohol-dehydrogenase	546	4039115	100	8	25	39.0	6.1	36.4	6.1	*Gm*
3	aldehyde dehydrogenase family 7 member A1	3036	29893325	43	4	8	54.0	5.6	54.7	5.5	*Gm*
3	ALDH2B4 (ALDEHYDE DEHYDROGENASE 2)	451	15228319	38	3	6	50.0	6.6	58.6	7.1	*At*
4	ALDH2B4 (ALDEHYDE DEHYDROGENASE 2)	620	15228319	57	4	8	50.0	6.3	58.6	7.1	*At*
4	aldo/keto reductase family protein	534	4895205	43	3	7	32.0	6.3	39.0	6.8	*At*
2	ascorbate peroxidase	1888	4406539	21	1	7	19.0	5.9	27.2	5.7	*Gm*
3	ascorbate peroxidase 2	354	1336082	150	9	52	26.5	5.6	27.1	5.7	*Gm*
4	CAT2 (CATALASE 2)	2616	15236264	28	3	6	52.0	7.1	56.9	6.6	*At*
4	Chain A The Structure Of Soybean Peroxidase	719	13399943	54	4	23	20.0	4.6	20.0	4.4	*Gm*
3	cytosolic ascorbate peroxidase 1	355	37196683	31	3	12	52.9	6.9	27.9	5.8	*Gm*
4	cytosolic ascorbate peroxidase 1	99	37196683	88	5	28	26.0	5.5	27.9	5.8	*Gm*
4	cytosolic ascorbate peroxidase 1	530	37196683	37	3	12	26.0	5.3	27.9	5.8	*Gm*
3	dehydroascorbate reductase	561	28192427	21	1	8	23.0	5.8	23.7	7.7	*Nt*
4	dehydroascorbate reductase	183	28192427	20	1	8	23.0	5.6	23.7	7.7	*Nt*
4	DHAR2	183	21593056	16	2	10	23.0	5.6	23.4	6.0	*At*
4	glutathione peroxidase Phaseolus lunatus	554	62946785	13	3	16	20.0	5.5	11.9	4.9	*Pl*
2	In2-1 protein	2633	11385579	164	10	49	49.8	5.0	27.0	5.2	*Gm*
2	In2-1 protein	2636	11385579	159	11	54	26.0	5.3	27.0	5.2	*Gm*
3	In2-1 protein	441	11385579	85	6	33	90.0	6.3	27.0	5.2	*Gm*
3	In2-1 protein	556	11385579	25	3	12	25.0	4.8	27.0	5.2	*Gm*
3	In2-1 protein	2652	11385579	59	5	20	24.0	5.2	27.0	5.2	*Gm*
2	L ascorbate peroxidase Medicago sativa	2605	168067	14	2	17	29.0	5.4	0.0	14.0	*14*
1	mitochondrial peroxiredoxin	330	47775654	13	1	4	17.0	6.9	21.5	8.4	*Ps*
1	peroxidase	3287	17467210	15	2	4	48.0	4.9	38.1	5.0	*Gm*
2	peroxidase	165	17467210	26	2	6	43.0	4.1	38.1	5.0	*Gm*
2	peroxidase	9	17467210	26	2	6	39.5	4.1	38.1	5.0	*Gm*

Table continues next page

Table 3.2 (Continued from previous page)

Expression cluster	Protein	Spot No. PG220	NCBI Accesion Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	In gel MW	In gel pI	Theoretical MW	Theoretical pI	Species
2	peroxidase	417	17467210	29	2	10	22.5	4.5	38.1	5.0	*Gm*
2	peroxidase	222	17467210	16	2	4	36.0	4.8	38.1	5.0	*Gm*
5	peroxidase 73 (PER73) (P73) (PRXR11)	1729	15240737	14	1	3	34.0	9.3	35.9	9.4	*At*
2	phospholipid hydroperoxide glutathione peroxidase	1888	46200528	15	1	6	18.0	6.0	19.4	8.2	*Zm*
4	ripening regulated protein DDTFR10-like	352	78191406	17	2	6	23.0	4.2	25.3	4.5	*St*
4	short chain alcohol dehydrogenase	385	2739279	14	1	5	27.7	5.4	29.8	6.2	*Nt*
2	thioredoxin fold Arachis hypogaea	2664	115187464	17	5	34	14.8	5.9	17.4	5.5	*Ah*
4	thioredoxin fold Arachis hypogaea	593	115187464	17	5	34	16.8	5.4	17.4	5.5	*Ah*
1	thioredoxin M-type, chloroplast precursor (TRX-M)	1615	3334376	24	2	11	14.0	5.1	18.1	8.7	*Zm*
	12 Unclear classification										
2	dormancy related protein, putative	638	12322163	13	1	2	31.0	6.5	31.2	5.9	*At*
1	Kunitz trypsin inhibitor	3225	13375349	82	7	34	18.0	4.6	24.1	5.0	*Gm*
2	Kunitz trypsin inhibitor	575	13375349	48	4	19	17.8	4.9	24.1	5.0	*Gm*
	13 Unclassified										
4	dd15	2364	28542706	40	3	50	34.0	6.5	5.8	5.3	*Gm*
2	root border cell-specific protein-like protein	2031	82400120	22	2	4	32.0	5.4	36.0	8.3	*St*
1	unnamed protein product	729	18615	50	3	7	52.0	6.1	55.5	6.2	*Gm*
2	unnamed protein product	2264	18615	34	3	8	31.0	5.1	55.5	6.2	*Gm*
2	unnamed protein product	454	18615	17	2	6	29.0	5.4	55.5	6.2	*Gm*
3	unnamed protein product	2752	18615	27	3	7	17.0	6.9	55.5	6.2	*Gm*
4	unnamed protein product	170	18615	48	4	10	34.0	5.2	55.5	6.2	*Gm*
	20 Secondary metabolism										
	20.1 Phenylpropanoids/phenolics										
2	chalcone isomerase	4444	14582263	63	4	20	23.0	6.0	23.3	6.2	*Gm*
2	isoflavone reductase homolog 1	484	6573169	25	2	8	19.0	5.5	33.9	5.8	*Gm*
4	isoflavone reductase homolog 1	338	6573169	110	7	32	33.0	5.6	33.9	5.8	*Gm*
4	isoflavone reductase homolog 1	2712	6573169	75	6	28	35.0	6.1	33.9	5.8	*Gm*
3	isoflavone reductase homolog 2	2232	6573171	85	8	28	33.0	5.6	33.9	5.6	*Gm*
4	isoflavone reductase homolog 2	1402	6573171	89	7	23	34.8	5.5	33.9	5.6	*Gm*
4	isoflavone reductase homolog 2	1719	6573171	69	7	23	33.0	5.5	33.9	5.6	*Gm*
	20 Secondary metabolism										
	20.99 Others										
4	myo-inositol-3-phosphate synthase	328	13936691	119	10	23	48.0	5.4	56.5	5.3	*Gm*

Table ends

3.3.5. Cluster analysis of 565 differentially expressed protein spots

To obtain a general idea of protein expression changes, we clustered 565 differentially expressed protein spots into five groups according to their expression patterns. The general hypothesis of *k* – means cluster analysis is that proteins engaged in a similar function or similar metabolic pathway will have similar profile of expression and thus likely to be grouped into the same group. The average pattern of the protein spots that are included in each specific set are shown in Figure 3.2. Each group consists of several proteins with some functional correlations. The corresponding spot identification by mass spectrometry is presented in Table 3.2.

Group A included proteins that accumulated to a moderate level until 50 DPA and then were substantially up regulated. Proteins that belong to the destination and storage class like beta-conglycinin alpha subunit and several isoforms of glycinin (storage) and 26S proteasome AAA-ATPase subunit (proteolysis), Kunitz trypsin inhibitor (unclear classification), 24 kDa 24SC (defence-related) and actin (cell structure) were grouped together in cluster A.

Group B included proteins that sharply increased until 50 DPA and then their levels decreased dramatically. The most abundant proteins found in this cluster were endoplasmic reticulum heat shock protein (protein destination and storage, folding and stability); In 2-1 protein (herbicide safener inducible 27 kDa) and alcohol dehydrogenase (disease/defence, detoxification); chloroplast ferritin (ion transporter).

Group C included proteins that were expressed at moderate levels at 20 DPA and their expression was slowly down-regulated during the next developmental stages. Representative proteins of this cluster include elongation factor EF-2 (transcription, mRNA synthesis), subtilisin-protease precursor (protein destination and storage, proteolysis), enolase (energy, glycolysis), isoflavone reductase homolog 1 and 2 (secondary metabolism, phenylpropanoids), 6-phosphogluconate dehydrogenase (energy, penthose

phosphate), rhamnose biosynthesis 1 (metabolism, sugars and polysaccharides), homoglutathione synthetase (disease/defence, detoxification) and calreticulin-1 (protein synthesis, translation factor).

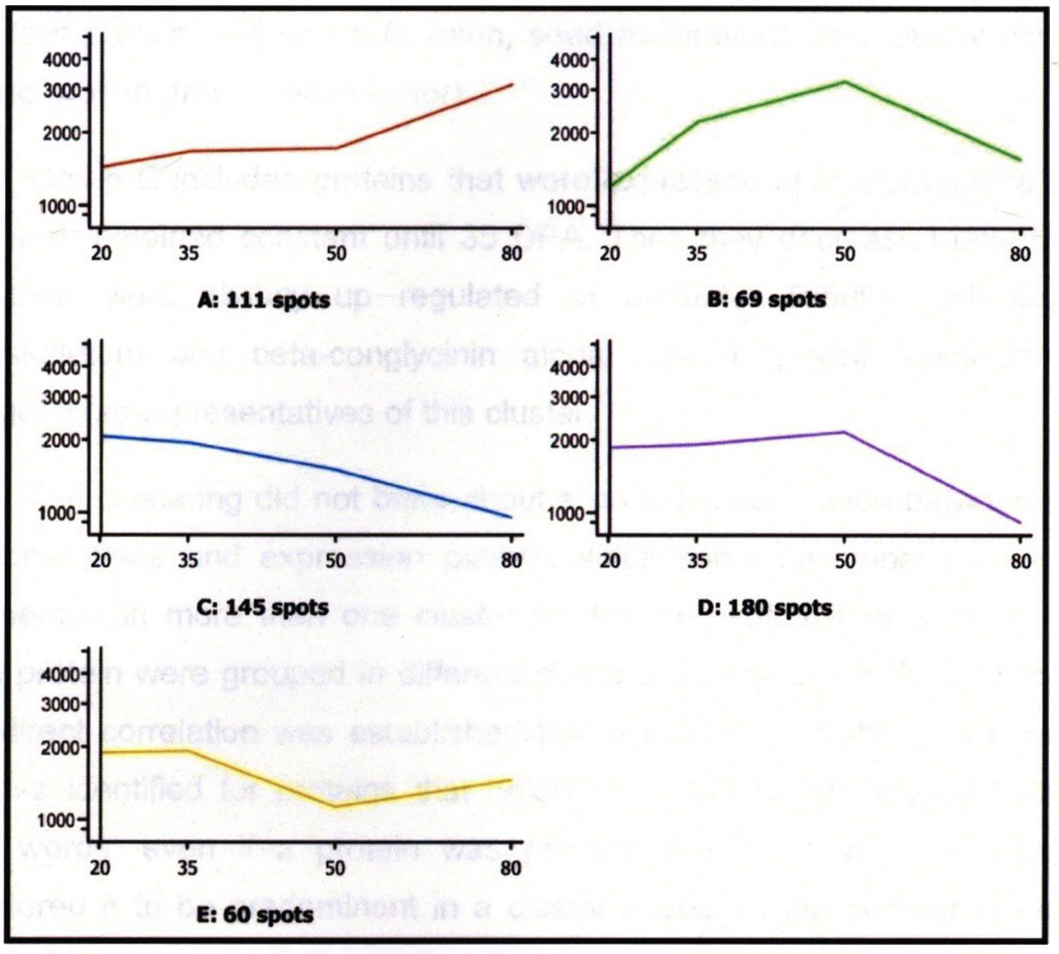


Figure 3.2 Cluster analysis of 565 differentially expressed protein spots in soybean seed coats. The proteins were classified using the *k*-means technique into five groups. Spots belonging to each group were averaged together and presented. The y axis is the normalized level of expression as a function of developmental stages (20, 35, 50 and 80 DPA).

Group D was constituted by proteins whose expression starts at relatively moderate levels that plateau until 50 DPA, and then decrease sharply. Most spots clustered together in this group. The most abundant spots corresponded to BiP (protein destination and storage, folding and stability), methionine synthase (amino acid metabolism), glutamate dehydrogenase 1 (sugar metabolism), soybean peroxidase SBP (disease/defence, detoxification), late embryogenesis abundant protein (cell growth/division, seed maturation), and glycine-rich RNA-binding protein (transcription factor).

Group E included proteins that were expressed at moderate levels at 20 DPA and remained constant until 35 DPA. Then they decreased until 50 DPA and then were slightly up regulated at maturity. Tubulin (cell structure, cytoeskeleton) and beta-conglycinin alpha subunit (protein destination and storage) were representatives of this cluster.

The clustering did not bring about a clear-cut correlation between protein functional class and expression pattern, since some functional classes were represented in more than one cluster or different isoforms or subunits of the same protein were grouped in different clusters. To overcome this shortcoming, an indirect correlation was established taking into account the total number of peptides identified for proteins that belonged to the same functional class. In other words, even if a protein was present in more than one cluster, we considered it to be predominant in a cluster based on the number of peptides identified as presented in the next section.

3.3.6. Composite expression profiles of plant functional classes reveal different expression trends

To characterize global expression trends of proteins involved in different processes, we established composite expression profiles by summing protein abundance, expressed as normalized volume, for each protein in each functional class for the four seed coat developmental stages (Figure 3.3).

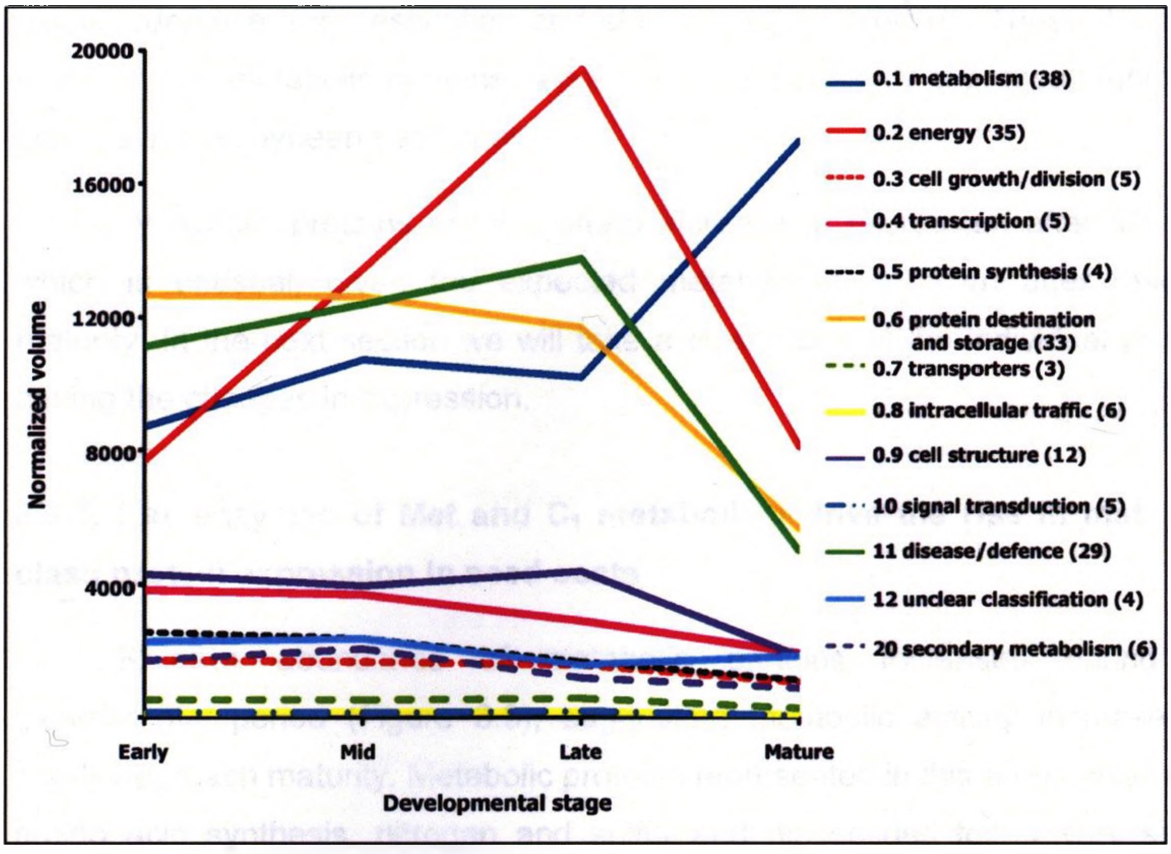


Figure 3.3 Composite protein expression profiles of gene functional categories. The combined expression profiles were calculated as the sum of all normalized volumes for each protein in the functional category. Gene functional classes are the ones presented in Table 3.2 following classification according to Bevan et al, (1988). Shown in parenthesis are the number of proteins engaged in each of the functional categories whose normalized volume was added to create the composite.

The expression trend for energy-related proteins reaches a maximum at 50 DPA, to sharply decrease thereafter. This same trend is followed by plant disease/defence and destination and storage-related proteins. These 3 groups, together with metabolic proteins, represent the most highly expressed functional classes in the soybean seed coat.

Metabolic proteins show a sharp increase in expression after 50 DPA, which is unusual, given the expected metabolic slow down after reaching maturity. In the next section we will take a closer look at the individual proteins driving the changes in expression.

3.3.7. Key enzymes of Met and C_1 metabolism drive the rise in metabolic class protein expression in seed coats

Relative abundance of metabolic proteins increased during the experimental period (Figure 3.3), suggesting metabolic activity increases as seeds approach maturity. Metabolic proteins represented in this group involved in amino acid synthesis, nitrogen and sulfur and nucleotides follow very similar patterns, but the ones involved in metabolism of sugars and lipids did not increase at later developmental stages (Figure 3.4A). Within the group of proteins involved in amino acid metabolism, methionine synthase (33325957) and serine hydroxymethyltransferase 4 (11762130) follow the same trend (Figure 3.4B); that is, after a down regulation at 50 DPA, their expression peaks at 80 DPA.

Other proteins involved in amino acid metabolism, such as aspartate aminotransferase (2654094), glutamate—ammonia ligase (547508) and glutamine synthase (10946357) were also up-regulated at maturity but in a less dramatic fashion. Cysteine synthase (126508784) remained constant throughout development. The slight differences in expression profiles both in terms of trends and relative amount of expression of proteins involved in amino acid synthesis is an indication of a fine tuned mechanism of amino acid regulation in the seed coat.

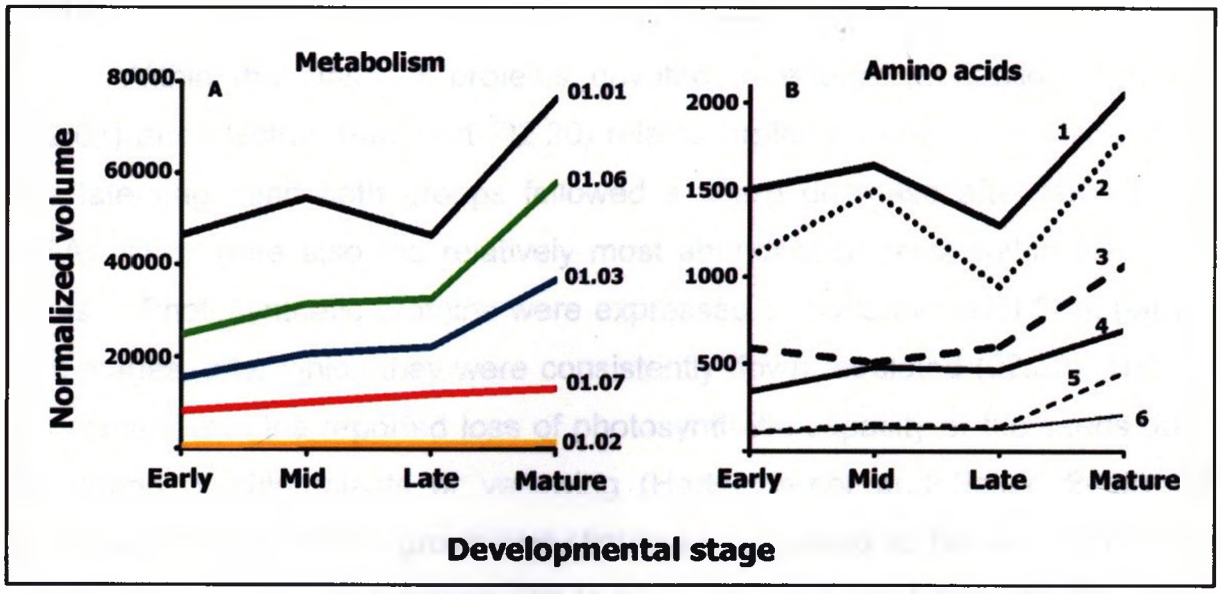


Figure 3.4 Composite expression profiles of metabolic related proteins. A) The combined expression profiles were calculated as the sum of all relative volumes for each protein in the metabolic functional category. The combined expression of 11 proteins involved in metabolism of amino acids (01.01), 16 in sugars and polysaccharides (01.06), 3 in nucleotides (01.03), 3 in lipids (01.07), 1 in nitrogen and sulphur (01.02) are presented. B) Composite expression profiles of most abundant proteins from amino acid metabolism class are presented, methionine synthase (1), serine hydroxymethyltransferase (2), glyoxysomal aspartate aminotransferase (3), glutamate--ammonia ligase (4), glutamine synthase (5) and cysteine synthase (6).

3.3.8. Glycolysis is the dominant energy-related process in soybean seed coats.

Within the class of proteins devoted to energy production, glycolysis (02.01) and electron transport (02.20) related proteins were up-regulated during the late-stage and both groups followed a sharp decrease afterwards (Figure 3.5A). They were also the relatively most abundant proteins within the energy class. Photosynthetic proteins were expressed a moderate level from early- to mid-stages, after which they were consistently down-regulated (02.30). This is in agreement with the reported loss of photosynthetic capacity of the seeds due to disruption of chloroplasts or yellowing (Hortensteiner and Feller, 2002). The gluconeogenesis (02.02) group was slightly up-regulated at the late- and mature-stages, but overall, their expression levels were the lowest amongst the energy group.

The expression of most abundant glycolytic enzymes is presented in Figure 3.5B. Cytosolic phosphoglycerate kinase (9230771) was by far the glycolytic protein with highest expression (4-fold) as compared with the other proteins within the glycolysis class. The expression trend follows steady up regulation until-late stage, followed by sharp down regulation at maturity. At lower normalized volumes, enolase (42521309) and glyceraldehyde-3-phosphate dehydrogenase (85720768) follow the same trend of up-regulation at late stage followed by down regulation at maturity. Other glycolytic enzymes such as subunit A of glyceraldehyde-3-phosphate dehydrogenase 77540210), triosephosphate isomerase (48773765) and glycine hydroxymethyltransferase (7433553) were resolved in 2D spots and their expression during overall development did not present noticeable changes.

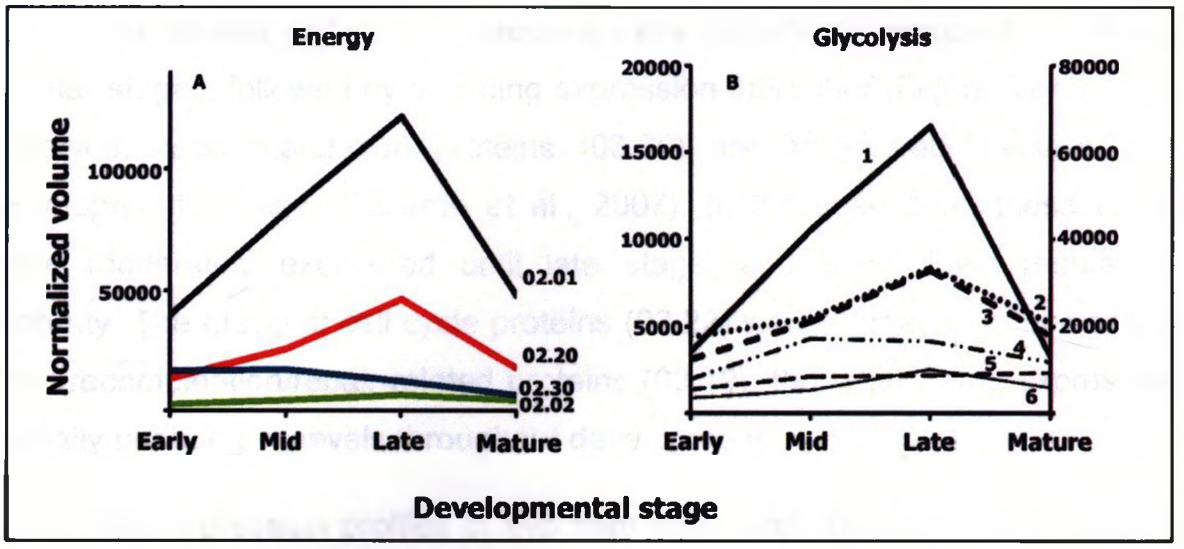


Figure 3.5 Composite expression profiles of energy related proteins. A) The combined expression profiles were calculated as the sum of all relative volumes for each protein in the energy-related functional category. The combined expression of 15 proteins involved in glycolysis (02.01), 9 in electron-transport (02.20), 5 in photosynthesis (02.30) and 2 in gluconeogenesis (02.02) are presented. B) Composite expression profiles of most abundant proteins from energy related class are presented, cytosolic phosphoglycerate kinase (1), enolase (2), glyceraldehyde-3-phosphate dehydrogenase (3), glyceraldehyde-3-phosphate dehydrogenase subunit A (4), triosephosphate isomerase (5) and glycine hydroxymethyltransferase (6). Graph B is presented in double y axis format to realistically represent the difference in levels of expression between cytosolic phosphoglycerate kinase and the rest of proteins presented.

3.3.9. Seed maturation proteins follow a similar expression profile throughout development.

Cell growth and division proteins were abundantly expressed from early- to late- stages, followed by declining expression thereafter (Figure 3.6A). In seed embryos, seed maturation proteins (03.20) are considered markers of the developmental stage (Gallardo et al., 2007). In the seed coat, these proteins were moderately expressed until late stage, and were down regulated at maturity. The group of cell cycle proteins (03.22) was expressed at higher levels than recombination/repair-related proteins (03.19), the latter being expressed at virtually unchanging levels throughout development.

The expression profiles of two maturation proteins are presented (Figure 3.6B). Seed maturation protein PM34 (9622153) was moderately expressed from early- to late- stages, after which its expression declined. Its expression was previously reported in soybean embryos (Hajduch et al., 2005) without level changes in development (spot 2537, (http://oilseedproteomics.missouri.edu/). A late-embryogenesis-abundant protein homolog to Lea14 (472850) was expressed at constant levels from early- to mid-stages, undergoing down-regulation thereafter. Such an expression pattern differs from that reported in developing soybean embryos, in that case embryonic Lea proteins were sharply down-regulated after 21 DPA, during the early stage of development (spots 900, 2225, 2233, 2304, etc). A potential explanation could be that the water-binding activity of these proteins (Maitra and Cushman, 1994) is required to preserve a certain degree of moisture for a prolonged time in the remaining layers of seed coat.

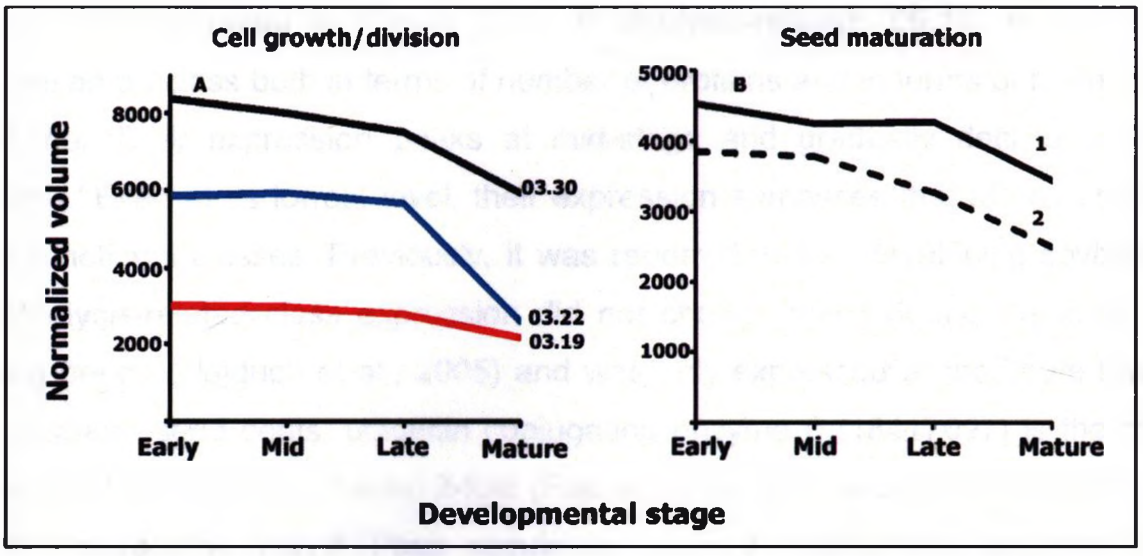


Figure 3.6 Composite expression profiles of cell growth and division related proteins. A) The combined expression profiles were calculated as the sum of all relative volumes for each protein in the respective functional category. The combined expression of 2 proteins involved in seed maturation (03.30), 2 in cell cycle (03.22) and 1 in recombination/repair (03.19) are presented. B) Expression profiles of the two seed maturation proteins from the functional class are presented, desiccation protectant protein homolog of Lea14 (1) and seed maturation protein PM34 (2).

3.3.10. Protein destination and storage sub classes show different expression profile trends.

The expression profiles of the protein destination and storage functional class are presented in Figure 3.7A. Proteolysis-related (06.13) is the most represented class both in terms of number of proteins and in terms of normalized volume. Their expression peaks at mid-stage and gradually decline in later stages. Even at its lowest level, their expression surpasses that of any other of the functional classes. Previously, it was reported that in developing soybeans, proteolysis-related class expression did not change much during the seed the filling period (Hajduch et al., 2005) and was only expressed at moderate levels. In soybean seed coats, ubiquitin conjugating enzyme 9 (18417097) is the most abundant protease by at least 2-fold (Figure 3.7B). Less abundant proteases are 26S proteasome non-ATPase regulatory subunit (21592398), subtilisin-type protease precursor (11611651), which is considered a seed coat tissue marker (Gallardo et al., 2007), 20S proteasome alpha subunit B1 (4) and cytosolic amino peptidase (5).

In the soybean seed coat, storage proteins (06.20) are expressed moderately from early- to mid-stages (Figure 3.7A), after which, there is a considerable increase in their levels to reach a peak at the late-stage, followed by slow decline. Storage proteins have previously been described to accumulate at the beginning of physiological maturity, in fact, they are considered a marker of such a stage. In soybean seeds, 54 storage proteins were reported to accumulate steadily starting at 21-28 DPA (early stage) until the end of seed filling (Hajduch et al., 2005); whereas, in *Medicago truncatula* the two major storage proteins reported (vicilin and legumin B) started accumulating at 20 DPA (seed filling period) but decreased sharply thereafter. The prolonged expression of storage proteins in the seed coat seems contradictory to the notion of protein turnover in the embryo surrounding tissues (endosperm and seed coat) that was previously reported (Gallardo et al., 2007). In soybean seed coat, the most abundant storage protein is glycinin with the different isoforms and precursors, which resolved in at least 7 different spots reported in Figure 3.7C.

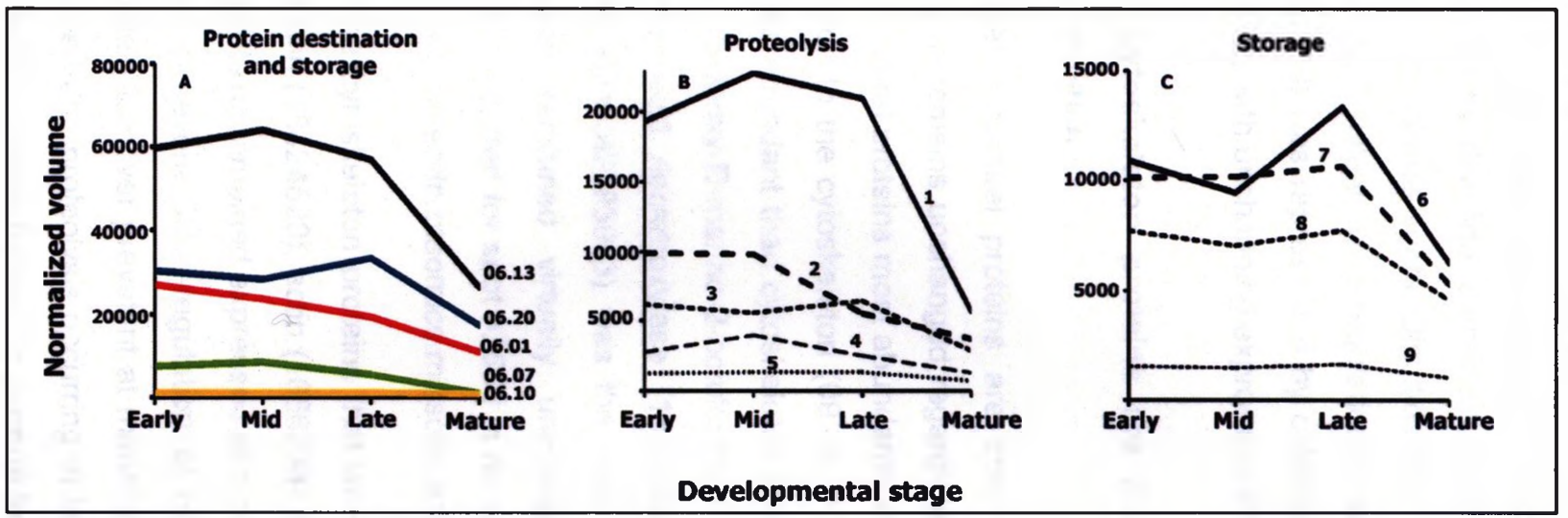


Figure 3.7 Composite expression profiles of protein destination and storage proteins. A) The combined expression profiles were calculated as the sum of all relative volumes for each protein in the respective functional category. The combined expression of 12 proteins involved in proteolysis (06.13), 8 in storage (06.20), 8 in folding and stability (06.01), 2 in protein modification (06.07) and 1 in complex assembly (06.10) are presented. B) Composite expression profiles of most abundant proteins from proteolytic class are presented, ubiquitin conjugating enzyme 9 (1), 26S proteasome non-ATPase regulatory subunit (2), subtilisin-type protease precursor (3), 20S proteasome alpha subunit B1 (4) and cytosolic amino peptidase (5). C) Composite expression profiles of most abundant proteins from storage protein class are presented, glycinin A1bB2-784 (6), glycinin (7) and prepro beta-conglycinin alpha prime subunit (8), beta conglycinin alpha subunit (9).

Chaperones and stability-related proteins (06.01) (Figure 3.7A) were most up-regulated earlier in development and their expression followed a declining trend throughout the study period. The group of protein modification (06.07) was represented by disulfide isomerase (49257111) which was resolved in at least two high-abundance spots. Disulfide isomerase expression was moderate up to mid stage, to follow down regulation in subsequent stages. Complex assembly class (06.10) was represented by chloroplast 60 kDa chaperonin alpha subunit (21554572) with unchanging expression levels throughout the study period.

3.3.11. Cytoskeleton proteins are the most abundant proteins in cell structure class.

Cell structural proteins are considered housekeeping proteins whose expression remains unchanged regardless experimental conditions. In the seed coat, structural proteins most abundantly expressed at late-stage of development are related to the cytoskeleton (09.04) (Figure 3.8A). Cell wall (09.01) proteins were less abundant than cytoskeleton proteins and were represented by only 2 entries (3-deoxy-D-manno-2-octulosonic-8-acid phosphate (32169731); UDP-glucoronic acid decarboxylase (145334845). Outer mitochondrial membrane porin 34 kDa (83283993) was the only mitochondrial protein (09.16) and its expression remained virtually unchanging during seed development, as previously reported for spot #2560 in developing soybean seeds (Hajduch et al., 2005)(http://oilseedproteomics.missouri.edu/).

The cytoskeleton proteins that were identified are tubulin A (62546341), beta tubulin (18424620), actin (1666234) and profilin (156938901) (Figure 3.8B). These proteins remained expressed at consistent levels from early to late stages, undergoing severe down-regulation at maturity. Considering the overall seed coat protein turnover prevalent at maturity, in part due to the desiccation of the whole seed and proteolysis occurring at levels of epidermis and hourglass cells, cytoskeleton proteins follow the general trend of housekeeping proteins and are conformed by smaller protein subunits (Higgins, 1984).

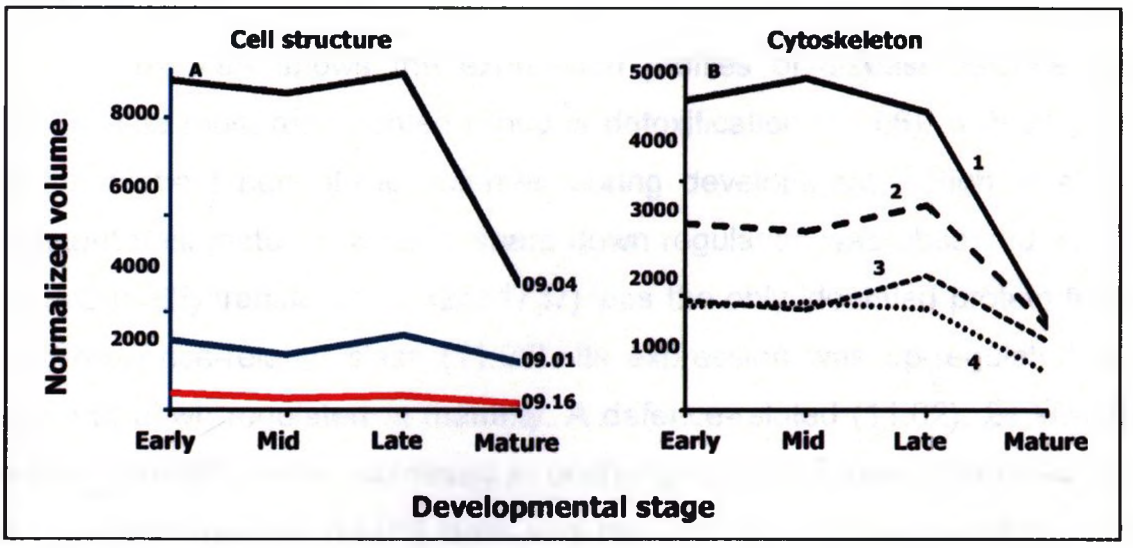


Figure 3.8 Composite expression profiles of cell structure proteins. A) The combined expression profiles were calculated as the sum of all relative volumes for each protein in the respective functional category. The combined expression of 9 cytoskeleton proteins (09.04), 2 in cell wall (09.01) and 1 in mitochondria (09.16) are presented. B) Expression profiles of the most abundant cytoskeleton proteins are presented, tubulin A (1), tubulin B2 (2), actin (3) and profilin (4).

3.3.12. Important role of detoxification proteins in the disease/defence functional class.

Figure 3.9A shows the expression profiles of disease/defence related proteins. The most represented group is detoxification (11.06), with 21 proteins and the highest normalized volumes during development. High levels were consistent until maturity, when a sharp down regulation was observed. Caffeoyl-CoA 3-O-methyltransferase 5 (2511737) was the only identified protein from the stress response-related class (11.05). Its expression was up-regulated at late stage and down regulated at maturity. A defence-related (11.02), 24 kDa SC24 protein (18448973) was expressed at unchanging levels during the study period. The resistance-related (11.01) class was represented by universal stress protein (USP) (30693971) with consistent low expression levels all along seed development.

Soybean peroxidase (17467210) (number 1 in Figure 3.9B) was consistently expressed at high levels from early through late stages, and underwent a sharp decline at maturity. Peroxidase resolved as the most abundant protein in at least 6 highly abundant 2D spots with MW within a range of 20 to 48 kDa, most likely representing different glycosylation events typical of this protein (Gray and Montgomery, 2006). Soybean peroxidase is one of the most abundant proteins present in the soybean seed coat, where it may constitute up to 10% of the total protein (Gillikin and Graham, 1991) and is sequestered mainly to the hourglass cells (Gijzen, 1997).

Another defence-related protein expressed in the seed coat is catalase (15236264) with the highest expression level early in development, and shows a decreasing trend thereafter. Ripening regulated protein DDTRF10-like (78191406) increased gradually until late-development but was expressed at relatively low levels.

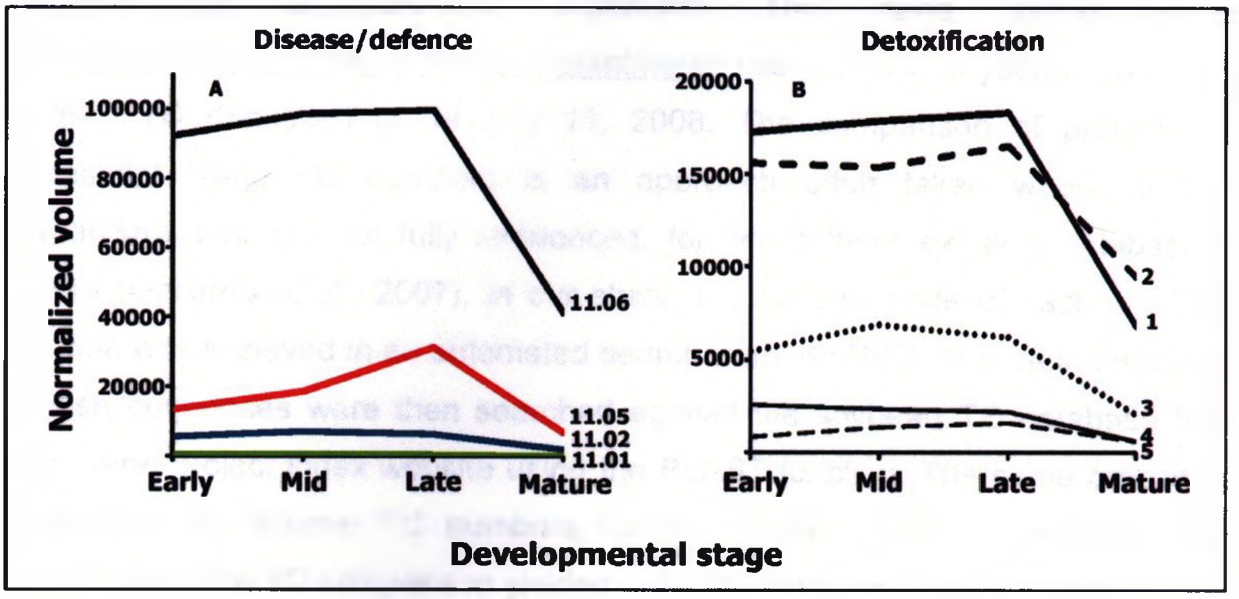


Figure 3.9 Composite expression profiles of disease and defence related proteins. A) The combined expression profiles were calculated as the sum of all relative volumes for each protein in the respective functional category. The combined expression of 21 proteins involved in detoxification (11.06), 1 in stress responses (11.05), 1 in defence-related (11.02) and 1 in resistance-related (11.01) are presented. B) Composite expression profiles of most abundant proteins from disease/defence class are presented, peroxidase (1), alcohol dehydrogenase 1 (2), 24 kDa protein SC24 (3), catalase (4) and ripening regulated protein DDTRF10-like (5).

3.3.13. Comparison of protein profiles with transcript patterns during seed coat development

In an attempt to compare seed coat proteomic data with transcriptome data, every identified protein was attributed a tentative contig number (TC), which is an alignment of overlapping expressed sequence tags (ESTs) for a particular organism. The gene index project (http://compbio.dfci.harvard.edu/tgi/plant.html) has 381524 soybean ESTs and 42647 TC deposited as of July 11, 2008. The comparison of protein with transcript using TC numbers is an approach often taken when studying organisms that are not fully sequenced, for which there exists a database of ESTs (Gallardo et al., 2007). In our study, the full sequence of each identified protein was retrieved in an automated search from the NCBI nr protein database. These sequences were then searched against the soybean TC database from the Gene project index website using the BLAST function. The same procedure was used to retrieve TC numbers for the 18,462 ESTs in the seed coat microarray. The TC comparison yielded only 348 matches (data not shown), only 8 corresponding to identified proteins. This mismatch could be due to the fact that ESTs represented in the microarray had often a gene identifier from a species different than *Glycine max*, making it difficult to find them in a soybean TC database (Dr. M. Stromvik, personal communication).

Selected identified proteins were correlated with transcript data from the microarray utilizing keyword searches. Figures 3.10 to 3.14 show comparisons of protein and transcript levels. Three different trends were observed, a coordinated expression of protein and transcripts (Figure 3.10-11); an apparent preferential transcript turnover during seed coat development (Figure 3.12) and a poor correlation between protein and transcript levels during the developmental period (Figure 3.13-3.14).

a) Protein and transcript levels coordinately expressed

Specific proteins whose transcripts were coordinately expressed during seed coat development are presented in Figures 3.10-11. This coordinated

expression suggests that protein accumulation is primarily regulated by transcript abundance. Aspartate aminotransferase (Figure 3.10A, 2654094) provides the source of aspartate used in the synthesis of the aspartate family of amino acids, including the agronomically important essential amino acids methionine and lysine (Gebhardt et al., 1998). Glutamate—ammonia ligase also called glutamine synthase (Figure 3.10B, 547508) is the enzyme responsible for the assimilation of ammonia into organic compounds (Marsolier *et al.*, 1995). D-apiose serves as the binding site for borate cross-linking of rhamnogalacturonan II in the plant cell wall, and biosynthesis of D-apiose involves UDP-D-apiose/UDP-D-xylose synthase 2 (Figure 3.10C, 18390863) catalyzing the conversion of UDP-D-glucuronate to a mixture of UDP-D-apiose and UDP-D-xylose (Ahn *et al.*, 2006). Phosphoglycerate mutase (PGM) (Figure 3.10D, 551288) is the catalyst of step 8 in glycolysis. It catalyzes the internal transfer of a phosphate group from C-3 to C-2 which results in the conversion of 3-phosphoglycerate to 2-phosphoglycerate through a 2,3-bisphosphoglycerate intermediate. This cofactor independent PGM is present in all green plants, is monomeric and quite unstable (Perez De La Ossa et al., 1994). This enzyme was detected in the absence of mRNA in dry maize embryos, it was found that new synthesis of the protein was required to permit the progress of germination (Grana et al., 1993).

Isoflavone reductase homolog 1 (Figure 3.10E, 6573169) and chalcone isomerase (Figure 3.10F, 14582263) are involved in the biosynthesis of isoflavonoids in soybean and other legumes (Paiva et al., 1994; McGonigle, 2002; Shoji et al., 2002; Dhaubhadel et al., 2003). Their transcripts were reported to decrease sharply after late-stage of development in soybean embryos (Dhaubhadel et al., 2007), which would correspond to the trend found in the seed coat, both at the protein and transcript level.

Detoxification protein alcohol dehydrogenase (Figure 3.11A, 22597178) was steadily up-regulated, both at transcript and protein level, until the late-stage, to sharply decrease at maturity. Ascorbate peroxidase (Figure 3.11B, 4406539) on the other hand, was abundantly expressed early in development; it decreased

at mid- and late-stages to substantially increase at maturity. Stress responsive protein caffeoyl-CoA-O-methlytransferase (Figure 3.11C, 2511737) was moderately expressed throughout development. Kunitz trypsin inhibitor protein (Figure 3.11D, 13375349) was steadily up-regulated throughout the study period, it resolved in 2 spots (3225, 575). This food allergen was reported to start accumulating in soybean embryos as early as 21 DPA and to continue steadily until maturity or to decline at mid-stage (Hajduch et al., 2005).

Dormancy related protein (Figure 3.11E, 12322163) was expressed at relatively low levels with corresponding values for the transcript. The highest level was observed at late stage, after which it declined until maturity. The functional classification of this protein is unclear, but it displays high homology to chloroplast short chain oxidoreductase from *Arabidopsis thaliana* (79366418) involved in sugar metabolism.

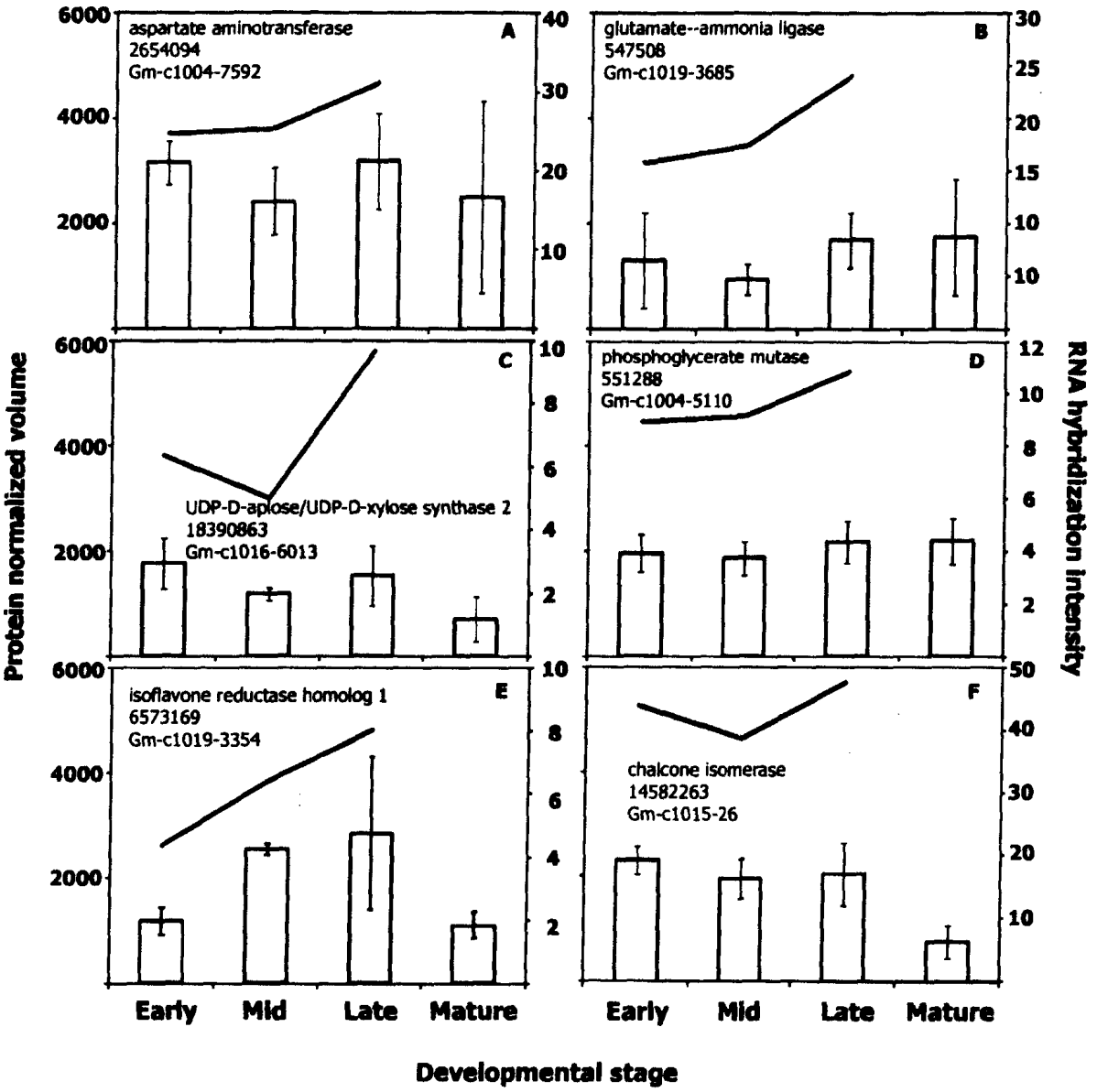


Figure 3.10 Comparison of protein profile with transcript patterns with coordinated expression during seed coat development. Panels A-B amino acid metabolism; C sugar metabolism; D energy (glycolysis); E-F secondary metabolism (isoprenylpropanoids). Bars represent protein normalized volume (with SD) and lines represent transcript hybridization intensity. Individual labels show protein name, NCBI protein accession number and systematic cDNA source clone from microarray.

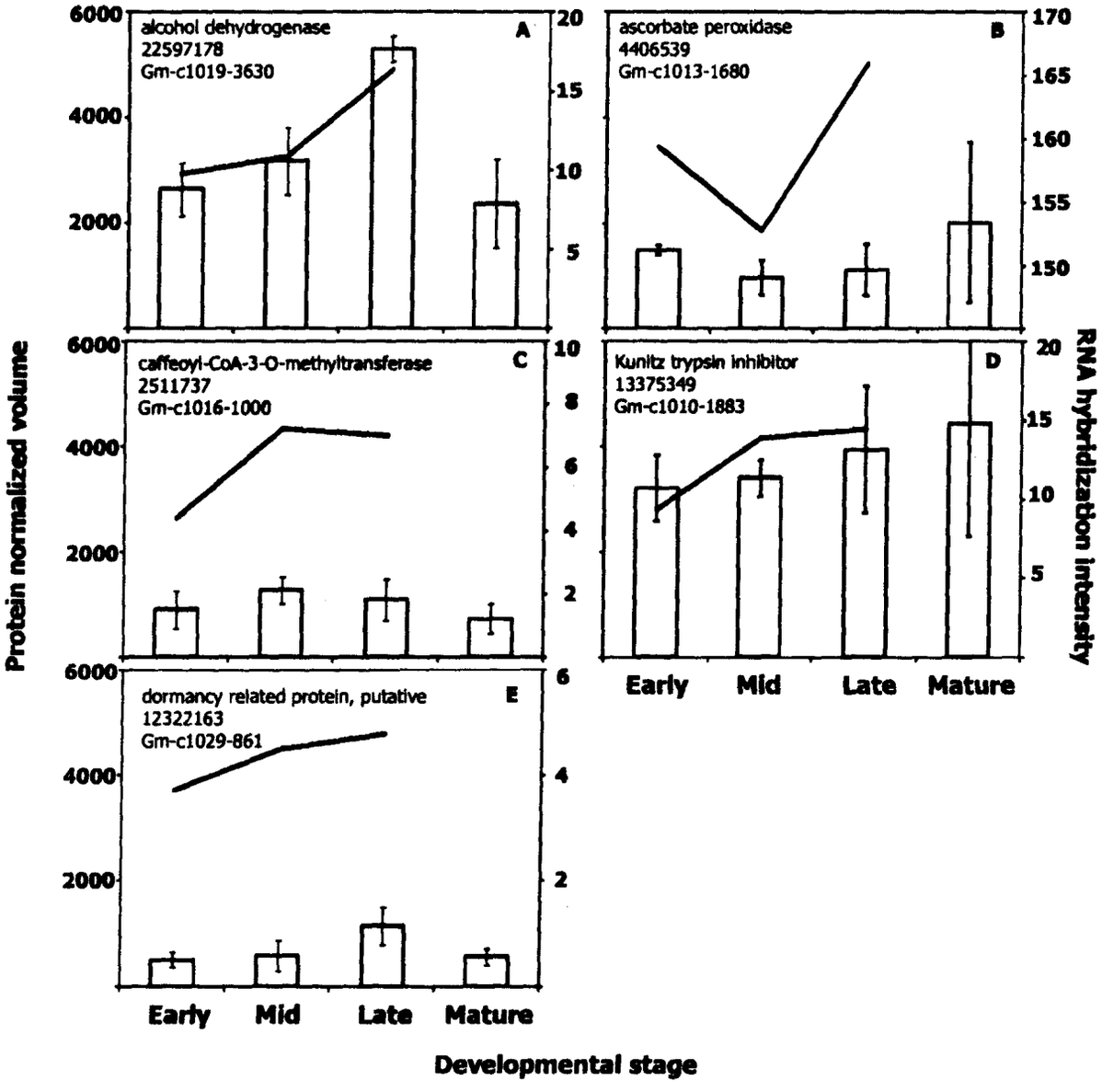


Figure 3.11 Comparison of protein profile with transcript patterns with coordinated expression during seed coat development. Panels A-B disease/defence responses (detoxification); C stress responses; D-E unclear classification. Bars represent protein normalized volume (with SD) and lines represent transcript hybridization intensity. Individual labels show protein name, NCBI protein accession number and systematic cDNA source clone from microarray.

b) Protein with apparent transcript turnover

Enolase (42521309) also known as phosphopyruvate dehydratase, was the only selected protein that displayed apparent transcript turnover, while protein levels remained constant during seed coat development until the late-stage of development. In soybean embryos, enolase (spots 557, 558; http://oilseedproteomics.missouri.edu/) levels were reported to be high early in development (14-20 DPA) and to decrease throughout development (Hajduch et al., 2005). This trend is different to what we observe in soybean seed coats, in which enolase was expressed at constant levels until the late-stage of development, when it started down-regulation (Figure 3.12).

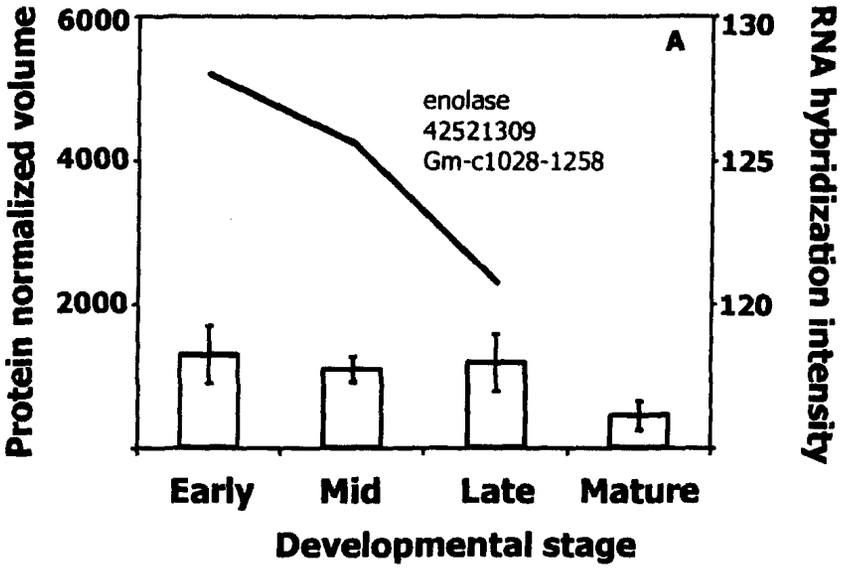


Figure 3.12 Comparison of protein profile with transcript patterns with apparent preferential transcript turnover during seed coat development. enolase: involved in glycolysis. Bars represent protein normalized volume (with SD) and lines represent transcript hybridization intensity. Individual labels show protein name, NCBI protein accession number and systematic cDNA source clone from seed coat microarray.

c) Protein and transcript levels poorly correlated

In Figures 3.13 and 3.14 we report several identified proteins whose transcript levels were poorly correlated to protein levels. Within this group we find that proteins that belong to the same functional class expressed at contrasting levels. This is the case for catalase (15236264; Figure 3.13A) expressed at relatively constant low levels and peroxidase (17467210; Figure 3.13B) expressed in an increasing fashion until late-stage and then sharply decreasing. The increasing levels of peroxidase despite the decreasing transcript trend could be due to the high stability of this protein (Kamal and Behere, 2003)

Cysteine synthase (126508784, Figure 3.13C) was expressed at increasing levels; whereas, the transcript was sharply reduced at mid-stage. Methionine synthase (33325957, Figure 3.13D) was expressed at relatively low and constant levels during seed coat development, despite the increasing trend of transcript levels, suggesting translational control of this protein's expression. Homoglutathione synthetase (7799808; Figure 3.13D) showed the highest protein levels early in development, and the trend was decreasing thereafter, whereas, the transcript levels were low early in development, reached a maximum at mid-stage to decrease at late stage. Protein and transcript levels of serine hydroxymethyltransferase (30690400) were coordinately expressed until mid-stage at very high levels; it is after mid stage that protein expression decreases and transcript increases, pointing to a developmentally regulated translational control of this protein.

Allene oxide cyclase (40644130; Figure 3.13G) and 14-3-3-like protein (4775555; Figure 3.13E) were expressed in similar trends, both at protein and transcript level. Interestingly, the transcripts were up-regulated early to sharply decrease at mid-stage, followed buy a steady up-regulation at late stages, suggesting translational control of these proteins involved in lipid metabolism and signal transduction. 14-3-3-like protein was reported in soybean embryos in several 2D spots, even following different trends along development, which led to

believe its regulatory involvement in different processes in embryo development (Hajduch et al., 2005).

The protein and transcript expression trends of cinnamyl alcohol dehydrogenase (6026516), subtilisin-type protease precursor (11611651), seed maturation protein (9622153) and universal stress protein (30693971) and beta-ketoacyl-ACP-synthetase 1-2 (7385203) were similar in that protein levels were moderately expressed and transcript levels increased as development progressed, suggesting the regulation of protein levels by translational control (Figure 3.14A, C, D, E, F).

Cytosolic phosphoglycerate kinase (15230595, Figure 3.14B) showed an increasing protein expression until late-stage, which started to decrease afterwards, which is in agreement with the decrease in metabolic activity associated with the down-regulation of energy related proteins (Gallardo *et al.*, 2003). At the transcript level, the highest level was observed at early-stage and the trend was decreasing as development progressed, suggesting high protein stability. This protein was reported as very abundant in the endosperm and not in the seed coat of *Medicago truncatula* seeds (Gallardo et al., 2007).

Isoflavone reductase homolog 2 (6573171; Figure 3.14F) showed the highest protein level at the mid stage of development, after which it was down-regulated. It was resolved in 4 2D spots (Table 3.2). At the transcript level, the trend continued to increase until late-development, suggesting high transcript stability. This was the only protein isoform reported in soybean embryos (Hajduch *et al.*, 2005) with similar expression profile (spot 912 and 920; http://oilseedproteomics.missouri.edu/). Notice that protein level of isoflavone reductase homolog 1 (6573169; Figure 3.10E) started to decrease only at maturity, showing a slight variation in the duration of the expression of the protein. The subtle control of protein expression of isoflavonoid synthesis related proteins has been previously reported (Zabala et al., 2006) and could be considered evidence of a sophisticated expression mechanism in the seed coat. However, isoflavone reductases are a large gene family with many isoforms.

There is the possibility that many of them do not participate in isoflavonoid biosynthesis and that their annotation carries errors due to conceptual translations (putative proteins).

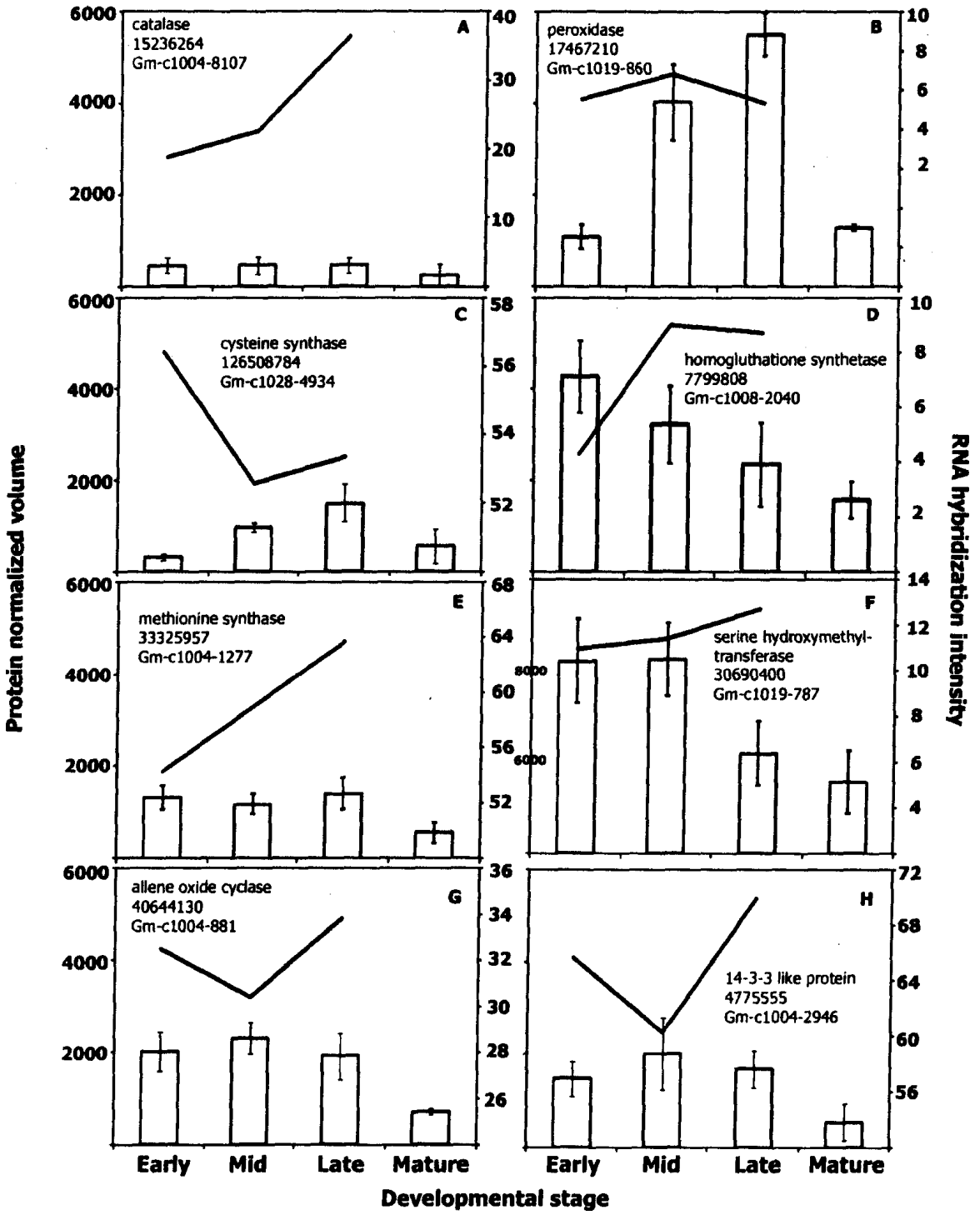


Figure 3.13 Comparison of protein profile with poorly correlated transcript and protein expression profile patterns during seed coat development. Panels A-B disease/defence; C-F amino acid metabolism; G lipid metabolism and H signal transduction. Bars represent protein normalized volume (with SD) and lines represent transcript hybridization intensity. Individual labels show protein name, NCBI protein accession number and systematic cDNA source clone from seed coat microarray.

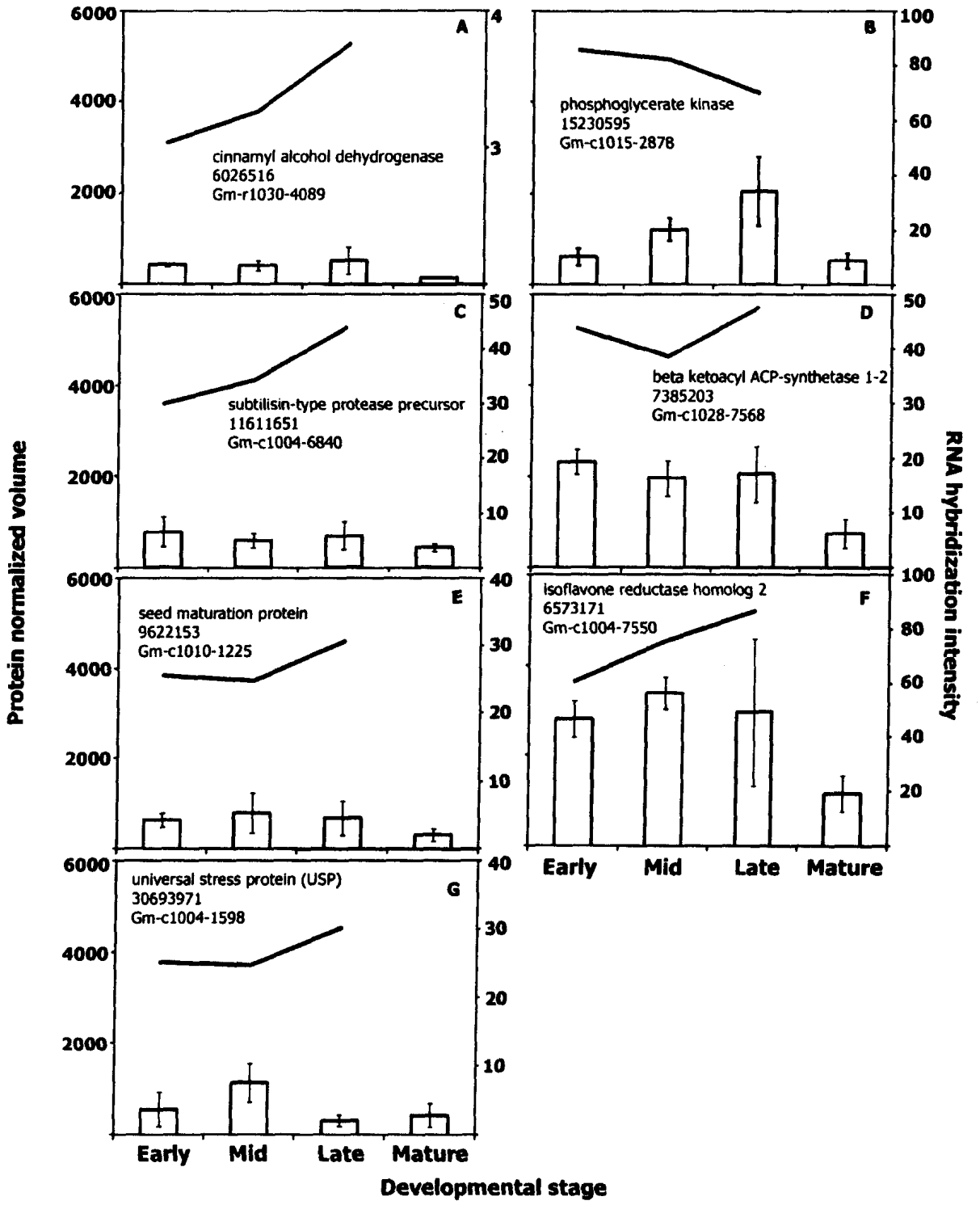


Figure 3.14 Comparison of protein profile with poorly correlated transcript and protein expression profile patterns during seed coat development. Panels A-B energy production; C-D protein destination and storage; E cell growth/seed maturation; F secondary metabolism (phenylpropanoids) and G resistance proteins. Bars represent protein normalized volume (with SD) and lines represent transcript hybridization intensity. Individual labels show protein name, NCBI protein accession number and systematic cDNA source clone from seed coat microarray.

3.4. Discussion

Soybean seed development has been divided into four major phases: a) morphogenesis and cell division; b) cell enlargement; c) seed maturation; and d) desiccation and dormancy (Mienke, 1981). By 8-10 DPA, the morphogenetic phase of early embryogenesis is completed and the most active period of cell division occurs from 8 to 15 DPA. The overall frequency of cell division in developing cotyledons declines sharply after 18 DPA. The transition from a period of morphogenesis and cell division to a period of cell enlargement occurs between 16 and 19 DPA and lasts for 6-9 days. Later stages of seed maturation are characterized by an increase in both fresh weight and dry weight. The maximum fresh weight is achieved late in development and it declines at maturity due to desiccation.

Major changes occur in the protein profiles of seed coats during development. The *k*-cluster analysis (Figure 3.2) shows how intense these changes are, which was confirmed by the Dunett's test (Table 3.1) that reported the changes between late- and mature-stages as the most noticeable. Dramatic changes between early- and mid-stages were reported in the development of *Brassica napus* (Hajduch et al., 2006), *M. truncatula* (Gallardo et al., 2003; Lei et al., 2007), sunflower (Hajduch et al., 2007) and soybeans (Hajduch et al., 2005) invariably driven by accumulation of storage proteins. Our results provide an indication that in the seed coat, the major protein changes are driven by metabolic (0.1 in Figure 3.3) and energy (0.2 in Figure 3.3) related proteins, providing a new insight into the functional role of this organ. Accumulation of protein destination and storage proteins (0.6 in Figure 3.3) is also important group that was mostly affected by proteases as opposed to storage proteins.

The use of relative volume as a means to measure protein abundance was found to be quite useful; however, it should be considered that it is an approximation based on the staining intensity, which could be affected by the presence of specific amino acids; also, the method does not measure total protein. The use of internal markers (such as specific amino acids) could be of

potential use if the addition of such addition of amino acids and the further complication of the sample could be treated in the analysis.

3.4.1. Key enzymes of Met and C_1 metabolism drive the rise in metabolic class protein expression in seed coats

The rise in expression of metabolic proteins during the experimental period (Figures 3.3-4) especially approaching maturity suggests that metabolism is highly active in the seed coat at the end of seed development. This seems contradictory with the general decrease in metabolic activity reported for seeds of *Arabidopsis* (Ruuska et al., 2002), *Brassica napus* (Hajduch et al., 2006), *M. truncatula* (Gallardo et al., 2003), *Pisum sativum* (Golombek et al., 2001) and soybean (Hajduch et al., 2005) during late stages of development. Our results indicate that proteins involved in amino acid synthesis are very abundant at later stages of development, especially methionine synthase, serine hydroxymethyltransferase and aspartate aminotransferase (Figure 3.4B). The abundance of expression of amino acid related proteins have been reported in *M. truncatula* seed coats and endosperm (Gallardo et al., 2007). However, these were reported at decreasing levels during seed filling; that is, relative abundance of amino-acid related proteins decreased as development progressed, contrary to what we observed in the soybean seed coat. This is an indication that the regulation of amino acid synthesis in the seed coat is highly species-specific, and could point to more complex regulatory pathways in which each amino acid is involved after biosynthesis. Examples may be: ethylene production for fruit ripening and C and N fluxes within the seed compartments. Overall, our data suggests that the seed coat is highly active in the synthesis of amino acids, which are transported to the embryo, where they help meet the high demand for protein synthesis. We know that there is no vascular connection between the seed coat and the embryo (Yaklich et al., 1984), which would suggest that transport events should be active and involve specific amino acid transporters.

Another possibility is that, there is no effective change in the relative amount of protein expression, but there is a change in the relative stability of specific proteins as they are affected by proteolytic processes of yellowing and desiccation taking place in the seed; which could account for an apparent increase of protein expression of metabolic proteins.

3.4.2. Glycolysis is the dominant energy-related process in soybean seed coats

Within the energy-production functional class that showed very high normalized volumes, glycolytic proteins are the most influential class in the overall rise of expression (Figure 3.5A). At least 11 glycolytic proteins were identified in the seed coat (Table 3.2) from which at least 3 are developmentally-regulated phosphoproteins in *Brassica napus* (phosphoglycerate kinase, phosphoglycerate mutase, triosephate isomerase) (Agrawal and Thelen, 2006). This is evidence of the importance of glycolysis in the seed coat and points to its potential regulatory mechanisms.

Cytosolic phosphoglycerate kinase is a specific protein whose expression was down-regulated at maturity and raised considerably the overall trend of the class. It transfers a phosphate group from 1,3-biphosphoglycerate to ADP, forming ATP and 3-phosphoglycerate. In *M. truncatula* seeds, its expression was confined to the endosperm, was down-regulated before 24 DPA and was considered evidence of commencing of metabolic quiescence. In isogenic sunflower lines, it was found that plastidic phosphoglycerate kinase was up-regulated in high oil varieties; whereas, the cytosolic isoform was the predominant one in seeds with lower oil content (Hajduch et al., 2007).

Enolase was detected at moderate levels (Figures 3.5B and 3.12) with slight decrease at maturity. Enolase catalyzes the 9th step of glycolysis, turning 2-phosphoglycerate (2-PGA) into phosphoenolpyruvate (PEP). Because PEP is the likely precursor for plastid *de novo* fatty acid synthesis during seed development, it was proposed that enolase could be important in "pulling" glycolytic and RuBisCo-generated 3-PGA into PEP and therefore, *de novo* fatty acid synthesis

in sunflower seeds (Hajduch et al., 2007). This model would presume that 3-PGA and 2-PGA are in equilibrium through phosphoglycerate mutase, which we found at consistent levels in the seed coat (Figure 3.10B; spots # 3185 and 2015 Table 3.2). The apparent differential expression of multiple enolase isoforms (spot # 1849; 445; 567; 2492) may point to a complex regulation for PEP production in the seed coat. Moreover, the transcript turn over of this protein (Figure 3.12) could be an indication of high protein stability and the notion of different isoforms expressed in a coordinated manner should not be ruled out. Further study of the regulation of glycolysis in the seed coat promises to be fruitful to understand the impact of the seed coat contribution to the seed fatty acids content.

3.4.3. Seed maturation proteins follow a similar expression profile throughout development.

Proteins involved in cell growth and division were few and expressed at relatively low levels during seed coat development (0.3 in Figure 3.3). Two different forms of transitional endoplasmatic reticulum ATPase were the only cell cycle-related proteins identified in the seed coat. This group of proteins were reported as abundant early in the seed development of *Brassica napus* (Hajduch et al., 2006) and *M. truncatula* (Gallardo et al., 2003). Proliferating cell nuclear antigen (Table 3.2) was the only DNA recombination and repair protein. This could probably be due to the fact that 2D spots were harvested from the late-stage technical replicates, a stage in which metabolic proteins were predominant in the 2D map and other groups (Appendix II), like cell cycle proteins that were present at lower abundance, were not picked for analysis.

In the case of seed maturation proteins (Figure 3.6B), desiccation protectant homolog of Lea 14 and seed maturation protein PM34, were highly expressed early in development and were slowly down regulated. These proteins are considered markers of protein maturation, since their expression in embryos peaks after 28 DPA at the commencement of maturation. In the seed coat,

however, their expression starts earlier than 20 DPA, and therefore, can not be associated with the timing of maturation.

3.4.4. Proteolysis-related proteins are the dominant class within destination and storage sub classes.

The importance of proteases in the seed coat has been highlighted in a comprehensive proteomic study of different seed compartments in *M. truncatula* (Gallardo et al., 2003; Gallardo et al., 2007). In this study, the authors associated the persistence of several proteases in the seed coat and endosperm with the supplementary source of amino acids for protein synthesis identified in the embryo. Although, this is an interesting hypothesis and very feasible, we propose that in the soybean seed coat the main role of the numerous proteases is related to tissue remodeling instead of that of amino acid recycling. This is illustrated by the dramatic changes in structure reported in the seed coat of legumes, such as the crushing of the parenchyma due to tensile tension that is so intense, that at maturity, the seed coat possesses only the remnants of parenchyma and some aleurone cells (Figure 1.1, all panels) (Miller et al., 1999), (Yaklich et al., 1990; Yaklich and Barlaszabo, 1993). At maturity the seed coat is formed of epidermis and hourglass cells that due to their sclereid nature are able to provide protection and resistance to environmental factors along with their ability to almost encapsulate their contents, preserving protein stability for longer periods of time.

The diversity of proteases found in the seed coat points to several regulatory mechanisms that are known to be driven by the proteolytic pathway. Ubiquitin binding proteins (UBC) and proteasome-related enzymes that have often vital and versatile functions (reviewed by Sakamoto, 2006) are prevalent in the seed coat. The UBC complex represents the specificity component of the ubiquitin-mediated pathway, and is therefore of especial interest, and has been studied in *Arabidopsis*. Some proteases act as a chaperone whose activity is inducible under specific conditions. Such is the case of the subtilisin-type

protease precursor in response to auxin signaling (Brechenimacher et al., 2008) or in response to light (Barnaby *et al.*, 2004). Previously it was proposed that a subtilisin-type protease (SCS1) expressed in the soybean seed could be involved in the regulation of thick-walled parenchyma cell differentiation before the tissue is crushed by the expanding embryo (Batchelor et al., 2000; Beilinson et al., 2000). Their prevalent presence in the seed coat (Table 3.2; Figure 3.7B; Figure 3.14C) suggests their involvement in such regulatory processes and could be subject of further study. Altogether, regulated proteolysis can be regarded as fine tuning at the last step of gene expression (Gottesman *et al.*, 1997; Wickner *et al.*, 1999). The ubiquitin-dependent degradation pathway through 26S proteasome provides a regulatory circuit with many developmental phenomena in plants (Hellmann and Estelle, 2002; Moon et al., 2004) and this work shows that the seed coat is no exception.

Within the protein destination and storage class, storage proteins in the seed coat accumulate at a relatively low level (Figure 3.7A and C) with glycinin (14 spots) and beta conglycinin (6 spots) of different isoforms and precursors being the main two forms, previously reported for the embryo (Hajduch et al., 2005). The moderate abundance of storage proteins in the seed coat is an indication of the relative low ability of the seed coat to store reserves.

3.4.6. Important role of detoxification proteins in the disease/defence functional class.

Disease and defence related proteins in the seed coat are of importance, as they enable the execution of the protective role of this organ. From this class, detoxification proteins are mostly represented in a steady manner from early to late stages, to decline at maturity (Figure 3.9). Soybean peroxidase has been reported for its high stability (Kamal and Behere, 2003; Welinder and Larsen, 2004), and potential use in immunoassays due to its thermal stability superior to that of horseradish peroxidase (Berlina *et al.*, 2007) and phenolic removal from water (Bassi et al., 2004; Geng et al., 2004; Mao et al., 2006; Bodalo et al., 2007; Magri et al., 2007). This becomes important considering the abundance of this

protein in the seed coat, which is around 10% of the total protein content (Gillikin and Graham, 1991) and its confined location to the hourglass cells (Gijzen, 1997), a common feature to legume seed coats. Protein expression under the control of a suitable promoter could be targeted to these cells and used for the accumulation of soybean peroxidase or other economically important proteins, using the hourglass cells and the seed coat as a bioreactor.

Cell structure, transcriptional and protein synthesis related proteins are expressed at similar levels (Figure 3.3), providing an indication of the close control of transcription and translation exerted in the seed coat. Glycine-rich RNA-binding proteins from *Arabidopsis* were found to delay seed germination during salinity stress and to accelerate it during cold-stress situations (Kim et al., 2007) and to affect stomata opening and closing during abiotic stress (Kim et al., 2008). ADP-ribosylation factors (Arf1-Arf5 protein families) have been implicated in endocytic and secretory membrane traffic and microtubule dynamics (Kahn et al., 2005) in *Arabidopsis*. Clearly, further study on the protein synthesis and transcriptional control in the seed coat could be beneficial for the applied use of seed coats as target organ for heterologous protein expression.

3.4.7. Comparison of protein and transcript trends during seed coat development

In this study, specific proteins were compared to their transcript levels, providing indication of the regulatory mechanism governing protein expression. A comprehensive comparison of the abundant proteomic data generated for the seed coat to that of the custom made seed coat microarray was in a way unsuccessful and required tedious use of tentative contigs (TC) for the comparison. Once the predicted soybean proteome becomes available for blast searches, the comparison of the full proteome and transcriptome components would be feasible, opening an exciting door to the details of protein expression and its regulation in the seed coat.

A large protein database is made available through our work for scientific investigation of the seed. The use relative volume as an indirect method to estimate protein expression allowed the semi-quantitative analysis of selected 2D spots. Undoubtedly, it will be of importance in the elucidation of protein regulation mechanics in the seed coat, important for the applied use of controlled heterologous protein expression and seed development manipulation.

3.5 References

Agrawal, G.K., and Thelen, J.J. (2006). Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. Mol Cell Proteomics **5**, 2044-2059.

Ahn, J.W., Verma, R., Kim, M., Lee, J.Y., Kim, Y.K., Bang, J.W., Reiter, W.D., and Pai, H.S. (2006). Depletion of UDP-D-apiose/UDP-D-xylose synthases results in rhamnogalacturonan-II deficiency, cell wall thickening, and cell death in higher plants. J Biol Chem **281**, 13708-13716.

Barnaby, N.G., He, F.L., Liu, X.W., Wilson, K.A., Wilson, K.A., and Tan-Wilson, A. (2004). Light-responsive subtilisin-related protease in soybean seedling leaves. Plant Physiol Bioch **42**, 125-134.

Bassi, A., Geng, Z., and Gijzen, M. (2004). Enzymatic removal of phenol and chlorophenols using soybean seed hulls. Engineering in Life Sciences **4**, 125-130.

Batchelor, A.K., Boutilier, K., Miller, S.S., Labbe, H., Bowman, L., Hu, M., Johnson, D.A., Gijzen, M., and Miki, B.L.A. (2000). The seed coat-specific expression of a subtilisin-like gene, SCS1, from soybean. Planta **211**, 484-492.

Beilinson, V., Moskalenko, O.V., Livingstone, D.S., Reverdatto, S.V., Jung, R., and Nielsen, N.C. (2000). A subtilisin-like protease from soybean seed coats. Plant Biology (Rockville) 2000, 51.

Berlina, A.N., Zherdev, A.V., Dzantiev, B.B., and Sakharov, I.Y. (2007). Use of soybean peroxidase for the enzyme immunoassay of sulfamethoxipyridazine in milk. Applied Biochemistry and Microbiology **43**, 550-555.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Bergkamp, R., Dirkse, W., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana. Nature **391**, 485-

488.

Bodalo, A., Gomez, J.L., Gomez, E., Hidalgo, A.M., Gomez, M., and Yelo, A.M. (2007). Elimination of 4-chlorophenol by soybean peroxidase and hydrogen peroxide: Kinetic model and intrinsic parameters. Biochemical Engineering Journal **34**, 242-247.

Brechenimacher, L., Kim, M.Y., Benitez, M., Li, M., Joshi, T., Calla, B., Lee, M.P., Libault, M., Vodkin, L.O., Xu, D., Lee, S.H., Clough, S.J., and Stacey, G. (2008). Transcription profiling of soybean nodulation by Bradyrhizobium japonicum. Molecular Plant-Microbe Interactions **21**, 631-645.

Dhaubhadel, S., McGarvey, B.D., Williams, R., and Gijzen, M. (2003). Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol **53**, 733-743.

Dhaubhadel, S., Gijzen, M., Moy, P., and Farhangkhoee, M. (2007). Transcriptome analysis reveals a critical role of CHS7 and CHS8 genes for isoflavonoid synthesis in soybean seeds. Plant Physiol **143**, 326-338.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., and Burstin, J. (2003). Proteomics of Medicago truncatula seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol **133**, 664-682.

Gallardo, K., Firnhaber, C., Zuber, H., Hericher, D., Belghazi, M., Henry, C., Kuster, H., and Thompson, R. (2007). A combined proteome and transcriptome analysis of developing Medicago truncatula seeds. Mol Cell Proteomics **6**, 2165-2179.

Gebhardt, J.S., Wadsworth, G.J., and Matthews, B.F. (1998). Characterization of a single soybean cDNA encoding cytosolic and glyoxysomal isozymes of aspartate aminotransferase. Plant Mol Biol **37**, 99-108.

Geng, Z.H., Bassi, A.S., and Gijzen, M. (2004). Enzymatic treatment of soils contaminated with phenol and chlorophenols using soybean seed hulls. Water Air And Soil Pollution **154**, 151-166.

Gijzen, M. (1997). A deletion mutation at the ep locus causes low seed coat peroxidase activity in soybean. Plant J **12**, 991-998.

Gillikin, J.W., and Graham, J.S. (1991). Purification And Developmental Analysis Of The Major Anionic Peroxidase From The Seed Coat Of Glycine-Max. Plant Physiol **96**, 214-220.

Golombek, S., Rolletschek, H., Wobus, U., and Weber, H. (2001). Control of storage protein accumulation during legume seed development. J Plant

Physiol **158**, 457-464.

Gottesman, S., Wickner, S., and Maurizi, M.R. (1997). Protein quality control: Triage by chaperones and proteases. Genes & Development **11**, 815-823.

Grana, X., Broceno, C., Garriga, J., Delaossa, P.P., and Climent, F. (1993). Phosphoglycerate Mutase Activity And Messenger-Rna Levels During Germination Of Maize Embryos. Plant Sci **89**, 147-151.

Gray, J.S.S., and Montgomery, R. (2006). Asymmetric glycosylation of soybean seed coat peroxidase. Carbohydrate Research **341**, 198-209.

Grusak, M.A., and Minchin, P.E.H. (1988). Seed coat unloading in Pisum sativum - osmotic effects in attached versus excised empty ovules. J Exp Bot **39**, 543-559.

Hajduch, M., Ganapathy, A., Stein, J.W., and Thelen, J.J. (2005). A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol **137**, 1397-1419.

Hajduch, M., Casteel, J.E., Hurrelmeyer, K.E., Song, Z., Agrawal, G.K., and Thelen, J.J. (2006). Proteomic analysis of seed filling in Brassica napus. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol **141**, 1159-1159.

Hajduch, M., Casteel, J.E., Tang, S.X., Hearne, L.B., Knapp, S., and Thelen, J.J. (2007). Proteomic analysis of near-isogenic sunflower varieties differing in seed oil traits. Journal of Proteome Research **6**, 3232-3241.

Haughn, G., and Chaudhury, A. (2005). Genetic analysis of seed coat development in Arabidopsis. Trends in Plant Science **10**, 472-477.

Hellmann, H., and Estelle, M. (2002). Plant development: Regulation by protein degradation. Science **297**, 793-797.

Higgins, T.J.V. (1984). Synthesis and regulation of major proteins in seeds. Briggs, W. R. (Ed.). Annual Review of Plant Physiology, Vol. 35. 736p. Annual Reviews, Inc.: Palo Alto, Calif., Usa. Illus, 191-222.

Hill, J.E.B.R.W. (1974). Proteins of soybean seeds.2. Accumulation of major protein components during seed development and maturation. Plant Physiol **53**, 747-751.

Hortensteiner, S., and Feller, U. (2002). Nitrogen metabolism and remobilization during senescence. J Exp Bot **53**, 927-937.

Kahn, R.A., Volpicelli-Daley, L., Bowzard, B., Shrivastava-Ranjan, P., Li, Y., Zhou, C., and Cunningham, L. (2005). Arf family GTPases: roles in membrane traffic and microtubule dynamics. Biochemical Society Transactions **33**, 1269-1272.

Kamal, J.K.A., and Behere, D.V. (2003). Activity, stability and conformational flexibility of seed coat soybean peroxidase. Journal of Inorganic Biochemistry **94**, 236-242.

Kim, J., Jung, H., Lee, H., Kim, K., Goh, C., Woo, Y., Oh, S., Han, Y., and H., K. (2008). Glycine-rich RNA-binding protein7 affects abiotic stress responses by regulating stomata opening and closing in Arabidopsis thaliana. Plant J, Jun 2. [Epub ahead of print].

Kim, J.Y., Park, S.J., Jang, B.S., Jung, C.H., Ahn, S.J., Goh, C.H., Cho, K., Han, O., and Kang, H.S. (2007). Functional characterization of a glycine-rich RNA-binding protein 2 in Arabidopsis thaliana under abiotic stress conditions. Plant J **50**, 439-451.

Lei, Z., Nagaraj, S., Watson, B.S., and Sumner, L.W. (2007). Proteomics of Medicago truncatula. Plant Proteomics, 121-136.

Ma, F.S., Cholewa, E., Mohamed, T., Peterson, C.A., and Gijzen, M. (2004). Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. Ann Bot **94**, 213-228.

Magri, M.L., Loustau, M.D., Miranda, M.V., and Cascone, O. (2007). Immobilisation of soybean seed coat peroxidase on its natural support for phenol removal from wastewater. Biocatalysis and Biotransformation **25**, 98-102.

Maitra, N., and Cushman, J.C. (1994). Isolation and characterization of a drought-induced soybean cDNA encoding a D95 family Late-Embryogenesis-Abundant Protein. Plant Physiol **106**, 805-806.

Mao, X., Buchanan, I.D., and Stanley, S.J. (2006). Development of an integrated enzymatic treatment system for phenolic waste streams. Environmental Technology **27**, 1401-1410.

Marsolier, M.C., Debrosses, G., and Hirel, B. (1995). Identification of several soybean cytosolic glutamine-synthetase transcripts highly or specifically expressed in nodules - expression studies using one of the corresponding genes in transgenic Lotus Corniculatus. Plant Mol Biol **27**, 1-15.

McFarlane, H.E., Young, R.E., Wasteneys, G.O., and Samuels, A.L. (2008). Cortical microtubules mark the mucilage secretion domain of the plasma membrane in Arabidopsis seed coat cells. Planta **227**, 1363-1375.

McGonigle, B. (2002). Metabolic engineering of the isoflavonoid pathway. Abstracts of papers of the American Chemical Society **224**, U91-U91.

Mienke, D.W., Chen, J., Beachy, R.N. (1981). Expression of storage-protein genes during soybean seed development. Planta **153**, 130-139.

Miller, S.S., Bowman, L.A.A., Gijzen, M., and Miki, B.L.A. (1999). Early development of the seed coat of soybean (Glycine max). Ann Bot **84**, 297-304.

Moon, J., Parry, G., and Estelle, M. (2004). The ubiquitin-proteasome pathway and plant development. Plant Cell **16**, 3181-3195.

Mullin, W.J., and Xu, W.L. (2001). Study of soybean seed coat components and their relationship to water absorption. J.Agric Food Chem **49**, 5331-5335.

Murray, D.R. (1979a). Nutritive role of the seed coats during embryo development in Pisum Sativum. Plant Physiol **64**, 763-769.

Murray, D.R. (1979b). Nutritive role of the seed coats during embryo development in Pisum sativum L. Plant Physiol **64**, 763-769.

Offler, C.E., and Patrick, J.W. (1993). Pathway of photosynthate transfer in the developing seed of Vicia faba L: a structural assessment of the role of transfer cells in unloading from the seed coat. J Exp Bot **44**, 711-724.

Paiva, N.L., Sun, Y.J., Dixon, R.A., Vanetten, H.D., and Hrazdina, G. (1994). Molecular cloning of isoflavone reductase from pea (Pisum-Sativum L) - Evidence for a 3r-isoflavanone intermediate in pisatin biosynthesis. Arch Biochem Biophys **312**, 501-510.

Perez De La Ossa, P., Grana, X., Ruiz-Lozano, P., and Climent, F. (1994). Isolation and characterization of cofactor-independent phosphoglycerate mutase gene from maize. Biochemical and Biophysical Research Communications **203**, 1204-1209.

Qutob, D., Ma, F., Peterson, C.A., Bernards, M.A., and Gijzen, M. (2008). Structural and permeability properties off the soybean seed coat. Botany-Botanique **86**, 219-227.

Rautengarten, C., Usadel, B., Neurnetzler, L., Hartmann, J., Buessis, D., and Altmann, T. (2008). A subtilisin-like serine protease essential for mucilage release from Arabidopsis seed coats. Plant J **54**, 466-480.

Rolletschek, H., Hosein, F., Miranda, M., Heim, U., Gotz, K.P., Schlereth, A., Borisjuk, L., Saalbach, I., Wobus, U., and Weber, H. (2005). Ectopic expression of an amino acid transporter (VfAAP1) in seeds of Vicia narbonensis and pea increases storage proteins. Plant Physiol **137**, 1236-

1249.

Ruuska, S.A., Girke, T., Benning, C., and Ohlrogge, J.B. (2002). Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell **14**, 1191-1206.

Sakamoto, W. (2006). Protein degradation machineries in plastids. Annu. Rev. Plant Biol **57**, 599-621.

Schuurmans, J., van Dongen, J.T., Rutjens, B.P.W., Boonman, A., Pieterse, C.M.J., and Borstlap, A.C. (2003). Members of the aquaporin family in the developing pea seed coat include representatives of the PIP, TIP, and NIP subfamilies. Plant Mol.Biol **53**, 655-667.

Shackel, K.A., and Turner, N.C. (2000). Seed coat cell turgor in chickpea is independent of changes in plant and pod water potential. J Exp Bot **51**, 895-900.

Shao, S.Q., Meyer, C.J., Ma, F.S., Peterson, C.A., and Bernards, M.A. (2007). The outermost cuticle of soybean seeds: chemical composition and function during imbibition. J Exp Bot **58**, 1071-1082.

Shoji, T., Winz, R., Iwase, T., Nakajima, K., Yamada, Y., and Hashimoto, T. (2002). Expression patterns of two tobacco isoflavone reductase-like genes and their possible roles in secondary metabolism in tobacco. Plant Mol.Biol **50**, 427-440.

Thelen, J.J., and Ohlrogge, J.B. (2002). Metabolic engineering of fatty acid biosynthesis in plants. Metab Eng **4**, 12-21.

Thorne, J.H. (1981). Morphology and ultrastructure of maternal seed tissues of soybean in relation to the import of photosynthate. Plant Physiol **67**, 1016-1025.

Truernit, E., and Haseloff, J. (2008). Arabidopsis thaliana outer ovule integument morphogenesis: Ectopic expression of KNAT1 reveals a compensation mechanism. Bmc Plant Biology [Epub ahead of print]

Van Dongen, J.T., Ammerlaan, A.M.H., Wouterlood, M., Van Aelst, A.C., and Borstlap, A.C. (2003). Structure of the developing pea seed coat and the post-phloem transport pathway of nutrients. Ann Bot **91**, 729-737.

Walker, N.A., Patrick, J.W., Zhang, W.H., and Fieuw, S. (1995). Efflux of photosynthate and acid from developing seed coats of Phaseolus vulgaris L.: a chemiosmotic analysis of pump-driven efflux. J Exp Bot **46**, 539-549.

Wang, X.D., Harrington, G., Patrick, J.W., Offler, C.E., and Fieuw, S. (1995). Cellular pathway of photosynthate transport in coats of developing seed of

Vicia faba L. and Phaseolus vulgaris L. II. Principal cellular site(s) of efflux. J Exp Bot **46**, 49-63.

Weber, H., Borisjuk, L., Heim, U., Buchner, P., and Wobus, U. (1995). Seed coat-associated invertases of fava bean control both unloading and storage functions: cloning of cDNAs and cell type-specific expression. Plant Cell **7**, 1835-1846.

Welinder, K.G., and Larsen, Y.B. (2004). Covalent structure of soybean seed coat peroxidase. Biochimica Et Biophysica Acta-Proteins And Proteomics **1698**, 121-126.

Wickner, S., Maurizi, M.R., and Gottesman, S. (1999). Posttranslational quality control: Folding, refolding, and degrading proteins. Science **286**, 1888-1893.

Yaklich, R., Wergin, W., and Vigil, E. (1990). Freeze Fracture Study Of Soybean Glycine-Max L. Merr. Seed Coat Cells. Physiologia Plantarum **79**, A17.

Yaklich, R.W., and Barlaszabo, G. (1993). Seed coat cracking in soybean. Crop Science **33**, 1016-1019.

Yaklich, R.W., Vigil, E.L., and Wergin, W.P. (1984). Scanning Electron-Microscopy of soybean seed coat. Scanning Electron Microscopy 991-1000.

Yaklich, R.W., Vigil, E.L., Erbe, E.F., and Wergin, W.P. (1992). The fine structure of aleurone cells in the soybean seed coat. Protoplasma **167**, 108-119.

Zabala, G., Zou, J., Tuteja, J., Gonzalez, D., Clough, S., and Vodkin, L. (2006). Transcriptome changes in the phenylpropanoid pathway of Glycine max in response to Pseudomonas syringae infection. BMC Plant Biology **6**.

Zhou, Y.C., Qu, H.X., Dibley, K.E., Offler, C.E., and Patrick, J.W. (2007). A suite of sucrose transporters expressed in coats of developing legume seeds includes novel pH-independent facilitators. Plant J **49**, 750-764.

CHAPTER 4

DISCUSSION

In the studies described in the preceding chapters, it has been demonstrated that the seed coat proteome is very complex and possesses numerous proteins. This is evidence of intricate mechanisms taking place in this organ, and led to an exploration of the general cellular pathways represented by functional classes. The association of the seed coat proteins with such pathways led to the establishment of detailed enzymatic pathways that represent the mechanisms of biosynthesis of cell walls, lipids, isoflavonoids and C_1 metabolism pathways in the seed coat. Although, we used well established pathways, the association of specific seed coat proteins with such important biosynthetic pathways is novel, confirming at the protein level the major functions of this tissue.

The developmental study that followed, allowed the identification and relative quantification of proteins that are differentially expressed during seed development. As the seed matures, the trend of metabolic proteins increases in the seed coat, leading to the conclusion that although metabolic quiescence is generally the situation in the embryo, the seed coat remains active to later stages of physiological maturity. It was also demonstrated that energy-related proteins along with the detoxification proteins, are the most abundant proteins expressed at physiological maturity in seed coats.

4.1 Protein identification

Large-scale identification of seed coat proteins is reported in Chapter 2. A combination of methods was utilized to extract, pre-fractionate and identify seed coat proteins, using TCA precipitation, SDS-PAGE and 2-SDS-PAGE, followed by iterative exclusion lists and tandem mass spectrometry. The combination of these methods allowed the identification of over a thousand proteins, which surpasses the number of proteins reported for whole soybean seeds. The use of iterative exclusion lists allowed the identification of lower

abundance proteins. These are normally not detected by conventional spectral processing due to the high abundance of house-keeping proteins, such as storage and structural proteins. The combination of these methods promises to be fruitful in the study of single cell proteomes, such as epidermal and hourglass cells, which are highly differentiated and specialized cell types (Yaklich et al., 1986).

4.2 Gene ontology assignment

The gene ontology assignment of the identified proteins is relevant in that it allows proteins to be grouped according to their cellular function, involvement in major pathways or at least their sub cellular localization. Due to soybean's economic importance, many efforts have been made to develop genetic and genomic resources for soybean, including genetic linkage maps and EST collections. Furthermore, sequencing of the entire soybean genome was completed and made available to the scientific community on January 18, 2008 (http://phytozome.net/soybean). However, the information is still preliminary and is not yet suitable for protein searches yet. Once the predicted proteome becomes available, the protein data that we report could be probed against the predicted proteome first to confirm our report, but more importantly, to identify more proteins within the ~130,000 spectral profiles that did not find any match in the NCBI nr protein database.

Gene ontology assignment was done manually, given the current lack of automated gene ontology assignment tools for other plants other than *Arabidopsis thaliana* and *Oryza sativa*. Genes associated with the *Rhizobium* colonization capacity are believed to account for 30% of the total genome of legume plants. That implies that if one was to use a non-legume protein database to search for legume proteins, there is a 30% less chance to find it before starting the search. Also, plant genomes are quite different. The 125-Mbp *Arabidopsis* genome is one of the smallest known among higher plants (Arumuganathan and Earle, 1991); whereas, the soybean haploid genome contains 1,115 Mbp. This almost 8-fold difference in genome size could be due

to ancient polyploidization events during the evolution of the family and the high level of repetitive sequences in the soybean genome (Grant *et al.*, 2000). Another factor is that there is considerable difference even between the genomes of model legume species such as *Medicago truncatula* and soybean (Choi et al., 2004) making comparisons even within the same family not robust. One example demonstrated here is the difference in expression of metabolic proteins that we reported in Chapter 3 between soybean and *M. truncatula*, this suggests different C_1 metabolism trends, which could be quite significant in terms of seed physiology and might be due to the intensive selection for agronomic traits in soybean as opposed to lower selective pressure exerted on *M. truncatula.*

Protein identification in tandem mass spectrometry is carried out by virtue of database searches (Sadygov et al., 2004). One should not forget that the current state of plant protein databases in general is still under developed (Yates et al., 2004). Although major efforts are currently being undertaken to resolve this issue, a large proportion of plant protein databases rely on the theoretical translation of nucleotide sequences, which only give putative proteins. Undoubtedly, seed research will benefit from the further development of protein databases and predicted proteomes as these resources become available.

4.3 Important cellular pathways confirmed in the seed coat

The identification of most of the enzymes involved in cell wall biosynthesis in the seed coat proteome is evidence of the importance of this pathway in this organ. Hourglass and palisade layers are formed by sclereid type cells much differentiated plant cells with thickened cell walls. The hourglass cells are the most prominent anatomical feature in mature soybean seed coats. The thickened cell walls provide structural support for the seed and allow them to withstand the tensile pressure of the growing embryo. Their rigidity limits seed size and results in crushing of some of the inner seed coat layers as the embryo grows. Thickened cell walls are also observed in the vascular parenchyma and

aerenchyma, where they may enhance the apoplastic transport of nutrients to the embryo during seed filling. The presence of extensins and proline-rich proteins provides further evidence for the cross-link to the extracellular matrix and solidification of cell walls.

Several cell-wall related proteins are reported that point to processes different than cell wall biosynthesis, such as cell wall invertases and their known role in the creation of sink strength. Two examples are: rhamnose synthase in pectin production and UDP-glucose epimerase in cell wall integrity. Serine proteases potentially involved in parenchyma tissue remodeling and cell wall loosening proteins are also reported. Other cell wall proteins, such as proteases, polysaccharide hydrolytic enzymes, and lipases were reported to contribute to the generation of defense signals and response to the environment and many still unknown proteins may fall in this category. Our results establish a baseline for further scientific investigation and discovery of key players in the mechanism of seed coat response to the environment. The spectral data should be further analyzed against the soybean cell wall proteome once the resources become available.

Our results demonstrate that at the protein level the seed coats are capable of *de novo* lipid synthesis. It is also noteworthy that seed coats seem to have the capability to synthesize tocopherols, from a branching of the phenylpropanoid pathway and FA synthesis. In *Arabidopsis* seeds tocopherols inhibit the oxidation of polyunsaturated fatty acids during dormancy and germination, increasing germination fitness (Sattler *et al.*, 2004), which would be fitting with the postulated function of the seed coat in germination enhancement. A metabolic engineering approach taken to increase the vitamin E content in soybean utilizing the over expression of 2-methyl-6-phytylbenzoquinol methyltransferase (VTE3) from *Arabidopsis*, the levels of α-tocopherol, which is the active vitamin E, increased its expression by 7-fold (Van Eenennaam et al., 2003). This demonstrates that data obtained from engineering tocopherols synthesis in a model system can be readily transferred to crop plants, marking the beginning of exciting times in which plant metabolic engineering can be used

to have a positive impact on human nutrition and health. In that sense, our results provide the protein information necessary for the evaluation and implementation of biotechnological efforts to modify seed coat and seed lipid synthesis.

Isoflavonoids have been reported to accumulate mostly in developing seeds and leaves, and in some extent in the seed coats (Dhaubhadel et al., 2003). It is known that soybean embryos have the capability to synthesize isoflavonoids *de novo* from simple precursors and it was proposed that the isoflavonoids from the seed coat are transported to the embryo, helping to increase the total amount of these metabolites in the seed. It was also noted that the inheritance of isoflavonoids in soybean seeds presents a maternal effect; that is, it is transmitted from plant to progeny in the maternal integuments, from which the seed coat arises. Our results confirm the *de novo* synthesis of isoflavonoid in maternal tissues and support the notion that seed coat isoflavonoid biosynthesis contributes to the overall content of the soybean seed. We also demonstrated the presence of several types of transporters that have been previously reported in the transport of isoflavonoids. The detailed complement of biosynthetic enzymes and potential transporters provide a baseline for the closer inspection of the specific mechanisms of isoflavonoid accumulation and transport in the seed coats and will certainly be helpful in any metabolic engineering of the isoflavonoid contents of soybean.

Proteolysis is one of the main processes taking place in soybean seed coats. Its presence is undoubtedly related to tissue remodeling, which is very dramatic during seed development, and even at maturity. At maturity the seed coat is formed of epidermis and hourglass cells that due to their sclereid nature are able to provide protection and resistance to the seed coat, besides their ability to almost encapsulate their contents, preserving protein stability for longer periods of time. There is also evidence to support the notion of nitrogen remobilization and protein degradation by subtilisin-type proteases, Clp-proteases, 20S proteasome. Also, the seed coat proteome is equipped with at least 54 cysteine proteases, half of the protease component, which have been

reported in chloroplast degradation or yellowing that takes place at physiological maturity. Glycoside hydrolases are involved in the hydrolysis of cell wall polyssacharides and signaling (Minic *et al.*, 2007); their presence is an indication of the cell wall degradation and remodeling that takes place in the seed coat. There is a wealth of proteases in the soybean seed coat that could be related to the regulation of several processes such as chloroplast biogenesis and local systemic defense responses. Our data provides a detailed prospecting of the proteolytic complement in soybean seed coats and will be helpful in future biotechnological efforts to modify the protein composition of seed coats. Within this same functional class, the moderate abundance of storage proteins in the seed coat is an indication of the relative low ability of the seed coat to store reserves.

At physiological maturity, the most abundant protein in the seed coat is methionine synthase. Together with *S*-adenosylmethionine synthase (AdoMet), these two enzymes were previously associated with the status of metabolic activity in seeds (Gallardo et al., 2003; Rajjou et al., 2004). It is noteworthy to find these metabolic enzymes as the most abundant in the seed coat, as an indication of the importance of this process. Their decreased levels could be considered an indication of the switch from active metabolism to a quiescent state. In our study, the high abundance of this protein indicates that seed coats are metabolically active for a prolonged period in comparison to the embryo. We propose that the high levels of methionine synthase in the seed coat could also be related to its participation in the production of the fruit ripening hormone ethylene. Several other enzymes found in the seed coat proteome, such as *S*-adenosyl-L methionine (SAM), aminocyclopropane 1-carboxylic acid (ACC) oxidase (ACO) and various forms of ACC synthases (late embryogenesis and maturation proteins), support the idea of active production of ethylene in the seed coat.

The finding that proteins involved in amino acid synthesis that are very abundant at later stages of development, especially methionine synthase, serine

hydroxymethyltransferase and aspartate aminotransferase is contradictory to the declining trend reported for metabolic proteins (Gallardo et al., 2007). The difference in the expression patterns of these proteins between the soybean seed coat and *M. truncatula* seed coats and endosperm is substantial. This is an indication that the regulation of amino acid synthesis in the seed coat is highly species-specific, and could indicate more complex regulatory pathways in which each amino acid is involved after biosynthesis, such as ethylene production for fruit ripening and C and N fluxes within the seed compartments. Overall, our data suggests that the seed coat is highly active in amino acid synthesis. The amino acids made in the seed coat may be transported to the embryo, where they help meet the high demand for protein synthesis. We know that there is no vascular connection between the seed coat and the embryo (Yaklich et al., 1984), that would infer that transport events are expected to be active and involving specific amino acid transporters, presenting a very interesting opportunity for the study of the manipulation of amino acid composition in the seeds.

Energy-production related proteins are important in the seed coat, as they are probably in every other plant organ. In the seed coat, glycolysis takes a noteworthy place, with the enzymes expressed at high relative volumes. Some glycolytic proteins are developmentally regulated phosphoproteins such as phosphoglycerate kinase, phosphoglycerate mutase and triosephate isomerase. Further study of the regulation of glycolysis in the seed coat promises to help understand the impact of the seed coat's contribution to the fatty acid content of the seed.

Disease and defence related proteins in the seed coat are of importance, as they enable the execution of the protective role of this organ. From this class, detoxification proteins are mostly represented in a steady manner from early to late stages and decline at maturity. Soybean peroxidase has been investigated for its high stability, potential use in immunoassays due to its thermal stability superior to that of horseradish peroxidase, and in the removal of phenolic waste from water. These uses become important considering the abundance of this protein in the seed coat, which is around 10% of the total protein content (Gillikin

and Graham, 1991) and its confined location to the hourglass cells (Gijzen, 1997), common feature to legume seed coats. Protein expression under the control of a suitable promoter could be targeted to these cells and used for the accumulation of soybean peroxidase or other economically important proteins, using the hourglass cells and the seed coat as a bioreactor.

In this study, specific proteins were compared to their transcript levels, providing an indication of the regulatory mechanism governing protein expression, as in the case of transcript or protein turnover during seed coat development. A large protein database is made available through our work for scientific investigation of the seed. Undoubtedly, it will be of importance in the elucidation of protein regulation mechanism in the seed coat.

4.4 Summary

The results presented here provide useful resources that can be exploited in future studies of the seed coat and its protein composition modification during development. Also, it is also an important contribution to the general understanding of seed development and the involvement of the seed coat in this process. These resources include a comprehensive list of over a thousand proteins classified into functional classes, representing a 5-fold increase in the number of proteins identified in whole soybean seeds. Over 300 hundred of these proteins were followed up during seed coat development, and there is information available on the expression levels at each stage with transcript level comparisons for some of them. Also, the allocation of several enzymes to important cellular biosynthetic pathways, such as C_1 metabolism, lipid synthesis and proteolysis, gives very detailed information on the mechanics and particularities of their regulation in the soybean seed coat.

4.5 References

Arumuganathan, K., and Earle, E.D. (1991). Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter **9,** 208-218.

Choi, H.K., Mun, J.H., Kim, D.J., Zhu, H.Y., Baek, J.M., Mudge, J., Roe, B., Ellis, N., Doyle, J., Kiss, G.B., Young, N.D., and Cook, D.R. (2004). Estimating genome conservation between crop and model legume species. Proceedings of the National Academy of Sciences of the United States of America **101,** 15289-15294.

Dhaubhadel, S., McGarvey, B.D., Williams, R., and Gijzen, M. (2003). Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol **53,** 733-743.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., and Burstin, J. (2003). Proteomics of Medicago truncatula seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol **133,** 664-682.

Gallardo, K., Firnhaber, C., Zuber, H., Hericher, D., Belghazi, M., Henry, C., Kuster, H., and Thompson, R. (2007). A combined proteome and transcriptome analysis of developing Medicago truncatula seeds. Mol Cell Proteomics **6,** 2165-2179.

Gijzen, M. (1997). A deletion mutation at the ep locus causes low seed coat peroxidase activity in soybean. Plant J **12,** 991-998.

Gillikin, J.W., and Graham, J.S. (1991). Purification and developmental analysis of the major anionic peroxidase from the seed coat of Glycine-Max. Plant Physiol **96,** 214-220.

Grant, D., Cregan, P., and Shoemaker, R.C. (2000). Genome organization in dicots: Genome duplication in Arabidopsis and synteny between soybean and Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America **97,** 4168-4173.

Le, B.H., Wagmaister, J.A., Kawashima, T., Bui, A.Q., Harada, J.J., and Goldberg, R.B. (2007). Using genomics to study legume seed development. Plant Physiol **144,** 562-574.

Lei, Z., Nagaraj, S., Watson, B.S., and Sumner, L.W. (2007). Proteomics of Medicago truncatula. Plant Proteomics, 121-136.

Lopes, M.A., and Larkins, B.A. (1993). Endosperm origin, development and function. Plant Cell **5,** 1383-1309.

Minic, Z., Jamet, E., Negroni, L., der Garabedian, P.A., Zivy, M., and Jouanin, L. (2007). A sub-proteome of Arabidopsis thaliana mature stems trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. J Exp Bot **58,** 2503-2512.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of alpha-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. Plant Physiol **134,** 1598-1613.

Sadygov, R.S., Cociorva, D., and Yates, J.R. (2004). Large-scale database searching using tandem-mass spectra: Looking up the answer in the back of the book. Nature Methods **1,** 195-202.

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M., and DellaPenna, D. (2004). Vitamin E is essential for seed longevity, and for preventing lipid peroxidation during germination. Plant Cell **16,** 1419-1432.

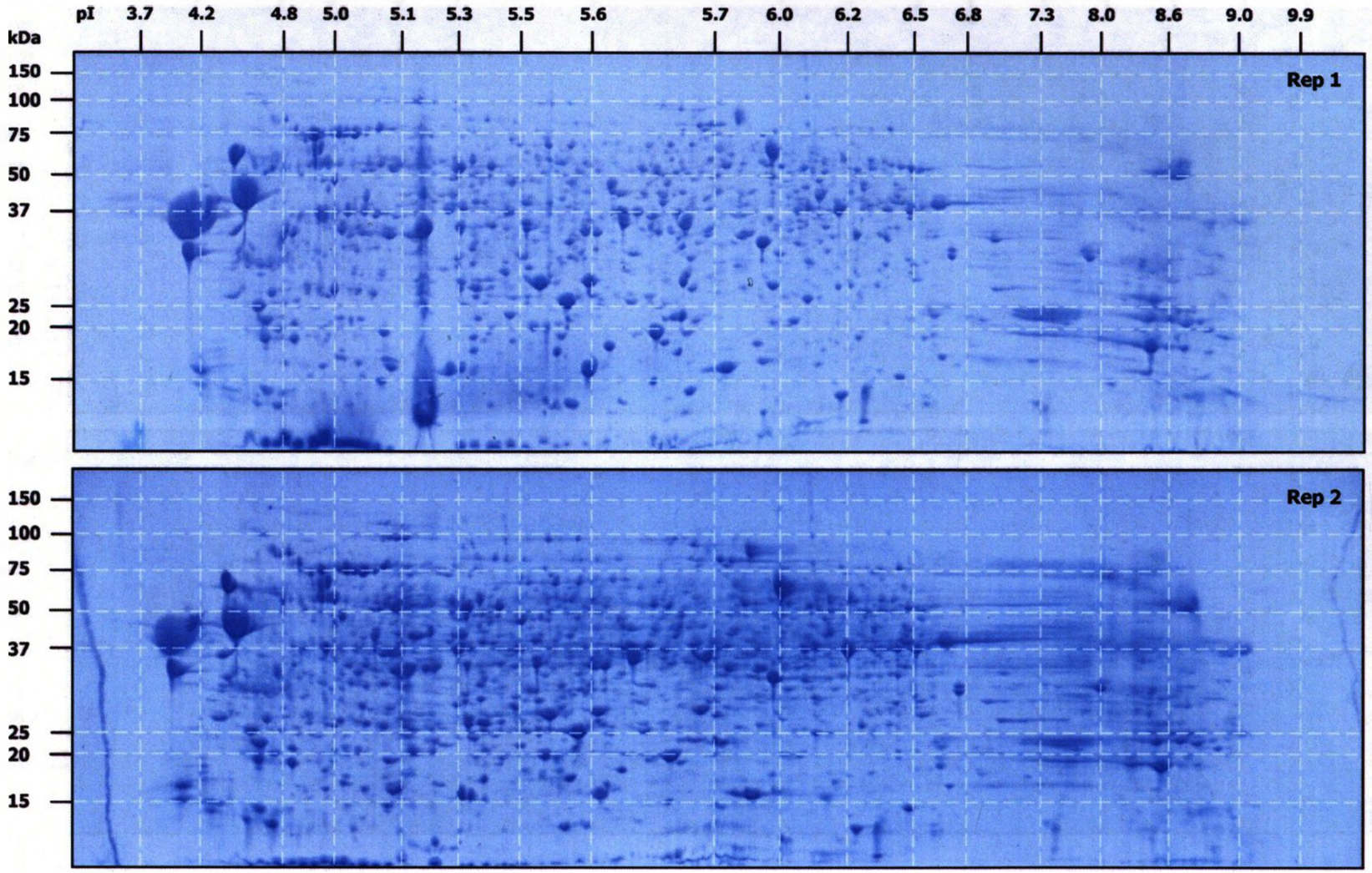
Van Eenennaam, A.L., Lincoln, K., Durrett, T.P., Valentin, H.E., Shewmaker, C.K., Thorne, G.M., Jiang, J., Baszis, S.R., Levering, C.K., Aasen, E.D., Hao, M., Stein, J.C., Norris, S.R., and Last, R.L. (2003). Engineering vitamin E content: From Arabidopsis mutant to soy oil. Plant Cell **15,** 3007-3019.

Yaklich, R.W., Vigil, E.L., and Wergin, W.P. (1984). Scanning electron-microscopy of soybean seed coat. Scanning Electron Microscopy, 991-1000.

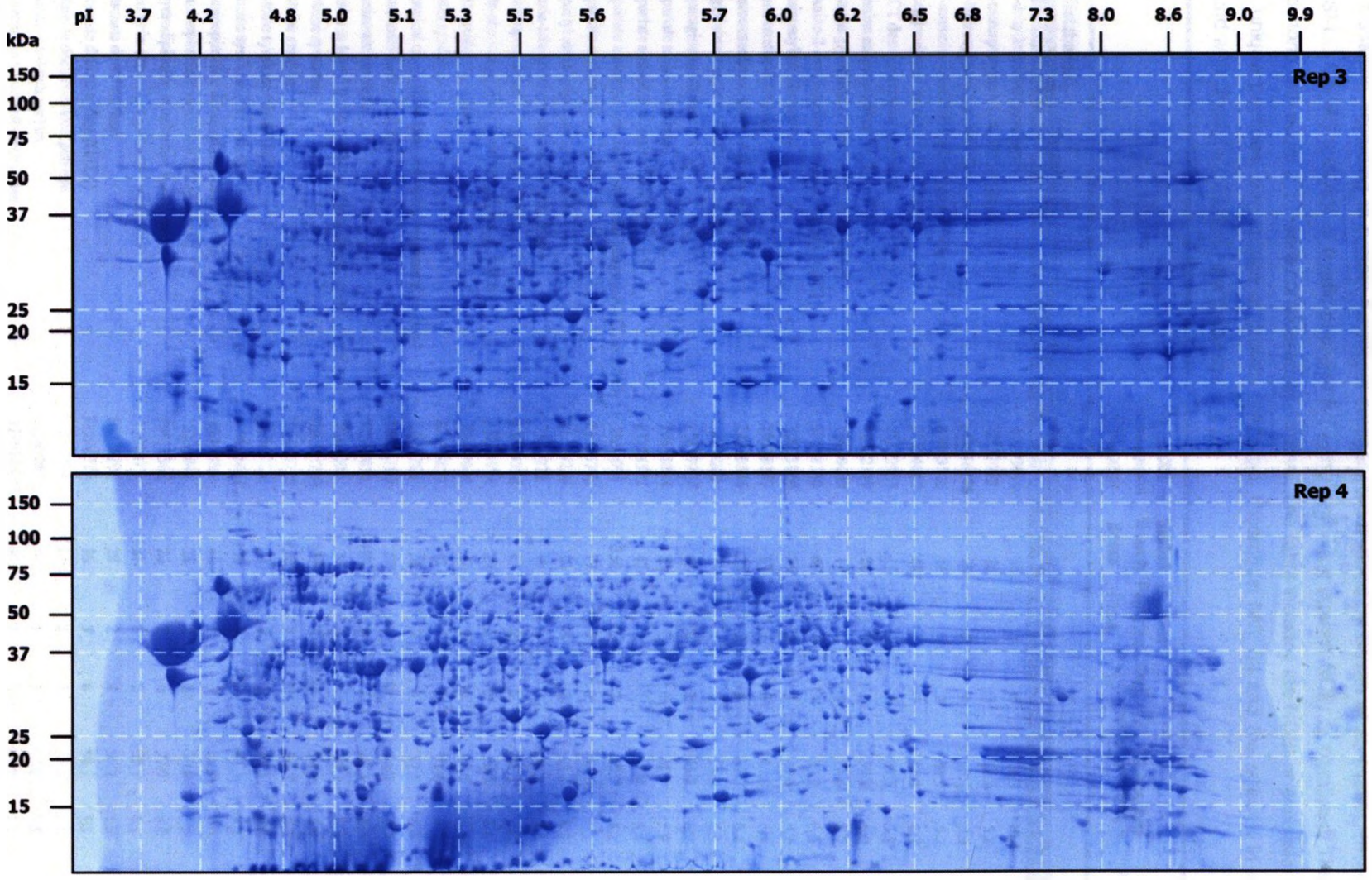
Yaklich, R.W., Wergin, W.P., and Vigil, E.L. (1986). Freeze fracture of secretory cells in soybean seed coat. Journal of Cell Biology **103,** 526A.

Yates, J.R., MacCoss, M., MacDonald, H., Wu, C., Sadygov, R., and Anderson, S. (2004). Plenary paper: High throughput proteomic analysis. Abstracts of Papers of the American Chemical Society **227,** U204-U204.

Appendix I. 2D-SDS-PAGE pre-fractionation of seed coat proteins. Four technical replicates of 500 µg of protein extracted from 35-50 DPA soybean seed coats. Images were analyzed using Progenesis PG220 and spots were excised from replicate 3.



Appendix I. *(continued from previous page).*



Appendix II. Proteins identified by LC-MS/MS from 35-50 DPA soybean seed coats 1D and 2D SDS-PAGE gels.

Proteins were classified according to protein functional categories described by Bevan et al. (1998). Proteins were identified by ESI-LC-MS/MS analysis of tryptic peptides following searching against NCBI nr database. The putative protein identifications with score ≥ 13 were considered as positive. The table includes the NCBI nr accession number, MS/MS score, unique peptides, percentage of protein coverage, theoretical MW /pI and the species in which the protein was found with closest similarity in the database.

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	01 Metabolism							
1	3-dehydroquinate synthase	57283679	14	1	9	16.3	10.4	*Hv*
2	3-mercaptopyruvate sulfurtransferase	7581972	16	1	2	40.2	5.8	*At*
3	4-hydroxyphenylpyruvate dioxygenase	148616206	28	2	4	48.4	5.6	*Gm*
4	5'-aminoimidazole ribonucleotide synthetase	37983566	16	1	7	42.9	5.2	*St*
5	acetylornithine transaminase, putative	21554043	48	3	8	48.8	6.3	*At*
6	ACT domain-containing protein	21592963	15	1	2	33.3	5.3	*At*
7	alanine aminotransferase (ALAAT1)	42562119	24	2	2	59.8	6.0	*At*
8	alanine:glyoxalate aminotransferase 2 (AGT2)	18420498	13	1	1	52.0	7.7	*At*
9	alanine-2-oxoglutarate aminotransferase 1 (OGT1)	30688330	69	4	11	53.3	6.5	*At*
10	aminoacylase like protein	30693849	25	2	5	49.0	5.7	*At*
11	aminomethyltransferase	79470337	13	1	2	43.5	6.3	*At*
12	aminotransferase class IV family protein	22330856	26	2	6	62.2	6.4	*At*
13	argininosuccinate synthase-like protein	4678262	13	1	2	54.9	6.3	*At*
14	asparagine synthetase	77819909	15	2	2	65.2	6.2	*Gm*
15	aspartate aminotransferase	29468084	20	2	4	45.9	5.9	*Os*
16	aspartate aminotransferase 1 (ASP1)	15224592	42	4	11	47.8	8.4	*At*
17	aspartate aminotransferase glyoxysomal isozyme AAT1 precursor	2654094	250	17	42	49.7	8.7	*Gm*
18	aspartate carbamoyltransferase	21535795	17	2	3	42.6	6.1	*St*
19	aspartyl aminopeptidase-like protein	21537290	14	1	2	52.5	6.3	*At*
20	ataxia-telangiectasia mutated protein (Atm)	7529272	19	2	0	435.1	7.0	*At*
21	auxin upregulated 1, yadokari 1 (YDK1)	15235538	13	2	3	68.2	5.9	*At*
22	beta-alanine-pyruvate aminotransferase, putative	15231974	29	2	5	52.5	8.1	*At*
23	beta-ureidopropionase	30698009	27	2	6	45.6	5.9	*At*
24	biotin synthase	82393851	13	1	2	41.6	6.8	*Gm*
25	branched chain alpha-keto acid dehydrogenase E2 subunit	7021284	14	2	3	52.8	6.4	*At*
26	chorismate mutase 1 (CM1)	18406100	17	2	12	38.2	5.7	*At*
27	chorismate synthase	77547031	106	8	17	47.3	6.3	*Gm*
28	similar to leucyl tRNA synthetase from Homo sapiens gb|D84223	8569090	34	2	4	96.3	6.8	*At*
29	cysteine synthase	126508784	127	8	35	34.7	5.3	*Gm*
30	cysteine synthase	126508778	120	8	32	34.5	5.5	*Gm*
31	cysteine synthase (OAS-TL3)	148562451	21	2	5	40.1	6.6	*Gm*
32	cysteine synthase (OAS-TL4)	148562457	139	9	32	41.3	8.1	*Gm*
33	diaminopimelate decarboxylase-like protein	9279586	15	1	3	58.5	7.6	*At*
34	diaminopimelate epimerase family protein	15231841	26	2	8	39.0	5.5	*At*
35	dihydrodipicolinate reductase family protein	18411368	26	2	6	37.9	6.8	*At*
36	ferredoxin-dependent glutamate synthase	1702872	21	3	2	180.1	6.1	*At*
37	glutamate decarboxylase	16226294	28	2	4	56.2	5.3	*At*
38	glutamate decarboxylase (GAD1)	31296711	36	3	6	54.3	5.3	*Hv*
39	glutamate decarboxylase, putative	21536919	29	2	3	55.8	5.4	*At*
40	glutamic acid-rich protein	15240907	13	1	1	72.0	5.8	*At*
41	glutamine amidotransferase/cyclase	3219164	40	3	5	64.7	6.5	*At*
42	glutamine synthetase	286124	22	2	6	39.3	5.3	*Zm*

(Table continues on following page)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
43	glutamine synthetase isoform GSe1	40317420	25	3	17	39.5	5.5	*Ta*
44	glutamine synthetase precursor	13877511	63	4	12	47.7	6.7	*Gm*
45	homologue to thiamine pyrophosphate (T25K16.8)	6715645	17	1	1	75.3	8.9	*At*
46	homologue to tyrosine/serine/threonine phosphatase	16612246	27	3	6	42.6	6.5	*At*
47	hypothetical protein	20043048	25	3	21	17.7	11.1	*Os*
48	hypothetical protein	156763838	23	2	5	74.2	8.3	*Nt*
49	hypothetical protein	1402889	21	2	15	24.6	9.9	*At*
50	hypothetical protein	4581179	19	2	2	129.3	5.3	*At*
51	hypothetical protein	20197288	18	2	2	79.7	5.2	*At*
52	hypothetical protein	19920070	18	2	11	43.1	10.9	*Os*
53	hypothetical protein	12321084	17	2	4	69.7	6.4	*At*
54	hypothetical protein	33321045	17	2	5	70.1	6.5	*Zm*
55	hypothetical protein	2829867	17	2	3	91.2	9.6	*At*
56	hypothetical protein	4680491	16	2	1	98.2	7.7	*Os*
57	hypothetical protein	33321046	16	2	3	99.5	5.9	*Zm*
58	hypothetical protein	11079488	16	2	10	47.5	9.6	*At*
59	hypothetical protein	6562308	16	2	2	168.9	6.0	*At*
60	hypothetical protein	12321684	15	2	6	72.9	9.9	*At*
61	hypothetical protein	2244937	15	2	2	56.6	4.0	*At*
62	hypothetical protein	4733986	15	2	3	97.8	6.2	*At*
63	hypothetical protein	19920121	15	2	3	81.2	5.5	*Os*
64	hypothetical protein	21805747	14	2	3	44.8	9.6	*At*
65	hypothetical protein	110740236	14	1	11	11.9	10.3	*At*
66	hypothetical protein	19881563	14	2	1	95.8	11.5	*Os*
67	hypothetical protein	20197076	13	1	7	12.6	8.8	*At*
68	hypothetical protein	9802760	13	1	1	58.8	9.3	*At*
69	hypothetical protein STB1_54t00008	53749455	15	2	6	61.5	8.4	*St*
70	IAA-leucine resistant (ILR)-like gene 1 (ILL1)	15241892	16	1	3	47.7	6.2	*At*
71	INCOMPLETE ROOT HAIR ELONGATION (IRE)	15241795	17	2	2	130.1	5.4	*At*
72	ISOVALERYL-COA-DEHYDROGENASE (IVD)	15230664	27	3	12	44.8	7.5	*At*
73	light induced protein like	79325005	44	3	20	16.5	9.6	*At*
74	lysine ketoglutarate reductase/saccharopine dehydrogenase	10716965	27	3	4	116.3	5.5	*Zm*
75	membrane alanyl aminopeptidase	22330618	23	3	3	155.7	6.2	*At*
76	methionine over-accumulator (MTO2)	18410191	31	2	5	56.9	8.2	*At*
77	methionine synthase	33325957	644	39	61	84.3	5.9	*Gm*
78	methionine synthase	8439545	34	2	3	84.7	5.9	*St*
79	methionine--tRNA ligase, putative	15236350	19	1	1	89.9	6.7	*At*
80	methylenetetrahydrofolate reductase 1 (MTHFR1)	15232215	38	3	7	66.3	5.6	*At*
81	mitogen-activated protein kinase	78096654	25	3	12	42.9	6.1	*Nt*
82	NADH dependent glutamate synthase	4008156	31	4	2	236.7	6.7	*Os*
83	nodulin-like protein	110736366	15	2	9	36.5	8.9	*At*
84	ornithine aminotransferase	77540214	14	1	2	51.4	8.1	*Gm*
85	ornithine carbamoyltransferase (OTC)	15222192	16	1	2	41.0	7.2	*At*
86	peptidase M3 family protein / thimet oligopeptidase family protein	18424970	30	2	4	88.8	5.9	*At*
87	peptidemethionine sulfoxide reductase 1 (PMSR1)	12597894	37	3	14	22.8	6.2	*Os*
88	peptidemethionine sulfoxide reductase 2 (PMSR2)	15240795	15	1	4	24.4	5.1	*At*
89	phosphatidylserine decarboxylase	29465780	15	2	5	50.2	9.3	*Le*
90	plastid transcriptionally active 17 (PTAC17)	15218287	16	1	3	50.3	5.3	*At*
91	prephenate dehydratase family protein	18390869	27	2	5	44.8	6.1	*At*
92	proline 4-dioxygenase (P4H ISOFORM 2)	18397528	27	2	10	33.0	5.9	*At*
93	protein-serine/threonine kinase	505146	13	2	7	47.0	9.0	*Nt*
94	putative aminopeptidase	12324950	30	3	2	108.1	6.0	*At*
95	putative aspartate aminotransferase	4102887	19	1	15	10.2	5.0	*Hv*
96	putative carbamoyl phosphate synthase large subunit	21535791	16	2	4	132.1	5.7	*Nt*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
97	putative glutamate decarboxylase	32493114	149	10	27	51.3	5.7	*Gm*
98	putative glutamine synthetase	121489623	23	2	6	39.0	5.5	*Ps*
99	putative methionine synthase	14532772	29	3	6	84.6	6.1	*At*
100	putative protein	4467128	33	2	2	83.4	5.3	*At*
101	putative serine carboxypeptidase	11967861	31	2	4	55.6	6.1	*Ps*
102	pyridoxine biosynthesis 1.3 (PDX1)	46399269	29	2	7	33.1	5.9	*Nt*
103	pyridoxine biosynthesis 2 (ATPDX2/EMB2407/PDX2)	18424366	16	1	3	27.4	5.2	*At*
104	pyridoxine biosynthesis protein	72256519	138	9	31	33.2	5.6	*Gm*
105	S-adenosyl-L-homocysteine hydrolase 1 mutant	60266729	176	13	32	53.5	5.6	*At*
106	S-adenosyl-L-methionine:carboxyl methyltransferase family protei	22330992	16	2	3	43.4	5.4	*At*
107	S-adenosylmethionine synthetase-2	37051117	89	6	23	37.6	6.3	*Ps*
108	serine carboxypeptidase 1 precursor-like protein	22531054	31	2	4	55.9	5.8	*At*
109	serine carboxypeptidase II, putative	12322985	22	2	3	51.3	6.9	*At*
110	serine carboxypeptidase-like 50 (SCPL50)	15223991	18	1	2	49.2	5.4	*At*
111	serine hydroxymethyltransferase 2 (SHM2)	30690400	143	10	23	59.1	8.8	*At*
112	serine hydroxymethyltransferase 3 (SHM3)	18418028	34	3	6	58.0	9.0	*At*
113	serine hydroxymethyltransferase 4 (SHM4)	15236375	131	8	23	51.7	6.8	*At*
114	serine hydroxymethyltransferase 5 (SHM5)	15236371	62	4	6	52.3	5.7	*At*
115	serine hydroxymethyltransferase 6 (SHM6)	15219182	16	2	1	66.6	6.4	*At*
116	serine protein kinase like protein	110736589	16	2	1	61.1	6.4	*At*
117	serine/threonine protein kinase	145339108	15	2	5	106.4	6.8	*At*
118	serine/threonine protein kinase (WNK4)	15237174	18	2	10	64.9	5.6	*At*
119	serine/threonine-protein phosphatase PP2A-1 catalytic subunit	126517972	32	3	10	35.5	5.1	*Ta*
120	serine-type peptidase	145358557	30	3	5	81.3	5.6	*At*
121	serine-type peptidase	30690669	28	2	3	106.3	6.0	*At*
122	ser-thr protein kinase-like protein	9294588	15	2	1	141.6	4.9	*At*
123	similar to fumarylacetoacetate hydrolase	7485073	14	1	2	40.4	8.8	*At*
124	similar to lysine decarboxylase (T3B23.2/T3B23.2)	18401696	14	1	5	23.2	6.1	*At*
125	spermidine synthase	33340515	32	2	11	28.6	4.8	*Nt*
126	tetrahydrofolylpolyglutamate synthase-like protein	10177571	19	2	6	63.1	5.7	*At*
127	threonine aldolase 2 (THA2)	7547111	19	1	3	39.9	5.9	*At*
128	threonine synthase	83272147	30	2	4	56.3	6.7	*Gm*
129	tryptophan synthase alpha subunit	107599348	32	2	20	18.3	6.4	*Gm*
130	type one serine/threonine protein phosphatase 8 (TOPP8)	30690815	21	2	5	36.8	5.4	*At*
131	tyrosine phosphatase I	3413473	76	5	18	38.3	6.1	*Gm*
132	unnamed protein product	18543	122	8	31	46.3	8.6	*Gm*
133	unnamed protein product	9294643	32	4	4	115.8	5.7	*At*
134	unnamed protein product	9758523	20	2	2	64.3	9.5	*At*
135	unnamed protein product	9294460	16	2	3	77.3	6.8	*At*
136	unnamed protein product	10176995	14	2	3	82.7	6.3	*At*
137	unnamed protein product	9294510	13	2	2	67.0	5.5	*At*
138	xylan serine peptidase 1 (XSP1)	18411254	14	2	1	80.3	9.4	*At*
	01 Metabolism							
	01.02 Nitrogen and sulphur							
139	ATP sulfurylase	90194295	50	4	8	51.8	7.3	*Gm*
140	auxin amidohydrolase	51538213	18	1	3	47.3	5.5	*Ta*
141	glutamate-1-semialdehyde 2,1-aminomutase	19875	55	4	16	51.0	7.1	*Nt*
142	mercaptopyruvate sulfurtransferase 1 (ST1)	18412307	41	3	7	41.9	6.0	*At*
143	nitrate reductase	4389417	22	3	5	100.1	6.4	*Gm*
144	nitrate reductase	5020385	14	2	3	101.5	6.4	*Zm*
145	nitrite reductase	1906002	24	3	8	66.9	7.0	*Gm*
146	ricske iron-sulfur protein-like	77416965	14	1	4	28.9	9.1	*Si*
147	sulfate permease	2738752	15	2	9	25.3	10.1	*Zm*
148	sulfite oxidase (SOX)	15232230	16	2	5	43.3	8.8	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	01 Metabolism							
	01.03 Nucleotides							
149	3'-5' exonuclease/ nucleic acid binding	4585987	23	3	7	74.1	5.4	*At*
150	3'-5'-exoribonuclease/ RNA binding	15227488	20	2	10	33.4	4.7	*At*
151	adenine phosphoribosyltransferase-like	82621166	60	4	24	19.8	5.2	*St*
152	adenylate kinase family protein	30686829	22	2	8	27.8	7.0	*At*
153	adenylosuccinate synthetase	9858775	56	5	10	64.5	8.1	*Le*
154	adenylosuccinate synthetase	21537345	29	2	6	53.0	6.7	*At*
155	ATRZ-1A, nucleotide binding	15231557	15	2	12	26.9	8.9	*At*
156	beta-1,3-glucanase-like protein	9758115	17	1	1	50.7	5.7	*At*
157	chloroplast alpha-glucan water dikinase isoform 3	53771834	17	2	1	131.3	5.9	*At*
158	cytosolic glutamine synthetase GSbeta1	10946357	265	16	57	39.0	5.5	*Gm*
159	dcncter-like protein 3 (DMI3)	2911056	15	2	3	106.3	9.5	*At*
160	dihydroorotase, pyrimidin 4 (PYR4)	15235865	19	1	4	41.9	8.6	*At*
161	dihydroorotate dehydrogenase family protein	15229529	53	3	9	46.8	6.4	*At*
162	endo-1,3-beta-glucanase	38640795	104	7	33	36.6	8.9	*Gm*
163	endonuclease	145442288	32	2	9	33.5	6.1	*Gm*
164	exonuclease family protein	15239167	16	2	4	53.2	8.7	*At*
165	formate tetrahyrofolate ligase	17017271	36	3	9	31.6	7.2	*Zm*
166	GDP dissociation inhibitor	2501850	115	8	21	49.7	5.4	*Nt*
167	GDP-D-mannose-4,6-dehydratase MUR1	21536808	28	3	8	42.0	5.7	*At*
168	glycinamide ribonucleotide transformylase	32815066	44	4	14	34.5	9.1	*Gm*
169	GTP-binding protein, putative	8439910	25	3	2	116.3	6.2	*At*
170	guanine nucleotide exchange family protein	15233734	20	2	1	188.8	5.5	*At*
171	guanine nucleotide regulatory protein	1208537	58	4	23	23.2	7.7	*Gm*
172	guanosine polyphosphate pyrophosphohydrolase (RSH1)	110742286	16	2	2	98.7	6.6	*At*
173	hypothetical protein	62319315	19	1	7	22.6	6.3	*At*
174	inosine monophosphate dehydrogenase	4468193	24	2	7	53.4	5.5	*Gm*
175	isopentenyl transferase	47498592	15	2	8	38.1	9.0	*Gm*
176	NAD+ ADP-ribosyltransferase	26451296	13	2	4	43.6	5.0	*At*
177	nucleoside diphosphate kinase	26245395	124	8	55	16.4	6.9	*Gm*
178	nucleoside diphosphate kinase	6435320	105	7	24	25.3	9.4	*Ps*
179	nucleoside diphosphate kinase	16396	16	1	8	16.2	7.0	*At*
180	nucleoside diphosphate kinase (NDPK2)	30697820	21	2	5	25.6	9.1	*At*
181	nucleotidyltransferase	30677893	19	2	6	42.3	6.1	*At*
182	nucleotidyltransferase family protein	18406841	17	2	2	85.8	6.4	*At*
183	pigment defective (PDE194)	15219681	16	2	8	38.5	9.3	*At*
184	pollen tube RhoGDI2	89473698	16	2	7	26.6	4.7	*Nt*
185	poly (ADP-ribose) polymerase, putative	30684908	13	2	2	111.2	8.8	*At*
186	pseudouridine synthase and archaeosine transglycosylase-(PUA) domain-containing protein	15217434	14	1	4	20.5	8.9	*At*
187	putative carbamoyl phosphate synthase small subunit	21535793	38	3	6	47.3	6.1	*Nt*
188	putative protein	3046704	15	2	3	129.7	9.1	*At*
189	R1 (homologue to putative disease resistance protein)	53831167	17	2	4	59.7	5.7	*St*
190	repressor of silencing 2b	138996990	16	2	2	186.5	6.4	*Nt*
191	ribonucleotide reductase	4151066	15	2	3	91.2	7.8	*Nt*
192	sec7 domain-containing protein	15241142	14	2	1	156.2	5.0	*At*
193	similar nucleotide excision repair proteins	9972383	19	2	4	71.3	8.1	*At*
194	tatD-related deoxyribonuclease family protein	42572645	14	1	3	36.0	6.1	*At*
195	tRNA-splicing endonuclease positive effector-related	10177999	18	2	2	100.2	8.6	*At*
196	UMP synthase	14582292	17	1	1	51.9	6.5	*At*
197	unknown protein, contains polynucleotidyl transferase domain	18399624	14	2	5	25.4	9.8	*At*
198	uracil phosphoribosyltransferase	8778301	51	5	7	61.0	6.8	*At*
199	uridylate kinase / uridine monophosphate kinase	30690243	63	5	21	23.1	6.4	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	01 Metabolism							
	01.06 Phosphate							
	01 Metabolism							
	01.08 Polysaccharide catabolism							
200	alpha-galactosidase I	53747927	90	5	16	45.0	6.5	*Ps*
201	alpha-glucosidase, putative	42562299	15	2	3	90.6	5.5	*At*
202	alpha-xylosidase 1 (ATXYL1/XYL1)	15221437	20	2	2	102.4	6.3	*At*
203	beta-amylase	62122635	95	7	18	56.1	5.3	*Gm*
204	beta-galactosidase	15451018	22	2	4	80.6	8.9	*At*
205	beta-galactosidase like protein	2961390	44	3	4	95.2	7.0	*At*
206	beta-glucosidase-like protein	4455284	17	2	4	59.9	7.2	*At*
207	beta-mannosidase enzyme	17226270	39	2	5	59.3	8.8	*Le*
208	BGAL12 (beta-galactosidase 12)	4538943	28	2	2	81.7	8.0	*At*
209	BMY7/TR-BAMY (beta-amylase 7)	15229544	22	2	4	63.8	5.7	*At*
210	BXL2 (beta-xylosidase 2)	18378991	14	2	4	83.0	8.6	*At*
211	endo-1,4-beta-xylanase	71142588	25	3	8	60.8	5.4	*Hv*
212	putative 1,4-beta-xylanase	19920134	20	3	5	59.7	8.8	*Os*
213	putative beta-galactosidase	34148077	70	5	9	102.5	6.1	*Gm*
214	putative beta-galactosidase	16905220	20	2	2	91.5	6.0	*Os*
	01 Metabolism							
215	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	24745903	14	2	3	59.7	9.0	*St*
216	3-isopropylmalate dehydrogenase, chloroplast, putative	15241338	18	2	5	44.2	5.8	*At*
217	ADP-glucose pyrophosphorylase	1237080	68	5	12	56.3	6.1	*Ps*
218	allene oxide cyclase	14423351	51	3	12	26.6	8.7	*Le*
219	allene oxide synthase	82795997	51	4	9	58.9	9.2	*Gm*
220	alpha 1,4-glucan phosphorylase L isozyme	21579	83	6	7	109.6	5.4	*St*
221	alpha-1,4-glucan-protein synthase [UDP-forming]	34588146	233	15	41	41.2	5.8	*Zm*
222	alpha-L-arabinofuranosidase/beta-D-xylosidase isoenzyme ARA-I	18025340	18	2	2	82.0	5.6	*Hv*
223	ATCWINV2 (cell wall invertase 2)	116831291	19	2	5	66.9	9.2	*At*
224	ATGSL03 (glucan synthase-like 3)	30685131	15	2	1	227.1	9.2	*At*
225	ATGSL09 (glucan synthase-like 9)	42568155	16	2	1	214.9	8.4	*At*
226	ATGSL11 (glucan synthase-like 11)	42566048	14	2	1	225.1	8.8	*At*
227	ATSPS2F (sucrose phosphate synthase 2F)	79510910	14	1	1	117.0	6.2	*At*
228	AXS1 (UDP-D-apiose/UDP-D-xylose synthase 1)	15226264	167	12	31	43.6	5.5	*At*
229	carbon-nitrogen hydrolase family protein	22326744	17	1	4	40.3	8.8	*At*
230	chloroplast NAD-MDH	3256066	41	2	7	42.4	8.5	*At*
231	galactokinase	53747925	50	3	7	54.6	5.4	*Ps*
232	galactosyltransferase family protein	4835784	25	3	11	45.6	9.2	*At*
233	gamma-aminobutyrate transaminase subunit precursor isozyme 1	29837282	64	4	10	56.7	7.7	*Le*
234	gamma-glutamylcysteine synthetase precursor	10130004	16	1	6	34.4	8.5	*Gm*
235	GATL5 (polygalacturonate-4-alpha-galactosyltransferase	15217851	17	2	8	41.2	8.9	*At*
236	GDH1 (glutamate dehydrogenase 1)	15238762	15	1	4	44.5	6.4	*At*
237	GDP-mannose pyrophosphorylase	13509287	77	5	16	39.6	6.3	*At*
238	GER2 (GDP-4-keto-6-deoxymannose-3,5-epimerase-4-reductase 1	18394547	18	1	3	36.2	6.3	*At*
239	glucosamine/galactosamine-6-phosphate isomerase-related	15238526	14	1	3	28.5	5.8	*At*
240	glucose-6-phosphate dehydrogenase	115394806	51	4	7	59.4	5.9	*Ps*
241	glucose-6-phosphate isomerase	51340062	143	11	24	67.7	5.5	*Le*
242	glucose-6-phosphate isomerase	51340060	20	2	7	62.9	6.5	*St*
243	glucose-6-phosphate-dehydrogenase	3452137	41	3	44	8.2	9.6	*Gm*
244	glucosyltransferase	82618888	32	2	4	53.2	6.7	*Gm*
245	glutamate dehydrogenase 1	59668638	250	15	53	44.5	6.0	*Gm*
246	GLX2-5 (glyoxalase 2-5)	15224661	17	2	3	35.8	9.0	*At*
247	glycoside hydrolase family 2 protein	42561840	17	2	1	107.7	6.0	*At*
248	glycosyl hydrolase family 1 protein	15224879	19	1	1	56.9	7.6	*At*
249	glycosyl hydrolase family 1 protein	30689724	15	2	2	51.7	6.1	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
256	glycosyl transferase family 43 protein	15240245	15	2	8	55.3	9.7	*At*
257	glycosyl transferase-related	15235222	16	2	1	147.4	6.2	*At*
258	glyoxalase I	4127862	101	8	47	21.0	5.6	*Gm*
259	glyoxalase I	37932483	14	1	3	32.4	5.6	*Zm*
260	glyoxalase I, putative	9828630	46	3	8	40.0	7.0	*At*
261	granule-bound starch synthase	119710158	13	2	5	68.1	6.6	*Gm*
262	hexokinase, putative	15222973	14	2	10	54.6	5.5	*At*
263	hydrolase family protein / HAD-superfamily protein	15231226	14	1	2	41.9	6.8	*At*
264	hydroxyisourate hydrolase	19569603	22	2	4	63.8	6.1	*Gm*
265	indole-3-glycerol phosphate synthase (IGPS)	21592587	14	2	2	44.6	7.0	*At*
266	iron sulfur subunit of succinate dehydrogenase	4803711	16	1	3	31.1	8.8	*Zm*
267	isoamylase isoform 2	73698627	15	2	1	96.1	5.4	*Ps*
268	latex abundant protein, putative (AMC5) / caspase family protein	15219340	21	2	10	40.2	4.7	*At*
269	L-galactono-gamma-lactone dehydrogenase	6519872	19	2	3	66.8	7.7	*Nt*
270	malate dehydrogenase, cytoplasmic	18202485	15	1	6	35.6	5.8	*Zm*
271	metacaspase 1	23343885	63	4	8	44.9	4.8	*La*
272	metacaspase 7	15219345	28	2	4	45.5	4.7	*At*
273	mitochondrial succinate dehydrogenase iron-sulphur subunit	21555840	72	5	16	31.3	8.8	*At*
274	nodule-enhanced sucrose synthase	3377764	179	12	17	92.4	6.0	*Ps*
275	pectate lyase precursor	127464581	15	1	3	34.5	6.4	*Gm*
276	phosphomannomutase	90762150	90	6	27	28.0	5.8	*Gm*
277	putative 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase	12597884	13	1	2	58.5	8.7	*Os*
278	putative glucosyltransferase	28302068	14	2	6	52.3	6.3	*Gm*
279	putative trehalose-6-phosphate synthase	12324075	14	1	2	96.5	5.8	*At*
280	RHM1/ROL1 (rhamnose biosynthesis 1)	15218420	191	13	19	75.4	6.8	*At*
281	ripening-related protein-like	8885554	14	1	4	31.5	4.7	*At*
282	SDH1-2 (succinate dehydrogenase 1-2)	15224174	54	4	7	69.4	5.8	*At*
283	SEX1 (starch excess 1)	6573745	31	3	3	172.1	5.8	*At*
284	similar to 3-deoxy-D-manno-2-octulosonate-8-phosphate synthase	4966354	63	4	15	32.6	5.8	*At*
285	SIP1 (seed imbibition 1-like)	15242680	13	1	1	86.2	4.9	*At*
286	soluble acid invertase	47969540	14	2	4	71.7	5.2	*Hv*
287	succinyl CoA ligase beta subunit-like protein	83284007	63	5	14	45.3	5.7	*St*
288	sucrose synthase	63852202	69	5	39	23.2	6.0	*Gm*
289	sucrose synthase 3	22121990	24	2	1	91.9	6.1	*Zm*
290	sucrose-phosphate synthase 2	33341083	18	2	2	111.3	6.2	*Ta*
291	sucrose-phosphate synthase isoform C	77176829	16	2	1	117.9	6.3	*Nt*
292	SUS2 (sucrose synthase 2)	15239816	19	2	1	92.1	5.7	*At*
293	thiamin biosynthetic enzyme	6552397	63	5	22	37.0	5.8	*Gm*
294	UDP-glucoronosyl/UDP-glucosyl transferase family protein	15221232	22	3	1	134.8	8.0	*At*
295	UDP-glucose-6-dehydrogenase, putative	48093457	118	7	17	61.0	6.5	*Nt*
296	UTP:alpha-D-glucose-1-phosphate uridylyltransferase	28863909	123	9	19	51.9	5.5	*St*
297	UXS2 (UDP-glucoronic acid decarboxylase 2)	48093465	63	5	14	49.9	9.4	*Nt*
298	xyloglucan endotransglycosylase precursor	89145876	15	1	7	19.5	6.6	*Gm*
299	xylose isomerase	21537178	57	4	9	53.7	5.5	*At*
	01 Metabolism							
	01.07 Lipid							
300	24 kDa oleosin isoform	18720	59	5	33	15.8	8.2	*Gm*
301	3-hydroxybutyryl-CoA dehydrogenase, putative	15232545	20	1	5	31.7	6.6	*At*
302	3-ketoacyl-CoA thiolase	37549269	128	8	26	47.0	8.8	*Gm*
303	AAE15 (acyl-activating enzyme 15)	22328609	17	2	3	81.5	8.9	*At*
304	acetyl-CoA carboxylase	8886469	54	5	11	58.8	7.2	*Gm*
305	acetyl-CoA carboxylase	14423251	15	2	3	120.6	6.0	*Zm*
306	acetyl-CoA carboxylase carboxyltransferase beta subunit	91214152	57	4	9	49.0	4.8	*Gm*
307	acetyl-CoA synthetase, putative	12323178	17	2	1	290.1	5.8	*At*
308	ADP-ribosylation factor	4324967	127	8	47	20.4	6.4	*Gm*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Identi.	Cov. %	Theoretical MW	Theoretical pI	Species
309	AIM1 (ABNORMAL INFLORESCENCE MERISTEM)	15235527	22	2	2	77.9	9.4	*At*
310	AMP-dependent synthetase and ligase family protein	15218839	16	2	3	64.9	7.6	*At*
311	ARF3/ARL1/ATARL1 (ADP-ribosylation factor 3)	30682545	17	1	8	20.2	5.2	*At*
312	beta-hydroxyacyl-ACP dehydratase, putative	15238069	17	1	3	24.1	9.3	*At*
313	beta-ketoacyl-ACP synthetase I	7385201	132	9	30	49.7	7.2	*Gm*
314	CAC3 (acetyl co-enzyme A carboxylase carboxyltransferase alpha	4895181	13	2	4	88.5	5.8	*At*
315	enoyl-ACP reductase	2204236	79	6	14	41.8	8.9	*Nt*
316	enoyl-Acp reductase	32400828	28	2	16	15.6	10.0	*Ta*
317	enoyl-CoA hydratase/isomerase	79473201	19	2	7	46.3	6.2	*At*
318	enoyl-CoA hydratase/isomerase family protein	42565158	17	1	2	45.7	6.1	*At*
319	enoyl-CoA hydratase/isomerase family protein	30683577	15	1	4	28.8	9.1	*At*
320	esterase, putative	15227376	36	3	5	31.7	5.9	*At*
321	fatty acyl coA reductase	22003082	16	2	4	57.5	8.8	*Ta*
322	GDSL-motif lipase/hydrolase family protein	15241404	21	2	5	43.6	8.7	*At*
323	GLIP7 (GDSL-motif lipase 7)	9755617	24	3	5	40.5	8.7	*At*
324	inorganic pyrophosphatase-like protein	21593570	99	7	27	24.6	5.3	*At*
325	lipase class 3 family protein	2244965	17	2	3	75.7	5.5	*At*
326	lipase, putative	13569989	26	3	4	56.8	8.7	*Os*
327	lipoxygenase	2598612	50	5	6	97.7	6.0	*Ps*
328	lipoxygenase	541746	25	2	3	97.1	6.3	*Ps*
329	lipoxygenase	493730	24	2	5	97.0	6.1	*Ps*
330	lipoxygenase	12620877	18	2	3	96.5	5.7	*Zm*
331	lipoxygenase	1407703	16	2	3	97.0	5.5	*St*
332	lipoxygenase, putative	15218506	15	2	3	104.8	7.1	*At*
333	lipoxygenase-9	152926332	304	22	32	96.4	6.5	*Gm*
334	LOX3 (lipoxygenase 3)	18394479	19	2	3	103.7	7.7	*At*
335	MFP2 (multifunctional protein)	15231317	23	2	3	78.8	9.2	*At*
336	PED1 (peroxisome defective 1)	15225798	36	3	9	48.6	8.6	*At*
337	phosphoesterase	8777472	15	2	2	87.3	6.0	*At*
338	phospholipase D alpha	6573119	17	1	1	92.2	5.4	*Le*
339	Phospholipase D alpha 1 (PLD alpha 1) (Choline phosphatase 1)	2499708	14	1	1	92.2	5.4	*Zm*
340	PLDALPHA1 (phospholipase D alpha 1)	15232671	28	2	2	91.8	5.5	*At*
341	stearoyl-acyl carrier protein desaturase B	62546347	35	3	7	47.2	6.0	*Gm*
342	WAVE3 (identical to irregular trichome branch 1)	7327832	17	2	2	131.9	4.5	*At*
	01 Metabolism							
	01.06 Sterol							
343	cytosolic acetoacetyl-coenzyme A thiolase	53854350	52	3	11	41.3	6.5	*Nt*
344	geranylgeranyl pyrophosphate synthetase-like protein	76363949	16	2	10	22.6	8.7	*La*
345	GGPS1 (geranylgeranyl phosphate synthase 1)	15234534	15	2	5	40.2	6.2	*At*
346	SMT1 (sterol methyltransferase 1)	15240691	30	2	7	38.3	5.9	*At*
	01 Metabolism							
347	GLP10 (germin-like protein 10)	15228673	15	1	7	23.6	8.9	*At*
348	pantothenate kinase family protein	21554145	21	1	5	41.0	4.8	*At*
	02 Energy							
	02.01 Glycolysis							
349	A_IG002N01.21 gene product	7267629	14	2	3	52.5	7.2	*At*
350	acyl-CoA oxidase	15553478	38	4	6	74.3	7.3	*Gm*
351	aldolase	77745483	101	7	23	37.7	7.7	*St*
352	aldolase C-1	786178	33	3	10	38.8	8.4	*Os*
353	aldose 1-epimerase family protein	15242099	44	3	10	35.4	5.7	*At*
354	aldose 1-epimerase family protein	30684727	16	1	3	37.2	5.9	*At*
355	chloroplastic aldolase	1781348	100	7	27	38.5	5.9	*St*
356	cytosolic phosphoglucomutase	15223226	123	8	14	63.5	5.6	*At*
357	cytosolic phosphoglycerate kinase	9230771	260	16	43	42.3	5.7	*Ps*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
358	enolase	42521309	468	28	72	47.7	5.3	*Gm*
359	enolase 2 (2-phosphoglycerate dehydratase 2)	1169528	66	4	14	48.2	5.7	*Zm*
360	fructose-1,6-bisphosphatase	5305145	17	1	3	35.9	6.3	*Ps*
361	fructose-bisphosphate aldolase	40457267	45	3	8	38.4	6.8	*Gm*
362	fructose-bisphosphate aldolase, putative	15226185	54	3	12	42.3	8.2	*At*
363	fructose-bisphosphate aldolase-like protein	15231715	79	6	12	38.5	6.1	*At*
364	GAPCP-1 (glyceraldehyde-3 phosphate dehydrogenase)	15219406	47	4	14	44.8	8.8	*At*
365	GAPCP-2 (glyceraldehyde-3-phosphate dehydrogenase)	21618027	31	2	5	44.9	8.7	*At*
366	glyceraldehyde-3-phosphate dehydrogenase	85720768	340	21	77	36.8	6.7	*Gm*
367	glyceraldehyde-3-phosphate dehydrogenase	19880027	97	7	17	54.2	6.8	*Os*
368	glyceraldehyde-3-phosphate dehydrogenase 1	3831571	31	2	28	10.2	4.8	*Gm*
369	glyceraldehyde-3-phosphate dehydrogenase A subunit	77540210	118	7	21	43.2	8.4	*Gm*
370	glyceraldehyde-3-phosphate dehydrogenase, cytosolic 3	6166167	96	6	16	36.4	7.0	*Zm*
371	inorganic pyrophosphatase-like protein	21592878	22	2	6	33.4	5.6	*At*
372	ketol-acid reductoisomerase	288063	44	3	7	63.9	6.5	*At*
373	L-lactate dehydrogenase (LDH)	126066	44	3	14	38.6	9.0	*Zm*
374	MEE51 (maternal effect embryo arrest 51)	3157931	55	4	5	62.4	5.8	*At*
375	pfkB-type carbohydrate kinase family protein	15221364	83	5	12	37.6	5.5	*At*
376	phosphoglycerate mutase	15231939	98	7	10	60.8	5.5	*At*
377	phosphoglycerate mutase family protein	15232324	15	1	2	30.4	6.1	*At*
378	pyrophosphate-dependent phosphofructo-1-kinase	4539423	96	7	17	55.3	8.1	*At*
379	pyrophosphate-dependent phosphofructokinase alpha subunit	110738773	22	2	3	67.6	6.8	*At*
380	pyrophosphate--fructose-6-phosphate-1-phosphotransferase-related	15218074	31	3	4	67.1	6.5	*At*
381	pyruvate kinase	59668642	189	13	33	54.4	6.8	*Gm*
382	pyruvate kinase-like protein	18409740	38	2	9	57.5	6.7	*At*
383	TIM (triosephosphate isomerase)	15226479	62	4	18	33.3	7.7	*At*
384	T-protein of the glycine decarboxylase complex	407475	159	10	31	44.3	8.8	*Ps*
385	triosephosphate isomerase	553107	14	1	5	27.6	6.6	*Os*
	02 Energy							
	02.02 Gluconeogenesis							
386	3-isopropylmalate dehydrogenase-like protein	7378609	62	5	13	40.4	6.3	*At*
387	cytosolic malate dehydrogenase	42521311	235	13	55	35.5	6.3	*Gm*
388	malate dehydrogenase	5929964	326	18	72	36.1	8.2	*Gm*
389	malate dehydrogenase-like protein	83283965	63	3	4	35.5	5.7	*St*
390	nodule-enhanced malate dehydrogenase	3377762	17	1	2	41.8	7.6	*Ps*
391	PMDH1 (peroxisomal NAD-malate dehydrogenase 1)	37725953	42	3	10	37.1	7.0	*Ps*
	02 Energy							
	02.07 Pentose phosphate							
392	6-phosphogluconate dehydrogenase	2529229	203	13	28	56.4	5.6	*Gm*
393	6-phosphogluconate dehydrogenase family protein	15222639	73	5	10	53.4	5.3	*At*
394	transaldolase-like	81076343	70	5	12	47.9	6.0	*St*
	02 Energy							
	02.10 TCA pathway							
395	2-oxoglutarate dehydrogenase, E1 subunit	4210330	104	7	9	116.7	7.0	*At*
396	aconitase family protein / aconitate hydratase family protein	18414006	31	2	4	55.0	8.1	*At*
397	aconitate hydratase, cytoplasmic / citrate hydro-lyase	15233349	217	15	22	98.2	6.0	*At*
398	coproporphyrinogen III oxidase	82469923	26	2	5	47.1	7.2	*Zm*
399	CSY2 (citrate synthase 2)	15231130	17	1	2	56.6	8.7	*At*
400	CSY5 (citrate synthase 5)	145339693	28	3	7	51.7	6.2	*At*
401	dihydrolipoamide dehydrogenase 2, plastidic	18414603	42	3	7	60.1	7.3	*At*
402	fumarase	1488652	17	2	4	53.4	6.5	*St*
403	isocitrate dehydrogenase (NADP) (EC 1.1.1.42)	479386	259	18	46	50.2	6.3	*Gm*
404	LPD2 (lipoamide dehydrogenase 2)	30684419	17	2	3	54.0	6.6	*At*
405	LTA2 (plastid E2 subunit of pyruvate decarboxylase)	15230922	54	4	8	50.1	8.3	*At*
406	malate synthase, putative	15237551	31	3	3	63.9	8.0	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Identified	Cov. %	Theoretical MW	Theoretical pI	Species
407	putative protein	14140144	15	1	8	26.4	7.6	*Os*
408	pyruvate decarboxylase	1616787	46	4	7	65.9	5.7	*At*
409	pyruvate decarboxylase family protein	15237954	33	3	4	61.5	5.7	*At*
410	pyruvate dehydrogenase E1 alpha subunit	3851005	54	4	7	42.9	8.4	*Zm*
411	pyruvate dehydrogenase E1 beta subunit	2454184	35	2	6	44.3	5.9	*At*
412	pyruvate dehydrogenase E1 beta subunit isoform 2	3851001	45	3	9	40.0	5.6	*Zm*
413	pyruvate dehydrogenase E1 beta subunit, mitochondrial	15241286	26	2	6	39.2	5.7	*At*
	02 Energy							
	02.13 Respiration							
414	ACLB-2 (ATP-citrate lyase B-2)	15239897	148	10	19	65.8	7.5	*At*
415	AT4g26970/F10M23_310	18416900	33	2	2	108.5	6.7	*At*
416	ATP synthase subunit beta, mitochondrial precursor	114420	274	17	46	59.1	6.0	*Zm*
417	dihydrolipoamide acetyltransferase	11994364	37	2	3	59.7	7.5	*At*
418	uncoupling protein 1a	18378376	26	2	7	25.8	9.6	*Gm*
	02 Energy							
	02.16 Fermentation							
	02 Energy							
	02.20 Electron-transport							
419	anion-transporting ATPase family protein	30681260	38	3	7	44.8	7.6	*At*
420	ATB5-A (cytochrome b5 A)	15238776	16	1	6	15.1	5.1	*At*
421	ATCBR (NADH:cytochrome B5 reductase 1)	15238025	18	1	6	31.5	8.6	*At*
422	ATLFNR1 (LEAF FNR 1)(similar to ferredoxin-NADP-reductase)	145334919	42	3	18	29.7	5.9	*At*
423	ATPase subunit 1	15226092	55	4	6	85.9	5.4	*At*
424	ATRFNR2 (ROOT FNR 2), oxidoreductase	30691910	37	3	6	42.8	8.8	*At*
425	CAD4 (cinnamyl alcohol dehydrogenase 4)	15230382	18	1	4	39.1	5.3	*At*
426	cinnamyl alcohol dehydrogenase 1	60265616	45	3	9	35.5	6.5	*Nt*
427	cinnamyl alcohol dehydrogenase, putative	12325359	19	2	1	91.8	8.8	*At*
428	COX6B (cytochrome c oxidase 6B)	15219886	33	3	12	21.2	4.3	*At*
429	CYP710A1 (cytochrome P450 710A1)	15226758	15	2	6	55.7	7.7	*At*
430	CYP712A1 (cytochrome P450 712A1)	15227911	19	2	4	58.1	9.0	*At*
431	CYP718 (cytochrome P450 718)	15228011	13	1	1	55.4	9.1	*At*
432	CYP71A13 (cytochrome P450 71A13)	42569483	17	2	1	56.8	8.7	*At*
433	CYP71B17 (cytochrome P450 71B17)	15231517	15	2	4	57.2	7.3	*At*
434	CYP71B2 (CYTOCHROME P450 71B2)	15222174	16	2	8	57.1	7.2	*At*
435	CYP71B34 (cytochrome P450 71B34)	15231538	18	2	7	57.1	6.9	*At*
436	cytochrome b-559 alpha subunit	27446512	29	2	36	6.5	4.4	*Ps*
437	cytochrome c	118004	16	1	7	12.0	9.3	*Zm*
438	cytochrome c oxidase subunit Vb	113734313	44	3	31	16.7	5.9	*Ps*
439	cytochrome c1	498789	34	2	13	28.6	5.3	*Si*
440	cytochrome f	91214155	34	3	15	35.3	8.9	*Gm*
441	cytochrome P450 family protein	42565543	16	2	7	54.9	9.3	*At*
442	cytochrome P450 like protein	2244889	17	2	8	36.9	7.1	*At*
443	cytochrome P450 monooxygenase CYP74A2	85001787	129	9	46	22.8	5.6	*Gm*
444	cytochrome P450 reductase	6503253	21	2	3	78.5	5.4	*Ps*
445	cytochrome P450-like protein	46798536	16	2	10	53.7	8.6	*Ta*
446	cytochrome-c oxidase	56675440	15	1	7	29.5	5.0	*Ps*
447	DNA (cytosine-5)-methyltransferase 2 (Chromomethylase 2)	75167623	16	2	2	101.6	5.4	*Zm*
448	ETFALPHA (electron transport flavoprotein alpha)	15223680	19	1	5	38.4	6.5	*At*
449	F27F5.14 (similar to copper ion binding / electron carrier)	7767674	15	2	7	43.7	5.6	*At*
450	ferredoxin:sulfite reductase precursor	12658639	89	6	13	63.8	9.1	*Gm*
451	ferredoxin-NADP(H) oxidoreductase	20302471	19	2	7	38.8	8.3	*Ta*
452	FQR1 (flavodoxin-like quinone reductase 1)	15239652	57	4	26	21.8	6.0	*At*
453	nad7 (NADH dehydrogenase subunit D)	81176545	14	1	2	44.3	6.6	*Ta*
454	NAD-dependent formate dehydrogenase	4760553	91	7	19	41.2	6.9	*Os*
455	NAD-dependent isocitrate dehydrogenase	54777949	75	5	23	29.0	5.6	*Zm*
456	NADH dehydrogenase, subunit 9	57337517	57	4	26	22.9	7.8	*Zm*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
457	NADH-cytochrome b5 reductase, putative	18420117	63	4	15	36.0	8.8	*At*
458	NADH-ubiquinone oxidoreductase 20 kDa subunit, mitochondrial	15239782	41	3	22	24.0	9.5	*At*
459	NADP-dependent malate dehydrogenase	18377751	73	5	15	48.3	5.8	*At*
460	NADPH-dependent mannose 6-phosphate reductase	21554266	57	4	13	35.1	6.1	*At*
461	ndhD gene product	12424	16	2	7	58.2	8.7	*Zm*
462	nucleotide-binding subunit of vacuolar ATPase	166627	300	18	55	54.7	5.0	*At*
463	OPR3 (OPDA-reductase 3)	15225045	13	2	6	42.7	7.7	*At*
464	oxidoreductase	15232542	16	1	3	37.2	5.7	*At*
465	oxidoreductase, zinc-binding dehydrogenase family protein	15234529	28	2	6	34.5	6.5	*At*
466	PDI-like protein	21592996	17	2	2	65.2	4.9	*At*
467	predicted NADH dehydrogenase 24 kD subunit	18411985	37	3	11	28.4	8.1	*At*
468	putative cytochrome	13786468	14	1	11	14.8	4.8	*Os*
469	Putative quinone oxidoreductase	15451578	14	1	2	35.5	5.5	*Os*
470	rieske Fe/S protein of cytochrome b6/f complex	19999	15	1	5	24.1	7.6	*Nt*
471	VHA-A (vacuolar H+-ATPase A2 subunit form)	15219234	268	17	33	68.8	5.1	*At*
472	V-type proton-ATPase	1143394	54	4	14	26.1	6.0	*At*
473	YUC2 (YUCCA2)	15235652	16	2	3	46.5	8.2	*At*
	02 Energy							
	02.30 Photosynthesis							
474	33 kDa polypeptide of oxygen-evolving complex	15451006	38	4	14	35.2	5.6	*At*
475	33kDa precursor protein of oxygen-evolving complex	809113	87	6	16	35.3	5.9	*St*
476	APH, carbonate dehydratase	15220153	20	2	8	30.1	6.7	*At*
477	ATP synthase CF1 alpha subunit	14017569	16	2	2	55.3	6.1	*Ta*
478	ATP synthase CF1 beta subunit	91214126	327	21	58	53.8	5.3	*Gm*
479	ATP synthase CF1 epsilon subunit	91214127	13	1	11	14.8	5.4	*Gm*
480	CA2 (beta carbonic anhydrase 2)	30685030	18	2	4	36.6	7.1	*At*
481	carbonic anhydrase	8096277	19	1	13	13.8	5.5	*Nt*
482	carbonic anhydrase, putative / carbonate dehydratase, putative	15220853	17	1	4	28.8	6.5	*At*
483	chlorophyll a/b binding protein type II	16805332	29	2	6	28.6	5.5	*Gm*
484	chlorophyll a-b binding protein	2570511	22	2	8	33.4	9.7	*Os*
485	chloroplastic ATP synthase gamma subunit	124294683	23	2	41	7.2	4.6	*Gm*
486	glycolate oxidase	2570515	15	2	4	40.2	8.7	*Os*
487	similar to phosphoenolpyruvate synthase (ppsA) (GB:AE001056)	3319357	17	2	3	70.8	5.3	*At*
488	CRTISO (carotenoid isomerase)	42561764	19	2	3	65.4	8.3	*At*
489	FKBP15-2 (FK506-binding protein 15 kD-2)	15239019	14	1	7	17.7	5.3	*At*
490	H+-transporting ATP synthase	19881545	38	3	10	41.0	8.5	*Os*
491	light-harvesting chlorophyll-a/b binding protein Lhcb1	56809379	23	1	6	28.4	5.5	*Ps*
492	NADPH-protochlorophyllide oxidoreductase	46019982	15	2	5	39.8	9.2	*Zm*
493	oxygen-evolving complex-related	7573402	18	2	9	35.8	9.0	*At*
494	phosphoenolpyruvate carboxylase	25901015	428	29	36	110.9	6.1	*Gm*
495	photosystem I subunit PsaD	148372347	14	1	14	23.0	9.6	*Gm*
496	photosystem I subunit VII	7525086	36	2	35	9.0	6.7	*At*
497	photosystem II protein D2	91214138	22	2	5	39.6	5.3	*Gm*
498	photosystem II protein H	91214170	16	1	21	7.8	5.0	*Gm*
499	photosystem II protein VI	11465974	14	1	20	4.5	10.7	*Nt*
500	PHYA (PHYTOCHROME A)	15217562	14	2	3	124.5	5.9	*At*
501	phytochrome B	85679505	14	2	3	125.3	5.7	*St*
502	phytochrome B2 apoprotein	37926881	14	2	2	128.3	5.8	*Zm*
503	PII protein	89357468	22	2	18	11.3	5.5	*Nt*
504	PORA (protochlorophyllide reductase A)	15239574	21	2	5	43.9	9.4	*At*
505	ppc2 (phosphoenolpyruvate carboxylase)	3777449	42	3	3	110.2	5.6	*St*
506	PSAH-1 (photosystem I subunit H-1)	15218186	17	1	7	15.3	9.9	*At*
507	PSBP-1 (oxygen-evolving enhancer protein 2)	15222166	18	1	5	28.1	6.9	*At*
508	ribose 5-phosphate isomerase-related	21594002	20	2	10	29.3	5.7	*At*
509	ribulose 1,5-bisphosphate carboxylase large subunit	3114769	250	19	43	52.4	6.1	*Gm*
510	ribulose bisphosphate carboxylase activase B	7960277	86	5	16	47.8	6.9	*Ta*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
511	ribulose-1,5-bisphosphate carboxylase small subunit rbcS1	10946375	171	11	48	20.0	8.9	*Gm*
512	succinyl-CoA ligase alpha 2 subunit	49617539	76	5	17	35.4	9.0	*Le*
513	transketolase 1	77563673	15	1	18	7.1	9.1	*Nt*
514	transketolase, chloroplast (TK)	75140229	67	5	8	73.0	5.5	*Zm*
	03 Cell [illegible]							
	03.01 Cell growth							
	03 Cell [illegible]							
	03.13 Meiosis							
515	PRD1, involved in meiotic DSB formation	15236434	19	2	2	142.3	5.2	*At*
516	unknown protein, contains Shugoshin domain, C-terminal	145334289	16	2	4	53.7	9.4	*At*
	03 Cell [illegible]							
	03.16 DNA synth/replication							
517	ania-6a type cyclin	20258846	17	2	2	47.6	9.0	*At*
518	At4g05190 (similar to kinesin-like protein)	34849893	18	2	2	89.2	6.5	*At*
519	C2 domain-containing protein	15238792	26	2	6	18.3	6.7	*At*
520	Ca+2-binding EF hand protein	2270994	28	2	10	27.0	6.0	*Gm*
521	calcium-binding EF hand family protein	15218021	32	3	3	109.8	5.0	*At*
522	calcium-binding EF hand family protein	22326598	14	1	5	37.1	6.4	*At*
523	cell division protein FtsH-like protein	21592745	19	2	4	69.4	9.7	*At*
524	centromeric protein-related	18417960	18	2	4	90.0	5.0	*At*
525	CIP1 (COP1-interactive protein 1)	15238181	18	2	2	150.0	4.9	*At*
526	copia-like retrotransposable element	10176701	28	3	4	152.9	6.9	*At*
527	cyclin, putative	18399278	18	2	3	56.9	9.3	*At*
528	EDA10 (embryo sac development arrest 10)	15217579	15	2	1	194.9	5.4	*At*
529	elongation factor-1 alpha	2130560	69	5	40	14.3	8.4	*Gm*
530	En/Spm-like transposon protein	4115353	23	2	2	85.5	6.2	*At*
531	F28J9.5 (similar to replication protein)	6272375	17	2	4	54.6	7.9	*At*
532	kinesin motor protein-related	15226915	18	2	2	120.3	6.1	*At*
533	mitotic cyclin a2-type	857397	17	2	2	53.8	8.2	*Gm*
534	N-ethylmaleimide sensitive fusion protein	1449179	14	1	1	80.9	5.8	*Nt*
535	origin recognition complex subunit 3	15866779	15	2	3	79.0	7.3	*Zm*
536	putative non-LTR retroelement reverse transcriptase	3327392	20	2	2	110.6	9.4	*At*
537	putative non-LTR retroelement reverse transcriptase	3785984	16	2	1	166.5	8.5	*At*
538	putative retroelement	13129465	22	3	1	244.2	8.7	*Os*
539	Ran GTPase binding / chromatin binding	79295482	18	2	7	33.6	5.5	*At*
540	retrotransposon - like protein	2827718	16	2	1	112.8	9.4	*At*
541	RPA 70kDa subunit (replication protein)	33621259	18	2	2	71.5	8.1	*Ps*
542	T1G11.3 (similar to AAA-type ATPase)	2494129	18	2	4	50.5	9.4	*At*
543	T25K16.4 (DEAD-like helicase)	6715634	16	2	0	227.4	6.3	*At*
544	unknown protein, contains domain non-LTR retroelement reverse transcriptase related	15219641	21	2	8	27.4	8.7	*At*
	03 Cell [illegible]							
	03.19 Recombination/repair							
545	ATM2 (myosin 4)	145334819	21	2	1	138.6	8.9	*At*
546	F21D18.26 (DNA repair)	8778524	14	2	3	80.0	5.6	*At*
547	MSH4 (DNA mismatch repair protein)	50083056	19	2	4	89.0	7.3	*At*
548	myosin heavy chain-related	8778462	26	3	2	182.0	8.5	*At*
549	myosin heavy chain-related	9294208	17	2	3	71.3	6.5	*At*
550	myosin heavy chain-related	15225947	17	2	2	54.2	9.5	*At*
551	myosin heavy chain-related	15236005	16	2	8	60.5	5.1	*At*
552	myosin-like protein	18266639	20	2	2	104.6	5.5	*Os*
553	probable myosin heavy chain [imported]	25408221	31	3	3	142.0	4.8	*At*
554	probable myosin-like protein [imported]	42566327	17	2	4	90.6	9.0	*At*
555	Putative myosin heavy chain	15451591	26	3	2	182.9	9.2	*Os*
556	putative myosin heavy chain class VIII A1 protein	40641595	16	2	10	20.8	10.5	*Ta*
557	putative retroelement	20043057	20	2	1	185.5	6.4	*Os*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
558	similar to microtubule-associated protein 7-2	22135818	24	3	6	70.2	6.7	*At*
559	unknown protein	15234295	19	2	1	110.4	6.1	*At*
560	unknown protein, ATPase involved in DNA repair	12324587	24	3	2	104.2	4.8	*At*
561	XIG (myosin-like protein XIG)	42569181	14	2	1	169.1	7.3	*At*
	03 Cell growth/division							
	03.22 Cell cycle							
562	ATG2 (G2p-related protein)	30693537	22	2	6	43.9	6.5	*At*
563	A-type cyclin	849070	16	2	6	53.9	8.1	*Nt*
564	cdc2 (cyclin dependent kinase 2)	3608177	43	3	11	33.9	6.8	*Ps*
565	CDC5 protein (cyclin dependent kinase 5)	18092653	16	2	4	104.4	5.3	*Zm*
566	cyclindependent kinase CDKB	42362295	15	2	10	35.8	9.0	*Gm*
567	kinesin heavy chain	15208451	18	2	7	36.2	6.4	*Zm*
568	kinesin motor protein-related	15231259	20	2	2	119.0	5.7	*At*
569	MEI1 (meiosis defective 1)	145337666	14	2	2	108.0	6.6	*At*
570	Mo25 family protein	15238126	14	2	4	39.6	6.3	*At*
571	SKP1 (kinetochore protein required for cell cycle progression)	51292007	29	2	18	17.5	4.6	*Nt*
572	transitional endoplasmic reticulum ATPase	98962497	460	30	40	89.9	5.1	*Nt*
	03 Cell growth/division							
	03.25 [illegible]							
573	division protein	33436339	21	3	5	88.3	4.8	*At*
574	FtsZ protein (similar to plastid-dividing ring protein)	3116020	33	2	6	44.4	7.7	*Ps*
575	FTSZ2-2 (FtsZ2-2)	15810585	26	2	7	50.3	5.7	*At*
576	POK1 (phragmoplast orienting kinesin 1)	145338627	32	4	2	233.9	5.3	*At*
577	POK2 (phragmoplast orienting kinesis 2)	145338697	20	2	0	315.1	5.1	*At*
578	SMC1 protein (structural maintenance of chromosomes)	27227801	16	2	1	145.4	6.7	*Os*
	03 Cell growth/division							
	03.26 Growth regulators							
	03 Cell growth/division							
	03.30 Seed maturation							
579	34 kDa maturing seed vacuolar thiol protease precursor	1199563	26	2	6	42.8	5.7	*Gm*
580	bHLH family protein (chromosome segregation ATPase)	42567496	31	3	2	175.0	5.5	*At*
581	desiccation-related protein, putative	21593191	32	2	7	34.3	8.8	*At*
582	EEA1 (early embryogenesis associated)	2570402	16	2	8	45.0	7.1	*Hv*
583	hydrophobic seed protein precursor	5019730	64	4	25	12.5	6.7	*Gm*
584	LEA14 (late embryogenesis abundant 14)	15223413	24	2	10	16.5	4.7	*At*
585	seed maturation protein PM22	4585271	37	3	19	16.7	5.2	*Gm*
586	seed maturation protein PM24	6648964	13	1	6	26.8	5.1	*Gm*
587	seed maturation protein PM31	4838149	32	2	17	17.7	6.1	*Gm*
588	seed maturation protein PM34	9622153	88	6	23	31.8	6.6	*Gm*
589	seed maturation protein PM37	5802244	71	5	15	46.3	5.9	*Gm*
	03 Cell growth/division							
	03.99 Other							
	04.01 rRNA synthesis							
590	glutamyl tRNA Reductase	150171033	15	2	6	59.3	8.5	*Nt*
	04.10 tRNA synthesis							
591	aspartate--tRNA ligase - like protein	30688944	32	2	4	62.9	5.9	*At*
592	elongation factor 1A SMV resistance-related protein	50263010	27	2	15	21.0	9.6	*Gm*
593	elongation factor 1-gamma	18958499	16	1	2	47.7	6.3	*Gm*
594	glycyl-tRNA synthetase / glycine--tRNA ligase	15292923	104	7	9	82.0	6.6	*At*
595	putative elongation factor 1-gamma-like	82623387	30	2	5	47.3	5.5	*St*
596	T5E21.11 (aminoacyl-tRNA ligase)	7527726	19	2	3	126.7	6.7	*At*
597	tRNA pseudouridine synthase family protein	15227735	19	2	4	59.0	6.0	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	04.19 mRNA synthesis							
598	elongation factor EF-2	6056373	324	22	28	94.2	5.9	*At*
599	AtRAD3 (ataxia telangiectasia-mutated)	10177961	14	2	0	314.1	6.7	*At*
600	RAD23-like protein	15221013	29	3	7	39.8	4.4	*At*
	04.1901 General TFs							
601	110 kDa 4SNc-Tudor domain protein	21929220	43	3	3	108.2	7.1	*Ps*
602	agenet domain-containing protein	30695584	16	2	3	80.7	6.9	*At*
603	ARF GAP-like zinc finger-containing protein ZIGA3	10441352	15	1	1	51.9	7.8	*At*
604	ATARD2 (acidoreductone dioxygenase)	2244827	35	2	2	108.0	6.3	*At*
605	ATMIN7 (hopm interactor 7)	42565399	28	3	2	194.8	5.4	*At*
606	ATP binding / DNA binding	15229022	14	2	1	215.8	7.2	*At*
607	ATP binding / DNA binding / helicase	145339485	13	1	0	226.3	5.6	*At*
608	auxin response factor 16	19352053	31	4	6	77.1	5.5	*Os*
609	binding / heme binding / protein binding / zinc ion binding	15237191	20	2	1	208.7	5.3	*At*
610	bZIP (transcription factor)	67906424	17	2	3	43.9	9.3	*Nt*
611	bZIP protein BZ3	30691978	16	2	5	34.3	8.4	*At*
612	bzip transcription factor	83853821	23	3	6	45.2	9.8	*Gm*
613	bZIP transcription factor bZIP117	113367212	24	3	9	35.7	5.5	*Gm*
614	bZIP transcription factor bZIP133	113367248	15	2	9	30.1	8.4	*Gm*
615	similar to reverse transcriptase	3377856	20	2	2	126.4	8.2	*At*
616	similar to reverse transcriptase	4325361	15	2	3	101.0	8.0	*At*
617	DNA binding	145332663	15	2	2	128.2	7.8	*At*
618	DNA binding / binding	145326646	17	2	3	91.9	5.8	*At*
619	DNA/RNA binding protein-like	9294614	13	1	1	54.3	5.7	*At*
620	DNA-binding bromodomain-containing protein	15230910	17	2	2	70.2	9.3	*At*
621	DNA-binding bromodomain-containing protein	15221424	15	2	3	84.1	5.2	*At*
622	DNA-binding protein, putative	18395518	18	2	2	70.9	8.2	*At*
623	F22G5.9 (transcription factor)	8778540	16	2	1	169.7	5.6	*At*
624	F27F5.21 (reverse transcriptase, putative)	7767672	19	2	1	115.7	8.8	*At*
625	F3M18.14 (homeobox-1 transcription factor)	6560763	20	2	1	205.7	5.4	*At*
626	F5O11.4 (similar to calmodulin, putative)	8778650	28	3	3	110.3	7.2	*At*
627	F-box family protein	22326994	24	2	4	53.9	5.7	*At*
628	F-box family protein	18398404	20	2	5	47.7	7.8	*At*
629	F-box family protein	15223362	17	2	2	64.1	9.0	*At*
630	F-box family protein	15229834	16	2	7	49.9	6.3	*At*
631	F-box family protein	15219627	16	2	3	52.7	8.3	*At*
632	F-box family protein	15217839	16	2	6	42.9	9.1	*At*
633	F-box family protein	12324759	15	2	8	40.5	9.8	*At*
634	F-box family protein	15232486	14	2	5	45.3	9.0	*At*
635	FES1 (FRIGIDA-ESSENTIAL 1)	30685865	15	2	4	64.4	6.1	*At*
636	FLK (flowering locus KH domain)	15229321	25	2	4	63.4	4.6	*At*
637	GBF3 (G-BOX BINDING FACTOR 3)	15225953	20	2	4	41.1	9.2	*At*
638	glycine-rich protein	2196542	18	2	15	20.6	9.5	*Os*
639	glycine-rich RNA-binding protein	5726567	176	11	83	15.8	6.6	*Gm*
640	glycine-rich RNA-binding protein PsGRBP	1778374	18	1	10	15.1	7.9	*Ps*
641	GTB1 (GLOBAL TRANSCRIPTION FACTOR GROUP B1)	145326656	17	2	2	185.9	5.1	*At*
642	heat shock transcription factor-like protein	7340657	16	2	5	53.8	5.9	*At*
643	mgp1 GTP-binding protein	1020092	15	2	6	23.2	6.0	*Zm*
644	Mutator transposable element-related protein, putative	113205434	24	3	4	70.3	9.3	*Si*
645	contains nascent polypeptide-associated complex (NAC) domain	15229149	37	2	13	23.7	4.4	*At*
646	contains nascent polypeptide-associated complex (NAC) domain	15230476	21	1	6	22.0	4.3	*At*
647	NIM1-like protein 2 (zinc finger)	49182278	17	2	4	64.2	6.0	*La*
648	nucleoid DNA-binding-like protein	21617933	13	1	2	45.5	9.3	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
649	OCL1 homeobox protein	5531484	14	2	3	84.5	5.6	*Zm*
650	PAPA-1-like family protein / zinc finger (HIT type) family protein	12323025	21	2	3	59.6	8.6	*At*
651	PHD finger protein-related	15233015	16	2	2	105.4	8.2	*At*
652	poly(A)-binding protein	83853808	89	7	14	68.5	6.2	*Gm*
653	PRHA (pathogenesis related homeodomain protein A)	15233766	19	2	2	90.7	4.9	*At*
654	PUR ALPHA-1 (purin-rich alpha 1)	18402871	31	2	6	32.2	5.8	*At*
655	putative protein	7340723	21	2	2	80.4	8.4	*At*
656	putative protein	2832678	16	2	4	53.4	6.3	*At*
657	putative transcription factor BTF3-like	82623431	22	2	14	17.5	6.3	*St*
658	putative WRKY-type DNA binding protein	32493108	14	2	5	53.6	6.8	*Gm*
659	regulator of chromosome condensation (RCC1) family protein	15218867	26	3	3	111.1	8.8	*At*
660	reverse transcriptase like protein	2244915	26	3	3	105.7	9.6	*At*
661	ribosomal protein L24-like protein	82400154	19	1	7	19.6	10.7	*St*
662	scc61beta family protein	15239337	16	1	9	10.9	11.6	*At*
663	SET domain protein 105	20977606	18	2	2	75.0	7.7	*Zm*
664	similar to nascent polypeptide associated complex alpha chain	4115918	20	1	6	25.4	4.2	*At*
665	SPT16 (GLOBAL TRANSCRIPTION FACTOR C)	15236899	18	2	1	120.6	5.7	*At*
666	TBP1 (telomeric DNA binding protein 1)	15240725	14	2	5	70.5	6.8	*At*
667	telomerase reverse transcriptase	13625302	16	2	3	143.7	9.5	*Os*
668	transcription factor	22328740	28	2	8	33.7	5.9	*At*
669	transcription factor	2342679	22	2	1	108.3	5.6	*At*
670	transcription factor bZIP38	145652329	18	2	6	48.0	9.6	*Gm*
671	transcription factor jumonji (jmjC) domain-containing protein	30699319	17	2	3	105.5	5.1	*At*
672	transcription factor-related	15218016	28	3	3	195.4	8.7	*At*
673	transcription factor-related	30685450	15	2	2	102.7	6.4	*At*
674	transcription initiation factor IIF (TFIIF-beta) family protein	15222264	23	2	8	29.7	6.5	*At*
675	translational inhibitor protein like	110739384	48	3	16	27.8	9.2	*At*
676	TTR1 (WRKY domain family protein 16)	30694675	16	2	2	155.7	6.0	*At*
677	WRKY60 (WRKY DNA-binding protein 60)	15224660	14	2	4	30.6	9.1	*At*
678	zinc finger (C2H2 type) family protein	15228685	16	2	1	72.8	9.1	*At*
679	zinc finger (C3HC4-type RING finger) family protein	15237223	33	4	0	527.3	5.4	*At*
680	zinc finger (C3HC4-type RING finger) family protein	15232143	19	2	4	104.1	9.6	*At*
681	zinc finger (C3HC4-type RING finger) family protein	15241188	19	2	7	44.4	5.4	*At*
682	zinc finger (C3HC4-type RING finger) family protein	15219544	18	2	9	40.6	9.4	*At*
683	zinc finger (C3HC4-type RING finger) protein-related	30686389	18	2	2	82.5	6.4	*At*
684	zinc finger protein-related	22330435	19	2	1	64.8	4.9	*At*
685	zinc finger transcription factor WRKY1	6689916	17	2	8	44.8	9.0	*Os*
686	zinc ion binding	145359513	18	2	6	48.0	10.1	*At*
	04.1904 Specific TFs							
687	ARF9 (auxin response factor 9)	15233647	20	3	4	72.3	6.5	*At*
688	ATBET9 (bromodomain and extraterminal domain protein 9)	18417335	18	2	3	75.9	4.9	*At*
689	ATBRM/CHR2 (AtBRAHMA)	42571243	27	3	1	245.5	8.9	*At*
690	EIL2 (transcription factor)	30016896	17	2	4	69.9	5.6	*Nt*
691	ethylene responsive protein	33331083	24	3	8	42.2	5.1	*Gm*
692	ETTb (auxin response factor)	85069283	15	2	13	37.2	9.7	*Nt*
693	GH1 protein (auxin inducible)	2388689	13	2	6	36.5	8.7	*Gm*
694	mitochondrial transcription termination factor-related	15220662	19	2	10	47.0	9.5	*At*
695	MS1 (male sterility 1)	15242181	18	2	4	77.0	7.8	*At*
696	myb family transcription factor	15231170	16	2	2	184.1	6.7	*At*
697	MYB transcription factor MYB61	110931660	15	2	8	33.7	10.1	*Gm*
698	MYB85 (myb domain protein 85)	116831385	14	2	16	30.5	5.2	*At*
699	MYB98 (myb domain protein 98)	116831373	15	2	5	50.2	6.1	*At*
700	pinoresinol-lariciresinol reductase, putative	21592830	16	1	3	35.6	6.0	*At*
701	PTAC2 (plastid transcriptionally active 2)	15221411	18	2	2	96.3	5.7	*At*
702	RNA binding protein-like	9293981	19	2	2	107.2	5.7	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	04.1807 Chromatin modification							
703	F24O1.3 (transcription factor)	7940293	19	2	3	104.0	5.8	*At*
704	heme activated protein	77999309	20	1	5	28.3	5.9	*Si*
705	histone H2A-like protein	21593507	23	2	11	14.9	10.3	*At*
706	histone H2B	15226943	23	2	16	16.5	10.0	*At*
707	Histone H4.3 (H4M4)	51315744	55	4	38	11.6	11.7	*Zm*
708	HUB1 (histone mono-ubiquitination 1)	30689877	21	3	4	99.7	6.5	*At*
	04.22 mRNA processing							
709	argonaute 1	119351183	49	4	5	120.9	9.4	*Pt*
710	ATCRS1/CRS1 (Zm chloroplast splicing factor CRS1)	15237295	22	3	4	83.3	9.4	*At*
711	binding (contains Armadillo-like helical domain)	15239621	15	2	2	93.8	6.8	*At*
712	binding (contains Armadillo-like helical)	6862925	17	2	2	94.5	6.7	*At*
713	binding (contains tetratricopeptide domain)	4455367	22	3	1	196.2	5.5	*At*
714	binding (contains tetratricopeptide domain)	15238238	20	1	3	62.4	4.7	*At*
715	binding (RNA stabilisation)	6522552	19	2	2	137.5	5.0	*At*
716	binding (RNA stabilisation)	79329389	17	2	4	68.9	7.1	*At*
717	binding (transcription factor)	15240293	14	2	1	71.0	7.1	*At*
718	CC-NBS-LRR protein	12322948	27	3	2	166.3	7.8	*At*
719	CC-NBS-LRR protein	149786540	14	2	1	144.1	5.7	*Si*
720	chloroplastic group IIA intron splicing facilitator CRS1	75173308	16	2	1	81.4	9.9	*Zm*
721	cinful1 polyprotein	33113968	15	2	2	109.6	5.3	*Zm*
722	CLIP-associating protein (CLASP) -related	42570286	27	3	2	159.0	6.7	*At*
723	copia-like polyprotein	6996255	31	4	4	155.5	7.3	*At*
724	DNA-binding protein	6958202	22	3	8	68.7	7.5	*Ta*
725	DNA-binding protein-related	18403397	15	1	9	20.1	9.4	*At*
726	F17F8.5 (homolgue to reverse transcriptase)	9755374	24	3	5	100.8	9.5	*At*
727	F27F5.19 (retrotransposon gag protein)	7767664	15	2	2	148.1	8.9	*At*
728	F7F22.15 (retrotransposon gag protein)	25354718	17	2	1	163.2	5.7	*At*
729	gag-pol (retrotransposon)	18092337	20	2	0	169.2	8.7	*Zm*
730	gag-pol polyprotein	29423282	14	2	1	179.8	8.4	*Gm*
731	haloacid dehalogenase-like hydrolase family protein	15241564	42	3	12	28.8	6.9	*At*
732	harpin binding protein 1	38679315	47	4	19	28.4	7.9	*Gm*
733	harpin-induced protein-related / HIN1-related	15234663	15	2	8	26.0	10.6	*At*
734	Highly similar to Tal-3 polyprotein	6623973	15	2	1	153.8	8.1	*At*
735	intron maturase, type II family protein	15220638	13	2	4	81.5	9.4	*At*
736	leucine-rich repeat family protein	111183161	16	2	3	65.7	6.4	*La*
737	leucine-rich repeat family protein/protein kinase family protein	15223459	16	2	4	114.6	6.0	*At*
738	maturase-like protein	7406415	18	2	2	80.5	8.4	*At*
739	MRH1 (morphogenesis of root hair 1)	7268658	15	2	3	76.6	6.6	*At*
740	NBS-LRR-like protein	15788510	14	2	7	46.7	8.8	*Hv*
741	NI.0E (contains LRR domain)	4235643	17	2	3	85.7	5.9	*La*
742	pentatricopeptide (PPR) repeat-containing protein	15240444	23	2	3	78.0	6.8	*At*
743	pentatricopeptide (PPR) repeat-containing protein	15242266	23	3	2	139.6	7.6	*At*
744	pentatricopeptide (PPR) repeat-containing protein	15227067	22	3	8	68.2	7.2	*At*
745	pentatricopeptide (PPR) repeat-containing protein	18407365	20	2	5	45.8	8.8	*At*
746	pentatricopeptide (PPR) repeat-containing protein	15222036	19	2	2	55.1	9.4	*At*
747	pentatricopeptide (PPR) repeat-containing protein	15233142	19	2	2	95.1	5.7	*At*
748	pentatricopeptide (PPR) repeat-containing protein	15228653	16	2	3	67.3	6.6	*At*
749	pentatricopeptide (PPR) repeat-containing protein	15240032	16	2	2	116.9	8.4	*At*
750	pentatricopeptide (PPR) repeat-containing protein	15231338	16	2	5	98.3	6.5	*At*
751	pentatricopeptide (PPR) repeat-containing protein	15233050	16	2	3	87.7	7.0	*At*
752	pentatricopeptide (PPR) repeat-containing protein	15222492	15	2	3	77.4	8.2	*At*
753	pentatricopeptide (PPR) repeat-containing protein	15223763	15	2	5	61.7	8.8	*At*
754	pentatricopeptide (PPR) repeat-containing protein	30686506	14	2	2	100.8	8.2	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
755	pentatricopeptide (PPR) repeat-containing protein	15221304	14	2	2	98.7	5.9	*At*
756	pentatricopeptide repeat-containing protein	91806015	15	2	5	40.3	7.9	*At*
757	polyprotein	10177485	19	2	0	159.5	9.3	*At*
758	polyprotein	4235644	18	2	1	175.6	9.0	*La*
759	polyprotein	18657020	16	2	1	205.8	8.8	*Os*
760	polyprotein	27764548	16	2	2	176.4	8.1	*Gm*
761	putative Athila retroelement ORF1 protein	4417310	17	2	7	56.5	5.9	*At*
762	putative Athila retroelement ORF1 protein	20197605	16	2	3	115.9	5.5	*At*
763	putative athila-like protein	7267490	16	2	2	65.9	6.4	*At*
764	putative gag-pol polyprotein	13435243	24	3	2	203.7	7.6	*Os*
765	putative gag-pol polyprotein	18855003	19	2	2	128.5	8.7	*Os*
766	putative gag-pol polyprotein	23928449	19	2	1	193.8	8.1	*Zm*
767	putative gag-pol polyprotein, 3'-partial	14018085	16	2	1	121.2	7.1	*Os*
768	putative gag-pol precursor	33113963	15	2	1	207.2	9.3	*Zm*
769	Putative mutator-like transposase	15451606	26	2	1	269.6	7.0	*Os*
770	putative polyprotein	14018106	29	4	2	169.3	8.6	*Os*
771	putative polyprotein	18568267	20	2	0	307.0	9.4	*Zm*
772	putative polyprotein	18657016	20	2	5	76.3	8.6	*Os*
773	putative polyprotein	16924110	17	2	10	32.4	9.2	*Os*
774	putative polyprotein	16519476	17	2	1	159.5	6.6	*Os*
775	putative protein	4972086	18	2	3	104.6	7.1	*At*
776	putative protein	7529264	18	2	9	31.3	5.6	*At*
777	putative protein	3269282	17	2	3	148.8	8.9	*At*
778	putative protein	7288029	13	2	2	48.5	6.6	*At*
779	putative protein (possibly fragment)	4753646	18	2	2	87.9	5.6	*At*
780	putative retroelement	16924051	16	2	4	48.1	6.2	*Os*
781	putative retroelement	15217240	15	2	1	188.8	9.1	*Os*
782	putative retroelement	19881671	14	2	2	84.4	6.8	*Os*
783	putative retroelement pol polyprotein	4388818	32	4	4	153.0	8.5	*At*
784	putative retroelement pol polyprotein	7523670	22	3	3	131.9	8.9	*At*
785	putative retroelement pol polyprotein	13129455	17	2	5	99.7	7.0	*Os*
786	putative retroelement pol polyprotein	4432797	16	2	1	98.7	9.8	*At*
787	putative RIRE2 orf3	45550145	17	2	2	90.8	5.0	*Zm*
788	putative transposase	3283026	16	2	5	84.2	6.5	*At*
789	putative transposase	16924109	15	2	4	77.4	8.6	*Os*
790	retroelement pol polyprotein-like	9759493	26	3	3	126.4	9.0	*At*
791	retroelement pol polyprotein-like	10177643	20	2	1	212.1	6.5	*At*
792	retroelement pol polyprotein-like	9294132	16	2	2	98.7	9.2	*At*
793	retroelement pol polyprotein-like	8777581	15	2	1	121.2	8.8	*At*
794	splicing factor SC35	9843653	25	3	9	35.2	11.5	*At*
795	splicing factor, putative	18403722	16	1	14	10.2	5.3	*At*
796	SRM102 (SR-rich pre-mRNA splicing activator)	9843651	16	2	1	102.4	11.7	*At*
797	T12C24.22 (contains helicase domian)	9502386	17	2	3	138.9	8.7	*At*
798	T14P8.20 (ribonucleoprotein)	3193300	18	2	2	98.5	5.8	*At*
799	transposable element activator uncharacterized	140171	18	2	10	23.0	9.4	*Zm*
800	transposase	7673677	18	2	2	81.9	8.7	*Zm*
801	unknown protein, contains DWNN domain	145340337	15	2	3	91.3	8.9	*At*
	04.31 RNA transport							
	06 Protein synthesis							
	06.01 Ribosomal proteins							
802	40S ribosomal protein S10 (RPS10C)	8953720	20	2	12	19.8	9.6	*At*
803	40S ribosomal protein S10-like	81074037	51	3	15	19.8	9.8	*Si*
804	40S ribosomal protein S12 (RPS12A)	15218373	17	1	6	15.4	5.4	*At*
805	40S ribosomal protein S13 (RPS13A)	18411716	16	1	7	17.1	10.4	*At*
806	40S ribosomal protein S14 (Clone MCH2)	131773	43	3	22	16.3	10.6	*Zm*
807	40S ribosomal protein S15A (RPS15aC)	116831147	18	2	21	15.4	9.3	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
808	40S ribosomal protein S17-like protein	76573345	16	1	7	16.3	10.0	*St*
809	40S ribosomal protein S19 (RPS19H)	15242322	47	4	30	15.8	10.1	*At*
810	40S ribosomal protein S2	21553851	55	3	9	30.9	10.3	*At*
811	40S ribosomal protein S23 (RPS23A)	23928437	16	1	3	34.0	9.9	*Zm*
812	40S ribosomal protein S25	33590374	15	1	10	10.5	10.8	*Gm*
813	40S ribosomal protein S25 (RPS25B)	15226590	15	1	12	12.1	10.7	*At*
814	40S ribosomal protein S3 (RPS3A)	82623397	73	6	26	26.4	9.7	*St*
815	40S ribosomal protein S3a-like protein	82400124	17	2	9	29.6	9.9	*St*
816	40S ribosomal protein S9 (RPS9C)	15242498	48	4	17	23.2	10.3	*At*
817	40S ribosomal S4 protein	22138108	129	10	35	30.0	10.3	*Gm*
818	4-coumarate:CoA ligase isoenzyme 2	4038975	27	2	5	60.2	6.3	*Gm*
819	60S acidic ribosomal protein P0 (RPP0B)	15232603	72	5	13	34.1	5.0	*At*
820	60S ribosomal protein L10 (RPL10C)	18408550	64	5	24	24.9	10.6	*At*
821	60S ribosomal protein L12 (RPL12B)	15231814	39	2	16	18.0	9.1	*At*
822	60S ribosomal protein L18 (RPL18C)	15241061	32	2	13	21.0	11.0	*At*
823	60S ribosomal protein L19 (RPL19C)	15235290	20	2	10	24.2	11.4	*At*
824	60S ribosomal protein L22-2 (RPL22B)	15230008	28	3	27	14.0	9.6	*At*
825	60S ribosomal protein L23 (RPL23B)	2341028	104	7	44	17.0	10.1	*At*
826	60S ribosomal protein L30	6094049	15	1	14	12.5	9.6	*Zm*
827	60S ribosomal protein L34 (RPL34A)	15223377	34	3	22	13.7	11.6	*At*
828	60S ribosomal protein L6 (RPL6A)	15221798	26	2	8	26.2	10.1	*At*
829	60S ribosomal protein L6, putative	21593910	22	2	7	26.1	10.1	*At*
830	60S ribosomal protein L7A-like	82623429	15	1	3	29.4	10.3	*St*
831	ATRPS5B (RIBOSOMAL PROTEIN 5B)	15228111	87	6	24	23.0	9.7	*At*
832	hypothetical protein F16M2.40	11277801	25	3	15	32.1	9.4	*At*
833	KH domain-containing protein	15241136	24	2	8	34.0	7.9	*At*
834	KH domain-containing protein	15225229	15	1	1	64.6	5.3	*At*
835	P40-like protein	15217294	96	7	30	33.1	4.9	*Os*
836	putative 60S ribosomal protein	19423912	18	2	9	23.4	10.4	*At*
837	ribosomal protein L11, cytosolic	7440684	60	4	21	21.1	9.9	*At*
838	ribosomal protein L13 family protein	30678423	21	2	4	23.4	10.4	*At*
839	ribosomal protein L2	12054507	19	2	11	28.0	10.5	*Gm*
840	ribosomal protein L2	17644112	18	2	7	21.9	10.7	*Le*
841	ribosomal protein L25-like protein	76573339	28	2	19	17.2	10.3	*St*
842	ribosomal protein L30	57471710	15	1	10	12.3	9.4	*Ta*
843	ribosomal protein L9	6015604	60	4	21	22.0	9.2	*Ps*
844	ribosomal protein S11	91214174	15	2	19	15.0	12.1	*Gm*
845	ribosomal protein S26	5706704	14	1	6	14.9	10.9	*Ps*
846	ribosomal protein S27	6850878	45	3	29	9.5	9.1	*At*
847	ribosomal protein S3	91214179	28	3	12	24.7	9.9	*Gm*
848	ribosomal protein S3	57013937	14	2	3	65.0	10.2	*Nt*
849	ribosomal protein S5 family protein	18408151	16	2	3	60.2	7.1	*At*
850	RPS1 (ribosomal protein S1)	30692346	18	1	2	45.1	5.1	*At*
851	RPS15A (RIBOSOMAL PROTEIN S15A)	76573307	59	4	30	14.8	9.9	*St*
852	rps3 (ribosomal protein S3)	81687782	22	2	8	64.2	10.4	*Ta*
853	RPS6 (ribosomal protein S6)	15236042	26	2	10	28.4	10.6	*At*
854	S1 RNA-binding domain-containing protein	15222076	18	2	2	85.7	5.3	*At*
855	S18.A ribosomal protein	13877525	27	2	13	17.6	10.5	*At*
856	S28 ribosomal protein	32400865	28	2	22	9.8	11.3	*Ta*
857	U2A' (U2 small nuclear ribonucleoprotein A)	15218274	59	4	21	28.0	5.8	*At*
	08 Protein synthesis							
	08.04 Translation factors							
858	AT5g23690/MQM1_4 (tRNA nucleotidyltransferase)	18377862	14	2	2	59.6	7.6	*At*
859	ATFIP1[V], RNA binding	15242929	14	2	1	133.0	5.3	*At*
860	ATP-dependent helicase	6522577	15	2	1	151.2	6.0	*At*
861	resistance to Pseudomonas syringae pv maculicola interactor 1	15242217	34	3	7	50.3	5.7	*At*
862	CAA30377.1 protein (contains Armadillo-like helical domain)	5777619	14	2	6	70.3	5.7	*Os*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
863	CETS1 (phosphatidylethanolamine binding protein)	99079228	33	2	12	19.0	9.1	*Gm*
864	CRP (CRYPTIC PRECOCIOUS)	79457834	27	3	2	237.5	9.0	*At*
865	DNA polymerase	76880150	22	3	4	128.6	8.6	*Nt*
866	EIF2 BETA (embryo defective 1401)	15242100	23	2	7	30.7	6.8	*At*
867	EIF4E (eukaryotic translation initiation factor 4E)	15236735	17	1	5	26.5	5.0	*At*
868	eIF4-gamma/eIF5/eIF2-epsilon domain-containing protein	9759036	30	2	5	49.3	5.7	*At*
869	EIF-5A (eukaryotic translation initiation factor 5A-1)	15223002	56	5	25	17.4	5.4	*At*
870	elongation factor 1-alpha-like protein	83283977	60	4	8	49.3	9.2	*St*
871	EMB1220 (embryo defective 1220)	3249066	16	2	4	55.5	5.7	*At*
872	EMB1353 (embryo defective 1353)	15234615	15	2	7	35.9	9.3	*At*
873	EMB1865 (embryo defective 1865)	15229636	14	2	3	96.0	5.8	*At*
874	EMB1974 (embryo defective 1974)	145338215	17	2	2	77.9	5.1	*At*
875	EMB2107 (embryo defective 2107)	42573323	15	2	6	50.9	7.6	*At*
876	EMB2221 (embryo defective 2221)	30695804	16	1	1	119.9	5.0	*At*
877	EMB2284/POL2A/TIL1 (EMBRYO DEFECTIVE 2284)	15223158	26	3	1	261.3	6.4	*At*
878	EMB2410 (embryo defective 2410)	42569320	30	3	1	238.0	7.6	*At*
879	EMB2719 (EMBRYO DEFECTIVE 2719)	8778979	17	2	3	59.1	8.5	*At*
880	EMB2755 (embryo defective 2755)	15241897	40	3	4	63.8	5.4	*At*
881	ERF1-3 (EUKARYOTIC RELEASE FACTOR 1-3)	42565216	26	2	4	49.0	5.4	*At*
882	ethylene-responsive RNA helicase	15231074	57	4	7	69.2	7.7	*At*
883	eukaryotic initiation factor 3I1 subunit	12407664	62	4	14	36.5	6.9	*At*
884	eukaryotic translation initiation factor 2B family protein/ eIF-2B family protein	18396170	33	2	9	39.6	5.9	*At*
885	eukaryotic translation initiation factor eIF4E	51599169	17	1	4	25.2	5.2	*Nt*
886	FH3.2, RNA binding	8778484	14	1	1	64.0	8.4	*At*
887	helicase-like protein	30694618	19	2	1	172.2	9.0	*At*
888	hypothetical EIF-2-Alpha	4588003	19	1	4	41.6	5.1	*At*
889	initiation factor eIF-4 gamma like	110741875	37	4	3	187.9	7.7	*At*
890	methyl-CpG-binding domain 13	22531130	16	2	3	82.6	9.4	*At*
891	NRPD2a (nuclear RNA polymerase D 2A)	79416709	18	2	1	132.7	8.6	*At*
892	oligouridylate-binding protein, putative	15231783	17	1	2	47.1	7.2	*At*
893	PAB2 (poly(A)-binding protein 2)	19347816	59	4	8	67.1	8.2	*At*
894	PIE1 (photoperiod-independent early flowering 1)	42564102	14	2	1	234.0	5.2	*At*
895	POLGAMMA2 (polymerase gamma 2)	12321800	16	2	3	119.4	6.7	*At*
896	poly(A) polymerase	2623246	17	2	4	50.2	5.3	*Ps*
897	polypyrimidine tract-binding protein, putative	22329999	26	2	6	48.2	6.4	*At*
898	probable protein ATP-dependent DNA helicase RecQ	25403040	18	2	2	104.2	8.9	*At*
899	chloroplast translation elongation factor EF-Tu precursor	23397095	32	3	11	51.7	5.8	*At*
900	putative DNA2-NAM7 helicase family protein	16924040	14	2	1	163.5	6.6	*Os*
901	putative protein	5262156	16	2	0	262.6	6.1	*At*
902	putative RNA-dependent RNA-polymerase	142942402	17	2	2	127.2	8.6	*St*
903	RDR1 (RNA-dependent RNA polymerase 1)	15223906	14	2	1	126.2	7.9	*At*
904	RDR6 (RNA-dependent RNA polymerase 6)	15229153	18	2	1	136.9	6.8	*At*
905	RNA helicase	3776005	28	2	3	51.1	5.7	*At*
906	RNA helicase	15231574	23	3	1	134.2	6.4	*At*
907	RNA helicase like protein	30683736	17	2	4	89.4	5.5	*At*
908	RNA polymerase beta" chain	11467184	27	3	2	176.1	6.2	*Zm*
909	RNA polymerase beta I chain	82754620	16	2	3	80.3	9.0	*St*
910	RNA polymerase beta subunit	108773122	16	2	2	120.6	8.5	*St*
911	RNA polymerase subunit	21592304	13	2	2	41.8	5.5	*At*
912	RNA recognition motif (RRM)-containing protein	42562492	19	1	4	26.2	7.7	*At*
913	RNA recognition motif (RRM)-containing protein	15242719	15	1	2	42.4	6.0	*At*
914	RNA recognition motif (RRM)-containing protein	15231193	15	2	2	108.4	5.9	*At*
915	RNA-directed DNA polymerase-like protein	4538901	25	3	2	143.9	9.2	*At*
916	RPE (embryo defective 2728)	15240250	15	1	5	30.0	8.2	*At*
917	rpoB	15778187	21	3	2	121.2	8.8	*Gm*
918	rpoC1 (RNA polymerase beta' subunit)	15778188	29	4	5	79.1	9.2	*Gm*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
919	RSZ32, nucleic acid binding	30693839	14	2	5	31.8	10.2	*At*
920	SNF2P, DNA repair	23193481	14	2	4	99.1	6.3	*Hv*
921	SYD (SPLAYED), DNA/ RNA unwinding	30683830	15	2	1	389.9	4.8	*At*
922	tobacco mosaic virus helicase domain-binding protein	16518974	19	2	4	59.2	9.3	*Nt*
923	translation elongation factor-TU	2546952	86	6	33	27.2	5.1	*Gm*
924	translation initiation factor	3024657	33	2	29	12.7	8.9	*Zm*
925	translation initiation factor	6143897	27	3	3	127.0	5.8	*At*
926	translational elongation factor 1 subunit Bbeta	38232568	78	5	17	25.3	4.4	*Ps*
927	translational elongation factor EF-TuM	11181616	161	10	24	48.5	6.0	*Zm*
928	U1 snRNP-specific protein, U1A	1050840	17	2	18	28.0	9.4	*St*
929	UBA2A, RNA binding	14596195	16	1	2	51.4	5.0	*At*
	06 Protein synthesis							
	06.07 Translation control							
	06 Protein synthesis							
	06.10 tRNA synthases							
	06 Protein synthesis							
	06.99 Others							
930	calnexin	56744207	40	4	8	62.0	4.8	*Gm*
931	calreticulin-1	117165712	408	27	68	48.2	4.4	*Gm*
932	uroporphyrinogen decarboxylase, chloroplast precursor	6014938	18	2	3	43.4	8.3	*Zm*
	06 Protein destination and storage							
	06.01 Folding and stability							
933	15.7 kDa class I-related small heat shock protein-like	15240308	22	2	9	15.7	7.9	*At*
934	70-kD heat shock protein	21481	51	3	46	9.2	6.1	*St*
935	allergen Cry m Bd 28K	12697782	136	9	21	52.6	5.7	*Gm*
936	AT4g28520/F20O9_210 (cruciferin 3)	19699273	16	2	4	58.2	6.6	*At*
937	ATCOAE, ATP binding, dephospho CoA kinase	15225886	14	1	5	25.7	9.5	*At*
938	ATP binding	6466963	73	6	6	91.0	5.3	*At*
939	ATP binding / protein binding / transmembrane receptor	5903073	16	2	2	125.8	6.0	*At*
940	ATP binding, DNA binding	15236340	17	2	1	240.1	5.6	*At*
941	ATP binding, heat shock protein, putative	15221072	15	2	2	86.3	5.5	*At*
942	ATPDIL1-3 (PDI-LIKE 1-3)	22331799	28	2	3	64.2	4.7	*At*
943	ATPDIL1-5 (PDI-LIKE 1-5) (thiol-disulfide exchange intermediat	12323134	19	2	3	61.2	5.0	*At*
944	BiP (binding immunoglobulin protein)	62433284	564	35	52	73.6	5.1	*Gm*
945	CLPB-MCLPB4/HSP98.7 (heat shock protein 98.7)	18400735	24	3	6	108.7	6.5	*At*
946	CONSTANS interacting protein 3	45544871	20	1	14	12.1	6.6	*La*
947	copper chaperone	30039180	15	1	16	8.5	6.2	*La*
948	CPHSC70-1 (chloroplast heat shock protein 70-1)	15233779	185	13	17	76.5	5.1	*At*
949	Cu/Zn-superoxide dismutase copper chaperone precursor	12711645	42	3	17	32.5	5.6	*Gm*
950	cyclophilin	17981611	175	10	71	18.2	8.7	*Gm*
951	cyclophilin-like protein	37788310	29	3	5	69.9	10.4	*Ta*
952	DNAJ heat shock N-terminal domain-containing protein	15234962	20	2	12	38.5	7.0	*At*
953	DNAJ heat shock N-terminal domain-containing protein	15231993	16	2	4	62.6	9.5	*At*
954	DnaJ-like protein	6782421	15	2	6	46.7	6.1	*La*
955	dnaJ-like protein	2230757	14	1	3	48.2	5.3	*At*
956	heat shock protein 26	453670	17	2	15	26.4	7.9	*Zm*
957	heat shock protein 70-3	38325815	408	25	47	71.0	5.1	*Nt*
958	heat shock protein 90	110083391	202	14	21	80.2	4.9	*Nt*
959	heat shock protein-related	15242850	21	2	3	108.7	8.3	*At*
960	heat shock transcription factor 29	671866	24	3	19	32.4	5.4	*Gm*
961	HSP91 (Heat shock protein 91)	15220026	46	3	5	91.8	5.2	*At*
962	legA class precursor	4218520	16	2	4	58.8	6.2	*Ps*
963	MTHSC70-1 (mitochondrial heat shock protein 70-1)	30691626	178	12	19	73.1	5.5	*At*
964	NTRA (NADPH-dependent thioredoxin reductase 2)	79557518	33	2	8	40.0	6.3	*At*
965	peptidyl-prolyl cis-trans isomerase (PPIase) (Rotamase)	118104	22	1	8	18.3	8.9	*Zm*
966	peptidyl-prolyl cis-trans isomerase / cyclophilin (CYP2) / rotamase	15226467	21	2	22	18.5	8.3	*At*
967	peptidyl-prolyl cis-trans isomerase cyclophilin-type family protein	30678211	43	3	16	17.5	8.4	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
968	peptidyl-prolyl cis-trans isomerase, putative / cyclophilin, putative	15224944	17	1	5	21.5	8.5	*At*
969	peptidylprolyl isomerase	133741925	43	3	13	25.9	9.4	*Ta*
970	peptidylprolyl isomerase	9294180	25	2	3	61.8	5.2	*At*
971	PLA IIIA/PLP7 (patatin-like protein 7)	15233136	16	2	5	53.1	8.6	*At*
972	protein disulfide isomerase	59861261	58	4	7	56.9	5.0	*Zm*
973	protein disulfide isomerase	59861273	17	1	2	46.9	5.5	*Zm*
974	ROC2 (rotamase CyP 2)	15228814	34	3	27	18.9	7.7	*At*
975	ROC7 (rotamase CyP 7)	15237739	28	2	8	22.0	9.1	*At*
976	SGT1-2, homolog to co-chaperone	126544456	18	2	3	41.3	5.0	*Ta*
977	similar to heat shock protein binding	15225676	17	2	7	52.0	5.4	*At*
978	superoxide dismutase [Cu-Zn] 4A	134597	69	4	26	15.1	5.7	*Zm*
979	trigger factor-like protein	9758119	16	2	5	65.2	5.2	*At*
	06 Protein destination and storage							
	06.04 Targeting							
980	putative coated vesicle membrane protein	21595553	22	1	10	24.3	5.9	*At*
	06 Protein destination and storage							
	06.07 Modification							
981	protein disulfide isomerase-like protein	49257111	242	15	46	40.4	5.7	*Gm*
	06 Protein destination and storage							
	06.10 Complex assembly							
982	chaperonin 21 precursor	7331143	38	3	11	26.6	6.9	*La*
983	chaperonin hsp60	16221	26	2	4	61.4	5.7	*At*
984	chaperonin, putative	15229866	97	6	12	59.8	6.0	*At*
985	chaperonin, putative	15240317	48	4	8	60.3	5.6	*At*
986	chaperonin, putative	15242093	27	2	7	57.3	5.6	*At*
987	chaperonin, putative	18396719	15	2	5	58.9	5.3	*At*
988	CPN20 (chaperonin 20)	15242045	33	3	8	26.8	8.9	*At*
989	CPN60A (chloroplast / 60 kDa chaperonin alpha subunit)	21554572	95	7	11	62.1	5.0	*At*
990	TCP-1 chaperonin-like protein	21536971	45	3	6	59.0	5.8	*At*
	06 Protein destination and storage							
	06.13 Proteolysis							
991	20S proteasome beta subunit PBH2	20260224	95	6	11	29.6	6.7	*At*
992	26S proteasome beta subunit	49175785	22	2	29	6.0	10.2	*Pa*
993	26S proteasome non-ATPase regulatory subunit	21592398	146	9	42	34.4	6.4	*At*
994	26S proteasome subunit 4-like	77745479	72	5	14	49.6	6.1	*St*
995	26S proteasome subunit RPN12	15217661	35	2	10	30.7	4.8	*At*
996	26S proteasome subunit RPN1b	32700012	38	3	5	98.0	5.1	*At*
997	AHSP (asprase)	79482708	18	2	1	244.8	6.8	*At*
998	aspartic proteinase 1	15186732	93	7	18	55.5	6.3	*Gm*
999	aspartic proteinase 2	15425751	58	4	12	55.5	6.3	*Gm*
1000	ATHMOV34 (asymmetri leaves enhancer 3)	77745499	84	5	18	34.8	5.9	*St*
1001	ATP-dependent Clp protease ClpH protein-related	145323770	17	2	3	107.8	8.1	*At*
1002	ATPREP1/ATZNMP (presequence protease 1)	22331173	35	3	4	121.0	5.5	*At*
1003	cathepsin B-like cysteine protease, putative	18378947	15	1	3	40.0	6.5	*At*
1004	ClpC (Clp protease ATP binding subunit)	2921158	38	3	4	103.5	6.3	*At*
1005	CLPP5 (nuclear encoded CLP protease 1)	18378982	27	2	9	32.4	8.4	*At*
1006	CLPX (Clp protease regulatory subunit X)	18423503	22	2	3	62.0	7.6	*At*
1007	CUL2 (cullin 2)	22329305	18	2	2	86.0	7.3	*At*
1008	cysteine protease TDI-65	5726641	14	1	3	51.1	5.9	*La*
1009	cysteine proteinase	479060	27	2	8	41.6	6.0	*Gm*
1010	cysteine proteinase	31559530	25	2	9	40.1	6.1	*Gm*
1011	cysteine proteinase inhibitor	1944319	82	5	29	27.6	7.3	*Gm*
1012	cysteine proteinase inhibitor	1277164	56	4	36	10.3	5.9	*Gm*
1013	cysteine proteinase inhibitor	1277168	25	2	27	11.1	5.8	*Gm*
1014	cysteine-type peptidase	15241982	17	2	6	56.1	4.7	*At*
1015	cytosol aminopeptidase family protein	15235763	84	5	9	61.3	6.6	*At*
1016	DEGP7 (DEGP protease 7)	30678834	15	2	1	119.9	5.7	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1017	DJ-1 family protein / protease-related	15232958	14	1	2	41.6	5.2	*At*
1018	DNA-damage inducible protein DDI1-like	21537297	38	3	7	45.4	4.8	*At*
1019	ECR1 (E1 C-terminal related 1)	18419850	19	1	2	50.5	5.5	*At*
1020	FIE22.3 (ubiquitin carboxyl-terminal hydrolase 1)	6686411	17	1	6	41.4	5.2	*At*
1021	FTSH3 (FtsH protease 3)	30684118	33	3	3	89.4	6.8	*At*
1022	Fstb-like protease	50892959	22	2	2	86.9	6.2	*Ps*
1023	FtsH-like protein Pftf precursor	4325041	39	3	6	74.4	6.0	*Nt*
1024	ICS1 (isochorismate synthase 1) (ubiquinone biosynthesis)	42572105	20	2	4	69.0	6.1	*At*
1025	leucine-rich repeat family protein	15220080	18	2	3	110.2	6.0	*At*
1026	mitochondrial processing peptidase	587564	14	1	2	59.3	6.2	*Si*
1027	mitochondrial processing peptidase alpha subunit, putative	21594004	23	2	3	54.5	5.9	*At*
1028	MMZ4 (MMS zwei homologue 4)	18409633	44	3	22	16.5	6.2	*At*
1029	MPPALPHA (mitochondrial processing peptidase alpha subunit)	14334534	30	3	5	54.1	6.2	*At*
1030	multicatalytic endopeptidase complex alpha subunit-like	20260140	96	6	25	27.3	5.4	*At*
1031	PAB1 (20S proteasome alpha subunit B1)	15219317	113	6	26	25.7	5.5	*At*
1032	PAE1 (20S proteasome alpha subunit E1)	15220961	125	8	46	25.9	4.7	*At*
1033	PAG1 (20S proteasome alpha subunit G1)	15225839	37	3	10	27.4	5.9	*At*
1034	PBA1 (20S proteasome beta subunit A 1)	79325892	81	5	18	25.3	5.3	*At*
1035	PBC1 (20S proteasome beta subunit C1)	21553663	58	4	20	22.8	5.3	*At*
1036	PBD1 (proteasome subunit PRCBI)	15228805	60	4	20	22.5	6.0	*At*
1037	PBE1 (20S proteasome beta subunit E1)	14594931	96	6	45	18.6	9.2	*Nt*
1038	PDG1 (20S proteasome beta subunit G1)	15223537	21	1	7	27.7	6.1	*At*
1039	pepsin A	79507883	16	1	2	48.7	9.7	*At*
1040	peptidase	12324166	18	2	8	34.9	9.0	*At*
1041	peptidase M20/M25/M40 family protein	42566909	25	2	9	48.2	5.2	*At*
1042	PEX4 (peroxin 4)	18420949	24	2	17	17.7	8.4	*At*
1043	phosphatidylinositol 3- and 4-kinase family protein	18407090	14	2	4	62.6	5.8	*At*
1044	polyubiquitin	3452083	101	6	41	12.8	9.7	*Gm*
1045	proteasome alpha subunit-like protein	76160982	111	7	33	28.1	5.4	*Si*
1046	proteasome-like protein alpha subunit	77999287	125	8	36	27.1	7.0	*Si*
1047	proteosome subunit	600387	32	2	10	25.3	7.8	*At*
1048	Putative 26S proteasome ATPase subunit	6056389	14	2	5	50.3	5.2	*At*
1049	putative alpha7 proteasome subunit	14594925	98	7	29	27.2	6.1	*Nt*
1050	putative beta7 proteasome subunit	14594935	41	3	23	14.7	8.2	*Nt*
1051	putative ubiquitin protein ligase	13174246	16	2	3	84.5	6.4	*Os*
1052	RPT1A (regulatory particle triple-A 1A)	15220930	131	9	26	47.8	6.3	*At*
1053	RPT5B (26S proteasome AAA-ATPase subunit RPT5B)	15217431	293	18	47	47.0	4.9	*At*
1054	RUB1 (related to ubiquitin 1)	30692436	43	3	17	17.4	5.8	*At*
1055	RUB1-conjugating enzyme-like protein	76573335	26	2	8	21.0	8.3	*Si*
1056	SFC (scarface, vascular network defective 3)	3914005	17	2	2	97.7	7.7	*Zm*
1057	subtilase family protein	42567017	25	3	5	78.5	9.4	*At*
1058	subtilase family protein	18416719	25	3	3	82.9	6.2	*At*
1059	subtilase family protein	18423316	14	1	1	85.0	9.4	*At*
1060	subtilisin-like protease	33621210	118	8	14	83.2	9.0	*Gm*
1061	subtilisin-like protease	86439745	15	2	3	76.5	9.3	*Ta*
1062	subtilisin-type protease precursor	11611651	397	25	40	82.7	6.9	*Gm*
1063	SUM2 (small ubiquitin-like modifier 2)	15240471	43	3	25	11.7	5.4	*At*
1064	thiol protease isoform B	1619903	40	3	11	35.0	7.6	*Gm*
1065	thiolprotease	3980198	18	1	2	51.3	6.1	*Ps*
1066	TPP2 (tripeptidyl peptidase II)	5262775	20	2	1	154.2	6.1	*At*
1067	UBC12 (ubiquitin-conjugating enzyme 12)	18398208	14	1	6	16.7	7.7	*At*
1068	UBC30 (ubiquitin-protein ligase)	18423829	16	1	7	16.5	6.8	*At*
1069	UBC36, ubiquitin-protein ligase	18394416	142	9	75	17.2	6.7	*At*
1070	UBC9 (ubiquitin conjugating enzyme 9)	18417097	30	2	19	20.2	7.0	*At*
1071	ubiquitin	1762935	13	1	23	8.7	8.1	*Nt*
1072	ubiquitin activating enzyme E1	1808656	56	4	5	120.3	5.4	*Nt*
1073	ubiquitin carboxyl-terminal hydrolase	42566353	14	2	6	46.6	9.5	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1074	ubiquitin carboxyl-terminal hydrolase-related	15242114	15	2	2	132.2	5.9	*At*
1075	ubiquitin carboxyl-terminal hydrolase-related	42572001	15	2	1	130.9	5.5	*At*
1076	ubiquitin conjugating enzyme-like	82623381	25	2	11	21.4	5.0	*St*
1077	ubiquitin family protein	15232924	25	2	5	44.2	4.8	*At*
1078	ubiquitin fusion-degradation protein-like	76160972	63	5	16	35.5	6.2	*St*
1079	ubiquitin isopeptidase T	11994150	14	1	1	88.4	5.0	*At*
1080	ubiquitin-conjugating enzyme family protein-like protein	76160962	92	6	34	16.6	6.2	*St*
1081	ubiquitin-conjugation enzyme	22597164	46	3	25	16.4	7.7	*Gm*
1082	ubiquitin-specific protease 6	11993465	14	1	1	53.7	5.8	*At*
1083	Ulp1 protease family protein	15232756	23	3	4	94.1	5.5	*At*
1084	Ulp1 protease family protein	15242433	18	2	2	105.8	8.6	*At*
1085	Ulp1 protease family protein	15229144	14	2	2	146.5	5.3	*At*
1086	Ulp1 protease family protein	15234224	14	2	2	82.1	5.4	*At*
1087	unknown protein, contains peptidase_C12 domain	15238875	21	2	5	49.7	8.2	*At*
1088	unknown protein, with PPPDE putative peptidase domain	15231383	17	1	5	28.8	5.6	*At*
1089	UPL2 (ubiquitin-protein ligase 2)	15223117	14	2	0	403.6	4.8	*At*
1090	urease	14599161	20	2	2	90.7	5.8	*Gm*
1091	urease	14599413	16	2	2	89.8	5.6	*St*
	06 Protein destination and storage							
	06.20 Storage proteins							
1092	albumin 1	32328738	23	2	30	11.0	5.5	*Gm*
1093	ATPGP1 (P-glycoprotein 1)	15228052	19	2	3	140.6	8.5	*At*
1094	beta-conglycinin alpha subunit	14245736	486	31	49	70.3	5.1	*Gm*
1095	fibrillin-precursor like protein	18377868	25	2	8	33.7	5.7	*At*
1096	glycinin	4249568	270	17	39	63.8	5.2	*Gm*
1097	glycinin G3 precursor	18609	362	21	64	54.3	5.6	*Gm*
1098	glycoprotein-like protein	76161008	37	3	14	15.4	10.4	*St*
1099	hydroxyproline-rich glycoprotein family protein	9454580	16	2	1	234.3	7.8	*At*
1100	MDR-like p-glycoprotein	26449438	16	2	1	136.9	8.3	*At*
1101	napin-type 2S albumin 1 precursor	4097894	27	2	12	17.8	6.0	*Gm*
1102	PGP13 (P-glycoprotein 13)	15217776	14	2	0	135.8	9.2	*At*
1103	PGP17 (P-glycoprotein 17)	15232977	16	2	1	136.1	8.6	*At*
1104	R 13 protein	27764543	15	2	4	72.8	6.8	*Gm*
	07 Transporters							
	07.01 Ions							
1105	AAA-type ATPase family protein	42562879	30	4	5	112.8	5.9	*At*
1106	AAA-type ATPase family protein	15234455	26	3	6	69.6	9.2	*At*
1107	AAA-type ATPase family protein	30696968	16	2	5	92.6	6.5	*At*
1108	AAA-type ATPase family protein	10178061	14	2	3	98.6	5.7	*At*
1109	AAA-type ATPase family protein	42567117	14	2	1	124.2	6.4	*At*
1110	AHG1-like ATPase family protein	18417605	35	4	7	56.5	8.6	*At*
1111	AHA7 (H(+)-ATPase 7)	15232300	14	2	5	105.5	6.4	*At*
1112	AHA8 (H(+)-ATPase 8)	15229126	20	2	2	104.1	5.5	*At*
1113	ATCHX18 (cation/hydrogen exchanger 18)	42573539	20	2	3	80.1	8.4	*At*
1114	ATPase, plasma membrane-type, putative / proton pump, putative	15234277	17	2	2	90.4	6.4	*At*
1115	ATP-binding region, ATPase-like domain-containing protein	18408874	18	2	6	66.8	5.4	*At*
1116	calcium ion binding	79595878	18	2	3	55.2	6.0	*At*
1117	calcium-transporting ATPase, plasma membrane-type, putative	15228891	14	2	1	112.5	8.7	*At*
1118	chloroplast ferritin	145442177	210	14	45	28.1	5.7	*Gs*
1119	ECA3 (ER-type calcium-transporting ATPase 3)	4808840	17	2	1	109.1	5.7	*At*
1120	FRS8 (FAR1-related sequence 8)	15220043	18	2	2	79.6	5.9	*At*
1121	haloacid dehalogenase-like hydrolase family protein	15232278	16	2	3	133.6	6.1	*At*
1122	high affinity sulfate transporter 1IVST1	1279876	14	2	5	72.6	9.0	*Hv*
1123	high-affinity nickel-transport family protein	18418571	18	2	7	38.5	9.5	*At*
1124	magnesium dependent soluble inorganic pyrophosphatase	2706450	73	6	17	24.8	5.0	*St*
1125	manganese-superoxide dismutase	147945633	173	11	44	26.7	8.6	*Gm*
1126	mechanosensitive ion channel domain-containing protein	15239859	23	3	4	83.0	9.2	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1127	mitochondrial phosphate transporter	15241291	17	2	5	40.1	9.3	*At*
1128	OEP37, ion channel	18406405	24	2	7	38.8	9.2	*At*
1129	pin-formed 8	15242212	16	2	5	40.5	6.4	*At*
1130	PIP2,2, aquaporin	62546339	14	1	4	30.8	8.3	*Gm*
1131	plasma membrane Ca2+-ATPase	11066054	33	4	5	110.7	5.7	*Gm*
1132	porin, putative	15240765	30	2	3	29.6	8.9	*At*
1133	proton P-ATPase	64460298	18	2	2	104.6	6.0	*Nt*
1134	putative ATP dependent copper transporter	48374954	17	2	6	108.1	6.3	*Zm*
1135	root border cell-specific protein-like protein	82400120	52	5	21	36.0	8.3	*Si*
1136	selenium binding protein	15485722	83	6	15	53.1	5.6	*Gm*
1137	selenium binding protein	13605825	27	2	5	54.0	5.8	*At*
1138	selenium binding protein	15236385	17	1	2	54.1	5.4	*At*
1139	voltage-dependent anion-selective channel protein hst2	15232074	35	3	15	29.4	8.8	*At*
	[illegible]							
	07.07 Sugars							
1140	C4-dicarboxylate transporter/malic acid transport family protein	18391400	17	2	6	63.2	9.3	*At*
1141	carbohydrate transporter/ sugar porter	30680865	15	2	6	52.2	6.3	*At*
1142	dicarboxylate/tricarboxylate carrier	19913109	29	2	8	32.1	9.5	*Nt*
1143	dicarboxylate/tricarboxylate carrier (DTC)	15241167	17	1	5	31.9	9.4	*At*
1144	sucrose-binding protein 2	29469054	171	12	34	55.8	6.1	*Gm*
1145	sugar transporter family protein	30678759	14	1	2	53.5	7.5	*At*
	07 Transporters							
	07.10 Amino Acids							
1146	EDA9 (embryo sac development arrest 9)	15235282	99	7	11	63.3	6.2	*At*
1147	XPO1B (exportin 1B)	30678764	16	2	2	123.2	5.6	*At*
	07 Transporters							
	07.13 Lipids							
1148	homeobox-leucine zipper family protein	30684155	14	2	4	79.3	6.2	*At*
1149	temperature-induced lipocalin'	77744861	74	5	28	21.2	7.8	*Gm*
	07 Transporters							
	07.16 Purine/pyramides							
1150	ADP/ATP translocator	1890116	62	4	11	42.1	9.8	*Le*
1151	plastid developmental protein DAG, putative	15226108	20	2	7	24.7	8.5	*At*
	07 Transporters							
	07.22 Transport ATPase							
	07 Transporters							
	07.26 ABC-type							
1152	(P-glycoprotein 7, PGP7)	15237456	23	2	1	136.1	7.2	*At*
1153	ABC transporter family protein	4204313	20	2	4	65.2	6.4	*At*
1154	ABC transporter family protein	15239420	14	2	3	117.8	8.8	*At*
1155	ATATH11 (ABC2 homolog 11)	15240334	18	2	3	104.5	7.9	*At*
1156	ATNAP6 (non-intrinsic ABC protein 6)	18398463	13	1	1	52.8	5.7	*At*
1157	ATP-binding cassette transporter MRP6	18031899	15	2	2	164.4	6.4	*At*
1158	ATWBC19 (white-brown complex homolog 19)	15233191	18	2	1	80.7	8.6	*At*
1159	multidrug resistance-associated protein (MRP)-like	9280227	41	5	6	144.9	7.0	*At*
1160	PDR-like ABC-transporter	94732079	39	4	3	162.7	9.0	*Gm*
1161	PGP10 (P-glycoprotein 10)	15220188	25	3	3	134.5	6.8	*At*
1162	putative protein	4490736	21	2	1	154.8	8.8	*At*
	07 Transporters							
	07.99 Others							
	08 Intracellular traffic							
	08.01 Nuclear							
1163	atranbp1a	2058282	21	1	6	26.4	5.0	*At*
1164	NBS1 (nijmegen breakage syndrome 1)	145338027	21	2	4	60.1	5.1	*At*
1165	nuclear fusion defective 1	2980773	17	2	3	88.0	5.7	*At*
1166	nuclear transport factor 2B	145324046	22	1	10	15.0	5.9	*At*
1167	nuclear-pore anchor	15219336	14	2	0	238.9	5.0	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1168	nucleoporin 155	8778227	15	2	1	161.3	5.4	*At*
1169	nucleoporin-related	17064886	18	2	2	81.7	9.3	*At*
1170	Ran1 (Ras-related nuclear protein)	123192431	172	11	53	25.3	6.4	*Ps*
1171	ras-GTPase-activating protein SH3-domain binding protein-like	21553535	32	3	7	49.4	5.7	*At*
1172	sterile alpha motif (SAM) domain-containing protein	15225548	20	2	5	81.0	6.0	*At*
	08 Intracellular traffic							
	08.02 Plastid							
	08 Intracellular traffic							
	08.04 Mitochondrial							
1173	mitochondrial substrate carrier family protein	15223098	17	2	6	33.4	9.8	*At*
1174	mitochondrial substrate carrier family protein	15236783	16	2	6	42.6	9.4	*At*
1175	TOM20-2 (translocase outer membrane 20-2)	30689902	15	1	3	23.2	5.5	*At*
	08 Intracellular traffic							
	08.07 Vesicular							
1176	binding (endomembrane system)	18416852	15	2	19	17.9	4.9	*At*
1177	binding (intracellular trafficking and secretion)	42568628	16	2	1	121.5	5.7	*At*
1178	CASP protein-like	11994111	17	2	3	84.4	5.7	*At*
1179	clathrin heavy chain, putative	30681617	18	2	1	193.2	5.3	*At*
1180	dynamin-like GTP binding protein	2267213	19	2	8	68.5	8.8	*At*
1181	dynamin-like protein	8778745	24	2	3	120.3	8.6	*At*
1182	dynamin-like protein	6850867	14	1	2	70.0	6.9	*At*
1183	dynamin-like protein 4	6651401	22	2	3	70.3	8.6	*At*
1184	Putative gamma-adaptin 1	15451585	14	2	1	148.5	6.3	*Os*
1185	Rab GTPase homolog G3f	15230211	94	7	37	23.1	5.0	*At*
1186	Rab GTPase homolog H1e	145357850	29	3	13	23.1	7.7	*At*
1187	transport protein particle (TRAPP) component Bet3 family protein	15229863	36	2	15	19.6	7.6	*At*
1188	unknown protein, contains Exocyst complex subunit SEC6	18424519	18	2	6	47.7	9.5	*At*
1189	unknown protein, vesicular transport	15237322	15	2	4	118.8	5.5	*At*
	08 Intracellular traffic							
	08.10 Perixosomal							
1190	peroxin 5	15241175	19	2	3	80.9	4.7	*At*
	08 Intracellular traffic							
	08.13 Vacuolar							
1191	SNF7 family protein	15220819	13	1	3	22.8	6.7	*At*
1192	vacuolar processing enzyme 2	37542692	17	1	2	53.5	5.4	*Gm*
1193	vacuolar protein sorting-associated protein 28 family protein	15234509	14	1	6	23.5	5.2	*At*
1194	vacuolar sorting receptor protein PV72-like protein	83284015	24	2	3	69.3	5.2	*Si*
1195	vacuolar targeting receptor bp-80	8886326	26	2	3	69.3	5.4	*Ta*
1196	VHS and GAT domain protein	82791812	15	1	2	73.0	5.4	*Gm*
1197	VHS domain-containing protein / GAT domain-containing protein	15237869	31	2	5	45.3	4.8	*At*
	08 Intracellular traffic							
	08.16 Extracellular							
1198	3-phosphoinositide dependent protein kinase 1	81538200	17	2	4	55.1	8.1	*Zm*
1199	oxysterol-binding family protein	15222204	18	2	3	92.3	6.1	*At*
1200	phox (PX) domain-containing protein	22327944	53	4	7	65.5	5.2	*At*
1201	polyphosphoinositide binding protein Ssh1p	2739044	16	2	5	36.9	6.8	*Gm*
1202	polyphosphoinositide-binding protein, putative	15238794	16	2	5	38.9	6.9	*At*
	08 Intracellular traffic							
	08.19 Import							
1203	importin	33337497	63	4	10	57.6	5.1	*Os*
1204	super sensitive to ABA and drought 2	30685014	16	1	1	119.2	4.8	*At*
	08 Intracellular traffic							
	08.99 Others							
1205	exocyst complex subunit Sec15-like family protein	18411920	14	2	2	86.5	6.0	*At*
1206	SEC14 cytosolic factor family protein / phosphoglyceride transfer	30695991	24	3	5	72.0	8.9	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	09 Cell structure							
	09.01 Cell wall							
1207	alpha-expansin 2	14193753	21	2	6	28.8	9.7	*Zm*
1208	cellulose synthase catalytic subunit	62318989	19	2	3	28.4	8.9	*At*
1209	GAE6 (UDP-D-glucoronate 4-epimerase 6)	24417280	21	2	7	50.6	9.7	*At*
1210	polygalacturonase inhibiting protein	37051109	14	1	6	20.9	9.4	*Ps*
1211	polygalacturonase inhibiting protein precursor	110836643	14	1	5	36.9	8.6	*Gm*
1212	UDP-glucuronate decarboxylase 1	48093461	119	8	23	38.7	7.1	*Nt*
1213	UXS1 (UDP-glucoronic acid decarboxylase 1)	48093467	32	3	8	45.9	9.0	*Nt*
1214	wall-associated kinase	15220882	15	1	1	85.2	6.6	*At*
	09 Cell structure							
	09.04 Cytoskeleton							
1215	actin	1498334	298	19	66	37.2	5.5	*Gm*
1216	actin	1666228	54	3	10	41.7	5.3	*Ps*
1217	actin	1498393	21	2	12	37.2	5.6	*Zm*
1218	actin depolumerizing factor 2	18408116	27	2	21	15.7	5.3	*At*
1219	actin-depolymerizing factor, putative	15223471	27	2	13	16.3	5.5	*At*
1220	actin-related protein C4	79325095	18	1	8	19.9	7.6	*At*
1221	alpha tubulin-2A	90289596	107	6	22	49.7	4.9	*Ta*
1222	alpha tubulin-5D	90289614	51	3	10	49.9	5.1	*Ta*
1223	ankyrin repeat family protein	15229331	18	2	4	49.3	5.2	*At*
1224	ankyrin repeat family protein	9280657	16	2	3	70.7	7.0	*At*
1225	ankyrin repeat family protein	15236310	16	2	5	70.5	6.8	*At*
1226	apospory-associated protein	19070767	15	1	3	23.1	4.5	*Os*
1227	ARPC p20	34148072	41	4	30	14.7	8.5	*Gm*
1228	cyclase associated protein 1	15236128	46	4	6	51.0	6.2	*At*
1229	microtubule motor	22331291	15	2	1	146.6	5.3	*At*
1230	microtubule organization 1	30686489	17	2	1	217.6	7.6	*At*
1231	microtubule-associated protein MAP65-1b	11558254	14	1	1	65.9	5.2	*Nt*
1232	microtubule-associated proteins 70-3	15226344	17	2	4	69.8	9.4	*At*
1233	profilin	156938901	42	3	24	14.1	4.7	*Gm*
1234	putative protein	4455199	19	2	5	53.7	5.0	*At*
1235	suppressor of actin 9	42566068	16	2	0	180.1	6.3	*At*
1236	tetraspore	30690898	16	2	2	106.5	8.6	*At*
1237	TGB12K interacting protein 2	29826242	91	6	18	37.3	4.5	*Nt*
1238	TUB8 (tubulin beta-8)	18420724	99	6	22	50.6	4.7	*At*
1239	tubulin A	62546341	323	20	64	49.7	5.0	*Gm*
1240	tubulin I34	62546343	399	25	66	50.4	4.7	*Gm*
1241	tubulin beta-5 chain	18394812	42	3	8	50.3	4.7	*At*
1242	villin 2	18405794	20	2	1	107.8	5.2	*At*
	09 Cell structure							
	09.07 ER/Golgi							
1243	beta-lg-H3 domain-containing protein / fasciclin domain-containing protein	18399319	15	2	6	50.8	6.7	*At*
1244	plant basic secretory protein (BSP) family protein	15226060	35	2	10	25.4	6.0	*At*
1245	unknown	116831573	29	2	7	23.5	9.3	*At*
1246	unknown	117670154	25	2	8	27.2	5.8	*Hv*
1247	unknown	77416945	21	1	6	32.5	9.4	*St*
1248	unknown	116830944	19	1	23	8.2	11.3	*At*
1249	unknown	78191396	17	2	7	30.1	11.5	*St*
1250	unknown	21592434	17	2	7	27.2	9.2	*At*
1251	unknown	116831277	16	2	3	66.7	9.4	*At*
1252	unknown protein, endomembrane system	15235414	23	2	11	27.9	9.5	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
	09 Cell structure							
	09.10 Nucleus							
1253	NAP1	15224782	28	2	5	43.5	4.3	*At*
	09 Cell structure							
	09.13 Chromosomes							
1254	DNA topoisomerase I beta	15240492	27	3	3	102.8	9.4	*At*
1255	DNA topoisomerase II family protein	30680387	19	2	4	60.9	8.2	*At*
1256	MAG2 (similar to chromosome structural maintenance)	15228233	16	2	3	90.2	5.4	*At*
1257	mini-chromosome maintenance protein MCM6	68236762	18	2	2	92.9	5.7	*Ps*
1258	telomere repeat binding factor 1	15229625	16	2	7	32.2	9.5	*At*
	09 Cell structure							
	09.16 Mitochondria							
1259	34 kDa outer mitochondrial membrane protein porin-like protein	83283993	59	4	14	29.4	7.7	*St*
1260	hypothetical protein ArthMp060	13449345	17	2	10	30.2	4.9	*At*
1261	mitochondrial F1-ATPase, gamma subunit	110740981	19	1	4	35.5	9.0	*At*
1262	NFU4 (NFU domain protein 4)	18402817	33	2	9	30.5	4.9	*At*
1263	unknown protein, contains domain ARM repeat	18410039	15	2	1	312.0	5.5	*At*
1264	unknown protein, similar to C2 domain-containing protein	79401911	23	2	3	78.1	8.5	*At*
	09 Cell structure							
	09.19 Perixosome							
1265	peroxisomal copper-containing amine oxidase	5230728	29	2	5	78.6	6.1	*Gm*
1266	uricase (Nod-35)	1498172	61	5	21	35.2	8.3	*Gm*
	09 Cell structure							
	09.26 Vacuole							
1267	GTP binding	15221444	103	7	21	44.5	6.4	*At*
1268	GTP-binding protein	303750	48	3	21	22.5	5.3	*Ps*
1269	NTGB2 (GTP-binding protein)	1184989	128	7	57	15.8	5.8	*Nt*
1270	ras-related GTP binding protein possessing GTPase activity	432607	28	2	13	22.6	5.3	*Os*
1271	small GTP-binding protein	22597172	55	5	31	24.1	7.7	*Gm*
1272	small GTP-binding protein	1381678	47	4	20	22.4	5.1	*Gm*
	09 Cell structure							
	09.28 Plastid							
1273	hypothetical chloroplast RF1	91214187	21	3	1	215.9	10.0	*Gm*
1274	translocon outer membrane complex 75-III	15232625	32	4	4	89.2	8.9	*At*
1275	unknown protein	30682877	19	2	2	105.1	6.2	*At*
1276	unknown protein, chloroplast	30689549	45	3	13	32.2	6.3	*At*
1277	unknown protein, chloroplast	145334497	26	2	14	17.9	7.7	*At*
1278	unknown protein, chloroplast	18411597	25	3	9	35.1	10.0	*At*
1279	unknown protein, chloroplast	15240659	16	2	3	64.4	9.8	*At*
1280	unknown protein, similar to mov34 family protein	145323832	17	2	6	24.8	5.0	*At*
1281	unknown protein, similar to pollen preferential protein	18399317	23	3	12	27.9	9.0	*At*
1282	unknown protein, similar to similar to Glutathione S-transferase	79325173	44	3	6	43.7	8.0	*At*
1283	unknown protein, similar to threonine endopeptidase	18411555	15	2	18	17.7	9.2	*At*
	09 Cell structure							
	09.99 Others							
1284	Arabidopsis homolog of nucleolar protein NOP56	15223458	17	2	2	58.7	8.8	*At*
1285	Nrap protein, nucleolus	145337144	14	2	1	120.2	6.4	*At*
1286	nucleolin, putative	15222009	15	2	4	58.8	5.1	*At*
1287	prohibitin	7716458	43	4	15	30.7	6.6	*Zm*
1288	suppressor of lin-12-like protein-related / sel-1 protein-related	14532716	15	2	3	75.9	5.8	*At*
	10 Signal transduction							
	10.01 Receptors							
1289	CBL-interacting protein kinase 4	15233500	13	2	5	47.8	8.2	*At*
1290	ethylene response sensor 1	15226788	17	2	10	68.3	6.1	*At*
1291	F17L21.26, contains WD40 domain	9802540	23	2	1	113.9	7.1	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1292	histidine kinase 2	38347686	26	3	2	112.4	8.7	*Zm*
1293	phototropic-responsive protein, putative	15229647	14	2	4	59.9	6.4	*At*
1294	receptor kinase 2	110341794	21	2	3	81.6	7.6	*Ta*
1295	receptor protein kinase-related	15233211	18	2	9	30.8	9.2	*At*
1296	receptor protein-like	11994328	24	3	5	92.2	9.5	*At*
1297	receptor serine/threonine kinase, putative	30684346	18	2	3	95.3	6.6	*At*
1298	receptor-like kinase	8575543	15	2	3	74.2	8.9	*Os*
1299	receptor-like protein kinase-related	116831228	13	2	10	31.0	9.2	*At*
1300	signal transducer	79579427	16	2	3	67.6	8.2	*At*
1301	similar to receptor-like protein kinase precursor	6049881	21	2	3	87.5	8.4	*At*
	10 Signal transduction							
	10.04 Mediators							
1302	phosphoserine aminotransferase	110737825	90	6	13	46.6	8.6	*At*
1303	transducin family protein / WD-40 repeat family protein	6091750	37	4	2	147.7	6.0	*At*
1304	transducin family protein / WD-40 repeat family protein	15242311	16	2	2	65.8	6.5	*At*
1305	transducin family protein / WD-40 repeat family protein	18400838	15	2	9	37.4	6.0	*At*
1306	WD-40 repeat protein	2289095	52	4	8	35.8	7.6	*At*
1307	WD-40 repeat protein	69207914	36	3	7	48.7	4.7	*Ps*
	10 Signal transduction							
	10.0400 Kinases							
1308	adenosine kinase	21698922	16	1	4	32.2	5.3	*Os*
1309	adenosine kinase isoform 2S	51949802	44	3	14	37.6	5.2	*Nt*
1310	adenylate kinase	7630193	88	6	34	22.9	8.2	*Os*
1311	aspartate kinase-homoserine dehydrogenase I	26452575	24	3	5	99.5	6.3	*At*
1312	AT5g52920/MXC20_15	15081612	18	1	1	63.5	6.6	*At*
1313	ATMPK4 (MAP KINASE 4)	15234152	20	2	6	42.9	5.7	*At*
1314	HON3 (HONZA13)	79339171	16	2	4	64.2	5.7	*At*
1315	carbohydrate kinase family	18419840	55	4	12	39.3	6.6	*At*
1316	CDG1 (constitutive differential growth 1)	116831240	18	2	6	48.6	9.2	*At*
1317	CIPK13 (CIPK13)	15226241	20	2	4	56.7	8.4	*At*
1318	CIPK5 (CBL-INTERACTING PROTEIN KINASE 5)	30683398	15	2	6	50.9	6.1	*At*
1319	FAT domain-containing protein	22329206	45	5	1	434.0	6.8	*At*
1320	geranylgeranylated protein ATGP1	15242933	24	2	19	22.5	7.0	*At*
1321	hexokinase 1a	45387403	18	1	2	53.8	6.1	*Nt*
1322	LSTK-1-like kinase	15637110	16	2	5	68.5	9.4	*Le*
1323	mitogen-activated protein kinase 2	33340593	43	3	9	44.8	5.5	*Gm*
1324	NADK2 (NAD kinase 2)	18395013	17	2	3	109.2	7.5	*At*
1325	PAK (PHOSPHATIDIC ACID KINASE)	18391394	17	2	2	100.9	6.2	*At*
1326	phosphatidylinositol-4-phosphate 5-kinase family protein	15230176	24	3	6	89.2	6.8	*At*
1327	phosphatidylinositol-4-phosphate 5-kinase family protein	15227305	19	2	3	88.2	7.0	*At*
1328	protein kinase family protein	22331335	20	2	3	108.0	6.9	*At*
1329	protein kinase family protein	17529280	19	2	4	75.4	5.9	*At*
1330	protein kinase family protein	42566214	17	2	2	74.9	6.0	*At*
1331	protein kinase family protein	15242731	16	2	3	69.6	5.9	*At*
1332	protein kinase family protein	42568974	15	2	2	78.1	6.2	*At*
1333	protein kinase family protein	22327904	14	2	3	88.5	8.8	*At*
1334	protein kinase family protein	51847836	14	2	6	56.2	8.9	*Gm*
1335	protein kinase family protein	42563282	14	2	2	87.5	5.3	*At*
1336	protein kinase family protein	15224775	14	1	1	90.9	9.1	*At*
1337	protein kinase family protein / glycerophosphoryl diester phosphodiesterase family protein	9755449	14	1	0	123.0	6.1	*At*
1338	protein kinase like	7801691	16	2	2	143.9	5.6	*At*
1339	protein kinase SPK-2	21593534	16	2	6	41.1	4.7	*At*
1340	protein kinase, putative	11120796	17	2	3	129.8	5.0	*At*
1341	protein kinase, putative	15221465	16	2	2	64.9	9.2	*At*
1342	protein kinase-like	2924514	16	2	4	52.2	6.4	*At*
1343	protein kinase-like protein	18415205	18	2	5	52.6	5.7	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1344	putative arabinose kinase	2326372	16	2	8	108.4	5.7	*At*
1345	putative protein kinase	120400397	19	2	4	51.9	6.9	*Zm*
1346	putative protein kinase	13324795	15	2	3	139.5	5.8	*Os*
1347	putative receptor protein kinase PERK1	77403742	17	2	4	48.6	6.6	*Gm*
1348	serpin family protein / serine protease inhibitor family protein	15225956	24	2	8	45.9	5.7	*At*
1349	S-locus protein kinase, putative	15237047	15	2	3	91.9	6.5	*At*
1350	SP3D	28200390	29	2	9	20.1	6.7	*Le*
1351	SRC2	2055230	52	4	13	31.0	6.7	*Gm*
1352	thymidylate kinase family protein	145334853	27	2	7	30.4	8.7	*At*
1353	wee1	42362341	21	3	6	55.9	6.9	*Gm*
	10 Signal transduction							
	10.0407 Phosphatases							
1354	At1g78200/T11I11_14	15081703	15	2	12	30.8	8.6	*At*
1355	ATPAP24/PAP24 (purple acid phosphatase 24)	30686692	23	2	2	69.1	5.7	*At*
1356	ATUK/UPRT1	15237512	48	4	11	54.4	6.1	*At*
1357	C2 domain-containing protein	5882720	16	2	1	141.4	8.5	*At*
1358	calcineurin-like phosphoesterase family protein	15222942	23	2	5	68.2	6.3	*At*
1359	calcineurin-like phosphoesterase family protein	15238894	14	2	3	67.6	4.9	*At*
1360	catalytic/ coenzyme binding	18404496	47	3	7	34.9	8.4	*At*
1361	catalytic/ coenzyme binding	18399328	34	3	8	43.9	9.3	*At*
1362	catalytic/ coenzyme binding	7340698	23	2	5	41.6	6.1	*At*
1363	catalytic/ coenzyme binding	10129651	19	2	6	32.7	9.3	*At*
1364	CPL2 (CTD phosphatase-like 2)	6759450	14	2	2	86.0	6.1	*At*
1365	CTD phosphatase-like 3	22212705	35	4	3	136.5	5.6	*At*
1366	inositol 1,4,5-trisphosphate 5-phosphatase	28393619	23	3	2	121.8	6.0	*At*
1367	inositol monophosphatase family protein	21537207	15	2	9	40.4	7.1	*At*
1368	inositol monophosphatase family protein	15234590	14	2	8	43.5	5.8	*At*
1369	kelch repeat-containing serine/ threonine phosphoesterase family protein	42569377	18	1	1	107.5	5.5	*At*
1369	Met-10+ like family protein / kelch repeat-containing protein	22328346	15	2	5	110.9	6.2	*At*
1370	phosphatase-like protein Psc923	13123659	24	2	5	43.4	5.5	*Ps*
1371	protein phosphatase 2A	1568511	97	7	17	65.4	5.0	*Nt*
1372	protein phosphatase 2A catalytic subunit	34398263	29	2	10	35.0	4.8	*Le*
1373	protein phosphatase type 2C	79326653	16	2	13	42.0	5.5	*At*
1374	putative phosphatase	27527030	105	8	30	28.9	5.7	*Gm*
1375	putative tyrosine phosphatase	8926334	17	2	8	27.2	6.7	*Os*
1376	TOR (target of rampamycin)	8569097	24	3	1	282.9	6.4	*At*
1377	unknown protein, similar to kelch repeat-containing F-box family	79527293	13	1	2	46.4	6.6	*At*
1378	uracil phosphoribosyl transferase like protein	21554263	49	4	8	52.6	6.4	*At*
	10 Signal transduction							
	10.0410 G proteins							
1379	G protein beta-subunit-like protein	125987958	32	2	18	15.7	6.1	*Nt*
	10 Signal transduction							
	10.99 Others							
1380	14-3-3 protein	22597176	99	6	36	24.9	4.9	*Gm*
1381	14-3-3-like protein	4850247	204	12	46	29.4	4.7	*Ps*
1382	14-3-3-like protein	46946654	47	4	15	28.9	4.8	*Zm*
1383	ATMP2 (At membrane associated progesterone binding protein 3)	4960154	32	2	9	28.2	8.6	*At*
1384	calcium-binding protein, putative	15233402	17	2	15	21.1	4.6	*At*
1385	calmodulin binding	79328260	19	2	11	21.0	6.4	*At*
1386	calmodulin-binding protein	12324199	16	2	5	50.7	7.2	*At*
1387	CAM2 (CALMODULIN-2)	30683369	59	4	23	20.6	4.7	*At*
1388	flowering locus T	56694632	17	1	6	19.8	7.7	*Ta*
1389	FPA	2288985	13	2	2	117.4	7.5	*At*
1390	PHD finger family protein	9294545	23	3	2	154.7	6.7	*At*
1391	PNCBP	17933110	16	2	2	146.2	5.9	*Si*
1392	protein transport SEC13-like protein	83283979	45	3	17	32.7	5.8	*Si*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1393	SNF1-related protein kinase regulatory beta subunit 1	66710734	19	1	5	30.7	5.6	*Ps*
	11 Disease/defence							
	11.01 Resistance genes							
1394	At1g11360/T23J18_35	15215774	13	1	6	11.3	4.8	*At*
1395	ATPDR5/PDR5 (PLEIOTROPIC DRUG RESISTANCE 5)	15228112	14	2	2	160.3	8.3	*At*
1396	ATPDR9/PDR9 (PLEIOTROPIC DRUG RESISTANCE 9)	15231821	20	2	1	164.1	6.8	*At*
1397	disease resistance protein (CC-NBS-LRR class), putative	15218003	18	2	3	121.0	6.9	*At*
1398	disease resistance protein (NBS-LRR class), putative	15237022	15	2	2	103.9	6.6	*At*
1399	disease resistance protein (TIR-NBS-LRR class), putative	15222558	20	2	1	115.4	6.3	*At*
1400	disease resistance protein (TIR-NBS-LRR class), putative	15242300	16	2	1	132.4	8.4	*At*
1401	disease resistance protein RPP13 variant	46410197	21	2	2	96.7	6.2	*At*
1402	disease resistance protein RPP13 variant	46410195	17	2	2	96.0	6.5	*At*
1403	disease resistance protein RPP1-WsB	15230846	14	2	1	223.6	6.6	*At*
1404	disease resistance protein SLH1	51555866	17	2	2	156.1	5.6	*At*
1405	disease resistance-responsive protein-related	15217886	13	1	6	20.6	9.5	*At*
1406	functional candidate resistance protein KR1	18033111	27	3	2	128.7	6.7	*Gm*
1407	nematode resistance protein-like protein	21553529	18	2	7	48.9	5.5	*At*
1408	Pi-b protein	6172381	19	2	1	141.6	8.6	*Os*
1409	probable resistance gene	7488170	38	5	3	280.6	5.3	*At*
1410	putative disease resistance protein	142942426	16	2	1	151.1	5.8	*St*
1411	Putative disease resistance protein, identical	54261825	18	2	4	96.1	8.1	*St*
1412	resistance protein KR4	27463527	14	2	1	137.0	6.0	*Gm*
1413	resistance protein Lcr3	32364514	17	2	1	104.3	6.2	*At*
1414	resistence protein-like	30684780	22	2	1	148.9	5.7	*At*
1415	rust resistance protein Rp1-dp7	12744961	25	3	2	144.7	6.6	*Zm*
1416	rust resistance protein rp3-1	45934295	19	2	1	142.3	6.5	*Zm*
1417	T7N9.23	10121908	24	3	3	177.0	6.1	*At*
1418	universal stress protein (USP) family protein	30693971	48	3	15	17.8	5.7	*At*
1419	universal stress protein (USP) family protein	8885586	36	2	12	26.9	6.3	*At*
	11 Disease/defence							
	11.02 Defence-related							
1420	24 kDa protein SC24	18448973	108	7	38	24.6	9.1	*Gm*
1421	AGD2 (ABERRANT GROWTH AND DEATH 2)	18418270	30	2	4	50.4	7.0	*At*
1422	aminoaldehyde dehydrogenase	15131692	65	4	10	54.7	5.5	*Ps*
1423	chitinase class I	6573210	84	5	26	34.3	7.4	*Gm*
1424	class III ADH enzyme	1675394	37	2	8	40.8	6.8	*Os*
1425	oxidoreductase, zinc-binding dehydrogenase family protein	15220854	39	3	10	41.0	8.5	*At*
1426	oxidoreductase, zinc-binding dehydrogenase family protein	30697873	37	3	6	45.5	7.6	*At*
	11 Disease/defence							
	11.03 Cell death							
	11 Disease/defence							
	11.04 Rescue							
	11 Disease/defence							
	11.05 Stress responses							
1427	ALDH10A9 (Aldehyde dehydrogenase 10A9)	15228346	24	2	4	54.9	5.4	*At*
1428	ALDH2B4 (aldehyde dehydrogenase 2)	15228319	80	5	10	58.6	7.1	*At*
1429	ALDH5F1 (succinic semialdehyde dehydrogenase)	15219379	23	2	6	56.6	6.5	*At*
1430	ATOMT1 (O-methyltransferase 1)	15239571	24	2	5	39.6	5.6	*At*
1431	caffeoyl-CoA 3-O-methyltransferase 5	2511737	77	5	32	27.2	5.4	*Nt*
1432	catechol O-methyltransferase	396589	23	2	6	39.4	5.4	*Nt*
1433	CCA1 (circadian clock associated 1)	30690518	16	2	2	67.0	5.7	*At*
1434	cold acclimation protein WCS19	19352333	14	2	38	14.6	5.6	*Ta*
1435	dimethylmenaquinone methyltransferase family protein	15232963	17	1	6	17.8	5.7	*At*
1436	FtsJ-like methyltransferase family protein	15240727	19	2	12	24.3	9.3	*At*
1437	glutaredoxin family protein	15237554	13	2	8	43.1	9.1	*At*
1438	methylmalonate semialdehyde dehydrogenase	30144414	29	2	4	41.1	5.5	*Ta*
1439	osmotin-like protein, putative	15226956	16	1	5	27.0	7.6	*At*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1440	unknown protein, reponse to salt stress	15239835	15	2	3	56.8	5.7	*At*
	11 Disease/defense							
	11.06 Detoxification							
1441	1-Cys peroxiredoxin PER1 (Thioredoxin peroxidase)	146325682	15	1	4	24.9	6.3	*Zm*
1442	2,4-D inducible glutathione S-transferase	2920666	56	4	17	25.6	6.2	*Gm*
1443	2-oxoacid dehydrogenase family protein	15240454	15	1	1	50.1	9.2	*At*
1444	5,10-methylenetetrahydrofolate dehydrogenase -5,10-methenyltetrahydrofolate cyclohydrolase	4103987	48	3	11	31.3	8.3	*Ps*
1445	ACC oxidase	25989506	15	2	6	35.6	6.1	*St*
1446	alcohol dehydrogenase	12862754	44	3	5	41.2	5.8	*At*
1447	alcohol dehydrogenase 1	22597178	287	19	57	40.0	6.2	*Gm*
1448	alcohol-dehydrogenase	4039115	198	13	37	36.4	6.1	*Gm*
1449	aldehyde dehydrogenase family 7 member A1	29893325	218	15	44	54.7	5.5	*Gm*
1450	aldo/keto reductase family protein	4895205	65	4	11	39.0	6.8	*At*
1451	aldo/keto reductase family protein	15232354	43	3	6	35.0	6.5	*At*
1452	aldo/keto reductase family protein	30696459	40	3	11	37.9	5.9	*At*
1453	allyl alcohol dehydrogenase	6692816	31	3	6	38.1	6.6	*Nt*
1454	alpha-DOX1	37962659	17	2	3	73.8	6.5	*Le*
1455	APX1 (ascorbate peroxidase 1, maternal effect embryo arrest 6)	15223049	15	1	5	27.6	5.7	*At*
1456	arginase	15236640	38	3	7	37.3	6.1	*At*
1457	arginase 2	54648782	33	3	10	36.9	5.6	*Le*
1458	ascorbate oxidase precursor	22218270	33	2	6	48.1	8.0	*Gm*
1459	ascorbate peroxidase	21039134	42	3	9	42.2	8.7	*Le*
1460	ascorbate peroxidase 2	1336082	230	13	64	27.1	5.7	*Gm*
1461	AT5g06290	18415155	17	1	9	29.8	5.6	*At*
1462	ATFRO8/FRO8 (ferric reduction oxidase 8)	22327681	15	2	4	83.2	9.6	*At*
1463	monodehydroascorbate reductase 1 (ATMDAR1)	15231702	15	2	3	46.5	6.4	*At*
1464	ATY2 (Arabidopsis thioredoxin y2)	30693659	15	1	4	18.6	8.5	*At*
1465	catalase	2661023	115	8	33	35.8	6.4	*Gm*
1466	catalase	3929924	96	6	13	56.2	7.4	*Os*
1467	catalase	40950550	27	2	7	56.4	6.6	*St*
1468	Contains Thioredoxin domain PF00085. ESTs gb|T42351	6899648	14	2	5	60.8	8.2	*At*
1469	copper chaperone homolog CCH	6525011	52	3	29	13.6	4.7	*Gm*
1470	cytosolic glutathione reductase	157362219	30	2	3	53.1	6.1	*Hv*
1471	cytosolic NADP-malic enzyme	2150029	87	6	14	64.1	5.7	*Le*
1472	dehydroascorbate reductase	68131811	150	10	45	28.9	8.5	*Gm*
1473	dehydroascorbate reductase	28192427	21	1	8	23.7	7.7	*Nt*
1474	dehydroascorbate reductase, putative	15239354	17	1	5	24.1	4.9	*At*
1475	dehydroascorbate reductase-like protein	76160951	23	2	13	23.4	6.1	*St*
1476	DHAR2	21593056	22	1	7	23.4	6.0	*At*
1477	Dihydrodipicolinate synthase, chloroplast precursor (DHDPS)	118241	25	2	5	41.2	6.9	*Zm*
1478	F3F9.13	8052535	22	2	4	50.1	6.3	*At*
1479	glutathione peroxidase 6 (ATGPX6)	30681827	47	3	12	25.6	9.4	*At*
1480	glutathione peroxidase 7 (ATGPX7)	22329066	60	4	12	25.8	9.5	*At*
1481	glutathione peroxidase-like protein	22268405	31	2	23	18.6	5.7	*Zm*
1482	glutathione reductase	6714837	27	3	5	58.7	8.1	*Gm*
1483	glutathione S-transferase	37051105	17	1	4	26.7	5.0	*Ps*
1484	glutathione S-transferase GST 12	11385439	16	2	7	26.6	5.9	*Gm*
1485	homoglutathione synthetase	7799808	201	14	32	55.7	6.0	*Gm*
1486	hsr203J homolog	6092014	16	1	2	38.4	5.8	*Ps*
1487	IMS1	15221125	32	3	6	68.1	6.0	*At*
1488	In2-1 protein	21537338	20	2	11	27.1	5.0	*At*
1489	LYTB-like protein precursor-like	81076565	22	2	3	51.8	5.8	*St*
1490	mitochondrial peroxiredoxin	47775654	15	1	4	21.5	8.4	*Ps*
1491	NTRB (NADPH-dependent thioredoxin reductase B)	42566473	16	2	3	116.0	9.1	*At*
1492	peroxidase	17467210	142	8	40	38.1	5.0	*Gm*
1493	peroxidase	5002236	123	8	31	37.1	5.5	*Gm*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1494	peroxidase	5002234	38	2	11	37.5	8.6	*Gm*
1495	peroxidase ATP22a	15234648	48	4	16	35.8	9.3	*At*
1496	peroxin-3 family protein	12597815	16	2	2	90.8	6.0	*At*
1497	phospholipid hydroperoxide glutathione peroxidase	31872080	29	2	9	18.9	6.4	*Le*
1498	pseudo-atpA	22742	138	10	32	34.8	5.6	*Gm*
1499	putative lactoylglutathione lyase	15810219	35	3	9	32.0	5.1	*At*
1500	putative NADP-dependent malic enzyme	46200525	15	2	4	34.7	6.6	*Zm*
1501	putative peroxiredoxin	28393058	16	1	7	24.1	6.2	*At*
1502	putative thioredoxin m2	15594012	16	1	6	19.8	9.2	*Ps*
1503	respiratory burst oxidase protein	9757922	15	2	2	100.6	9.3	*At*
1504	respiratory burst oxidase protein D	3242789	15	2	2	103.9	9.3	*At*
1505	respiratory burst oxidase protein B	3242787	15	2	2	107.3	9.0	*At*
1506	short chain alcohol dehydrogenase	27804441	32	3	14	18.2	7.1	*At*
1507	short chain alcohol dehydrogenase	2739279	20	1	5	29.8	6.2	*Nt*
1508	short-chain dehydrogenase/reductase (SDR) family protein	15229203	26	3	14	28.1	6.1	*At*
1509	short-chain dehydrogenase/reductase (SDR) family protein	18398539	24	2	10	34.3	9.3	*At*
1510	superoxidase dismutase	33327349	34	2	12	22.3	6.0	*Le*
1511	tetrahydrofolate dehydrogenase/cyclohydrolase, putative	15236888	28	2	8	38.7	8.6	*At*
1512	thioredoxin h	8980491	20	2	14	13.4	5.1	*Ta*
1513	thioredoxin M-type, chloroplast precursor (TRX-M)	3334376	33	2	11	18.1	8.7	*Zm*
1514	thioredoxin peroxidase	21912927	20	1	6	29.8	8.2	*Nt*
1515	thioredoxin-related	15224267	29	2	9	24.5	6.4	*At*
1516	thylakoid-bound ascorbate peroxidase	25992557	23	2	4	41.3	5.4	*Ta*
1517	TPX1 (thioredoxin-dependent peroxidase 1)	15218877	29	2	6	17.4	5.2	*At*
1518	TTL4 (tetratricopeptide-repeat thioredoxin-like 4)	42566029	15	2	2	74.4	9.1	*At*
	12 Unclear classification							
1519	armadillo/beta-catenin repeat family protein	15232329	16	1	2	40.9	5.1	*At*
1520	armadillo/beta-catenin repeat family protein	15235456	16	2	9	43.5	7.7	*At*
1521	armadillo/beta-catenin repeat family protein	15220457	15	2	1	86.3	5.3	*At*
1522	At2g19940/F6F22.3	16604368	29	2	6	44.2	8.5	*At*
1523	AT3g05350/T12H1_32	19310478	17	1	0	78.7	6.1	*At*
1524	ATK4 (AtKINESIN 4)	18421069	13	2	2	110.0	6.5	*At*
1525	ATPHB2 (PROHIBITIN 2)	15219569	16	2	8	31.8	9.4	*At*
1526	BTB/POZ domain-containing protein	62319919	15	2	3	92.5	5.7	*At*
1527	Bowman-Birk proteinase inhibitor	4191566	16	1	35	4.5	6.7	*Gm*
1528	Bowman-Birk type proteinase isoinhibitor C	19698249	16	1	11	12.1	5.2	*Gs*
1529	Bowman-Birk type proteinase isoinhibitor D	19698251	15	1	11	12.2	5.7	*Gs*
1530	C2 domain-containing protein	2244932	17	2	3	72.3	9.4	*At*
1531	C2 domain-containing protein beta-catenin repeat family protein	15223965	17	2	0	228.6	5.4	*At*
1532	catalytic	79537390	22	2	4	73.9	5.2	*At*
1533	CBS domain-containing protein	15238284	33	3	16	22.7	9.1	*At*
1534	circumsporozoite protein-like protein	18087675	15	2	8	17.2	12.1	*Os*
1535	delta-COP	7677262	20	2	6	57.5	5.5	*Zm*
1536	dienelactone hydrolase family protein	15225693	32	2	8	25.9	5.3	*At*
1537	dormancy related protein, putative	12322163	14	1	2	31.2	5.9	*At*
1538	exostosin family protein	15241150	24	3	4	69.1	9.5	*At*
1539	F5M15.25	8778608	31	4	6	100.2	6.6	*At*
1540	fasciclin-like protein FLA26	115349936	18	2	8	49.6	6.1	*Ta*
1541	GTP-binding protein	303742	18	2	7	24.1	5.7	*Ps*
1542	kinesin motor protein-related	145334367	15	2	1	141.1	5.9	*At*
1543	kinesin motor protein-related	15227596	15	2	2	119.3	5.8	*At*
1544	kinesin-like protein	10177316	14	2	3	149.7	7.3	*At*
1545	kinesin-related protein (MKRP1)	5263326	31	4	5	100.4	6.2	*At*
1546	Kunitz trypsin inhibitor	13375349	246	15	52	24.1	5.0	*Gm*
1547	MUR3 (MURUS 3)	30680972	18	2	5	70.7	8.3	*At*
1548	NBS-LRR-like protein	15788516	19	2	3	144.4	6.5	*Hv*
1549	P23 protein	587546	44	3	11	18.8	4.5	*St*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1550	putative kinesin heavy chain	4567265	16	2	5	65.7	9.6	*At*
1551	putative protein	2864615	20	2	2	93.6	5.2	*At*
1552	remorin family protein	42571771	16	2	5	60.9	9.7	*At*
1553	serine-rich protein-related	15238753	23	2	12	24.1	11.8	*At*
1554	serpin, putative / serine protease inhibitor, putative	15220298	21	2	6	42.6	5.0	*At*
1555	T22C5.22	6693025	15	2	25	9.7	9.7	*At*
1556	trypsin inhibitor	9367042	30	2	17	18.0	6.1	*Gm*
1557	tumor protein-like protein	125660616	18	1	12	14.3	5.0	*Ps*
1558	unknown protein, contains nucleic acid-binding domain	15223907	15	2	4	80.7	9.6	*At*
1559	unknown protein, agenet domain containing protein	18873854	18	2	1	217.2	5.8	*Os*
1560	unknown protein, aminotransferase-like protein	15222573	25	3	2	134.2	9.4	*At*
1561	unknown protein, calcineurin-like phosphoesterase	6714410	18	2	0	231.2	7.7	*At*
1562	unknown protein, contains domain mannose-binding lectin	15224348	16	2	8	35.6	9.1	*At*
1563	unknown protein, contains HGWP repeat	19881564	17	2	2	103.9	11.3	*Os*
1564	unknown protein, contains SAM binding motif	18400797	17	2	3	32.1	8.7	*At*
1565	unknown protein, contains WD40 domain	21539489	16	2	4	88.0	6.6	*At*
1566	unknown protein, contains WD40 domain	30678740	16	2	2	108.0	4.9	*At*
1567	unknown protein, endomembrane system	15239555	14	1	2	39.3	8.7	*At*
1568	unknown protein, similar to bystin (51.6 kD)-like	145336313	14	2	6	51.5	6.0	*At*
1569	unknown protein, similar to IMP dehydrogenase/GMP reductase	15239100	16	2	2	84.5	5.0	*At*
1570	unknown protein, similar to kinectin-related	15227941	18	2	13	38.3	4.8	*At*
1571	unknown protein, similar to nucleic acid binding	18397819	13	1	3	36.4	5.8	*At*
1572	unknown protein, similar to putative 200 kDa antigen p200	42568499	18	2	4	84.0	7.9	*At*
1573	unknown protein, similar to roothairless 1	42562587	17	2	4	100.1	5.6	*At*
1574	Ycf2	91214183	21	2	1	269.2	8.8	*Gm*
1575	ZmGR2c	4240041	20	2	10	29.9	5.2	*Zm*
	13 Unclassified							
1576	At1g17880/F2H15_10	16323127	34	3	22	18.0	6.6	*At*
1577	At3g26750	42565219	14	2	3	57.9	5.0	*At*
1578	AT3g53350/F4P12_50	17978993	19	2	7	45.7	5.7	*At*
1579	AT5g56980/MHM17_10	21592758	18	2	4	42.9	9.2	*At*
1580	ATBRCA1 (BREAST CANCER SUSCEPTIBILITY1)	5262797	21	3	1	166.6	8.9	*At*
1581	AtPIP5K1	3702691	17	2	5	78.4	9.2	*At*
1582	COP1-interacting protein-related	42563167	23	2	3	130.0	6.5	*At*
1583	DC1 domain-containing protein	15232189	16	2	3	77.9	6.2	*At*
1584	ddl5	28542706	54	4	64	5.8	5.3	*Gm*
1585	DGCR14-related	15231871	19	2	2	56.8	9.1	*At*
1586	DNA-binding bromodomain-containing protein	8885596	16	2	3	57.2	9.4	*At*
1587	EST gb|T21788 comes from this gene.	2341027	17	2	4	46.2	8.9	*At*
1588	F15O4.13	8778340	26	3	2	217.3	8.6	*At*
1589	F20N2.6	8778500	16	2	5	62.7	8.5	*At*
1590	F21J9.22	9743341	26	3	3	88.3	5.2	*At*
1591	F22G5.33	8778556	18	2	2	84.9	4.7	*At*
1592	F28C11.19	8778583	16	2	0	362.5	9.5	*At*
1593	F8K7.24	5263333	20	2	3	72.2	5.0	*At*
1594	GAMM1 protein-like	79527307	38	3	6	42.3	5.9	*At*
1595	gene 11-1 protein - like	79476972	16	2	1	222.1	4.3	*At*
1596	gigantea 3	50593496	17	2	2	126.1	6.6	*Ta*
1597	hAT dimerisation domain-containing protein	15230660	18	2	2	68.2	6.8	*At*
1598	IAA23	1711205	16	2	6	64.9	6.5	*At*
1599	integral membrane family protein	15234654	16	2	14	20.3	10.0	*At*
1600	LNG1 (LONGIFOLIA1)	15242342	17	2	3	104.1	9.1	*At*
1601	LSH1 (LIGHT-DEPENDENT SHORT HYPOCOTYLS 1)	11994313	32	3	13	21.4	10.2	*At*
1602	MBD9 (METHYL-CPG-BINDING DOMAIN 9)	6016705	26	3	3	106.0	5.5	*At*
1603	pg1	55792422	17	2	2	102.5	7.2	*Hv*
1604	pollen-specific lysine-rich protein SBgLR	39653215	16	2	14	23.2	4.7	*St*
1605	pore protein of 24 kD (OEP24)	2764574	30	2	9	23.4	9.1	*Ps*

(Table continues on following page.)

(Table continued from previous page.)

Appendix II (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1606	PR10-like protein	18643	16	1	7	16.8	4.9	*Gm*
1607	predicted by genscan and genefinder	4262243	16	2	19	20.3	6.1	*At*
1608	predicted proline-rich protein	7269684	13	1	2	55.6	5.8	*At*
1609	putative protein	4454007	21	3	9	50.4	10.1	*At*
1610	putative protein	7572912	14	2	4	83.4	6.7	*At*
1611	putative protein	6735376	14	1	0	90.9	6.5	*At*
1612	RGH1A	20513867	16	2	3	109.2	6.3	*Hv*
1613	seven transmembrane protein Mlo1	44458502	15	2	3	63.4	9.7	*Zm*
1614	seven transmembrane protein Mlo2	13784977	17	2	3	64.3	9.4	*Zm*
1615	SH3 domain-containing protein 2 (SH3P2)	3096935	15	1	2	43.9	9.1	*At*
1616	Similar to auxin-independent growth promoter	6587842	26	3	5	63.6	9.8	*At*
1617	T10O22.24	8671773	20	2	2	160.2	5.6	*At*
1618	T17H3.8	5668769	14	2	7	44.4	6.5	*At*
1619	TPA: TPA_exp: SAB	39930307	17	2	1	292.9	8.6	*At*
1620	UNE1 (unfertilized embryo sac 1)	22329840	14	2	3	51.4	9.0	*At*
1621	unknown protein	15222278	15	2	13	19.2	9.6	*At*
1622	Unknown protein	13899069	15	1	5	15.8	6.8	*At*
1623	unknown protein	145334205	15	2	7	22.9	7.7	*At*
1624	unknown protein	30678586	17	2	11	31.8	9.7	*At*
1625	unknown protein	16905165	16	2	1	108.4	6.9	*Os*
1626	unknown protein	15239374	16	2	5	40.8	7.2	*At*
1627	unknown protein	15236196	16	2	1	137.4	5.2	*At*
1628	unknown protein	42570797	16	2	20	9.5	10.4	*At*
1629	unknown protein	51970260	25	3	6	79.6	5.8	*At*
1630	unknown protein	15237323	23	2	9	24.5	9.6	*At*
1631	unknown protein	15224752	25	3	8	39.0	7.8	*At*
1632	unknown protein	15229445	23	3	3	130.7	5.9	*At*
1633	unknown protein	15235902	22	3	7	50.7	5.5	*At*
1634	unknown protein	18396354	21	2	9	35.1	6.4	*At*
1635	unknown protein	14192867	18	2	6	50.0	6.1	*Os*
1636	unknown protein	15217540	18	2	1	164.0	5.7	*At*
1637	unknown protein	15234138	18	2	2	86.9	7.4	*At*
1638	unknown protein	15229938	17	2	4	62.2	6.0	*At*
1639	unknown protein	15241545	17	2	7	46.6	10.0	*At*
1640	unknown protein	79364304	17	2	6	58.2	8.2	*At*
1641	unknown protein	15229750	17	2	3	59.2	5.1	*At*
1642	unknown protein	18403592	17	2	7	23.4	8.2	*At*
1643	unknown protein	27311771	20	2	2	127.2	8.3	*At*
1644	unknown protein	15231790	19	2	4	59.2	9.4	*At*
1645	unknown protein	15235638	14	2	5	38.0	5.9	*At*
1646	unknown protein	15222268	14	2	20	6.8	4.5	*At*
1647	unknown protein	48374982	14	2	2	78.5	6.3	*Zm*
1648	unknown protein	15234869	14	2	2	129.2	5.5	*At*
1649	unknown protein	22327950	14	2	2	108.9	5.0	*At*
1650	unknown protein	30683681	14	2	4	124.9	5.5	*At*
1651	unknown protein	42566521	14	2	7	57.2	8.9	*At*
1652	unknown protein	18390717	14	1	2	40.6	8.9	*At*
1653	unknown protein	18399883	14	1	4	31.1	6.5	*At*
1654	unknown protein	15238754	28	3	2	133.4	5.3	*At*
1655	unknown protein	18400785	32	2	9	32.6	6.4	*At*
1656	unknown protein	30686923	36	2	14	22.8	9.2	*At*
1657	unknown protein, contains cyclin-like F-box domain	79532250	18	2	7	37.5	9.2	*At*
1658	unknown protein, contains DNA clamp	79315859	28	4	8	45.0	5.4	*At*
1659	unknown protein, contains domain of unknown function DUF287	42568967	16	2	4	52.9	4.6	*At*
1660	unknown protein, contains domain of unknown function DUF295	18423797	15	2	6	41.2	8.7	*At*
1661	unknown protein, contains domain of unknown function DUF593	15235613	14	2	7	33.1	4.6	*At*
1662	unknown protein, contains DUF543 domain	12323664	14	1	5	32.8	5.7	*At*

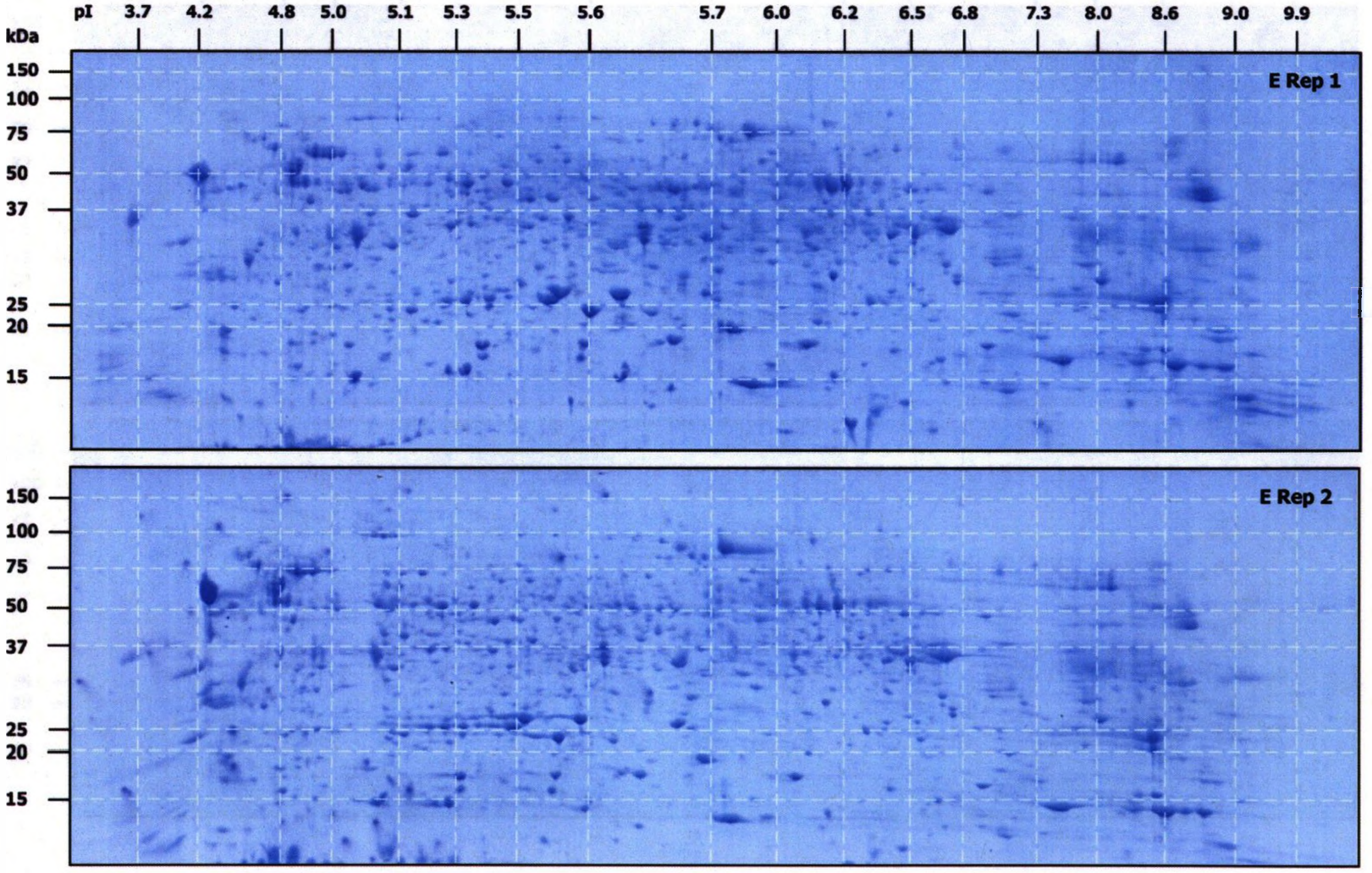
(Table continues on following page.)

(Table continued from previous page.)

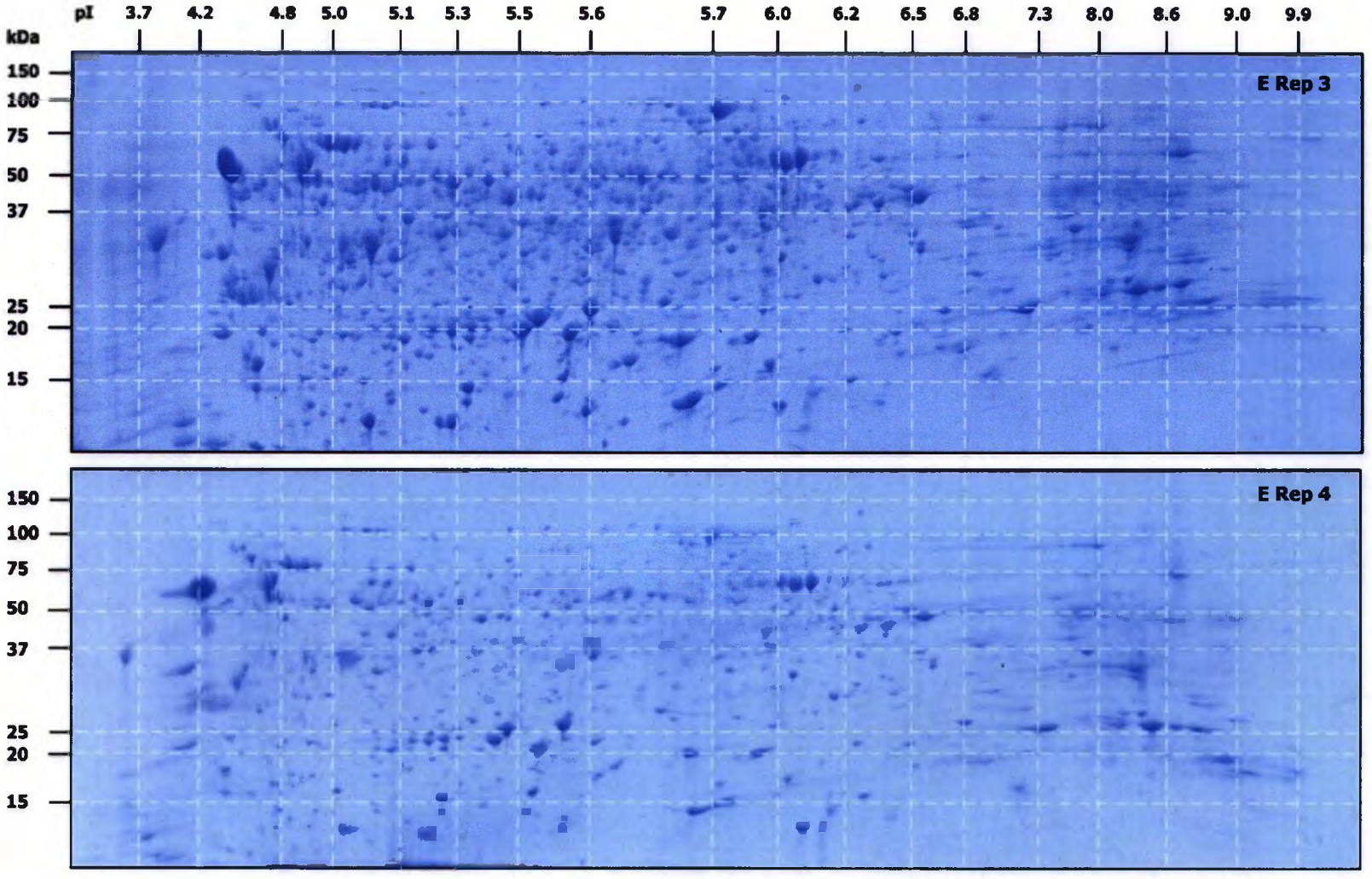
Appendix II. (Continued from previous page)

		NCBI Accession Number	MS/MS Search Score	Distinct Peptides Ident.	Cov. %	Theoretical MW	Theoretical pI	Species
1663	unknown protein, N-terminal protein myristoylation	30693012	23	3	5	69.2	9.6	*At*
1664	unknown protein, plant protein of unknown function (DUF827).	15221217	14	2	6	62.7	5.4	*At*
1665	unknown protein, protein of unknown function DUF266	15221386	15	2	6	47.1	8.7	*At*
1666	unknown protein, protein of unknown function DUF287	15227841	17	2	5	79.1	4.9	*At*
1667	unknown protein, protein of unknown function DUF567	79398116	16	2	18	22.0	7.7	*At*
1668	unknown protein, similar to adenyl cyclase	42565082	26	3	8	44.1	9.1	*At*
1669	unknown protein, similar to alanine transaminase	15220083	16	2	12	12.7	9.5	*At*
1670	unknown protein, similar to Avr9/Cf-9 rapidly elicited protein 75	18399561	18	2	17	17.2	11.1	*At*
1671	unknown Protein, similar to CRM family member 3	10086489	18	2	1	114.5	5.3	*At*
1672	unknown protein, similar to similar to Leucine Rich Repeat protein	15242427	29	4	4	132.6	5.5	*At*
1673	unknown protein, similar to transcription factor	30684987	18	2	4	59.6	9.2	*At*
1675	unknown protein, similar to viral A-type inclusion protein repeat	15229907	18	2	2	73.3	5.1	*At*
1676	unknown protein, uncharacterized conserved protein	18419744	19	2	28	8.4	7.7	*At*
1677	unknown protein, uncharacterized integral membrane protein	12322994	18	2	5	32.0	7.7	*At*
1678	VPS13-like protein	145358407	30	4	2	369.4	6.0	*At*
	20 Secondary metabolism							
	20.1 Phenylpropanoids/phenolics							
1679	2'-hydroxydihydrodaidzein reductase	6573167	68	5	17	36.1	5.7	*Gm*
1680	anthocyanidin synthase	38679407	15	2	7	40.0	5.6	*Gm*
1681	chalcone isomerase	14582263	162	10	56	23.3	6.2	*Gm*
1682	chalcone isomerase 2	51039626	33	2	15	24.7	6.0	*Gm*
1683	cytochrome P450 monooxygenaseCYP93D1	5059126	20	2	4	58.3	8.7	*Gm*
1684	dihydroflavonol-4-reductase 2	121755811	18	1	2	39.5	6.1	*Gm*
1685	flavanone 3-hydroxylase-like protein	21553527	18	2	6	39.4	5.6	*At*
1686	hydroxycinnamoyl transferase	27475616	18	1	2	48.3	5.9	*Nt*
1687	isoflavone reductase homolog 1	6573169	198	11	52	33.9	5.8	*Gm*
1688	isoflavone reductase, putative	15222191	15	1	3	35.6	6.1	*At*
1689	putative NAD(P)H oxidoreductase, isoflavone reductase	19310585	18	1	3	34.1	6.6	*At*
1690	N-hydroxycinnamoyl/benzoyltransferase 4	83853813	14	2	4	52.5	6.5	*Gm*
	20 Secondary metabolism							
	20.2 Terpenoids							
1691	6-deoxocastasterone oxidase	89146804	16	2	4	51.1	9.2	*La*
1692	isopentenyl diphosphate isomerase I	13603406	72	5	17	33.2	6.0	*Nt*
1693	MRN (marneral synthase)	15239009	17	2	8	87.2	6.3	*At*
1694	putative phorbol ester / diacylglycerol binding protein	12323292	18	2	2	166.4	8.4	*At*
1695	terpene cyclase like protein	2982437	14	2	6	66.6	6.0	*At*
	20 Secondary metabolism							
	20.3 Alkaloids							
	20 Secondary metabolism							
	20.4 Non-protein amino acids							
1695	At1g52120	34222072	18	2	2	49.9	5.2	*At*
1696	F1E22.15	6686398	16	2	1	188.5	8.6	*At*
1697	jacalin lectin family protein	15228218	17	2	5	72.5	5.3	*At*
1698	lectin	81238245	18	2	14	22.5	9.6	*Gm*
	20 Secondary metabolism							
	20.5 Amines							
	20 Secondary metabolism							
	20.6 Glucosinolates							
	20 Secondary metabolism							
	20.99 Others							
1700	ANNAT1 (ANNEXIN ARABIDOPSIS 1)	15220216	32	2	8	36.2	5.2	*At*
1701	BGAL8 (beta-glactosidase 8)	7939623	102	7	9	93.2	6.8	*La*
1702	dioxygenase RAMOSUS1	45504725	19	2	3	62.4	6.2	*Ps*
1703	LUE1 protein	118767197	16	2	4	57.3	7.0	*Ta*
1704	myo-inositol-1-phosphate synthase	84311235	364	22	59	56.4	5.4	*Gm*
1705	pyridoxal kinase	68131817	32	2	11	33.7	5.5	*Gm*

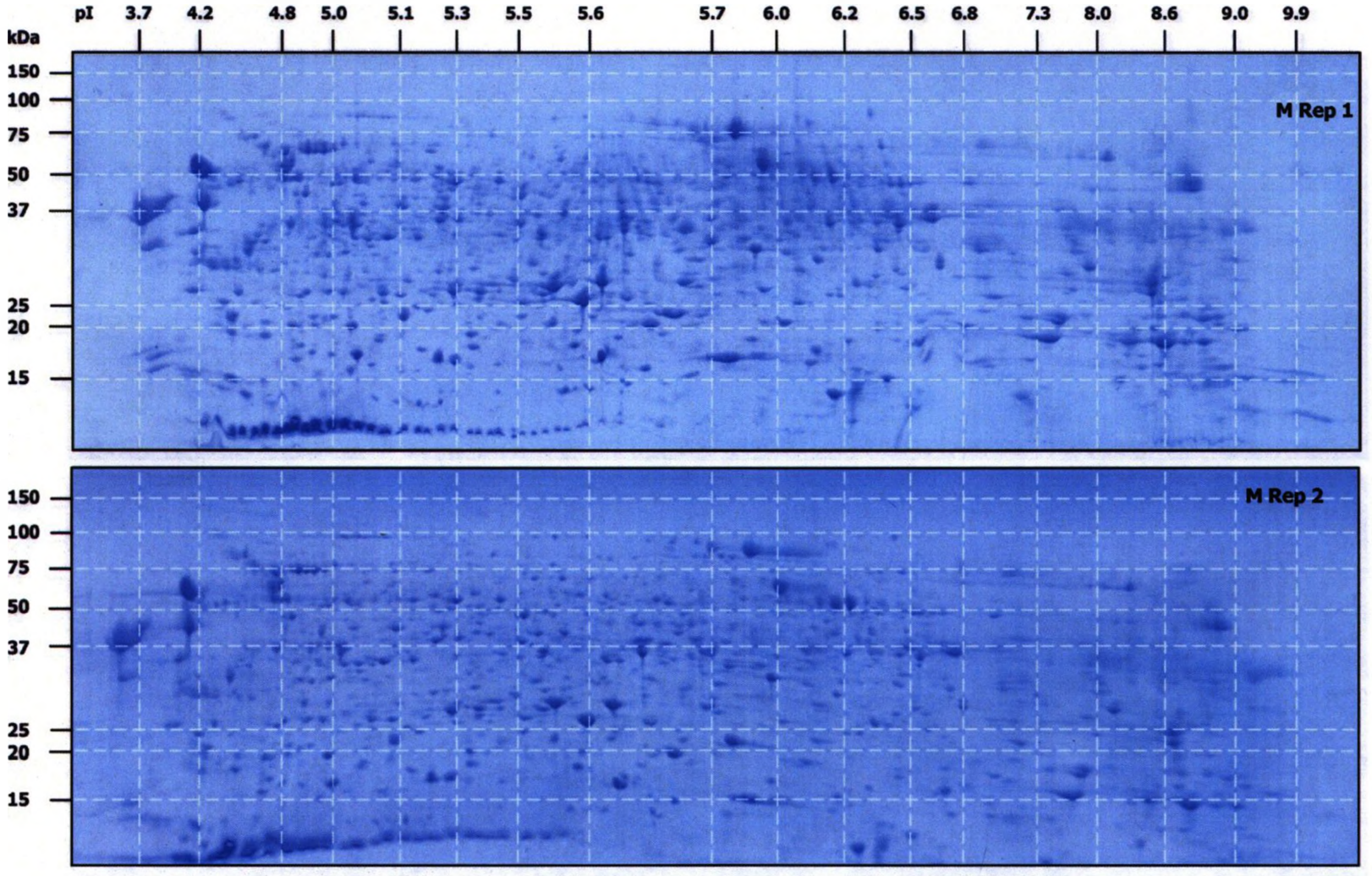
Appendix III. 2D-SDS-PAGE pre-fractionation of seed coat proteins. Four technical replicates of 500 µg of protein extracted from 10-20 DPA (E), 21-35 DPA (M), 80 DPA (Mat) soybean seed coats. E Rep1, M Rep 3 and Mat Rep4 were chosen as reference gels.



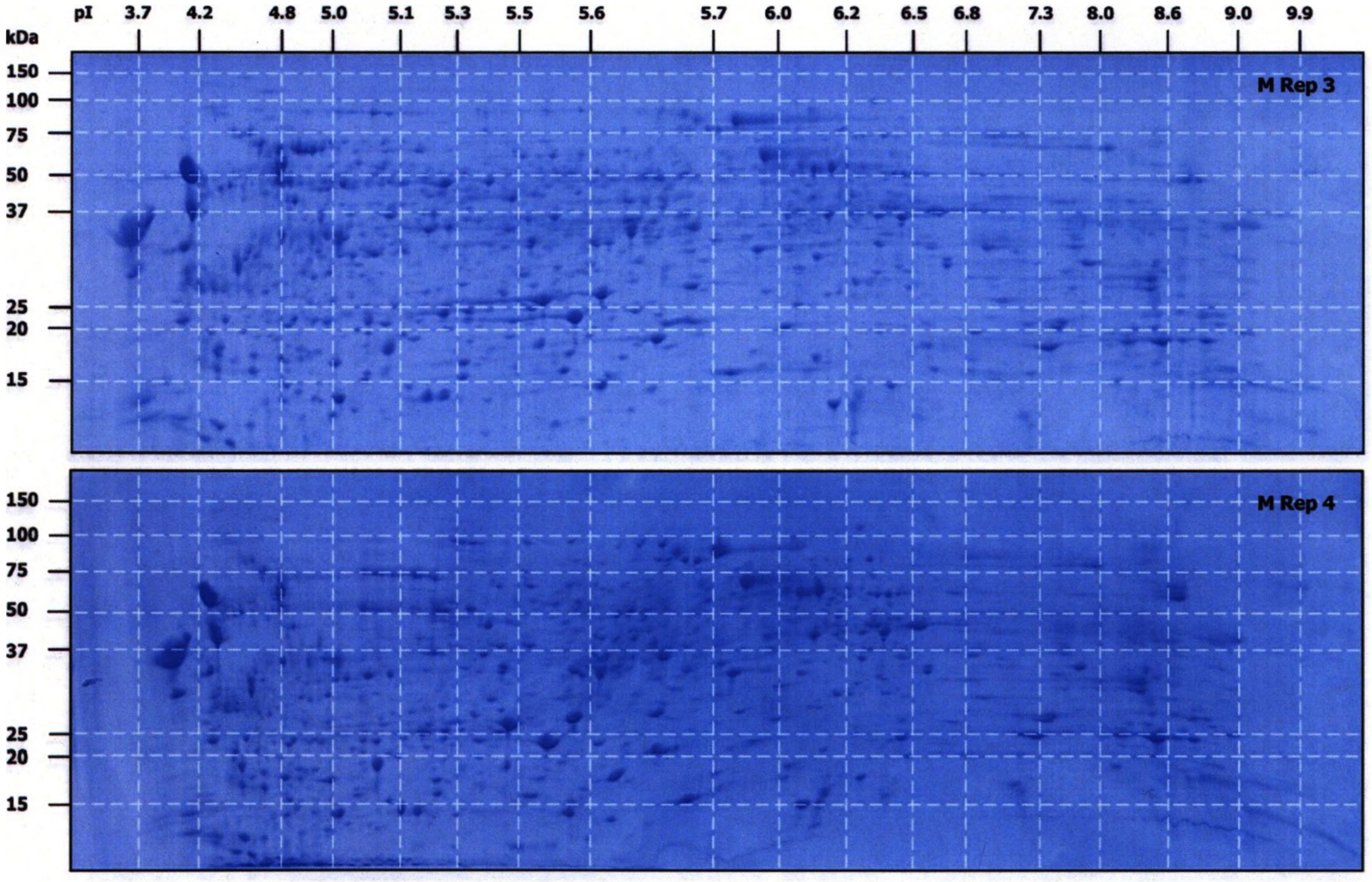
Appendix III. *(Continued from previous page).*



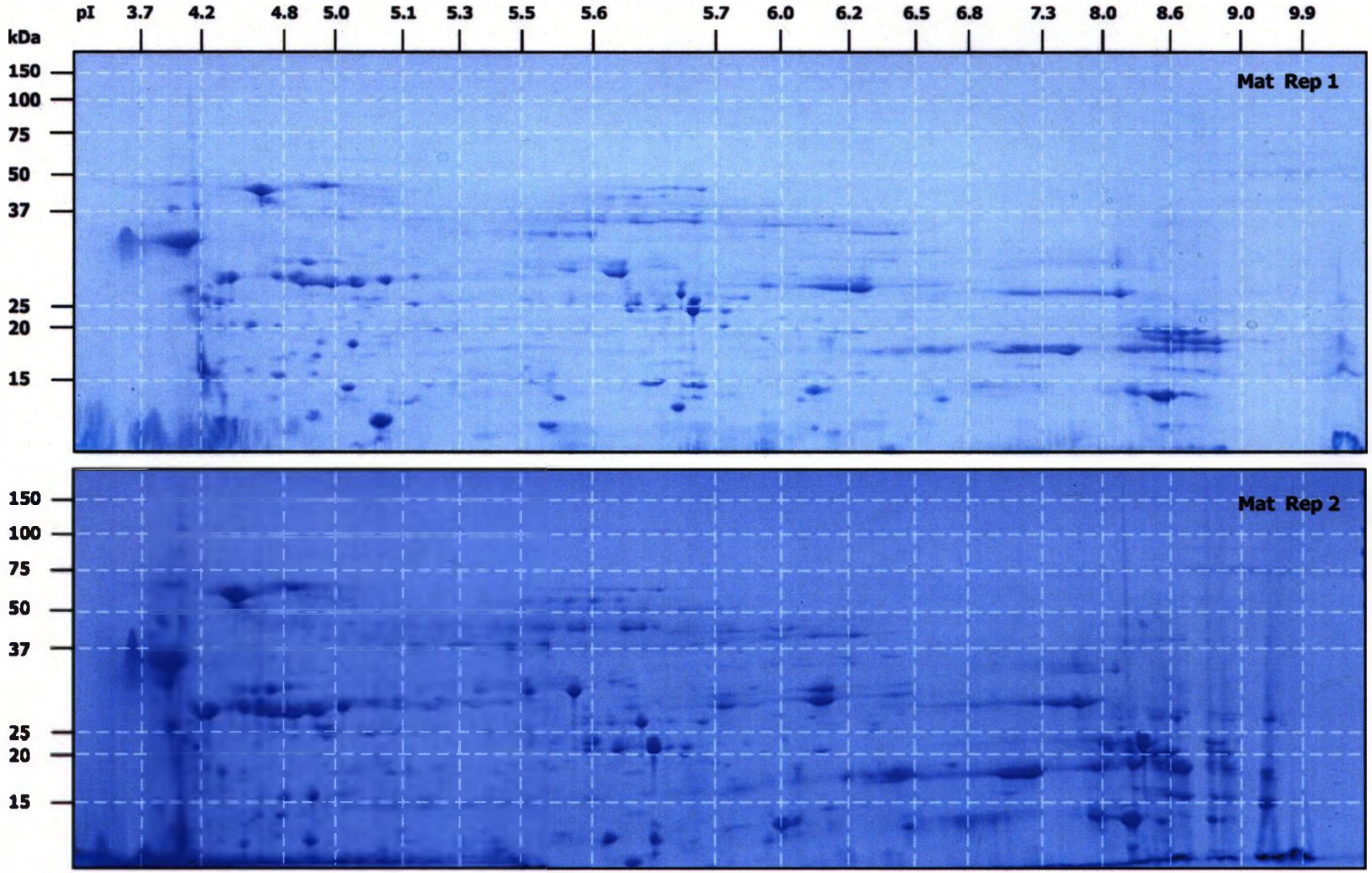
Appendix III. *(Continued from previous page).*



Appendix III. *(Continued from previous page).*



Appendix III. *(Continued from previous page).*



Appendix III. *(Continued from previous page).*