



HAL
open science

Plant cell wall functional genomics: Novelties from proteomics

Rafael Pont-Lezica, Zoran Minic, David Roujol, H el ene San Clemente, Elisabeth Jamet

► **To cite this version:**

Rafael Pont-Lezica, Zoran Minic, David Roujol, H el ene San Clemente, Elisabeth Jamet. Plant cell wall functional genomics: Novelties from proteomics. MA Osborne. Advances in Genetics, volume 1, NOVA Publishers, chapter 10, 2010. hal-00485293

HAL Id: hal-00485293

<https://hal.science/hal-00485293>

Submitted on 20 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

Published in:

Advances in Genetics Research (2010)

Ed. MA Osborne

Nova Science Publishers, Inc.

Hauppauge, NY, USA

**Plant cell wall functional genomics:
Novelties from proteomics**

Rafael Pont-Lezica ^a, Zoran Minic ^b, David Roujol ^a,

Hélène San Clemente ^a, Elisabeth Jamet ^a

^a Surfaces Cellulaires et Signalisation chez les Végétaux, UMR 5546 CNRS - UPS -
Université de Toulouse, Pôle de Biotechnologie Végétale, 24 chemin de Borde-Rouge, BP
42617 Auzeville, 31326 Castanet-Tolosan, France

^b Department of Chemistry, University of Saskatchewan, 110 Science Place, Saskatoon, SK
S7N 5C9, Canada

ABSTRACT

Proteomics has become an important contributor to the knowledge of plant cell wall structure and function by allowing the identification of proteins present in cell walls. This chapter will give an overview on recent development in the cell wall proteomic field. Results from proteomics show some discrepancies when compared to results from transcriptomics obtained on the same organ. It suggests that post-transcriptional regulatory steps involve an important proportion of genes encoding cell wall proteins (CWPs). Proteomics thus complements transcriptomic. The cell wall proteome of *Arabidopsis thaliana* is the most completely described at the moment with about one third of expected CWPs identified. CWPs were grouped in functional classes according to the presence of predicted functional domains to allow a better understanding of main functions in cell walls. The second best-described cell wall proteome is that of *Oryza sativa*. Same functional classes were found, with different compositions reflecting the differences in polysaccharide structure between dicot and monocot cell walls. All these proteomic data were collected in a new publicly accessible database called *WallProtDB* (<http://www.polebio.scsv.ups-tlse.fr/WallProtDB/>). In conclusion, some perspectives in plant cell wall proteomics are discussed.

INTRODUCTION

Several years after the launching of systematic programs of genome sequencing, the challenge of gene function discovery is still enormous, especially in the case of genes encoding cell wall proteins (CWPs). To date, only about 10% of them have a characterized function. Plant cell walls are mainly composed of networks of polysaccharides which represent up to 95% of cell wall mass. After completion of growth, secondary walls reinforce primary walls around cells. Models of primary cell wall structure describe the arrangement of their components into two structurally independent but interacting networks, embedded in a pectin matrix [9, 16]. Cellulose microfibrils and hemicelluloses constitute the first network; the second one is formed by structural proteins among which extensins [38]. CWPs only represent 5 to 10% of the cell wall mass [10], but they are playing many roles especially in polysaccharide network remodelling during plant development and in response to environmental stresses [25].

Many experimental approaches were developed to understand cell wall structure and function. Transcriptomics has greatly contributed to the understanding of gene regulation and to the identification of candidate genes for biogenesis of cell walls, especially secondary walls [19, 39, 55, 73]. Biochemistry of cell wall polysaccharides and proteins has been particularly studied, allowing a good knowledge of the cell wall structure [25]. Proteomics has recently been a newcomer in the field with the first significant results obtained since 2002. Since then, a large number of studies were performed, especially on *Arabidopsis thaliana*, the dicot model plant which genome was the first to be completely sequenced [1]. They led to the identification of about 500 CWPs, representing one third of the expected CWPs [31]. Moreover, improvement of proteomic tools allowed comparative and quantitative studies

between different physiological stages or in response to stresses [13, 68, 75]. Bioinformatics strongly helped in prediction of gene structure and function and allowed the building of sophisticated databases collecting all this information (<http://www.ncbi.nlm.nih.gov/>; <http://www.arabidopsis.org/>). In this chapter, three questions will be addressed: (i) What is the interest of proteomics as compared to transcriptomics? (ii) What is the current picture of *A. thaliana* cell wall proteome? (iii) What is expected from the knowledge of a monocot cell wall proteome?

Is proteomics redundant with regard to transcriptomics?

Transcript profiling is one of the most widespread methods to identify genes involved in a developmental process or in response to environmental changes. Although it is widely accepted that the rate of synthesis and degradation determines transcript abundance, the majority of studies only measure steady state transcript levels, largely due to the technical simplicity with which they are determined [48]. It is generally assumed that a high level of transcripts at a particular physiological stage means that the genes play an important role in the process studied, allowing their selection for in-depth studies. However, gene expression can be regulated at different levels: transcription, post-transcription, translation, post-translation, biological activity of the encoded protein, and its degradation.

The proteome is the full complement of proteins expressed by a genome at a specific time [70]. The development of proteomics during the last 20 years allowed the large scale identification of the working end of the cell: the protein machinery [52]. However, proteomics encounters great difficulties:

- (i) All proteins cannot be extracted by a simple method because they can have different properties like pI ranging from 2 to 12, and hydrophobicity/hydrophilicity.
- (ii) The dynamic range of proteins in the cell may be orders of magnitude different since one protein can be expressed as 10.000 copies and another as 10 copies only. Thus, since there is no polymerase chain reaction (PCR) equivalent for replicating proteins, for the vast majority of proteomic analyses, it is only the most abundant 10% to 20% of proteins which are monitored.
- (iii) The proteome of each living cell is dynamic, altering in response to the individual cell metabolic state and perception of intracellular and extracellular signal molecules.
- (iv) Many proteins undergo post-translational modifications (PTMs) which can interfere with their separation prior to mass spectrometry (MS) analysis, and/or their identification through standard MS protocols.

In this part, we will present results of transcriptomic and proteomic studies performed on the same *A. thaliana* organs to show that both approaches are not redundant, but rather complementary. Mature stems and dark-grown hypocotyls were studied [29, 33, 45, 46]. In the transcriptomic studies, around 3.000 genes coding for proteins predicted to be targeted to the secretory pathway (SPGs, for secretory pathway genes) were selected. Fifty-eight percent of them showed detectable level of transcripts, but only genes with high and moderate level of transcripts (*i.e.* 4- to 64-fold the background level) were compared to the proteomic data. The rationale for this choice was that given the restrictions of proteomics, only the more abundant proteins were identified. If the idea that highly expressed genes produce the most abundant proteins is correct, the results should fit. One hundred ninety-three and 433 SPGs were selected on the basis on their transcript levels in mature stems and etiolated hypocotyls respectively. Among those genes, only 23 (11.9%) and 48 (11%) were also identified in the

respective proteomic studies [29, 45]. This is illustrated in Figure 1A in the case of etiolated hypocotyls. This means that many CWPs escape proteomic analyses, but these results are in the low end of the normal results for proteomics [50]. In addition to the difficulties encountered by proteomics mentioned above, two features are specific to cell wall proteomics: (i) extraction of CWPs from the polysaccharide matrix of the wall can be difficult because they can be insolubilized within the matrix by several types of linkages [6, 62]; (ii) heavily glycosylated proteins are not easily identified. The latter case is illustrated in Table 1 with the arabinogalactan protein (AGP) gene family. These proteins are hydroxyproline-rich glycoproteins (HRGPs) which undergo many PTMs: most Pro residues are hydroxylated and *O*-glycosylated [61], thus preventing their identification by classical methods. Their identification requires a specific deglycosylation procedure using hydrogen fluoride (HF) [64, 74]. Indeed, no AGP has been identified although sixteen AGP genes show significant levels of transcripts in *A. thaliana* etiolated hypocotyls. However, there are some other cases where none of the proposed explanation is valid. As illustrated in Table 1, this is the case of the xyloglucan endotransglucosylase/hydrolase (XTH) gene family (<http://labs.plantbio.cornell.edu/xth/>). Five XTHs were identified in the cell wall proteomic study performed on *A. thaliana* etiolated hypocotyls. There is one case where the protein is detected when the level of transcripts of the gene is below background (AtXTH33), and twelve cases where transcript levels are above background with no protein identified. One has to assume that post-transcriptional events contribute to the regulation of such genes.

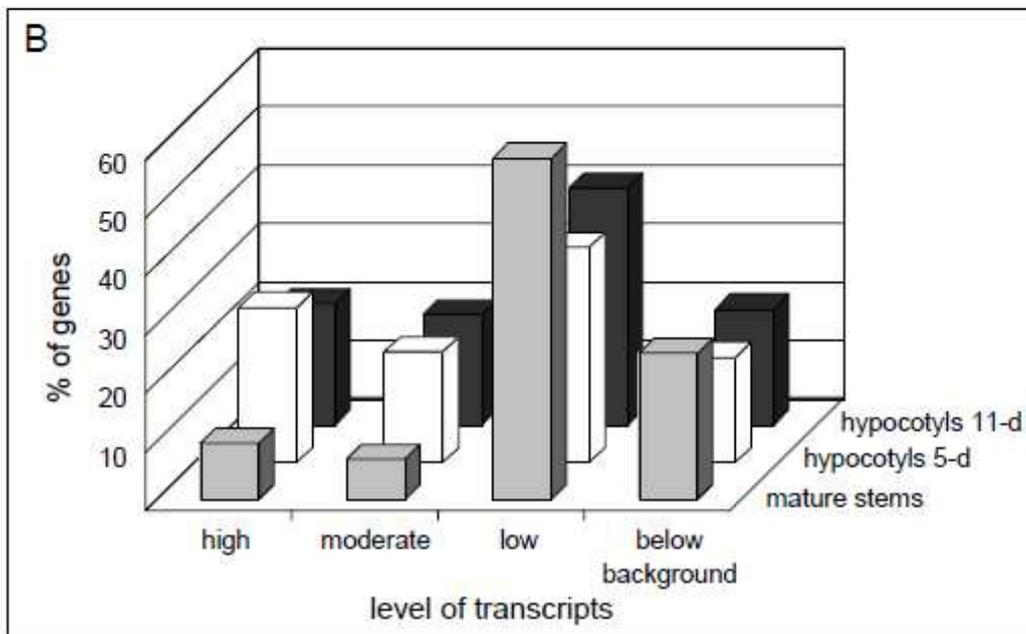
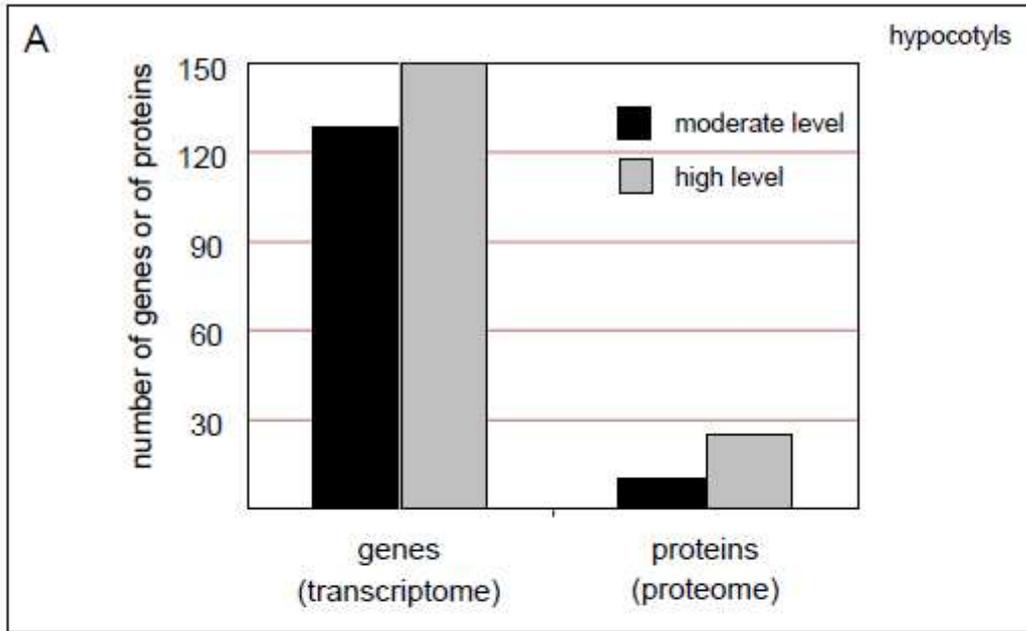


Figure 1. Transcriptomics vs proteomics

A. Distribution of SPGs having moderate or high levels of transcripts in *A. thaliana* etiolated hypocotyls: comparison between results of transcriptomics and proteomics [29, 33]. **B.** Levels of transcripts of *A. thaliana* genes encoding CWPs identified by proteomics in stems [45] and in 5- (5-d) and 11-day-old (11-d) etiolated hypocotyls [29]. Percentages of genes falling in the three following categories are represented: high: corresponds to \log_2 values of the mean signal intensity higher than 10; moderate: values between 9 and 10; low: values between background and 9; values under the background level [33, 46].

Table 1. Comparison of transcriptomic and proteomic data obtained with etiolated hypocotyls of *A. thaliana* for two gene families: XTHs and AGPs.

Transcriptomic data were obtained on CATMA microarrays with 5- and 11-day-old etiolated hypocotyls [33]. They are expressed as \log_2 of mean signal intensity. The background of the experiment was estimated to 6.83. Cell wall proteomic data are from [29]: + means that the protein has been identified; - means that the protein has not been found. Transcriptomic results which are consistent between transcriptomic and proteomic studies are in bold characters.

XTH family	transcriptomics		proteomics		AGP family	transcriptomics	
	5 days	11 days	5 days	11 days		5 days	11 days
AtXTH4	13.27	12.62	+	+	AtAGP9	12.43	12.55
AtXTH31	10.10	8.76	+	-	AtAGP15	12.04	12.70
AtXTH5	9.02	8.63	+	-	AtAGP31	11.98	11.89
AtXTH32	8.82	7.91	-	+	AtAGP12	11.93	12.97
AtXTH33	6.47	6.47	+	+	AtAGP4	11.76	11.86
AtXTH15	13.61	12.31	-	-	AtAGP22	9.98	10.94
AtXTH19	12.37	12.15	-	-	AtAGP1	9.62	9.86
AtXTH30	11.24	11.17	-	-	AtAGP26	9.07	9.62
AtXTH8	10.48	10.34	-	-	AtAGP25	8.68	8.80
AtXTH27	10.42	10.34	-	-	AtAGP18	8.64	8.41
AtXTH7	9.89	8.65	-	-	AtAGP10	8.18	8.88
AtXTH24	9.88	10.14	-	-	AtAGP19	7.94	7.93
AtXTH28	9.72	9.73	-	-	AtAGP5	7.34	7.49
AtXTH16	8.17	7.89	-	-	AtAGP30	6.95	6.57
AtXTH20	7.93	8.12	-	-	AtAGP17	6.94	6.84
AtXTH14	7.47	7.43	-	-	AtAGP41	6.84	6.93
AtXTH10	7.07	7.08	-	-			

Conversely, we looked at the level of transcripts of genes encoding CWPs identified through proteomics in etiolated hypocotyls (137 proteins) [29] and mature stems (87 proteins) [45]. The level of transcripts of some of the genes were not found in the microarray experiments (about 20%) since some have no gene specific tags (GSTs) or were eliminated because of poor signals of hybridization to the RNA probe. The results in Figure 1B are normalized plotting the percentage of genes for each level of transcripts. The big surprise was that most of the CWPs identified by proteomics originate from genes which level of transcripts was low (between 37 and 58%) or below the background (between 18 and 25%). Considering the limitations of proteomics, the identified proteins were expected to be the products of most abundant transcripts, but it was not the case. It suggests that transcripts could have short half-lives and/or that CWPs could have a low turnover. Post-transcriptional regulation seems to be important for more than 56% of CWPs in mature stems and etiolated hypocotyls. Only 16 to 44 % of the identified CWPs are the product of genes with moderate or high expression level, suggesting that transcription is the main regulatory step for that group of proteins.

These results are in agreement with previous ones obtained in yeast [28, 40], *A. thaliana* [34], and *Brassica napus* [75], showing that the quantification of transcripts does not always reflect the actual level of protein. We should consider that most of the genes encoding CWPs have a post-transcriptional regulation. This means that transcriptomics and proteomics are complementary, and not redundant [41].

A summary of present proteomic results obtained on *Arabidopsis thaliana*

Nearly 500 *A. thaliana* CWPs were identified in seventeen studies on different organs using various strategies. A new database (*WallProtDB*) collecting all this information was set up (<http://www.polebio.scsv.ups-tlse.fr/WallProtDB/>). *WallProtDB* allows searching for CWPs identified in the seventeen cell wall proteomes already published, and looking for the presence of proteins or protein families of interest in several proteomes. An exportation format is offered to download results of queries in the Microsoft Office Excel format (<http://www.microsoft.com/france/office/2007/programs/excel/overview.mspix>). Each proteomic study is described through a simplified flowchart showing its different steps from plant material to protein identification. For each proteomic study, the used strategy is highlighted in color. As illustrated in Figure 2, two types of methods can be used to prepare a CWP fraction. Non-destructive methods leave the cells alive and allow elution of CWPs from cell walls using different buffered solutions (*e.g.* [4, 5]). Destructive methods start with tissue grinding thus mixing CWPs and intracellular proteins (*e.g.* [2, 14, 21]). The CWP fraction needs to be fractionated to allow identification of proteins by mass spectrometry (MS). Proteins can be directly submitted to enzymatic digestion with appropriate proteases such as trypsin or to chemical treatment to get peptides of appropriate mass (usually between 750 and 4000 Da). Alternatively, proteins are separated prior to cleavage into peptides. Since most CWPs are basic glycoproteins poorly resolved by bi-dimensional electrophoresis (2D-E), the most efficient ways to separate them are mono dimensional electrophoresis (1D-E) (*e.g.* [5]) or cationic exchange chromatography followed by 1D-E of protein fractions eluted with a salt gradient [29]. Identification of proteins can be done either by peptide sequencing by liquid-chromatography (LC) coupled to MS (LC-MS/MS) or by peptide mass mapping using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS. Bioinformatic analysis

of results is required both for protein identification, prediction of sub-cellular localization, and presence of functional domains [57]. To facilitate the interpretation of data, a database called *ProtAnnDB* dedicated to structural and functional annotation of *A. thaliana* proteins was recently built up [57]. *ProtAnnDB* also provides a link to the NCBI reference protein sequences (RefSeq) containing curated sequences (<http://www.ncbi.nlm.nih.gov/RefSeq/>) [54]. Proteins present in *WallProtDB* were linked to their annotation in *ProtAnnDB*.

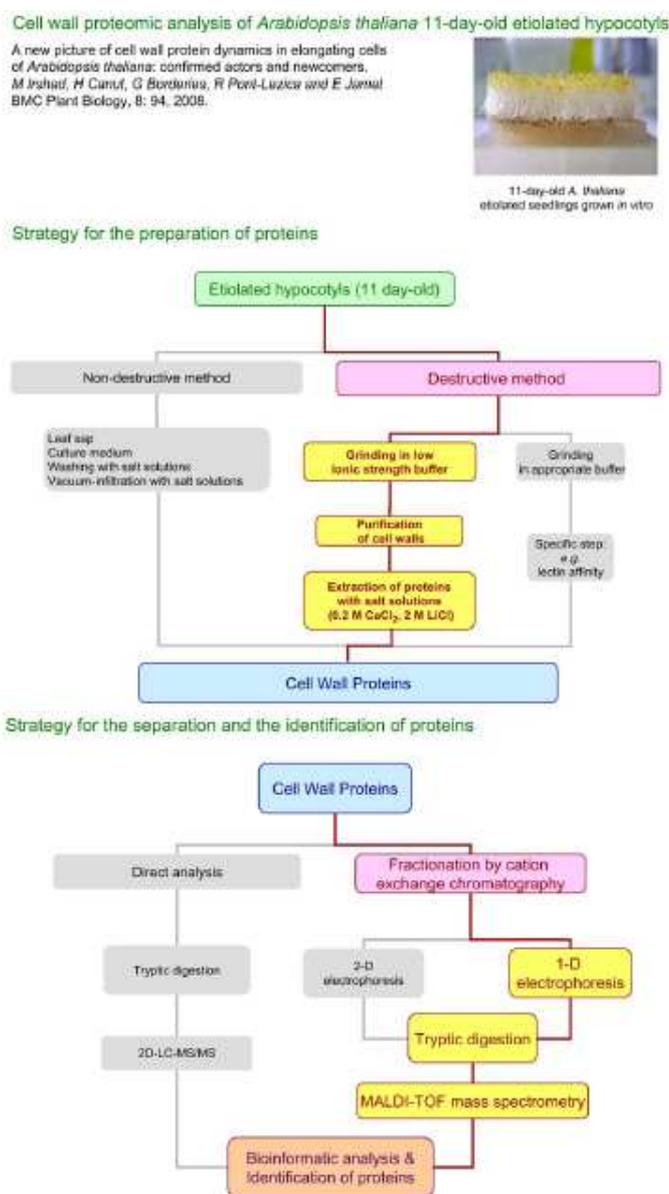


Figure 2. A strategy for cell wall proteomics as illustrated in *WallProtDB* (<http://www.polebio.scsv.ups-tlse.fr/WallProtDB/index.php>).

For each experimental work, a description of the procedure is provided as a simplified flowchart showing its main steps for the preparation of the CWP sample, the separation of CWPs, and their identification using MS. All strategies were put on the same scheme that is customized for each experiment. The example shown is taken from [21], with the used strategy highlighted in color.

The *Arabidopsis* CWPs were classified in nine groups on the basis on their predicted functional domains or known function [31]. The main class is represented by proteins acting on cell wall carbohydrates (26.1%) such as expansins (<http://www.bio.psu.edu/expansins/>), and glycoside hydrolases (GHs) among which xyloglycan endotransglucosylases/hydrolases (XTHs) [56], polygalacturonases (<http://cellwall.genomics.purdue.edu/families/4-3-3.html>), and β -1,4-glucanases (<http://cellwall.genomics.purdue.edu/families/4-3-2-1.html>). These proteins are assumed to modify polysaccharide networks *in muro* during growth and development or in response to environmental constraints. As the second most important group of CWPs (13.9%), oxido-reductases comprise peroxidases [51], multicopper oxidases [30], blue copper binding proteins [49] and berberine bridge oxido-reductases. Proteases (<http://merops.sanger.ac.uk/>) represent 11.6% of the CWPs. They can play roles in protein degradation, protein maturation and peptide signaling [69]. CWPs having interacting domains with polysaccharides and/or with proteins (11.1%) comprise proteins homologous to lectins, proteins with leucine-rich repeats (LRRs) domains, and enzyme inhibitors [17, 35, 67]. Proteins involved in signaling (6.7%) are arabinogalactan proteins (AGPs) [61] and receptor protein kinases which have transmembrane domains and intracellular kinase domains [63]. Considering the high number of CWPs (5.9%) related to lipid metabolism identified in cell wall proteomic studies, this group of proteins was recently introduced in our classification. They are thought to be involved in cuticle synthesis [42, 53]. Proteins of yet unknown function represent 12.2% of CWPs, and can have conserved structural domains or domains common to many proteins called domains of unknown function (DUFs). Such proteins or protein families are also called uncharacterized conserved proteins (UCPs) or uncharacterized protein families (UPFs). Finally, miscellaneous proteins are diverse proteins which cannot be assigned to any of the previous group of CWPs. In particular, they comprise pathogenesis-related proteins (PR-proteins) [18], purple acid phosphatases [37], proteins homologous to

strictosidine synthases [66], and proteins homologous to phosphate induced proteins like EXORDIUM [59]. Although we have now an overview of CWPs present in plant cell walls, the exact function of most of these proteins *in muro* is still unknown.

Why investigating additional cell wall proteomes?

As discussed above, plant cell wall proteomic studies have encountered several limitations. The choice of a strategy for a cell wall proteomic study leads to the selection of sub-proteomes. Elution of proteins around living cells limits the use of drastic treatments to extract CWPs that would be more tightly interacting with cell walls [4]. Conversely, purification of cell walls prior to extraction of CWPs leads to the loss of proteins weakly interacting with cell wall polysaccharides [21]. In *A. thaliana*, two thirds of the predicted CWPs remain to be identified. Additional physiological stages can be studied, as well as interactions with environmental factors. Proteomic strategies can also be customized for specific protein families, e.g. [60]. Cell walls from monocots have only recently started to be analyzed with first papers published in 2008 on *Oryza sativa* and the identification of about 300 CWPs [12, 15, 36]. Among monocots, *O. sativa* is the model plant for cereals, whereas *Brachypodium distachyon* will probably become the model plant for herbaceous grasses [26]. *WallProtDB* was extended to *O. sativa*, thus collecting nearly 800 CWPs from twenty studies. Due to differences in cell wall composition and structure, there is interest in characterizing monocot cell wall proteomes. Differences between available *A. thaliana* and *O. sativa* cell wall proteomes will be discussed mainly for proteins acting on cell wall polysaccharides.

Table 2. CAZy GH gene families and their putative substrates in plant cell walls.

GH families are named according to the CAZy nomenclature (<http://www.cazy.org/>) [8]. Putative substrates *in muro* are given according to [47].

GH family	Putative substrates <i>in muro</i>
GH1	glucan/cellulose xyloglucan
GH3	xylan arabinan arabinoxylan
GH5	glucan/cellulose mannan
GH9	glucan/cellulose
GH10	xylan
GH13	1,4- α -glucosidic linkages
GH16	xyloglucan
GH17	1,3- β -glucan (1,3)(1,4)- β -glucan
GH18	GlcNAc linkages
GH19	GlcNAc linkages
GH27	galactomannan
GH28	homogalacturonan
GH31	xyloglucan
GH35	galactan glycoproteins (AGPs)
GH43	xylan
GH51	arabinoxylan xylan arabinan
GH79	glycoproteins (AGPs)

In a recently published review [47], the families of GHs involved in the modification and/or degradation of cell wall polysaccharides of *A. thaliana* and *O. sativa* were selected on the basis of their classification using the CAZY (<http://www.cazy.org/>) and TAIR (<http://www.arabidopsis.org/>) databases, predicted cellular localization and experimentally-determined substrate specificities in various plants and microorganisms. A total of 200 genes were selected in the *A. thaliana* genome. They belong to thirteen different families (GH 1, 3, 5, 9, 10, 16, 17, 27, 28, 31, 35, 43, and 51). Similarly, 174 genes were selected in the *O. sativa* genome. Putative substrates *in muro* for these enzymes are listed in Table 2. The comparison between the *A. thaliana* and *O. sativa* GH genes shows differences in the number of genes in some GH families. For example, the number of GH28 genes differs significantly between *A. thaliana* and *O. sativa*, particularly for the genes encoding polygalacturonases [44, 72]. This is in agreement with the different amounts of pectins present in the cell walls of these two species. The cell walls of commelinoid monocots, such as *O. sativa*, have a low content in pectins in contrast to those of *A. thaliana*, a dicot, which are rich in this constituent. In addition, dicots contain higher levels of xyloglucans than monocots [20, 24]. The total number of genes encoding GH16 and GH31 which are involved in the hydrolysis of xyloglucans is slightly higher in *A. thaliana* than in *O. sativa*. In contrast, the *O. sativa* genome contains more genes in the GH3 and GH17 families. These GH families are known to be involved in hydrolysis of (1,3)(1,4)- β -D-glucans [23] which are, in addition to glucuronoarabinoxylans, major polysaccharides in monocot cell walls.

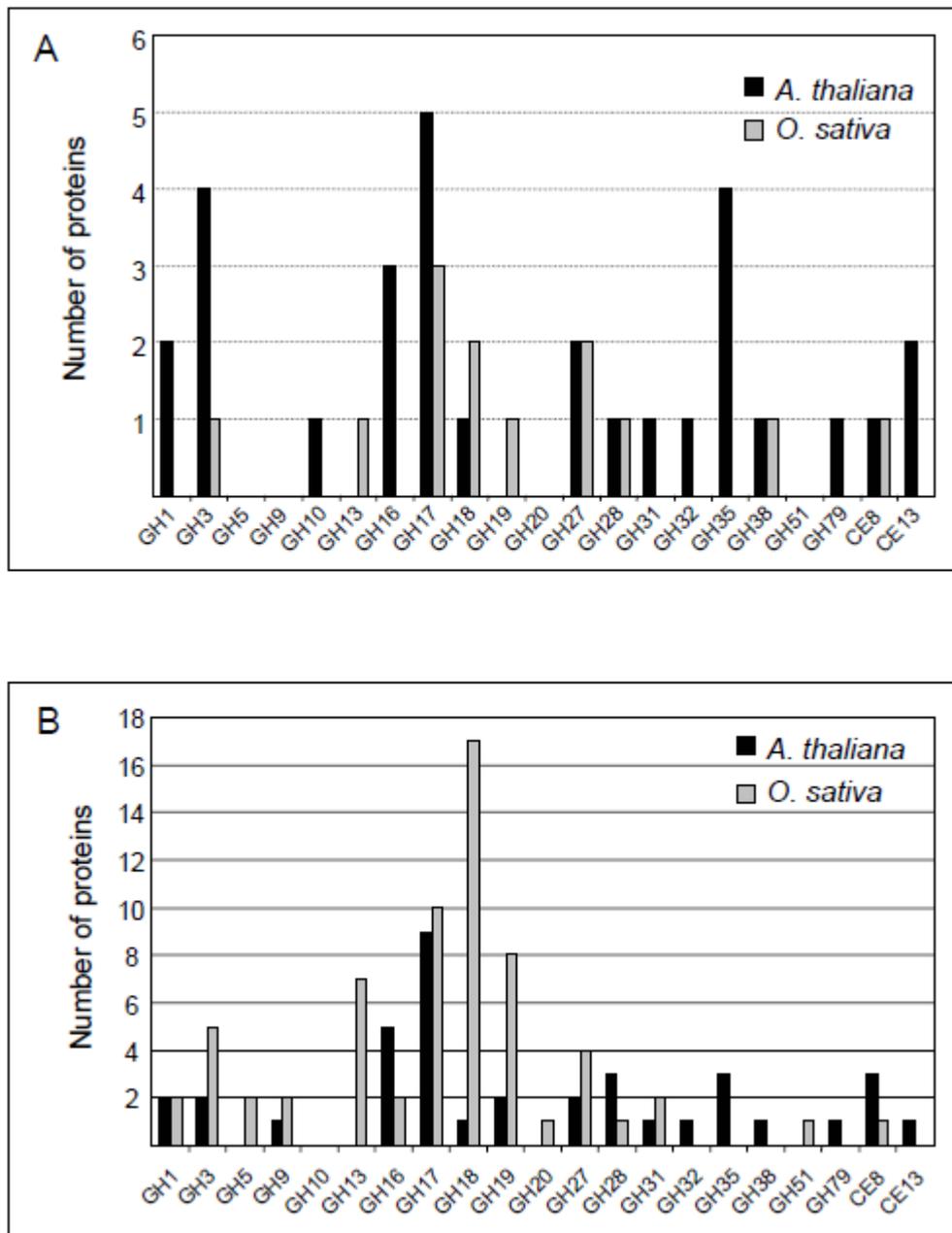


Figure 3. Comparison of *A. thaliana* and *O. sativa* GHs identified by proteomics.

GH families were identified by cell wall proteomics either in leaves (A) or in cell suspension cultures and culture medium of cell suspension cultures (B). GHs families are numbered according to the CAZy database (<http://www.cazy.org/>). The list of proteins was obtained from *WallProtDB* (<http://www.polebio.scsv-ups-tlse.fr/WallProtDB/>).

Comparison of the cell wall proteomic analyses of *O. sativa* [12, 15, 36] and *A. thaliana* [31] leads to the conclusion that proteins acting on polysaccharides constitute the major functional class in both plants. Inside this class, GHs represent the largest number of proteins. Figure 3 presents a comparison between the number of GH families identified by cell wall proteomics in *A. thaliana* and *O. sativa* from leaves, cell suspension culture and culture medium of cell suspension cultures. Although the plant materials and the strategies used for proteomic analyses in *A. thaliana* and *O. sativa* were not wholly the same, these analyses give useful information on the relative abundance of GH families in the cell walls of both plants. For example, two GH5 enzymes were identified in the cell wall of rice. They might be involved in modification of mixed glycan polymers, only found in monocot cell walls [44]. In contrast, a higher number of GH16 enzymes, which are XTHs involved in the modification of the structure of xyloglucans [56], were identified in cell walls of *A. thaliana*. Similarly, a higher number of GH28 and GH35 enzymes were found in cell walls of *A. thaliana*. As mentioned above, such enzymes are assumed to be involved in modification of pectins. In addition, a higher number of carbohydrate esterases (CE8 and CE13) were identified in cell walls of *A. thaliana* cell suspension cultures and leaves respectively. These families comprise respectively pectin methylesterases and pectin acylesterases which modify pectins [43, 71].

Most of the identified *O. sativa* GH families are also found in *A. thaliana*. However, GH13 enzymes, which are predicted to be α -amylases, were only found in *O. sativa* cell walls. α -amylases are endo-amylolytic enzymes which hydrolyze the 1,4- α -glucosidic linkages of starch. They are found in most storage tissues during periods of starch mobilization [3, 22]. Eight of them were predicted to be extracellular proteins (<http://psort.ims.u-tokyo.ac.jp/form.html>). Subcellular localization of α -amylases in *O. sativa* cell suspension cultures revealed that these enzymes are localized in cell walls as well as in starch granules

within amyloplasts [11]. The dual localization of α -amylases in rice is in agreement with the mobilization of starch in the endosperm by secreted amylases, and the mobilization of chloroplastic starch in leaves during the dark periods by chloroplastic α -amylases [65].

Important difference in the number of identified enzymes between *A. thaliana* and *O. sativa* were found for the GH18 and GH19 families. The number of proteins identified in *O. sativa* is much higher than in *A. thaliana*, especially in cell suspension cultures which are stressed cells. GH18 and GH19 are predicted to be chitinase-like enzymes. Chitinases catalyze the hydrolysis of N-acetyl- β -D-glucosaminide 1,4-linkages in chitins and chitodextrins which are not found in plants [58]. Thus, they are assumed to be involved in the protection of plants against pathogens [7, 58]. Hence, chitinases can inhibit fungal growth and kill fungi presumably by degradation of their cell walls made of chitin. Phylogenetic analysis of identified chitinases by proteomic analyses reveals that many of them are predicted to belong to class III, according to their primary structures [36]. Although precise functions of class III chitinases cannot be predicted, it is known that GH18 enzymes of class III also have lysozyme activity. These lysozymes/chitinases show higher activity on bacterial cell walls peptidoglycan (murein) [7]. On the other hand, some class III chitinases, of the TAXI-type (*Triticum aestivum* xylanase inhibitor), were found to be inhibitors of fungal and bacterial xylanases (GH11) in cereals [27]. However, chitinases were shown to be induced in response to abiotic stress and during development. It was suggested that class III chitinases could act on GlcNAc-containing glycolipids or glycoproteins, and be involved in signal transduction [58].

In conclusion, comparative genomic and proteomic analyses between the *A. thaliana* dicot plant, and the *O. sativa* monocot plant, show some similarity in number of cell wall enzymes

involved in cell wall polysaccharide modifications, but also some differences with regard to the presence/absence of some GH families. These features could be related with variations in their cell wall compositions, stages of development, and environmental factors.

Conclusion

Plant cell wall proteomics has brought a new vision of CWPs and cell wall functions. New questions now need to be addressed. A comprehensive understanding of gene regulation requires all steps from gene transcription to protein degradation to be taken into account. Indeed, comparisons of transcriptomics and proteomics data show that the amount of an mRNA is not always strictly correlated with that of the translated protein. Major biological roles for proteolytic activities were only recently demonstrated in maturation of enzymes or production of extracellular peptide signals [69]. But there is still no information about CWP turnover which might be of critical importance in the regulation of extracellular functions. We are still far from the exhaustive description of cell wall proteomes. Additional information on CWPs could be obtained using specific methods for extracting CWPs strongly bound to cell wall components. Efforts should be made to extract CWPs that are physically-linked to cell wall components, such as polysaccharides or lignins [32]. On the other hand, the cell wall proteomes of monocots are only begun to be described. Many new CWPs having predicted enzymatic activities toward cell wall polysaccharides specific to monocots should be found [44]. A complete description of CWP biological functions will require complementary approaches including genetics, biochemistry, and the study of patterns of gene expression. Special attention should be drawn on the numerous CWPs (12.2%) with yet unknown predictable function. Unexpected cell wall functions during plant development and response to environmental factors will certainly arise from such studies.

Acknowledgements

This work was supported by the Université Paul Sabatier in Toulouse and Centre National de la Recherche Scientifique (France). The authors wish to thank the *Plate-forme de Bioinformatique GenoToul Midi-Pyrénées* in Toulouse (France) for providing calculation facilities for the *O. sativa* sequences (<http://www.bioinfo.genotoul.fr>).

References

- 1 Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408, 796-815.
- 2 Bayer, E. M., Bottrill, A. R., Walshaw, J., Vigouroux, M., Naldrett, M. J., Thomas, C. L. and Maule, A. J. (2006). *Arabidopsis* cell wall proteome defined using multidimensional protein identification technology. *Proteomics*, 6, 301-11.
- 3 Beck, E. and Ziegler, P. (1989). Biosynthesis and degradation of starch in higher plants. *Annu Rev Plant Physiol Plant Mol Biol*, 40, 95-117.
- 4 Borderies, G., Jamet, E., Lafitte, C., Rossignol, M., Jauneau, A., Boudart, G., Monsarrat, B., Esquerré-Tugayé, M. T., Boudet, A. and Pont-Lezica, R. (2003). Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: A critical analysis. *Electrophoresis*, 24, 3421-3432.
- 5 Boudart, G., Jamet, E., Rossignol, M., Lafitte, C., Borderies, G., Jauneau, A., Esquerré-Tugayé, M.-T. and Pont-Lezica, R. (2005). Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics. *Proteomics*, 5, 212-221.
- 6 Brady, J. D., Sadler, I. H. and Fry, S. C. (1996). Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: a new potential cross-link. *Biochem J*, 315, 323-7.
- 7 Brunner, F., Stintzi, A., Fritig, B. and Legrand, M. (1998). Substrate specificities of tobacco chitinases. *Plant J*, 14, 225-234.
- 8 Cantarel, B., Coutinho, P., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*, 37, D233-D238.
- 9 Carpita, N. C. and Gibeaut, D. M. (1993). Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J*, 3, 1-30.
- 10 Cassab, G. I. and Varner, J. E. (1988). Cell wall proteins. *Annu Rev Plant Physiol Plant Mol Biol*, 39, 321-353.
- 11 Chen, M., Huang, L., Li, H., Chen, Y. and Yu, S. (2004). Signal peptide-dependent targeting of a rice alpha-amylase and cargo proteins to plastids and extracellular compartments of plant cells. *Plant Physiol*, 135, 1367-1377.
- 12 Chen, X., Kim, S., Cho, W., Rim, Y., Kim, S., Kim, S., Kang, K., Park, Z. and Kim, J. (2008). Proteomics of weakly bound cell wall proteins in rice calli. *J Plant Physiol*, doi:10.1016/j.jplph.2008.09.010.

- 13 Cheng, F.-Y., Blackburn, K., Lin, Y.-M., Goshe, M. and Williamson, J. (2009). Absolute protein quantification by LC/MS for global analysis of salicylic acid-induced plant protein secretion responses. *J Proteome Res*, 8, 82-93.
- 14 Chivasa, S., Ndimba, B. K., Simon, W. J., Robertson, D., Yu, X.-L., Knox, J. P., Bolwell, P. and Slabas, A. R. (2002). Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis*, 23, 1754-1765.
- 15 Cho, W., Chen, X., Chu, H., Rim, Y., Kim, S., Kim, S., Kim, S.-W., Park, Z.-Y. and Kim, J.-Y. (2009). The proteomic analysis of the secretome of rice calli. *Physiol Plant*, 135, 331-341.
- 16 Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nat Rev Mol Cell Biol*, 6, 850-61.
- 17 Di Matteo, A., Federici, L., Mattei, B., Salvi, G., Johnson, K., Savino, C., De Lorenzo, G., Tsernoglou, D. and Cervone, F. (2003). The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc Natl Acad Sci USA*, 100, 10124-10128.
- 18 Edreva, A. (2005). Pathogenesis-related proteins: research progress in the last 15 years. *Gen Appl Plant Phys*, 31, 105-124.
- 19 Ehrling, J., Mattheus, N., Aeschliman, D., Hamberger, B., Cullis, I., Zhuang, J., Kaneda, M., Mansfield, S., Samuels, L., Ritland, K. et al. (2005). Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. *Plant J*, 42, 618-640.
- 20 Farrokhi, N., Burton, R., Brownfield, L., Hrmova, M., Wilson, S., Bacic, A. and Fincher, G. (2006). Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant Biotechnol J*, 4, 145-167.
- 21 Feiz, L., Irshad, M., Pont-Lezica, R. F., Canut, H. and Jamet, E. (2006). Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis hypocotyls*. *Plant Methods*, 2, 10.
- 22 Fincher, G. (1989). Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains. *Annu Rev Plant Physiol Plant Mol Biol*, 40, 305-346.
- 23 Fincher, G. (2009). Exploring the evolution of (1,3;1,4)-beta-D-glucans in plant cell walls: comparative genomics can help! *Curr Opin Plant Biol*, 12, 140-147.
- 24 Fincher, G. (2009). Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. *Plant Physiol*, 149, 27-37.
- 25 Fry, S. C. (2004). Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. *New Phytol*, 161, 641-675.
- 26 Garvin, D., Gu, Y.-Q., Hasterok, R., Hazen, S., Jenkins, G., Mockler, T., Mur, L. and Vogel, J. (2008). Development of genetic and genomic resources for *Brachypodium distachyon*, a new model for grass crop research. *Crop Sci*, 48, S69-S84.
- 27 Gebruers, K., Brijs, K., Courtin, C., Fierens, K., Goesaert, H., Rabijns, A., Raedschelders, G., Robben, J., Sansen, S., Sørensen, J. et al. (2004). Properties of TAXI-type endoxylanase inhibitors. *Biochim Biophys Acta*, 1696, 213-221.
- 28 Gygi, S., Rochon, Y., Franz, B. and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*, 19, 1720-1730.
- 29 Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R. and Jamet, E. (2008). A new picture of cell wall protein dynamics in elongating cells of *Arabidopsis*: confirmed actors and newcomers. *BMC Plant Biol*, 8:94.

- 30 Jacobs, J. and Roe, J. L. (2005). *SKS6*, a multicopper oxidase-like gene, participates in cotyledon vascular patterning during *Arabidopsis thaliana* development. *Planta*, 222, 652-666.
- 31 Jamet, E., Albenne, C., Boudart, G., Irshad, M., Canut, H. and Pont-Lezica, R. (2008). Recent advances in plant cell wall proteomics. *Proteomics*, 8, 893-908.
- 32 Jamet, E., Canut, H., Boudart, G. and Pont-Lezica, R. F. (2006). Cell wall proteins: a new insight through proteomics. *Trends in Plant Sci.*, 11, 33-39.
- 33 Jamet, E., Roujol, D., San Clemente, H., Irshad, M., Soubigou-Taconnat, L., Renou, J.-P. and Pont-Lezica, R. (2009). Cell wall biogenesis of *Arabidopsis thaliana* elongating cells: transcriptomics complements proteomics. (submitted).
- 34 Jones, A. M., Thomas, V., Truman, B., Lilley, K., Mansfield, J. and Grant, M. (2004). Specific changes in the *Arabidopsis* proteome in response to bacterial challenge: differentiating basal and R-gene mediated resistance. *Phytochemistry*, 65, 1805-16.
- 35 Juge, N. (2006). Plant protein inhibitors of cell wall degrading enzymes. *Trends Plant Sci*, 11, 359-367.
- 36 Jung, Y.-H., Jeong, S.-H., Kim, S., Singh, R., Lee, J.-E., Cho, Y.-S., Agrawal, G., Rakwal, R. and Jwa, N.-S. (2008). Systematic secretome analyses of rice leaf and seed callus suspension-cultured cells: Workflow development and establishment of high-density two-dimensional gel reference maps. *J Proteome Res*, 7, 5187-5210.
- 37 Kaida, R., Hayashi, T. and Kaneko, T. (2008). Purple acid phosphatase in the walls of tobacco cells. *Phytochemistry*, 69, 2546-2551.
- 38 Kieliszewski, M. J. and Lamport, D. T. (1994). Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J*, 5, 157-72.
- 39 Ko, J., Han, K., Park, S. and Yang, J. (2004). Plant body weight-induced secondary growth in *Arabidopsis* and its transcription phenotype revealed by whole-transcriptome profiling. *Plant Physiol*, 135, 1069-1083.
- 40 Kolkman, A., Daran-Lapujade, P., Fullaondo, A., Olsthoorn, M., Pronk, J., Slijper, M. and Heck, A. (2006). Proteome analysis of yeast response to various nutrient limitations. *Mol Syst Biol*, 2, 2006.0026.
- 41 Kolkman, A., Daran-Lapujade, P., Fullaondo, A., Olsthoorn, M. M., Pronk, J. T., Slijper, M. and Heck, A. J. (2006). Proteome analysis of yeast response to various nutrient limitations. *Mol. Syst. Biol.*, 2, 2006.0026.
- 42 Kurdyukov, S., Faust, A., Nawrath, C., Bär, S., Voisin, D., Efremova, N., Franke, R., Schreiber, L., Saedler, H., Métraux, J. et al. (2006). The epidermis-specific extracellular BODYGUARD controls cuticle development and morphogenesis in *Arabidopsis*. *Plant Cell*, 18, 321-339.
- 43 Micheli, F. (2001). Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci*, 6, 414-419.
- 44 Minic, Z. (2008). Physiological roles of plant glycoside hydrolases. *Planta*, 227, 723-740.
- 45 Minic, Z., Jamet, E., Negroni, L., der Garabedian, P. A., Zivy, M. and Jouanin, L. (2007). A sub-proteome of *Arabidopsis thaliana* trapped on Concanavalin A is enriched in cell wall glycoside hydrolases. *J Exp Bot*, in press.
- 46 Minic, Z., Jamet, E., San Clemente, H., Pelletier, S., Renou, J.-P., Rihouey, C., Okinyo, D., Proux, C., Lerouge, P. and Jouanin, L. (2009). Transcriptomic analysis of *Arabidopsis* developing stems: a close-up on cell wall genes. *BMC Plant Biol*, 9, 6.
- 47 Minic, Z. and Jouanin, L. (2006). Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol Biochem*, 44, 435-449.

- 48 Narsai, R., Howell, K., Millar, A., O'Toole, N., Small, I. and Whealan, J. (2007). Genome-wide analysis of mRNA decay rates and their determinants in *Arabidopsis thaliana*. *Plant Cell*, 19, 3418-3436.
- 49 Nersissian, A. M. and Shipp, E. L. (2002). Blue copper-binding domains. *Adv Protein Chem*, 60, 271-340.
- 50 Newton, R., Brenton, A., Smith, C. and Dudley, E. (2004). Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry*, 65, 1449-1485.
- 51 Passardi, F., Penel, C. and Dunand, C. (2004). Performing the paradoxical: how plant peroxidases modify the cell wall. *Trends Plant Sci*, 9, 534-540.
- 52 Peck, S. (2005). Update on proteomics in *Arabidopsis*. Where do we go from here? *Plant Physiol*, 138, 591-599.
- 53 Pollard, M., Beisson, F., Li, Y. and Ohlrogge, J. (2008). Building lipid barriers: biosynthesis of cutin and suberin. *Trends Plant Sci*, 13, 236-246.
- 54 Pruitt, K., Tatusova, T. and Maglott, D. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*, 34.
- 55 Ranik, M., Creux, N. and Myburg, A. (2006). Within-tree transcriptome profiling in wood-forming tissues of a fast-growing *Eucalyptus* tree. *Tree physiol*, 26, 365-375.
- 56 Rose, J., Braam, J., Fry, S. and Nishitani, K. (2002). The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. *Plant Cell Physiol*, 43, 1421-1435.
- 57 San Clemente, H., Pont-Lezica, R. and Jamet, E. (2009). Bioinformatics as a tool for assessing the quality of sub-cellular proteomic strategies and inferring functions of proteins: plant cell wall proteomics as a test case. *Bioinform Biol Insights*, 3, 15-28.
- 58 Sasaki, C., Vårum, K., Itoh, Y., Tamoi, M. and Fukamizo, T. (2006). Rice chitinases: sugar recognition specificities of the individual subsites. *Glycobiology*, 16, 1242-1250.
- 59 Schröder, F., Lisso, J., Lange, P. and Müssig, C. (2009). The extracellular EXO protein mediates cell expansion in *Arabidopsis* leaves. *BMC Plant Biol*, 13, 9-20.
- 60 Schultz, C. J., Ferguson, K. L., Lahnstein, J. and Bacic, A. (2004). Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. *J. Biol. Chem.*, 279, 455103-45511.
- 61 Seifert, G. and Roberts, K. (2007). The biology of arabinogalactan proteins. *Annu Rev Plant Biol*, 58, 137-161.
- 62 Shah, K., Penel, C., Gagnon, J. and Dunand, C. (2004). Purification and identification of a Ca²⁺-pectate binding peroxidase from *Arabidopsis* leaves. *Phytochemistry*, 65, 307-312.
- 63 Shiu, S. H. and Bleecker, A. B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci USA*, 98, 10763-8.
- 64 Shpak, E., Leykam, J. and Kieliszewski, M. (1999). Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. *Proc Natl Acad Sci USA*, 21, 14736-14741.
- 65 Smith, A., Zeeman, S. and Smith, S. (2005). Starch degradation. *Annu Rev Plant Biol*, 56, 73-98.
- 66 Sohani, M., Schenk, P., Schultz, C. and Schmidt, O. (2009). Purple acid phosphatase in the walls of tobacco cells. *Plant Biol*, 11, 105-117.
- 67 Spadoni, S., Zabolina, O., Di Matteo, A., Mikkelsen, J., Cervone, F., De Lorenzo, G., Mattei, B. and Bellincampi, D. (2006). Polygalacturonase-inhibiting protein interacts

- with pectin through a binding site formed by four clustered residues of arginine and lysine. *Plant Physiol*, 141, 557-564.
- 68 Thelen, J. and Peck, S. (2007). Quantitative proteomics in plants: choices in abundance. *Plant Cell*, 19, 3339-3346.
- 69 van der Hoorn, R. (2008). Plant proteases: from phenotypes to molecular mechanisms. *Annu Rev Plant Biol*, 59, 191-223.
- 70 Wasinger, V., Cordwell, S., Cerpa-Poljak, A., Yan, J., Gooley, A., Wilkins, M., Duncan, M., Harris, R., Williams, K. and Humphery-Smith, I. (1995). Progress with gene-product mapping of the molliculites: *Mycoplasma genitalium*. *Electrophoresis*, 16, 1090-1094.
- 71 Willats, W., McCartney, L., Mackie, W. and Knox, J. (2001). Pectin: cell biology and prospects for functional analysis. *Plant Mol Biol*, 47, 9-27.
- 72 Yokoyama, R. and Nishitani, K. (2004). Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and *Arabidopsis*. *Plant Cell Physiol*, 45, 1111-1121.
- 73 Yokoyama, R. and Nishitani, K. (2006). Identification and characterization of *Arabidopsis thaliana* genes involved in xylem secondary cell walls. *J Plant Res*, 119, 189-194.
- 74 Zhao, Z., Tan, L., Showalter, A., Lamport, D. and Kieliszewski, M. (2002). Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. *Plant J*, 31, 431-444.
- 75 Zhu, M., Dai, S., McClung, S., Yan, X. and Chen, S. (2009). Functional differentiation of *Brassica napus* guard cells and mesophyll cells revealed by comparative proteomics. *Mol Cell Proteomics*, 8, 752-766.