



Cell Wall Proteome

Georges Boudart, Zoran Minic, Cécile Albenne, Hervé Canut, Elisabeth Jamet, Rafael F Pont-Lezica

► To cite this version:

Georges Boudart, Zoran Minic, Cécile Albenne, Hervé Canut, Elisabeth Jamet, et al.. Cell Wall Proteome. Samaj, J & Thelen, J. Plant Proteomics, Springer, pp.169-185, 2007. hal-00169309

HAL Id: hal-00169309

<https://hal.science/hal-00169309>

Submitted on 3 Sep 2007

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 destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Published in: *Plant Proteomics* (J. Šamaj and J. Thelen (eds.), Springer, pp. 169-185 (2007).

Chapter 12

Cell Wall Proteome

Georges BOUDART¹, Zoran MINIC², Cécile ALBENNE¹, Hervé CANUT¹, Elisabeth JAMET¹, Rafael F. PONT-LEZICA^{1*}

¹ Surfaces Cellulaires et Signalisation chez les Végétaux, UMR 5546 CNRS/Université Paul Sabatier Toulouse III, 24 Chemin de Borde Rouge, BP 42617, 31326-Castanet-Tolosan, France

² Laboratoire de Biochimie des Signaux Régulateurs Cellulaires et Moléculaires, FRE 2621, CNRS, Université Pierre et Marie Curie, 96 Boulevard Raspail, 75006 Paris, France

* Corresponding author

Abstract

In this chapter, we will focus on the contribution of proteomics to the identification and determination of the structure and function of CWPs as well as discussing new perspectives in this area. The great variety of proteins found in the plant cell wall is described. Some families, such as glycoside hydrolases, proteases, lectins, and inhibitors of cell wall modifying enzymes, are discussed in detail. Examples of the use of proteomic techniques to elucidate the structure of various cell wall proteins, especially with post-translational modifications such as *N*-glycosylations, proline hydroxylation and *O*-glycosylations, addition of GPI anchors, and phosphorylation, are given. Finally, the emerging understanding of the functions of cell wall proteins is discussed, as well as proposals for future research.

12.1 Introduction: Cell Wall Proteins Before Proteomics

Several reports between 1924 and 1960 had presented evidence for the presence of proteins in the cell walls of higher plants. However, with the discovery by Lamport and Northcote (1960) of hydroxyproline (Hyp)-containing proteins in the cell walls of bean and sycamore, the concept of a dynamic structure, containing many more proteins than expected, progressively emerged.

Throughout the decade 1985–1995, sequencing and transcription studies of many genes encoding cell wall proteins (CWPs) in different physiological situations have increased our knowledge of these genes, particularly those encoding structural proteins, e.g. Hyp-rich glycoproteins (HRGPs), Hyp/Pro-rich proteins (H/PRPs), and Gly-rich proteins (GRPs) (Cassab 1998; Ringli et al. 2001). Such proteins are unusually rich in a few specific amino acids, contain highly repetitive sequence domains, and can be highly glycosylated. During the past 5 years, the availability of the complete sequence of the *Arabidopsis thaliana* genome has contributed greatly to the description of other well known CWP families such as endo-1,4- α -D-glucanases (del Campillo 1999), glycoside hydrolase (GH) family 1 (Xu et al. 2004), xyloglucan endotransglucosylase/hydrolases (XTHs) (Fry 2004), pectin methylesterases (PMEs) (Micheli

2001), expansins (Cosgrove et al. 2002), peroxidases (Welinder et al. 2002), leucinerich repeat-extensins (LRXs) (Baumberger et al. 2003), and arabinogalactan proteins (AGPs) (Gaspar et al. 2001). It was found that most of the known CWPs are encoded by multigene families, which can be very large – up to more than 70 members.

Recently, the emergence of proteomics has been supported by improvements in mass spectrometry (MS) techniques, the complete sequencing of the *A. thaliana* genome, and new bioinformatic tools. A new vision of the cell wall proteome is thus progressively emerging, with the description of new CWP families as well as the characterisation of post-translational modifications (PTMs). In this chapter, we will focus on the contribution of proteomics to the identification and determination of the structure and function of CWPs as well as discussing new perspectives in this area.

12.2 The Surprising Variety of Cell Wall Proteins

Cell wall proteomics is faced with several difficulties, linked mainly to the structure of cell walls as well as to the nature of CWPs (most CWPs have a basic pI, and can be heavily glycosylated). These facts make extraction, separation, and identification procedures challenging (Jamet et al. 2006). Nevertheless, despite this complexity, cell wall proteomics has become a dynamic discipline. Most of the available data has been obtained with *Arabidopsis*. At present, 365 *Arabidopsis* CWPs have been identified from various organs and other sources such as rosettes (Boudart et al. 2005), hypocotyls (Feiz et al. 2006), stems at the flowering stage (Minic et al. 2007), cell suspension cultures (Robertson et al. 1997; Chivasa et al. 2002; Borderies et al. 2003; Bayer et al. 2006), and cell wallregenerating protoplasts (Kwon et al. 2005). Pieces of information are also now beginning to come from other species such as *Zea mays* and *Medicago sativa* (Zhu et al. 2006; Watson et al. 2004).

The great majority of *Arabidopsis* CWPs (88.5%) can be distributed into seven classes based on the presence of functional domains as predicted by bioinformatics (Table 12.1). The remaining 11.5% consists of proteins with unknown function. As expected, proteins acting on carbohydrates constitute the main class (27.4% of identified CWPs), with GHs as the major representatives (20.5%). Carbohydrate esterases/lyases, expansins and glycoside transferases are less abundant. Two functional classes are of equal importance: (1) oxido-reductases (13.7%), include peroxidases, multicopper oxidases, berberine-bridge enzyme (S)-reticulon: oxygen oxido-reductases, laccases and germins; and (2) proteases (12.6%) of several types, i.e. serine proteases (subtilases), aspartic proteases, cysteine proteases and serine carboxypeptidases. Proteins containing interacting domains (8.5%) include carbohydrate- binding proteins (lectins), and protein-binding proteins through leucinerich repeat (LRR) domains (e.g. polygalacturonase inhibitors), or enzyme inhibitors (e.g. PME and protease inhibitors). Proteins involved in signalling (8.2%) are mainly AGPs and LRR-receptor protein kinases that have been identified through their extracellular LRR domains. Structural proteins are poorly represented (1.9%). Other proteins of various functions (16.2%) are lumped together in a functional class called “miscellaneous”. These include proteins awaiting additional experimental data to allow them to be more precisely classified, such as proteins homologous to acid phosphatases, blue copper binding proteins and proteins with lipase/acyl hydrolase domains. The final class encompasses proteins of unknown function (11.5%) that contain no previously characterised functional domains; some contain common “domains of unknown function” that can be found in databases, and some are plant-specific proteins.

Table 12.1 Functional classification of CWP identified by proteomic approaches in *Arabidopsis*.

Functional classes ^a Subclasses	Number of proteins	% of identified proteins
Proteins acting on carbohydrates	100	27.4
Glycoside hydrolases	75	20.5
Esterases and lyases	17	4.7
Expansins	8	2.2
Oxido-reductases	50	13.7
Peroxidases	23	6.3
Others	27	7.4
Proteases	46	12.6
Proteins with interacting domains	31	8.5
Carbohydrate-binding	9	2.5
Protein-binding	22	6.0
Signaling	30	8.2
Structural proteins	7	1.9
Miscellaneous	59	16.2
Unknown function	42	11.5

^a Proteins are distributed in functional classes according to the presence of functional domains. Data originate from cell suspension cultures (Robertson et al. 1997, Chivasa et al. 2002, Borderies et al. 2003, Bayer et al. 2006); rosettes (Boudart et al. 2005); hypocotyls (Feiz et al. 2006); stems at the flowering stage (Minic et al., 2007); and cell wall-regenerating protoplasts (Kwon et al. 2005).

Bioinformatic analysis of the *Arabidopsis* genome has led to the identification of many gene families encoding CWPs. Table 12.2 presents a comparison between the number of genes predicted to encode CWPs from several families annotated by experts, and the present number of identified proteins. Although this comparison does not give an exhaustive view of the cell wall proteome, it shows that about one-quarter of predicted CWPs have been identified. Based on this figure, we can estimate that around 1,500 genes in the *Arabidopsis* genome code for CWPs. A great disparity can be observed between percentages of identified proteins in each family. Eighty-two percent of the predicted cell wall lectins (legume lectins and monocot mannose-binding lectins) have been identified. In contrast, among structural proteins, not a single extensin has yet been identified by proteomics. There are several likely explanations: extensins are cross-linked into a protein network by covalent bonds (Brady et al. 1996), their high level of glycosylation can prevent accessibility to proteolysis prior to MS analysis, and glycosylated peptides are not identified by the bioinformatic tools used in proteomics. In addition to structural CWPs, low abundant, low molecular mass, short half-life and highly tissue-specific CWPs still escape identification.

Table 12.2 Number of predicted genes encoding CWP compared to the number of proteins identified by proteomics in *Arabidopsis*.

Protein families	N° of genes predicted	N° of identified proteins	% identified proteins	References
Glycoside hydrolases family 1	26	6	23.1	Xu et al. 2004
Cellulases	25	2	6.0	del Campillo 1999
Polygalacturonases	69	8	11.6	http://afmb.cnrs-mrs.fr/CAZY/
Pectin methylesterases	67	10	14.9	http://afmb.cnrs-mrs.fr/CAZY/
Pectin acetylsterases	11	3	27.3	http://afmb.cnrs-mrs.fr/CAZY/
XTH	33	7	21.2	http://labs.plantbio.cornell.edu/xth/genes.htm
Expansins	36	8	22.2	http://www.bio.psu.edu/expansins/arabidopsis.htm
Legume and mannose-binding lectins	11	9	81.8	this chapter ^a
Blue copper binding proteins (plantacyanins, stellacyanins, uclacyanins, early nodulins, plastocyanins)	42	7	16.7	Nersissian and Shipp (2002)
Extensins	19	0	0.0	this chapter ^a
LRX	11	4	36.4	Baumberger et al. (2003)
Subtilases	50	18	36.0	Rautengarten et al. (2005)
Fasciclin-like arabinogalactan proteins	13	5	38.5	Schultz et al. (2002)
AGP	26	12	46.2	Schultz et al. (2002)
SKS (SKU-similar)	19	8	42.1	Jacobs and Roe (2005)
Peroxidases (class III)	73	23	31.5	Welinder et al. (2002)
Germin-like proteins	25	7	28.0	this chapter ^a
Total	556	137	24.6	

^aInterProScan searches for motifs or profiles stored in databases (<http://www.ebi.ac.uk/InterProScan/>).

12.2.1 Glycoside Hydrolases

A large proportion of proteins identified by cell wall proteomic analyses act on carbohydrates. These includes GHs, which have been classified according to CAZy nomenclature based on sequence homology (<http://afmb.cnrs-mrs.fr/CAZY/>). The genome of *Arabidopsis* contains 379 genes for GHs from 29 families (Henrissat et al. 2001). Plant GHs are supposed to play a role in various functions in cell wall metabolism, plant defence, signalling, mobilisation of storage reserves (Henrissat et al. 2001) and reorganisation of glycans (Minic et al. 2007). They participate in the regulation of cell wall expansion and alteration during development (Fry 2004; Minic and Jouanin 2006). However, the physiological and biochemical functions of only a small proportion of identified GHs have been demonstrated experimentally. Plant cell wall polysaccharides are very heterogeneous complex polymers. Consequently, spectra of GH activities are very diverse. The 75 GHs identified by proteomics can be sub-divided into four groups. The first group of GHs comprises enzymes that might be involved in reorganisation of cell wall carbohydrates during growth and development (Minic and Jouanin 2006). Of these, 42 proteins belonging to 10 different families were identified by proteomic analyses (Table 12.3). Their possible *in muro* substrates have also been indicated. For the majority of these enzymes, the substrates are pectins (homogalacturonan and rhamnogalacturonan I) and xyloglucans. Interestingly, these two types of polysaccharide are soluble in water, in contrast to cellulose, xylan and galactomannan, which are not. The majority of these enzymes are exo-GHs. A lower number of enzymes are endo-GHs belonging to GH families 9, 10, 16 and 28, which act on cellulose, xylan, xyloglucans and homogalacturonan, respectively. These results suggest that xyloglucans and pectins might undergo important structural changes after their deposition in cell walls. As shown in Table 12.3, some GH families have a broad substrate range. This low specificity has been reported for several purified plant cell wall GHs (Lee et al. 2003; Minic et al.

2006). It was hypothesised that the multifunctionality of plant GHs allows effective modification of complex cell wall carbohydrates without requiring a large number of enzymes.

The second largest group of identified GHs could be involved in defence against pathogens. Chitinases (GH18, GH19) and β -1,3-glucanases (GH17) have been shown to possess antifungal activity (Schlumbaum et al. 1986). Cell walls of various fungi contain chitin and β -1,3- or β -1,6-glucan. Chitinases and β -1,3- glucanases produced by plants can be secreted into intercellular spaces and inhibit fungal growth by degrading their cell wall (Jach et al. 1995; Zhu et al. 1994). Some chitinases (GH18, GH19) also show lysozyme activity hydrolysing β -1,4 linkages between GlcNAc and *N*-acetylmuramic acid (Boller et al. 1983; Brunner et al. 1998). These lysozymes/chitinases show higher activity on bacterial cell wall peptidoglycan.

Table 12.3 Families of Arabidopsis glycoside hydrolases (GHs) identified by cell wall proteomics which are potentially involved in reorganisation of cell wall carbohydrates"

CAZy GH family ^a	Predicted enzymatic activity	Cell wall carbohydrates as possible substrates <i>in muro</i>	Number of proteins
1	β -D-glucosidase	glucan, xyloglucan, cellulose	5
3	β -D-xylosidase α -L-arabinofuranosidase	arabinan, xylan, arabinoxylan	6
9	endo-1,4- β -glucanase	1,4- β -glucan (cellulose)	2
10	1,4- β -xylan endohydrolase	xylan	1
16	xyloglucan endotransglycosylase (XTH)	xyloglucan	7
27	α -D-galactosidase	galactomannan	3
28	endo-polygalacturonase exo-polygalacturonase	homogalacturonan	8
31	α -D-xylosidase α -D-glucosidase	xyloglucan	2
35	β -D-galactosidase	galactan	7
51	α -L-arabinofuranosidase	arabinoxylan, xylan, arabinan	1

^a GH have been classified according to CAZy (<http://afmb.cnrs-mrs.fr/CAZY/>). Predicted enzymatic activities, possible substrates *in muro* and numbers of identified proteins in each family are indicated. Data originate from cell suspension cultures as in Table I.

A third group of GHs could be involved in glycoprotein PTMs. Such enzymes include α -L-arabinofuranosidases (GH3), chitinases (GH18, GH19), β -D-galactosidases (GH35), β -D-mannosidases (GH38), and β -D-glucuronidases (GH79). The role of GHs in hydrolysis of carbohydrate moieties of AGPs has been particularly well studied. Degradation of AGPs seems to occur by the concerted action of several glycosidases. A bifunctional α -L-arabinofuranosidase/ β -D-xylosidase (GH3) and a β -D-galactosidase (GH35) from immature seeds of radish (*Raphanus*

sativus L.) were shown to be involved in the hydrolysis of carbohydrate moieties of AGP (Kotake et al. 2005). Recently, a β -D-glucuronidase from *A. thaliana* (AtGUS, GH79) was shown to hydrolyse glucuronic acid from carbohydrate chains of AGPs (Minic et al. 2007).

Finally, three GHs identified by proteomics could be involved in the mobilisation of storage reserves. In higher plants, sucrose is one of the predominant initial products of photosynthesis and serves as an intermediate storage system as an alternative to starch. In addition, sucrose is the major carbohydrate translocator, osmoticum, and regulator of gene expression, as well as playing a role in cellular signalling. Use of sucrose depends on its cleavage into glucose and fructose. In plants, this enzymatic reaction is performed by invertases (GH31), which are also called α -D-fructofuranoside- fructohydrolases (Roitsh and Gonzalez 2004). Invertase isozymes are distinguished by their subcellular localisation, such as cell wall, vacuole and cytoplasm. Cell wall invertases are considered to be key enzymes involved in sucrose unloading, cell differentiation, and the response to wounding or pathogen attack.

12.2.2 Proteases

Proteases are necessary for protein turnover, maturation of enzymes, and defence against pathogens (Schaller 2004). Consequently, proteases may play a role in CWP turnover – still a poorly understood process. Previous studies localised protease activities to leaf intercellular fluids (Huangpu and Graham 1995). Results from proteomics have revealed the existence of a great diversity of proteases, such as subtilases, carboxypeptidases, and aspartyl- and cysteine-proteases (Jamet et al. 2006). The subtilase family – the best represented (40%) among the cell wall proteases identified by proteomics – is particularly interesting because in mammals, subtilases are involved in the formation of peptide hormones and growth factors from precursor polypeptides (Rautengarten et al. 2005). Three main functions are proposed for subtilases: control of development, protein turnover, and as components of signalling cascades. Proteases certainly play key roles in the maturation of CWPs and in the generation of active peptides in the cell wall. For example, using MS techniques, the barley ARAI α -L-arabinofuranosidase/ β -D-xylosidase was shown to be processed *in vivo* by removing about 130 amino acids at its C-terminus (Lee et al. 2003). On the other hand, a role(s) for plant peptides as peptide hormones in intercellular signalling has recently been proposed (Matsubayashi and Sakagami 2006). The group of *CLE* (CLV3/ESR-related) genes encode small, basic, secreted proteins with a conserved stretch of 14 amino acids close to their C-termini that are found only in plants. Interestingly, the 14-amino-acid peptides released by N- and C- terminal processing harbour the biological activity of these proteins (Ito et al. 2006; Kondo et al. 2006).

12.2.3 Lectins

Based on their structure, lectins so far identified in cell walls belong to two out of the seven groups of plant lectins, namely legume lectins and monocot mannose-binding lectins (van Damme et al. 1998). Around 80% of the predicted cell wall lectins from *Arabidopsis* were identified by proteomics (Table 12.2). One mannose-binding lectin has been identified in almost all tissues analyzed. In contrast, a legume-like lectin was only found in the stems of flowering plants. The characteristic feature of lectins is their ability to recognise and bind specific carbohydrates, thus serving as translators of the sugar code. These recognition functions include lectin involvement in interactions with cells or extracellular materials from the same organism

(which could be considered as recognition of self, or of endogenous ligands) and interactions with foreign particles or cells (recognition of non-self, or of exogenous ligands). The functions of plant lectins remain elusive, but a role in protection and symbiosis has been proposed (Sharon and Lis 2004). Cell wall lectins are probably involved in recognition and defence against pathogens, but it cannot be ruled out that cell wall lectins are also involved in the assembly of cell wall polysaccharides.

12.2.4 Protein Inhibitors of Cell Wall Modifying Enzymes

Inhibitors of cell wall modifying enzymes (CWMEIs) represent an expanding family of plant proteins that presumably play a role in defence by limiting the rate of degradation of the cell wall by microbial enzymes, thus reducing colonisation of plant tissues by pathogens. Several *Arabidopsis* CWMEIs, such as inhibitors of polygalacturonases (PGIPs), PMEIs and xyloglucanases, have been identified by proteomics.

PGIPs are widespread in the plant kingdom (Gomathi and Gnanamanickam 2004). They inhibit only fungal polygalacturonases (PGs), and the degree of susceptibility of the PG depends on the mode of cleavage of the substrate. Many plants possess more than one PGIP with different abilities to inhibit pathogen PG. Overexpression of *PGIP* genes in plants resulted in reduced susceptibility to a fungal pathogen. PGIPs belong to the superfamily of LRR-proteins. The LRR motif has been shown to be essential for affinity and specificity of PGIP for the PG ligand. Interestingly, bean PGIP 2 interacts with polygalacturonic acid via a motif of four clustered Arg and Lys residues located in the vicinity of the PG-binding site, suggesting that PGIPs can be immobilised on the PG substrate (Spadoni et al. 2006).

Pectins are demethylated to polygalacturonic acid by PME, thus favouring further cleavage of the acidic polygalacturonic chains by polygalacturonase. PME can be inhibited by a protein inhibitor (PMEI). Unlike in the case of PGIP, PMEIs exclusively inhibit plant PMEs (Bellincampi et al. 2004). Although *Arabidopsis* PMEI shares strong sequence and structural homology with an inhibitor of tobacco invertase (Nt-CIF), each inhibitor recognises different target enzymes, as recently elucidated by crystallographic analysis (Hothorn et al. 2004).

Inhibitors of hemicellulose-degrading enzymes, such as monocot endoxylanase inhibitors (XIP- and TAXI-type) and dicot xyloglucan endoglucanase inhibiting proteins (XEGIPs), are another interesting family of CWMEIs (Juge 2006). XIP and TAXI are specific inhibitors of microbial endoxylanases but do not inhibit plant endoxylanases. The microbial endoxylanases GH10 and GH11 are inhibited by XIP-I from wheat. The crystal structure of XIP-I in complex with xylanases revealed that the inhibitor possesses two independent enzyme-binding sites, an unusual feature allowing binding to two xylanases that display different folding (Payan et al. 2004). TAXI inhibits only GH11 endoglucanases. Tomato XEGIP inhibits a fungal GH12 xyloglucanase, but not endoxylanases of the GH10 and GH11 families. Based on gene sequence homology and inhibitory activity, XEGIP presumably belongs to an inhibitor protein superfamily that includes the carrot extracellular dermal glycoprotein, the tobacco nectar NEC4 protein, and monocot TAXI-type inhibitors.

The structural similarities of some CWMEI with glucanases (Durand et al. 2005), of wheat XIP-I with chitinases of the GH18 family (Payan et al. 2003), and of TAXI with an aspartyl protease (Sansen et al. 2004) raise the question of their origin. It is probable that gene sequence data associated with elucidation of the biochemical activities of CWMEIs will allow the rapid characterisation of novel families of CWMEIs (Durand et al. 2005).

12.3 Proteomics as an Efficient Way to Depict Cell Wall Protein Structure

MS applied to proteins is being used to explore and elucidate the structure of some CWP, especially with respect to PTMs. At present, the types of PTMs most studied in plant CWPs are *N*- or *O*-glycosylation, hydroxylation of Pro residues, addition of glycosylphosphatidylinositol (GPI) anchors, and phosphorylation. The following sections present some examples in which PTMs of secreted proteins have been elucidated.

12.3.1 *N*-Glycosylation

Only a few experimental data are available to describe *N*-glycosylations in CWPs. Nevertheless, site-directed mutagenesis preventing correct *N*-glycosylation of a peanut peroxidase showed that *N*-glycosylation is essential for enzymatic activity and appropriate folding of the protein (Lige et al. 2001). MS analysis of a soybean peroxidase showed considerable heterogeneity in its pattern of *N*-glycosylation (Gray and Montgomery 2006). Soybean peroxidase is a glycoprotein with a carbohydrate content of 18% (w/w). It includes seven putative Asn-X-Thr consensus glycosylation sites for *N*-linked glycans on Asn at positions 56, 130, 144, 185, 197, 211 and 216. Matrix-assisted laser desorption/ionisation-time-of-flight (MALDITOF) MS analyses of tryptic peptides showed that although all these sites in soybean peroxidase can potentially be glycosylated, not all sites are equally populated. *N*-Glycosylation is complete at Asn⁵⁶, Asn¹³⁰ and Asn¹⁴⁴, since the corresponding non-glycosylated peptides were not detected. The other *N*-glycosylation consensus sites occur in closely spaced pairs: Asn¹⁸⁵–Asn¹⁹⁷ and Asn²¹¹–Asn²¹⁶. The Asn¹⁸⁵–Asn¹⁹⁷ pair was found to be 90% glycosylated. In contrast, Asn²¹¹–Asn²¹⁶ is primarily monoglycosylated at Asn²¹¹. This can be explained by the fact that these two potential glycosylation sites are separated by only two amino acids. Since translation and glycosylation both occur in the lumen of the endoplasmic reticulum (ER), it is suggested that glycosylation at Asn²¹¹ will interfere with that of Asn²¹⁶ by steric hindrance. Mapping the glycans onto the crystal structure of soybean peroxidase showed that they are asymmetrically distributed on the molecule, occurring on the substrate-channel face of the enzyme.

12.3.2 Proline Hydroxylation and *O*-Glycosylation

The 4-hydroxylation of Pro is a PTM found in many plant glycoproteins. However, prolyl hydroxylase does not hydroxylate all Pro residues in HRGPs (extensins and AGPs) or H/PRPs. Modification of Pro residues is thought to occur co-translationally as the protein is inserted in the ER and can be followed by *O*-Hyp glycosylation. A series of experiments using synthetic genes provided the basis for a Hyp-*O*-glycosylation code (Kieliszewski 2001). A range of simple repetitive polypeptides encoding putative glycosylation motifs was designed. These experiments showed that Hyp-*O*-glycosylation is sequence-driven. Hyp-*O*-glycosylation occurs in two ways. Hyp-*O*-arabinosylation is correlated with Hyp contiguity in sequences such as (Ser-Hyp2)_n and (Ser-Hyp4)_n in *Nicotiana tabacum*. Hyp-*O*-arabinosylation results in short, neutral, linear oligosaccharides such as Hyp-(Ara)1–4, which are found in extensins. Hyp-*O*-galactosylation is predicted to occur on clustered non-contiguous Hyp residues (Ser-Hyp)_n typically found in AGPs. Hyp-*O*-galactosylation results in addition of large arabinogalactan heteropolysaccharides (Hyp-polysaccharides). In AGP, these polysaccharides consist of a β-1→3-linked galactan backbone with 1→6-linked side chains containing galactose, arabinose, and often rhamnose and

glucuronic acid. Their length is diverse, ranging from 30 sugars/Hyp in gum arabic AGP, up to 150 sugars/Hyp in a radish leaf AGP (Nothnagel 1997).

Two dodecapeptides from the CLE (CLAVATA 3/ESR-related) family were recently shown to contain Hyp by MS: a tracheary differentiation inhibitory factor (TDIF) and the active peptide of CLAVATA 3. TDIF was found in the *Zinnia elegans* xylogenic in vitro culture system (Ito et al. 2006), where it inhibits the formation of tracheids, and promotes cell division. Tandem MS/MS proved that the final product was a dodecapeptide containing two Hyp residues: HEVHypSGHypNPISN. The peptide is produced from a protein of 132 amino acids by removal of 119 amino acids from its N-terminus, and a single Arg residue from its C-terminus. The TDIF sequence from *Z. elegans* was found to be the same as that of Arabidopsis Cterminal 12 amino acids of CLE41 and CLE44.

The Arabidopsis *CLAVATA 3* gene (*CLV3*) encodes a stem cell-specific protein controlling the size of the shoot apical meristem. *CLV3* encodes a 96-amino-acid secreted protein that shares 14 conserved amino acids with other members of the CLE family in its C-terminal region. MALDI-TOF MS performed on calli overexpressing *CLV3* detected a typical region corresponding to a 12-amino-acid peptide from Arg⁷⁰ to His⁸¹ in *CLV3*, in which two of three proline residues were modified to Hyp. Comparison of MS/MS analysis of this natural peptide to the fragmentation patterns of three synthetic peptides containing two Hyp residues covering all possible combinations allowed the two Hyp in the protein sequence to be localised (Kondo et al. 2006).

12.3.3 Addition of GPI Anchors

GPI-anchored proteins are targeted to the plasma membrane. The anchor can be cleaved by specific phospholipases; consequently, the proteins can exist in both soluble and membrane-associated forms. GPI-anchored proteins belong to numerous protein families that have been identified in Arabidopsis (Borner et al. 2003). Integral membrane proteins were prepared by Triton X-114 phase partitioning and assessed for sensitivity to phospholipase C prior to proteomic analysis. Among the identified proteins are classical AGPs and arabinogalactan peptides. The PTMs of arabinogalactan peptides were more precisely characterised, and include addition of a GPI-anchor, cleavage of signal peptide and GPI-anchor, as well as hydroxylation of Gly-Pro motifs (Schultz et al. 2004). The bulk of Arabidopsis AGPs was isolated from seedlings by β -Yariv precipitation, followed by deglycosylation with anhydrous hydrogen fluoride. The resulting AG-peptides were only 10–17 amino acids long, thus allowing MS analysis without tryptic digestion. Peptide sequencing was performed on the deglycosylated peptides and the precise cleavage sites for both N- and C-terminal signals were determined using a combination of MALDITOF MS and tandem MS/MS. It was also shown that 8 out of 12 arabinogalactan peptides have all Pro residues fully hydroxylated.

12.3.4 Phosphorylation

Three reports suggest that phospho-Tyr proteins are present in the cell walls of Arabidopsis and maize, and that changes in their level of phosphorylation is induced by pathogen elicitors (Ndimba et al. 2003; Chivasa et al. 2005b). The presence of phosphorylated CWPs was revealed with antibodies raised against synthetic phospho-Tyr. The fact that phosphorylation events can occur outside the cytoplasm was surprising and interesting. The regulation of cell viability via the presence of extracellular ATP in plants suggests that an extracellular

phosphorylation mechanism exists (Chivasa et al. 2005a; see Chap. 6 by Chivasa et al., this volume). However, all protein kinases identified in Arabidopsis are predicted to be located in the cytosol (Rudrabhatla et al. 2006). The two extracellular kinases claimed in Arabidopsis (At1g53070 and At1g78850) are actually lectins. This mistake resulted from erroneous annotation since no kinase domain can be found in these proteins by bioinformatic searches for functional motifs (<http://www.ebi.ac.uk/InterProScan/>). Nevertheless, MS provides an excellent tool to prove the presence of phosphate in peptides derived from phosphoproteins. This evidence is still missing for CWP.

12.4 Emerging Functions for Cell Wall Proteins

Proteomic data cannot provide information about the roles of the identified proteins. In fact, the functions of most CWPs remain unknown. Bioinformatic searches for functional domains provide only some clues. Additional data (genetic, biochemical and cell biological) are required to fully understand protein function as exemplified in the following.

The functions assigned to GHs are provisional because their biological role, as well as the enzymatic activities of many of them, have not yet been established (Jamet et al. 2006). For example, an enzyme of the GH3 family (XYL3) shares amino acid homology with β -D-xylosidase, but XYL3 was identified as efficiently hydrolysing arabinosyl residues from arabinans, suggesting that it works as an α -L-arabinofuranosidase (Minic et al. 2006). In addition, some GH family members could have multiple physiological functions. For example, β -D-glucosidases (GH1) are involved in lignification, signalling, defence and other functions related to hydrolysis of secondary metabolites. Chitinase-like enzymes are able to degrade chitin in the cell walls of fungal pathogens (Xue et al. 1995). However, the substrates and functional roles of most chitinase-like enzymes are not completely known. For example, a mutation in a chitinase-like gene classified in GH19 (AtCTL1) caused aberrant patterns of lignification with incomplete cell walls in the stem pith (Zhong et al. 2002).

New roles for secreted proteases in the control of plant development and during plant-pathogen interactions are emerging with the description of three Arabidopsis mutants: *sdd1-1* (*stomatal density and distribution1-1*), *ale1* (*abnormal leaf shape1*) and *CDR1-D* (*constitutive disease resistant1-Dominant*). SDD1 and ALE1 belong to a family of 56 members, 46 of which are predicted to be secreted (Plant Subtilase Database, <http://csbdb.mpimp-golm.mpg.de/psdb.htm>). The establishment of a normal stomatal pattern is disrupted in *sdd1-1*, resulting in stomata clustering and increased stomatal density (von Groll et al. 2002). Overexpression of *SDD1* produced the opposite phenotype. It was proposed that SDD1 may generate an extracellular signal regulating the number of asymmetric divisions in satellite meristemoids. In the *ale1* mutant, epidermal surfaces of embryos and juvenile plants were defective, resulting in excess water loss and organ fusion in plantlets (Tanaka et al. 2001). Remnants of the endosperm remain attached to the developing embryos. It was assumed that ALE1 produces a peptide ligand required for proper differentiation of epidermis. CDR1 encodes an apoplastic aspartic protease (Xia et al. 2004). Overexpression of CDR1 causes dwarfing and resistance to virulent *Pseudomonas syringae*. Antisense *CDR1* plants were less resistant to avirulent *P. syringae* and more susceptible to virulent strains than wild type. Induction of *CDR1* generates a small mobile signal (3–10 kDa) sensitive to heat and protease. CDR1 action is blocked by the protease inhibitor pepstatin and by mutations in the protease active site. CDR1 is assumed to mediate a peptide signal system involved in the activation of inducible resistance mechanisms (Xia et al. 2004). The nature of the peptides involved in all these processes is still unknown.

Several multicopper oxidase-like proteins identified through proteomic analysis are thought to catalyse full, four-electron reduction of dioxygen (O₂) to water (H₂O) using a variety of substrates. However, since they lack a functional copper-binding site, biochemical tests should be done to characterise their enzymatic activities (Jacobs and Roe 2005). These proteins belong to a large gene family with 19 members in *Arabidopsis*. Functions for only two members of this family have been assessed by studying mutant phenotypes: SKU5 is involved in the control of root growth (Sedbrook et al. 2002) and SKS6 contributes to cotyledon vascular patterning during development (Jacobs and Roe 2005).

12.5 Conclusions

During the last 5 years, the application of proteomic techniques to the cell wall has delivered an overview of CWPs, a description of some CWP structures, and identification of new CWPs, thus opening new horizons for studies on cell wall function. To date, cell wall proteomic studies have focused essentially on CWP identification by peptide mass fingerprinting and characterisation of PTMs by peptide sequencing. Additional cell wall proteomics efforts will permit investigation of CWPs in other plants, organs, and tissues, as well as during biotic or abiotic stresses, allowing the identification of CWPs that are specific to such physiological situations. However, many CWPs might be missing in these analyses due to, e.g., cross-linking to the polysaccharide matrix, fast turnover, low abundance, or large carbohydrate moieties. Progress in MS will facilitate access to these missing proteins. Since proteins, and not genes, are the real actors in the cellular machinery, such future studies are of high importance. They are notably involved in molecular recognition, catalysis, signalling, and protein turnover, critical steps of regulation in many cellular processes.

12.6 References

- Baumberger N, Doesseger B, Guyot R, Diet A, Parsons RL, Clark MA, Simmons MP, Bedinger P, Goff SA, Ringli C, Keller B (2003) Whole-genome comparison of leucine-rich repeat extensins in *Arabidopsis* and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade. *Plant Physiol* 131:1313-1326
- Bayer EM, Bottrill AR, Walshaw J, Vigouroux M, Naldrett MJ, Thomas CL, Maule AJ (2006) *Arabidopsis* cell wall proteome defined using multidimensional protein identification technology. *Proteomics* 6:301-311
- Bellincampi D, Camardella L, Delcour JA, Desseaux V, D'Ovidio R, Durand A, Elliot G, Gebruers K, Giovane A, Juge N, Sorensen JF, Svensson B, Vairo D (2004) Potential physiological role of plant glycosidase inhibitors. *Biochim Biophys Acta* 1696:265-274
- Boller T, Gehri A, Mauch F, Vögeli U (1983) Chitinase in bean leaves: induction by ethylene, purification, properties, and possible function. *Planta* 157:22-31
- Borderies G, Jamet E, Lafitte C, Rossignol M, Jauneau A, Boudart G, Monsarrat B, Esquerré-Tugayé MT, Boudet A, Pont-Lezica R (2003) Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: a critical analysis. *Electrophoresis* 24:3421-32
- Borner GHH, Lilley KS, Stevens TJ, Dupree P (2003) Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. *Plant Physiol* 132:568-577

- Boudart G, Jamet E, Rossignol M, Lafitte C, Borderies G, Jauneau A, Esquerré-Tugayé M-T, Pont-Lezica R (2005) Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics. *Proteomics* 5:212-221
- Brady JD, Sadler IH, Fry SC (1996) Di-isodityrosine, a novel tetrameric derivative of tyrosine in plant cell wall proteins: a new potential cross-link. *J Biochem* 315:323-327
- Brunner F, Stintzi A, Fritig B, Legrand M (1998) Substrate specificities of tobacco chitinases. *Plant J* 14:225-234
- Cassab GI (1998) Plant cell wall proteins. *Annu Rev Plant Physiol Plant Mol Biol* 49:281-309
- Chivasa S, Ndimba B, Simon W, Robertson D, Yu X-L, Knox J, Bolwell P, Slabas A (2002) Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* 23:1754-1765
- Chivasa S, Ndimba B, Simon WJ, Lindsey K, Slabas AR (2005a) Extracellular ATP functions as an endogenous external metabolite regulating plant cell viability. *Plant Cell* 17:3019-3034
- Chivasa S, Simon WJ, Yu X-L, Yalpan N, Slabas AR (2005b) Pathogen elicitor-induced changes in the maize extracellular matrix proteome. *Proteomics* 5:4894-4904
- Cosgrove DJ, Li LC, Cho HT, Hoffmann-Benning S, Moore RC, Blecker D (2002) The growing world of expansins. *Plant Cell Physiol* 43:1436-1444
- del Campillo E (1999) Multiple endo-1,4-beta-D-glucanase (cellulase) genes in *Arabidopsis*. *Curr Top Dev Biol* 46:39-61
- Durand A, Hughes R, Roussel A, Flatman R, Henrissat B, Juge N (2005) Emergence of a subfamily of xylanase inhibitors within glycoside hydrolase family 18. *FEBS Letters* 272:1745-1755
- Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E (2006) Evaluation of cell wall preparations for cell wall proteomics. *Plant Methods* 2:10
- Fry SC (2004) Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. *New Phytol* 161:641-675
- Gaspar Y, Johnson KL, McKenna JA, Bacic A, Schultz CJ (2001) The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Mol Biol* 47:161-186
- Gomathi V, Gnanamanickam SS (2004) Polygalacturonase-inhibiting proteins in plant defence. *Curr Sci* 87:1211-1217
- Gray JS, Montgomery R (2006) Asymmetric glycosylation of soybean seed coat peroxidase. *Carbohydr Res* 341:198-209
- Henrissat B, Coutinho PM, Davies GJ (2001) A census of carbohydrate-active enzymes in the genome of *Arabidopsis thaliana*. *Plant Mol Biol* 47:55-72
- Hothorn M, Wolf S, Aloy P, Greiner S, Scheiffzek K (2004) Structural insights into the target specificity of plant invertase and pectin methylesterase inhibitory proteins. *Plant Cell* 16:3437-3447
- Huangpu J, Graham JS (1995) Purification and developmental analysis of an extracellular proteinase from young leaves of soybean. *Plant Physiol* 108:969-974
- Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, Fukuda H (2006) Dodeca-CLE peptides as suppressors of plant stem differentiation. *Science* 313:842-8845
- Jach G, Görnhardt B, Mundy J, Logemann J, Pinsdorf E, Leah R, Schell J, Maas C (1995) Enhanced quantitative resistance against fungal disease by combinatorial expression of different barley antifungal proteins in transgenic tobacco. *Plant J* 8:97-109
- Jacobs J, Roe JL (2005) SKS6, a multicopper oxidase-like gene, participates in cotyledon vascular patterning during *Arabidopsis thaliana* development. *Planta* 222:652-666

- Jamet E, Canut H, Boudart G, Pont-Lezica RF (2006) Cell wall proteins: a new insight through proteomics. *Trends Plant Sci* 11:33-39
- Juge N (2006) Plant protein inhibitors of cell wall degrading enzymes. *Trends Plant Sci* 11:359-367
- Kieliszewski MJ (2001) The latest hype on Hyp-*O*-glycosylation codes. *Phytochem* 57:319-323
- Kondo T, Sawa S, Kinoshita A, Mizuno S, Kakimoto T, Fukuda H, Sakagami Y (2006) A plant peptide encoded by CLV3 identified by in situ MALDI-TOF MS analysis. *Science* 313:845-848
- Kotake T, Dina S, Konishi T, Kaneko S, Igarashi K, Samejima M, Watanabe Y, Kimura K, Tsumuraya Y (2005) Molecular cloning of a {beta}-galactosidase from radish that specifically hydrolyzes {beta}-(1->3)- and {beta}-(1->6)-galactosyl residues of Arabinogalactan protein. *Plant Physiol* 138:1563-1576
- Kwon HK, Yokoyama R, Nishitani K (2005) A proteomic approach to apoplastic proteins involved in cell wall regeneration in protoplasts of *Arabidopsis* suspension-cultured cells. *Plant Cell Physiol* 46:843-857
- Lamport DTA, Northcote DH (1960) Hydroxyproline in primary cell walls of higher plants. *Nature* 18:665-666
- Lee RC, Hrmova M, Burton RA, Lahnstein J, Fincher GB (2003) Bifunctional family 3 glycoside hydrolases from barley with α -L-arabinofuranosidase and α -D-xylosidase activity. *J Biol Chem* 278:5377-5387
- Lige B, Ma S, Huystee RB (2001) The effects of the site-directed removal of *N*-glycosylation from cationic peanut peroxidase on its function. *Arch Biochem Biophys* 386:17-24
- Matsubayashi Y, Sakagami Y (2006) Peptide hormones in plants. *Annu Rev Plant Biol* 57:649-674
- Micheli F (2001) Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci* 6:414-419
- Minic Z, Do C-T, Rihouey C, Morin H, Lerouge P, Jouanin L (2006) Purification, functional characterization, cloning and identification of mutants of a seed specific arabinan hydrolase in *Arabidopsis*. *J Exp Bot* 57:2339-2351
- Minic Z, Jouanin L (2006) Plant glycoside hydrolases involved in cell wall polysaccharides degradation. *Plant Physiol Biochem*, in press
- Ndimba BK, Chivasa S, Hamilton JM, Simon WJ, Slabas AR (2003) Proteomic analysis of changes in the extracellular matrix of *Arabidopsis* cell suspension cultures induced by fungal elicitors. *Proteomics* 3:1047-1059
- Nersissian AM, Shipp EL (2002) Blue copper-binding domains *Adv Prot Chem* 60:271-340
- Nothnagel E (1997) Proteoglycans and related components in plant cells. *Int Rev Cytol* 174:195-291
- Payan F, Flatman R, Porciero S, Williamson G, Roussel A (2003) Structural analysis of xylanase inhibitor protein I (XIP-I), a proteinaceous xylanase inhibitor from wheat (*Triticum aestivum*, var. Soisson). *Biochem J* 372:399-405
- Payan F, Leone P, Porciero S, Furniss C, Tahir T, Williamson G, Durand A, Manzanares P, Gilbert HJ, Juge N, Roussel A (2004) The dual nature of wheat xylanase protein inhibitor XIP-I- structural basis for the inhibition of family 10 and family 11 xylanases. *J Biol Chem* 279:36029-36037
- Rautengarten C, Steinhauser D, Büssis D, Stinzi A, Schaller A, Kopka J, Altmann T (2005) Inferring hypotheses on functional relationships of genes: analysis of the *Arabidopsis thaliana* subtilase gene family. *PLoS Compu Biol* 1:e40

- Ringli C, Keller B, Ryser U (2001) Glycine-rich proteins as structural components of plant cell walls. *Cell Mol Life Sci* 58:1430-1441
- Robertson D, Mitchell GP, Gilroy JS, Gerrish C, Bolwell GP, Slabas AR (1997) Differential extraction and protein sequencing reveals major differences in patterns of primary cell wall proteins from plants. *J Biol Chem* 272:15841-15848
- Roitsch T, Gonzalez MC (2004) Function and regulation of plant invertases: sweet sensations. *Trends Plant Sci* 9:606-613
- Rudrabhatla P, Reddy MM, Rajasekharan R (2006) Genome-wide analysis and experimentation of plant serine/threonine/tyrosine-specific protein kinases. *Plant Mol Biol* 60:293-319.
- Sansen S, De Ranter CJ, Gebruers K, Brijs K, Courtin CM, Delcour JA, Rabijns A (2004) Structural basis for inhibition of *Aspergillus niger* xylanase by *Triticum aestivum* xylanase inhibitor-I. *J Biol Chem* 278:36022-36028
- Schaller A (2004). A cut above the rest: the regulatory function of plant proteases. *Planta* 220:183-97
- Schlumberg A, Mauch F, Vogeli U, Boller T (1986) Plant chitinases are potent inhibitors of fungal growth. *Nature* 324:365-367
- Schultz C, Rumsewicz M, Johnson K, Jones B, Gaspar Y, Bacic A (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiol* 129:1448-1463
- Schultz CJ, Ferguson KL, Lahnstein J, Bacic A (2004) Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. *J Biol Chem* 279:45503-45511
- Sedbrook JC, Carroll KL, Hung KF, Masson PH, Sommerville CR (2002) The Arabidopsis SKU5 gene encodes an extracellular glycosyl phosphatidylinositol-anchored glycoprotein involved in directional root growth. *Plant Cell* 14:1635-1648
- Sharon N, Lis H (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* 14:53R-62R
- Spadoni S, Zabotina O, Di Matteo A, Mikkelsen JD, Cervone F, De Lorenzo G, Mattei B, Bellincampi D (2006) Polygalacturonase-inhibiting protein interacts with pectin through a binding site formed of four clustered residues of arginine and lysine. *Plant Physiol* 141:557-564
- Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y (2001) A subtilisin-like serine protease is required for epidermal surface formation in Arabidopsis embryos and juvenile plants. *Development* 128:4681-4689
- Van Damme E, Peumans W, Barre A, Rougé P (1998) Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Crit Rev Plant Sci* 17:575-692
- von Groll U, Berger D, Altmann T (2002) The subtilisin-like serine protease SDD1 mediates cell-to-cell signaling during Arabidopsis stomatal development. *Plant Cell* 14:1527-1539
- Watson BS, Lei Z, Dixon RA, Sumner LW (2004) Proteomics of *Medicago sativa* cell walls. *Phytochem* 65:1709-1720
- Welinder KG, Justesen AF, Kjaersgard IVH, Jensen RB, Rasmussen SK, Jespersen HM, Duroux L (2002) Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*. *Eur J Biochem* 269:6063-6081
- Xia Y, Suzuki H, Borevitz J, Blount J, Guo Z, Patel K, Dixon RA, Lamb C (2004) An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. *EMBO J* 23:980-988

- Xu Z, Escamilla-Trevino L, Zeng L, Lalgondar M, Bevan D, Winkel B, Mohamed A, Cheng CL, Shih MC, Poulton J, Esen A (2004) Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 1. *Plant Mol Biol* 55:343-367
- Xue J, Jorgensen M, Pihlgren U, Rask L (1995) The myrosinase gene family in *Arabidopsis thaliana*: gene organization, expression and evolution. *Plant Mol Biol* 27:911-922
- Zhong R, Kays SJ, Schroeder BP, Ye ZH (2002) Mutation of a chitinase-like gene causes ectopic deposition of lignin, aberrant cell shapes, and overproduction of ethylene. *Plant Cell* 14:165-179
- Zhu J, Chen S, Alvarez S, Asirvatham VS, Schachtman DP, Wu Y, Sharp RE (2006) Cell wall proteome in the maize primary root elongation zone. I. Extraction and identification of water-soluble and lightly ionically bound proteins. *Plant Physiol* 140:311-325
- Zhu Q, Maher EA, Masoud S, Dixon RA, Lamb CJ (1994) Enhanced protection against fungal attack by constitutive co-expression of chitinase and glucanase genes in transgenic tobacco. *Biotechnology* 12:807-812