Combining various strategies to increase the coverage of the plant cell wall glycoproteome.

Yu Zhang, Aurélie Giboulot, Michel Zivy, Benoît Valot, Elisabeth Jamet, Cécile Albenne

Combining various strategies
to increase the coverage of the plant cell wall glycoproteome

Yu Zhang [a, b, e #], Aurélie Giboulot [a, b #], Michel Zivy [c], Benoît Valot [d], Elisabeth Jamet [a, b],

Cécile Albenne[a, b, *]

[#] co-first authors

[a.] Université de Toulouse; UPS; UMR 5546, Surfaces Cellulaires et Signalisation chez les Végétaux; BP 42617, F-31326 Castanet-Tolosan, France
[b.] CNRS; UMR 5546; BP 42617, F-31326 Castanet-Tolosan, France
[c.] CNRS, PAPPSO, UMR 0320 / UMR 8120 Génétique Végétale, F-91190 Gif sur Yvette, France
[d.] INRA, PAPPSO, UMR 0320 / UMR 8120 Génétique Végétale, F-91190 Gif sur Yvette, France

[*] Corresponding author: Tel: +33 534 32 38 27; fax: +33 534 32 38 02
 Email address: albenne@scsv.ups-tlse.fr
[e.] Present address : Institut Jean-Pierre Bourgin, UMR 1318 INRA-AgroParisTech,
INRA Centre de Grignon, F-78850 Thiverval-Grignon, France

ABSTRACT

Glycoproteomics recently became a very active field, mostly in mammals. The first part of this paper consists of a mini-review on the strategies used in glycoproteomics, namely methods for enrichment in glycoproteins and mass spectrometry (MS) techniques currently used. In a second part, these strategies are applied to the cell wall glycoproteome of etiolated hypocotyls of *Arabidopsis thaliana*, showing their complementarity. Several sub-glycoproteomes were obtained by: (i) affinity chromatography on concanavaline A (ConA) and analysis of glycoproteins by MALDI-TOF MS; (ii) multidimensional lectin chromatography (using AIL, PNA, ConA and WGA lectins) and subsequent identification of glycoproteins by MALDI-TOF MS and LC-MS/MS; (iii) boronic acid chromatography followed by identification of glycoproteins by MALDI-TOF MS. Altogether, 127 glycoproteins were identified. Most glycoproteins were found to be putative *N*-glycoproteins and *N*-glycopeptides were predicted from MS data using the *ProTerNyc* bioinformatics software.

KEYWORDS:

cell wall protein, post-translational modification, glycoproteomics, glycan prediction

1. INTRODUCTION

The plant cell wall, mainly composed of polysaccharides (cellulose, hemicelluloses, and pectins), is involved in different biological processes such as growth, environmental sensing, signaling, and defense (Ellis et al., 2002; Roberts, 1990, 2001; Vogel et al., 2004). Cell wall proteins (CWPs), which represent less than 10% of the cell wall mass, are very diverse. From previous estimation, 1500 to 2000 cell wall proteins are expected in *Arabidopsis thaliana* (Jamet et al., 2006). Most of them probably undergo glycosylation during their transit through the secretory pathway. Several cell wall proteomic studies performed on *A. thaliana* (for reviews, see Jamet et al., 2008; Jamet et al., 2006) and other species (Alvarez et al., 2006; Bhushan et al., 2006; Chen et al., 2008; Cho et al., 2009; Jung et al., 2008; Watson et al., 2004; Zhu et al., 2006) enabled to predict a variety of functions in the wall. A database containing about 800 CWPs classified into nine functional classes is available at our web site (www.polebio.scsv.ups-tlse.fr/WallProtDB/) (Jamet et al., 2006; Pont-Lezica et al., 2010). However, the biochemical function of most CWPs was not demonstrated and their wall substrates or partners were not identified so far.

The post-translational modification (PTMs) of CWPs should be considered to fully understand their structure/function relationships. Glycosylation is a widespread protein PTM that controls various processes such as protein folding, activity, stability, signaling or interaction (Spiro, 2002). The two main types of protein glycosylations are *N-* and *O-* glycosylations. In plant cells, as in other eukaryotic cells, *N-*glycosylation starts in the endoplasmic reticulum (ER) by a co-translational transfer of an oligosaccharide precursor, $Glc_3Man_9GlcNAc_2$, onto specific Asn residues within the consensus sequence Asn-X-Ser/Thr (where X is any amino acid except Pro). During protein transport along the secretory pathway, specific glycosidases and glycosyltransferases perform maturation steps leading to characteristic complex and high Man *N-*glycans (Faye et al., 2005). Unlike *N-*glycan addition, *O-*glycosylation takes place at a post-translational level, by a step by step process. *O-* glycosylation in plants mainly concerns cell wall hydroxyproline (Hyp)-rich glycoproteins (HRGPs) such as extensins or ArabinoGalactan Proteins (AGPs), and Hyp/Pro-rich proteins (H/PRPs) (Faye et al., 2005; Kieliszewski, 2001). Ser residues can be monogalactosylated (Cho and Chrispeels, 1976; Faye et al., 2005). Pro hydroxylation is performed in the ER and/or in the Golgi apparatus by a prolyl-4-hydroxylase (Robinson et al., 1985). Hyp-glycosylation, which is unique to plants, occurs through a process still not completely

understood. Pro hydroxylation and Hyp glycosylation are directed by a code depending on the contiguity of Pro along the protein primary sequence (Kieliszewski, 2001). Extensins, characterized by clusters of contiguous Hyp, carry arabinooligosaccharides whereas AGPs are hyperglycosylated onto Hyp alternating with Ala, Ser or Thr residues. Finally, H/PRPs contain repetitive Pro/Hyp-rich sequences which are assumed to be galactosylated in *Nicotiana alata* (Sommer-Knudsen et al., 1996) and *A. thaliana* (Liu and Mehdy, 2007).

Little information is available on plant protein glycosylation in the literature, as compared to mammal glycoproteins that have been extensively characterized. However, it was shown that *N*-glycans on the jack bean α-mannosidase influence the enzyme activity and its oligomerization (Kimura et al., 1999). In the case of the cationic peanut peroxidase, it was reported that *N*-glycans are required for full catalytic activity as well as for the control of thermostability and folding (Lige et al., 2001). Recently, it was proposed that *N*-glycosylation motifs on a plant cell wall polygalacturonase inhibitor protein (PGIP) favor the interaction with its fungal polygalacturonase target during host-pathogen interaction (Lim et al., 2009). Concerning the role of *O*-glycosylations in plants, it was demonstrated that *O*-glycans permit to keep extensins in a rigid extended conformation (Stafstrom and Staehelin, 1986). It has also been suggested that *O*-glycans increase HRGP solubility, resistance to proteolytic degradation (Lamport, 1967), and thermal stability (Ferris et al., 2001; Homer and Roberts, 1979). Recently, biophysical data provided, for the first time, quantitative evidence that *O*-glycosylation causes a dramatic increase in the thermal stability of the polyproline type II helix characteristic of some HRGPs (Owens et al., 2010).

Although glycoproteomics recently became a very active field (Tissot et al., 2009), such approaches have never been performed on plants. The global pattern of protein glycosylation inside the wall may reflect the regulation processes occurring *in muro* to modulate biological activity of CWPs and cell wall dynamics. Considering the numerous methods available (described in the next paragraph) and the lack of information about global plant protein glycosylation patterns, we decided to explore the cell wall glycoproteome of *A. thaliana*.

We will describe for the first time the cell wall glycoproteome of *A. thaliana* etiolated hypocotyls, a model previously used for cell wall proteomics (Irshad et al., 2008). The strategy consisted in purifying cell walls prior to extracting proteins with salt solutions (Feiz et al., 2006). Then, mono-dimensional or multi-dimensional lectin affinity chromatography,

combined with boronic acid chromatography were performed to get sub-proteomes enriched into different classes of glycoproteins. Altogether, 127 glycoproteins were identified, with a preponderance of *N*-glycoproteins. *N*-glycopeptides were predicted for some of *N*-glycoproteins, revealing different patterns of *N*-glycans. The information obtained on the identity of glycoproteins and the prediction of glycan motif structures enabled to go beyond protein identification provided by mass spectrometry (MS) and bioinformatics, and to give new insight into plant CWP structure.

2. GLYCOPROTEOMICS STRATEGIES

The "glycoproteomics" field concerns the identification of glycoproteins in a complex mixture, but also tends to locate glycosylation sites and to characterize glycan structures (An et al., 2009). A wide range of methodological and instrumental techniques can be used to obtain glycoproteomes and to analyze them by MS (Fig. 1). The starting material is enriched in glycoproteins by affinity or chemical methods. Affinity methods are based on the use of lectin conjugate resins. The specificity of the lectin determines the type of glycan captured (Table 1). Concanavalin A (ConA) affinity chromatography is the most commonly used (Bunkenborg et al., 2004; Faye et al., 2005; Minic et al., 2007; Wang et al., 2006). Lectin chromatography using other lectins such as Wheat Germ Agglutinin (WGA) (Zhao et al., 2006), *Sambucus nigra* lectin (SNA) (Qiu and Regnier, 2005), Peanut agglutinin (PNA) (Schwientek et al., 2007) or *Artocarpus integrifolia* lectin (AIL) (Schwientek et al., 2007) were carried out, either in mono- or multi-dimensional ways. Chemical methods rely on binding glycoproteins onto a specific resin by carbohydrate chemical reactivity, without any preliminary treatment (*e.g.* boronic acid-functionalized beads) or with a preliminary stage of modification or labeling of carbohydrates to facilitate their isolation onto dedicated matrixes. Under basic conditions, boronic acid can form diesters with all glycans containing *cis*-diol groups (Sparbier et al., 2005). The hydrazide-based glycoprotein capture method starts by a periodic oxidation that cleaves the carbon-carbon linkage between the diols of saccharides. The aldehyde groups produced can be bound onto hydrazine functional groups immobilized on a solid support (Zhang et al., 2003; Zhang et al., 2009). An alternative strategy is the enzymatic modification or the metabolic labeling which add a bioorthogonal moiety onto target carbohydrates, allowing highly efficient purification of glycoproteins (Bond and Kohler, 2007). Considering the plurality and the complementarities of available methods for glycoprotein capture, combining different strategies was shown to substantially improve the

coverage of the glycoproteome (McDonald et al., 2009), as well as quantification of glycopeptides (Zhang et al., 2003).

Following glycoprotein capture, their identification is performed through a classical proteomics approach (Aebersold and Mann, 2003), consisting in a proteolysis stage (commonly using trypsin), a MS analysis of peptides and a bioinformatics validation. Protein identification can be performed by Matrix-Assisted Laser Desorption/Ionisation – Time Of Flight (MALDI-TOF) MS or Liquid Chromatography – ElectroSpray Ionisation tandem MS (LC-ESI-MS/MS). MALDI-TOF MS enables to identify proteins by peptide mass fingerprinting (PMF) and is used for sequenced genomes. LC-MS/MS provides peptide sequence information. It is usually more sensitive and allows enlarging the coverage of identified proteins and increasing the number of proteins identified. Single MS approaches also permit to gain insight into the glycan structures. Indeed, since MALDI and ESI are soft ionization sources, glycopeptide integrity is preserved during MS analysis and MS spectra enclose information about glycan masses and heterogeneity. Bioinformatic tools such as GlycoWorkbench (Ceroni et al., 2008) or *ProTerNyc* (Albenne et al., 2009) were developed to facilitate the identification of glycopeptides on MS spectra. New fragmentation methods, such as electron capture dissociation (ECD) and electron transfer dissociation (ETD), allow glycopeptide analysis since they cleave peptide bonds while leaving the glycan attached to the peptide, allowing the location of glycosylation sites (Catalina et al., 2007; Hakansson et al., 2001).

To go further into the glycan location and structure, additional methods can be implemented. On the one hand, a step of glycopeptide capture can be introduced after proteolysis in order to focus on glycopeptides, improving the global glycosylation pattern analysis: (i) the lectin chromatography and hydrazide approaches described above can be carried out; ii) hydrophilic interaction liquid chromatography (HILIC), using cellulose or sepharose matrixes, can be used since most glycopeptides are more hydrophilic than unmodified peptides (Hagglund et al., 2004; Wada et al., 2004; Wuhrer et al., 2009); iii) size-exclusion chromatography has also been applied to glycopeptide enrichment since most tryptic glycopeptides have high masses (Alvarez-Manilla et al., 2006). On the other hand, a deglycosylation step can be performed, using chemical (Edge, 2003; Mirgorodskaya et al., 1999; Patel et al., 1993) or enzymatic approaches (King et al., 2006; Küster et al., 1996; Tarentino et al., 1985). Analysis of fully or partially deglycosylated peptides enables to confirm the glycosylation site occupancy and/or

to characterize the type of glycosylation. The released glycan can also be analyzed by tandem MS providing additional information about its structure (Harvey, 2005; Morelle et al., 2006).
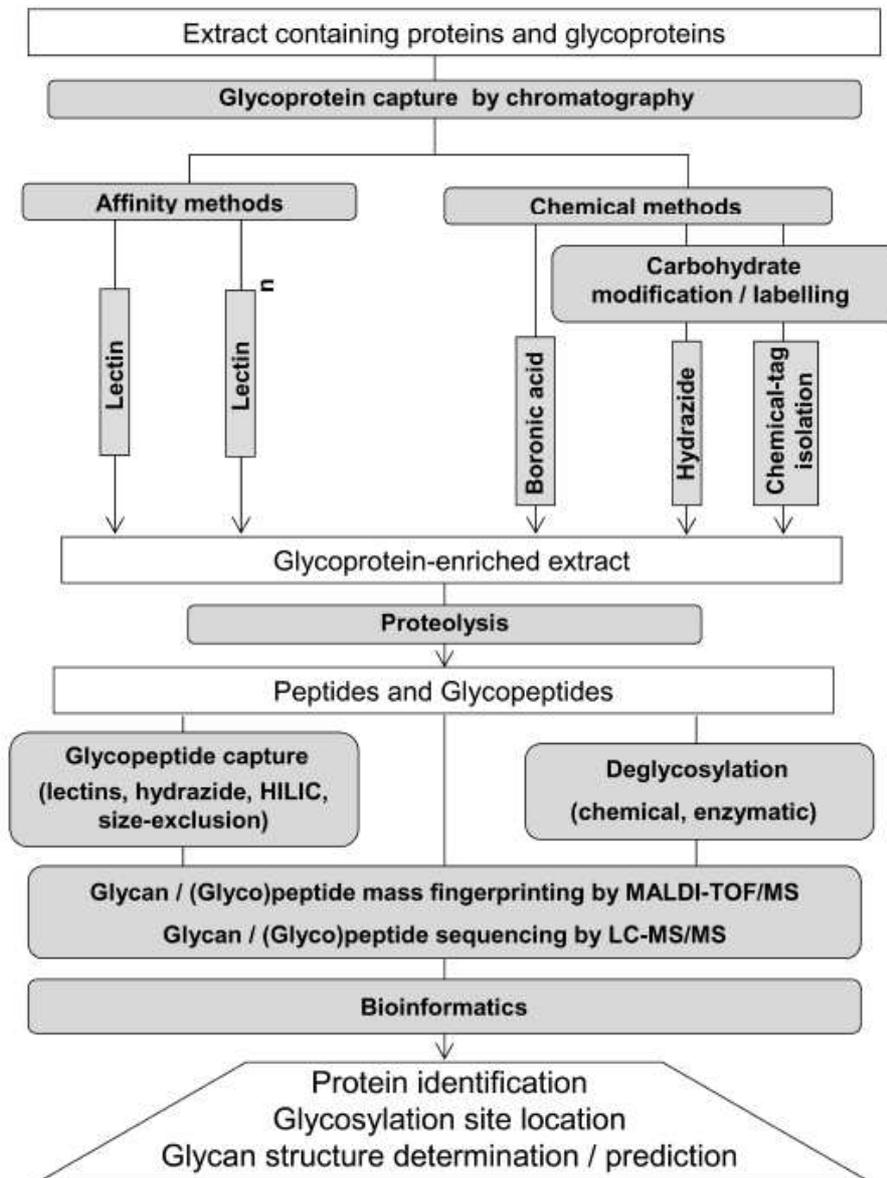

Figure 1. Glycoproteomics approaches: a simplified flowchart.

In a glycoproteomics strategy, the first step consists in getting sub-proteomes enriched in glycoproteins through affinity or chemical chromatography methods. After a proteolysis step, peptides and glycopeptides are analysis by either PMF using MALDI-TOF MS or peptide sequencing using LC-MS/MS. Glycopeptide capture as well as deglycosylation experiments can also be performed before MS analysis. MS data analyses of (glyco)peptides and/or glycans, assisted by bioinformatics software, enable to identify proteins, to locate glycosylation sites and to predict or determine glycan structures.

3. RESULTS AND DISCUSSION

3.1. Glycoprotein capture and identification

CWPs were extracted by salt solutions from cell walls of 11-day-old *A. thaliana* etiolated hypocotyls (Irshad et al., 2008). Different glycoprotein capture strategies and MS strategies were employed to obtain several sub-proteomes (Fig. 2): i) ConA affinity chromatography and analysis of glycoproteins by MALDI-TOF MS; ii) multidimensional lectin chromatography (using AIL, PNA, ConA, and WGA lectins; Table 1) followed by analysis of glycoproteins by MALDI-TOF MS or LC-MS/MS; iii) boronic acid chromatography and analysis of glycoproteins by MALDI-TOF MS.

Table 1. Sugar specificity of plant lectins used for glycan capture and detection. The specificities targeted in this study are written in bold.

| Abbreviation	Lectin name	Origin	Specificity
Mannose-specific lectins			
ConA	Concanavalin A	*Concanavalia ensiformis*	**α-Man** > α-Glc > GlcNac
GNA	Galanthus Nivalis Agglutinin	*Galanthus nivalis*	**α-Man**
Galactose-specific lectins			
PNA	PeaNut Agglutinin	*Arachis hypogaea*	**Gal**β1-3GalNAc > **α-Gal** and **β-Gal**
AIL	Artocarpus Integrifolia Lectin	*Artocarpus integrifolia*	**Gal**β1-3GalNAc > **α-Gal**, α-GalNac
***N*-acetyl-glucosamine specific lectin**			
WGA	Wheat Germ Agglutinin	*Triticum vulgaris*	**GlcNAc**(β1-4**GlcNAc**)$_{1-2}$ > **βGlcNAc**
DSA	Datura Stramonium Agglutinin	*Datura stramonium*	Gal1-4**GlcNAc** > **GlcNAc**
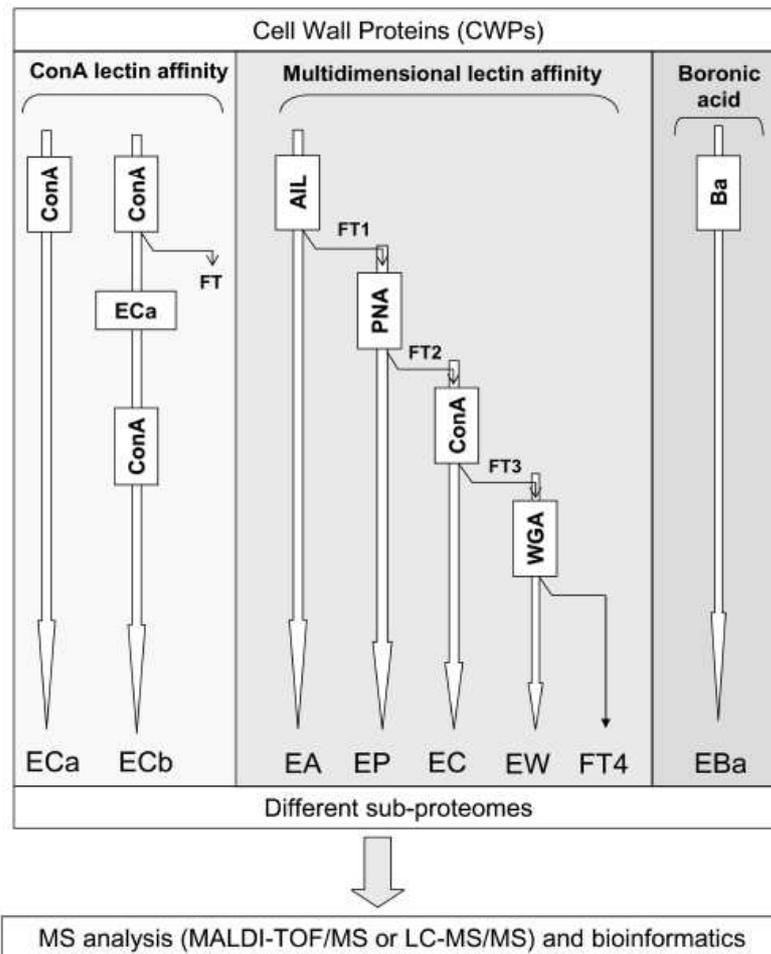
Figure 2. Enrichment of CWPs into glycoproteins using different chromatographic procedures.

Our cell wall glycoproteomic strategy consisted in obtaining different sub-proteomes enriched in glycoproteins from a CWP extract of *A. thaliana* etiolated hypocotyls. ConA lectin affinity chromatography led to the elution fractions ECa. ECa was then loaded onto a second ConA matrix, leading to the elution fractions ECb. A multidimensional lectin affinity approach was carried out using AIL, PNA, ConA, and WGA. Flow-through fractions (FTn) from a matrix (n) were loaded onto the following matrix (n+1). Each lectin-conjugate resin was eluted, producing elution fractions EA, EP, EC, and EW, from AIL, PNA, ConA, and WGA, respectively. Boronic acid-functionalized beads (Ba) were also used, leading to the elution fractions EBa. All these sub-proteomes were subsequently analyzed using MALDI-TOF/MS or LC-MS/MS and bioinformatics software.

3.1.1. ConA lectin affinity

A ConA affinity chromatography, specific for Man residues, was carried out in order to capture *N*-glycoproteins. The unbound fractions (flow-through and washes) were estimated to contain more than 50% (in mass) of the CWPs put in contact with the matrix. An aliquot of CWPs (1/160), a part of the flow-through fraction (3/8) as well as the totality of the wash and

elution fractions were analyzed by SDS-PAGE (Fig. 3A). The ECa elution fractions display a specific pattern compared to the flow-through one, suggesting an efficient separation between Man-containing glycoproteins and other proteins. A total of 38 bands were excised from the ECa (1) lane for MALDI-TOF MS analyses. Thirty-eight proteins were identified (Tables S1 and S5, supplementary data). All of them had a predicted signal peptide, showing that they were putative secreted proteins.
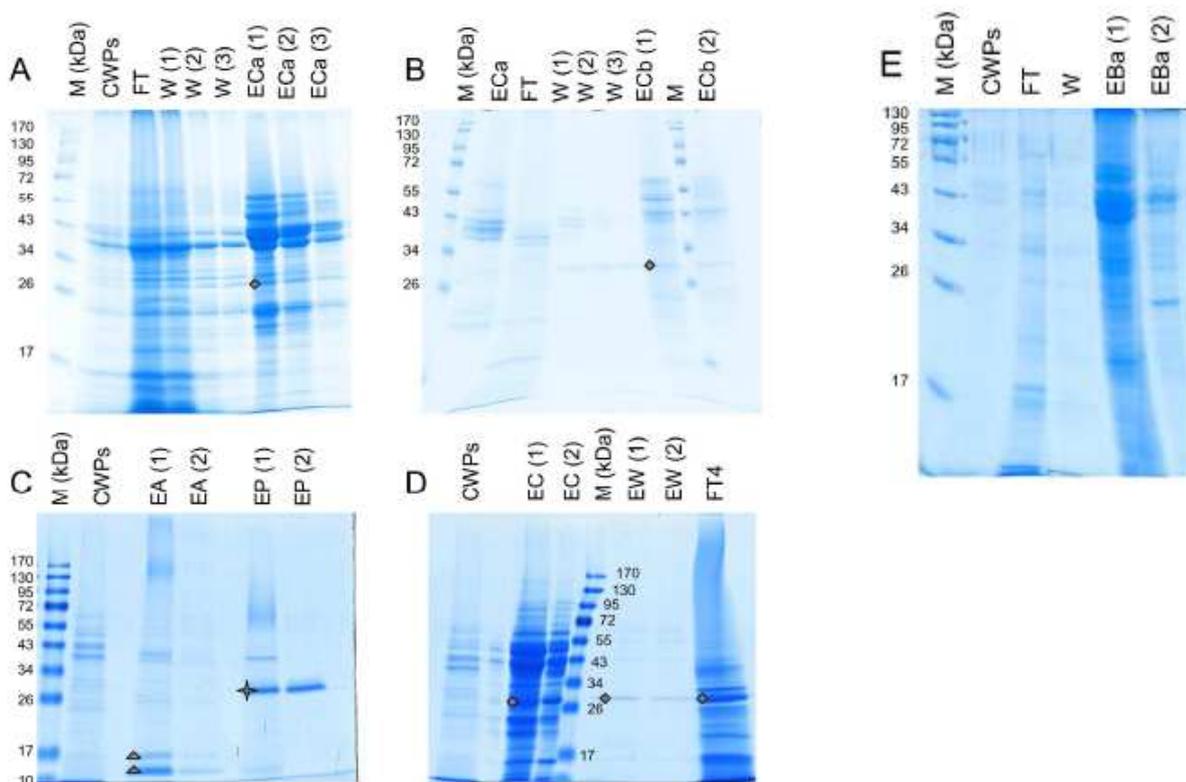

Figure 3. Electrophoretic analysis of proteins separated by different chromatographic methods for enrichment of glycoproteins.

A. Fractions obtained from ConA chromatography; **B**. Fractions obtained from two successive ConA chromatographies; **C**. Fractions obtained from AIL and PNA chromatographies with the multi-dimensional lectin affinity approach; **D**. Fractions obtained from ConA and WGA chromatographies with the multi-dimensional lectin affinity approach; **E**. Fractions obtained from boronic acid chromatography.
M: molecular mass markers; CWPs: cell wall proteins; FT: flow-through fraction; W(1, 2, 3): wash fractions; ECa (1, 2, 3), ECb (1, 2), EA (1, 2), EP (1, 2), EC (1, 2), EW (1, 2) and EBa (1, 2) : elution fractions according to the nomenclature used on Fig. 2.
Lectins identified by MALDI-TOF MS in the elution fractions are indicated with the following symbols: ConA: ◆ ; AIL: ▲ ; PNA.: ✦

In a second experiment, two successive steps of ConA affinity chromatography were performed, the elution fractions from the first step (ECa) being loaded onto a new ConA matrix (Fig. 2). This experiment aimed at increasing the specificity of *N*-glycoprotein binding. As for the first ConA step, an aliquot of ECa (1/240), a part of the flow-through fraction (3/8) and the totality of the wash and elution fractions were analyzed by SDS-PAGE (Fig. 3B). The pattern of ECb shows a specific enrichment into some proteins of ECa. More than 70 % in mass of the proteins were lost between the two ConA steps. MALDI-TOF MS analyses of ECb (1) allowed the identification of 17 proteins (Tables S1 and S5, supplementary data). All these proteins were already found in ECa (1). The low recovery observed between ECa and ECb could be explained by non-specific interactions during the first step or by glycan loss.

3.1.2. Multidimensional lectin affinity

A multidimensional lectin affinity chromatography experiment was carried out in order to get specific sub-glycoproteomes and to enlarge the coverage of the cell wall glycoproteome. The flow-through fraction from a matrix was loaded onto the next one (Fig. 2). First, AIL and PNA lectins were used to capture Gal-containing *N*- and/or *O*-glycoproteins (Table 1). AIL is assumed to mostly recognize α-Gal, and PNA α-Gal and β-Gal (Banerjee et al., 1994; Jeyaprakash et al., 2005). A ConA chromatography was introduced in third position and a WGA resin, specific for GlcNAc residues, was used as the last step (Table 1). Elution fractions (EA(1), EP(1), EC(1), EW(1)) were divided into equal parts which were separated by SDS-PAGE for three distinct analyses: Coomassie Brilliant Blue (CBB) staining (Fig. 3C-D); Yariv phenylglucoside staining to reveal AGPs (Nothnagel, 1997; Yariv et al., 1967) (Fig. 4A); and lectin-blot (Fig. 4B1-4). In each case, an aliquot of CWPs (1/200) and of the last flow-through fraction FT4 (1/18) were also analyzed. Positive and negative controls were added.

EA and EP display a very weak CBB staining, revealing smears and few thin bands (Fig. 3C). The most intense bands corresponded to the lectins (AIL and PNA) that were weakly attached onto the resin. Eighteen and 10 bands were excised from EA and EP respectively (Fig. S2, supplementary data). However, MALDI-TOF MS signals were very weak and only two proteins were identified in each sub-proteome (Table S2, supplementary data). At5g14920 (AtGASA14, Roxrud et al., 2007) and At5g06860 were found in EA (1). At5g14920 has a large Hyp/Pro-rich region and a short C-terminal sequence containing the tryptic peptides

used for protein identification. Previous MS analyses suggested that the Hyp/Pro-rich region carries O-glycosylations (Albenne et al., 2009). In this new experiment, At5g14920 is observed as a smear with an apparent molecular mass between 147 and 175 kDa, although its theoretical molecular mass is 24 kDa (Fig. 3C). This is much higher than the molecular mass previously observed for this protein after cationic exchange chromatography (Albenne et al., 2009; Irshad et al., 2008). This suggests that the separation process can lead to different glycoform isolation and/or that glycan degradation events can occur during separation steps. Binding of At5g14920 onto the AIL lectin confirms that it is a glycoprotein and suggests that it contains α-Gal residues (see Table 1). Lectin-blot analysis of the EA (1) fraction using the PNA lectin gave a very weak signal (Fig. 4B1), suggesting a glycosylation pattern poorly recognized by this lectin. Besides, Yariv staining was negative (Fig. 4A), confirming previous results (Albenne et al., 2009). The second protein found in EA (1) is encoded by *At5g06860* and is predicted to be a polygalacturonase inhibiting protein (PGIP1). It displays an apparent molecular mass of 43 kDa, *i.e.* 9 kDa above its theoretical molecular mass. *O*-glycans on Ser residues may occur. This glycan pattern is recognized neither by the PNA lectin, nor by the Yariv reagent (Fig. 4B1, Fig. 4A). These results suggest possible new glycosylation patterns not yet described for such proteins which could be responsible for the different affinities observed with PNA and AIL lectins.

The first elution fraction obtained after the PNA chromatography (EP (1)) displays a smear between 57 and 85 kDa where a protein called AGP31 could be identified. MS criteria were modified for identification of this protein by PMF. Indeed, only two different peptides, already found in a previous study (Irshad et al., 2008) were used for AGP31 identification. However, its characteristic electrophoretic pattern, as well as the strong signal observed on the PNA lectin-blot (Fig. 4B2), strengthen the identification of AGP31, previously described as a Gal-rich glycoprotein (Liu and Mehdy, 2007). AGP31 displays a short AGP motif and is expected to be galactosylated. However, unlike previously reported (Liu and Mehdy, 2007), AGP31 was not stained by the Yariv reagent, suggesting that, in our assay conditions, arabinogalactan motifs were not present (Fig. 4A). The second protein found in EP (1) was identified in a thin band at an apparent molecular mass of 43 kDa (Fig. 3C; Fig. S2, supplementary data). This is a protein of unknown function encoded by *At5g25460*. It may carry Gal-containing *N*- or *O*-glycans.
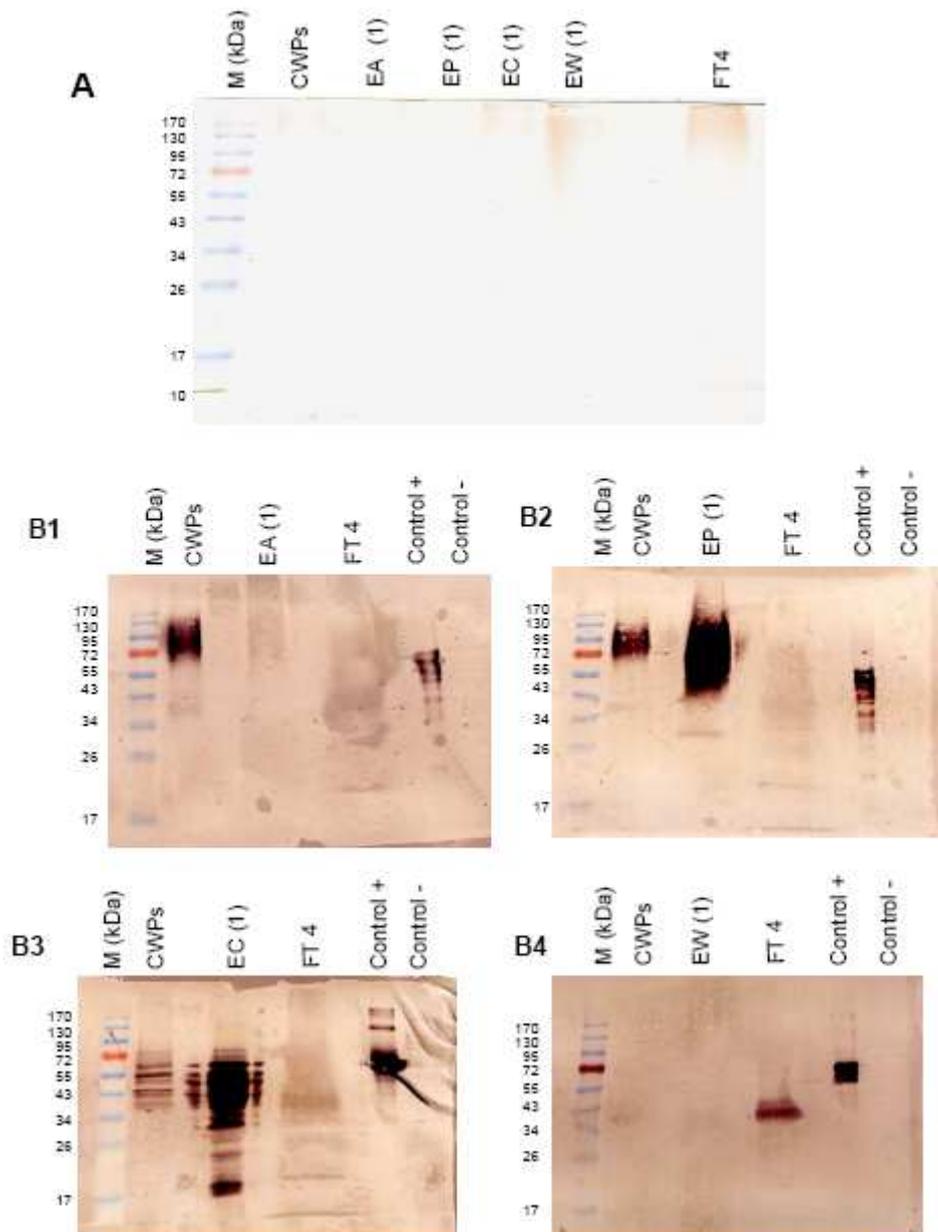
Figure 4. Analysis of the sub-proteomes obtained by multi-dimensional lectin chromatography using Yariv staining or lectin blot.

A. Yariv stained SDS-polyacrylamide gel. **B**. Lectin-blot analysis using PNA digoxigenin-labeled lectin (**B1** and **B2**), GNA digoxigenin-labeled lectin (**B3**) and DSA digoxigenin-labeled lectin (**B4**) for detection of glycans. Positive controls are provided in the DIG Glycan Differentiation Kit (Roche) used for these analyses. Negative control is an *E. coli* protein extract.
M: molecular mass markers; CWPs: cell wall proteins; EA(1), EP(1), EC(1), EW(1): first elution fractions obtained from AIL, PNA, ConA and WGA chromatographic steps, respectively; FT4 : flow-through fraction obtained from WGA chromatography.

The third stage of this multi-dimensional lectin procedure consisted in a ConA chromatography similar to that reported in the previous paragraph. SDS-PAGE analysis of EC (1) and EC (2) followed by CBB staining displays a pattern similar to that shown on Fig. 3A (Fig. 3D). Lectin-blot analysis of EC (1) showed that most proteins, excepted low molecular mass ones, were recognized by GNA, a lectin specific for Man residues (Fig. 4B3). ConA thus appears as an efficient tool to specifically capture *N*-glycoproteins. As expected, Yariv staining of EC (1) gave no signal (Fig. 4A). In order to increase the number of identified proteins in EC (1), LC-MS/MS analyses were performed from a liquid sample similar to that analyzed by SDS-PAGE. One hundred fifteen proteins were identified in EC, with only two proteins having no predicted signal peptide (Table S4, supplementary data), attesting the efficiency of ConA to specifically capture secreted proteins. Three times more proteins were identified in EC (1) than in ECa (1) using LC-MS/MS, since LC-MS/MS is a more efficient tool than MALDI-TOF MS for protein identification in a complex mixture. However, among the 38 proteins identified in ECa (1), only 30 were also found in EC (1). MALDI-TOF MS allowed specific identification of 8 proteins, showing that both approaches are complementary. In addition, unlike LC-MS/MS, MALDI-TOF MS analyses permit to keep PTMs, enabling to get information on glycan structures from spectra (see paragraph 3.3).

The last stage of the multi-dimensional lectin strategy was carried out using the WGA lectin, specific for GlcNAc (Table 1). SDS-PAGE of EW only permitted to detect three weak bands using CBB staining (Fig. 3D). They were excised from the EW (1) lane, but the only identified protein was ConA that was issued from the previous step (Figure S2B and Table S2, supplementary data). Lectin blot analysis using DSA, a lectin with specificity similar to that of WGA (Table 1), was also negative (Fig. 4B4). These results show that the WGA step was inefficient. It can be proposed that *N*-glycoproteins containing terminal GlcNAc were all retained on the ConA resin. However, a very weak signal could be seen on EW (1) after the Yariv staining (Fig. 4A). This suggests that some AGPs were retained onto WGA, although AGPs are not expected to contain GlcNAc (Seifert and Roberts, 2007).

Following the four stages of this multi-dimensional lectin procedure, the last flow-through fraction (FT4) corresponding to the overall unbound proteins was analyzed. CBB-stained SDS-polyacrylamide gel displayed a staining with a smear in its upper part and more focused bands in its downer part (Fig. 3D). MALDI-TOF MS analysis of the 18 bands excised from the gel only allowed identification of 6 proteins (Table S2, supplementary data). Proteolysis is

assumed to have occurred during the different stages of this multi-step process, limiting protein identification by PMF using MALDI-TOF MS. All the proteins identified except one (At4g29270) were found in at least one sub-glycoproteome, suggesting that they exist in both glycosylated and non glycosylated forms. It should be noted that At4g29270 has no predicted *N*-glycosylation site and was previously found in the cell wall proteome of 11-day-old etiolated hypocotyls (Irshad et al., 2008). Lectin-blot analyses using PNA and GNA did not permit to detect any protein in FT4 (Fig. 4B1-3). It shows that this fraction does not contain any Gal-rich proteins nor *N*-glycoproteins, showing the efficiency of the previous chromatographic steps carried out. On the contrary, using the DSA lectin, a signal is observed at 42 kDa where three proteins of unknown function (At5g11420, At5g25460, and At4g32460) were identified (Fig. 4B4). At least one of these proteins may carry GlcNAc residues. Since these proteins were very abundant in the CWP extract, it is possible that their capture by the different chromatographic steps was not complete. Finally, a Yariv staining was observed at the top of the FT4 lane (Fig. 4A). This suggests that the CWP extract contains some AGPs, but that the chromatographic steps performed were not efficient to capture them.

3.1.3. Boronic acid chromatography

Chemical glycoprotein capture methods were carried out in order to take advantage of the plurality of available methods to enlarge the cell wall glycoproteome. In our hands, the hydrazide-based method was unsuccessful. On the contrary, boronic acid chromatography enabled to bind more than 50 % in mass of the CWPs loaded on the resin. An aliquot of CWPs (1/200), a part of the flow-through fraction (3/10) as well as the totality of the wash and elution fractions were analyzed by SDS-PAGE (Fig. 3E). Elution fractions (EBa (1) and EBa (2)) display an electrophoretic pattern distinct from those obtained by lectin affinity chromatography. Indeed, boronic acid resin is assumed to be less specific than lectin matrixes, since it can bind any cis-diol containing carbohydrate. However, MALDI-TOF MS analyses of EBa (1) permitted to identify only 20 proteins, among which 2 proteins with no predicted signal peptide. It shows that EBa is mostly composed of some major proteins, reducing the diversity of this sub-proteome.

### 3.2. Comparative analysis of the glycoproteomes

A comparison of the different sub-proteomes obtained is illustrated by a Venn diagram shown on Fig. 5A. Altogether, 127 CWPs were identified combining lectin affinity (ConA, AIL, and

PNA) and boronic acid chromatographies. The first evidence is that ConA permitted to capture most of the glycoproteins identified. Indeed, 121 out of the 127 proteins are found in at least one ConA sub-proteome (EC/ECa/ECb). It should be noted that the overall ConA sub-proteome was enlarged by the combination of LC-MS/MS and MALDI-TOF MS tools. The use of AIL, PNA and WGA lectins in the multi-dimensional lectin approach permitted to identify only one additional protein in EA (AtGASA14). This protein is assumed to be *O*-glycosylated as previously reported (Albenne et al., 2009). Besides, the absence of putative *N*-glycosylation sites on AtGASA14 is in agreement with the fact that it was not found in ConA fractions. The second protein of EA (At5g06860) was also found in EBa and EC, possibly because of the presence of diverse glycan structures on it. The two proteins of EP (At5g25460 and At1g28290) were also found in ConA sub-proteomes. Their presence in EP and ConA sub-proteomes suggests that they carry both *N*- and *O*-glycans. Boronic acid chromatography enabled to identify 5 additional proteins (two peroxidases, one glycine-rich protein and two proteins of unknown function). Although four of them possess putative *N*-glycosylation sites, they may carry *O*-glycans and not *N*-glycans.

Figure 5. Numerical analysis of the cell wall glycoproteome of *A. thaliana* etiolated hypocotyls using Venn diagrams.

A. Comparison of the different sub-glycoproteomes obtained by lectin affinity chromatography on ConA (EC/ECa/ECb), PNA (EP), AIL (EA), and by boronic acid chromatography (EBa). The numbers indicated correspond to the number of glycoproteins identified in each sub-proteome.
B. Comparison of the cell wall glycoproteome (this study) to the cell wall proteome (Irshad *et al.,* 2008) of 11-day-old etiolated hypocotyls of *A. thaliana*.
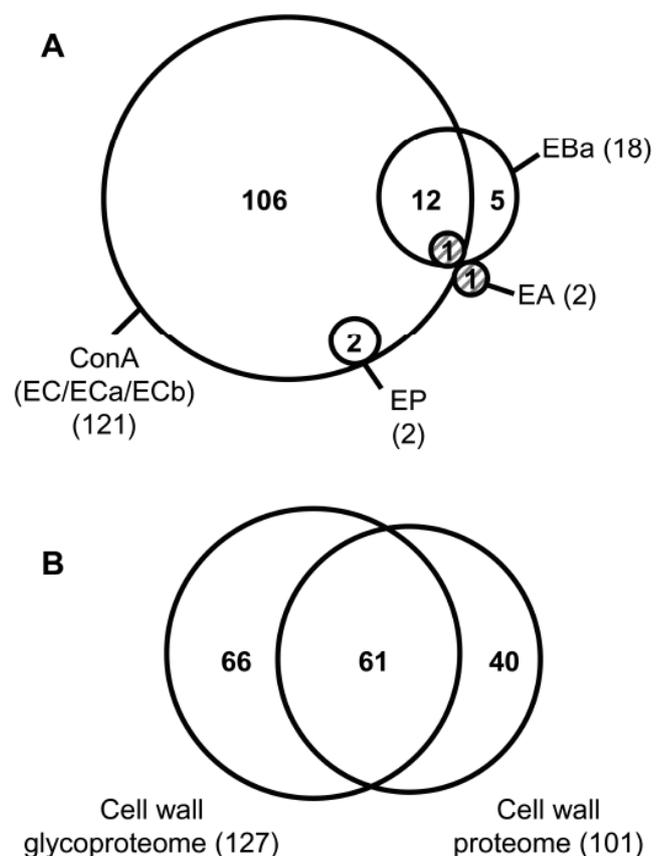
The coverage of the cell wall glycoproteome of *A. thaliana* 11 day-old etiolated hypocotyls is difficult to estimate. Protein extraction procedure is assumed to introduce a bias since most labile proteins are probably lost and strongly bound proteins among which HRGPs are not extracted (Feiz et al., 2006). Glycan maturation and degradation events are expected in cell walls (Minic, 2008). Finally, glycan loss can occur during the extraction or the chromatographic steps. However, our results clearly show that our glycoproteome is essentially an *N*-glycoproteome. Interestingly, 119 out of the 127 glycoproteins possess putative *N*-glycosylation sites. The different strategies employed to capture *O*-glycoproteins (AIL, PNA, boronic acid chromatographies) were found to be far less efficient. Only a few structural proteins were identified in our glycoproteome (Table S5, supplementary data). Although *O*-glycoproteins, like HRGPs and H/PRPs, are assumed to be abundant proteins in the cell wall, they are under-represented. Several limitations can explain this situation: i) their extraction from the cell wall may be difficult because of strong or even covalent interactions with cell wall polysaccharides or CWPs (Brady et al., 1996, 1998; Ringli, 2010); ii) their affinity towards lectins may be too weak for an efficient binding; iii) their high carbohydrate content may prevent their migration in SDS-polyacrylamide gels as well as their CBB staining; iv) trypsin digestion may be inefficient because of a low frequency of Lys and Arg residues along the primary sequence and/or steric hindrance of glycans; v) large *O*-glycopeptides may be difficult to ionize and their identification may be limited since they are hardly predictable. In the case of AGPs, none of the chromatographic procedures used seem to be efficient for capturing them since Yariv staining gave a signal mostly in the FT4 fraction (Fig. 4A). In addition, their identification would require other methods such as deglycosylation and N-terminus sequencing (Schultz et al., 2004).

The 127 proteins of our glycoproteome belong to the 9 functional classes previously proposed (Jamet et al., 2008): proteins acting on carbohydrates (20.5 %); oxido-reductases (16.6 %); proteases (9.4 %); proteins with interacting domains (14.2 %); proteins related to lipid metabolism (3.9 %); signaling proteins (3.9 %); structural proteins (3.9 %); miscellaneous proteins (14.2 %); and proteins of unknown function (13.4 %). The possible biological functions of these proteins were discussed in a previous paper (Irshad et al., 2008). Among the 127 proteins identified in the present study, 61 were already found in our previous cell wall proteome from 11-day old *A. thaliana* etiolated hypocotyls (Irshad et al., 2008) (Fig. 5B). Among the 66 newly identified proteins, 56 are unique to EC, suggesting that the enlargement of the cell wall proteome is mostly due to the use of LC-MS/MS. Besides, glycoprotein capture strategies may have permit to enrich some minor proteins, making

possible their identification by MS. On the contrary, 40 proteins previously identified in the cell wall proteome are absent from the glycoproteome (Fig. 5B). They may correspond to non-glycosylated proteins. Interestingly, only 25 proteins out of these 40, *i.e.* 62.5 %, have putative *N*-glycosylation sites. This rate is clearly lower than that determined for our glycoproteome (93.7 %), strengthening the hypothesis of the lack of glycosylation for these proteins.

3.3. Glycoprotein validation and glycan prediction

Analysis of the electrophoretic migration of proteins from ECa and ECb fractions (Fig. 3A-B and Table S1, supplementary data) shows that 81 % and 89 % of them, respectively, display an experimental molecular mass above the theoretical one, suggesting the presence of PTMs. The proteins for which experimental mass is lower than expected may be degraded or matured (Albenne et al., 2009). For EBa, 9 proteins among the 20 found in this sub-proteome, *i.e.* 45 %, have an experimental molecular mass lower than the theoretical one, suggesting that proteolysis occurred during the preparation and the analysis of this sample (Table S3, supplementary data). On the contrary, the proteins identified in the EA and EP sub-proteomes display an apparent molecular mass above the theoretical one (Table S2, supplementary data). In particular, the proteins At5g14920 and At1g28290 are found as smears at apparent masses much higher than theoretical ones, suggesting high carbohydrate contents for both proteins (Fig. 3C).

Bioinformatics data processing was performed in order to check the specificity of the glycoprotein capture and to predict *N*-glycan structure. Among the 127 (121) proteins identified in our glycoproteome (ConA glycoproteome), 119 (115) possess at least one putative *N*-glycosylation site (Table S5, supplementary data). Surprisingly, among the 6 proteins devoid of *N*-glycosylation site from the ConA proteome, 5 are alpha-expansins. It can be suggested that they may be captured by ConA through *N*-glycosylated protein partners.
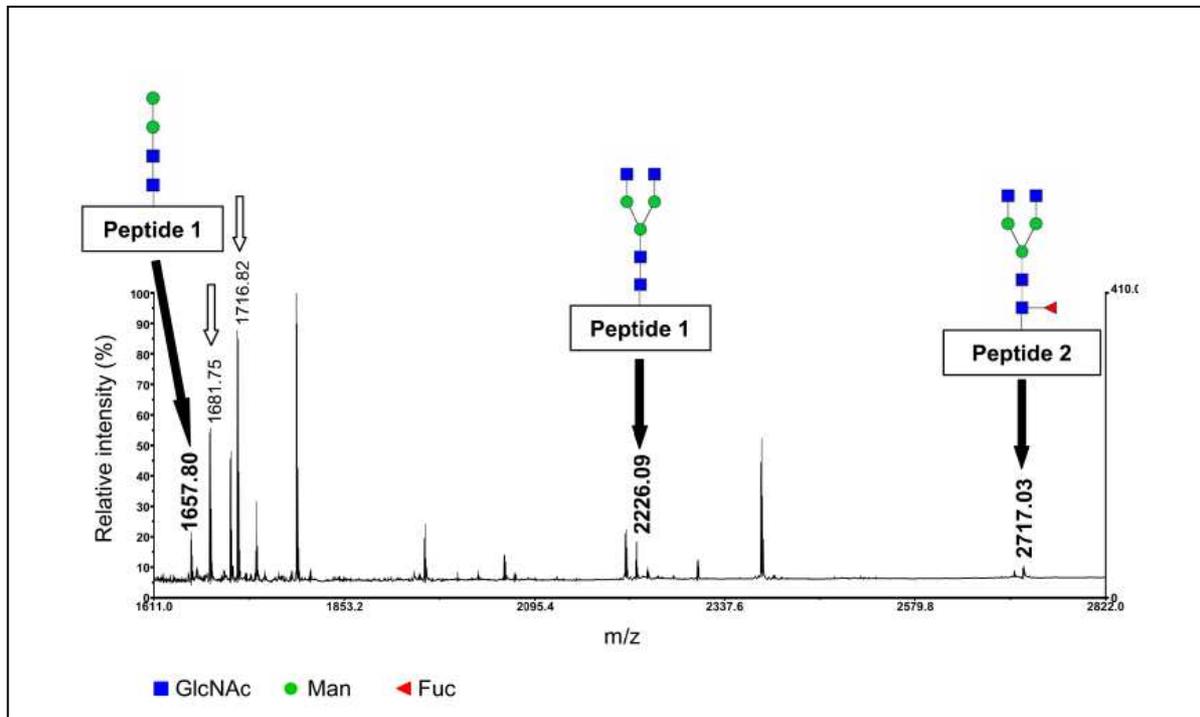
Figure 6. Prediction of *N*-glycopeptides of At1g68560 alpha-xylosidase (XYL1) on MALDI-TOF MS spectrum using *ProTerNyc.*

The protein was identified in the ECa1-2 bands from SDS-PAGE analysis of *N*-glycoproteins separated by ConA affinity chromatography.
Non-modified peptides used for identification are shown with white arrows and predicted *N*-glycopeptides with black arrows. Two peptides (1 and 2) carrying *N*-glycosylation consensus sequences (shown on Table 2) were predicted to carry *N*-glycosylations using *ProTerNyc.* The mass of these *N*-glycopeptides corresponds to the addition of the mass of the peptide to that of the *N*-glycans schematically represented.

To go further into the *N*-glycoproteome analysis, we performed bioinformatics analyses using *ProTerNyc* (Albenne et al., 2009). This software was developed to predict *N*-glycosylated peptides using the data obtained by MS. Exploring the MALDI-TOF MS for the 43 putative glycoproteins found in ECa, ECb and EBa, *N*-glycopeptides were predicted for 20 proteins (Table 2). The fact that prediction was not possible for 23 proteins is not really surprising since glycopeptides are difficult to analyze by MS. Indeed, glycans may cause steric hindrance, preventing trypsin action. Besides, glycopeptides can be difficult to ionize. The 20 proteins for which predictions were possible belong to the 7 functional classes represented by the 43 proteins considered for this search, suggesting that *N*-glycosylation is ubiquitous. For each protein, between 1 and 5 glycopeptides were predicted: multiple predictions concern different glycosylation sites, but also different glycan structures at a given site. Such an heterogeneity is well known for *N*-glycoproteins and was particularly well documented in the case of the horseradish peroxidase (Gray, 1998). In the case of the alpha-xylosidase encoded by *At1g68560*, the peptide 1 (DFNMSWK) is predicted to carry 2 different *N*-glycans (N2M2 and N2M3N2) whereas peptide 2 (INATGVVAPVGFK) is predicted to carry only one kind
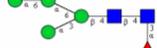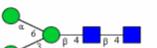
of *N*-glycan (Fig. 6, Table 2). Overall, among the complex or hybrid structures, the N2M3F1X1 motif is the major species. N2M4N2 and N2M3N2F1X1 were also found. Very few high Man-type glycans were predicted (two N2M5 and one N2M7). These patterns of glycosylation corroborate those reported for soybean and horseradish peroxidases whose glycosylations were fully characterized (Gray, 1998; Gray et al., 1996). Finally, some short glycans such as N2M2, N2M2F1 or N2M2X1 can be predicted. These structures may correspond to degradation of *N*-glycans. Glycoside hydrolases, which are abundant in the wall, may have cell wall *N*-glycoproteins as substrates and may be involved in the regulation of their function (Minic, 2008). These *N*-glycan predictions give new insight into the structure of some *N*-glycoproteins and may be considered in the frame of the study of their structure/function relationships.
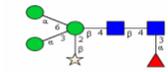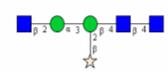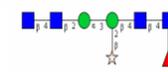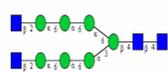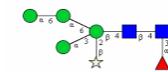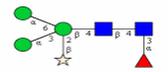
4. CONCLUSIONS

This work constitutes the first plant cell wall glycoproteome described so far. Several glycoprotein capture approaches combined with different MS methods were carried out successfully to identify 127 putative glycoproteins, some of them being identified for the first time in cell walls of 11-day-old *A. thaliana* etiolated hypocotyls. ConA chromatography was found to be the most efficient method, most of the glycoproteins identified being putative *N*-glycoproteins. The use of other lectin and boronic acid chromatographies permitted to slightly increase the overall glycoproteome. The *ProTerNyc* software, recently developed in our lab (Albenne et al., 2009) was found to be an efficient tool for *N*-glycan predictions and permitted to gain new insight into the structure of some *N*-glycoproteins. In the future, enrichment methods based on metabolic labeling of carbohydrates and hydrazide chemistry should be developed to efficiently capture plant *O*-glycoproteins and improve their identification (Bond and Kohler, 2007). Besides, advanced tandem MS technologies, namely ECD and ETD, are very promising and will permit to go further into the structural characterization of plant cell wall glycoproteins (Catalina et al., 2007; Hakansson et al., 2001).
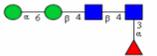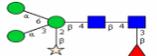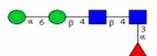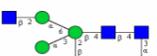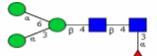
Table 2. Prediction of *N*-glycopeptides on glycoproteins using *ProTerNyc* (Albenne et al., 2009)

Glycoproteins were isolated by chromatography on ConA or boronic acid resins and were identified by PMF using MALDI-TOF/MS.

Accession number	Functional annotation	Sequence of putative glycopeptide [a]	Theoretical m/z		Experimental m/z of predicted glycopeptides	Predicted *N*-glycan [b]	
			unmodified peptide	predicted *N*-glycopeptide			
Proteins acting on carbohydrates							
At1g68560	glycoside hydrolase family 31 (GH1) (alpha-xylosidase AtXYL1)	DF**NMSW**K (897-903) (Peptide 1)	927.4035	1657.6835	1657.8003	N2M2	
		DF**NMSW**K (897-903) (Peptide 1)	927.4035	2225.9035	2226.0922	N2M3N2	
		I**NATG**VVAPVGFK (481-493) (Peptide 2)	1272.7316	2717.2916	2717.0380	N2M3N2F1	
Oxido-reductases							
At1g71695	peroxidase (AtPrx12)	S**NCSA**R (333-338)	637.2728	1837.7328	1837.3118	N2M4F1	
		S**NCSA**R (333-338)	637.2728	1848.7328	1848.3540	N2M2N1F1X1	
		TCPTA**NSSN**TQVNDIR (246-261)	1720.7924	2451.0724	2451.4043	N2M2	
		RTCPTA**NSSN**TQVNDIR (245-261)	1876.8935	2769.2335	2769.1876	N2M3	
		MGQMSVLTGTQGEIRS**NCSA**R (318-338)	2226.0429	3686.6029	3686.6965	N2M4N2	

At2g30210	homologous to laccase (AtLAC3)	CQGP**NGTR** (379-386)	832.3736	1708.7136	1708.7366	N2M2F1
At3g49120	peroxidase (AtPrx34)	SCP**NVTN**IVR (40-49)	1102.5679	1964.8879	1964.9476	N2M2X1
		SCP**NVTN**IVRETIVNELR (40-57)	2057.0813	3065.4613	3065.5449	N2M2F1X1
		GLIQSDQELFSSP**NATD**TIPLVR (272-294)	2501.2888	3687.7288	3687.7991	N2M4X1
At4g12880	homologous to blue copper binding protein (AEEN20)	**NYTR** (88-91)	553.2734	1415.5934	1415.3922	N2M2X1
		NYTRGAGR (88-95)	894.4546	1827.8146	1827.5085	N2M2N1
		FWNPNI**NYTL**WAQGK (36-50)	1851.9182	2931.3382	2931.2651	N2M2N1F1
		FWNPNI**NYTL**WAQGK (36-50)	1851.9182	3312.4782	3312.3093	N2M4N2
		KFWNPNI**NYTL**WAQGK (35-50)	1980.0132	3661.6532	3661.0480	N2M3N1F1X1F1G1

Proteases

| At3g02110 | homologous to serine carboxypeptidase D (SCPL25) (Peptidase family S10, S10.005, MEROPS) | NW**NDTD**STVLPIYR (365-378) | 1693.8186 | 3270.4186 | 3270.4210 | N2M3N2F1X1 |
| At3g54400 | homologous to Asp protease (pepsin family) (Peptidase family A1, subfamily A1B unassigned peptidases, MEROPS) | SS**NFSG**SLR (239-247) | 954.4645 | 1992.8645 | 1993.0300 | N2M3F1 |

		Peptide				Composition	Structure
		SS**NFSG**SLR (239-247)	954.4645	2124.9045	2124.9189	N2M3F1X1	
		SS**NFSG**SLR (239-247)	954.4645	2286.9645	2286.4467	N2M4F1X1	
At3g63470	homologous to serine carboxypeptidase (SCPL40)	M**NLTA**K (262-267)	677.3656	1610.7256	1610.5756	N2M2N1	
		KM**NLTA**K (261-267)	805.4606	1870.8606	1870.5458	N2M2N1X1	
		KM**NLTA**K (261-267)	805.4606	2220.0006	2219.5480	N2M2N2F1X1	
		KM**NLTA**K (261-267)	805.4606	2225.0006	2224.4596	N2M5N1	
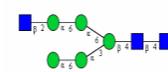
Proteins with interacting domains (with proteins or polysaccharides)

		Peptide				Composition	Structure
At1g03220	homologous to carrot EDGP and tomato XEGIP	QTNCANF**NFTS**TA (400-412)	1418.6011	2635.0611	2635.3037	N2M5	
		QTNCANF**NFTS**TA (421-433)	1418.6011	3365.3411	3364.6962	N2M7N2	
At1g78830	homologous to lectin (curculin-like)	DA**NTSS**VAYIK (444-454)	1168.5850	2501.0850	2500.3921	N2M4F1X1	
		NNPVGE**NATL**SLGR (100-113)	1441.7399	2612.1799	2612.1033	N2M3F1X1	
		DA**NTSS**VAYIKY (444-455)	1331.6483	2573.1283	2573.5324	N2M3N1F1	
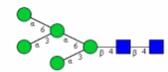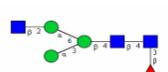
At1g78850	homologous to lectin (curculin-like)	H**NATL**SFIR (268-276)	1058.5747	1934.9147	1934.6541	N2M2F1	
		GSPVKE**NATL**TFGEDGNLVLAEADGR (96-121)	2660.3168	3830.7568	3830.5284	N2M3F1X1	
At1g78860	homologous to lectin (curculin-like)	H**NATL**SFLR (268-276)	1058.5747	1934.9147	1934.6541	N2M2F1	
		AP**NASK** (432-437)	587.3153	1960.8353	1960.6692	N2M3N1F1X1	
		H**NATL**SFLR (268-276)	1058.5747	2599.1547	2598.3817	N2M7	
		TGDTSLVAYVKAP**NASK** (421-437)	1721.9074	2760.3074	2760.3701	N2M3F1	
		GSPVKE**NATL**TFGEDGNLVLAEADGR (96-121)	2660.3168	3830.7568	3830.5284	N2M3F1X1	
At3g20820	homologous to PGIP (LRR domains)	SLT**NLSS**LMHLDLR (170-183)	1599.8529	2932.3529	2932.7810	N2M4F1X1	
		SLT**NLSS**LMHLDLR (170-183)	1599.8529	3014.3929	3014.3935	N2M2N2F1X1	
At5g12940	homologous to *Phaseolus vulgaris* PGIP2 (LRR domains)	NN**NISG**VIPR (191-200)	1083.5911	2254.0311	2253.4454	N2M3F1X1	
		NN**NISG**VIPR (191-200)	1083.5911	2660.1911	2660.2128	N2M3N2F1X1	
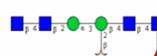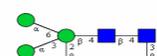
Structural proteins

23

		Peptide				Composition	Structure
At1g28290	PRP/AGP (AtAGP31)	F**NRSL**VAVR (181-189)	1061.6220	2273.0800	2273.2976	N2M2N1F1X1	
		F**NRSL**VAVR (225-233)	1061.6220	2303.1020	2302.9716	N2M3N1F1	
		SKK**NITA**ETTTDK (222-234)	1436.7597	2867.2997	2867.2985	N2M3N2X1	
		LVCKSKK**NITA**ETTTDK (218-234)	1880.0163	3324.5763	3325.3717	N2M3N2F1	
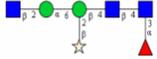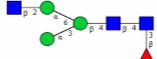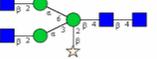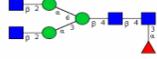
Miscellaneous

At2g27190	homologous to purple acid phosphatase	**NRTH**AYFSWNR (429-439)	1451.6933	2314.0133	2313.4018	N2M2X1	
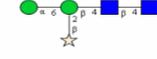
At4g25900	homologous to apospory-associated protein (aldose 1-epimerase domain)	**NRTR** (213-216)	546.3112	1716.7512	1716.5568	N2M3F1X1	
		NRTR (213-216)	546.3112	1757.7712	1757.5659	N2M2N1F1X1	
		NRTR (213-216)	546.3112	2122.9112	2122.6203	N2M3N2F1X1	
		NRTRFTDHEK (213-222)	1303.6507	2342.0507	2341.6471	N2M3F1	
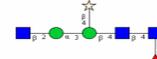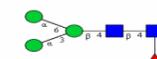
Unknown function

At2g28490	expressed protein (cupin domain)	TL**NLTS**LSIAFGVDEETMRR (462-481)	2253.1549	3145.4949	3145.6553	N2M3	
		TL**NLTS**LSIAFGVDEETMRR (462-481)	2253.1549	3261.5349	3262.1740	N2M2F1X1	
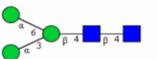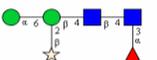
		TL**NLTS**LSIAFGVDEETMR (462-480)	2097.0538	3313.5138	3314.4323	N2M5
At3g22640	expressed protein (cupin domain)	LFTSQ**NYSY**FAPTSR (450-464)	1781.8499	2952.2899	2951.9541	N2M3F1X1
At5g11420	expressed protein (DUF642)	L**NVTK** (97-101)	574.3564	1760.7964	1760.8636	N2M4X1
		L**NVTK** (97-101)	574.3564	1774.8164	1774.8811	N2M4F1

[a] Putative *N*-glycosylation sites predicted by PROSITE are in bold (http://www.expasy.org/prosite/)

[b] Glycan structures are designed using the GlycoWorkbench software (Ceroni et al., 2008): Man (M) ●, GlcNAc (N) ■, Fuc (F) ▲, Xylose (X) ☆, ○ Gal (G).

25

5. EXPERIMENTAL

5.1. Plant culture, cell wall isolation and cell wall protein extraction

A. thaliana seeds (ecotype Columbia 0) were grown in the dark as described (Feiz et al., 2006). Etiolated hypocotyls were collected at 11 days. Cell walls were isolated as previously reported (Feiz et al., 2006). Proteins were extracted from the cell wall preparation using 0.2 M $CaCl_2$ and 2 M LiCl solutions, as described (Irshad et al., 2008). After desalting, proteins were lyophilized and solubilized in binding buffer suitable for the subsequent chromatography step. Protein content was estimated using the Bradford method with bovine serum albumin as a standard (Bradford, 1976).

5.2. Glycoprotein enrichment

5.2.1. Lectin affinity chromatography

Four lectin-agarose matrixes (Sigma, St Louis, MO) were used, *e.g.* ConA, AIL, PNA and WGA. All chromatography steps were carried out in batch mode and the different samples (flow-through, wash and elution) were obtained by decantation. Before equilibration, ConA, AIL and PNA resins were conditioned with 10X volume of pre-wash buffer 1 (20 mM Tris-HCl pH 7.4, 1 M NaCl, 3.33 mM $MgCl_2$/$MnCl_2$/$CaCl_2$) and the WGA resin with 10X volume of pre-wash buffer 2 (20 mM Na_2HPO_4/NaH_2PO_4 pH 7, 1 M NaCl) in order to remove unbound lectins.

5.2.1.1. ConA affinity selection

After resin equilibration with 10X volume of binding buffer 1 (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM $MgCl_2$/$MnCl_2$/$CaCl_2$), CWPs (8 mL at 125 µg/mL in binding buffer 1) were put in contact with the matrix (0.6 mL) for 2 h at room temperature. After flow-through removal, the resin was washed three times by 1.5 mL of binding buffer 1. Proteins were eluted with 3X 1.5 mL of binding buffer 1 supplemented with 1 M methyl-α-D-Glc (Sigma). A fraction of flow-through (3 mL) and the totality of each wash and elution samples were desalted using Econo-Pac 10 DG columns (Bio-Rad, Hercules, CA) and lyophilized.

In a second experiment, two successive steps of ConA affinity chromatography were performed (Fig. 2). The first chromatography was carried out like described above. The three elution fractions were mixed, desalted using Econo-Pac 10 DG columns (Bio-Rad) and resuspended into binding buffer 1 prior to a second ConA chromatography performed as the first one.

5.2.1.2. Multidimensional lectin affinity selection

Four lectin-agarose resins (AIL, ConA, PNA and WGA) were used for multidimensional lectin affinity selection. The flow-through fraction from a matrix was loaded onto the next one in the following order: AIL, PNA, ConA and WGA (Fig. 2). CWPs (10 mL at 300 µg/mL in binding buffer 1) were incubated with 1 mL of the AIL matrix, beforehand conditioned using 10 mL of binding buffer 1, for 8 h at 4°C under mild agitation. Flow-through was collected. After two consecutive washes using 3 mL of binding buffer 1, proteins were eluted by 2X 3.2 mL of 1 M Gal in binding buffer 1.

Chromatography on PNA and ConA were performed, using 1 mL of resin in each case. Equilibration and wash steps were performed with binding buffer 1, under similar conditions as for AIL chromatography. The incubation times with proteins were adjusted to 14 h for PNA and 4 h for ConA. Elutions were carried out with binding buffer 1, supplemented with 0.5 M Gal and 1 M methyl-α-D-Glc for PNA and ConA respectively. The ConA flow-through fraction was dialyzed against binding buffer 2 (50 mM Na_2HPO_4/NaH_2PO_4 pH 7, 200 mM NaCl) required for WGA-affinity chromatography. In this case, incubation was performed for 8 h at 4°C. Washes and elution steps were done similarly to other lectin chromatography, using 0.5 M GlcNAc in binding buffer 2. All the elution samples, as well as a part (one fifth) of the last flow-through fraction issued from the WGA resin, were desalted using Econo-Pac 10 DG columns and lyophilized prior to electrophoretic analyses.

5.2.2. Boronic acid chromatography

A phenylboronic acid-agarose conjugate resin (Clontech, Saint-Germain-en-Laye, France) was used. After a conditioning step of the resin (2 mL) with 2X 10 mL of binding buffer 3 (50 mM HEPES pH 8.4, 0.5 M NaCl), CWPs (10 mL at 100 µg/mL in binding buffer 3) were incubated in batch with the matrix for 20 min on a shaker at room temperature. The flow-through fraction was collected by gravity and a wash step was performed with 2X 10 mL of binding buffer 3. The resin, in suspension with binding buffer 3 was transferred into an empty column. After an additional wash step with 5 mL of binding buffer 3, proteins were eluted using 2X 3 mL of binding buffer 3 supplemented with 100 mM Tris. Part of flow-through and wash fractions as well as the totality of each elution fractions were desalted using Econo-Pac 10 DG columns (Bio-Rad) and lyophilized prior to electrophoretic analysis.

5.3. Electrophoretic analyses of glycoproteins by SDS-PAGE and lectin blots

Dried samples from the different chromatography steps were dissolved in 50 to 200 µL of UHQ water, depending on the number of electrophoretic analyses to be performed. In the case of ECb, dried sample was dissolved in 100 µL of 25 mM DTT and incubated at 50°C for 30 min. After iodoacetamide addition to a final concentration of 100 mM, ECb sample is incubated at room temperature for 30 min prior loading. Fifty µL of each sample were loaded on 10 x 12 x 0.15 cm or 15 x 22 x 0.15 cm SDS-polyacrylamide gel with a concentration of 12.5 % of acrylamide/bisacrylamide. Separation was performed according to Laemmli (1970). The gel staining was carried out with CBB (Scheler et al., 1998) or with the β-glycosyl Yariv reagent (Willats and Knox, 1996). The experimental molecular masses were determined using the Melanie3 software (Geneva Bioinformatics SA, Geneva, Switzerland) with marker masses as standards. For lectin blots, proteins separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) were transferred onto a PVDF membrane (Schleicher & Schuell, VWR, Fontenay-sous-Bois, France) using a liquid transfer cell (Mini Trans-Blot Cell, Bio-Rad). Transfer was carried out at 4°C in transfer buffer (Tris-HCl 48 mM, glycine 39 mM, SDS 0.0375%) for 2 h 15 at 75 V. Glycoproteins bound to the membrane were revealed, through their glycans, by a lectin blot method, using different digoxigenin-labeled lectins (DSA, GNA and PNA) provided in the DIG Glycan Differentiation Kit, following the manufacturer's recommendations (Roche, Mannheim, Germany). A total protein extract from *Escherichia coli* (30 µL at 0.4 µg/µL) was used as negative control for lectin blots.

5.4. MS analyses

5.4.1. MALDI-TOF MS

Bands stained with CBB were excised from gels and digested with trypsin as described (Borderies et al., 2003; Boudart et al., 2005). MALDI-TOF MS analyses were performed using a Voyager-DE STR mass spectrometer (Applied Biosystems/MD SCIEX). Spectra were acquired in reflectron mode as previously reported (Borderies et al., 2003). Peptide mass fingerprints were compared to the non-redundant database of *A. thaliana* at NCBI (http://www.ncbi.nlm.nih.gov/) using ProteinProspector MS-FIT (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard).

Monoisotopic mass accuracy of 20 ppm, minimum of 4 matching peptides, maximum one missed cleavage, and possible oxidation of Met residues and hydroxylation of Pro residues were the criteria used for database searches. Cys carboxyamidomethylation was allowed for reduced alkylated samples.

5.4.2. LC-MS/MS

Prior to analysis, proteins were shortly separated by SDS-PAGE to get three samples in order to increase the efficiency of tryptic digestion. In-gel digestion was performed with the Progest system (Genomic Solution, Digilab, Cambridgeshire, UK) according to a standard trypsin protocol. Gel pieces were washed twice by successive baths of 10% acetic acid, 40% ethanol, and acetonitrile (ACN). They were then washed twice successively with 25 mM NH_4CO_3 and ACN. Digestion was subsequently performed for 6 h at 37°C with 125 ng of modified trypsin (Promega, Charbonnières-les-Bains, France) dissolved in 20% methanol and 20 mM NH_4CO_3. The peptides were extracted successively with 2% trifluoroacetic acid (TFA) and 50% ACN, and then with ACN. Peptide extracts were dried in a vacuum centrifuge and suspended in 20 µL of 0.05% TFA, 0.05% HCOOH, and 2% ACN.

HPLC was performed on an NanoLC-Ultra system (Eksigent, Dublin, CA). A 4 µL sample was loaded at 7.5 µL.min^{-1} on a precolumn cartridge (stationary phase: C18 PepMap 100, 5 µm; column: 100 µm i.d., 1 cm; Dionex, Voisins le Bretonneux, France) and desalted with 0.1% HCOOH. After 3 min, the precolumn cartridge was connected to the separating PepMap C18 column (stationary phase: C18 PepMap 100, 3 µm; column: 75 µm i.d., 150 mm; Dionex). Buffers were 0.1% HCOOH in water (A) and 0.1% HCOOH in ACN (B). The peptide separation was achieved with a linear gradient from 5 to 30% B for 28 min at 300 nL.min^{-1}. Including the regeneration step at 95% B and the equilibration step at 95% A, one run took 45 min.

Eluted peptides were analyzed on-line with a LTQ XL ion trap (Thermo Electron, Thermo Fisher Scientific Inc, Courtaboeuf, France) using a nanoelectrospray interface. Ionization (1.5 kV ionization potential) was performed with liquid junction and a noncoated capillary probe (10 µm i.d.; New Objective). Peptide ions were analyzed using Xcalibur 2.07 (Thermo Fisher Scientific Inc) with the following data-dependent acquisition steps: (1) full MS scan (mass-to-charge ratio (m/z) 300 to 1400, centroid mode) and (2) MS/MS (qz = 0.25, activation time = 30 ms, and collision energy = 35%; centroid mode). Step 2 was repeated for the three major ions detected in step 1. Dynamic exclusion was set to 30 s.

A database search was performed with X! Tandem (version 2009.04.01.1) (http://www.thegpm.org/TANDEM/). Enzymatic cleavage was declared as a trypsin digestion with one possible miscleavage. Cys carboxyamidomethylation and Met oxidation were set to static and possible modifications, respectively. Precursor mass and fragment mass tolerance were 2.0 and 0.8, respectively. A refinement search was added with similar parameters except

that semi-tryptic peptides and possible N-ter proteins acetylation were searched. The Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org/) database (32825 entries, version 8) and a contaminant database (trypsin, keratins, ...) were used. Only peptides with an E-value smaller than 0.1 were reported.

Identified proteins were filtered and grouped using X! Tandem parser (http://pappso.inra.fr/bioinformatique.html) according to: (1) a minimum of two different peptides was required with an E-value smaller than 0.05, (2) a protein E-value (calculated as the product of unique peptide E-values) smaller than 10^{-4}. To take redundancy into account, proteins with at least one peptide in common were grouped. Within each group, proteins with at least one specific peptide relatively to other members of the group were reported as sub-groups.

5.5. Bioinformatics

Sub-cellular location, length of signal peptides, prediction of transmembrane domains, homologies to other proteins and protein functional domains were predicted as previously described (Minic et al., 2007) or using the *ProtAnnDB* database (http://www.polebio.scsv.ups-tlse.fr/ProtAnnDB/) (San Clemente et al., 2009). *N*-glycosylation consensus sequences were determined using Prosite (http://www.expasy.org/prosite/). *N*-glycans were predicted by *ProTerNyc* (http://www.polebio.scsv.ups-tlse.fr/ProTerNyc/) (Albenne et al., 2009). The list of masses obtained after theoretical digestion of protein, considering possible *N*-glycosylations on predicted consensus sites was compared to the list of experimental masses obtained after MALDI-TOF MS analysis with a mass tolerance up to 200 ppm. A manual validation of glycan prediction was performed on MS spectra, considering only peaks with a significant signal/noise ratio and not used for any protein identification in the corresponding band.

References

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

Albenne, C., Canut H, Boudart, G., Zhang, Y., San Clemente, H., Pont-Lezica, R., Jamet, E., 2009. Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships. Mol. Plant 2, 977-989.

Alvarez, S., Goodger, J.Q., Marsh, E.L., Chen, S., Asirvatham, V.S., Schachtman, D.P., 2006. Characterization of the maize xylem sap proteome. J. Proteome Res. 5, 963-72.

Alvarez-Manilla, G., Atwood, J.r., Guo, Y., Warren, N.L., Orlando, R., Pierce, M., 2006. Tools for glycoproteomic analysis: size exclusion chromatography facilitates identification of tryptic glycopeptides with *N*-linked glycosylation sites. J. Proteome Res. 5, 701-708.

An, H.J., Froehlich, J.W., Lebrilla, C.B., 2009. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. Curr. Opin. Chem. Biol. 13, 421-426.

Banerjee, R., Mande, S.C., Ganesh, V., Das, K., Dhanaraj, V., Mahanta, S.K., Suguna, K., Surolia, A., Vijayan, M., 1994. Crystal structure of peanut lectin, a protein with an unusual quaternary structure. Proc. Natl. Acad. Sci. USA 91, 227-231.

Bhushan, D., Pandey, A., Chattopadhyay, A., Choudhary, M.K., Chakraborty, S., Datta, A., Chakraborty, N., 2006. Extracellular matrix proteome of chickpea (*Cicer arietinum* L.) illustrates pathway abundance, novel protein functions and evolutionary perspect. J. Proteome Res. 5, 1711-20.

Bond, M.R., Kohler, J.J., 2007. Chemical methods for glycoprotein discovery. Curr. Opin. Chem. Biol. 11, 52-58.

Borderies, G., Jamet, E., Lafitte, C., Rossignol, M., Jauneau, A., Boudart, G., Monsarrat, B., Esquerré-Tugayé, M.T., Boudet, A., Pont-Lezica, R., 2003. Proteomics of loosely bound cell wall proteins of *Arabidopsis thaliana* cell suspension cultures: A critical analysis. Electrophoresis 24, 3421-3432.

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.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254.

Brady, J., Sadler, I., Fry, S., 1996. Di-isodityrosine, a novel tetrametric derivative of tyrosine in plant cell wall proteins: a new potential cross-link. Biochem J 315, 323-327.

Brady, J., Sadler, I., Fry, S., 1998. Pulcherosine, an oxidatively coupled trimer of tyrosine in plant cell walls: its role in cross-link formation. Phytochemistry 47, 349-353.

Bunkenborg, J., Pilch, B.J., Podtelejnikov, A.V., Wisniewski, J.R., 2004. Screening for *N*-glycosylated proteins by liquid chromatography mass spectrometry. Proteomics 4, 454-465.

Catalina, M.I., Koeleman, C.A.M., Deelder, A.M., Wuhrer, M., 2007. Electron tranfer dissociation of *N*-glycopeptides: loss of the entire N-glycosylated asparagine side chain. Rapid Commun. Mass Spectrom. 21, 1053-1061.

Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., Haslam, S.M., 2008. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. J. Proteome Res. 7, 1650-1659.

Chen, X., Kim, S., Cho, W., Rim, Y., Kim, S., Kim, S., Kang, K., Park, Z., Kim, J., 2008. Proteomics of weakly bound cell wall proteins in rice calli. J. Plant Physiol. 166, 665-685.

Cho, W., Chen, X., Chu, H., Rim, Y., Kim, S., Kim, S., Kim, S.-W., Park, Z.-Y., Kim, J.-Y., 2009. The proteomic analysis of the secretome of rice calli. Physiol. Plant. 135, 331-3

Cho, Y., Chrispeels, M., 1976. Serine-*O*-galactosyl linkages in glycopeptides from carrot cell walls. Phytochemistry 15, 165-169.

Edge, A.S.B., 2003. Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. Biochem J. 376, 339-350.

Ellis, C., Karafyllidis, I., Wasternack, C., Turner, J., 2002. The Arabidopsis mutant *cev1* links cell wall signaling to jasmonate and ethylene responses. Plant Cell 14, 1557-1566.

Faye, L., Boulaflous, A., Benchabane, M., Gomord, V., Michaud, D., 2005. Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. Vaccine 23, 1770-1778.

Feiz, L., Irshad, M., Pont-Lezica, R.F., Canut, H., Jamet, E., 2006. Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from *Arabidopsis* hypocotyls. Plant Methods 2, 10.

Ferris, P.J., Woessner, J.P., Waffenschmidt, S., Kilz, S., Drees, J., Goodenough, U.W., 2001. Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. Biochemistry 40, 2978-2987.

Gray, J., 1998. Heterogeneity of glycans at each *N*-glycosylation site of horseradish peroxidase. Carbohydr. Res. 311, 61-69.

Gray, J., Yang, B., Hull, S., Venzke, D., Montgomery, R., 1996. The glycans of soybean peroxidase. Glycobiology 6, 23-32.

Hagglund, P., Bunkenborg, J., Elortza, F., Jensen, O.N., Roepstorff, P., 2004. A new strategy for identification of *N*-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. J. Proteome Res. 3, 556-566.

Hakansson, K., Cooper, H.J., Emmett, M.R., Costello, C.E., Marshall, A.G., Nilsson, C.L., 2001. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an *N*-glycosylated tryptic peptide to yield complementary sequence information. Anal. Chem. 73, 4530-4536.

Harvey, D.J., 2005. Proteomic analysis of glycosylation: structural determination of *N*- and *O*-linked glycans by mass spectrometry. Expert Rev. Proteomics 2, 87-101.

Homer, R.B., Roberts, K., 1979. Glycoprotein conformation in plant cell walls. Planta 146, 217-222.

Irshad, M., Canut, H., Borderies, G., Pont-Lezica, R., Jamet, E., 2008. A new picture of cell wall protein dynamics in elongating cells of Arabidopsis thaliana: Confirmed actors and newcomers. BMC Plant Biol. 8, 94.

Jamet, E., Albenne, C., Boudart, G., Irshad, M., Canut, H., Pont-Lezica, R., 2008. Recent advances in plant cell wall proteomics. Proteomics 8, 893-908.

Jamet, E., Canut, H., Boudart, G., Pont-Lezica, R., 2006. Cell wall proteins: a new insight through proteomics. Trends Plant Sci. 11, 33-39.

Jeyaprakash, A., Jayashree, G., Mahanta, S., Swaminathan, C., Sekar, K., Surolia, A., Vijayan, M., 2005. Structural basis for the energetics of jacalin-sugar interactions: promiscuity *versus* specificity. J. Mol. Biol. 347, 181-188.

Jung, Y.-H., Jeong, S.-H., Kim, S., Singh, R., Lee, J.-E., Cho, Y.-S., Agrawal, G., Rakwal, R., 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.

Kieliszewski, M., 2001. The latest hype on Hyp-*O*-glycosylation codes. Phytochemistry 57, 319-323.

Kimura, Y., Hess, D., Sturm, A., 1999. The *N*-glycans of jack bean α-mannosidase. Eur. J. Biochem. 264, 168-175.

King, S.J., Hippe, K.R., Weiser, J.N., 2006. Deglycosylation of human glycoconjugates by the sequential activities of exoglycosidases expressed by *Streptococcus pneumoniae*. Mol. Microbiol. 59, 961-974.

Küster, B., Naven, T.J., Harvey, D.J., 1996. Rapid approach for sequencing neutral oligosaccharides by exoglycosidase digestion and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. J. Mass Spectrom. 31, 1131-1140.

Laemmli, U.K., 1970. Cleavage of the structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.

Lamport, D.T.A., 1967. Hydroxyproline-*O*-glycosidic linkage of the plant cell wall glycoprotein extensin. Nature 216, 1322-1324.

Lige, B., Shengwu, M., van Huystee, R., 2001. The effects of the site-directed removal of *N*-glycosylation from cationic peanut peroxidase on its function. Arch. Biochem. Biophys. 386, 17-24.

Lim, J.-M., Aoki, K., Angel, P., Garrison, D., King, D., Tiemeyer, M., Bergmann, C., Wells, L., 2009. Mapping glycans onto specific *N*-linked glycosylation sites of *Pyrus communis* PGIP redefines the interface for EPG-PGIP interactions. J. Proteome Res. 8, 673-680.

Liu, C., Mehdy, M., 2007. A nonclassical arabinogalactan protein gene highly expressed in vascular tissues, AGP31, is transcriptionally repressed by methyl jasmonic acid in Arabidopsis. Plant Physiol. 145, 863-874.

McDonald, C.A., Yang, J.Y., Marathe, V., Yen, T.-Y., Macher, B.A., 2009. Combining results from lectin affinity chromatography and glycocapture approaches substantially improves the coverage of the glycoproteome. Mol. Cell. Proteomics 8, 287-301.

Minic, Z., 2008. Physiological roles of plant glycoside hydrolases. Planta 227, 723-740.

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

Mirgorodskaya, E., HAssan, H., Wandall, H.H., Clausen, H., Roepstorff, P., 1999. Partial vapor-phase hydrolysis of peptide bonds: a method for mass spectrometric determination of *O*-glycosylated sites in glycopeptides. Anal. Biochem. 269, 54-65.

Morelle, W., Canis, K., Chirat, F., Faid, V., Michalski, J.-C., 2006. The use of mass spectrometry for the proteomic analysis of glycosylation. Proteomics 6, 3993-4015.

Nothnagel, E., 1997. Proteoglycans and related components in plant cells. Int Rev Cytol 174, 195-291.

Owens, N.W., Stetefeld, J., Lattova, E., Schweizer, F., 2010. Contiguous *O*-galactosylation of 4(*R*)-hydroxy-L-proline residues forms very stable polyproline II helices. J. Am. Chem. Soc. 132, 5036-5042.

Patel, T., Bruce, J., Merry, A., Bigge, C., Wormald, M., Jaques, A., Parekh, R., 1993. Use of hydrazine to release in intact and unreduced form both *N*- and *O*-linked oligosaccharides from glycoproteins. Biochemistry 32, 679-693.

Pont-Lezica, R., Minic, Z., San Clemente, H., Roujol, D., Jamet, E., 2010. Plant cell wall functional genomics: Novelties from proteomics. In *Advances in Genetics Research*, vol. 1, chapter 10 (ed. M. A. Osborne). Hauppauge, NY, USA: Nova Science Publishers, Inc.

Qiu, R., Regnier, F.E., 2005. Use of multidimensional lectin affinity chromatography in differential glycoproteomics. Anal. Chem. 77, 2802-2809.

Ringli, C., 2010. The hydroxyproline-rich glycoprotein domain of the Arabidopsis LRX1 requires Tyr for function but not for insolubilization in the cell wall. Plant J (in press), doi: 10.1111/j.1365-313X.2010.04270.x.

Roberts, K., 1990. Structures at the plant cell surface. Curr. Opin. Cell Biol. 2, 920-928.

Roberts, K., 2001. How the cell wall acquired a cellular context. Plant Physiol. 125, 127-30.

Robinson, D., Andreae, M., Sauer, A., 1985. Hydroxyproline-rich glycoprotein biosynthesis: a comparison with that of collagen. In *Biochemistry of plant cell walls*, eds. C. T. Brett and J. R. Hillman), pp. 155-176. Cambridge: Cambridge University Press.

Roxrud, I., Lid, S., Fletcher, J., Schmidt, E., Opsahl-Sorteberg, H.-G., 2007. *GASA4*, one of the 14-member *Arabidopsis* GASA family of small polypeptides, regulates flowering and seed development. Plant Cell Physiol. 48, 471-483.

San Clemente, H., Pont-Lezica, R., 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.

Scheler, C., Lamer, S., Pan, Z., Li, X.P., Salnikow, J., Jangblut, P., 1998. Peptide mass fingerprinting sequences coverage from differentially stained proteins in two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). Electrophoresis 19, 918-927.

Schultz, C.J., Ferguson, K.L., Lahnstein, J., Bacic, A., 2004. Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. J. Biol. Chem. 279, 455103-45511.

Schwientek, T., Mandel, U., Roth, U., Müller, S., Hanisch, F.-G., 2007. A serial lectin approach to the mucin-type *O*-glycoproteome of *Drosophila melanogaster* S2 cells. Proteomics 7, 3264-3277.

Seifert, G., Roberts, K., 2007. The biology of arabinogalactan proteins. Annu. Rev. Plant Biol. 58, 137-161.

Sommer-Knudsen, J., Clarke, A.E., Bacic, A., 1996. A galactose-rich, cell-wall glycoprotein from styles of *Nicotiana alata*. Plant J. 9, 71-83.

Sparbier, K., Koch, S., Kessler, I., Wenzel, T., Kostrzewa, M., 2005. Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. J. Biomol. Tech. 16, 407-413.

Spiro, R.G., 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 12, 43R-56R.

Stafstrom, J.P., Staehelin, L.A., 1986. The role of carbohydrate in maintaining extensin in an extended conformation. Plant Physiol. 81, 242-246.

Tarentino, A.L., Gomez, C.M., Plummer, T.H.J., 1985. Deglycosylation of asparagine-linked glycans by peptide:*N*-glycosidase F. Biochemistry 24, 4665-4671.

Tissot, B., North, S.J., Ceroni, A., Pang, P.-C., Panico, M., Rosati, F., Capone, A., Haslam, S.M., Dell, A., Morris, H.R., 2009. Glycoproteomics: Past, present and future. FEBS Lett. 583, 1728-1735.

Vogel, J., Raab, T., Somerville, C., Somerville, S., 2004. Mutations in *PMR5* result in powdery mildew resistance and altered cell wall composition. Plant J. 40, 968-978.

Wada, Y., Tajiri, M., Yoshida, S., 2004. Hydrophilic affinity isolation and MALDI multiple-stage tandem mass spectrometry of glycopeptides for glycoproteomics. Anal. Chem. 76, 6560-6565.

Wang, L., Li, F., Sun, W., Wu, S., Wang, X., Zhang, L., Zheng, D., Wang, J., Gao, Y., 2006. Concanavalin A-captured glycoproteins in healthy human urine. Mol. Cell. Proteomics 5, 560-562.

Watson, B.S., Lei, Z., Dixon, R.A., Sumner, L.W., 2004. Proteomics of *Medicago sativa* cell walls. Phytochemistry 65, 1709-1720.

Willats, W.G., Knox, J.P., 1996. A role for arabinogalactan-proteins in plant cell expansion: evidence from studies on the interaction of β-glucosyl Yariv reagent with seedlings of *Arabidopsis thaliana*. Plant J. 9, 919-925.

Wuhrer, M., de Boer, A.R., Deelder, A.M., 2009. Structural glycomics using hydrophilic interaction chromatography (HILIC) with mass spectrometry. Mass Spectrom. Rev. 28, 192-206.

Yariv, J., Lis, H., Katchalski, E., 1967. Precipitation of arabic acid and some

seed polysaccharides by glycosylphenylazo dyes. Biochim J 105, 1C-2C.

Zhang, H., Li, X.-J., Martin, D.B., Aebersold, R., 2003. Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nature Biotechnol. 21, 660-666.

Zhang, S., Yang, C., Peng, J., Sun, S., Wang, X., 2009. *GASA5*, a regulator of flowering time and stem growth in *Arabidopsis thaliana*. Plant Mol. Biol. 69, 745-759.

Zhao, J., Simeone, D.M., Heidt, D., Anderson, M.A., Lubman, D.M., 2006. Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: Application to pancreatic cancer serum. J. Proteome Res. 5, 1792-1802.

Zhu, J., Chen, S., Alvarez, S., Asirvatham, V.S., Schachtman, D.P., Wu, Y., Sharp, R.E., 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, 311-25.

Supplementary data

Table S1. Identification by MALDI-TOF MS of *N*-glycoproteins separated by ConA affinity chromatography.

Table S2. Identification by MALDI-TOF MS of glycoproteins separated by multidimensional lectin affinity chromatography.

Table S3. Identification by MALDI-TOF MS of glycoproteins separated by boronic acid chromatography.

Table S4. Identification by LC-MS/MS of proteins retained on ConA with the multidimensional lectin affinity chromatography.

Table S5. Proteins identified in the glycoproteome of 11-day-old etiolated hypocotyls of *A. thaliana.*

Figure S1. Analysis of CWPs by SDS-PAGE after separation by ConA chromatography.

Figure S2. Analysis of CWPs by SDS-PAGE after separation by multidimensional lectin chromatography.

Figure S3. Analysis of CWPs by SDS-PAGE after separation by boronic acid chromatography.