Identification of glycosyltransferases involved in cell wall synthesis of wheat endosperm

M. Suliman, Anne-Laure Chateigner-Boutin, M. Francin-Allami, A. Partier, B. Bouchet, J Salse, C. Pont, J. Marion, H. Rogniaux, D. Tessier, et al.

Identification of Golgi-localized glycosyltransferases involved in cell wall synthesis of wheat endosperm

M. Suliman[a], AL. Chateigner-Boutin[a], M. Francin-Allami[a], A. Partier[b], B. Bouchet[a], J. Salse[b], C. Pont[b], J. Marion[c], H. Rogniaux[a], D. Tessier[a], F. Guillon[a], C. Larré[a*]

[a]INRA, UR1268 Biopolymères, Interactions Assemblages, F-44316 Nantes, France.
[b]INRA, UMR1095 Génétique, Diversité & Ecophysiologie des Céréales, F-63100 Clermont-Ferrand, France.
[c]Institut des Sciences du Végétal, F-91198 Gif-sur-Yvette, France.

MS and ALCB contributed equally to the work.

Abstract:

Plant cell walls are complex structures critical for plant fitness and valuable for human nutrition as dietary fibre and for industrial uses such as biofuel production. The cell wall polysaccharides in wheat endosperm consist of two major polymers, arabinoxylans and beta-glucans, as well as other minor components. Most of these polysaccharides are synthesized in the Golgi apparatus but the mechanisms underlying their synthesis have yet to be fully elucidated and only a few of the enzymes involved have been characterized. To identify actors involved in the wheat endosperm cell wall formation, we used a subcellular fractionation strategy to isolate Golgi-enriched fractions from endosperm harvested during active cell wall deposition. The proteins extracted from these Golgi-enriched fractions were analysed by LC-MS/MS. We report the identification of 1135 proteins among which 64 glycosyltransferases distributed in 17 families. Their potential function in cell wall synthesis is discussed. In addition, we identified 63 glycosylhydrolases, some of which may be involved in cell wall remodeling. Several glycosyltransferases were validated by showing that when expressed as fusion proteins with a fluorescent reporter, they indeed accumulate in the Golgi apparatus. Our results provide new candidates potentially involved in cell wall biogenesis in wheat endosperm.

Keywords:

Glycosyltransferases / Golgi apparatus / subcellular fraction / wheat endosperm / proteomics

Abbreviations:

AX arabinoxylan, MLG (1-3) (1-4) β-D-glucan or mixed-linked glucan, GT glycosyltransferases, GH glycosylhydrolase

1 Introduction

Plant cells are surrounded by cell walls composed of polysaccharides, proteins and, in some specialized cells, various non-carbohydrate substances (e.g., lignins). Cell wall composition varies between species and between cell types and developmental stages within a given plant species. Cell wall polysaccharides have been grouped into cellulose, hemicelluloses and pectins [1]. Pectins and hemicelluloses are components of the wall 'matrix', within which the cellulose framework is embedded. The walls of most dicots and non-grass monocots (type I walls) contain about equal amounts of xyloglucan and cellulose. Grasses are noteworthy for the unusual composition of their cell walls (type II walls), which contain less pectin and xyloglucan, but more heteroxylan than the walls of other higher plants. The cell wall can accumulate extensive interconnecting networks of phenylpropanoids, particularly as the cells stop expanding and start developing secondary cell walls. The chemical structure of the components and the way that they are assembled are key factors underlying cell wall properties. These properties are important not only for plant fitness but also for the processing quality of crop materials (e.g., degradation potential for biofuel production and bread-making quality of wheat flour).

The cell wall composition of wheat grain endosperm is very specific, essentially composed of arabinoxylan (AX) (70%) and (1-3) (1-4) β-D-glucan or mixed-linked glucan (MLG) (20%) with minor amounts of cellulose (4%) and glucomannan (7%) [2, 3]. The synthesis of cell wall polysaccharides is activated during wheat grain development. Drastic changes in cell wall composition and structure were observed in the developing endosperm [4-6]. In particular, a transient accumulation of (1-3) β-D-glucan (callose) and xyloglucan has been reported [4, 7].

The metabolism of cell wall polysaccharides requires the intervention of numerous enzymes. Enzymes acting on carbohydrates are listed and classified on the basis of their sequence similarities and activities in the Carbohydrate-Active enZYmes

database (CAZy, http://www.cazy.org/) [8]. Currently, CAZy contains 94 glycosyltransferase (GT) families covering 122 enzymatic activities, 130 glycosylhydrolase (GH) families covering 158 activities, as well as 22 families of polysaccharide lyases, 16 of carbohydrate esterases (CEs) and 64 families of carbohydrate-binding modules (CBMs). Plants are under-represented in this database, essentially because there are less fully-sequenced genomes of plant species compared to other kingdoms such as bacteria. However, in the sequenced model *Arabidopsis thaliana,* about 3.3% of all of the genes encode carbohydrate-active enzymes [9, 10].

Cell wall polysaccharides are synthesized from activated precursors (nucleotide sugars) by enzymatic complexes. Several groups have accumulated evidence for two major subcellular locations for cell wall polysaccharide synthesis. Cellulose microfibrils and callose are synthesized at the plasma membrane surface and the polysaccharides are deposited directly into the cell wall [11-13]. Pectin and hemicellulose are synthesized in the Golgi apparatus to then be transported to the wall [14-18].

The GT superfamily contains enzymes that synthesize polysaccharides such as starch and cell wall polysaccharides. GTs are also involved in the glycosylation of many other substrates such as proteins, lipids and secondary metabolites. Several GTs involved in cell wall polysaccharide synthesis have been localized in the Golgi apparatus [19-21]. GTs involved in protein glycosylation may also be localized in the Golgi apparatus and in the endoplasmic reticulum (ER) [22, 23], whereas GTs involved in starch synthesis were localized in plastids [24].

To gain access to the cellular machinery involved in the synthesis of the cell wall hemicelluloses in the wheat grain endosperm, we used a subcellular proteomic approach that targeted the Golgi apparatus to focus on the direct actors of the polysaccharide biosynthesis, i.e., the enzymes. Subcellular fractionation has been widely described across all cell types and tissues, and has been shown to be particularly efficient for organelles such as mitochondria [25] and chloroplasts [26], leading to pure fractions of these organelles. However, subcellular proteomics remains challenging for the endomembrane system. In the case of the Golgi apparatus and the ER, their similar density makes it very tedious to isolate them as pure fractions. Moreover, the large number of proteins trafficking from the ER to the Golgi apparatus inevitably leads to the identification of ER proteins, even in an ER-free Golgi

preparation [27]. Efficiently monitoring the subcellular fractionation and assessing the possible contamination of the isolated fractions are additional difficulties [23, 28]. Finally, many GTs are membrane proteins characterized by issues such as poor solubility, low abundance and recalcitrance to classical proteomic techniques.

In the present work, we have developed a subcellular strategy based on wheat endosperm Golgi fractionation from developing grain harvested at a stage where cell wall polysaccharides are accumulating. We partitioned integral membrane and water soluble proteins associated with the endomembranes. Using mass spectrometry, we identified 1135 proteins, 64 of which were predicted to be GTs involved in glycosylation processes in the wheat endosperm. We also identified 63 GHs potentially involved in the remodelling or degradation of the endosperm carbohydrates. The proteins identified comprise Golgi residents and proteins in transit in the Golgi apparatus.

2. Materials and methods

2.1. Plant materials:

The plants *Triticum aestivum* cv. Recital were grown in pots in a greenhouse under conditions of natural day length (UMR Amélioration des Plantes et Biotechnologies Végétales, INRA-Rennes, France). To harvest grains at defined developmental stages, individual ears were tagged at flowering. Seed development was calculated on the basis of cumulated temperature in Celsius degrees days (°D) after flowering. Grains were harvested between 250 and 275°D and endosperms were isolated manually and maintained in buffer A (250 mM sucrose, 10 mM HEPES-NaOH, pH 7.4, 1 mM EDTA, 1 mM DTT and anti-protease cocktail) (Roche) on ice before further treatment.

For microscopy studies, the plants were grown in pots in a growth chamber at 23°C day/15°C night at 65% relative humidity under a 16h/8h day/dark photoperiod (UPR 2355/IFR 87, Gif Sur Yvette, France). Wheat grains were harvested at 250°D.

2.2. Electron transmission microscopy

2.2.1 Sample preparation

Pieces of tissue (1 mm^3) were sampled from halved wheat grains. Samples were prepared as follows for morphological assay. Samples were loaded into copper hats filled with 1-hexadecene and high-pressure frozen using a HPF-EM PACT 2 freezer (Leica, http://www.leica-microsystems.com). The hats were immediately stored in

liquid nitrogen until the freeze substitution procedure was initiated. Samples were transferred to a freeze substitution automate (AFS2/FSP, Leica) and pre-cooled to -90°C. Samples were substituted in anhydrous acetone with 2% osmium tetroxyde at -90°C for 125 h. The temperature was gradually raised to -30°C for 60 h and stabilized for 12 h, then gradually raised to -4°C (34 h) and stabilized for 3 h. Samples were rinsed in anhydrous acetone several times and slowly infiltrated and embedded in Epon's resin.

For immunolabeling, samples were prepared as described by Chevalier et al. [29]. High-pressure freezing was performed with the HPF-EM PACT I freezer. Samples were then transferred to an AFS. Samples were finally embedded in London Resin White (LRW, LRWhite R1280 HARD GRADE, London Resin Company Limited) and polymerization was performed in the AFS apparatus at -15°C under UV light for 48 h.

2.2.2 Polysaccharide staining and immunolabeling

Periodic acid-thiosemicarbazide-silver proteinate staining (PATAg) [30] for polysaccharide detection.

Ultra-thin sections (80 nm) were cut from Epon's embedded samples using an ultramicrotome (MICROM MT-7000) equipped with a diamond knife and directly floated in 1% periodic acid (MERCK) solution for 30 min. They were then rinsed 5 times in deionized water before incubation in thiosemicarbazide (0.2%) (MERCK) diluted in 20% acetic acid for 17 h. Sections were rinsed 5 times in decreasing acetic acid concentrations (20, 10, 5 and 2.5%), 4 times in deionized water, and stained with 1% aqueous silver proteinate (PROLABO) in the dark for 30 min. After rinsing 5 times in deionized water, sections were mounted on copper grids. Sections were examined with a JEOL 1230 transmission electron microscope with an accelerating voltage of 80 KeV (IBISA/BioGenOuest Biopolymers, Interations, Structural Biology platform (BIBS), UR 1268 BIA, INRA Angers-Nantes).

Immunolabeling

Ultra-thin sections (80 nm) were cut from LRW embedded samples and collected on nickel grids. Sections were incubated in a blocking solution of 3% (w/v) BSA in 20 mM PBS, pH 7.2, to block non-specific labeling for 30 min at room temperature. Sections were then incubated in a solution containing the mouse monoclonal antibody anti-AX1 [31] and the rabbit polyclonal antibody raised against the pea reversibly glycosylated polypeptide 1 (anti-RGP1) (a gift from K. Dhugga), diluted 1:20 and 1:2500,

respectively, in 20 mM PBS supplemented with 1% BSA and 0.05% Tween-20 (Sigma-Aldrich) for 1 h at room temperature. The sections were extensively washed in buffer used for diluting the primary antibodies and then incubated 1 h at room temperature in the dark with goat-anti-mouse IgG or goat–anti rabbit conjugated with 1 nM colloidal gold complexes diluted 1:20 (v/v) in the respective buffer chosen for the primary antibodies. Labeling was then intensified with the silver enhancement kit (Aurion) according to the manufacturer's instructions. After washing, the grids were stained with 2% uranyl acetate. Sections were examined with a JEOL 1230 transmission electron microscope with an accelerating voltage of 80 KeV.

2.3. Preparation of the microsomal fraction

Six experiments were conducted. For each experiment, approximately 20 g of endosperm were ground with a mortar and pestle on ice in buffer A and then homogenized on ice using a polytron (Kinematica AG, Dispersing and Mixing Technology) for two pulses of 7 s at 7500 rpm. The endosperm lysate was then centrifuged twice for 5 min at 2,200 g to separate cell debris. The supernatant was subsequently loaded on the top of an SW41 centrifuge tube that contained 8 mL of a 18% iodixanol prepared from Optiprep solution (Sigma-Aldrich) and centrifuged at 100,000 g for 2 h at 4°C (Beckman Coulter SW41 rotor). The interfacial fraction was collected as the microsome fraction.

2.4. Generation of Golgi-enriched fractions

Four mL of the recovered microsomal fraction were then collected from the interface and suspended in a final solution of 2.8 mM HEPES, pH 7.4, 0.28 mM EDTA, 0.28 mM DTT, 14% iodixanol and gently mixed to avoid organelle damage. These organelles were further sub-fractionated by generating an iodixanol gradient by ultracentrifuging at 200,000 g for 12 h in a Ti 70 rotor (Beckman Coulter). Fifteen fractions of 1.8 mL were harvested from the top of the gradient and their densities were determined using a refractometer (Maselli Misure; LR-01 Digital Refractometer). The protein concentration of each fraction was then determined using the NI™ Protein Assay (G-Biosciences).

2.5. Separation of organelle membranes

The collected fractions, microsomal and Golgi-enriched fractions, included organelles and their associated proteins. Cold deionized water was added to each fraction to reach a final volume of 22 mL in a centrifuge bottle and ultracentrifuged at 100,000 g for 2 h. The pellet consisted of organelle membranes from which membrane proteins were extracted for further identification; they will be referred to as MP (membrane proteins), MPG when issued from Golgi-enriched fractions. The supernatants of the microsomal and fraction samples were extensively dialyzed against cold water at 4°C using a semi-permeable membrane with a molecular weight cut-off of 3,500 (Spectra Pro). The dialysis water was replaced several times over the course of three days at 4°C. The dialyzed proteins were then collected and lyophilized; these fractions will be referred to as AP (proteins associated with the membranes).

2.6. Monitoring of ER and Golgi enrichment in fractions by Western blot analysis

The hydrophobic nature of membrane proteins led us to choose an anionic detergent for their solubilization. Both membrane pellets and lyophilized AP were therefore solubilized in 100 µL of Tris–HCl buffer, pH 6.8, containing 2% (w/v) SDS, 10% (v/v) glycerol and bromophenol blue. Solubilized proteins were reduced by the addition of 5 µL of β-mercaptoethanol and separated by SDS–PAGE on 12% acrylamide gels. After separation, gels were electroblotted onto nitrocellulose membranes (Amersham Biosciences) in 25 mM Tris, 192 mM glycine, 0.1% SDS and 20% ethanol. Semi-dry transfer was achieved at 100 V for 1.5 h.

Membranes were washed in PBS solution, blocked at room temperature in 5% dry milk in PBS, 0.1% Tween-20, and then washed twice for 10 min in PBS, pH 7.4, 0.1% Tween-20. Two primary antibodies were used for incubation. The polyclonal antibody anti-RGP1 (see above for immunolabeling), which recognizes a Golgi-localized protein, was used at 1:10,000 dilution for 2 h at room temperature. The antibody anti-BiP2 (Binding immunoglobulin protein2, Agrisera AB, Sweden), specific for Arabidopsis ER, was used at 1:2,000 dilution for 2 h at room temperature. After washing 2 times in PBS, 0.1% Tween-20, the secondary antibodies, consisting of peroxidase-conjugated anti-rabbit IgG, were used at 1:5,000 and 1:200,000 dilution, respectively, for 2 h at room temperature. After washing, nitrocellulose membranes were incubated in a chemiluminescent substrate (Super Signal West Dura Extended Duration Substrate, Pierce) for 6 min and dried between two paper sheets according

to the manufacturer's instructions. Luminescence was then detected using a camera (Luminescent Image Analyzer LAS 3000; Fuji Film).

2.8. Analysis of peptides by mass spectrometry

Proteins extracted from fractions that exhibited a high Golgi apparatus content as revealed by RGP1 were separated by SDS–PAGE on 12% acrylamide gels. The migration was performed on 2 cm of gel in order to increase the efficiency of tryptic digestion. After separation, proteins were stained with Coomassie Brilliant Blue G250 (Sigma-Aldrich), as described by Devouge et al. [32].

2.8.1. Protein digestion

For each SDS-PAGE lane, 20 bands were manually excised from the gels to be hydrolyzed. In brief, the gel pieces were first washed three times in 50% (v/v) ACN/25 mM ammonium bicarbonate, and the proteins were then reduced using DTT and alkylated with iodoacetamide prior to tryptic digestion, according to Larré et al. [33].

2.8.2. Liquid chromatography and mass spectrometry

Nanoscale capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses of the digested proteins were performed using an Ultimate 3000 RSLC system (Dionex) coupled with an LTQ-Orbitrap VELOS mass spectrometer (Thermo Fisher). Chromatographic separation was conducted on a reverse-phase capillary column (Acclaim Pepmap C18 2 µm 100A, 75-µm i.d. x 15-cm length; Dionex) at a flow rate of 300 nL.min-1.

Mobile phases were composed as indicated: A (99.9% water, 0.1% formic acid), B (90% acetonitrile, 0.08% formic acid). The gradient consisted of a linear increase from 4% to 45% of B in 30 min, followed by a rapid increase to 70% within 1 min. Composition was maintained at 70% B for 5 min and then decreased to 4% B for re-equilibration of the column.

Mass data acquisitions were performed using Xcalibur 2.1 software. Full MS scans were acquired at high resolution (FWMH 30,000) in the Orbitrap analyzer (mass-to-charge ratio (m/z): 400 to 2000), while collision-induced dissociation (CID) spectra were recorded on the five most intense ions in the linear LTQ traps.

2.9. Database search and interpretation

Raw data collected during LC-MS/MS analyses were processed into mgf format files and further searched against databanks using X!Tandem Spectrum Modeler. Protein

identification was achieved by confronting mass data (MS and MS/MS spectra) against four databases: (i) UniProt Knowledgebase (http://www.uniprot.org/) : restricted to Viridiplantae (version 2011_03), (ii) Gene Index Databases Wheat release 12.0, August 2010 (Wheat GI) (http://compbio.dfci.harvard.edu/tgi/), (iii) an in-house glycosyltransferases databank (GTIDB) (http://wwwappli.nantes.inra.fr:8180/GTIDB) [34], and (iv) a contaminant database including keratins and trypsin.

A database search was performed with XTandem 2008.02.01 (http://www.thegpm.org/TANDEM/) via the X!tandem pipeline available at http://pappso.inra.fr/bioinfo/xtandempipeline/ . Enzymatic cleavage was declared as a trypsin digestion with one possible miscleavage event. Fixed modifications of cysteine residues by iodoacetamide as well as oxidized methionins were considered. Precursor mass and fragment mass tolerance were set at 5 ppm and 0.8 Da, respectively. One missed trypsin cleavage was allowed. Identified proteins were filtered according to the following specifications: at least two peptides with an E value below 0.001 and a protein E value below 10^{-4}. To take redundancy into account, proteins with at least one peptide in common were grouped.

Protein identifications were compared in the three databanks. In the event of identification in the three databanks, the identifier from UniProt will be preferentially reported, followed by the one from Wheat GI and, finally, the one from GTIDB. The search against the decoy database (Reverse Uniprot restricted to Viridiplantae) gave a mean false-discovery rate of 0.3% for each query when using the $p < 0.01$ significance level.

Tentative functional annotations were assessed by running a BLAST homology search of sequences identified from mass spectrometry against the NCBI nr database using Blast2GO (http://www.blast2go.org). The cut-off stringency for BLAST was $E = 10^{-10}$. For each protein, the best hit from each species was collected if present. The two sets of proteins (protein sequences obtained from spectrometry analysis and their best homologs) were analyzed by applying the Phobius (http://phobius.sbc.su.se/) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) programs to predict transmembrane domains (TM) and signal peptides.

A refined search enabling protein identification from single peptide assignment was performed to obtain more putative GTs. This extended protein list was then blasted against the entire non-redundant sequences of the CAZy database using sequence similarity or protein domains (Pfam) association rules [35]. Those for which an ortholog

was found in the CAZy database (e-value < 0.01) were reported. Proteins with homology to GTs and GHs were listed, and when identified with only one peptide in MS, the quality of the peptide was checked and validated manually.

For a subset of GTs and GHs, the best homolog (best Blast hit) in the rice and *Brachypodium distachyon* genomes were searched using the Blast tool of the Rice Genome Annotation website (http://rice.plantbiology.msu.edu/analyses_search_blast.shtml). Nucleotide sequences were translated with the ExPASy Translate Tool (http://web.expasy.org/translate/) and amino acid sequences were aligned with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

2.10 Transcriptomic analysis

A subset of GT and GH sequences were blasted against the sequence of the wheat Affymetrix probe set using the PLEXdb Blast tool (http://www.plexdb.org/modules/tools/plexdb_blast.php) [36]. The expression pattern of the corresponding “genes” was retrieved from expression studies covering wheat (whole) grain development [37] and different organs and tissues (PlaNet platform WheatNet http://aranet.mpimp-golm.mpg.de/wheatnet) [38]. These analyses may not discriminate all possible homologous copies within the wheat genome such as highly similar paralogs and homoeologous forms of genes from the three genomes of hexaploid wheat. Therefore, we will consider one “gene/transcript” when possibly several forms co-exist.

2.11. Subcellular localization

2.11.1 cDNA templates

Total RNA was isolated from manually dissected endosperm of wheat grains at 250°D. About 1 g of tissue was ground in liquid nitrogen and extracted in 4.5 mL of buffer (0.1 M NaCl; 10 mM Tris HCl, pH 7.4, 1mM EDTA pH8; 1% SDS) with 3 mL phenol-chloroform-IAA (25:24:1). After mixing and centrifugation, the supernatant was extracted twice with 3 mL phenol-chloroform-IAA (25:24:1). The aqueous phase was precipitated by the addition of 300 µL of sodium acetate (3M, pH 5.2) and 6 mL of ethanol. RNA pellets were rinsed in 70% ethanol and dissolved in water. RNA was treated twice with the DNase set (Qiagen) and then purified through the RNeasy MinElute Cleanup kit (Qiagen), following the manufacturer’s recommendations.

Reverse transcription was carried out with 2 µg of total RNA from random hexamers using the Transcriptor First Strand cDNA synthesis kit (Roche).

2.11.2 Constructs

All clones used in this study were constructed using Gateway technology (Invitrogen). The cDNA encoding the four GTs was amplified by PCR using a high fidelity KOD DNA polymerase (Novagen) with the primers listed in Supplemental Table 1. A two-step PCR procedure was carried out. A first PCR with gene-specific primers extended with part of the recombination regions attB1 and attB2 was achieved, and the resulting PCR products served as template for a second PCR, with the primers attB1 and attB2 for the GT47/Wheat GI ID: TC390543 and the GT61/UniProt ID: Q5QPY3 constructs, and the primers U5ATGSDK and U3 for the GT47/UniProt ID: E0ZPV1 and GT75/ Wheat GI ID: TC386312 constructs, to obtain the full recombination regions. Entry clones were obtained via BP reaction in pDONR-207 (Invitrogen). The binary pK7FWG2 and pK7RWG2 vectors containing the 35S promoter of cauliflower mosaic virus (CaMV) and the eGFP (enhanced Green Fluorescent Protein) or mRFP (monomeric Red Fluorescent Protein) respectively were used to construct C terminal fusions to the different GTs.

2.11.3 Plant material and Agrobacterium-mediated transient expression

Nicotiana tabacum sp. plants were cultivated in a controlled growth chamber at 21°C with 14h light/10h dark for 5-6 weeks prior to Agrobacterium infiltration.

Transient expression of fluorescent protein fusions in tobacco leaf epidermal cells was performed as previously described by Sparkes et al. [39]. Briefly, *Agrobacterium tumefasciens* strains GV3101::mp90 were transformed by heat-shock with the binary plasmids. The transformants were then centrifuged and washed once with infiltration buffer (5 g/L glucose, 50 mM MES, 2 mM Na3PO4.12H2O, 100 µM acetosyringone) before being resuspended with the same buffer at an optical density (λ=600 nm) of 0.1 and 0.05 for GT constructs and Golgi markers (rat sialyl transferase signal anchor sequence (ST) fused to eGFP or mRFP), respectively. Lower leaves of *Nicotiana tabacum* sp. plants were infiltrated with the diluted bacteria using a syringe. For co-expression, the bacteria were mixed in appropriate volumes of infiltration buffer prior to injection into the leaves. Fluorescent protein expression was studied 2-3 days post-infiltration.

2.11.4 Confocal laser scanning microscopy

Confocal microscopy experiments were performed at the facilities of the Biopolymers-Structural Biology platform at INRA Nantes. An inverted Nikon A1 confocal laser scanning microscope was used to examine the subcellular localization of GFP and mRFP fluorescence. The samples were examined with a water-immersion x40 objective. GFP and mRFP channels were acquired by simultaneous scanning using 488-561 nm laser lines for excitation: emissions were collected via a photomultiplier through band-pass filters, a 500-530 nm filter and a 570-620 nm filter. Images were processed using NIS-Element Software (Nikon) and Adobe Photoshop 6.0.

3. Results

3.1. Validation of the starting material and antibody used in the subcellular proteomic approach

The wheat grain endosperm at 250°D was manually dissected and examined by fluorescence microscopy, it contains the starchy endosperm as well as the aleurone layer (Supplemental Fig. 1). At this stage, Golgi stacks are scarce compared to ER in aleurone (Fig. 1A) and in starchy endosperm (Fig. 1B) cells, making the isolation of fractions enriched in Golgi membranes challenging. Cell wall PATAg staining (Fig. 1C) and AX immunolabeling (Fig. 1D) indicate that at this stage, cell wall polysaccharides and AX, in particular, are actively produced in Golgi stacks and secreted into the wall. The antibody used to detect Golgi membranes in our fractions was initially raised against pea RGP1, a protein that has been localized in the Golgi cisternae of pea stem subapical cells [12]. We checked the reactivity of this anti-RGP1 against wheat endosperm. Immunogold labeling was detected in the Golgi apparatus of wheat aleurone cells, validating the use of anti-RGP1 to detect Golgi apparatus-enriched fractions (Fig. 2).

3.2. Subcellular fractionation

We developed a subcellular fractionation procedure to obtain fractions enriched in Golgi apparatus from wheat endosperm. We first obtained a microsomal fraction that was further fractionated into 15 subfractions that were characterized for their relative content in Golgi stacks and ER. The fractions obtained in preliminary tests presented a low organelle content, so we decided to pool the fractions obtained from four

experiments. The results described in the next paragraphs were obtained from approximately 120 g of endosperm isolated from wheat grains harvested between 250 and 275°D.

3.2.1. Generation of the microsomal fraction containing Golgi membranes

In order to obtain a sufficient amount of protein for characterization and proteomic analysis from Golgi membrane-enriched fractions, six fractionations were conducted. For each fractionation experiment, we obtained approximately 20 g of endosperm from 1200 wheat grains that were used to generate approximately 4 mL of microsomal fraction by concentration on an iodixanol cushion.

The enrichment of these fractions in organelles was examined by Western blot experiments performed with markers for two subcellular compartments, Golgi and ER with RGP1 and BiP2, respectively (Fig. 3). Western blot analysis revealed that RGP1 protein was mainly present in the associated protein subfraction of the microsomal interface, whereas BiP2 protein was only found in the membrane subfraction. Accordingly, in the subsequent fractionation, the presence of BiP2 was only tested in membranes and that of RGP1 in associated protein fractions (Fig. 3).

3.2.2. Generation of Golgi-enriched fractions

Microsomal fractions were further separated on a gradient generated from an iodixanol solution of 14%, and 15 fractions were collected along the gradient. In each experiment, the shape of the gradient was controlled by measuring the refractive index of the fractions, and a good linearity was observed from densities between 1.05 and 1.1 (results not shown). The Western blots (Fig. 3) show the relative distribution of Golgi stacks and ER along the gradient fractions. The good reproducibility of these gradients led us to pool the fractions produced.

For protein identification, fractions revealing a high content in RGP1 and, at the same time, a low content in BiP2, were selected (Fig. 3). Fractions 6, 7 and 8 were mixed in one pool and referred to as G1, and fractions 11, 12, 13 were pooled and referred to as G2. Both G1 and G2 were further fractionated into membranes and associated proteins. The proteins extracted from the membranes pelleted from G1 and G2 will be referred to as MPG1 and MPG2, whereas the associated proteins recovered from G1 and G2 will be referred to as APG1 and APG2. 1-D electrophoresis was used to reduce the sample complexity; each electrophoresis gel was cut into 20 pieces before in-gel

digestion and peptide extraction. These peptides were then analyzed by nano LC MS/MS.

3.3. Protein identification

The sequencing of the wheat genome is underway. To identify proteins, we therefore used UniProt restricted to Viridiplantae and, in parallel, the wheat GI database and GTIdb, an in-house database listing GT sequences from wheat, Arabidopsis and rice [34], in which each protein is assigned to a GT family according to the CAZy classification. Proteins identified from Wheat GI were annotated by running the BLAST alignment algorithm [40] against the NCBI nr database.

3.3.1 Global analysis

First, a global analysis of all spectra generated from the four fractions was performed by combining the result files corresponding to 81 bands cut in the previously described electrophoresis experiment. This set of data generated a total of 1135 identified proteins validated with at least two peptides, and all proteins with at least one common peptide were grouped to avoid redundancy.

With these criteria, 668 proteins were selected (Supplemental Table 2), and the vast majority of them were identified from the Wheat GI database. Their best homologous proteins were then searched against the GenBank non-redundant protein database (NCBI), and with the exception of one sequence, homologous proteins were found with an e value < 10^{-10}. Only 20% of the sequences were identified in wheat. The others were identified by best homologs mainly found in *Hordeum vulgare* (54%), *Oryza sativa* (15%), followed by other less represented species.

In a second analysis the spectra were grouped on the basis of the fraction they came from, membranes (MPG) or associated (APG) Golgi proteins. Among the 668 proteins, 262 were found exclusively in membrane fractions (MPG1 and MPG2), 196 in the associated protein fractions (APG1 and APG2) and 210 were found to be present in both membranes and soluble fractions. These three sub-datasets were further inspected for the presence of transmembrane domains (TM) and of signal peptides (SP) using two different software programs, Phobius and TMHMM. TM and SP predictions were examined in the sequences retrieved from the MS analysis as well as in that of their best homologs. The output of these predictions [41] is reported in Table 1. Among the 262 sequences identified from the membrane fractions, 43% (112) had

at least one predicted TM domain. This proportion increased to 59% when the analysis was performed on their best homologs retrieved from the NCBI nr database. In the case of sequences identified exclusively in the APG fractions, the proportion of protein predicted with at least one TM drastically decreased to 4%, while among those found in both fractions, 14% were reported with at least one TM domain predicted.

3.3.2 Search for carbohydrate-active enzymes

In order to identify all putative GTs and GHs, we ran BLAST search against the entire non-redundant sequences of the CAZy database by taking, as inputs, all identified protein sequences (1135) which returned a set of 78 sequences with a pfam domain (Supplemental Table 3). Our primary targets were the GTs, but many hits were found in the other classes of carbohydrate-active enzymes. In fact, we identified 29 putative GT homologs in 11 GT families. We also identified seven sequences for carbohydrate-binding modules, five of which also included a glycosyl hydrolase "GH13" domain. We therefore decided to assign them to the GH13 family. We thus identified 46 putative GH homologs distributed in 17 GH families. In addition, one sequence belonging to Ces was retrieved. It was similar to a pectin acetylesterase from the Ces13 family.

In order to identify all putative GTs and GHs, we ran a new BLAST search against the CAZy database based on similarity. This secondary search resulted in identifying six additional GT families (GT2, GT30, GT41, GT61, GT77 and GT92) and two GH families (GH2 and GH12) (Supplemental Table 3). In order to avoid missing low abundant proteins we also examined the GT and GH sequences retrieved by a BLAST search on proteins identified with a single peptide. Three of them were retained as they presented a peptide with a good quality score, reported at the end of Supplemental Table 2.

Finally, our search resulted in an increase in the number of families identified to 17 GT families and 19 GH families (Fig. 4). Further analysis revealed that GTs and GHs were identified in a specific fraction, e.g., either GH18, GT34 in MGP (in blue, Fig. 4) or GH77, GT75 in AGP (in red, Fig. 4). Surprisingly, some of them were found in both fractions.

Among the GT and GH families identified, the GT2, GT34, GT47, GT48, GT61, GT75, GH3, GH9 and GH17 were reported in the literature to contain members potentially involved in cell wall metabolism. The wheat sequences identified in these families were

assigned to their best BLAST hit in the genome of rice and *Brachypodium distachyon* and their putative or characterized functions are reported in Table 2.

3.4 Expression pattern of a subset of candidate genes

We analyzed the expression pattern of the subset listed in Table 2 by mining published and publicly available transcriptomic data. In Supplemental Fig. 2, the results show that several “genes” such as members of the GT34, GT61, GH3, GH9 and GH17 families exhibit a pattern of expression relatively specific to the grain. This result confirmed the importance of choosing the target tissue of interest as starting material for our analysis.

3.5 Experimental validation of the Golgi localization for three chosen GTs

To validate the Golgi localization of candidates identified with the subcellular proteomic approach, we chose GT families known for their potential involvement in cell wall biosynthesis. In particular, we selected GTs for which a complete nucleotide coding sequence was available at the time we did our analysis, and validated them *in silico* after nucleotide and amino acid sequence alignments with their best BLAST hit in rice and *Brachypodium distachyon*. The coding sequences corresponding to the GT75/Wheat GI ID: TC386312, GT47/UniProt ID: E0ZPV1, GT61/UniProt ID: Q5QPY3 (GenBank ID AK333763.1), and GT47/Wheat GI ID: TC390543 were inspected and the first three were found to be complete. In the case of GT47/Wheat GI ID: TC390543, sequence alignments with the best Brachypodium homolog and with wheat ESTs revealed a missing nucleotide in the coding sequence that shifted the reading frame (Supplemental Document 1). The *in silico* validated coding sequences of the three GTs were cloned using cDNA obtained from the endosperm of wheat grain at 250°D. The intracellular localization of the GT proteins was determined by transiently expressing them as fluorescent protein fusions in tobacco leaf epidermal cells. GT75/Wheat GI ID: TC386312, GT47/UniProt ID: E0ZPV1, GT61/UniProt ID: Q5QPY3, and GT47/Wheat GI ID: TC390543 were all expressed as small and mobile bodies in the cells, highly similar to the Golgi bodies. Co-expression of the GTs with a Golgi marker confirmed their localization in the Golgi apparatus (Fig. 5).

4. Discussion

The wheat endosperm cell wall composition is unusual and relatively simple, and appears to be a good model to reach enzymes implicated in the biosynthesis of their constituting polymers. It is thought that part of these enzymes and especially the GTs are located in the Golgi apparatus. Many of these GTs are membrane proteins present in low abundance; they were not detected in global proteomic analyses of the wheat grain endosperm [42-44]. The challenge of this work was to identify carbohydrate-active wheat enzymes at the proteomic level with a special focus on GTs potentially involved in the synthesis of cell wall polysaccharides in the grain endosperm. Additional difficulties had to be overcome. Only a few Golgi apparatus are visible in the starchy endosperm at 250°D. We chose this developmental stage because the endosperm is then easily isolated and because there is an active deposition of cell wall polysaccharides (in particular, the AX starts to accumulate), whereas grain filling with storage compounds (starch and storage proteins) is only beginning. Another difficulty lays in the fact that wheat proteins and especially wheat GTs are not well represented in protein databases.

We developed a procedure for the isolation of Golgi-enriched fractions adapted from that of Graham [45] and Sadowski et al. [46] followed by 1D protein analysis, in-gel digestion and mass spectrometry analysis of the resulting peptides.

Isolation of Golgi-enriched fractions

The microsomal fraction we obtained revealed an enrichment in Golgi stacks and ER. The BiP2 was used as an ER marker and was only revealed in the membranes of our microsomal fraction, corroborating the results obtained by Mitra et al. [47] who identified BiP2 in the membrane sub-proteome of whole Arabidopsis seedlings.

RGP1 was used as a Golgi marker. We, however, detected RGP in both the soluble and the membrane subfractions of the microsome, which confirmed the presence of Golgi stacks in the microsomal fraction and corroborates previous results that report that RGPs have been found in both the membrane and soluble fractions of many species [48, 49]. A RGP was identified in the associated fraction (Wheat GI ID: TC386312) in our proteomic analysis.

Variation of the relative enrichment in Golgi or ER was observed along the gradient, allowing us to collect the most Golgi-enriched fractions for protein identification. The presence of BiP2 in these fractions can be either due to ER contamination or to a possible localization in Golgi as reported recently in Arabidopsis cell culture [50].

Proteomic analysis

The subcellular fractionation and LC–MS/MS analysis of wheat endosperm resulted in the identification of 1135 unique proteins (at least two peptides with an E value < 0.001) that, after grouping, gave 668 non-redundant proteins, 262 from the membrane fractions and 196 from the associated fractions, meaning that 210 were found in both fractions.

Membrane and associated proteins

Proteins only detected in the membrane fractions are expected to be strongly linked to the membranes (integral membrane proteins and proteins anchored to the membrane), whereas proteins within the membrane-associated fractions may be proteins in transit within the endomembrane system and proteins linked to membrane proteins.

The proteins identified in both fractions may be due to a strong association between a protein and a membrane protein within a complex where some but not all proteins are integral membrane proteins as already described by Zeng et al. [28].

Furthermore, as recently reported by Atmodjo et al. [51], the transmembrane domain of some type II proteins are cleaved in the Golgi apparatus, although the processed proteins are retained in the Golgi apparatus within protein complexes. We cannot exclude the possibility that for some of the proteins, the dual location may reflect a contamination due to insufficient membrane washing, although the fractions were extensively diluted (30 times) before membrane pelleting.

The prediction analysis of the transmembrane domain reflects this distribution: only several percent of the proteins within the associated fraction are predicted to contain a TM domain, whereas half of the proteins found only in the membrane fractions are predicted to contain such a domain. This percentage would undoubtedly increase if the wheat protein sequences were complete since many Golgi proteins have their transmembrane domain in the N-terminal region.

Golgi localization

Many proteins transit via the endomembrane system (ER and Golgi) before being directed to their final destination. In particular, secreted proteins undergo glycosylation in the ER and the Golgi apparatus, and it has been postulated that about 50% of plant proteins could be glycosylated [52]. Some proteins are retained in the Golgi apparatus where they carry on their function.

In our Golgi-enriched fractions, we identified proteins that are not supposed to be Golgi residents: abundant proteins (histones, ribosomal proteins, ribulose-1,5-bis-phosphate

carboxylase/oxygenase) that contaminate our fractions, as well as GTs such as the callose synthase (Wheat GI ID: TC451164), most probably localized in the plasma membrane and the starch synthase (UniProt ID: Q43654|SSY1_WHEAT) in plastids. In addition to their transit in the Golgi apparatus, proteins can be recycled via the Golgi apparatus. For example, the cellulose synthase and possibly the callose synthase, both located in the plasma membrane, could be stored in the Golgi apparatus or recycled to be directed where needed [50, 53]. In order to validate the Golgi location of some of the GTs identified, four of them were co-expressed with a Golgi marker in tobacco cells.

Identification of proteins of interest

Storage proteins

Our study focused on one specific developmental stage that corresponds to the very early stage of storage deposition (starch and protein). The quality of our samples was confirmed by the low amount of storage proteins found by our analysis. Less than 1% of the identified proteins were storage proteins - five gliadins and three glutenins - which are secretory proteins synthesized and folded into the ER before being deposited into organelles known as protein bodies [54]. However, these proteins were also detected in Golgi-derived vesicles early in the grain development, indicating that the Golgi apparatus is also involved in prolamin transport [55, 56].

Carbohydrate-active enzymes

Within the numerous proteins identified, it was predicted that 78 sequences would be active on carbohydrates, among them, proteins from 17 GT families (Fig. 4).

Several of the GTs identified are clearly not involved in cell wall synthesis. GT4 and GT5 are enzymes that catalyze the metabolism of sucrose (sucrose synthase UniProt ID: Q8W1W4) and starch (starch synthase UniProt ID: Q43654). Interestingly, they were only identified in the associated fraction. Members of the GT66 family (STT3 homolog GenBank ID: AAT69659) were implicated in protein N-glycosylation and localized in the ER [23].

GTs potentially involved in cell wall polysaccharide biosynthesis

Other GTs identified in our proteomic analysis have already been implicated in cell wall synthesis. This is the case for the GT2 CslF6 protein of wheat (Wheat GI ID:

TC390697). Homologs were identified as MLG glucan synthases following the insertion of rice *CslF* genes into the *Arabidopsis thaliana* genome, which resulted in the detection of grass-specific MLG in the walls of the transgenic lines [57]. The wheat gene *CslF6* was identified and silenced by RNA interference, resulting in a decrease of the MLG content in wheat grain endosperm [58].

Conversely, we identified GTs known to be involved in AX synthesis. The GT47-13 protein UniProt ID: E0ZPV1_WHEAT was found in a complex with glucurono(arabino)xylan synthase activity in a wheat seedling microsome fraction [28]. Although no transmembrane domain is predicted for this protein, we localized it in the Golgi apparatus. The TaGT47-13 protein is the potential ortholog of Arabidopsis IRX10 and IRX10-L [59, 60], which have been implicated in the elongation of the xylan backbone.

RGP proteins are members of the GT75 family, although their role in glycosylation remains to be proven. Indeed, several recombinant rice and *Arabidopsis thaliana* RGPs were found to have a UDP-Ara*p* mutase (UAM) activity, which is required to convert the UDP-arabinopyranose into UDP-arabinofuranose [61, 62]. Down-regulation of rice UAM reduces the proportion of arabinofuranose in rice cell walls [63]. Zeng et al. [28] detected a GT75 protein (UniProt ID: Q9ZR33) in the glucurono(arabino)xylan complex that they isolated from a wheat seedling microsomal fraction. We identified three GT75 in the wheat endosperm (Wheat GI ID: TC386312, Wheat GI ID: TC371253, Wheat GI ID: TC399081), the sequence of Wheat GI ID: TC386312 is highly similar but not identical to UniProt ID: Q9ZR33. Our fusion experiment revealed that Wheat GI ID: TC386312 was localized in the Golgi apparatus, despite a lack of transmembrane domain.

Very recently, several rice and wheat GT61 of the subgroup A were shown to be xylan O3 arabinosyltransferases (XATs) [19]. We identified three different GT61 of the subgroup A in the wheat endosperm (Table 2) and localized one of them, the UniProt ID: Q5QPY3 corresponding to XAT1, in the Golgi apparatus. Transcriptomic analysis showed that it seems specifically expressed in the endosperm. A GT61 from the subgroup B (UniProt ID: Q5QPZ5_WHEAT) was also found in our analysis (Table 2). The corresponding gene exhibits an expression pattern specific to the grain endosperm and floral bracts before anthesis (Supplemental Fig. 2). This GT61 may be an arabinosyltransferase required for the double substitution of xylose residues in the

endosperm AX at position O2 and O3. It is noteworthy that no GT61 was identified in the glucurono(arabino)xylan complex isolated from wheat seedlings by Zeng et al. [28]. A GT48 protein was revealed in our analysis. The GT48 family is known to include the glucan synthase-like proteins (GSL) or callose synthase (CalS) [64, 65]. Callose plays an important role in plant development and in response to stress. Callose is deposited at cell plates during cytokinesis. It accumulates transiently in the wheat endosperm early in the development [4]. Voigt et al. [66] identified eight *GSL* genes in wheat and studied their expression in different organs. The protein we detected in the endosperm at this stage (Wheat GI ID: TC451164) corresponds to the *TaGSL12* gene.

It was very recently revealed that members of the CslD family have mannan synthase activity [67]. Mannans are present in the wheat endosperm. Nemeth et al. [58] produced RNA interference lines targeting the wheat *CslD2* gene potentially encoding the CslD2 protein that we detected by proteomic analysis (Wheat GI ID: TC394160). They only investigated a potential effect on AX and MLG. Thus, it is possible that the wheat CslD2 has a mannan synthase activity. Virtually nothing is known about the structure, i.e., the possible substituents of mannans in wheat grain. Among the proteins we identified in our proteomic analysis, several GT34 share similarity with a galactomannan galactosyltransferase identified in lotus seeds [68]. The transcripts encoding the GT34 Wheat GI ID: TC370787 were detected in the grain but not in the vegetative organs investigated, and the transcript level increases steeply in the developing grain (Supplemental Figure 2). Mannans have a storage function in the seeds of lotus and many other species [69]. Nothing is known about their function in the wheat grain. Another activity was described for members of the GT34 family. Several Arabidopsis genes of this family were found to encode xyloglucan xylosyltransferase [70-72]. Xyloglucan were recently detected transiently in the wheat endosperm [7].

Another interesting GT identified is the GT47 Wheat GI ID: TC390543 (best hit in rice LOC_Os11g03410), a protein containing a single transmembrane domain and that we localized in the Golgi apparatus. The GT47 family is encoded by 39 genes in Arabidopsis [73]. Phylogenetic analyses of the GT47 family revealed that LOC_Os11g03410 (and Wheat GI ID: TC390543) is in a different subgroup of GT47 (subgroup C) than GT47-13 (UniProt ID: E0ZPV1) and Arabidopsis IRX10-10L, which are in subgroup A according to the Zhong and Ye classification [7, 73]. Subgroup C contains several rice and Arabidopsis proteins. Only two of these proteins (ARAD1 and

ARAD2) have an ascribed function based on mutant characterization: they are putative arabinosyltransferases acting on pectin [74]. Wheat GI ID: TC390543 and its rice ortholog LOC_Os11g03410 are not really close to the ARAD proteins, and although pectin is a component of the rice endosperm [75], no pectin has been reported in the wheat grain endosperm [3]. Thus, it is possible that the enzyme encoded by Wheat GI ID: TC390543 has a different activity.

We achieved our aim by identifying GTs potentially involved in wheat endosperm cell wall synthesis through a proteomic approach.

GHs potentially involved in cell wall polysaccharide remodelling

Surprisingly, we also identified many GH proteins that we did not expect to be located within the Golgi apparatus. To our knowledge, Golgi location was described for some GHs, such as members of the families GH38 (Wheat GI ID: TC416977) and GH47 (Wheat GI ID: TC393888), which would be an alpha-mannosidase involved in protein glycosylation [23, 50, 76]. It is thought that most of the other GHs are secreted. Many GHs are glycosylated and probably pass through the Golgi apparatus. Half of these GHs were identified in the associated fraction and only a few of them were predicted to contain a TM domain, consolidating the idea that we identified them during their transit through the endomembrane system.

The GH proteins identified in the endosperm fell into 18 of the 34 families that are represented in the rice, Arabidopsis and Brachypodium genomes [77].

The putative function of some of the individual GHs was searched by studying the annotation of their best homologs in other species. However, the occurrence within the same family of numerous enzymatic activities makes the task of identifying the potential activity of individual GHs difficult. Some GHs identified here are clearly not involved in cell wall polysaccharide modification/turnover. In fact, the GH14 proteins are beta-amylases (UniProt ID: Q7X9M2_WHEAT). Other families include enzymatic activities that could modify cell walls.

The GH17 family contains many characterized activities, including β-1,3-glucanases (EC 3.2.1.39) that specifically degrade callose, as shown in Arabidopsis plasmodesmata [78]. We identified several GH17 in our experiment. BJ242908 is similar to the rice putative β-1,3-glucanase (UniProt ID: Q9ZNZ1) and may be related to the early and transient deposition of callose in the grain cell wall that is degraded during development since callose is not detected in the mature grain endosperm. The

expression of the gene encoding GenBank ID: BJ242908 seems to be specific to the grain endosperm and increases during grain development (Supplemental Fig. 2). Fused with GFP, the callase identified in the plasmodesmata was localized in the membrane and the ER [78].

Among the GH9, some were predicted to have an endo-β-glucanase activity. We identified several GH9 sequences (GenBank ID: CD888519_WHEAT, Wheat GI ID: TC382188, Wheat GI ID: TC442952). Best hit results by similarity search are endo-β-glucanase (cellulase) implicated in leaf and fruit abscission in *Prunus persica* ppEG1, *Citrus sinensis* and *Malus x domestica* [79]. In *Arabidopsis thaliana*, KORRIGAN is a GH9 endo-β-glucanase that was proposed to play a role in cellulose synthesis [80-82]. The cell wall of wheat grain endosperm contains a minor amount of cellulose [2]. The transcript corresponding to Wheat GI ID: TC382188 seems to be specific to the grain endosperm and its expression increases during grain development (Supplemental Fig. 2).

Several GH3 were revealed in our analysis. Several activities were reported for enzymes within the GH3 family. Among the GH3 identified in our screen, Wheat GI ID: TC369308 is homolog to the barley β-D-glucan exohydrolase isoenzyme ExoII (UniProt ID: Q42835_HORVU), and Wheat GI ID: TC422719 and Wheat GI ID: TC384225 are homologs to a xylan β-D-xylosidase characterized in barley (Genbank ID: AAK38482). Transcripts encoding the putative xylosidases are well represented in the endosperm, and the abundance of Wheat GI ID: TC422719 increases considerably during early development (Supplemental Fig. 2). Genbank ID: CK196562 is homolog to the arabinofuranosidase/xylosidase identified in germinating barley (Lee et al. 2003). The substrates of all these enzymes are most likely to be MLGs and AXs, the two main components of the wheat endosperm cell wall. At the onset of their synthesis, the AXs of wheat endosperm are highly substituted with arabinose and later in the endosperm development, the level of AX substitution decreases [4, 5, 83, 84]. Toole et al. [84] proposed that a GH with arabinofuranohydrolase or arabinofuranosidase activity would remove arabinose residues to decrease the level of AX substitution. Recent studies using films of pure AXs as models of endosperm cell walls revealed the influence of AX structure on water diffusion [85, 86]. Higher water diffusion was obtained for films made with highly-substituted AXs, compared to films made with lowly-substituted AXs. This suggests that AX-decreased substitution may have a physiological function during

development, probably in the regulation of water/solute diffusion during grain filling/dessication.

In summary, using this subcellular proteomic approach, we have successfully identified over a thousand proteins from wheat starchy endosperm with many new enzymes probably involved in cell wall biogenesis. They are members of families already assumed to be involved in cell wall synthesis and modification or degradation, but information about their function is still lacking. The expression pattern of the candidates reveals that all of them were expressed in the wheat endosperm, with a few of them exhibiting an expression profile specific to the endosperm. Synthesis of cell wall hemicellulose takes place in the Golgi apparatus. We report here GTs potentially involved in this mechanism, four of which were localized in the Golgi apparatus.

Acknowledgements:

This research was financed by the AIC and a CEPIA post-doc. We are grateful to INRA Rennes for providing the wheat used in this experiment, and to G. Deshayes for assistance with proteomics, P.E. Bodet and A. Geairon for assistance in mass spectrometry and A. Bouder for her technical assistance. We thank our colleagues of the Cell Biology Unit of the Imaging platform (Imagif) of the Institut des Sciences du Végétal Gif sur Yvette, and of the “plateforme de Recherche en Imagerie Cellulaire de Haute Normandie”, Université de Rouen, in particular, Laurence Chevalier for their help and support with sample preparation for ultrastructural and immunocytochemistry TEM studies, respectively.

References

[1] Carpita N, McCann M. The Cell Wall. In: Buchanan B, Gruissem W, Jones R, editors. Biochemistry & Molecular Biology of Plants: American Society of Plant Physiologists; 2000. p. 52-108.

[2] Bacic A, Stone BA. Isolation and ultrastructure of cell walls from wheat and barley. Australian Journal of Plant Physiology. 1981;8:453-74.

[3] Saulnier L, Guillon F, Chateigner-Boutin A-L. Cell wall deposition and metabolism in wheat grain. Journal of Cereal Science. 2012.

[4] Philippe S, Robert P, Barron C, Saulnier L, Guillon F. Deposition of cell wall polysaccharides in wheat endosperm during grain development: Fourier transform-infrared microspectroscopy study. J Agric Food Chem. 2006;54:2303-8.

[5] Philippe S, Saulnier L, Guillon F. Arabinoxylan and (1-->3),(1-->4)-beta-glucan deposition in cell walls during wheat endosperm development. Planta. 2006;224:449-61.

[6] Philippe S, Tranquet O, Utille JP, Saulnier L, Guillon F. Investigation of ferulate deposition in endosperm cell walls of mature and developing wheat grains by using a polyclonal antibody. Planta. 2007;225:1287-99.

[7] Pellny TK, Lovegrove A, Freeman J, Tosi P, Love CG, Knox JP, et al. Cell walls of developing wheat starchy endosperm: comparison of composition and RNA-Seq transcriptome. Plant Physiol. 2012;158:612-27.

[8] Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37:D233-8.

[9] Coutinho PM, Stam M, Blanc E, Henrissat B. Why are there so many carbohydrate-active enzyme-related genes in plants? Trends Plant Sci. 2003;8:563-5.

[10] Henrissat B, Coutinho PM, Davies GJ. A census of carbohydrate-active enzymes in the genome of Arabidopsis thaliana. Plant Mol Biol. 2001;47:55-72.

[11] Cai G, Faleri C, Del Casino C, Emons AM, Cresti M. Distribution of callose synthase, cellulose synthase, and sucrose synthase in tobacco pollen tube is controlled in dissimilar ways by actin filaments and microtubules. Plant Physiol. 2011;155:1169-90.

[12] Dhugga KS, Tiwari SC, Ray PM. A reversibly glycosylated polypeptide (RGP1) possibly involved in plant cell wall synthesis: purification, gene cloning, and trans-Golgi localization. Proc Natl Acad Sci U S A 1997;94:7679-84.

[13] Saxena IM, Brown RM, Jr. Cellulose biosynthesis: current views and evolving concepts. Ann Bot. 2005;96:9-21.

[14] Carpita NC, McCann MC. The maize mixed-linkage (1->3),(1->4)-beta-D-glucan polysaccharide is synthesized at the golgi membrane. Plant Physiol. 2010;153:1362-71.

[15] Harris PJ, Northcote DH. Polysaccharide formation in plant Golgi bodies. Biochim Biophys Acta. 1971;237:56-64.

[16] Northcote DH, Pickett-Heaps JD. A function of the Golgi apparatus in polysaccharide synthesis and transport in the root-cap cells of wheat. Biochem J. 1966;98:159-67.

[17] Ray PM, Shininger TL, Ray MM. Isolation of beta-glucan synthetase particles from plant cells and identification with Golgi membranes. Proc Natl Acad Sci U S A. 1969;64:605-12.

[18] Zhang GF, Staehelin LA. Functional compartmentation of the Golgi apparatus of plant cells : immunocytochemical analysis of high-pressure frozen- and freeze-substituted sycamore maple suspension culture cells. Plant Physiol. 1992;99:1070-83.

[19] Anders N, Wilkinson MD, Lovegrove A, Freeman J, Tryfona T, Pellny TK, et al. Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. Proc Natl Acad Sci U S A. 2012;109:989-93.

[20] Pena MJ, Zhong R, Zhou GK, Richardson EA, O'Neill MA, Darvill AG, et al. Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. Plant Cell. 2007;19:549-63.
[21] Zhong R, Pena MJ, Zhou GK, Nairn CJ, Wood-Jones A, Richardson EA, et al. Arabidopsis fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell. 2005;17:3390-408.
[22] Bencur P, Steinkellner H, Svoboda B, Mucha J, Strasser R, Kolarich D, et al. Arabidopsis thaliana beta1,2-xylosyltransferase: an unusual glycosyltransferase with the potential to act at multiple stages of the plant N-glycosylation pathway. Biochem J. 2005;388:515-25.
[23] Dunkley TP, Hester S, Shadforth IP, Runions J, Weimar T, Hanton SL, et al. Mapping the Arabidopsis organelle proteome. Proc Natl Acad Sci U S A. 2006;103:6518-23.
[24] Olinares PD, Ponnala L, van Wijk KJ. Megadalton complexes in the chloroplast stroma of Arabidopsis thaliana characterized by size exclusion chromatography, mass spectrometry, and hierarchical clustering. Mol Cell Proteomics. 2010;9:1594-615.
[25] Millar AH, Heazlewood JL, Kristensen BK, Braun HP, Moller IM. The plant mitochondrial proteome. Trends Plant Sci. 2005;10:36-43.
[26] van Wijk KJ. Plastid proteomics. Plant Physiol Biochem. 2004;42:963-77.
[27] Agrawal GK, Bourguignon J, Rolland N, Ephritikhine G, Ferro M, Jaquinod M, et al. Plant organelle proteomics: Collaborating for optimal cell function. Mass Spectrometry Reviews. 2011;30:772– 853.
[28] Zeng W, Jiang N, Nadella R, Killen TL, Nadella V, Faik A. A glucurono(arabino)xylan synthase complex from wheat contains members of the GT43, GT47, and GT75 families and functions cooperatively. Plant Physiol. 2010;154:78-97.
[29] Chevalier L, Bernard S, Ramdani Y, Lamour R, Bardor M, Lerouge P, et al. Subcompartment localization of the side chain xyloglucan-synthesizing enzymes within Golgi stacks of tobacco suspension-cultured cells. Plant J. 2010;64:977-89.
[30] Thiery JP. Mise en évidence de polysaccharides sur coupes fines en microscopie électronique. Journal de Microscopie. 1967;6:987-1018.
[31] Guillon F, Tranquet O, Quillien L, Utille J-P, Ordaz Ortiz JJ, Saulnier L. Generation of polyclonal and monoclonal antibodies against arabinoxylans and their use for immunocytochemical location of arabinoxylans in cell walls of endosperm of wheat. Journal of Cereal Science. 2004;40:167-82.
[32] Devouge V, Rogniaux H, Nesi N, Tessier D, Gueguen J, Larre C. Differential proteomic analysis of four near-isogenic Brassica napus varieties bred for their erucic acid and glucosinolate contents. J Proteome Res. 2007;6:1342-53.
[33] Larre C, Penninck S, Bouchet B, Lollier V, Tranquet O, Denery-Papini S, et al. Brachypodium distachyon grain: identification and subcellular localization of storage proteins. J Exp Bot. 2010;61:1771-83.
[34] Sado PE, Tessier D, Vasseur M, Elmorjani K, Guillon F, Saulnier L. Integrating genes and phenotype: a wheat-Arabidopsis-rice glycosyltransferase database for candidate gene analyses. Funct Integr Genomics. 2009;9:43-58.
[35] Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC. CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. 2010;20:1574-84.
[36] Dash S, Van Hemert J, Hong L, Wise RP, Dickerson JA. PLEXdb: gene expression resources for plants and plant pathogens. Nucleic Acids Res. 2012;40:D1194-201.
[37] Pont C, Murat F, Confolent C, Balzergue S, Salse J. RNA-seq in grain unveils fate of neo- and paleopolyploidization events in bread wheat (Triticum aestivum L.). Genome Biol. 2011;12:R119.

[38] Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, et al. PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell. 2011;23:895-910.
[39] Sparkes IA, Runions J, Kearns A, Hawes C. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc. 2006;1:2019-25.
[40] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-402.
[41] Kall L, Krogh A, Sonnhammer EL. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res. 2007;35:W429-32.
[42] Merlino M, Leroy P, Chambon C, Branlard G. Mapping and proteomic analysis of albumin and globulin proteins in hexaploid wheat kernels (Triticum aestivum L.). Theor Appl Genet. 2009;118:1321-37.
[43] Skylas DJ, Mackintosh JA, Cordwell SJ, Basseal DJ, Walsh BJ, Harry J, et al. Proteome approach to the characterisation of protein composition in the developing and mature wheat-grain endosperm. Journal of Cereal Science. 2000;32:169-88.
[44] Skylas DJ, Van Dyk D, Wrigley CW. Proteomics of wheat grain. Journal of Cereal Science. 2005;41:165–79.
[45] Graham JM. Isolation of Golgi membranes from tissues and cells by differential and density gradient centrifugation. Curr Protoc Cell Biol. 2001;Chapter 3:Unit 3 9.
[46] Sadowski PG, Dunkley TP, Shadforth IP, Dupree P, Bessant C, Griffin JL, et al. Quantitative proteomic approach to study subcellular localization of membrane proteins. Nat Protoc. 2006;1:1778-89.
[47] Mitra SK, Gantt JA, Ruby JF, Clouse SD, Goshe MB. Membrane proteomic analysis of Arabidopsis thaliana using alternative solubilization techniques. J Proteome Res. 2007;6:1933-50.
[48] Drakakaki G, Zabotina O, Delgado I, Robert S, Keegstra K, Raikhel N. Arabidopsis reversibly glycosylated polypeptides 1 and 2 are essential for pollen development. Plant Physiol. 2006;142:1480-92.
[49] Langeveld SM, Vennik M, Kottenhagen M, Van Wijk R, Buijk A, Kijne JW, et al. Glucosylation activity and complex formation of two classes of reversibly glycosylated polypeptides. Plant Physiol. 2002;129:278-89.
[50] Parsons HT, Christiansen K, Knierim B, Carroll A, Ito J, Batth TS, et al. Isolation and proteomic characterization of the Arabidopsis Golgi defines functional and novel components involved in plant cell wall biosynthesis. Plant Physiol. 2012;159:12-26.
[51] Atmodjo MA, Sakuragi Y, Zhu X, Burrell AJ, Mohanty SS, Atwood JA, 3rd, et al. Galacturonosyltransferase (GAUT)1 and GAUT7 are the core of a plant cell wall pectin biosynthetic homogalacturonan:galacturonosyltransferase complex. Proc Natl Acad Sci U S A. 2011;108:20225-30.
[52] Apweiler R, Hermjakob H, Sharon N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta. 1999;1473:4-8.
[53] Wightman R, Turner S. Trafficking of the plant cellulose synthase complex. Plant Physiol. 2010;153:427-32.
[54] Shewry PR, Napier JA, Tatham AS. Seed storage proteins: structures and biosynthesis. Plant Cell. 1995;7:945-56.
[55] Loussert C, Popineau Y, Mangavel C. Protein bodies ontogeny and localization of prolamin components in the developing endosperm of wheat caryopses. Journal of Cereal Science. 2008;47:445-56.

[56] Tosi P, Parker M, Gritsch CS, Carzaniga R, Martin B, Shewry PR. Trafficking of storage proteins in developing grain of wheat. J Exp Bot. 2009;60:979-91.
[57] Burton RA, Wilson SM, Hrmova M, Harvey AJ, Shirley NJ, Medhurst A, et al. Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. Science. 2006;311:1940-2.
[58] Nemeth C, Freeman J, Jones HD, Sparks C, Pellny TK, Wilkinson MD, et al. Down-regulation of the CSLF6 gene results in decreased (1,3;1,4)-beta-D-glucan in endosperm of wheat. Plant Physiol. 2010;152:1209-18.
[59] Brown DM, Zhang Z, Stephens E, Dupree P, Turner SR. Characterization of IRX10 and IRX10-like reveals an essential role in glucuronoxylan biosynthesis in Arabidopsis. Plant J. 2009;57:732-46.
[60] Wu AM, Rihouey C, Seveno M, Hornblad E, Singh SK, Matsunaga T, et al. The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation. Plant J. 2009;57:718-31.
[61] Konishi T, Takeda T, Miyazaki Y, Ohnishi-Kameyama M, Hayashi T, O'Neill MA, et al. A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. Glycobiology. 2007;17:345-54.
[62] Rautengarten C, Ebert B, Herter T, Petzold CJ, Ishii T, Mukhopadhyay A, et al. The interconversion of UDP-arabinopyranose and UDP-arabinofuranose is indispensable for plant development in Arabidopsis. Plant Cell. 2011;23:1373-90.
[63] Konishi T, Aohara T, Igasaki T, Hayashi N, Miyazaki Y, Takahashi A, et al. Down-regulation of UDP-arabinopyranose mutase reduces the proportion of arabinofuranose present in rice cell walls. Phytochemistry. 2011;72:1962-8.
[64] Richmond TA, Somerville CR. The cellulose synthase superfamily. Plant Physiol. 2000;124:495-8.
[65] Verma DP, Hong Z. Plant callose synthase complexes. Plant Mol Biol. 2001;47:693-701.
[66] Voigt CA, Schafer W, Salomon S. A comprehensive view on organ-specific callose synthesis in wheat (Triticum aestivum L.): glucan synthase-like gene expression, callose synthase activity, callose quantification and deposition. Plant Physiol Biochem. 2006;44:242-7.
[67] Verhertbruggen Y, Yin L, Oikawa A, Scheller HV. Mannan synthase activity in the CSLD family. Plant Signal Behav. 2011;6:1620-3.
[68] Edwards ME, Choo TS, Dickson CA, Scott C, Gidley MJ, Reid JS. The seeds of Lotus japonicus lines transformed with sense, antisense, and sense/antisense galactomannan galactosyltransferase constructs have structurally altered galactomannans in their endosperm cell walls. Plant Physiol. 2004;134:1153-62.
[69] Buckeridge MS. Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. Plant Physiol. 2010;154:1017-23.
[70] Cavalier DM, Keegstra K. Two xyloglucan xylosyltransferases catalyze the addition of multiple xylosyl residues to cellohexaose. J Biol Chem. 2006;281:34197-207.
[71] Faik A, Price NJ, Raikhel NV, Keegstra K. An Arabidopsis gene encoding an alpha-xylosyltransferase involved in xyloglucan biosynthesis. Proc Natl Acad Sci U S A. 2002;99:7797-802.
[72] Zabotina OA, van de Ven WT, Freshour G, Drakakaki G, Cavalier D, Mouille G, et al. Arabidopsis XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. Plant J. 2008;56:101-15.
[73] Zhong R, Ye ZH. Unraveling the functions of glycosyltransferase family 47 in plants. Trends Plant Sci. 2003;8:565-8.

[74] Harholt J, Jensen JK, Sorensen SO, Orfila C, Pauly M, Scheller HV. ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in Arabidopsis. Plant Physiol. 2006;140:49-58.
[75] Shibuya N, Nakane R. Pectic polysaccharides of rice endosperm cell walls. Phytochemistry. 1984;23:1425-9.
[76] Strasser R, Schoberer J, Jin C, Glossl J, Mach L, Steinkellner H. Molecular cloning and characterization of Arabidopsis thaliana Golgi alpha-mannosidase II, a key enzyme in the formation of complex N-glycans in plants. Plant J. 2006;45:789-803.
[77] Tyler L, Bragg JN, Wu J, Yang X, Tuskan GA, Vogel JP. Annotation and comparative analysis of the glycoside hydrolase genes in Brachypodium distachyon. BMC Genomics. 2010;11:600.
[78] Levy A, Erlanger M, Rosenthal M, Epel BL. A plasmodesmata-associated beta-1,3-glucanase in Arabidopsis. Plant J. 2007;49:669-82.
[79] Trainotti L, Spolaore S, Ferrarese L, Casadoro G. Characterization of ppEG1, a member of a multigene family which encodes endo-beta-1,4-glucanase in peach. Plant Mol Biol. 1997;34:791-802.
[80] Lane DR, Wiedemeier A, Peng L, Hofte H, Vernhettes S, Desprez T, et al. Temperature-sensitive alleles of RSW2 link the KORRIGAN endo-1,4-beta-glucanase to cellulose synthesis and cytokinesis in Arabidopsis. Plant Physiol. 2001;126:278-88.
[81] Nicol F, His I, Jauneau A, Vernhettes S, Canut H, Hofte H. A plasma membrane-bound putative endo-1,4-beta-D-glucanase is required for normal wall assembly and cell elongation in Arabidopsis. EMBO J. 1998;17:5563-76.
[82] Szyjanowicz PM, McKinnon I, Taylor NG, Gardiner J, Jarvis MC, Turner SR. The irregular xylem 2 mutant is an allele of korrigan that affects the secondary cell wall of Arabidopsis thaliana. Plant J. 2004;37:730-40.
[83] Saulnier L, Robert P, Grintchenko M, Jamme F, Bouchet B, Guillon F. Wheat endosperm cell walls: Spatial heterogeneity of polysaccharide structure and composition using micro-scale enzymatic fingerprinting and FT-IR microspectroscopy. Journal of Cereal Science. 2009;50:312-7.
[84] Toole GA, Le Gall G, Colquhoun I, Nemeth C, Saulnier L, Lovegrove A, et al. Temporal and spatial changes in cell wall composition in developing grains of wheat cv. Hereward. Planta. 2010;232:677-89.
[85] Ying R, Barron C, Saulnier L, Rondeau-Mouro C. Water mobility within arabinoxylan and beta-glucan films studied by NMR and dynamic vapour sorption. J Sci Food Agric. 2011;91:2601-5.
[86] Ying R, Saulnier L, Rondeau-Mouro C. Films of arabinoxylans and beta-glucans extracted from cereal grains: Molecular motions by TD-NMR. Carbohydrate Polymers. 2011;86:812-22.

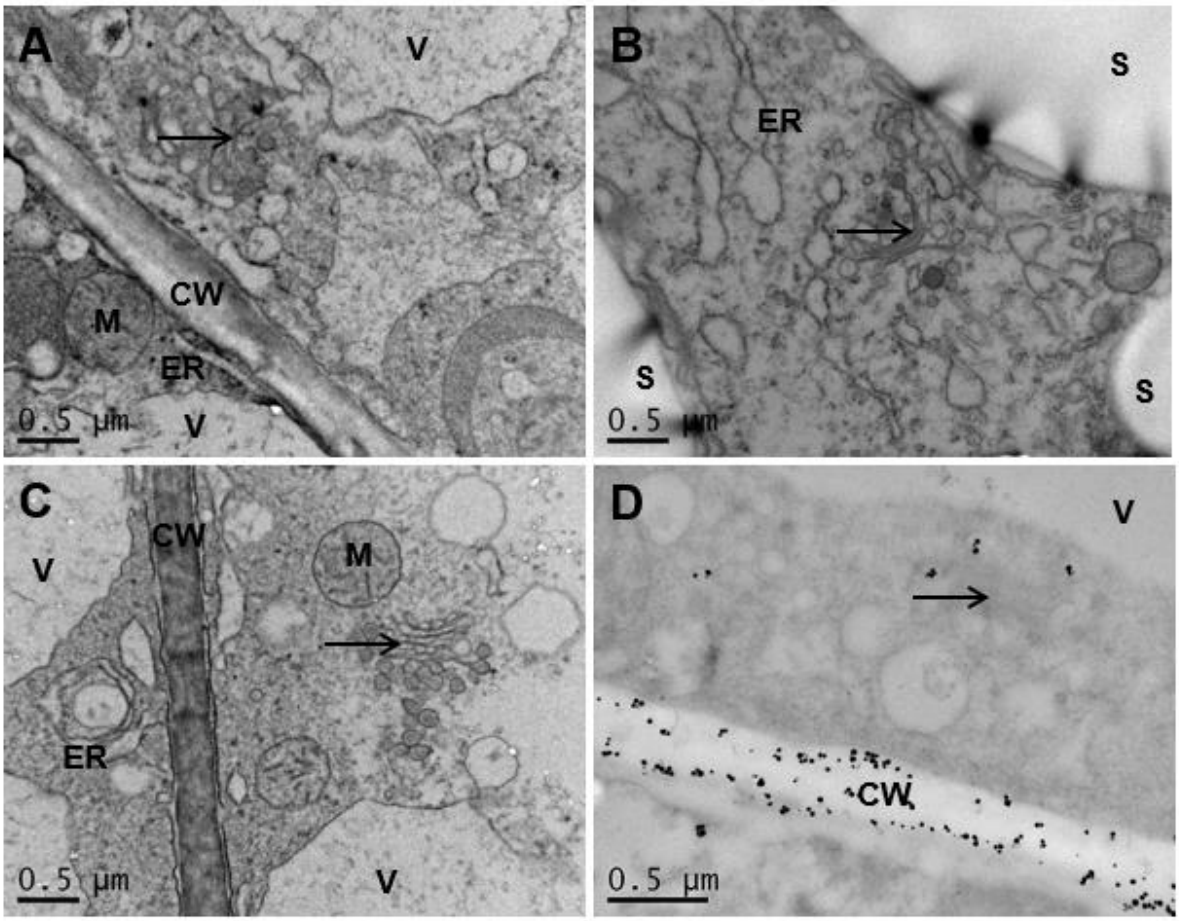

Figure 1: Transmission electron micrographs of developing wheat grain at 250°D. A view of the cytoplasm showing different organelles in aleurone cells (A) and in the starchy endosperm (B). The aleurone cells exhibit dense granular cytoplasm with numerous vacuoles, ER and mitochondria. A few Golgi stacks are observed. In the starchy endosperm, ER tubules, some of which show dilation, are abundant, whereas Golgi structures with stacked cisternae are barely visible. PATAg staining of aleurone cell walls (C). PATAg-stained materials are detected within the cell walls and in Golgi-derived vesicles. Immunogold labeling of arabinoxylan in aleurone cells (D). Arabinoxylan epitopes are visualized in the cell wall. Labeling is mainly localized in the inner region of the primary cell walls and found over Golgi stacks.

CW, cell wall; ER, endoplasmic reticulum; M, mitochondria; S, starch granule, V, vacuole; arrow, Golgi stacks.

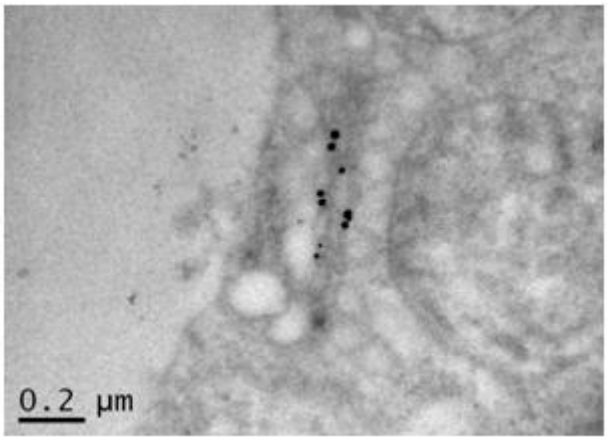

Figure 2: Immunogold localization of RGP1 in aleurone cells. Gold particles occur over Golgi stacks but none over other cell structures.

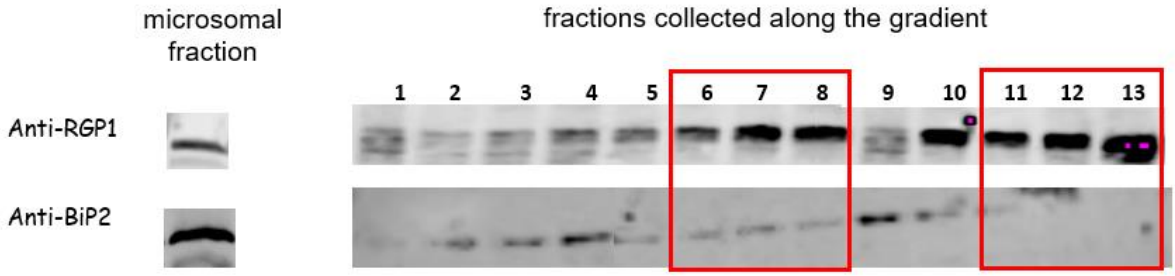

Figure 3: Presence of Golgi apparatus and ER in microsomal fraction and fractions collected along the gradient revealed by western blot with antibodies against Golgi protein (Anti-RGP1 antibody) and ER protein (Anti-BiP2 antibody) as indicated. The anti-RGP1 results presented here were obtained on the associated protein fractions and those of Anti-BiP2 on membrane protein fractions.

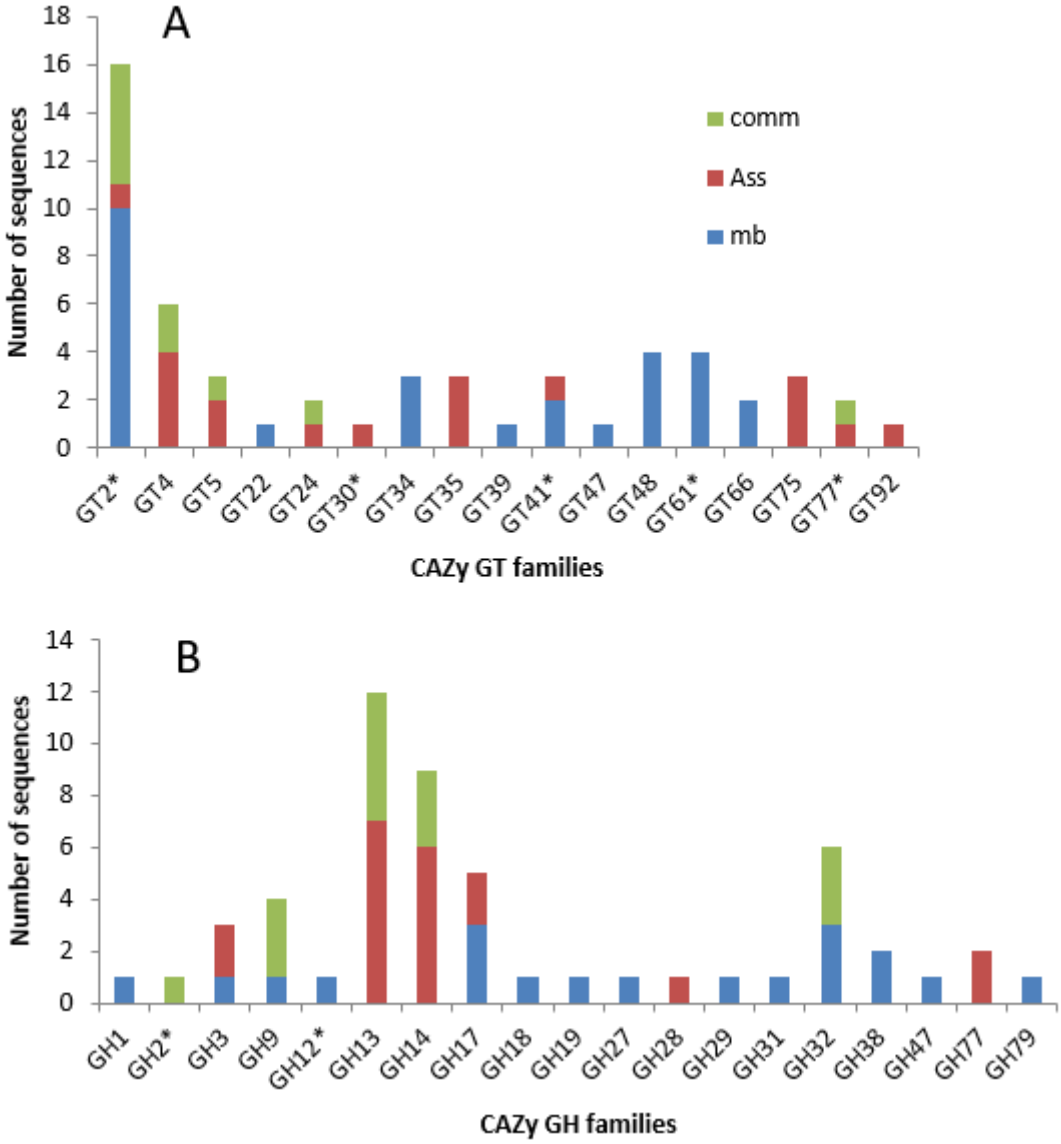

Figure 4: Distribution of blasted sequences in (A) glycosyltransferase (GT) families and in (B) glycosylhydrolase (GH) families; number of hits obtained with the Pfam domain except for GTs and GHs labeled with a star for which the hits were obtained only by sequence similarity. Colors correspond to the fractions in which the sequences were identified: blue: present in membrane fractions; red: present in associated fractions; green: present in both fractions.

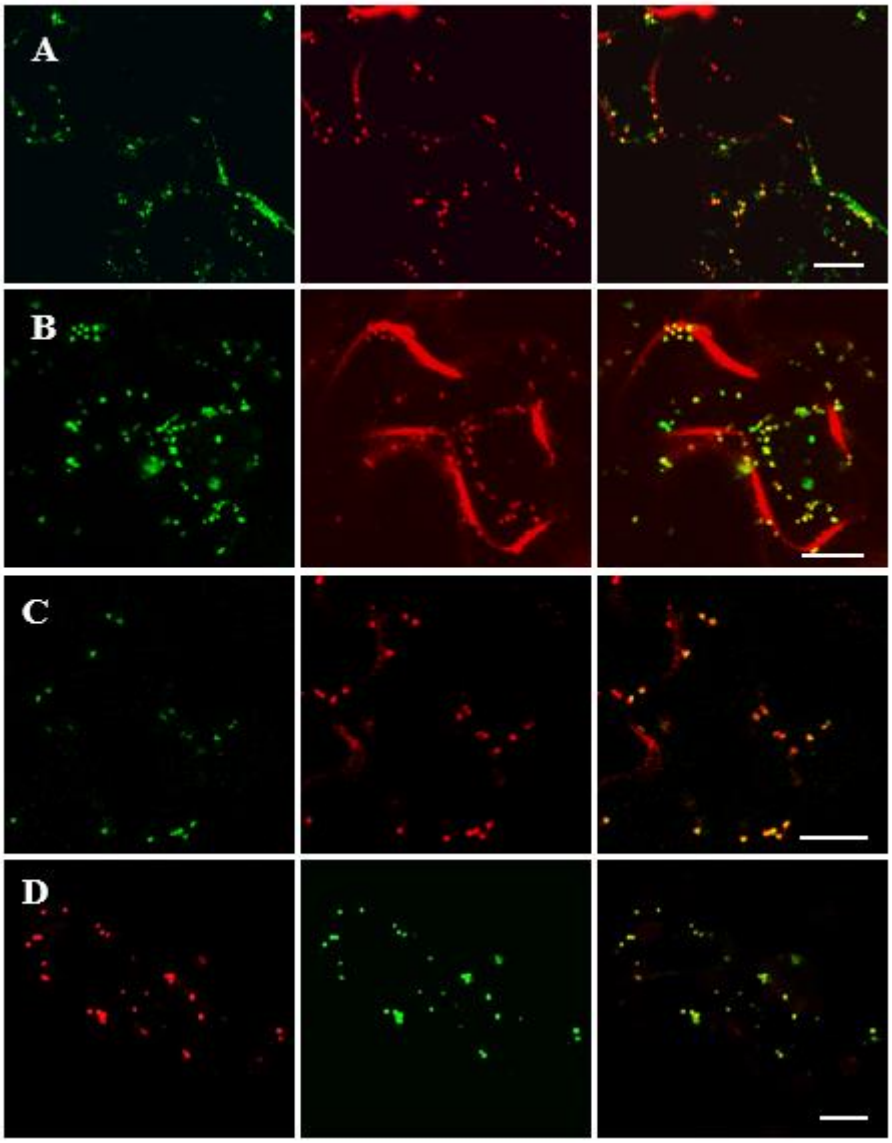

Figure 5: The subcellular localization of some glycosyltransferases fused to fluorescent proteins in tobacco epidermal cells. Confocal images showing eGFP (green) and mRFP (red) fusion proteins coexpressed in tobacco leaf epidermal cells 2-3 days after agrobacterium infiltration. Images of cells expressing GT47-TC390543-eGFP (green) co-expressed with a Golgi marker, ST-mRFP (red) and merged image (A), GT75-TC386312-eGFP (green) co-expressed with ST-mRFP (red) and merged image (B), GT61-Q5QPY3-eGFP (green) co-expressed with ST-mRFP (red) and merged image (C), GT47-E0ZPV1-mRFP (red) co-expressed with ST-eGFP (green) and merged image (D),Scale bars = 10 µm.

			Phobius			TMHMM	
		Total number of sequences	TM	SP	TM (%)	TM	TM (%)
Membrane (MPG)	sequences	262	106	39	40.6	112	42.9
	best Homologues	262	137	87	52.5	152	58.2
Associated (APG)	sequences	196	7	20	3.6	12	6.2
	best Homologues	196	5	34	2.6	12	6.2
both fractions	sequences	210	30	36	14.3	35	16.7
	best Homologues	210	35	63	16.7	35	16.7

Table 1 : Prediction of signal peptides (SP) and transmembrane domains (TM) using Phobius and TMHMM.

1

	proteomic hit	group	Families	best blast hit rice	E-value	annotation/putative fonction (rice)	brachypodium hit	Activity in family and/or subgroup
Wheat GI ID:	TC390697	359	GT2	LOC_Os08g06380	0	CSLF8 - cellulose synthase-like family F; beta1,3;1,4 glucan synthase	Bradi1g25130	MLG synthase
Wheat GI ID:	TC394160	242	GT2	LOC_Os06g02180	0	CSLD2 - cellulose synthase-like family D	Bradi1g50170	mannan synthase
Wheat GI ID:	TC370787	535	GT34	LOC_Os02g49140	1.0e-77	glycosyltransferase, putative	Bradi3g56510	galactomannan galactosylTase
Wheat GI ID:	TC413183	958	GT34	LOC_Os11g34390	1.4e-75	glycosyltransferase, putative	Bradi4g16670	galactomannan galactosylTase, xyloG xylosylTase
Wheat GI ID:	TC390543	444	GT47	LOC_Os11g03410	2.0e-209	exostosin family domain containing protein	Bradi4g26150	subgoup C arabinosyltransferase acting on pectins
UniProt ID:	E0ZPV1	1091	GT47	LOC_Os01g70190	4.3e-219	exostosin family domain containing protein	Bradi2g59400	subgroup A xylan xylosylTase (xylan synthase?)
Wheat GI ID:	TC451164	137	GT48	LOC_Os06g02260	6.2e-125	callose synthase, putative	Bradi1g51760	callose synthase
UniProt ID:	Q5QPY3	299	GT61	LOC_Os02g22650.2	2.5e-184	glycosyltransferase, putative	Bradi3g11300	XAT1 clade A xylan O3 arabinosyltransferase
Wheat GI ID:	TC437977	597	GT61	LOC_Os02g22650	1.5e-142	glycosyltransferase, putative	Bradi3g11300	clade A xylan O3 arabinosyltransferase
UniProt ID:	D8L9S8	663	GT61	LOC_Os01g02930	3.5e-130	glycosyltransferase, putative	Bradi2g01420	clade A xylan O3 arabinosyltransferase
UniProt ID:	Q5QPZ5	511	GT61	LOC_Os01g72610	6.4e-170	glycosyltransferase, putative	Bradi2g61230	clade B unknow n function
Wheat GI ID:	TC386312	203.1	GT75	LOC_Os03g40270	1.1e-196	alpha-1,4-glucan-protein synthase, putative	Bradi1g15050	RGP/UDP-Arapmutase
Wheat GI ID:	TC371253	203.2	GT75	LOC_Os03g40270	5.1e-133	alpha-1,4-glucan-protein synthase, putative	Bradi1g15050	RGP/UDP-Arapmutase
Wheat GI ID:	TC399081	203.3	GT75	LOC_Os07g41360	1.3e-176	alpha-1,4-glucan-protein synthase, putative	Bradi1g21990	RGP/UDP-Arapmutase
Wheat GI ID:	TC384225	289	GH3	LOC_Os02g51620	2.3e-235	glycosyl hydrolase family 3 protein, putative	Bradi3g59020	xylan B-D-xylosidase
Wheat GI ID:	TC422719	594	GH3	LOC_Os11g18730	1.2e-179	glycosyl hydrolase family 3 protein, putative	Bradi4g20200	xylan B-D-xylosidase
Wheat GI ID:	TC369308	197	GH3	LOC_Os03g53800	1.2e-240	beta glucosidase	Bradi1g08560	MLG endohydrolase
GenBank ID:	CK196562	1195	GH3	LOC_Os04g54810	0	beta-D-xylosidase, putative	Bradi5g23470	arabinofuranosidase
GenBank ID:	CD888519	655	GH9	LOC_Os05g03840	1.8e-28	endoglucanase, putative	Bradi2g38360	endo-1,4 beta-glucanase activity (cellulase)
Wheat GI ID:	TC382188	84	GH9	LOC_Os04g36610	1.4e-83	endoglucanase, putative	Bradi5g11310	endo-1,4 beta-glucanase activity
Wheat GI ID:	TC442952	317	GH9	LOC_Os04g36610	2.1e-45	endoglucanase, putative	Bradi5g11310	endo-1,4 beta-glucanase activity
GenBank ID:	BJ242908	641	GH17	LOC_Os01g71810	5.9e-59	glycosyl hydrolases family 17, putative	Bradi2g60490	beta 1,3 glucanase callase
Wheat GI ID:	TC395632	237	GH17	LOC_Os01g71380	7.7e-86	glycosyl hydrolases family 17, putative	Bradi2g60490	beta 1,3 glucanase callase
Wheat GI ID:	TC394973	466	GH17	LOC_Os01g71680	4.9e-37	glycosyl hydrolases family 17, putative	Bradi2g60490	beta 1,3 glucanase callase

Table 2 : List of candidate GTs and GHs for a role in cell wall metabolism in the developing wheat endosperm

2