

Assessment of Proteomics Strategies for Plant Cell Wall Glycosyltransferases in Wheat, a Non-Model Species: Glucurono(Arabino)Xylan as a Case Study

Faik Ahmed
Ohio University
USA

1. Introduction

The proteome of a tissue is a snapshot of the entire set of proteins expressed by that tissue at a given developmental stage. This set of proteins can be considered as an integrated and complex response of the tissue to a set of environmental and growth conditions. Thus, the analysis of a proteome would provide valuable clues of the physiological status of a tissue (or a cell) at a given time. Proteomics, an expanding scientific field, represents a promising tool to investigate such proteomes in a high-throughput manner. Although proteomics proved to be helpful in elucidating difficult cellular processes, its use in plant cell wall polysaccharides biosynthesis in non-model plants remains challenging. In this chapter, we evaluated the capabilities of two proteomics strategies in identifying specifically three Golgi-resident glycosyltransferases (GTs), TaGT43-4, TaGT47-13, and TaGT75-4, involved in glucurono(arabino)xylan (GAX) biosynthesis in wheat, an economically important non-model crop plant (Zeng et al., 2008, 2010). GAX polymer is the second most abundant polymer in the biomass from grass plants and there is an urgent need to elucidate its biosynthetic pathways to allow engineering of plant biomass for biofuel and other human needs (Faik, 2010).

1.1 Proteomics in plant cell wall polysaccharides biosynthesis

In the model plant *Arabidopsis* (*Arabidopsis thaliana*) about 10% of the genome (~2,500 genes) is dedicated to cell wall metabolism and function (Carpita 2011). Among these genes 17% are putative GTs identified through a homology search, and only a dozen have characterized biochemical functions. The sequences of these enzymes are available through the Carbohydrate Active enZyme database (CAZy) that classified them into GT families on the basis of protein sequence similarity (Coutinho 2003). It is anticipated that more GTs (yet to be identified) are needed to synthesize all polysaccharides currently found in plant cell walls. Most of the GTs are predicated to be integral membrane proteins, however recent works indicate that some GTs have a cleavable signal peptide and are closely associated with other integral membranes as a complex to secure their proper subcellular localization (Zeng et al., 2010). GAX polymer, like all hemicelluloses with the exception of callose, is

synthesized in the Golgi by a multi-protein complex called xylan synthase complex (XSC). Golgi apparatus is a multifunctional organelle used by a plant cell not only to synthesize plant cell wall polymers, but also to process and modify glycoproteins and sorting their destination. Thus, the proteins that populate the membranes of the Golgi are very complex in nature and content (Simon et al., 2008). The challenges in studying the Golgi-resident GTs involved in plant cell wall biosynthesis comes from the fact that they are difficult to purify in an active state and are usually considered low abundant proteins. This may explain why there are very limited reports on purification of these GTs (Perrin et al., 1999; Faik et al., 2002; Zeng et al., 2010). Instead, researchers put more efforts in developing strategies to isolate membrane preparations relatively enriched in endo-membranes or organelles (plasma membranes, endoplasmic reticulum, Golgi) along with enriched GT activities. However, these GT activities are still contaminated with other unrelated proteins. Hence, optimization of high-throughput MS methods for protein identification on these partially purified GT activities is currently lacking.

1.2 Proteomics in non-model plants

Economically important non-model crop plants such as wheat are currently lacking publicly available genomic/protein sequence information. Applying proteomics methods to these species is challenging and requires cross-species identification. Fortunately, genomes and their protein-encoding gene sequences are currently available for several plant species including *Arabidopsis thaliana*, *Oryza sativa*, *Populus trichocarpa*, *Sorghum bicolor*, *Brachypodium distachyon*, *Vitis vinifera*, *Zea mays*, *Medicago*, and *Glycine max* (Figure 1). These nucleotide sequences as well as the predicted amino acid sequences are available through the Phytozome website (<http://www.phytozome.net/>) or NCBI (<http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html>), both of which are considerably larger than the total protein entries available through the Uniprot database which only contains a total of 53,1192 protein entries from different plant species.

The key factor to consider in the cross-species identification is gene annotation in databanks. For example, Uniprot database has only ~29,500 protein entries that are reviewed (annotated). In the case of *Arabidopsis* for which the genome sequence was completed and published more than a decade ago (*Arabidopsis* Genome Initiative, 2000), ~10% of proteins are still not annotated and ~30% are listed with unknown biochemical function. Nevertheless, with all the wealth in genomics resources, the identification of a protein from a non-model plant is now possible through BLAST search using MS/MS spectra and *de novo* sequencing. Even if a homologous protein is not listed in protein databases, a similarity at nucleotide sequence level (~35%) would be sufficient for a reliable identification of a homolog to the peptide sequence by cross-comparison to translated available plant genes (tBLASTn).

2. Advantages and disadvantages of proteomics strategies

Proteomics methods for high-throughput identification of proteins involve three main steps: the first step is the proteolytic (mostly trypsin) digestion of the protein samples. This digestion step would need optimization as incomplete digestion may result in a loss of protein information. The second step consists of three processes: (i) fractionation of the resulting peptides by liquid chromatography (LC), (ii) ionization of the separated peptides (fractions) through a source such as Matrix-Assisted LASER Desorption/Ionization

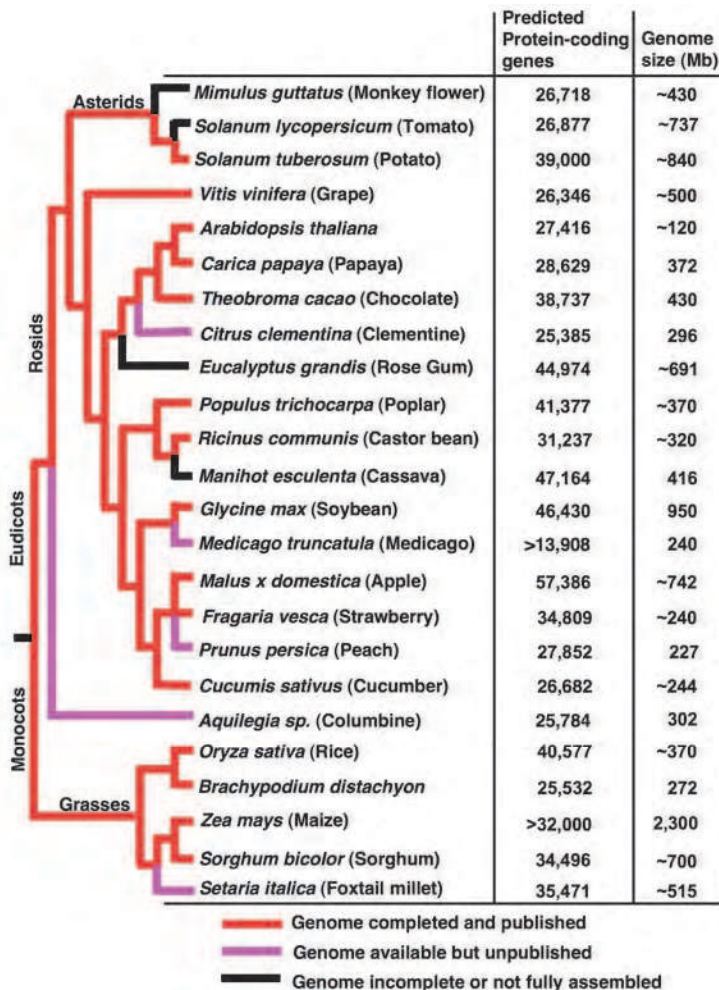


Fig. 1. Predicted number of protein-coding genes and sizes of plant genomes currently available in public databases. Red lines indicate genome sequencing completed and published; purple lines indicate unpublished genomes but publicly available; and black lines indicate plants with incomplete genomes or not fully assembled (less available). Species names are indicated along with the size of their genomes. These sequences are available through Phytosome and/or NCBI websites. Adapted from (http://synteny.cnr.berkeley.edu/wiki/index.php/Sequenced_plant_genomes)

(MALDI) or ElectroSpray Ionization (ESI), (iii) mass spectrometry (MS) analysis of the ionized peptides (received from the source) by tandem mass spectrometry (MS/MS), also called collision-induced dissociation (CID). This step generally delivers large information-rich files of MS/MS spectra (Bodnar et al., 2003). The third and critical step concerns the use of these MS/MS spectra in the identification of the exact proteins or at least their closest homologous proteins from the databases. Depending of the complexity of the samples,

separation of the proteins prior to digestion can increase the chances of identifying these target proteins even in complex samples. Gel-based methods such as SDS-PAGE and 2-dimensional (2-D)-PAGE are widely used to separate the proteins of a sample and thus reduce its complexity. Although tremendous progress in LC and MS techniques has been made in the recent years, the application of these methods is still challenging in many complex biological processes such as plant cell wall biosynthesis (see section 1.1). There are two general proteomic strategies: gel-free and gel-based strategies.

2.1 Gel-free proteomics strategy

Direct analysis of protein composition of any biological sample (without prior separation by electrophoresis) can be achieved by a non-gel shotgun approach called MudPIT (multidimensional protein identification technology) (Washburn et al., 2001; McDonald et al., 2002). The method consists of a 2-D liquid chromatography (LC/LC) separation of the peptides (as opposed to the 2-D-PAGE of the proteins) before MS/MS analysis. The first dimension separates the peptides on the basis of their charges using a strong cation exchange (SCX) column. In the second dimension the peptides eluted from SCX column are fractionated further according to their hydrophobicity by reverse phase chromatography. Depending of the mass spectrometer used, the MudPIT strategy may have some limitations in identifying low abundant proteins. For example, there is a loading limit of protein sample onto the SCX column and it saturates easily, which can significantly impact the identification of low abundant proteins (i.e., GTs) in complex samples. Also, complex samples would require optimization of running time (retention time) for optimal protein identification (Chen et al., 2005). Lastly, although MudPIT has been successfully used to identify more than a 1000 proteins from a complex biological sample (Washburn et al., 2001), the dynamic range between the most abundant and least abundant proteins/peptides (10,000 to 1) can still limit the identification of very low abundant proteins (Wolters et al., 2001). The other key factor to consider when using gel-free proteomics methods is the precipitation step, which is used to not only concentrate the proteins, but also to remove small soluble compounds (including detergents) that are easily ionized and would mask low abundant peptide ions (Newton et al., 2004). Thus, optimizing this step is critical and may also result in a loss of some proteins from the sample. Ethanol, chloroform/methanol, cold acetone, trichloroacetic acid (TCA), or TCA/acetone are all MS-compatible reagents that can be used in proteins precipitation (Ferro et al., 2003). However, because an individual protein has specific physicochemical characteristics, any precipitation method has its own advantages and disadvantages. For example, although ethanol and TCA will precipitate around 90% of proteins, several proteins can be eliminated, reduced, or enriched in the precipitated samples (Zellner et al., 2005; Chen et al., 2010). Furthermore, ethanol is also known to precipitate salts along with the proteins, which may require dialysis of the samples before any further analysis. In our previous work, cold acetone was found to precipitate sucrose along with the proteins when sucrose amounts is >25% w/w in a fraction (Zeng et al., 2010). The increase in sucrose content in the sample decreased peptide ionization and produced low quality MS/MS spectra.

2.2 Gel-based proteomics strategy

The most direct and easy way to separate proteins in a complex sample is using a 1-D SDS-PAGE method (Laemmli, 1970). The main advantage of 1-D SDS-PAGE over the 2-D-PAGE

method is that virtually all proteins can be solubilized and separated on the gel (up to 10-15 μ g per sample). Loading larger amounts of protein on 1-D SDS-PAGE can be detrimental, as abundant proteins can spread over a large area of the gel and may mask low abundant proteins (Zhu et al., 2010). This can be problematic when working with proteins from photosynthesizing organs rich in ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) protein, an abundant and difficult protein to eliminate (Komatsu et al., 1999). Regarding 2-D-PAGE, it is known that protein solubility is the Achilles' heel of this technique, as it does result in the loss of low abundant proteins, proteins with extreme pI, and hydrophobic proteins (integral membrane proteins) (O'Farrell, 1975; Santoni et al., 2000; Ephritikhine et al., 2004). In relation to this approach, Golgi proteins are also known to easily aggregate during sample preparation/precipitation and do not run well on 2-D-PAGE gels (Asakura et al., 2006). Once 1-D SDS-PAGE separation is completed, the gels are usually cut into small equal slices or individual visible bands are excised. These slices/bands are subjected to trypsin digestion, and released peptides further fractionated by LC, and their structures analyzed by MS/MS (see next section).

2.3 Protein identification methods (MS/MS spectra processing)

Despite the availability of a large repository of protein-coding gene sequences from many plant species that can be used as databases for proteomic studies (Figure 1), the most daunting task in proteomics is still the identification of proteins with high sensitivity and accuracy from these databases using MS/MS spectra (Nesvizhskii et al., 2007; Patterson 2003; Service 2008). The processing of these MS/MS spectra for protein identification can be performed via two main strategies:

- i. **Database-dependent strategy** in which the experimental masses of peptide ions (parent and its fragmentation product ions) from MS/MS spectra are compared to the theoretical peptide masses derived from *in silico* digestion of proteins in a database. This search would result in the identification of peptide hits. These hits are scored and ranked from best to worse matches, and any proteins assembling two or several hits are considered candidates. In the case of non-model plants, protein identification can be challenging since either the exact protein may not exist in the database or the database may not contain an evolutionary similar protein. This underscores the need for continuous effort to sequence the genomes of as many plants as possible to allow success of proteomics projects in non-model plants. Many bioinformatics algorithms such as SEQUEST, Mascot, X!Tandem, OMSSA, and Phenyx have been developed for database-dependent search using MS/MS spectra (for details about these algorithms I refer the readers to the following link: http://www.proteomesoftware.com/Proteome_software_link_software.html).
- ii. **Database-independent strategy** consists of converting the fragmentation MS/MS spectra to possible *de novo* amino acid sequences. Several bioinformatics algorithms such as PepNovo, NovoHMM, Mascot, and PEAKS were designed for the heavy computation needed for the extraction of amino acid sequence information and can make MS/MS spectrum interpretable. These *de novo* peptide sequences are then used in BLAST searches of non-redundant protein database (*i.e.*, NCBI, Swissprot). For non-model plants this combination (*de novo* and BLAST search) was proved to improve protein identification in terms of accuracy and rate increase (Shevchenko et al., 2001). Mascot program is a powerful search engine that can use the data from MS/MS spectra in different combinations. For example, this program can perform a search in three

modes: (a) peptide mass fingerprint mode using peptide mass values determined experimentally from MS/MS spectra; (b) sequence query mode using the combination of peptide mass data, amino acid sequence, and amino acid composition (called also sequence tag). In this situation, Mascot program extracts three or four amino acid residues of sequence data from a series of m/z ion peaks and build on them during the BLAST search. This search mode supports both standard and error tolerant sequence tags, and allows arbitrary combinations of fragment ion mass values, amino acid sequence data and amino acid composition data to be searched; and (c) the program can also use directly the raw MS/MS spectra to search a database. Because Mascot program is able to combine different types of searching modes in a single identification search, we decided to use this program for this work.

3. Application of proteomics strategies to the identification of GAX synthesizing GTs

In our previous work we demonstrated that GAX synthase is a multi-enzyme complex (XSC) formed of at least two known GTs (TaGT43-4 and TaGT47-13) and a mutase (TaGT75-4 called also reversibly glycosylated polypeptide, RGP) (Zeng et al., 2010). According to native gel electrophoresis data, this XSC has an apparent MW of ~250 kDa. In this chapter, two proteomics strategies (MudPIT and Gel-LC-MS/MS) were evaluated for efficient identification of these three proteins in membrane fractions enriched in wheat GAX synthase activity (partially purified activity). The goal is to develop a proteomics workflow for optimal identification of candidate proteins involved in plant cell wall biosynthesis in organisms with no available genome sequence information.

3.1 Experimental procedures

3.1.1 Preparation of membrane fractions enriched in GAX synthase complex

GAX synthase activity is routinely monitored as the amount of [^{14}C]glucuronic acid (GlcA) transferred from UDP- ^{14}C]GlcA into ethanol-insoluble GAX polymer in the presence or absence of UDP-xylose (UDP-Xyl) (Zeng et al., 2008, 2010). It has been shown that the rice Golgi complex has distinct compartments that could be separated by density gradient centrifugation in presence of EDTA or MgCl_2 (Mikami et al., 2001). Thus, we fractionated Golgi-enriched membranes prepared from etiolated wheat seedlings on a continuous 25%-40% (w/v) sucrose density gradient supplemented with 1mM EDTA and used our enzyme assay to monitor GAX synthase activity distribution (Figure 2). According to the specific activity and protein content in fraction #3 (as measured via spectrophotometer), we have achieved ~11 fold purification with a 59% yield (Table 1). However, according to SDS-PAGE analysis of this fraction, only limited number of protein bands is visible on the gel (Figure 3), rather suggesting a higher purification was achieved. Together these results indicate that fraction #3 is a good starting material for proteomics analyses for our study.

3.1.2 Nanospray tandem mass spectrometry procedures

All proteomics analyses were carried out at the Mass spectrometry and Proteomics Facility (<http://www.ccic.ohio-state.edu/MS/>) at Campus Chemical Instrument Center (CCIC) of The Ohio State University (Columbus, OH). Liquid chromatography-nanospray tandem mass spectrometry (nano-LC/MS/MS) of global protein identification was performed on an LTQ XL or an LTQ Orbitrap XL (Thermo Scientific) mass spectrometers using different protocols:

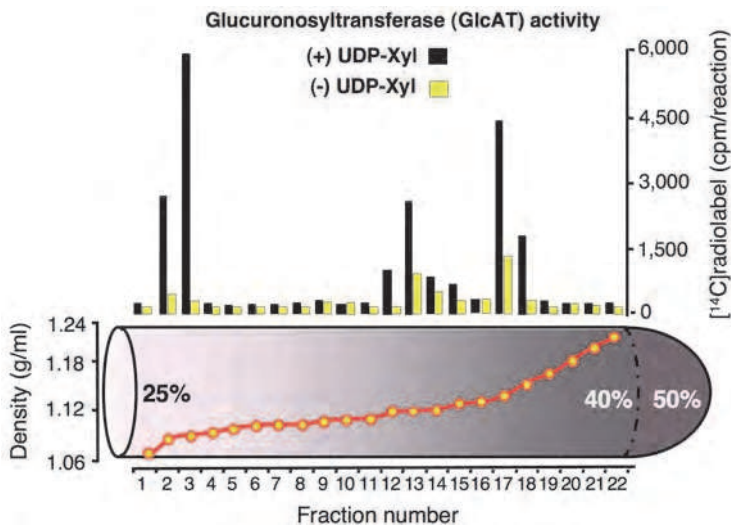


Fig. 2. Distribution of GAX synthase activity after fractionation of wheat Golgi-enriched microsomal membranes on a linear (25%-40%) sucrose density gradient. Microsomal membranes were prepared from 6-day old etiolated wheat seedlings. Fractions of 1ml were collected and each fraction was tested for GAX synthase activity ($[^{14}\text{C}]\text{GlcA}$ incorporation in presence of UDP-xylose, as cpm/reaction) and for sucrose density (g/ml).

Fraction	Volume (ml)	Total Protein (mg)	Total Activity (pmol/h)	Specific Activity (pmol/h/mg)	Fold Purification	Yield (%)
Golgi-enriched membranes	12	60	~3060	51	1	100
Fraction #3 (after sucrose gradient)	1	0.3	~1833	~550	~11	~59

Table 1. Partial purification of wheat GAX synthase activity from Golgi-enriched microsomal membranes. The activity was measured as $[^{14}\text{C}]\text{GlcA}$ incorporation from UDP- $[^{14}\text{C}]\text{GlcA}$ (900cpm/pmol) into ethanol-insoluble materials and expressed as pmol GlcA incorporation per hour per milligram of protein (pmol/h/mg). Protein content was estimated using Bradford reagent.

3.1.2.1 Nano-LC/MS/MS on LTQ XL

Nano-liquid chromatography tandem mass spectrometry (Nano-LC/MS/MS) was performed on a Thermo Scientific LTQ XL mass spectrometer (Linear Quadrupole Ion Trap MSn) equipped with a nanospray source operated in positive ion mode. The LC system was an UltiMate™ 3000 system from Dionex (Sunnyvale, CA). The solvent A was water containing 50 mM acetic acid and the solvent B was acetonitrile. 5 μL of each sample was first injected on to the μ -Precolumn Cartridge (Dionex, Sunnyvale, CA), and washed with 50 mM acetic acid. The injector port was switched to inject and the peptides were eluted off of the trap onto the column. A 5 cm 75 μm ID ProteoPep II C18 column (New Objective, Inc. Woburn, MA) packed directly in the nanospray tip was used for chromatographic

separations. Peptides were eluted directly off the column into the LTQ system using a gradient of 2-80%B over 45 min, with a flow rate of 300 nL/min. The total run time was 65 min. The MS/MS was acquired according to standard conditions established in the lab. Briefly, a nanospray source operated with a spray voltage of 3 KV and a capillary temperature of 200°C is used. The scan sequence of the mass spectrometer was based on the TopTen™ method; the analysis was programmed for a full scan recorded between 350 and 2000 Da, and a MS/MS scan to generate product ion spectra to determine amino acid sequence in consecutive instrument scans of the ten most abundant peaks in the spectrum. The CID fragmentation energy was set to 35%. Dynamic exclusion was enabled with a repeat count of 2 within 10 seconds, a mass list size of 200, an exclusion duration 350 seconds, the low mass width was 0.5 and the high mass width was 1.5 Da.

3.1.2.2 Capillary-LC/MS/MS on LTQ Orbitrap XL

Capillary-liquid chromatography-nanospray tandem mass spectrometry (Capillary-LC/MS/MS) of global protein identification was performed on a Thermo Scientific LTQ Orbitrap XL mass spectrometer equipped with a microspray source (Michrom Bioresources Inc, Auburn, CA) operated in positive ion mode. Samples were separated on a capillary column (0.2 X 150mm Magic C18AQ 3 μ 200A, Michrom Bioresources Inc, Auburn, CA) using an UltiMate™ 3000 HPLC system from LC-Packings A Dionex Co (Sunnyvale, CA). Each sample was injected into the μ -Precolumn Cartridge (Dionex, Sunnyvale, CA) and desalted with 50 mM acetic acid for 10 min. The injector port was then switched to inject and the peptides were eluted from the trap onto the column. Mobile phase A was 0.1% formic acid in water, and 0.1% formic acid in acetonitrile was used as mobile phase B. Flow rate was set at 2 μ L/min. Typically, mobile phase B was increased from 2% to 50% in 90-250 min, depending on the complexity of the sample, to separate the peptides. Mobile B was then increased from 50% to 90% in 5 min and then kept at 90% for another 5 min before being brought back quickly to 2% in 1 min. The column was equilibrated at 2% of mobile phase B (or 98% A) for 30 min before the next sample injection. MS/MS data was acquired with a spray voltage of 2 KV and a capillary temperature of 175°C is used. The scan sequence of the mass spectrometer was based on the data dependant TopTen™ method: the analysis was programmed for a full scan recorded between 300 - 2000 Da and a MS/MS scan to generate product ion spectra to determine amino acid sequence in consecutive scans of the ten most abundant peaks in the spectrum. The resolution of full scan was set at 30,000 to achieve high mass accuracy MS determination. The CID fragmentation energy was set to 35%. Dynamic exclusion is enabled with a repeat count of 30 seconds, exclusion duration of 350 seconds and a low mass width of 0.5 and high mass width of 1.5 Da. Multiple MS/MS detection of the same peptide was excluded after detecting it three times.

3.1.2.3 MS/MS data processing

Sequence information from the MS/MS spectra was processed by converting the raw data files into a merged file (.mgf) using an in-house program, RAW2MZXML_n_MGF_batch (merge.pl, a Perl script). The resulting mgf files were searched using Mascot Daemon by Matrix Science version 2.2.1 (Boston, MA) and the database searched against the full SwissProt database version 54.1 (283,454 sequences; 104,030,551 residues) or NCBI database. The mass accuracy of the precursor ions were set to 2.0 Da given that the data was acquired on an ion trap mass (LTQ) analyzer and the fragment mass accuracy was set to 0.5 Da (for analysis by high resolution Orbitrap, the mass accuracy of the precursor ions were set to 0.1 Da). Considered modifications (variable) were methionine oxidation and carbamidomethyl cysteine. Two

missed cleavages for the enzyme were permitted. Peptides with a score less than 40 were filtered. Protein identifications were checked manually and proteins with a Mascot score of 50 or higher with a minimum of two unique peptides from one protein having *a -b* or *-y* ion sequence tag of five residues or better were accepted. Mascot also provides a “histogram of the MOWSE score distribution” for hits and a value of significance threshold (represented by a green region). A score is defined as $-10 \cdot \text{LOG}_{10}(P)$, where *P* is the absolute probability that the observed match is a random event. Therefore, significant matches would give scores that are higher than the significance threshold (outside of the green region). It is important to know that significance threshold values and scores of the hits are greatly affected by an increase in mass tolerance values in a MS/MS fingerprint search.

3.1.2.4 Gel-LC-MS/MS analysis

One-D SDS-PAGE was carried out according to (Laemmli, 1970). Separation gels were 10% or 12% acrylamide (150V until the loading dye reached the bottom) and molecular weight markers (Precision Plus Protein™, kaleidoscope, Bio-Rad) were used. Proteins were visualized by coomassie blue or silver staining. Briefly, the protein sample (50-60µg) was solubilized in 20 µL Invitrosol (Invitrogen) plus 1 µL of 10% amidosulphobetaine 14 (ASB-14) detergent, 2 µL of 10X SDS-PAGE loading dye, and heated at 90°C for 7 min before fractionation on 1-D SDS-PAGE acrylamide gel. The gel was rinsed with water for 10 min to remove excess SDS detergent, before staining with coomassie blue for 1 h and destained overnight. Figure 3 shows a typical silver-stained 1-D SDS-PAGE gel of fraction #3.

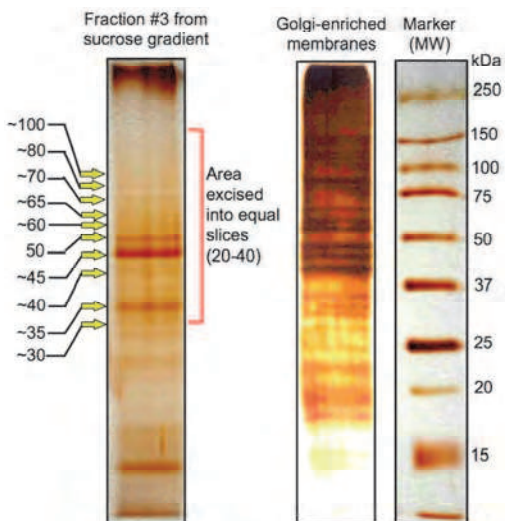


Fig. 3. SDS-PAGE (10%) and silver staining analysis of fraction #3 obtained from Figure 2. The arrows on the left indicate the position of the visible bands that were manually excised and digested with trypsin before ESI-LC-MS/MS-TRAP analysis. The area indicated on the right of the same gel corresponds to the area excised into 20-40 equal slices that were subjected to “In-gel” trypsin digestion followed by ESI-LC-MS/MS-TRAP analysis. Protein identification was carried out using Mascot program to search green plant databases at NCBI. For comparison, the original Golgi-enriched membranes were analyzed. Molecular mass markers (MW) are indicated on the right.

3.1.2.5 In gel digestion

The gels were excised in two ways (see Figure 3): (i) only the most visible protein bands in the gel were manually excised and individually digested with trypsin, or (ii) the gel area from 30 to 180 kDa was excised into 20 to 40 equal small slices and each slice was individually digested with trypsin. The resulting peptides were then analyzed by LC-MS/MS as described above.

For manual excision of visible bands, gels were digested with sequencing grade trypsin from Promega (Madison, WI) using the Multiscreen Solvinert Filter Plates from Millipore (Bedford, MA). Briefly, bands were trimmed as close as possible to minimize background polyacrylamide material. Gel pieces were then washed twice in nanopure water for 5 min each before destaining with 1:1 v/v methanol:50 mM ammonium bicarbonate for 10 min twice. The gel pieces were dehydrated with 1:1 (v/v) acetonitrile:50 mM ammonium bicarbonate, and then rehydrated by incubation with dithiothreitol (DTT) solution (25 mM in 100 mM ammonium bicarbonate) for 30 min prior to incubation in iodoacetamide solution (55 mM in 100 mM ammonium bicarbonate) for 30 min in dark. The gel bands were washed again with two cycles of water and dehydrated with 1:1 (v/v) acetonitrile:50 mM ammonium bicarbonate. Protease digestion was carried out by rehydrating gel pieces in 12 ng/mL trypsin in 0.01% ProteaseMAX Surfactant for 5 min. The gel pieces are then overlaid with 40 mL of 0.01% ProteaseMAX surfactant:50 mM ammonium bicarbonate and gently mixed on a shaker for 1 h. The digestion is stopped with addition of 0.5% TFA. The MS analysis is immediately performed to ensure high quality tryptic peptides with minimal non-specific cleavage or frozen at -80°C until samples can be analyzed.

For gel excision into 20-40 equal slices, robotic digestion was carried out using the Ettan Spot Handling Workstation (Amersham Biosciences). The slices were placed in a 96 well plate that was robotically washed, and slices digested according to the Ettan Spot Handling Workstation 2.1 User Manual. Briefly, gel slices were washed in 100 µL of 50% methanol/5% acetic acid for 30 min. This washing step was repeated 3 times and slices placed in a storage solution of 50 µL of 50% methanol/5% acetic acid until digestion. Gel slices were digested with sequencing grade trypsin from Promega (Madison, WI). Digestion was carried out by adding 100 µL water for 10 min followed by 100 µL acetonitrile for 10 min. The water and acetonitrile were removed and the gel slices were rehydrated and incubated with DTT (prepared as 32 mM in 100 mM ammonium bicarbonate) for 30 min prior incubation in iodoacetamide solution (prepared as 80 mM in 100 mM ammonium bicarbonate solution) in dark for 30 min. The gel slices were washed again with cycles of acetonitrile and 100 mM ammonium bicarbonate in 10 and 5 min increments. The slices were dried for 10 min and then incubated in 25 µL of 50 mM ammonium bicarbonate containing trypsin (5 µg/mL) for 180 min. The peptides were extracted from the polyacrylamide with 50 µL 50% acetonitrile and 5% formic acid (3 times) and the extracts pooled together. The extracted pools were mixed with 50 µL of acetonitrile and incubated for 15 min before drying for 15 min. The MS analysis is immediately performed to ensure high quality tryptic peptides with minimal non-specific cleavage or frozen at -80°C until samples can be analyzed.

3.1.2.6 MudPIT analysis

The first MudPIT analysis on fraction #3 was carried out on a freeze-dried sample that was re-suspended in 5X Invitrosol protein solubilizer (Invitrogen) and diluted to the final 1X Invitrosol by adding 25mM ammonium bicarbonate solution. Ten microliter (10µL) of DTT (5

mg/mL solution prepared in 100mM ammonium bicarbonate) was added to the sample and incubated for 15 min. Ten microliter (10 μ L) of Iodoacetamide solution (15 mg/mL in 100 mM ammonium bicarbonate) was added to the sample and incubated for another 15 min in the dark. However, we noticed that this freeze-dried sample contained higher concentrations of sucrose and salts, which contributed to the production of low quality MS/MS spectra. Therefore, removal of sucrose and salts by precipitation was necessary before trypsin digestion. Two precipitation methods were evaluated: TCA/acetone or chloroform/methanol. These two methods are known to be effective in desalting and lipid removal from protein samples. In our case, TCA/acetone combination gave a better result in removing sucrose and salts from our protein sample (fraction #3), as the quality of MS/MS spectra was improved and scores were higher than the significance threshold (see section 2.3). Trypsin was prepared in 50 mM ammonium bicarbonate and added to the protein solution in trypsin:protein ratio of 1:25 (w/v) and the mixture was incubated for 60 min at 37°C (no difference was observed when incubation time was extended up to 16 h).

3.2 Results

3.2.1 Protein composition analysis of fraction #3 via MudPIT strategy

In our first MudPIT trial, the analysis was carried out on the LTQ XL (ion trap) instrument using MS/MS ion search mode with carbamidomethylation of cysteine residue as fixed modification and methionine oxidation as variable modification. Peptide mass tolerance was set to ± 2 Da, fragment mass tolerance was set to ± 0.8 Da and protein mass was unrestricted. Under these conditions, LTQ analysis of fraction #3 produced a total of 13,504 peptide queries that were used to search the green plant databases at NCBI (6,350,093 sequences). The search yielded 74 peptide matches above significance threshold (set to: Individual ions scores >59 and $p < 0.05$) that were matched by only 501 peptide sequences ($\sim 4\%$ of the 13,504 peptide queries). The decoy value was 41 corresponding to a false discovery rate of 11.33%. Most of the identified hits were protein entries with similar annotation (redundancy) but from different related species (orthologous proteins). Therefore, for simplicity, entries having similar annotation were eliminated, which reduced the number of total protein identified to 25 proteins (Table 2). The low protein identification rate could be due to the inhibition of peptide ionization from low abundant proteins by more abundant proteins. The other possibility could be the interference of some small molecules (i.e., sucrose, salts) present in the sample, which could produce MS/MS spectra of low quality. However, it is possible that some of these peptide queries are species-specific and do not have matches in the database. Therefore, to improve protein identification rate, we repeated the analysis of the same sample (fraction #3) using an Orbitrap high resolution mass spectrometer (LTQ Orbitrap XL). The high resolving power of the Orbitrap (determines masses with very high accuracy) has been useful in analyzing complex peptide mixtures by improving the resolution (narrow peak width), which in turn would increase the detection of more m/z ions. Thus, ions of similar m/z and proteins with multiple charges produced by ESI can be detected distinctly. Our results indicated that, even with reduced peptide mass tolerance (± 0.1 Da) and fragment mass tolerance (± 0.5 Da) values, Orbitrap analysis generated 20,123 peptide queries (6,619 more peptides compared to LTQ analysis). This analysis resulted in the identification of a total of 51 non-redundant proteins that included 37 new proteins (marked by "*" in Table 2). This increase in the rate of protein identification is the result of an increase in the number of matching peptides, 908 peptides (~ 408 more peptides

Accession No	Annotation	MW (Da)	emPAI	Score	Total peptides matched	Peptides with score<59
LTQ analysis						
gi 10720235	Chloroplast precursor (PCR A)	41,358	6.34	1380	145	84
gi 3318722	Leech-Derived Tryptase Inhibitor	24,142	3.35	660	134	101
gi 2493131	V-ATPase subunit B 1	54,107	0.95	685	32	18
gi 90025017	Vacuolar proton-ATPase subunit A	68,754	0.61	672	35	26
gi 10720236	Chloroplast precursor (PCR B)	42,350	0.45	460	25	23
gi 18274925	Vacuolar membrane proton pump	80,108	0.31	376	20	13
gi 476003	HSP70	67,146	0.11	349	12	10
gi 461465	Actin	41,940	0.19	335	13	9
gi 8928416	Tubulin beta-8 chain (Beta-8-tubulin)	50,596	0.07	232	7	6
gi 7960277	RuBisCO activase B	48,012	0.16	215	12	9
gi 461988	Elongation factor 1-alpha	49,489	0.16	202	19	15
gi 90903441	Lipoxygenase 2	17,788	ND	110	4	4
gi 125548120	Hypothetical protein OsI_015175	92,128	0.04	154	4	3
gi 118498764	Glyceraldehyde-3-phosphate dehydrogenase	30,091	0.13	136	6	5
gi 7579064	Cytosolic glyceraldehyde-3-phosphate dehydrogenase	25,448	0.15	135	5	4
gi 13925731	Cyclophilin A-1	18,721	0.21	100	3	3
gi 20302473	Ferredoxin-NADP(H) oxidoreductase	40,491	0.08	102	2	1
gi 914031	WBP2=lipid transfer protein homolog	3,003	1.56	98	4	4
gi 6911146	Glycine-rich RNA-binding protein 2	16,312	0.24	95	8	5
gi 303844	Eukaryotic initiation factor 4A	47,187	0.08	93	2	1
gi 944842	ATP/ADP carrier protein	36,013	0.10	90	5	3
gi 56606827	Calreticulin-like protein	47,404	0.14	82	2	1
gi 56713236	Non-specific lipid transfer protein	12,807	ND	66	2	2
gi 4099408	Tonoplast intrinsic protein	25,431	0.15	65	3	2
gi 115349890	Fasciclin-like protein FLA3	41,600	ND	59	1	0
Total peptides					505	
Orbitrap analysis (Hits marked by "*" are unique hits to from Orbitrap analysis)						
gi 10720235	Chloroplast precursor (PCR A)	41,358	24.30	3667	142	14
gi 90025017	Vacuolar proton-ATPase subunit A	68,754	2.48	2670	88	8
gi 10720236	Chloroplast precursor (PCR B)	42,350	1.60	1111	48	6
*gi 157339402	ATPase subunit A	69,767	0.51	961	29	2
gi 2493131	V-ATPase subunit B 1	54,107	1.28	906	35	4
gi 32481063	Rubisco activase	47,371	0.82	871	28	3
gi 7960277	RuBisCO activase B	48,012	0.97	759	28	5
gi 19880505	Aquaporin PIP1	31,083	1.26	648	27	20
gi 18274925	Vacuolar membrane proton pump	80,108	0.43	637	37	15
*gi 162457723	Binding protein homolog1	73,211	0.32	522	20	11
gi 56606827	Calreticulin-like protein	47,404	0.41	486	12	6
*gi 115383189	Plasma membrane intrinsic protein	30,046	1.85	471	21	14
*gi 34922469	Lipoxygenase 2.1	106,254	0.31	409	24	18
*gi 6525065	Chloroplast TE factor Tu	50,551	0.38	385	9	2
gi 303844	Eukaryotic initiation factor 4A	47,187	0.29	381	12	5
*gi 2827002	HSP70	71,385	0.58	380	23	10
*gi 26986186	Hexose transporter	79,631	0.17	348	13	7
gi 461465	Actin	41,940	0.62	323	21	15
*gi 125535913	Hypothetical protein OsI_036360	17,064	1.00	318	19	15
*gi 8928423	Tubulin beta-4 chain	50,791	0.49	316	12	6
*gi 115874	Serine carboxypeptidase III	55,869	0.16	304	10	4

Accession No	Annotation	MW (Da)	emPAI	Score	Total peptides matched	Peptides with score<59
*gi 14017584	<i>Cytochrome f</i>	35,425	0.77	276	15	12
*gi 14017579	<i>ATP synthase CF1 beta subunit</i>	53,881	0.83	265	33	29
*gi 739292	<i>Oxygen-evolving complex protein 1</i>	26,603	0.57	261	18	15
*gi 7452979	<i>Allene oxide synthase</i>	53,616	0.16	253	7	1
*gi 5123910	<i>HSP80-2</i>	80,533	0.22	252	12	7
*gi 122022	<i>Histone H2B.1</i>	16,423	0.62	248	8	3
gi 90903441	<i>Lipoxygenase 2</i>	17,788	2.05	241	13	12
*gi 127519390	<i>Lipid transfer protein</i>	11,629	0.45	197	6	4
*gi 493591	<i>Disulfide isomerase precursor</i>	56,845	0.44	195	11	10
*gi 1658313	<i>Os07g0683900</i>	39,146	0.23	183	10	9
*gi 73912433	<i>Aspartic proteinase</i>	54,965	0.25	179	9	8
*gi 4158232	<i>Reversibly glycosylated polypeptide</i>	41,985	0.28	168	9	6
*gi 111073715	<i>Triticain alpha</i>	51,572	0.17	160	7	4
*gi 15289940	<i>Putative actin</i>	41,929	0.64	149	17	15
*gi 1556446	<i>Alpha tubulin</i>	50,266	0.38	140	10	9
*gi 32400836	<i>Elongation factor</i>	18,638	0.24	139	4	1
*gi 72256525	<i>Geranylgeranyl hydrogenase</i>	50,966	0.17	139	7	6
*gi 6013196	<i>Heat shock protein 101</i>	100,949	0.13	128	4	2
gi 944842	<i>ATP/ADP carrier protein</i>	36,013	0.12	108	5	4
*gi 3646373	<i>RGP1 [Oryza sativa (indica cultivar-group)]</i>	40,079	0.17	108	4	3
*gi 4158221	<i>Reversibly glycosylated polypeptide [Oryza sativa (indica cultivar-group)]</i>	41,861	0.17	108	4	3
*gi 108709682	<i>Alpha-1,4-glucan-protein synthase [Oryza sativa (japonica cultivar-group)]</i>	30,549	0.17	108	4	3
*gi 3334138	<i>Calnexin homolog precursor</i>	62,270	0.07	100	4	4
*gi 12585487	<i>V-ATPase subunit C</i>	40,131	0.22	91	3	2
*gi 18650668	<i>Temperature stress-induced lipocalin</i>	21,809	0.15	90	2	1
*gi 399414	<i>Elongation factor 1-alpha</i>	49,480	0.28	75	12	12
*gi 90652740	<i>Beta-glucosidase</i>	64,980	0.06	75	5	5
gi 115349890	<i>Fasciclin-like protein FLA3</i>	44,024	0.07	69	2	2
*gi 13027362	<i>Unknown protein (rice)</i>	30,690	0.11	69	2	2
*gi 58339283	<i>Leucine zipper protein zip1</i>	48,111	0.29	57	3	3
Total peptides					908	

Table 2. List of proteins identified in fraction #3 by MudPIT using LTQ or Orbitrap instruments. Fraction #3 was obtained from sucrose density gradient in Figure 2. Hits corresponding to wheat proteins are in italic. Unique hits identified in the Orbitrap analysis are marked by “*”.

compared to LTQ) and generation of new MS/MS spectra by the higher resolution of the Orbitrap. In addition, the Orbitrap analysis improved the scores (i.e., better MOWSE scores) of the identified proteins. Interestingly, some of the proteins identified in LTQ analysis were absent (or have lower scores) in Orbitrap analysis, which may indicate a higher rate of false positive in LTQ analysis. In addition, the decoy value for Orbitrap analysis was 54 corresponding to ~7% false discovery rate (compared to 11.33% in LTQ analysis). It is worth mentioning that we were able to improve protein identification by reducing false positive rate to ~3% after cleaning our samples from residual sucrose, lipids and slats without major increase in protein identification rate.

In the context of GAX biosynthesis, only the Orbitrap analysis resulted in the identification of several hits corresponding to reversibly glycosylated polypeptides (RGPs, GT75 family) from rice (gi|3646373, gi|4158221, gi|108709682) and wheat (gi|4158232). These proteins were identified by the following peptides (all with scores higher than 100): GIFWQEDIIPFFQNATIPK, NLDLFLEMRPFFQPYHLIIVQDGDPSK, YVDAVLTIPIK, TGLPYLWHSK, VPEGFDYELYNR, NLLSPSTPFFNTLYDPYR, and EGAPTAVSHGLWLNIPDYDAPTQMVKPR. However, both LTQ and Orbitrap analyses failed to identify TaGT43-4 and TaGT47-13 or any members of the GT43 or GT47 families. These results underscore the limitations of MudPIT in identifying some GTs involved in plant cell wall biosynthesis.

3.2.2 Gel-LC-MS/MS strategy to analyze protein content of fraction #3

3.2.2.1 Analysis of individual visible bands on 1-D SDS-PAGE

When individual visible bands on the gel (according to coomassie blue staining) were trypsin-digested and the resulting peptides analyzed by LTQ, a total of 169 proteins were matched by 2,106 peptide sequences (~5% of the total peptide sequences, Table 3). Among these 169 proteins, only three proteins (gi|4158232, gi|2218152, and Os01g0926600) were identified as GTs involved in GAX biosynthesis, and one of these three GTs (Os01g0926600)

Approximate MW of the bands (kDa)	Total No of peptide queries	No of peptide sequences with hits	No of hits identified	Top hit [No of peptides matched]	GTs involved in GAX biosynthesis [No of peptides matched]
30	4,356	42	14	gi 10720235 [15]	None
35	4,766	116	17	gi 10720235 [48]	None
40	4,790	122	21	gi 10720235 [21]	gi 4158232, gi 2218152, GT75 [3]
48	3,972	36	10	gi 10720235 [11]	None
50	4,124	127	19	gi 5917747 [15]	Os01g0926600, GT47 [1]
59	4,526	790	29	gi 2493131 [83]	None
64	4,275	376	20	gi 14017569 [62]	None
70	4,739	267	9	gi 90025017 [114]	None
80	4,430	169	17	gi 90025017 [28]	None
100	3,306	61	13	gi 544242 [3]	None
Total	43,284	2,106	169		

Table 3. Proteins identified from Gel-LC-MS/MS analysis of visible bands on SDS-PAGE gel of fraction #3 (Figure 3) involved in GAX biosynthesis. Fraction #3 was from EDTA-supplemented sucrose gradient (see Figure 2). The approximate MW of gel bands is indicated along with the number of proteins identified under the same band.

was newly identified by this analysis (not present in MudPIT analysis). Table 4 lists non-redundant proteins identified by each analysis of individual proteins band. There were 57 new proteins identified by this strategy (not present in MudPIT analysis). Among these new hits, Os01g0926600 (MW 47,271) was identified from the analysis of gel band around 50 kDa by the following peptide **IEGSAGDVLEDDPVGR** (score 79). This rice protein has an exostosin domain belonging to the family GT47 that also contains wheat members known to be involved in GAX synthesis (Zeng et al., 2010). Again, this strategy failed in identifying any members of the GT43 family.

3.2.2.2 Analysis of equal slices covering 30-180 kDa area of the 1-D SDS-PAGE

When the gel area between 30 and 180 kDa was sliced into 20-40 equal slices and each slice subjected to trypsin digestion, a total of 233 proteins were identified through LC-MS/MS using LTQ analysis. These hits were matched by 1,283 peptide sequences, a ~0.9% of the total peptide sequences (Table 5). Table 5 lists the total peptide queries resulted from each slice along with the number of hits identified in NCBI databases, and the number of peptides that matched these hits. This table also lists the top hit from the analysis of each slice (often the top hit is similar in many slices). Table 6 lists the 41 identified proteins

Accession No	Annotation	Score	Peptides matched
Band around 25 kDa			
gi 22204124	Putative 40S ribosomal protein S3 [Triticum aestivum]	117	2
gi 115458216	Os04g0405100 [Oryza sativa (japonica cultivar-group)]	116	2
gi 115475680	Os08g0277900 [Oryza sativa (japonica cultivar-group)]	103	2
gi 41529149	Putative acid phosphatase [Hordeum vulgare subsp. vulgare]	97	4
gi 22022400	Glutathione-S-transferase 19E50 [Triticum aestivum]	72	2
gi 47607439	Mitochondrial ATP synthase precursor [Triticum aestivum]	61	3
Band around 30 kDa			
gi 162459667	Annexin p33 [Zea mays]	188	4
gi 115436780	Os01g0501800 [Oryza sativa (japonica cultivar-group)]	171	4
gi 162462814	Toc34-2 protein [Zea mays]	170	8
gi 82502214	Vacuolar proton ATPase subunit E [Triticum aestivum]	166	13
gi 2586127	β -keto acyl reductase [Hordeum vulgare]	124	4
gi 146231063	Hypersensitive response protein [Triticum aestivum]	97	4
gi 32308080	α -SNAP [Hordeum vulgare subsp. vulgare]	70	2
gi 15226197	Leucine-rich repeat transmembrane protein kinase [Arabidopsis thaliana]	58	3
Band around 35 kDa			
gi 1658313	osr40g2 [Oryza sativa Indica Group]	163	5
gi 20302471	Ferredoxin-NADP(H) oxidoreductase [Triticum aestivum]	162	7
gi 115474135	Os07g0683600 [Oryza sativa (japonica cultivar-group)]	141	5
gi 21952858	Putative 60S ribosomal protein L5 [Oryza sativa Japonica Group]	132	3
gi 115434012	Os01g0104400 [Oryza sativa (japonica cultivar-group)]	118	6
gi 218155	Chloroplastic aldolase [Oryza sativa Japonica Group]	97	4
Band around 40 kDa			
gi 146552329	Glyceraldehyde 3-phosphate dehydrogenase [Zehneria baueriana]	155	6
gi 120670	Glyceraldehyde-3-phosphate dehydrogenase [Zea mays]	130	5
gi 125553548	Hypothetical protein OsI_020490 [Oryza sativa (indica cultivar-group)]	120	9
gi 115437984	Os01g0587000 [Oryza sativa (japonica cultivar-group)]	119	2
gi 110278822	ATP synthase subunit gamma [Zea mays]	111	6

Accession No	Annotation	Score	Peptides matched
gi 157328342	Unnamed protein product [Vitis vinifera]	97	6
gi 2218152	Type IIIa membrane protein cp-wap13 [Vigna unguiculata]	86	3
gi 218157	Cytoplasmic aldolase [Oryza sativa Japonica Group]	63	2
gi 115473689	Os07g0643700 [Oryza sativa (japonica cultivar-group)]	62	3
gi 53791484	DUF642 containing protein, unknown protein [Oryza sativa Japonica Group]	58	2
Band around 45 kDa			
gi 46911561	Putative vacuolar ATPase subunit H [Populus deltoides x 140 maximowiczii]		2
gi 6015084	Elongation factor Tu [Pisum sativum]	137	3
gi 75243541	Probable V-type proton ATPase subunit H [Oryza sativa Japonica Group]	111	2
gi 50251779	Citrate synthase, glyoxysomal precursor [Oryza sativa Japonica Group]	78	2
Band around 50 kDa			
gi 5917747	Elongation factor-1 alpha 3 [Lilium longiflorum]	241	15
gi 7452981	Allene oxide synthase [Hordeum vulgare subsp. vulgare]	125	6
gi 115465711	Os05g0585500 [Oryza sativa (japonica cultivar-group)]	91	2
gi 115441967	Os01g0926600 [Oryza sativa (japonica cultivar-group)] GT47	79	1
gi 37361675	RubisCO large subunit [Aloe niebuhriana]	64	5
gi 90289596	Alpha tubulin-2A [Triticum aestivum]	58	1
gi 2274988	Unnamed protein product [Hordeum vulgare subsp. vulgare]	57	2
Band around 59 kDa			
gi 15233891	Vacuolar ATP synthase B2 [Arabidopsis thaliana]	1999	75
gi 115589736	Serine hydroxymethyltransferase [Triticum monococcum]	551	21
gi 162462751	Mitochondrial F-1-ATPase subunit 2 [Zea mays]	528	19
gi 47522360	Putative calcium-dependent protein kinase [Triticum aestivum]	167	3
gi 81176509	atp1 [Triticum aestivum]	82	7
gi 77554943	Aminoacyl-tRNA synthetase [Oryza sativa (japonica cultivar-group)]	63	2
Band around 64 kDa			
gi 14017569	ATP synthase CF1 alpha subunit [Triticum aestivum]	1363	62
Band around 70 kDa			
gi 1181331	Calnexin [Zea mays]	95	3
gi 51592190	Nucleotide pyrophosphatase/phosphodiesterase [Hordeum vulgare]	57	2
Band around 80 kDa			
gi 50897038	Methionine synthase [Hordeum vulgare subsp. vulgare]	821	28
gi 8134568	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase [common iceplant]	243	13
gi 82582811	Heat shock protein 90 [Triticum aestivum]	198	8
gi 32492578	RNA binding protein Rp120 [Oryza sativa (japonica cultivar-group)]	83	2
gi 115452177	Os03g0271200 [Oryza sativa (japonica cultivar-group)]	64	2
Band around 100 kDa			
gi 544242	Endoplasmic homolog [Hordeum vulgare]	93	3
gi 2429087	Lipoxygenase 2 [Hordeum vulgare subsp. vulgare]	63	2

Table 4. Newly identified proteins by LC-MS/MS analysis from individual visible bands on SDS-PAGE gels (not identified in MudPIT analysis). Only proteins with scores >55 and/or two peptide matches are listed.

Gel area covered MW (kDa)	Slice #	Total No of peptide queries	No of sequences with hits	No of hits identified	Top hit [No of peptides matched]	GTs involved in GAX biosynthesis	
~70 - ~180	1	4,683	88	13	gi 6715512	None	
	2	4,845	141	11	gi 90025017	None	
	3	4,761	41	7	gi 90025017	None	
	4	5,262	137	23	gi 10720235	None	
	5	5,229	37	5	gi 10720235	None	
	6	4,971	17	5	gi 10720235	None	
	7	4,866	23	7	gi 6715512	None	
	8	5,009	23	6	gi 10720235	None	
	9	4,249	3	1	gi 10720235	None	
	10	4,804	12	1	gi 10720235	None	
	11	4,897	23	4	gi 10720235	None	
	12	4,623	160	15	gi 2493132	None	
	13	4,283	38	11	gi 10720235	None	
	16	3,937	10	3	gi 13375563	None	
	17	3,793	3	2	gi 129707	None	
	~30 to ~70	1	5,236	21	5	gi 10720235	None
		2	4,722	16	4	gi 739292	None
3		4,552	13	4	gi 10720235	None	
4		4,564	2	1	gi 10720235	None	
5		4,623	49	4	gi 14017569	None	
6		4,243	30	5	gi 6715512	None	
7		3,950	4	2	gi 20322	None	
8		4,165	4	2	gi 439586	None	
9		4,596	40	8	gi 10720235	gi 159470791, gi 159471277, GT47 family	
10		3,834	18	5	gi 57471704	None	
11		3,481	5	2	gi 10720235	None	
12		4,171	68	9	gi 90025017	None	
13		4,852	17	3	gi 129708	None	
14		3,880	21	5	gi 10720235	None	
15		3,764	283	14	gi 2493131	None	
16		1,007	26	4	gi 10720235	gi 2218152, gi 4158232, GT75 family	
17		1,994	37	6	gi 10720235	None	
18		2,212	40	8	gi 10720235	gi 2218152, gi 4158232, GT75 family	
19		4,766	116	17	gi 10720235	None	
20		4,790	122	21	gi 10720235	None	
Total	37	149,614	1283	233			

Table 5. Proteins involved in GAX biosynthesis identified in fraction #3 by Gel-LC-MS/MS and LTQ strategy. The gel area between 30 and 180 kDa of SDS-PAGE was sliced into 20-40 slices (see Figure 3) and each slice was trypsin-digested and analyzed.

(among the 233 hits) that were unique to this strategy (not in previous analyses). Among the unique hits identified by this strategy, two *Chlamydomonas reinhardtii* GTs (gi | 159470791, and gi | 159471277) belonging to the GT47 family (both annotated as exostosin-like glycosyltransferase) were identified by the following peptide RVAEADIPRL (score 56). This

Accession No	Annotation	Score	Peptides matched
gi 6715512	V-type H ⁺ ATPase B subunit [Nicotiana tabacum]	117	18
gi 2493650	RuBisCO large subunit-binding protein subunit beta	77	2
gi 475600	BiP isoform B [Glycine max]	102	5
gi 123656	Heat shock-related protein [Spinacia oleracea]	94	7
gi 115458184	Calreticulin family, Os04g0402100	62	2
gi 42541152	Delta tonoplast intrinsic protein TIP2;2 [Triticum aestivum]	120	3
gi 1709846	Photosystem II 22 kDa protein [Lycopersicon esculentum]	109	2
gi 4099406	Camma-type tonoplast intrinsic protein [Triticum aestivum]	95	3
gi 28569578	Allene oxide synthase [Triticum aestivum]	59	3
gi 1709358	Nucleoside-triphosphatase [Pisum sativum]	81	2
gi 904147	Adenosine triphosphatase [Sinofranchetia chinensis]	237	11
gi 15010616	AT4g38510/F20M13_70 [Arabidopsis thaliana]	235	8
gi 24496452	Actin [Hordeum vulgare]	172	5
gi 115589744	S-adenosylmethionine synthetase 1 [Triticum monococcum]	74	2
gi 13375563	Lipid transfer protein precursor [Triticum aestivum]	154	4
gi 14017578	ATP synthase CF1 epsilon subunit [Triticum aestivum]	99	2
gi 16225	Calmodulin [Arabidopsis thaliana]	69	2
gi 75108545	Peroxioredoxin Q, chloroplastic [Triticum aestivum]	95	2
gi 464517	50S ribosomal protein L12-1 [Secale cereale]	67	4
gi 115442509	Cyt-b5 family, Os01g0971500 [Oryza sativa (japonica cultivar-group)]	65	3
gi 42565453	Cyclophilin [Hyacinthus orientalis]	59	2
gi 68566191	Cytochrome b6-f complex [Triticum aestivum]	129	5
gi 118104	Peptidyl-prolyl cis-trans isomerase [Zea mays]	89	2
gi 154761388	Cyclophilin [Triticum aestivum]	89	2
gi 231496	Actin-58 [Solanum tuberosum]	166	8
gi 115467154	Os06g0221200 annexin family [Oryza sativa (japonica cultivar-group)]	86	2
gi 52548250	ADP-ribosylation factor [Triticum aestivum]	162	6
gi 74048999	Eukaryotic translation initiation factor 5A1 [Triticum aestivum]	71	2
gi 57471704	Ribosomal protein L11 [Triticum aestivum]	186	8
gi 432607	Ras-related GTP binding protein possessing GTPase activity [Oryza sativa]	119	4
gi 115441299	Os01g0869800, PsbS subunit [Oryza sativa (japonica cultivar-group)]	73	2
gi 115444503	Os02g0171100 [Oryza sativa (japonica cultivar-group)]	283	4
gi 16304127	Glyceraldehyde 3-phosphate dehydrogenase 1 [Fragaria x ananassa]	132	3
gi 18071421	Putative dehydrogenase [Oryza sativa (japonica cultivar-group)]	125	4
gi 166627	Nucleotide-binding subunit of vacuolar ATPase [Arabidopsis thaliana]	115	2
gi 8272480	Fructose 1,6-bisphosphate aldolase precursor [Avena sativa]	92	5
gi 159470791	Exostosin-like glycosyltransferase [Chlamydomonas reinhardtii]	56	2
gi 159471277	Exostosin-like glycosyltransferase [Chlamydomonas reinhardtii]	56	1
gi 115451383	Os03g0200800 [Oryza sativa (japonica cultivar-group)]	202	4
gi 147858623	Hypothetical protein [Vitis vinifera]	65	2

Table 6. List of unique hits identified through Gel-LC-MS/MS and LTQ analysis of 20-40 slices covering 30-180 kDa area of the SDS-PAGE. Only hits with scores >55 and/or two peptide matches are listed.

strategy, however, identified the exact wheat RGP protein (TaGT75-4, gi|4158232) with the following peptides VPEGFYELYNR and YVDAVLTIPK (both with score 59). Therefore, this strategy successfully identified TaGT75-4 protein and homolog to TaGT47-13 but failed to identify the exact TaGT47-13 protein or any homolog to TaGT43-4 protein.

4. Discussion

Hemicellulosic polymers such as GAX represent up to 40% (w/w) of grass cell walls (in particular from growing tissues). In sharp contrast with the abundance of these polymers, the GTs that synthesize these compounds are present in low amounts in Golgi membranes of the plant cell. This observation suggests that these enzymes are highly active and may not be required in large quantities in the plant cell. This low abundance of GTs has been the main limiting factor in applying proteomics approaches to plant cell wall biosynthesis. To further complicate the issue, isolation of GTs from Golgi membranes (or simply disrupting these membranes) generally results in a drastic reduction or loss of transferase activity *in vitro*. To detect this weak transferase activity *in vitro*, it is necessary to use very sensitive biochemical assays (i.e., [¹⁴C]radiolabeled sugars-based assay). Since the loss of transferase activity is GT-dependent, the biochemical assays are not the best way to estimate the abundance of these enzymes in a particular protein preparation. Therefore, when working with plant cell wall GTs, all these factors should be taken in consideration. In this work, such *in vitro* assay was used to monitor the distribution of GAX synthase activity (from Golgi-enriched membranes) on a linear sucrose density gradient supplemented with EDTA as described earlier (Zeng et al., 2010). According to our *in vitro* assay, fraction #3 was substantially enriched in GAX synthase activity (Figure 2), and it can be assumed that this fraction is also enriched in TaGT43-4, TaGT47-13, and TaGT75-4 proteins. Therefore, fraction #3 is an excellent starting material to evaluate proteomics strategies in identifying these three GTs among a mixture of proteins. Furthermore, because genome and protein sequence information from five grass species are currently publicly available (Figure 1), it can be expected that proteomics analysis on wheat would be successful.

Our analyses indicated that gel-based proteomics approach (gel-LC-MS/MS) has a superior result compared to gel-free approach (i.e., MudPIT). In the MudPIT strategy, LTQ and Orbitrap analyses identified a total of 83 non-redundant proteins, but only 14 of these proteins were in common (Figure 4). However, the Orbitrap gave higher scores and protein identification rates. On the other hand, the Gel-LC-MS/MS strategy resulted in the identification of a total of 180 non-redundant proteins, among which 83 proteins were in common with MudPIT analyses (97 new proteins) (Figure 4).

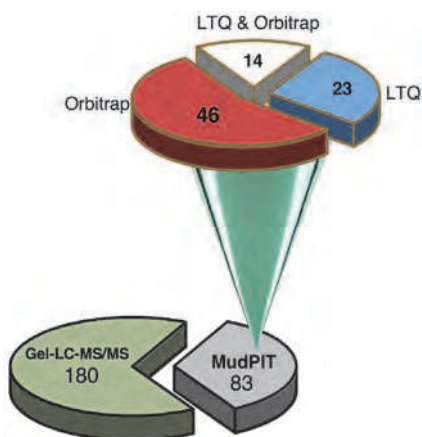


Fig. 4. Distribution of protein hits identified by Gel-LC-MS/MS and MudPIT strategies.

Regarding the ability to identify GTs, the Gel-LC-MS/MS strategy identified most of the GTs associated with GAX biosynthesis. Intriguingly, all the strategies used failed to identify TaGT43-4 or any closest homolog from the NCBI database. Three possibilities could explain this result: (i) the TaGT43-4 protein may be lost during the precipitation step (preparation of the sample); (ii) TaGT43-4 is a very active enzyme and is present in only small amounts in fraction #3, which may not be detectable by the LC-MS/MS methods used in this work, or (iii) TaGT43-4 protein is somehow resistant to trypsin digestion.

Our hypothesis is that most of the TaGT43-4 protein was lost during the precipitation step, as Golgi proteins are known to easily aggregated during precipitation and are very difficult to re-solubilize in a buffer containing detergent. Although it has been shown that ASB-14 and SDS detergents are suited for solubilizing hydrophobic proteins (Herbert, 1999), their use in this study may not be efficient in re-solubilizing freeze-dried or TCA/acetone precipitated wheat Golgi proteins. In support of this hypothesis, fraction #3 should be enriched in Golgi proteins (Zeng et al., 2010), but our analysis indicates that fraction #3 was actually enriched in endoplasmic reticulum (14%), tonoplast (17%), and plastid (28%) proteins, and Golgi proteins represented only 2% of the total hits (according to NCBI annotation of possible subcellular localizations) (Figure 5). Therefore, a reliable 'precipitation-re-solubilization' strategy appears to present a crucial step that must be optimized for minimal protein loss. Alternatively, improving enrichment strategies to overcome protein loss during the 'precipitation-re-solubilization' step should be developed.

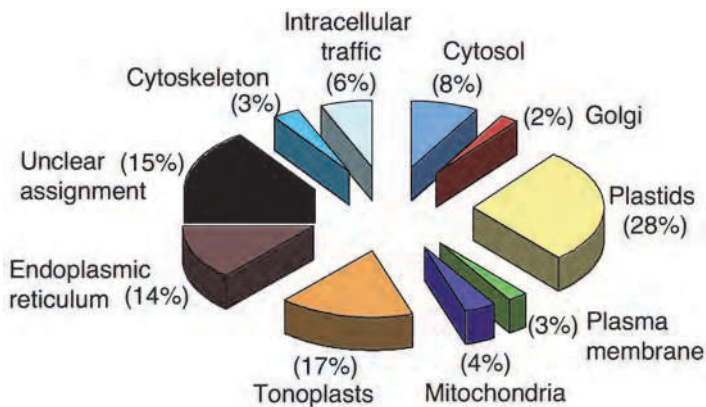


Fig. 5. Classification of proteins identified in fraction #3 according to NCBI annotation of their possible sub-localization.

Although all proteomics strategies employed here have failed to reveal the exact identity of some GTs associated with GAX biosynthesis, proteomics is still a powerful tool, as many low abundant GTs (among the 2% proteins from the Golgi) could be identified. Furthermore, this work demonstrated that working with a non-model species without a fully sequenced genome such as wheat did not seem to be an issue, as most (40-60%) of the proteins identified were either from wheat sequences available in the databases, or were closest homologs to the anticipated wheat proteins from grass species (rice, barley, maize, or sorghum). The other limitation in applying proteomics to plant cell wall biosynthesis is the capacity of a mass spectrometer analyzer to extract as many MS/MS spectra as possible to increase the detection rate of proteins. To overcome all these issues and depending of the

complexity of the protein sample, we are proposing a workflow to carry out a successful proteomics analysis (Figure 6). In this workflow, the first step is to assess the quality of the sample by optimizing the precipitation step (removal of salts and contaminants) without any protein loss. Our work demonstrated that “precipitation-re-solubilization” step is crucial in a successful proteomics analysis of Golgi membrane proteins. Depending of the complexity of the samples, the simplest proteomics strategy to try is the combination of MudPIT fractionation with LTQ analysis. If the sample contains more than 500 proteins, the use of high resolution mass spectrometry (e.g. Orbitrap) in combination with MudPIT could

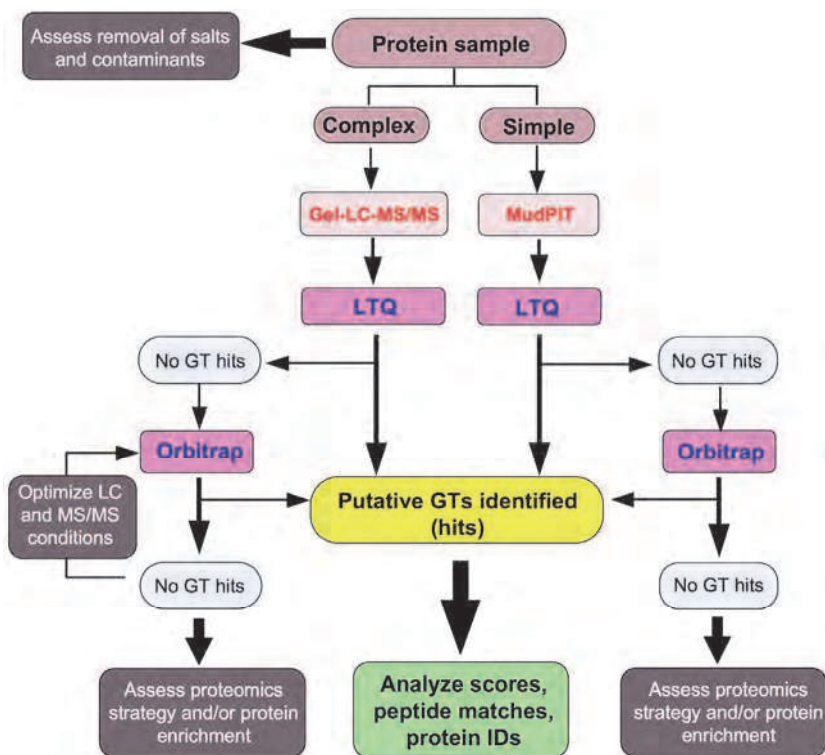


Fig. 6. A proteomics workflow pipeline for efficient protein identification from unknown samples.

be the easiest strategy to test. For more complex protein samples (more than 1000 proteins) it may be necessary to combine 1-D SDS-PAGE fractionation, LC-MS/MS and the high resolving power of the Orbitrap analyzer for optimal protein identification (Figure 6).

The distribution of GAX synthase over sucrose density gradient was intriguing. In the absence of EDTA, all GAX synthase activity stabilized at the expected density of ~1.16g/mL (fractions 17 and 18 in Figure 2) along with the Golgi marker activity IDPase (Zeng et al., 2010). The inclusion of EDTA in the sucrose gradient resulted in splitting of the activity into three density areas, namely around density 1.09g/mL (fractions 2 and 3 in Figure 2), around density 1.14g/mL (fractions 12 and 13 in Figure 2), and around density 1.16g/mL (fractions 17 and 18 in Figure 2). Fraction #3 contained the highest GAX synthase activity. This shift in the density

may suggest that GAX synthase activity is associated with various Golgi compartments. The presence of such Golgi compartments was reported earlier by Mikami et al. (2001) in rice but not in tobacco cells. They showed that rice Golgi complex fractionated into several compartments by simple centrifugation on density gradient in presence of EDTA or $MgCl_2$. Recently, Asakura et al., (2006) used this strategy to isolate (and analyze by proteomics) rice cis-Golgi membranes labeled with green fluorescent protein (GFP) fused to a cis-Golgi marker SYP31 (which belongs to a family of SNARE proteins; soluble N-ethyl-melaeimide sensitive factor attachment protein receptor). Interestingly, their proteomics results gave very similar protein composition to our data, except that no members of the GT43, 47, and 75 were identified in their study (Asakura et al., 2006). Taking together, these results suggest that the cis-Golgi is less tightly attached to the medial and trans-Golgi compartments in grasses. Therefore, we are tempted to propose a possible explanation of the effect of EDTA on the dissociation of these Golgi compartments (Figure 7). Our hypothesis is that some ions (*i.e.*, Ca^{2+}) are involved in linking cis-Golgi (and probably the trans Golgi network [TGN]) to the medial and trans-Golgi cisternae. In the absence of EDTA, the whole Golgi complex would stabilize at an apparent high density ($\sim 1.16g/mL$, fraction 17 and 18 in Figure 2). Upon addition of EDTA into sucrose gradient, the ions are chelated leaving different Golgi compartments stabilized at their corresponding densities (1.09, 1.12, and 1.16g/mL in Figure 2). In any case, the fractionation of the Golgi complex from grasses in the presence of EDTA is an excellent tool for isolating different Golgi compartments.

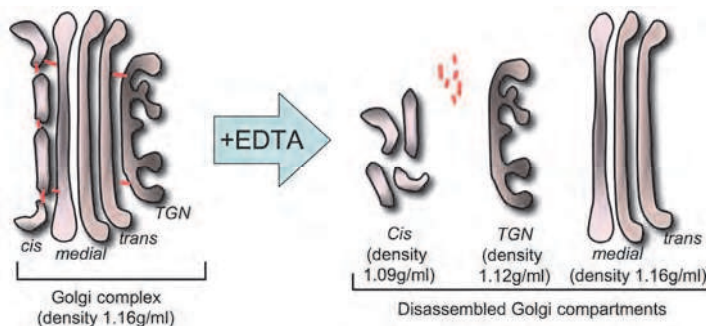


Fig. 7. A diagrammatic representation of the possible effect of EDTA on the dissociation of Golgi compartments, cis-, medial-, trans-Golgi, and trans Golgi network (TGN). EDTA would chelate metal ions that mediate the attachment (red lines) of cis-Golgi compartment to the medial and trans-Golgi cisternae.

5. Conclusion

We evaluated two proteomics strategies for MS/MS identification of GTs associated with GAX synthase complex in wheat for which little genome sequence is available. The evaluation of these strategies is based on their capacity to identify the exact wheat proteins TaGT43-4, TaGT47-13, and TaGT75-4, or at least their closest homologous proteins from other grass species such as rice, barley, maize, or sorghum. Therefore, these strategies are MS/MS spectra quality-dependent and cross-species-dependent using error tolerant BLAST search and *de novo* peptide sequences generated from the MS/MS spectra. Our data indicated that the highest number of unique hits identified (180 proteins) was obtained

through Gel-LC-MS/MS strategy using the Orbitrap analyzer, but the fact that TaGT43-4 and/or its homologous proteins were not identified by this method underscores the importance of optimizing sample preparation step (precipitation-re-solubilization). Based on our results and interpretations, we have proposed a workflow chart that includes routinely used proteomics methods and optimization steps to help increase the detection rate of proteins of plant cell wall GTs from non-model plants.

6. Acknowledgment

The author would like to thank Dr. Green-Church for her valuable comments and editing of the experimental procedures dealing with proteomics analysis. Thanks to Wei Zeng and Nan Jiang for the excellent work preparing protein samples for proteomics. This work was supported by the National Science Foundation (grant no. IOS-0724135 to A.F.)

7. References

- Asakura, T., Hirose, S., Katamine, H., Kitajima, A., Hori, H., Sato, M.H., Fujiwara, M., Shimamoto, K. & Mitsui, T. (2006) Isolation and proteomic analysis of rice Golgi membranes: cis-Golgi membranes labeled with GFP-SYP31. *Plant Biotechnology* 23: 475-485
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815
- Bodnar, W.M., Blackburn, R.K., Krise, J.M. & Moseley, M.A. (2003) Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage. *J. Am. Soc. Mass Spectrom.* 14: 971-979
- Carpita, N.C. (2011) Update on Mechanisms of Plant Cell Wall Biosynthesis: How Plants Make Cellulose and Other (1-4)- β -D-Glycans. *Plant Physiol.*, 155: 171-184
- Chen, H.S., Rejtar, T., Andreev, V., Moskovets, E. & Karger, B.L. (2005) Enhanced characterization of complex proteomic samples using LC-MALDI MS/MS: exclusion of redundant peptides from MS/MS analysis in replicate runs. *Anal. Chem.*, 77: 7816-7825
- Chen, Y.Y., Lin, S.Y., Yeh, Y.Y., Hsiao, H-H., Wu, C-Y., Chen, S-T. & Wang A. H-J. (2005) A modified protein precipitation procedure for efficient removal of albumin from serum. *Electrophoresis*, 26: 2117-27
- Coutinho, P.M., Deleury, E., Davies, G.J. & Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* 328: 307-317
- Ephritikhine, G., Ferro, M. and Rolland, N. (2004) Plant membranes proteomics: *Plant Physiol. Biochem.*, 42: 943-962
- Faik, A. (2010) xylan biosynthesis: news from the grass. *Plant Physiol.*, 153: 396-40
- Faik, A., Bar-Peled, M., DeRocher, A.E., Zeng, W., Perrin, R.M., Wilkerson, C., Raikhel, N.V. & Keegstra K. (2000) Biochemical characterization and molecular cloning of an alpha-1,2-fucosyltransferase that catalyzes the last step of cell wall xyloglucan biosynthesis in pea. *J. Biol. Chem.*, 275: 15082-15089.
- Ferro, M., Salvi, D., Brugiére, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J. & Rolland, N. (2003) Proteomics of the chloroplast envelop membranes from *Arabidopsis thaliana*. *Mol. Cell Proteomics*, 2: 325-345
- Herbert, B. (1999) Advances in protein solubilization for two-dimensional electrophoresis. *Electrophoresis*, 20: 660-663

- Komatsu, S., Muhammad, A. & Rakwal, R. (1999) Separation and characterization of proteins from green and etiolated shoots of rice (*Oryza sativa* L.): towards a rice proteome. *Electrophoresis*, 20: 630-636
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227: 680-685
- McDonald, W.H., Ohi, R., Miyamoto, D.T., Mitchison, T.J. & Yates III, J.R. (2002) Comparison of three directly coupled HPLC MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int. J. Mass Spectrom.* 219: 245-251
- Mikami, S., Hori, H. & Mitsui, T. (2001) Separation of distinct compartments of rice Golgi complex by sucrose density gradient centrifugation. *Plant Sci.*, 161: 665-675
- Newton, R.P., Breton, A.G., Smith, C.J. & Dudley, E. (2004) Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry*, 65: 1449-1485
- Nesvizhskii, A.I., Vitek, O. & Aebersold, R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* 4: 787-797
- O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, 250: 4007-4021
- Patterson, S.D. (2003) Data analysis-the Achilles heel of proteomics. *Nat. Biotechnol.*, 21: 221-222
- Perrin, R.M., DeRoche, A.E., Bar-Peled, M., Zeng, W., Norambuena, L., Orellana, A., Raikhel, N.V. & Keegstra, K. (1999) Xyloglucan fucosyltransferase, an enzyme involved in plant cell wall biosynthesis. *Science* 284: 1976-1979
- Santoni, V., Kieffer, S., Desclaux, D., Masson, F. & Rabilloud, T. (2000) Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. *Electrophoresis*, 21: 3329-3344
- Service, R.F. (2008) Proteomics. Proteomics ponders prime time. *Science* 321: 1758-1761
- Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W. & Standing, K.G. (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time of flight mass spectrometry and Blast homology searching. *Anal. Chem.*, 73: 1917-1926
- Simon, W.J., Maltman, D.J. & Slabas, A.R. (2008) Isolation and fractionation of the endoplasmic reticulum from castor bean (*Ricinus communis*) endosperm for proteomic analysis. *Methods Mol. Biol.*, 425: 203-215
- Washburn, M.P., Wolters, D. & Yates III, J.R. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.*, 19: 242-247
- Wolters, D.A., Washburn, M.P. & Yates III, J.R. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.*, 73: 5683-5690
- Zellner, M., Winkler, W., Hayden, H., Diestinger, M., Eliassen, M., Gesslbauer, B., Miller, L., Chang, M., Kungl, A., Roth, E. & Oehler, R. (2005) Quantitative validation of different protein precipitation methods in proteome analysis of blood platelets. *Electrophoresis*, 26: 2481-2489
- Zhu, W., Smith, J.W. & Huang C.M. (2010) Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.*, 840518. Doi:10.1155/2010/840518
- Zeng, W., Jiang, N., Nadella, R., Killen, T.L., Nadella, V. & Faik, A. (2010) A glucurono(arabino)xylan synthase complex from wheat contains members of the GT43, GT47, and GT75 families and functions cooperatively. *Plant Physiol.*, 154: 78-97



Proteomic Applications in Biology

Edited by Dr. Joshua Heazlewood

ISBN 978-953-307-613-3

Hard cover, 264 pages

Publisher InTech

Published online 18, January, 2012

Published in print edition January, 2012

The past decade has seen the field of proteomics expand from a highly technical endeavor to a widely utilized technique. The objective of this book is to highlight the ways in which proteomics is currently being employed to address issues in the biological sciences. Although there have been significant advances in techniques involving the utilization of proteomics in biology, fundamental approaches involving basic sample visualization and protein identification still represent the principle techniques used by the vast majority of researchers to solve problems in biology. The work presented in this book extends from overviews of proteomics in specific biological subject areas to novel studies that have employed a proteomics-based approach. Collectively they demonstrate the power of established and developing proteomic techniques to characterize complex biological systems.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Faik Ahmed (2012). Assessment of Proteomics Strategies for Plant Cell Wall Glycosyltransferases in Wheat, a Non-Model Species: Glucurono(Arabino)Xylan as a Case Study, *Proteomic Applications in Biology*, Dr. Joshua Heazlewood (Ed.), ISBN: 978-953-307-613-3, InTech, Available from:

<http://www.intechopen.com/books/proteomic-applications-in-biology/assessment-of-proteomics-strategies-for-plant-cell-wall-glycosyltransferases-in-wheat-a-non-model-sp>

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.