

# **Enrichment proteomics challenges and perspectives**

**Analysis of the *N*-glycoproteome and plasma membrane proteome in glycosylation mutants and plant-pathogen interactions**

**Wei Song**

## **Thesis committee**

### **Promotor**

Prof. Dr H.J. Bouwmeester  
Professor of Plant Physiology  
Wageningen University

### **Co-promotors**

Dr A.R. van der Krol  
Associate professor, Laboratory of Plant Physiology  
Wageningen University

Dr A.H.P. America  
Researcher, Plant Research International  
Wageningen University & Research

### **Other members**

Prof. Dr G.C. Angenent, Wageningen University  
Dr M. Rep, University of Amsterdam  
Dr C.H. Hokke, Leiden University Medical Center  
Dr M.H.A.J. Joosten, Wageningen University

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences (EPS)

# **Enrichment proteomics challenges and perspectives**

## **Analysis of the *N*-glycoproteome and plasma membrane proteome in glycosylation mutants and plant-pathogen interactions**

**Wei Song**

Thesis  
submitted in fulfilment of the requirements for the degree of doctor  
at Wageningen University  
by the authority of the Rector Magnificus  
Prof. Dr A.P.J. Mol,  
in the presence of the  
Thesis Committee appointed by the Academic Board  
to be defended in public  
on Wednesday 19 October  
at 4 p.m. in the Aula

## **Wei Song**

Enrichment proteomics challenges and perspectives: analysis of the *N*-glycoproteome and plasma membrane proteome in glycosylation mutants and plant-pathogen interactions

172 pages

PhD thesis, Wageningen University, Wageningen, NL (2016)  
With references, with summary in English

ISBN 978-94-6257-872-2  
DOI 10.18174/386853

## Table of contents

<b>Chapter 1</b>	7
General introduction	
<b>Chapter 2</b>	27
<i>N</i> -glycoproteomics in plants: challenges and perspectives	
<b>Chapter 3</b>	55
<i>N</i> -glycan occupancy of <i>Arabidopsis</i> <i>N</i> -glycoproteins	
<b>Chapter 4</b>	85
Comparative proteomics of <i>Arabidopsis</i> plants with different susceptibility to <i>Phytophthora brassicae</i> identifies CYSTEINE-RICH PROTEIN KINASE 37 as resistance component	
<b>Chapter 5</b>	109
Plasma membrane proteomics of tomato MsK8 suspension cells infected with <i>Phytophthora infestans</i>	
<b>Chapter 6</b>	139
General discussion	
<b>Summary</b>	157
<b>Acknowledgements</b>	161
<b>Curriculum vitae</b>	167
<b>List of publications</b>	169
<b>Education certificate</b>	170

**To my beloved parents, relatives and friends**

献给我的父母，亲戚和朋友

艺术求美，科学求真

# Chapter 1

## General Introduction

Wei Song

## Background

This thesis was carried out within the research program of the Centre of Biosystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative of the Netherlands Organization for Scientific Research. The program is based on two Technology Development (TD) projects: *N*-glycoproteomics in *Arabidopsis* and its applications (TD7) and Plasma membrane proteomics in plant pathogen interactions (TD5). Both projects were focused on the development of plant proteomics methodology, utilizing different types of proteome enrichment techniques. In Liquid Chromatography coupled to Mass Spectroscopy (LC-MS)-based proteomics studies, proteins are isolated from the tissue or cells of interest, digested with trypsin and the resulting peptides are separated by LC and measured by MS. In proteomics of a total cell protein extract, peptide detection may be dominated by peptides from the most abundant proteins, resulting in an underrepresentation of low abundant peptides/proteins. Moreover, in very complex peptide samples the separation of peptides may not be optimal and different peptides with overlapping elution during LC may hinder ionization of the peptide and hence mass spectrometer detection. Therefore, proteomics can benefit from specific enrichment procedures with allow focusing on a particular type of proteins in the cell. In this project we developed the techniques to measure proteins that are modified by *N*-glycosylation or which specifically reside in the plasma-membrane (PM). Because *N*-glycosylation takes place on proteins that enter the secretory pathway, which includes proteins that migrate to the PM, many PM proteins are also glycosylated. Therefore, results from *N*-glycoproteomics and PM proteomics can be complementary. The role of protein *N*-glycosylation in plants is still not fully understood and for most plant proteins it is not even known whether they are *N*-glycosylated. The developed glycoproteomics protocol was therefore applied to make an inventory of all *N*-glycosylated proteins in *Arabidopsis*. The PM proteomics was applied to quantify changes in the PM-proteome of plant cells and its importance in plant-pathogen interactions.

### 1. General approaches in proteomics

Proteomics is widely used for the characterization of cellular responses to different treatments that are not marked by transcriptional changes but may initially only be detectable at the protein level. In proteomics the level of different proteins is determined by measuring the abundance of peptides derived from each protein. Proteomics may be used to identify proteins in a protein extract, while in quantitative proteomics peptide abundance is also quantified to get information about quantitative changes in the proteome.

Until recently, gel-based proteomics was the main method for proteome profiling via one or two dimensional polyacrylamide gel electrophoresis (2DGE). Individual protein spots on the gel were isolated for trypsin digestion, mass spectrometry and subsequent



identification of the protein. 2DGE has a high resolving power, and a few thousand protein spots can be separated and visualized on a single gel (Tannu and Hemby, 2006). However, this number is still far below the actual number of proteins in most biological samples and low abundant proteins are not or difficult to detect by gel staining assays. A typical 2DGE can only visualize 30-50% of the entire proteome (Baggerman et al., 2005). Moreover, for quantitative proteomics, comparison of protein content in two different samples is needed and this requires alignment of protein spots from two different samples for which specialized software has been developed. Instead of comparing protein spots between different gels, samples may also be run together in a single gel to allow comparison of protein spot intensity from two samples by differential fluorescent labeling of the proteins in each sample by N-hydroxy succinimide fluorescent dyes (e.g. Cy3, Cy5) (Van den Bergh et al., 2003). 2D-difference gel electrophoresis (2D-DIGE) has been used successfully to detect differential protein abundances in two protein samples, however, it is only efficiently applied in comparison of two protein samples. In experiments where multiple conditions or time points have to be compared, the use of 2D-DIGE becomes a major problem due to limitations in dynamic range and insufficient separation of proteins (Ong and Pandey, 2001). For such cases gel free proteomics has become more popular.

Gel free proteomics makes use of separation of peptides derived from a trypsin digest of the protein extract instead of separation of the proteins by 2D PAGE. Peptides are separated by LC and when needed, separation of peptides in complex mixtures may be enhanced by multi-dimensional liquid chromatography, followed by tandem mass spectrometry. The peptides that are identified are then used to identify the corresponding proteins. A peptide based proteomic analysis can be performed much faster and cost effectively than a complete gel-based analysis. For relative peptide quantitation in multiple samples there are many peptide labeling approaches that can be used in gel free proteomics such as ICAT (isotope-coded affinity tagging), SILAC (stable isotope labeling in cell culture), iTRAQ (Isobaric tag for relative and absolute quantitation), etc. However, in these peptide labeling procedures not all the protein/peptides are tagged and different efficiencies of peptide labeling may hinder accurate relative quantification in different samples (Patel et al., 2009). Such problems that are introduced by the peptide/protein labeling for quantitative proteomics may be avoided by using label-free comparative proteomics.

## 2. Label-free comparative proteomics

Label-free comparative proteomics is a popular tool to reveal changes in the whole protein pool of samples following different treatments, as it does not suffer the constraints and costs of labeling methods. In this research we used label-free comparative proteomics for both the proteins selected for *N*-glycoproteomics and proteins selected for PM proteomics. While label-free proteomics sample preparation may be less time consuming and does not suffer from variations in labeling of proteins

with heavy isotope labeling methods such as SILAC, ICAT and iTRAQ, the label-free proteomics does require extra care for reliable relative quantification. Therefore, label free proteomics needs multiple replicates per protein sample, which in the post-processing of data introduces a challenge for the alignment between samples. Still, with the help of recently developed software for protein identification and quantification, in combination with advanced statistical tools, label-free proteomics can be used as a reliable and powerful tool to unravel the proteome profile in multiple protein samples.

## 1

In LC-MS analysis of a peptide sample the machine can switch between the data acquisition mode for peptide identification: DDA (data dependent acquisition) and for peptide quantification:  $MS^E$  (data independent acquisition, DIA). These MS modes provide different types of output, each with their own constraints and missing data. In DDA, during each spectrum scan the most abundant precursor ions are selected for  $MS/MS$ . Therefore good fragmentation patterns can be obtained, which helps in determining the peptide amino acid sequence and thus identifying the corresponding protein. The trade-off for this mode is that other masses that co-elute at the same time are selected and therefore not identified. Also, at lower frequency of MS spectrum collection, for each detected mass, the intensity signal covers only part of the related eluting peak. DDA data (at lower MS frequency) are therefore not quantitative, but good for protein identification. In contrast, in  $MS^E$  mode all precursor ions are measured more frequently (every 1 or 2 seconds) and data on intensity of the signal over the entire eluting peak are captured in the  $MS^E$  mode. Therefore, the  $MS^E$  mode provides good quantitative information for each of the detected peptide masses, but  $MS^E$  spectra are less suitable for peptide identification.

In this project we analyzed protein/peptide samples by both DDA and  $MS^E$  and one of the major challenges was to link DDA data (peptide identity) to  $MS^E$  data (peptide abundance) in multiple samples. In MS-based proteomics the efficiency of peptide identification and quantification is also influenced by the complexity of the protein sample, due to co-elution of multiple peptides. Therefore, complex protein samples like from whole plant cell extracts may first be fractionated by a first dimension LC in multiple sub-pools, followed by separate second dimension LC-MS of these individual sub-pools. However, this also introduces the additional problem that peptide information belonging to the same protein may be distributed over multiple LC-MS runs.

The complexity of the protein samples may also be reduced by focusing on enrichment of particular types of proteins (e.g. proteins with an *N*-glycan or phosphorylated proteins, etc.). Alternatively, protein isolates may be reduced in complexity by enrichment of proteins located at particular sites of the cells (e.g. soluble proteins versus membrane located proteins, chloroplast proteins, etc.). In my thesis I explored two approaches to protein enrichment for label-free quantitative proteomics: selection of *N*-glycosylated proteins/peptides and selection of (plasma)membrane bound proteins. Each of these approaches may be linked to different types of biological

questions. Modification of proteins by *N*-glycans is one of the most common protein modifications, but has not extensively been studied in plants because mutants in *N*-glycosylation do not show a clear phenotype (in contrast to *N*-glycosylation mutants in other eukaryotes). Therefore, the function of *N*-glycosylation in plants is still not fully understood. Proteins in the plasma membrane are often under-represented in full proteome studies as these proteins are more difficult to extract and usually have a low abundance. Analysis of proteins from the plasma membrane of cells is especially of relevance when studying the responses of plants to microbial pathogens, of which the detection usually occurs in the extracellular space and for which the presence may be signaled to the cell by receptors in the plasma membrane.

### 3. *N*-glycoproteomics

Glycosylation is one of the most common post-translational modifications of proteins in eukaryotic cells. It is estimated that half of the expressed proteome in higher eukaryotes consist of secreted proteins which are modified by *N*-glycans (Varki, 1999). Glycoproteins are proteins modified with a glycan, consisting of one or several sugar groups. In eukaryotes, proteins may be modified at different positions and by different types of sugars. Within the pool of proteins modified by glycans we distinguish *N*-glycans, *O*-glycans and glycosyl phosphatidyl inositol (GPI) lipid anchor. Glycans may be attached to serine or threonine in the protein backbone (*O*-glycosylation) or to asparagine (*N*-glycosylation). Whereas *O*-glycosylation takes place in the cytosol, *N*-glycosylation takes place on proteins that enter the secretory pathway upon import into the ER on the consensus site Asn-X(P!)-Ser/Thr (N-P!-S/T, where X is any amino acid except proline) (Varki, 1999). *O*-glycan attachment sites seem to be less well conserved compared to the *N*-glycan attachment sites on proteins. GPI anchor attachment may take place on both cytosolic proteins and in the ER secretory pathway without the presence of a clear consensus site. In this thesis research we used special techniques to study *N*-glycoproteins, so we will first summarize the essential points of *N*-glycosylation in plants in comparison to *N*-glycosylation in animals.

#### 3.1 *N*-glycosylation in plants versus *N*-glycosylation in mammals

In both mammals and plants, *N*-linked glycan biosynthesis starts at the cytosolic face of the ER where two *N*-acetylglucosamine and five mannose residues are added onto the dolichol carrier. Then the intermediate dolichol linked oligosaccharide is flipped from the cytosolic to the luminal face of the ER. Eventually, four more mannoses and three glucose residues are transferred to the oligosaccharide and the oligosaccharide precursor is transferred from the dolichol to an asparagine residue of a nascent peptide. *N*-glycans are further processed in the ER, and later on in the Golgi apparatus. In the ER, the terminal glucoses and one or two mannoses residues of Glc<sub>3</sub>Man<sub>6</sub>GlcNAc<sub>2</sub> are removed. Then the glycoprotein may be transported to the Golgi apparatus and the remaining  $\alpha$ 1,2-linked mannoses are removed by  $\alpha$ 1,2-mannosidase (ManI) resulting

in a  $\text{Man}_5\text{GlcNAc}_2$  structure. In the Golgi, an additional GlcNAc residue is added to the  $\text{Man}_5\text{GlcNAc}_2$  catalyzed by *N*-acetylglucosaminyltransferase I (GnTI). This is an obligatory step, and only after this the *N*-glycans can be further modified by glycosidases and glycosyltransferases which can add additional sugars with different linkages to the *N*-glycan. Since many of these reactions are not always completed, this results in a heterogeneous glycan profile existing of specific glycoprotein forms in different tissues. The *N*-glycosylation pathway is evolutionary strongly conserved in all eukaryotes, but certain *N*-glycan core modifications like fucosylation and xylosylation are specific for higher plants, suggesting an important role for these plant specific glycan-modifications.

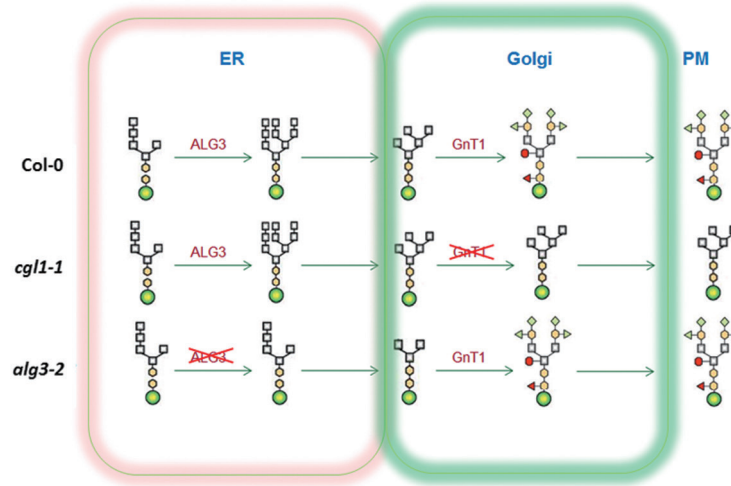
## 1

### 3.2 Function of protein *N*-glycosylation in ER relates to protein folding quality control

One general role of *N*-glycosylation for proteins secreted into the ER is in protein folding control, but for animal systems a role for *N*-glycans has also been implicated in polar protein localization, protein stability, ligand binding, endocytosis, immune recognition, inflammation and pathogenicity, cell signaling and cell motility (Varki, 1999). The importance of *N*-glycan modifications in mammalian cells is exemplified by the strong congenital disorders resulting from mutations in *N*-glycan processing. In plants, *N*-glycosylation has also been implicated in protein folding control, as complete absence of *N*-glycosylation is associated with induction of several protein folding stress responses (Varki, 1999). However, in contrast to animals, in plants changes in the structure of the mannose type *N*-glycans on glycoproteins imported into the ER seems to be less essential. For example, in the *alg3-2* mutant (Fig. 1), which has aberrant *N*-glycans for ER localized glycoproteins but normal *N*-glycans once the glycoprotein is transported to the Golgi, no clear growth phenotype was observed (Henquet et al., 2008).

### 3.3 Level and function of protein *N*-glycosylation for secreted proteins beyond ER not well understood

*N*-glycans on glycoproteins which are transported from the ER to the Golgi may be further modified into so-called complex *N*-glycans. Plants have specific types of complex *N*-glycan biosynthesis but the function of these modifications is not well understood. For instance, the complex glycan less (*cgl1-1*) mutant of *Arabidopsis*, which completely lacks complex *N*-glycans (Fig. 1), has no obvious growth phenotype and only displays a somewhat altered salt stress tolerance (Kang et al., 2008). The lack of a clear phenotype in *N*-glycosylation mutants hampers resolving the role of *N*-glycosylation in plants. Specific measurement of glycoproteins could help elucidate the role of this protein modification and quantitative glycoproteomics could for instance help establish whether there is a difference in lifetime of the same protein in *Arabidopsis* Col-0 (normal *N*-glycan) and in a glycosylation mutant background (e.g. in the *alg3-2* mutant background).



**Figure 1.** Illustration shows the difference in the *N*-glycosylation pathway of *Arabidopsis* Col-0 (WT), the complex glycan less (*cgl1-1*) mutant and the  $\alpha$ 1,3-mannosyltransferase (*alg3-2*) mutant.

### 3.4 Different strategies for *N*-glycoproteome enrichment

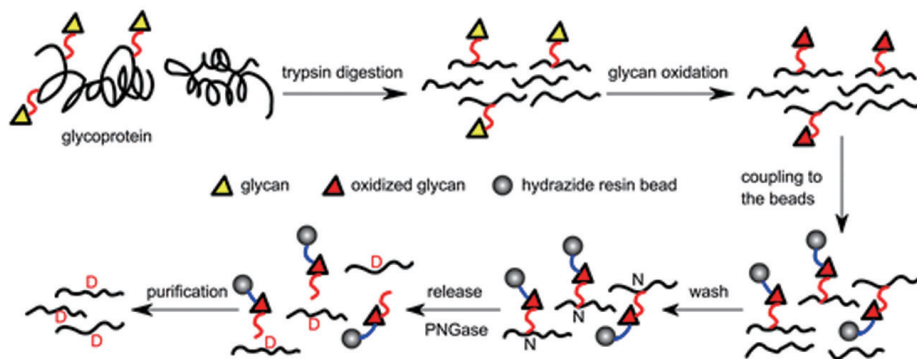
Due to the heterogeneity of the *N*-glycans on glycoproteins, glycopeptides are notoriously difficult to identify and sequence by mass spectrometry. Glycoproteomics tries to overcome these difficulties and is based on enrichment of the glycoprotein. Glycoprotein enrichment is achieved through lectin affinity capturing. Lectins are proteins that can bind carbohydrates and different lectins may show specificity for different sugar moieties. For instance, lectin ConA is mainly used for capturing proteins with mannose type of glycans. Wheat germ agglutinin (WGA) is often used for enrichment of glycoproteins with a terminal GlcNAc on *N*-glycans. Therefore, multiple lectins need to be used for isolation of the full *N*-glycan proteome (Kullolli et al., 2008; Zielinska et al., 2010; Zielinska et al., 2012).

Instead of enrichment of intact *N*-glycoproteins, enrichment of *N*-glycosylated peptides after trypsin digestion of the proteins may be used. Lectins may also be used for this (Xu et al., 2016). However, lectins mainly bind to mannose type *N*-glycans and not or weak to complex *N*-glycans, resulting in underrepresentation of glycoproteins with complex *N*-glycans. Glycopeptides may also be captured by covalently linking them to activated materials. For instance, after a glycan oxidation step, glycopeptides can be covalently coupled to hydrazide resin beads (Zhang et al., 2003; Tian et al., 2007; Zhang, 2007; Zhang and Cotter, 2008; Li et al., 2013; Sun et al., 2015; Yang et al., 2016). This method has no bias for different glycans. However, release of the captured glycopeptides highly depends on the efficiency and specificity of glycosidases which cleave the *N*-glycan

from the peptide and thus releases the captured peptide from the column for analysis by LC-MS. Similar to glycopeptide capturing to hydrazide beads, glycopeptides may be captured by boronic acid activated beads. Boronic acid can covalently react with cis-diols in the oligosaccharide and form five- or six-membered cyclic esters in an alkaline environment, while the cyclic esters dissociate when the environment changes to low pH. Boronic acid can be immobilized to monoliths, magnetic particles, mesoporous silica, polymer nanoparticles, gold nanoparticles or other solid-phase particles for glycopeptides enrichment (Wang et al., 2013). Also for this method, the release of the captured glycopeptides highly depends on the efficiency and specificity of glycosidases which release the deglycosylated peptide from the carrier material. Especially for plant *N*-glycoproteomics this step poses a problem because the plant specific complex *N*-glycans are not substrate for the glycosidase PNGase F, which is commonly used in these procedures and which for instance can cleave all types of *N*-glycans occurring in mammalian cells. Because the *N*-glycans are hydrophilic, a so called zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC) solid phase extraction (SPE) may also be used for enrichment of hydrophilic compounds such as glycan or glycopeptides (Pohlentz et al., 2016). In this case, glycopeptides are not covalently linked and eluted as intact glycopeptide, which requires special conditions for peptide/*N*-glycan identification.

### 3.5 Hydrazide beads glycopeptide affinity capture for glycoproteomics in plants

We chose to use the hydrazide beads affinity capturing method for glycopeptide enrichment to do glycoproteomics in *Arabidopsis* (Fig. 2). As special feature we made use of glycosylation mutants, such as complex glycan less (*cgl1-1*). The *cgl1-1* mutant was used to be able to study the full glycoproteome of *Arabidopsis*, which is not possible for *Arabidopsis* Col-0 because of complex *N*-glycans on glycoproteins that are in or go through the Golgi. Glycoproteins in the *cgl1-1* mutant do not have complex *N*-glycans and therefore potentially all glycopeptides can be released from hydrazide beads by PNGase F. Analysis of peptides released from *Arabidopsis* Col-0 and the *cgl1-1* mutant can give a first comprehensive inventory of glycopeptides/proteins in plants. Moreover, the identification of glycopeptides with enhanced signal in *cgl1-1* compared to *Arabidopsis* Col-0 potentially identifies glycopeptides with complex *N*-glycans, so identifies proteins that are present in or passed through the Golgi.



**Figure 2.** Schematic representation of an *N*-glycoproteomics workflow by using hydrazide resin beads as solid phase for affinity capture of *N*-glycoproteins.

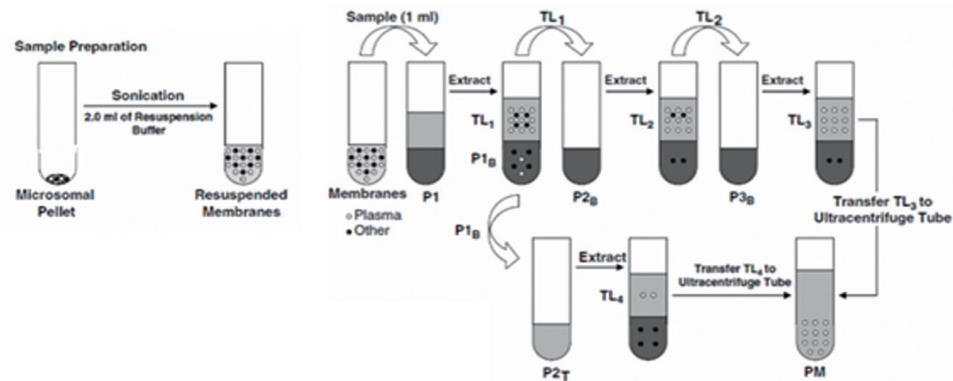
#### 4. Plasma membrane proteomics in plants

The plasma membrane in plant forms a semipermeable and selective barrier between the cell and its environment. In plants, the plasma membrane (PM) together with the cell wall (CW) forms the PM-CW continuum which separates the intracellular and extracellular space and regulates what enters and exits the cell. Proteins in the PM play an important role in many cellular processes such as signal transduction, exchange of metabolites, ion transport, endocytosis, and pathogen recognition (Shiu and Bleecker, 2001; Alberts et al., 2002; Uemura et al., 2004; Dodds and Rathjen, 2010). Proteins in the PM contain one or multiple transmembrane domains or they are anchored to the PM by a lipid anchor. The space for proteins in the PM is small compared to the volume of the cytosol of a cell. Therefore proteins in the PM membrane are less abundant than cytosolic proteins and studying the PM proteins therefore requires enrichment of the PM fraction of cells. For PM proteomics, membranes are isolated from cells and subsequently fractions are enriched for PMs. Because the lipid content of membranes of endogenous cellular compartments differs, membrane vesicles can be separated based on their different lipid composition by aqueous two-phase partitioning (Alexandersson et al., 2004) (Fig. 3).

##### 4.1 Plasma membrane proteomics application in plants

In addition to the glycoproteomics described above, in my thesis I used a second approach to enrich proteomics in plants and implemented and optimized a procedure for PM protein enrichment. The procedure of PM enrichment had been described for mammalian cells (Alexandersson et al., 2004; Fig. 3), but at the onset of

my thesis work had hardly been used for plants. The presence of the cell wall in plants poses additional problems for PM isolation and the procedure requires optimization for each plant species or tissue type (*Arabidopsis* leaf, potato leaf, tomato cells). In my thesis, PM proteomics is used to study different aspects of plant-pathogen interactions. In *Arabidopsis* I compare the plasma membrane composition of Col-0 (resistant to *P. brassicae*) and *lecrk-1.9* mutant and 35S-*ipiO1* transgenic plants (susceptible to *P. brassicae*). In addition, I study the PM proteome of tomato cell suspension cells with mock and real infection by *P. infestans* zoospores or zoospore supernatant which contains secreted effector proteins from the pathogen.



**Figure 3.** Scheme of plasma membrane enrichment from *Arabidopsis* by two phase partitioning as described by Alexandersson *et al.* (Alexandersson *et al.*, 2004).

## 5. PM proteomics in plant-pathogen interaction studies

Plants are affected by microbial pathogens such as fungi, bacteria and oomycetes. The interaction between plants and microbial pathogens can be viewed as two-way communication. On one side, when a pathogen has landed on a plant, the plant should recognize pathogen contact and initialize its defense system. On the other side, pathogens will try to manipulate plants for successful invasion by adapting/creating an environment within the host suitable for its growth and reproduction. The battle between plant and pathogen resulted in the evolution of genes that enable recognition and response between plant and pathogen. When a pathogen attacks a plant, a so called pathogen-associated molecular pattern (PAMP) can occur (Pel and Pieterse, 2013), which triggers the plant defense, also known as PAMP-triggered immunity (PTI) (Boller and Felix, 2009). In many cases, pathogens are able to suppress the PTI by secreting “effector” proteins into the plant. Subsequently, the plant defense system can recognize these effectors, which triggers additional plant defense responses, also known as effector-triggered immunity (ETI). Many responses to plant pathogens may only occur at the protein level and therefore are only detected by proteome analysis. Table 1 shows



that many proteomics studies have been done to study plant-pathogen interactions, but in none of these a PM enrichment procedure was used.

In studies of effector proteins from the oomycete tomato pathogen *P. infestans*, the *ipiO* gene was identified and shown to interact with a LecRK receptor kinase (Bouwmeester et al., 2011). Subsequent research using an *Arabidopsis* model with *Phytophthora brassicae* showed that both *Arabidopsis* plants with *ipiO* overexpression (35S-*ipiO1*) as well as the *Arabidopsis lecrk-I.9* mutant have a gain of susceptibility to *P. brassicae* while *Arabidopsis Col-0* is resistant to *P. brassicae*. In my thesis I use this system to study the PM proteome as I postulate that the similar phenotype of 35S-*ipiO1* and *lecrk-I.9* mutant is caused by similar changes in the PM proteome in these transgenic lines. In a more direct study of the plant pathogen interaction I apply PM proteomics to study the PM proteome changes in tomato MsK8 culture cells in response to either infection by *P. infestans* zoospores or treatment of supernatant from *P. infestans* zoospore cultures. PM enrichment is difficult, for example because of the presence of contaminating proteins. Another major problem is the limited link between identified peptides/proteins and quantitative data of peptide features. By using filtering procedures and alternatives to protein identification I try to solve these problems so I can still identify some potentially interesting candidate proteins which may play a role in the response to the plant pathogens *P. infestans* and *P. brassicae*.

**Table 1.** Summary of publications in *Arabidopsis* proteomics studies of plant-pathogen interactions, including pathogen name, plant material, protein separation technology, identification method, publication source and which post-translational modifications (PTMs) were introduced.

Pathogen/effector	Material	Method	Mass spectrometry	Journal	PTM in focus
fig22 (PTI)	Cell suspension culture	IMAC*	LC-MS/MS	Current Opinion in Plant Biology (Peck, 2003)	Phosphorylation
	Plasma membrane of suspension cultured cell	IMAC and iTRAQ	Q-TOF/TOF	The Plant Journal (Nuhse et al., 2007)	Phosphorylation
	Plasma membrane of suspension cell culture	SCX- Titanium oxide Enrichment	Nano HPLC-MS/MS	Molecular & Cellular proteomics (Benschop et al., 2007)	Phosphorylation
	Plasma membrane and detergent-resistant Membrane of suspension cell cultures	Isotope labelled gel free	LC-MS/MS	Journal of Biological Chemistry (Keinath et al., 2010)	
	Protein of total and cytoplasmic fractions from liquid cultured <i>Arabidopsis</i> (in dark)	C18-IMAC	LC-MS/MS	Journal of Proteome Research (Sun et al., 2013)	
	Cell suspension culture	iTRAQ	LC-MS/MS	New Phytologist (Hemsley et al., 2013)	Palmitoylation
	Cell suspension culture	C18-IMAC	LC-MS/MS	Journal of Proteomics (Rayapuram et al., 2014)	Phosphorylation

<i>Pseudomonas syringae</i> DC3000, DC3000AvrRpm1, DC3000hrpA-, DC3000AvrRpt2; (ETI)	Soluble fraction of leaf protein	2D gel	LC-MS/MS	Phytochemistry (Jones et al., 2004)	
	<i>Arabidopsis</i> leaves	phosphoprotein extraction kit 37101, Qiagen (affinity columns) ITRAQ	LC-MS/MS	Proteomics (Jones et al., 2006a)	Phosphorylation
	Cytosol, Chloroplast, Mitochondria proteins from leaves	2D gel	LC-MS/MS	Plant Physiology (Jones et al., 2006b)	
	Extracellular Proteins of <i>Arabidopsis</i> suspension cultures	ITRAQ	LC-MS/MS <sup>o</sup>	Molecular & Cellular Proteomics (Kaffarnik et al., 2009)	Phosphorylation
	S-nitrosylated protein from leaves	Biotin-labeling and pull down <sup>o</sup>	LTQ-MS/MS <sup>o</sup>	Acta Physioplant (Maldonado-Alconada et al., 2011)	Nitrosylation
	Total protein from wild-type and S-nitroglutathione reductase (GSNOR) knockout mutant	2D gel, DIGE	MALDI-TOF/TOF MS	Proteomics (Holzmeister et al., 2011)	
	<i>Arabidopsis</i> plasma membrane	8–16% precise protein gradient Gel (one-dimensional SDS-PAGE)	LC-MS/MS	Molecular & Cellular Proteomics (Elmore et al., 2012)	
Fungal elicitor (ETD)	Cell wall, extracellular matrix and soluble proteins of <i>Arabidopsis</i> suspension cultures	2D gel	MALDI-TOF MS, Q-TOF MS/MS	Proteomics (Ndimba et al., 2003)	Phosphorylation
	Total protein from <i>Arabidopsis</i> suspension cultures	2D gel	MALDI-TOF	Journal of Experimental Botany (Davies et al., 2006)	
	Total protein from leaves	2D gel	LCQ-Deca XP	Journal of Proteomics (Mukherjee et al., 2010)	
	Total leaf protein	2D gel, MOAC	MALDI-TOF-MS and/or nano LC-ESI-LIT-MS/MS	Journal of Proteomics (Huang et al., 2011)	Phosphorylation
	Total protein from leaves	2D gel, DIGE 18 cm (pH3-10)	MALDI-TOF/TOF	Proteome Science (Asano et al., 2012)	

## Abbreviations

PTI: Pathogen-associated molecular pattern (PAMP)-triggered immunity; ETI: Effector-triggered immunity; ETD: Effector-triggered defense

Flg22: A 22-amino acid sequence (flg22) of the conserved N-terminal part of flagellin

SDS-PAGE: Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

DIGE: Differential in-gel electrophoresis

MOAC: Metal oxide affinity chromatography

IMAC: Immobilized metal ion affinity chromatography

iTRAQ: Isobaric tags for relative and absolute quantitation

LC-MS: Liquid chromatograph and mass spectrometry

SCX: Strong cation exchange chromatography

MALDI-TOF: Matrix-assisted laser desorption/ionization and Time-of-Flight

ESI-LIT-MS: Electrospray ionization-linear ion trap mass spectrometry

LTQ-MS: Linear trap quadrupole mass spectrometry

LCQ: Liquid chromatography quadrupole

Q-TOF: Quadrupole Time-of-Flight

## 6. Scope of the thesis

In **Chapter 2** (*N*-glycoproteomics in plants: challenges and perspectives Song *et al.* (Song *et al.*, 2011) we give an overview of the current state of *N*-glycoproteomics in plants and perspectives of application. The difference between *N*-glycosylation in plant and mammalian is discussed and how this affects the possibilities of established methods for enrichment and analysis of glycoproteins in plants versus animals. One big difference between plants and mammalian organisms is that in mammalian cells the secreted proteome is estimated to constitute up to 50% of the overall proteins, while in plants the glycoproteome is only a small fraction of the total cellular protein pool. Enrichment of the glycoproteome in plants is therefore much more important than in mammalian cells. In plants, most intriguingly, mutations in glycan modification of glycoproteins do usually not have a clear effect on the plant phenotype, which is in sharp contrast to similar types of mutations in animals where they are often lethal or have high impact on development. The lack of a clear biological function, while at the same type glycosylation is strongly conserved in plants, makes the dedicated analysis of glycoproteins an important target for the future. Different glycoprotein or glycopeptide enrichment strategies are discussed and compared. Implications for the analysis of the biological role of these complex protein modifications are discussed.

In **Chapter 3** (*N*-glycan occupancy of *Arabidopsis* *N*-glycoproteins Song *et al.* (Song *et al.*, 2013) we describe the optimization of the protocol for *N*-glycoproteome enrichment in *Arabidopsis*. Using selective affinity capturing, 330 confirmed *N*-glycopeptides are identified which belong to 173 glycoproteins. Different methods for glycoprotein enrichment are described and compared. The limitations in the software for glycoprotein identification are discussed and the false positive discovery of protein identification estimated. Protein quantification is briefly introduced and applied. In application, the complex glycan less (*cgl1-1*) *Arabidopsis* mutant is used and results compared with Col-0 to generate an inventory of *N*-glycoproteins in *Arabidopsis*. As application of the results, the topology of several membrane proteins is verified, leading to identification of one membrane protein with wrongly annotated topology. Moreover, the results also reveal heterogeneity of *N*-glycans (both mannose type and complex type) on a single protein.

In **Chapter 4** we describe the protocol of plasma membrane proteomics in *Arabidopsis* using label-free comparative proteomics as a tool. In application, two *Arabidopsis* mutants are studied which show a gain of susceptibility under infection by *P. brassicae*. Plasma membrane proteins from those mutants show subtle protein composition differences on SDS-PAGE in comparison with *Arabidopsis* Col-0. With the help of high resolution 2D nano LC-MS, we are able to select the differential PM proteins quantitatively. Candidate genes are selected and validated. One *Arabidopsis* mutant *crk37* from the list of candidate genes showed a strong phenotype in gain of susceptibility under infection of *P. brassicae*, a *Phytophthora* species which is incompatible with *Arabidopsis* Col-0. This indicates that CRK37 might be involved in

plant-pathogen interaction.

In **Chapter 5** we describe the optimization of the protocol for plasma membrane enrichment from potato leaf and tomato MsK8 cultured cell. The in house build “SEDMAT” (single experiment data matching tool) a tool in the Galaxy toolshed is introduced, and shown to be instrumental in linking protein identification (DDA and MS<sup>E</sup> from ProteinLynx) with quantification (Progenesis). In this chapter, the difficulty of plasma membrane enrichment from potato is described and discussed. As an alternative, tomato MsK8 suspension cell are used for the study of plant-*P. infestans* interaction. Plasma membrane protein identification in treatments with zoospores and with supernatant of zoospores are compared quantitatively. Protein candidates are subsequently selected based on the peptide scoring and peptide/protein ranking from SEDMAT, the confirmation of peptide abundance profiles in Progenesis, specific domain prediction, the presence of a trans-membrane domain or GPI anchor and the presence of signal peptides. These eight protein candidates are manually confirmed and characterized by comparison with putative *Arabidopsis* homologs.

In **Chapter 6** I summarize and discuss the results from the different experimental chapters and give an overview of the major problems encountered and potential solutions for these problems. Finally I give perspectives for enrichment proteomics to study a number of different biological problems.

## References

- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P** (2002) Transport into the cell from the plasma membrane: Endocytosis.
- Alexandersson E, Saalbach G, Larsson C, Kjellbom P** (2004) *Arabidopsis* plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol* **45**: 1543–56
- Asano T, Kimura M, Nishiuchi T** (2012) The defense response in *Arabidopsis thaliana* against *Fusarium sporotrichioides*. *Proteome Sci.* doi: Artn 61Doi 10.1186/1477-5956-10-61
- Baggerman G, Vierstraete E, De Loof A, Schoofs L** (2005) Gel-based versus gel-free proteomics: a review. *Comb Chem High Throughput Screen* **8**: 669–77
- Benschop JJ, Mohammed S, O’Flaherty M, Heck AJR, Slijper M, Menke FLH** (2007) Quantitative phosphoproteomics of early elicitor signaling in *Arabidopsis*. *Mol Cell Proteomics* **6**: 1198–1214
- Van den Bergh G, Clerens S, Cnops L, Vandesande F, Arckens L** (2003) Fluorescent two-dimensional difference gel electrophoresis and mass spectrometry identify age-related protein expression differences for the primary visual cortex of kitten and adult cat. *J Neurochem* **85**: 193–205
- Boller T, Felix G** (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu Rev Plant Biol* **60**: 379–406
- Bouwmeester K, de Sain M, Weide R, Gouget A, Klammer S, Canut H, Govers F** (2011) The lectin receptor kinase LecRK-I.9 is a novel *Phytophthora* resistance component and a potential host target for a RXLR effector. *Plos Pathog.* doi: ARTN e100132710.1371/journal.ppat.1001327
- Davies DR, Bindschedler L V, Strickland TS, Bolwell GP** (2006) Production of reactive oxygen species in *Arabidopsis thaliana* cell suspension cultures in response to an elicitor from *Fusarium oxysporum*: implications for basal resistance. *J Exp Bot* **57**: 1817–1827
- Dodds PN, Rathjen JP** (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* **11**: 539–548
- Elmore JM, Liu J, Smith B, Phinney B, Coaker G** (2012) Quantitative proteomics reveals dynamic changes in the plasma membrane during *Arabidopsis* immune

signaling. *Mol Cell Proteomics*. doi: 10.1074/mcp.M111.014555

**Hemsley PA, Weimar T, Lilley KS, Dupree P, Grierson CS (2013)** A proteomic approach identifies many novel palmitoylated proteins in *Arabidopsis*. *New Phytol* **197**: 805–814

**Henquet M, Lehle L, Schreuder M, Rouwendal G, Molthoff J, Helsper J, Van Der Krol S, Bosch D (2008)** Identification of the gene encoding the  $\alpha$  1, 3-mannosyltransferase (ALG3) in *Arabidopsis* and characterization of downstream *N*-glycan processing. *Plant Cell Online* **20**: 1652

**Holzmeister C, Frohlich A, Sarioglu H, Bauer N, Durner J, Lindermayr C (2011)** Proteomic analysis of defense response of wildtype *Arabidopsis thaliana* and plants with impaired NO-homeostasis. *Proteomics* **11**: 1664–1683

**Huang C, Verrillo F, Renzone G, Arena S, Rocco M, Scaloni A, Marra M (2011)** Response to biotic and oxidative stress in *Arabidopsis thaliana*: Analysis of variably phosphorylated proteins. *J Proteomics* **74**: 1934–1949

**Jones AME, Bennett MH, Mansfield JW, Grant M (2006a)** Analysis of the defence phosphoproteome of *Arabidopsis thaliana* using differential mass tagging. *Proteomics* **6**: 4155–4165

**Jones AME, Thomas V, Bennett MH, Mansfield J, Grant M (2006b)** Modifications to the *Arabidopsis* defense proteome occur prior to significant transcriptional change in response to inoculation with *Pseudomonas syringae*. *Plant Physiol* **142**: 1603–1620

**Jones AME, Thomas V, Truman B, Lilley K, Mansfield J, Grant M (2004)** Specific changes in the *Arabidopsis* proteome in response to bacterial challenge: differentiating basal and R-gene mediated resistance. *Phytochemistry* **65**: 1805–1816

**Kaffarnik FAR, Jones AME, Rathjen JP, Peck SC (2009)** Effector proteins of the bacterial pathogen *Pseudomonas syringae* alter the extracellular proteome of the host plant, *Arabidopsis thaliana*. *Mol Cell Proteomics* **8**: 145–156

**Kang JS, Frank J, Kang CH, Kajiura H, Vikram M, Ueda A, Kim S, Bahk JD, Triplett B, Fujiyama K, et al (2008)** Salt tolerance of *Arabidopsis thaliana* requires maturation of *N*-glycosylated proteins in the Golgi apparatus (vol 105, pg 5933, 2008). *Proc Natl Acad Sci U S A* **105**: 7893

**Keinath NF, Kierszniowska S, Lorek J, Bourdais G, Kessler SA, Shimosato-Asano H, Grossniklaus U, Schulze WX, Robatzek S, Panstruga R (2010)** PAMP (pathogen-associated molecular pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity. *J Biol Chem* **285**:

39140–39149

**Kullolli M, Hancock WS, Hincapie M** (2008) Preparation of a high-performance multi-lectin affinity chromatography (HP-M-LAC) adsorbent for the analysis of human plasma glycoproteins. *J Sep Sci* **31**: 2733–2739

**Li QK, Gabrielson E, Askin F, Chan DW, Zhang H** (2013) Glycoproteomics using fluid-based specimens in the discovery of lung cancer protein biomarkers: Promise and challenge. *PROTEOMICS - Clin Appl* **7**: 55–69

**Maldonado-Alconada AM, Echevarria-Zomeno S, Lindermayr C, Redondo-Lopez I, Durner J, Jorrin-Novo J V** (2011) Proteomic analysis of *Arabidopsis* protein S-nitrosylation in response to inoculation with *Pseudomonas syringae*. *Acta Physiol Plant* **33**: 1493–1514

**Mukherjee AK, Carp M-JJ, Zuchman R, Ziv T, Horwitz BA, Gepstein S** (2010) Proteomics of the response of *Arabidopsis thaliana* to infection with *Alternaria brassicicola*. *J Proteomics* **73**: 709–720

**Ndimba BK, Chivasa S, Hamilton JM, Simon WJ, Slabas AR** (2003) Proteomic analysis of changes in the extracellular matrix of *Arabidopsis* cell suspension cultures induced by fungal elicitors. *Proteomics* **3**: 1047–1059

**Nuhse TS, Bottrill AR, Jones AME, Peck SC** (2007) Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. *Plant J* **51**: 931–940

**Ong SE, Pandey A** (2001) An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol Eng* **18**: 195–205

**Patel VJ, Thalassinos K, Slade SE, Connolly JB, Crombie A, Murrell JC, Scrivens JH** (2009) A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* **8**: 3752–3759

**Peck SC** (2003) Early phosphorylation events in biotic stress. *Curr Opin Plant Biol* **6**: 334–338

**Pel MJC, Pieterse CMJ** (2013) Microbial recognition and evasion of host immunity. *J Exp Bot* **64**: 1237–1248

**Pohlentz G, Marx K, Mormann M** (2016) Characterization of protein *N*-glycosylation by analysis of ZIC-HILIC-enriched intact proteolytic glycopeptides. *Methods Mol Biol* **1394**: 163–179

**Rayapuram N, Bonhomme L, Bigeard J, Haddadou K, Przybylski C, Hirt H, Pflieger D** (2014) Identification of novel PAMP-triggered phosphorylation and dephosphorylation events in *Arabidopsis thaliana* by quantitative phosphoproteomic analysis. *J Proteome Res* **13**: 2137–2151

**Shiu SH, Bleecker AB** (2001) Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE* **2001**: re22

**Song W, Henquet MGL, Mentink RA, van Dijk AJ, Cordewener JHG, Bosch D, America AHP, Van der Krol AR** (2011) *N*-glycoproteomics in plants: perspectives and challenges. *J Proteomics* **74**: 1463–1474

**Song W, Mentink RA, Henquet MGL, Cordewener JHG, Van Dijk ADJ, Bosch D, America AHP, Van der Krol AR** (2013) *N*-glycan occupancy of *Arabidopsis* *N*-glycoproteins. *J Proteomics* **93**: 343–355

**Sun HH, Fukao Y, Ishida S, Yamamoto H, Maekawa S, Fujiwara M, Sato T, Yamaguchi J** (2013) Proteomics analysis reveals a highly heterogeneous proteasome composition and the post-translational regulation of peptidase activity under pathogen signaling in plants. *J Proteome Res* **12**: 5084–5095

**Sun S, Shah P, Eshghi ST, Yang W, Trikannad N, Yang S, Chen L, Aiyetan P, Höti N, Zhang Z, et al** (2015) Comprehensive analysis of protein glycosylation by solid-phase extraction of *N*-linked glycans and glycosite-containing peptides. *Nat Biotechnol* **34**: 84–88

**Tannu NS, Hemby SE** (2006) Two-dimensional fluorescence difference gel electrophoresis for comparative proteomics profiling. *Nat Protoc* **1**: 1732–1742

**Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H** (2007) Solid-phase extraction of *N*-linked glycopeptides. *Nat Protoc* **2**: 334–339

**Uemura T, Ueda T, Ohniwa RL, Nakano A, Takeyasu K, Sato MH** (2004) Systematic analysis of SNARE molecules in *Arabidopsis*: dissection of the post-Golgi network in plant cells. *Cell Struct Funct* **29**: 49–65

**Varki A** (1999) *Essentials of glycobiology*. Cold Spring Harbor Laboratory Pr

**Wang XJ, Xia N, Liu L** (2013) Boronic acid-based approach for separation and immobilization of glycoproteins and its application in sensing. *Int J Mol Sci* **14**: 20890–20912

**Xu SL, Medzihradzky KF, Wang ZY, Burlingame AL, Chalkley RJ** (2016) *N*-glycopeptide profiling in *Arabidopsis* inflorescence. *Mol Cell Proteomics* **15**: 2048–



2054

**Yang W, Jackson B, Zhang H** (2016) Identification of glycoproteins associated with HIV latently infected cells using quantitative glycoproteomics. *Proteomics* **16**: 1872–1880

**Zhang H** (2007) Glycoproteomics using chemical immobilization. *Curr Protoc Protein Sci Chapter* **24**: Unit 24 3

**Zhang H, Cotter RJ** (2008) Glycoproteomics: New technology developments and applications provide renewed interest in glycoproteins. *Clin Proteomics* **4**: 1–4

**Zhang H, Li X-JJ, Martin DB, Aebersold R** (2003) Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**: 660–6

**Zielinska DF, Gnad F, Schropp K, Wisniewski JR, Mann M** (2012) Mapping *N*-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol Cell* **46**: 542–548

**Zielinska DF, Gnad F, Wisniewski JR, Mann M, Wiśniewski JR, Mann M** (2010) Precision mapping of an *in vivo* *N*-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**: 897–907



# Chapter 2

## ***N*-Glycoproteomics in Plants: perspectives and challenges**

Wei Song <sup>a,b</sup>, Maurice Henquet <sup>b</sup>, Remco Mentink <sup>b</sup>, Aalt Jan van Dijk <sup>b</sup>,  
Jan Cordewener <sup>b</sup>, Dirk Bosch <sup>b</sup>, Harro Bouwmeester <sup>a</sup>,  
Twan America <sup>b</sup>, Sander van der Krol <sup>a,\*</sup>

<sup>a</sup> Laboratory of Plant Physiology, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands

<sup>b</sup> Plant Research International, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB,  
Wageningen, The Netherlands

Published in: *J Proteomics*. 2011 Aug 12;74(8):1463-74

## Abstract

In eukaryotes, proteins that are secreted into the ER are mostly modified by *N*-glycans on consensus N-P!-S/T sites. The *N*-linked glycan subsequently undergoes varying degrees of processing by enzymes which are spatially distributed over the ER and the Golgi apparatus. The post-ER *N*-glycan processing to complex glycans differs between animals and plants, with consequences for *N*-glycan and glycopeptide isolation and characterization of plant glycoproteins. Here we describe some recent developments in plant glycoproteomics and illustrate how general and plant specific technologies may be used to address different important biological questions.

## Keywords

*N*-glycosylation, glycoproteomics, *N*-glycans, glycopeptides, cell specific proteomics, mass spectrometry

2

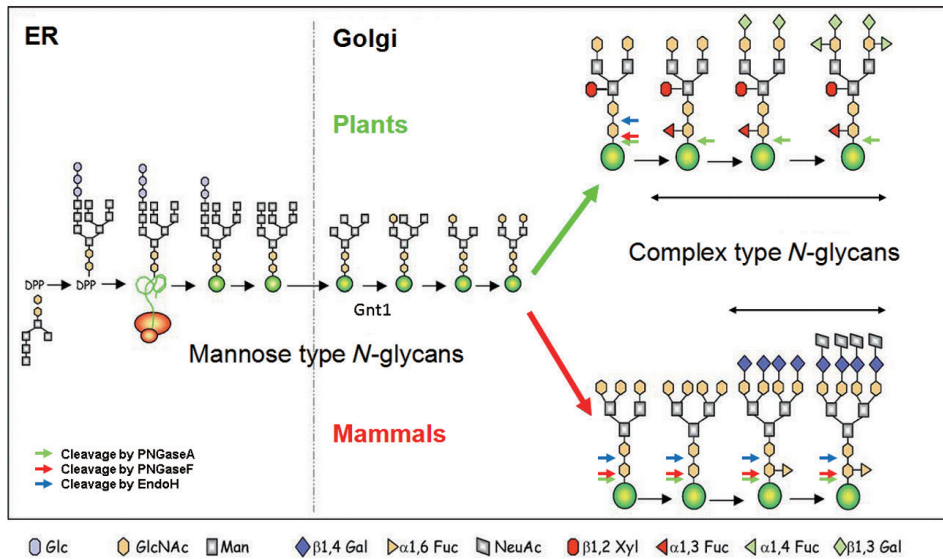
## Abbreviations

ER, Endoplasmatic Reticulum; *cgl*, complex glycan less; GlcNAc, *N*-acetylglucosamine; OST, oligosaccharyltransferase; ERAD, ER-associated degradation; PNGase, peptide-*N*-glycosidase; UPR: unfolded protein response; MALDI-TOF MS, matrix assisted laser desorption/ionization time-of-flight mass spectrometry; PAGE, polyacrylamide gel electrophoresis; ConA, Concanavalin A; WGA, wheat germ agglutinin; AIL, *Artocarpus integrifolia* lectin; PNA, peanut agglutinin; GalNAc, *N*-acetylgalactosamine; HRP, Horseradish peroxidase; LC-MS/MS, liquid chromatography tandem mass spectrometry; CE, capillary electrophoresis; LTP, lipid transfer protein; RBCS, ribulose biphosphate carboxylase; GO, Gene Ontology;

## N-glycosylation of proteins imported into ER

The process of *N*-glycosylation in plants was recently reviewed in the context of therapeutic glycoprotein production in plants by Gomord *et al.* (2010) (Gomord *et al.*, 2010). The process starts with dolichol lipid-linked glycan biosynthesis on the cytosolic side of the ER membrane. After flipping to the lumen side of the ER glycan synthesis continues to the full Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> structure (Trombetta, 2003). This fully assembled glycan is linked to the protein backbone via an amide bond (*N*-linked: *N*-glycosylation) of an asparagine residue in an Asn-X(P!)-Ser/Thr (N-P!-S/T) motif; where X can be any amino acid, except proline (Bause, 1983). Transfer from the lipid to the protein occurs during translation and transfer of secreted proteins into the ER at the Sec61 transmembrane channel complex by the oligosaccharyltransferase (OST) complex (Knauer and Lehle, 1999; Lehle *et al.*, 2006). In *Arabidopsis*, only three from the nine different components of the OST complex in yeast have been identified (Lerouxel *et al.*, 2005). Approximately 70% of all N-P!-S/T sites on secreted proteins are indeed glycosylated (Apweiler *et al.*, 1999; Petrescu *et al.*, 2004), suggesting additional requirements for *N*-glycosylation such as flanking peptide sequence, rate of the protein folding inside the ER lumen and capacity of the OST glycosylation machinery (Kasturi *et al.*, 1997; Mellquist *et al.*, 1998; Senger and Karim, 2005).

In the ER the terminal glucose and some mannoses are trimmed from the *N*-glycan (see next section). The lipid linked glycan biosynthesis and the first modification steps in ER and the Golgi apparatus are conserved between higher eukaryotic species, but further modifications in the Golgi apparatus may differ between species (Faye *et al.*, 2005). In the Golgi apparatus the first modification is the removal of one to four  $\alpha$ -1,2 mannose residues by  $\alpha$ -mannosidase I (Szumilo *et al.*, 1986). Up until this point, the *N*-linked glycans are termed high-mannose *N*-glycans structures. The first obligatory step in the synthesis of complex glycans is the addition of an *N*-acetylglucosamine moiety to the  $\alpha$ -1,3 mannose branch by *N*-acetylglucosaminyltransferase I (GnTI) (Fig. 1, (von Schaewen *et al.*, 1993)) which in plants subsequently allows for two independent *N*-glycan core modifications: an  $\alpha$ -1,3 linked fucose residue can be connected to the innermost GlcNAc residue and/or a  $\beta$ -1,2 xylose can be attached to the central mannose residue (Lerouge *et al.*, 1998). In mammals, the complex *N*-glycans have an  $\alpha$ -1,6 fucose instead of an  $\alpha$ -1,3 fucose linked to the innermost GlcNAc of the glycan core, while no  $\beta$ -1,2 xylose core modifications occur in mammals (Fig. 1) and glycans may be further modified by a  $\beta$ -1,4 galactose to which additional sialic acid residues may be attached (van Ree *et al.*, 2000; Fotisch and Vieths, 2001; Bencurova *et al.*, 2004). Since many of these reactions do not go to completion, a very heterogenous end-glycan profile may exist on a specific glycoprotein in a tissue. The difference in core glycan modifications between plants and animals is reviewed in detail by Gomord *et al.* (Gomord *et al.*, 2010) and of importance in glycoproteomics because the enzyme that is frequently used for deglycosylation of mammalian glycoproteins (PNGase F) does not recognize the plant glycan structure modified by an  $\alpha$ -1,3 linked fucose.



**Figure 1.** *N*-glycosylation in plants and animals. Differences in glycosylation pattern between plants and mammals appear in the Golgi apparatus. The most notable plant-specific *N*-glycan modifications are the  $\alpha$ -1,3 fucose attached to the first GlcNAc residue and the  $\beta$ -1,2 xylose attached to the central mannose residue. Mammalian *N*-glycans may be modified by  $\alpha$ -1,6 fucoses as opposed to the  $\alpha$ -1,3 of plants, as well as  $\beta$ -1,4 galactose residues, attached to terminal GlcNAc moieties. To these galactose residues sialic acid groups may be attached. (Adapted from (Faye et al., 2005)).

## Functions of *N*-glycans in the ER

In the ER, the mannose type *N*-glycan on the imported protein supposedly aids the protein folding process. Moreover, the attached glycan serves as a tag, which marks the quality of the protein folding state (Helenius and Aebi, 2004). The misfolded protein is retained in the ER or, when transported to the Golgi apparatus may be recycled back to the ER, till correctly folded (Spiro, 2004). However, a slow *N*-glycan trimming activity in the ER serves as a timer mechanism: when the *N*-glycan is trimmed too far, this indicates a problem with folding and the trimmed *N*-glycan with exposed hydrophobic domains, is recognized by ER associated degradation machinery (ERAD) (Hampton, 2002). The ERAD protein complex binds to the misfolded glycoprotein and facilitates export to the cytosol where the *N*-glycan is removed by PNGases and the protein is targeted for degradation by proteasomes (Yoshida, 2003; Helenius and Aebi, 2004; Spiro, 2004). Recently a set of *N*-glycan modifying mannosidase encoding genes (*Mns* 1-3) were identified in *Arabidopsis* (Liebminger et al., 2010). The triple mutant of the three *mns* genes forms only short roots, with radially swollen cortical cells and displays

alterations in development of cell walls and aerial plant parts, suggesting that several glycoproteins involved in cell wall formation are affected by the mutations. The effect of the mutations could be linked to the role of mannose trimming in glycoprotein folding quality control in the ER and ER-associated degradation of misfolded glycoproteins (Liebminger et al., 2010). Our knowledge of protein folding quality control and ERAD is mostly from yeast and mammalian systems, but recently the role of ERAD related processes in plants was reviewed by Liu and Howell (Liu and Howell, 2010). Because of the prominent role of *N*-glycans in the unfolded protein responses (UPR) and associated ERAD, glycoproteomics can play a key role in further elucidation of these processes in plants.

## Functions of *N*-glycans beyond the ER

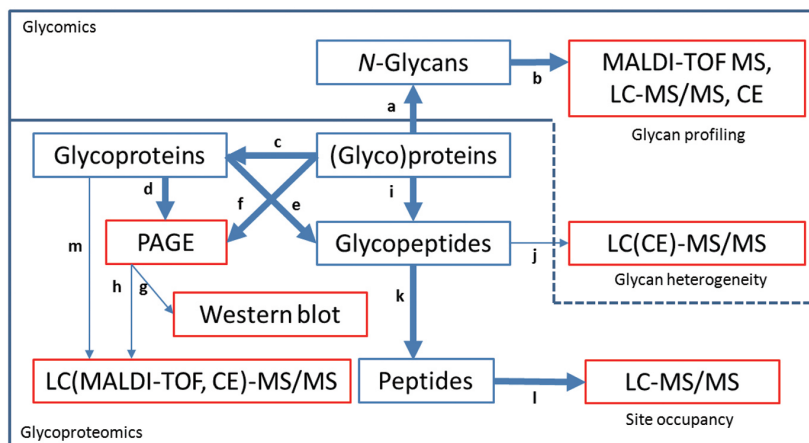
In mammalian cells the glycans on the glycoproteins have been proven to be involved in a wide range of biological functions such as receptor binding, cell signaling, protein folding, subcellular distribution and localization, protein stability, endocytosis, immune recognition, inflammation and pathogenicity (Klenk and Garten, 1994; Ihara et al., 1997; Takahashi et al., 2003; Lehle et al., 2006; Green et al., 2007; Chatterjee et al., 2008). In contrast, only very little is known about the role of *N*-glycan processing beyond the ER in plants. Only for isolated plant proteins the function of the processed *N*-glycan has been investigated in detail. *N*-glycans on plant glycoproteins have been shown to affect catalytic activity, thermostability and folding (Lige et al., 2001) or subcellular localization and secretion (Ceriotti et al., 1998). Recently it was found that *N*-glycans may also play a role in plant pathogen interactions (Pattison and Amtmann, 2009) and in functional pattern recognition receptors (Häweker et al., 2010). Some plant glycoproteins have a special *N*-glycan structure known as a “Lewis a” antigen (Fitchette-Laine et al., 1997; Melo et al., 1997). In mammals these glycan structures are located on glycoproteins at the cell surface and function in cell-cell recognition and adhesion processes (Lerouge et al., 1998). In plants these structures are also at the cell surface, but the function of this specific glycan structure in plants remains unknown.

Mutations in mammalian cells that affect *N*-glycan processing are linked to very severe phenotypes (Type II congenital disorders (Grunewald et al., 2002; Schollen et al., 2005)). In contrast, eliminating overall modifications to complex type *N*-glycans on plant glycoproteins has no or little effect on plant growth. For instance, the *Arabidopsis cgl1-1* mutant is unable to modify *N*-glycans on glycoproteins in the Golgi apparatus to complex glycans due to a non-functional GnTI gene (von Schaewen et al., 1993; Strasser et al., 2005). In the *cgl1-1* mutant all glycoproteins which normally have complex *N*-glycans now have mannose type *N*-glycans. However, plants grow normal and only display a higher sensitivity to salt stress (Kang et al., 2008). Also, when complex glycosylation was artificially suppressed by RNAi in tobacco and duckweed, the plants did not show an obvious growth phenotype (Cox et al., 2006; Strasser et al., 2008; Daskalova et al., 2010). There is also an altered *N*-glycan profile in the

*Arabidopsis* glycosylation mutant *alg3-2* (von Schaewen et al., 1993; Henquet et al., 2008). In the *alg3-2* mutant the *N*-glycans of proteins in the ER are replaced by aberrant mannose type *N*-glycans. Although it could be expected that this affects recognition of misfolded proteins in the ER, the *alg3-2* mutant shows no severe effect on plant growth (Henquet et al., 2008). Similar disorders in mammals result in severe phenotypes. The difference of the impact of mutated complex glycosylation in mammals and plants could be related to the fact that in mammalian cells this part of the glycoproteome has essential functions in protein-protein interactions in the circulatory system, while in plants there is only very limited long range transport of proteins.

### Strategies in *N*-glycoprotein analysis: *N*-Glycomics

There are two parts in the analysis of glycoproteins: one is related to the structure profiling of the *N*-glycan on the protein and the other is related to identification of the proteins which are modified by *N*-glycans. The strategies that can be used in glycoprotein analysis are summarized in Fig. 2. In this figure, path a-b and i-j describe the different options for obtaining information on the *N*-glycan structure based on MALDI-TOF MS, MALDI-TOF/TOF MS/MS, or LC-MS/MS analysis of either the cleaved *N*-glycans or the glycopeptides (Glycomics). Alternatively, capillary electrophoresis (CE) can be used for oligosaccharides profiling analysis. CE has the advantage of efficient separation and short run times allowing for HTP analysis of multiple samples (Taverna et al., 1992; Chen and Evangelista, 1998; Karamanos and Hjerpe, 1999; Kamoda et al., 2004; Balaguer and Neusüss, 2006). For path i-j information about the mass (amino acid sequence) of the peptide to which the glycan is attached is necessary for establishing the glycan structure. The other branches relate to different strategies in identifying glycoproteins or glycopeptides (Glycoproteomics). In Fig. 2 the thick arrows indicate strategies which potentially allow for bulk analysis of *N*-glycoproteins/*N*-glycans, the thin arrows indicate strategies which are usually limited to individually selected glycoproteins and thus are less compatible with an ‘omics’ type of approach.





**Figure 2.** Strategies in *N*-glycomics and *N*-glycoproteomics. The thick arrows indicate strategies which potentially allow for bulk analysis of *N*-glycoproteins/glycans, while the thin arrows indicate strategies which are usually limited to individually selected, purified *N*-glycoproteins. Path a: release of *N*-glycans by PNGase treatment; b: *N*-glycans analysis by MALDI-TOF MS, LC-MS/MS or Capillary Electrophoresis (CE); c: isolation of *N*-glycoproteins subfraction from whole (glyco)proteins pool; d: PAGE size separation of *N*-glycoprotein subfraction; e: tryptic digestion of *N*-glycoprotein subfraction; f: PAGE size separation of whole (glyco)proteins pool; g: specific *N*-glycoprotein detection by Western blot; h: analysis of tryptic digest of selected *N*-glycoprotein by LC-MS/MS, MALDI-TOF MS or CE-MS/MS after PAGE separation; i: *N*-glycopeptides isolation by selective affinity capture; j: *N*-glycan analysis of purified glycopeptides by LC-MS/MS or CE-MS/MS; k: release of peptides from *N*-glycopeptides by PNGase treatment; l: mapping the *N*-glyco-site occupancy of peptides obtained from path k by LC-MS/MS; m: analysis of intact *N*-glycoprotein by LC-MS/MS, MALDI-TOF MS or CE-MS/MS.

The total set of *N*-glycans present in the total glycoprotein pool of a plant tissue may be analyzed by MALDI-TOF MS, LC-MS/MS or CE after release of the *N*-glycans from the protein(s) by PNGase A treatment (Fig. 2 path a-b). For glycan identification, there is a free accessible tool called “*GlycoWorkbench*” available online, which can give glycan interpretation and annotation based on the glycan mass spectrometry data (Ceroni et al., 2008). An example of such a MALDI-TOF MS analysis of the *N*-glycan pool from two different plant tissues (leaf and seed) is shown in Table 1. The relative abundance of the different glycan structures clearly differs between leaves and seeds, and also changes upon aging in leaves. Alternatively, a specific glycoprotein may be isolated from a plant tissue extract, as shown in Table 1. Structures can be released with PNGase A for analysis with different approaches (Fig. 2, path a-b), or the (purified) protein can be digested with trypsin and analyzed by LC-MS/MS, MALDI-TOF MS or CE-MS/MS to determine the mass of the glycopeptide (Fig. 2, path h). A bioinformatics tool called “*ProTerNyc*” can be used to calculate all putative *N*-glycopeptides and identify them within the mass data obtained from mass spectrometry (Albenne et al., 2009). In many cases, a single glycoprotein shows a range of different glycan structures (Triguero et al., 2005; Lim et al., 2008), even when there is only a single *N*-glycosylation site present (Henquet et al., 2010). This most likely is the result of different subcellular resident times in ER and/or cis-, medial-, or trans-Golgi, resulting in different *N*-glycan processing intermediates. The *N*-glycan composition of glycoproteins isolated from different tissues (leaf, seedling, and seed) also shows distinct differences, which may be a reflection of different glycoprotein composition in these tissues (data not show). Remarkably, there are large differences in the ratio of ‘mannose’ and complex type glycans between different plant species. It could be that this is related to differences in ER:Golgi/plasma membrane ratio in large (tobacco) and small (*Arabidopsis*) plant cells. Experiments with the inhibitor of *N*-glycosylation tunicamycin shows that glycoproteins have a turnover varying from hours to days (Häweker et al., 2010).

However, whether the changes in the *N*-glycan composition in leaves upon aging are a reflection of different glycoproteins being expressed or reflect a maturation of *N*-glycans on the existing glycoproteins is presently not known.

**Table 1.** *N*-glycans profile in different plants, plant tissues and of isolated glycoprotein (numbers are % of the detected total *N*-glycan pool); Data was compiled from references (Elbers et al., 2001; Henquet et al., 2008; Henquet et al., 2010). Secr Ab: Protein from transgene expressing antibody with secretion signal; ER Ab: Protein from transgene expressing antibody with ER retention signal; n.d.: not detectable.

	Tobacco			<i>A. thaliana</i>		<i>A. thaliana</i>	<i>A. thaliana</i>
	Young	Old	Secr Ab	Young	Old	Seed	Seed
						Total	ER Ab
<b>'Mannose-type' <i>N</i>-glycans</b>							
Man3GlcNAc2						n.d.	
Man4GlcNAc2						n.d.	
Man5GlcNAc2	3.0	4.0	4.0	26.7	31.7	21.3	
Man6GlcNAc2	5.0	5.0		4.7	7.5	9.3	
Man7GlcNAc2	3.0	2.0	4.0	4.5	5.7	4.7	26.6
Man8GlcNAc2		2.0	4.0	5.9	5.2	8.9	73.4
Man9GlcNAc2			1.0	2.2	1.8	n.d.	
Man3XylGlcNAc2	6.0	10.0	4.0	8.3	5.0	4.8	
Man4XylGlcNAc2						2.1	
GlcNAcMan3GlcNAc2			4.0				
GlcNAcMan3XylGlcNAc2	3.0	3.0	3.0	8.5	7.7	n.d.	
GlcNAc2Man3GlcNAc2			10.0				
GlcNAc2Man3XylGlcNAc2			4.0				
<b>'Complex-type' <i>N</i>-glycans</b>							
Man3XylFucGlcNAc2	29.0	39.0	14.0	15.3	11.3	41.2	
Man3FucGlcNAc2		4.0	1.0				
GlcNAcMan3FucGlcNAc2			2.0				
GlcNAcMan3XylFucGlcNAc2	42.0	31.0	13.0	8.7	9.1	3.5	
GlcNAc2Man3FucGlcNAc2			5.0				
GlcNAc2Man3XylFucGlcNAc2	8.0		23.0	15.1	15.0	4.3	
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

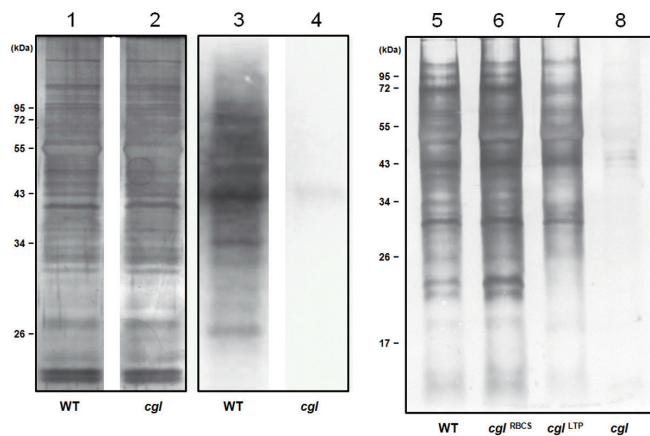
## Strategies in *N*-glycoprotein analysis: *N*-Glycoproteomics

### *Gel based glycoprotein analysis*

Most information of individual plant glycoproteins is coming from PAGE based glycoprotein analysis (Fig. 2, path d, g, h). To test whether a protein is glycosylated it can be treated with PNGase or Endo H, which (under denaturing conditions) can remove the glycan(s) from the glycoprotein. Treated and untreated protein fractions can be analyzed on Western blot after size separation on PAGE and difference in mobility of a specific glycoprotein can be detected if specific antibodies against this protein are

available. This approach only tells whether a protein is glycosylated. It provides no information on the attached *N*-glycan structure beyond the level of mannose type or complex type *N*-glycan (by comparing sensitivity to PNGase A, PNGase F and Endo H: Fig. 1), nor information on glycosylation site occupancy. Moreover, it is limited to single protein analysis, so not applicable for ‘omics’ approaches in glycoprotein analysis.

*N*-glycan modifications on proteins within a protein extract may be mapped by PAGE and Western blotting, using glycan specific lectins, antibodies or fluorescent staining techniques (Varki, 1999). Lectins are a group of proteins capable of binding various forms of carbohydrates with different specificities. The lectin ConA is often used to detect glycoproteins. It binds to mannose type *N*-glycans, but not to plant complex-type *N*-glycans (Varki, 1999). Some lectins may also bind to part of the *O*-glycoproteome. For instance, wheat germ agglutinin (WGA) binds to some GlcNAc- or sialic acid-containing *N*-glycans, but also to clusters of  $\alpha$ -GalNAc *O*-glycans in mucin-like glycoprotein(s) of *C.elegans* (Natsuka et al., 2005). The selective staining of part of the glycoproteome may make it easier to detect subtle changes in a total protein extract when glycoproteins are specifically affected. The anti-HRP polyclonal antibody directed against the Horseradish Peroxidase (HRP) recognizes the fucose and xylose residues on complex type *N*-glycans. Anti-HRP is therefore often used to detect proteins with complex *N*-glycans on Western blots. Fig. 3 shows an example of detection with anti-HRP of the complex glycoproteome in leaves of *Arabidopsis* Col-0 plants (Fig. 3, lanes 1, 3 and 5) and leaves of the glycosylation mutant *cgl1-1* (von Schaewen et al., 1993; Strasser et al., 2005), which is blocked in the formation of complex *N*-glycans (Fig. 3, lanes 2, 4 and 8). Glycoproteins may also be detected on gel by covalently linking a fluorescent molecule to periodate-oxidized carbohydrate groups (e.g. Pro-Q-Emerald, Molecular Probes™). This staining method enables detection of even low abundant glycoproteins on gel, but does not distinguish between different types of *N*-glycans or between *N*- and *O*-glycans (Zhang et al., 2008).



**Figure 3.** Complex *N*-glycan tagging in epidermal and mesophyll cells of *Arabidopsis* leaves. Lanes 1-2: SDS-PAGE of total protein extracts from leaves of *Arabidopsis* Col-0 (lane 1) and *cgl1-1* mutant (lane 2) as detected by silver staining. Lanes 3-4: Western blot of identical Col-0 (lane 3) and *cgl1-1* (lane 4) fractions and staining of the complex glycoproteome using anti-HRP; Lanes 5-8: Western blot analysis of the complex glycoproteome (probe anti-HRP) of a leaf tissue of *Arabidopsis* Col-0 (lane 5), leaf tissue of *cgl1-1* mutant complemented with RBCS-GnTI (lane 6) or LTP1-GnTI (lane 7), leaf tissue of *cgl1-1* mutant (lane 8). For technical details see supplement.

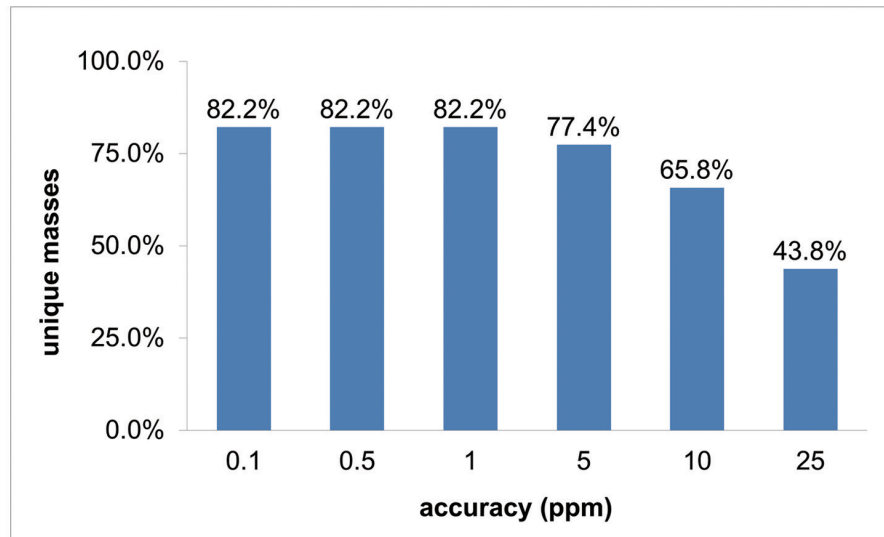
### ***Complexity reduction by selective glyco-proteomics***

Identification of a complete proteome of a tissue by LC-MS/MS is mostly hampered by the high complexity of the peptide mixture that is obtained after digestion with trypsin. This can be solved in part by fractionation of the peptide mixture using multidimensional LC-MS/MS. However, this makes comparison of numerous protein samples extra difficult. Proteomics by direct LC-MS/MS is therefore more feasible with selected proteome fractions with reduced peptide complexity, such as from isolated membranes or isolated glycoproteome fractions using different types of affinity capturing. For instance, only 3% of the tryptic fragments of the human proteome contain *N*-glycosylation sites; yet these peptides with an *N*-glycan represent the majority of extracellular proteins (over 70%) (Zhang et al., 2006). The putative *N*-glycoproteome of a plant species can be predicted from its genome sequence information. Of the 32,825 proteins that have been annotated in the *Arabidopsis* genome (TAIR9, www.arabidopsis.org), 5879 proteins have a secretion signal peptide (SP) at the *N*-terminus predicted by TargetP, indicating that these are putative secreted proteins and potential candidates for *N*-glycosylation. Of this putative secreted proteome, 3597 proteins actually contain one or more N-P!-S/T consensus glycosylation sites. Because not every N-P!-S/T consensus site on secreted proteins will be occupied by an *N*-glycan, the predicted 3597 proteins constitute the maximum estimate of the *N*-glycoproteome. Fig. 5 compares the annotated biological functions of the entire *Arabidopsis* proteome with that of the predicted secreted *N*-glycoproteome (see also Table 2). The pie-charts show that the putative secreted *N*-glycoproteome is enriched for hydrolase, kinase and receptor signaling proteins. Especially the receptor kinases with leucine rich repeats (which have been implemented in plant pathogen defense reactions (Kedzierski et al., 2004; Shanmugam, 2005; Chinchilla et al., 2007; Dodds and Rathjen, 2010)), have multiple (up to 32) consensus *N*-glycosylation sites, suggesting that isolation and analysis of the *N*-glycoproteome may be of assistance in studying plant pathogen interactions.

**Table 2.** Gene Ontology annotations for the whole proteome (All #) and putative N-glycoproteome (Glyco #) of *Arabidopsis*. Enrichment shows the increase (fold change) of proteins with specific molecular function in putative N-glycoproteome compared to whole proteome.

GO molecular function	All #	%	Glyco #	%	Enrichment
Unknown molecular functions	10,893	24.45%	744	18.60%	
Other binding	5345	12.00%	473	11.83%	
Other enzyme activity	4359	9.79%	389	9.73%	
Transferase activity	3764	8.45%	332	8.30%	
Hydrolase activity	3273	7.35%	805	20.13%	2.72
DNA or RNA binding	2910	6.53%	30	0.75%	
Kinase activity	2597	5.83%	439	10.98%	1.87
Nucleotide binding	2456	5.51%	196	4.90%	
Transporter activity	1864	4.18%	122	3.05%	
Transcription factor activity	1679	3.77%	18	0.45%	
Nucleic acid binding	1610	3.61%	16	0.40%	
Protein binding	1518	3.41%	233	5.83%	1.71
Other molecular functions	1502	3.37%	133	3.33%	
Structural molecule activity	536	1.20%	22	0.55%	
Receptor binding or activity	239	0.54%	48	1.20%	2.22
<b>Total</b>	<b>44,545</b>	<b>100%</b>	<b>4000</b>	<b>100%</b>	

After in silico trypsin digestion, the proteins of the predicted N-glycoproteome of *Arabidopsis* render 12399 peptide fragments which contain one or more N-P!-S/T sites. A tryptic digest of the full *Arabidopsis* proteome results in 1665744 (predicted) fragments and a selection of only the glycopeptides would give approximately a 100-fold reduction in peptide complexity, not regarding the additional complexity formed by different glycan structures. When glycopeptides from *Arabidopsis* can be selectively isolated and subsequently deglycosylated, the measured mass of the peptides (without the PTM) can be matched to a database of predicted (glyco)peptide masses (based on the tryptic peptide masses from all secreted proteins with a consensus N-P!-S/T site). When the mass alone would be measured with an accuracy of ~1ppm, a maximum of 82% of the peptides can be matched unequivocally to a single glycoprotein (see Fig. 4). For the remainder 18% of predicted (glyco)peptides additional information, like MS/MS sequence information is required for further identification, or protein isoforms could not be discriminated at (glyco)peptide sequence level.

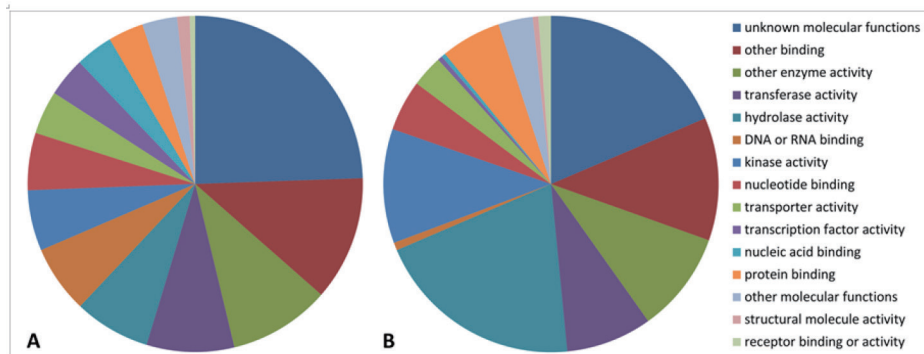


**Figure 4.** Percentage of unique putative glycopeptide masses per mass accuracy (ppm). Comparison of measured Mass of putative glycopeptides and a database of predicted glycopeptide Masses, based on the *Arabidopsis* sequence data from TAIR9. The percentage indicates how many of the masses can be uniquely assigned to a glycoprotein at the given accuracy of Mass measurement.

### ***Lectin based selection of glycoproteins and glycopeptides***

Antibodies or lectins can not only be used for detection of glycoproteins, but may also be used for isolation of glycoprotein pools. The isolated glycoprotein subfraction may then be size fractionated on PAGE, 2D-PAGE or Size-Exclusion Chromatography (SEC), and protein bands of interest may be isolated for analysis by LC-MS/MS, MALDI-TOF MS or CE-MS/MS (Fig. 2, path c-d-h). Alternatively, the glycoproteins can be identified directly by those different approaches without PAGE separation (Fig. 2, path m). Lectins are useful in selecting for groups of proteins carrying a specific type of glycan, but do not tolerate strong denaturing conditions for the interaction with glycans. Therefore, affinity capturing of intact glycoproteins may miss those proteins for which the glycan moiety is masked by the protein structure. The selectivity of lectins for certain glycans that affects detection on Western blots (see above) also affects the selection of the glycoprotein pool from a total protein extract. The lectin ConA is often used to capture proteins with mannose type glycans as ConA does not bind to complex type *N*-glycans (Varki, 1999). WGA can be used to isolate glycoproteins with a terminal GlcNAc on *N*-glycans, but glycoproteins modified by *O*-glycosylation may also bind to WGA (Jackson and Hardham, 1998). The lectins AIL and PNA can be used to capture Gal-containing *N*- and *O*-glycoproteins. AIL is assumed to recognize mostly  $\alpha$ -Gal, while PNA recognizes  $\alpha$ -Gal and  $\beta$ -Gal (Banerjee et al., 1994; Arockia Jeyaprakash et al., 2005). Recently, different lectin affinity purification methods were combined

for fractionation and identification of cell wall associated glycoproteins (Zhang et al., 2010). After tryptic digestion of the isolated protein fractions, peptides were identified by either LC-MS/MS or MALDI-TOF MS, resulting in the identification of 127 glycoproteins (Zhang et al., 2010). Of these 127 glycoproteins, 101 proteins were previously identified in a cell wall proteome analysis of 11-day-old etiolated hypocotyls of *Arabidopsis thaliana* (Irshad et al., 2008).



**Figure 5.** Ontology of *Arabidopsis* proteome and putative secreted *N*-glycoproteome. (A) Gene Ontology for molecular function of the full *Arabidopsis* proteome from TAIR9; (B) Gene Ontology for molecular function of the putative secreted *N*-glycoproteome of *Arabidopsis* (Berardini et al., 2004) based on signal peptides (SignalP: [www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/)) and N-P!-S/T consensus site.

### ***Hydrazide based selection of glycoproteins and glycopeptides***

Glycopeptides may also be captured by linking them through the glycan moiety to hydrazide resin (Fig. 2, path i-k-l). For this, the glycans of a tryptic digest of the isolated protein sample are first oxidized to create functional aldehyde groups and subsequently covalently linked to hydrazide resin (Tian et al., 2007). After washing away the non-linked peptides, the glycopeptides that are linked to the resin with *N*-glycans can be released by PNGase treatment, which results in a cleavage between the *N*-glycan and the peptide backbone. Although both PNGase A and PNGase F work for glycoproteins from mammals (Fig. 1), PNGase F is used mostly because it is (commercially) available at much higher concentration than PNGase A and can therefore be used in a much more robust way. PNGase F only cleaves ‘mannose type’ *N*-glycans, whereas peptides linked by complex *N*-glycans (with a glycan core modified by  $\alpha$ -1,3 fucose) will remain linked to the resin. To release the complex *N*-glycan containing peptides from the resin, the less efficient PNGase A has to be used. For quantitative analysis of *N*-linked glycopeptides and comparison between samples we use hydrazide based isolation of glycopeptides,

followed by label free quantitative LC-MS/MS. Alternatively, differential stable isotope labeling of samples can be used (Tian et al., 2007). In our hands, even after extensive washing, also non-glycopeptides are released from the hydrazide resin upon PNGase F treatment. The presence of an N-P!-S/T glycosylation consensus site, in combination with a 1 Da mass shift (N to D conversion) of the peptide, can be used to distinguish between real glycopeptides and contaminating non-glycopeptides.

## Applications of *N*-glycoproteome analysis

### *N*-glycan type mapping

*N*-glycan type mapping can give information on the subcellular location or transport of a glycoprotein. Because the different *N*-glycan modifying enzymes are spatially distributed over the ER and cis-, medial- or trans-Golgi, the *N*-glycan structure itself contains information about the subcellular location or transport pathway of a glycoprotein. For instance, when a glycoprotein only contains mannose type *N*-glycans it is indicative of ER localization or of direct transport to subcellular compartments that by-pass the Golgi apparatus. When a protein contains one or more complex *N*-glycans, it is an indication that the protein is resident in or has passed through the Golgi apparatus (e.g. many of the plasma membrane proteins are modified by complex *N*-glycans). Evidence of glycosylation of plant proteins is mostly obtained through gel-shift based assays in which the mobility of the target protein before and after PNGase treatment is compared. When treatment of the protein with either PNGase F or endoglucanase H (both do not cleave complex *N*-glycans of which the core is modified by fucose) or PNGase A (cleaves all *N*-glycans) results in a gel mobility shift, this can be taken as proof for the presence of glycan and a rough indication of the glycan type may be obtained. As mentioned above, such information is indicative for subcellular location or protein transport pathway. However, on the intact protein not all attached *N*-glycans may be accessible for deglycosylation. The use of some reducing and/or denaturing agent may be required to make buried glycans accessible for enzymatic action. Furthermore, gel-shift based assays do not provide information on the site of *N*-glycan attachment and only limited information on the number and type of *N*-glycans that are attached to the protein.

### *N*-glycan site occupancy mapping for transmembrane topology verification

For secreted proteins with multiple membrane spanning domains, the position of an *N*-glycan can give information on the membrane spanning domain topology, because *N*-glycan attachment is at the lumen site of the ER (Trombetta, 2003; Helenius and Aebi, 2004; Moremen and Molinari, 2006). Indeed, the topology of membrane proteins is often verified by introducing a consensus *N*-glycosylation sites in specific loops, located between the transmembrane spanning domains, and monitor the occurrence of *N*-glycosylation when expressed in the cell (Wo and Oswald, 1994; Kato et al., 2001).



*N*-glycoproteomics aimed at mapping *N*-glycan site occupancy of glycoproteins may therefore help to validate the correct topology of membrane proteins with one or more transmembrane spanning domains. Knowing the correct topology of membrane proteins is important for assigning the correct location of enzymatic activity (ER-lumen, cytosol, apoplast etc) (Kimura et al., 1999; Mühlenhoff et al., 2001; Yusa et al., 2005; Uemura et al., 2006; Yusa et al., 2006).

### ***N*-glycan site occupancy mapping for prediction tools**

Accurate prediction of *N*-glycan site occupancy on secreted proteins is of use for development of plants as production platform of highly value pharmaceutical glycoproteins like antibodies or vaccines with uniform glycosylation patterns. The efficiency with which the glycan is transferred onto the consensus sequence is determined by (1) the serine or threonine in the consensus sequence N-P!-S/T; (2) the amino acid x in the consensus sequence; (3) the location of the N-P!-S/T motif within the polypeptide chain and its proximity to the C terminus; (4) the amino acids flanking the N-P!-S/T sequence; and (5) the polypeptide structure surrounding the consensus sequence (Jones et al., 2005). A prediction tool was developed for variable Site-Occupancy Classification of N-Linked Glycosylation Using Artificial Neural Networks, based on only 48 N-P!-S/T sites from individual proteins of which site occupancy had been retrieved from literature (Senger and Karim, 2005). For such prediction tools, true glycoproteomics would greatly increase the input that is used for entraining the neural network and could thus vastly improve the accuracy of such prediction tools.

2

## **Unique opportunities for *N*-glycoproteomics in plants**

### ***Full N-glycoproteome mapping***

The difference between high mannose versus complex-type glycans can be used to discriminate proteins that have passed the Golgi apparatus or not (yet). For the study of changes in the glycoproteome of the ER and ER derived compartments that bypass protein transport through the Golgi apparatus, the selectiveness of PNGase F is actually an advantage when glycopeptides are captured through their glycan and can only be released by PNGase F treatment. In protein extracts from *Arabidopsis* Col-0 plants, the expensive PNGase A enzyme (which can cleave all *N*-glycans) is inefficient in releasing *N*-glycopeptides (our observation). This limits an integral analysis of all glycopeptides captured through their *N*-glycans. Because PNGase F can only cleave mannose-type *N*-glycans it will mostly release glycopeptides from ER compartments in Col-0 plant protein extracts. For an inventory of the full *N*-glycoproteome, mutant plants in which *N*-glycan modification to complex glycan has been blocked can be used (von Schaeuwen et al., 1993). In this genetic background glycoproteins only contain mannose type *N*-glycans and therefore, all glycopeptides can be released by PNGase F from their *N*-glycans. Alternatively, for plant species for which such mutants are not available,

different knockout strategies can be used to block complex glycosylation (Cox et al., 2006; Strasser et al., 2008; Daskalova et al., 2010; Gomord et al., 2010). The isolated glycopeptides can be analyzed by label free quantitative LC-MS/MS comparing mass peak intensities between (groups of) samples. Alternatively, glycopeptides may be labeled during the PNGase treatment. When the deglycosylation reaction is performed in  $^{18}\text{O}$ -water,  $^{18}\text{O}$  is incorporated at the *N*-glycosylation site and samples treated with PNGase F in normal water and  $^{18}\text{O}$ -water can be easily compared quantitatively (Shakey et al., 2010).

### ***Cell specific tagging of the complex N-glycoproteome***

As a tool to compare the glycoprotein pools of different cell types in plants, we used the *Arabidopsis* complex-glycan-less (*cgl1-1*) mutant (von Schaewen et al., 1993; Strasser et al., 2005). By restoring *N*-glycan processing in this mutant by expressing the *Arabidopsis* Col-0 GnTI gene under control of a cell specific promoters (LTP1 for epidermal expression and RBCS for mesophyll specific expression), the biosynthesis of complex *N*-glycans is only restored in epidermal cells or mesophyll cells, respectively. Only in the cells with Col-0 GnTI activity, glycoproteins carry complex type *N*-glycans with the xylose and fucose epitopes, which allows the study, characterization and isolation of the ‘complex-type’ glycoproteome of the specified cells. Fig. 3 shows as example the difference in complex glycoproteome detected in *Arabidopsis* leaves. Lanes 1 and 2 show the equal (silver) staining of Col-0 and *cgl1-1* protein that was loaded, while lanes 3 and 4 show the Western blot detection of proteins with complex *N*-glycans using anti-HRP. Lane 4 shows that in the *cgl1-1* mutant background virtually no proteins are detected by anti-HRP. Lanes 5-8 show an example of the cell type specific tagging with complex *N*-glycans in epidermal and mesophyll cells of *Arabidopsis* leaf tissue. The detected glycoprotein pattern in lane 6 and 7 shows there is quite some overlap of the complex glycoproteins in the two cell types. Even though Western blots only allow for the detection of abundant glycoproteins, already for each cell type uniquely tagged proteins are detected. We are currently optimizing the protocols for isolating the peptides that are tagged with complex *N*-glycans from *Arabidopsis* leaf tissue for further protein identification.

## **Bottlenecks in glycoproteomics**

### ***Multiple glycoprotein products***

The complex chemical structure, the different chemical properties of the peptide and glycan moieties and relative low abundance of most glycoproteins in plants, make analysis of glycoproteins or glycopeptides a difficult task. Integral approaches to glycoprotein analysis are further complicated by the dynamics of the glycosylation process itself, resulting in variable glycan structures on the same glycoprotein. The already relatively low signal of a single glycopeptide with different attached glycan

moieties may therefore be distributed over multiple peaks at different retention times using liquid chromatography, impeding detection or efficient identification in complex mixtures.

### ***Peptide identification***

The 1 Da (in water) or 2 Da (in  $^{18}\text{O}$ -water) mass shift from the PNGase treatment, being diagnostic for the glycosylation site, could alternatively originate from a deamidation of asparagine. This needs to be checked by control samples (eg. detecting the “normal” level of deamidation). Furthermore, because many glycoproteins contain only one or few glycosylation sites, linking the isolated glycopeptide mass to the presence and identity of a specific protein is no longer supported by multiple peptide masses. While common guidelines for publication require the identification of proteins by multiple peptides (Alves et al., 2010; Nesvizhskii, 2010), this is not feasible in many cases for affinity selected single peptides. We argue that if good quality MS/MS information is provided per peptide, these “one-hit” identifications are correct and valuable.

### ***Non-robust PNGase A***

General N-glycoproteomics research for plant glycoproteins would profit a lot with isolation of the PNGase A gene and biotechnological production of this more versatile PNGase. It would enable more efficient release of glycopeptides independent of glycosylation mutant background. We mentioned the unique possibility of performing cell specific glycan tagging in plants, by cell specific complementation of the *cgl1-1* mutant. Analysis of the proteins of the targeted cell type that are tagged by complex glycans, using the hydrazide capturing method, requires removal of ‘mannose’ type N-glycopeptides with PNGase F and subsequent release with PNGase A. This procedure is still not effective enough and we are currently testing whether isolation of the complex glycopeptides may be enriched by prior removal of all mannose type N-glycans from the peptide mix before binding to the hydrazide resin.

## **Summary and Perspectives**

With the new developments at the level of software and state of the art analytical equipment, it becomes feasible to dive deeply inside of the functional importance of N-glycosylation in plants. Recent newly developed MS technologies like Electron Capture Dissociation (ECD) and Electron Transfer Dissociation (ETD) which results in peptide backbone fragmentation while keeping the post-translational modification intact may further improve characterization of glycopeptides (Syka et al., 2004; Ahn et al., 2007; Jackson et al., 2009). The glycoproteome, and especially the part of the glycoproteome in the plasma membrane, which are (mostly) modified by complex N-glycans, is highly enriched in components that interact with extracellular signals during plant development, but also extracellular signals emanating from invading

pathogens. Glycoproteins may only be one small part of these biological processes, but the tools that are being developed to study this part of the proteome may be of great assistance to enhance our insight into these important biological processes.

### **Acknowledgements**

This project was carried out within the research program of the Centre of BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

## References

- Ahn NG, Shabb JB, Old WM, Resing KA** (2007) Achieving in-depth proteomics profiling by mass spectrometry. *ACS Chem Biol* **2**: 39–52
- Albenne C, Canut H, Boudart G, Zhang Y, San Clemente H, Pont-Lezica R, Jamet E** (2009) Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships. *Mol Plant* **2**: 977–989
- Alves G, Ogurtsov AY, Yu YK** (2010) Assigning statistical significance to proteotypic peptides via database searches. *J Proteomics* **74**: 199–211
- Apweiler R, Hermjakob H, Sharon N** (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**: 4–8
- Arockia Jeyaprakash A, Jayashree G, Mahanta SK, Swaminathan CP, Sekar K, Surolia A, Vijayan M** (2005) Structural basis for the energetics of jacalin-sugar interactions: promiscuity versus specificity. *J Mol Biol* **347**: 181–188
- Balaguer E, Neusüss C** (2006) Glycoprotein characterization combining intact protein and glycan analysis by capillary electrophoresis-electrospray ionization-mass spectrometry. *Anal Chem* **78**: 5384–5393
- Banerjee R, Mande SC, Ganesh V, Das K, Dhanaraj V, Mahanta SK, Suguna K, Surolia A, Vijayan M** (1994) Crystal structure of peanut lectin, a protein with an unusual quaternary structure. *Proc Natl Acad Sci U S A* **91**: 227
- Bause E** (1983) Active-site-directed inhibition of asparagine *N*-glycosyltransferases with epoxy-peptide derivatives. *Biochem J* **209**: 323–330
- Bencurova M, Hemmer W, Focke-Tejkl M, Wilson IB, Altmann F** (2004) Specificity of IgG and IgE antibodies against plant and insect glycoprotein glycans determined with artificial glycoforms of human transferrin. *Glycobiology* **14**: 457–466
- Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang PF, Mueller LA, Yoon J, Doyle A, Lander G, et al** (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol* **135**: 745–755
- Cerioti A, Duranti M, Bollini R** (1998) Effects of *N*-glycosylation on the folding and structure of plant proteins. *J Exp Bot* **49**: 1091–1103
- Ceroni A, Maass K, Geyer H, Geyer R, Dell A, Haslam SM** (2008) GlycoWorkbench: A tool for the computer-assisted annotation of mass spectra of glycans†. *J Proteome*

Res 7: 1650–1659

**Chatterjee S, Kolmakova A, Rajesh M** (2008) Regulation of lactosylceramide synthase (glucosylceramide 14 galactosyltransferase); implication as a drug target. *Curr Drug Targets* **9**: 272–281

**Chen FTA, Evangelista RA** (1998) Profiling glycoprotein *N*-linked oligosaccharide by capillary electrophoresis. *Electrophoresis* **19**: 2639–2644

**Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nürnberger T, Jones JDG, Felix G, Boller T** (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**: 497–500

**Cox KM, Sterling JD, Regan JT, Gasdaska JR, Frantz KK, Peele CG, Black A, Passmore D, Moldovan-Loomis C, Srinivasan M** (2006) Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat Biotechnol* **24**: 1591–1597

**Daskalova SM, Radder JE, Cichacz ZA, Olsen SH, Tsaprailis G, Mason H, Lopez LC** (2010) Engineering of *N. benthamiana* L. plants for production of *N*-acetylgalactosamine-glycosylated proteins--towards development of a plant-based platform for production of protein therapeutics with mucin type *O*-glycosylation. *BMC Biotechnol* **10**: 62

**Dodds PN, Rathjen JP** (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat Rev Genet* **11**: 539–548

**Elbers IJ, Stoopen GM, Bakker H, Stevens LH, Bardor M, Molthoff JW, Jordi WJ, Bosch D, Lommen A** (2001) Influence of growth conditions and developmental stage on *N*-glycan heterogeneity of transgenic immunoglobulin G and endogenous proteins in tobacco leaves. *Plant Physiol* **126**: 1314–1322

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

**Fitchette-Laine AC, Gomord V, Cabanes M, Michalski JC, Saint Macary M, Foucher B, Cavelier B, Hawes C, Lerouge P, Faye L** (1997) *N*-glycans harboring the Lewis a epitope are expressed at the surface of plant cells. *Plant J* **12**: 1411–1417

**Fotisch K, Vieths S** (2001) *N*- and *O*-linked oligosaccharides of allergenic glycoproteins. *Glycoconj J* **18**: 373–390

**Frank J, Kaulfurst-Soboll H, Rips S, Koiwa H, von Schaewen A** (2008) Comparative

analyses of *Arabidopsis cgl1* (complex glycan 1) mutants and genetic interaction with staurosporin and temperature sensitive3a. *Plant Physiol.* **148(3)**:1354–67

**Gomord V, Fitchette AC, Menu Bouaouiche L, Saint Jore Dupas C, Plasson C, Michaud D, Faye L** (2010) Plant specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol J* **8**: 564–587

**Green RS, Stone EL, Tenno M, Lehtonen E, Farquhar MG, Marth JD** (2007) Mammalian *N*-glycan branching protects against innate immune self-recognition and inflammation in autoimmune disease pathogenesis. *Immunity* **27**: 308–320

**Grunewald S, Matthijs G, Jaeken J** (2002) Congenital disorders of glycosylation: a review. *Pediatr Res* **52**: 618–624

**Hampton RY** (2002) ER-associated degradation in protein quality control and cellular regulation. *Curr Opin Cell Biol* **14**: 476–482

**Häweker H, Rips S, Koiwa H, Salomon S, Saijo Y, Chinchilla D, Robatzek S, von Schaewen A, Häweker H, Rips S, et al** (2010) Pattern recognition receptors require *N*-glycosylation to mediate plant immunity. *J Biol Chem* **285**: 4629–4636

**Helenius A, Aebi M** (2004) Roles of *N*-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* **73**: 1019–1049

**Henquet M, Lehle L, Schreuder M, Rouwendal G, Molthoff J, Helsper J, Van Der Krol S, Bosch D** (2008) Identification of the gene encoding the  $\alpha$ 1,3-mannosyltransferase (*ALG3*) in *Arabidopsis* and characterization of downstream *N*-glycan processing. *Plant Cell Online* **20**: 1652

**Henquet MGL, Eigenhuijsen J, Hesselink T, Spiegel H, Schreuder MEL, van Duijn E, Cordewener JHG, Depicker A, van der Krol AR, Bosch D, et al** (2010) Characterization of the single-chain Fv-Fc antibody MBP10 produced in *Arabidopsis alg3* mutant seeds. *N-glycosylation plants Sci Appl* 105

**Ihara Y, Sakamoto Y, Mihara M, Shimizu K, Taniguchi N** (1997) Overexpression of *N*-acetylglucosaminyltransferase III disrupts the tyrosine phosphorylation of Trk with resultant signaling dysfunction in PC12 cells treated with nerve growth factor. *J Biol Chem* **272**: 9629

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

**Jackson SL, Hardham AR** (1998) Dynamic rearrangement of the filamentous actin

network occurs during zoosporegenesis and encystment in the oomycete *phytophthora cinnamomi*. Fungal Genet Biol **24**: 24–33

**Jackson SN, Dutta S, Woods AS** (2009) The use of ECD/ETD to identify the site of electrostatic interaction in noncovalent complexes. J Am Soc Mass Spectrom **20**: 176–179

**Jones J, Krag SS, Betenbaugh MJ** (2005) Controlling *N*-linked glycan site occupancy. Biochim Biophys Acta (BBA)-General Subj **1726**: 121–137

**Kamoda S, Nomura C, Kinoshita M, Nishiura S, Ishikawa R, Kakehi K, Kawasaki N, Hayakawa T** (2004) Profiling analysis of oligosaccharides in antibody pharmaceuticals by capillary electrophoresis. J Chromatogr A **1050**: 211–216

**Kang JS, Frank J, Kang CH, Kajiura H, Vikram M, Ueda A, Kim S, Bahk JD, Triplett B, Fujiyama K, et al** (2008) Salt tolerance of *Arabidopsis thaliana* requires maturation of *N*-glycosylated proteins in the Golgi apparatus (vol 105, pg 5933, 2008). Proc Natl Acad Sci U S A **105**: 7893

**Karamanos NK, Hjerpe A** (1999) Strategies for analysis and structure characterization of glycans/proteoglycans by capillary electrophoresis. Their diagnostic and biopharmaceutical importance. Biomed Chromatogr **13**: 507–512

**Kasturi L, Chen H, Shakin-Eshleman SH** (1997) Regulation of *N*-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem J **323** ( Pt **2**): 415–419

**Kato Y, Sakaguchi M, Mori Y, Saito K, Nakamura T, Bakker EP, Sato Y, Goshima S, Uozumi N** (2001) Evidence in support of a four transmembrane-pore-transmembrane topology model for the *Arabidopsis thaliana* Na<sup>+</sup>/K<sup>+</sup> translocating AtHKT1 protein, a member of the superfamily of K<sup>+</sup> transporters. Proc Natl Acad Sci U S A **98**: 6488

**Kedzierski Ł, Montgomery J, Curtis J, Handman E** (2004) Leucine-rich repeats in host-pathogen interactions. Arch Immunol Ther Exp (Warsz) **52**: 104–12

**Kimura Y, Hess D, Sturm A** (1999) The *N*-glycans of jack bean alpha-mannosidase. Structure, topology and function. Eur J Biochem **264**: 168–175

**Klenk HD, Garten W** (1994) Host cell proteases controlling virus pathogenicity. Trends Microbiol **2**: 39–43

**Knauer R, Lehle L** (1999) The oligosaccharyltransferase complex from *Saccharomyces cerevisiae*. Isolation of the *OST6* gene, its synthetic interaction with *OST3*, and analysis of the native complex. J Biol Chem **274**: 17249–17256



**Lehle L, Strahl S, Tanner W** (2006) Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases. *Angew Chem Int Ed Engl* **45**: 6802–18

**Lerouge P, Cabanes-Macheteau M, Rayon C, Fischette-Laine AC, Gomord V, Faye L** (1998) *N*-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol Biol* **38**: 31–48

**Lerouxel O, Mouille G, Andeme-Onzighi C, Bruyant MP, Seveno M, Loutelier-Bourhis C, Driouch A, Hofte H, Lerouge P** (2005) Mutants in DEFECTIVE GLYCOSYLATION, an *Arabidopsis* homolog of an oligosaccharyltransferase complex subunit, show protein underglycosylation and defects in cell differentiation and growth. *Plant J* **42**: 455–468

**Liebming E, Veit C, Mach L, Strasser R** (2010) Mannose trimming reactions in the early stages of the *N*-glycan processing pathway. *Plant Signal Behav* **5**: 476–478  
**Lige B, Ma S, van Huystee RB** (2001) The effects of the site-directed removal of *N*-glycosylation from cationic peanut peroxidase on its function. *Arch Biochem Biophys* **386**: 17–24

**Lim JM, Aoki K, Angel P, Garrison D, King D, Tiemeyer M, Bergmann C, Wells L** (2008) Mapping glycans onto specific *N*-Linked glycosylation sites of *Pyrus communis*

**PGIP Redefines the interface for EPG- PGIP interactions. J Proteome Res** **8**: 673–680  
**Liu JX, Howell SH** (2010) Endoplasmic Reticulum protein quality control and its relationship to environmental stress responses in plants. *Plant Cell Online* **22**: 2930–2942

**Mellquist JL, Kasturi L, Spitalnik SL, Shakin-Eshleman SH** (1998) The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of *N*-linked core glycosylation efficiency. *Biochemistry* **37**: 6833–6837

**Melo NS, Nimtz M, Conradt HS, Fevereiro PS, Costa J** (1997) Identification of the human Lewis(a) carbohydrate motif in a secretory peroxidase from a plant cell suspension culture (*Vaccinium myrtillus* L.). *Febs Lett* **415**: 186–191

**Moremen KW, Molinari M** (2006) *N*-linked glycan recognition and processing: the molecular basis of endoplasmic reticulum quality control. *Curr Opin Struct Biol* **16**: 592–599

**Mühlenhoff M, Manegold A, Windfuhr M, Gotza B, Gerardy-Schahn R** (2001) The impact of *N*-glycosylation on the functions of polysialyltransferases. *J Biol Chem* **276**: 34066

**Natsuka S, Kawaguchi M, Wada Y, Ichikawa A, Ikura K, Hase S** (2005) Characterization of wheat germ agglutinin ligand on soluble glycoproteins in *Caenorhabditis elegans*. *J Biochem* **138**: 209

**Nesvizhskii AI** (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. *J Proteomics* **73**: 2092–2123

**Pattison RJ, Amtmann A** (2009) *N*-glycan production in the endoplasmic reticulum of plants. *Trends Plant Sci* **14**: 92–99

**Petrescu A-JJ, Milac A-LL, Petrescu SM, Dwek R a, Wormald MR** (2004) Statistical analysis of the protein environment of *N*-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**: 103

**van Ree R, Cabanes-Macheteau M, Akkerdaas J, Milazzo JP, Loutelier-Bourhis C, Rayon C, Villalba M, Koppelman S, Aalberse R, Rodriguez R, et al** (2000) Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J Biol Chem* **275**: 11451–11458

**von Schaewen A, Sturm A, O'Neill J, Chrispeels MJ, Vonschaewen A, Sturm A, Oneill J, Chrispeels MJ** (1993) Isolation of a mutant *Arabidopsis* plant that lacks *N*-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex *N*-linked glycans. *Plant Physiol* **102**: 1109–18

**Schollen E, Grunewald S, Keldermans L, Albrecht B, Korner C, Matthijs G** (2005) CDG-Id caused by homozygosity for an ALG3 mutation due to segmental maternal isodisomy UPD3(q21.3-qter). *Eur J Med Genet* **48**: 153–158

**Senger RS, Karim MN** (2005) Variable site occupancy classification of *N*-linked glycosylation using artificial neural networks. *Biotechnol Prog* **21**: 1653–1662

**Shakey Q, Bates B, Wu J** (2010) An approach to quantifying *N*-linked glycoproteins by enzyme-catalyzed <sup>18</sup>O<sub>3</sub>-labeling of solid-phase enriched glycopeptides. *Anal Chem* **82**: 7722–7728

**Shanmugam V** (2005) Role of extracytoplasmic leucine rich repeat proteins in plant defence mechanisms. *Microbiol Res* **160**: 83–94

**Spiro RG** (2004) Role of *N*-linked polymannose oligosaccharides in targeting glycoproteins for endoplasmic reticulum-associated degradation. *Cell Mol life Sci* **61**: 1025–1041

**Strasser R, Stadlmann J, Schähls M, Stiegler G, Quendler H, Mach L, Glössl**

**J, Weterings K, Pabst M, Steinkellner H** (2008) Generation of glyco engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human like *N* glycan structure. *Plant Biotechnol J* **6**: 392–402

**Strasser R, Stadlmann J, Svoboda B, Altmann F, Glossl J, Mach L, Glössl J, Mach L, Glossl J, Mach L** (2005) Molecular basis of *N*-acetylglucosaminyltransferase I deficiency in *Arabidopsis thaliana* plants lacking complex *N*-glycans. *Biochem J* **387**: 385–391

**Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF** (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **101**: 9528

**Szumilo T, Kaushal GP, Elbein AD** (1986) Purification and properties of glucosidase I from mung bean seedlings. *Arch Biochem Biophys* **247**: 261–271

**Takahashi M, Tsuda T, Ikeda Y, Honke K, Taniguchi N** (2003) Role of *N*-glycans in growth factor signaling. *Glycoconj J* **20**: 207–212

**Taverna M, Baillet A, Biou D, Schlüter M, Werner R, Ferrier D** (1992) Analysis of carbohydrate mediated heterogeneity and characterization of *N* linked oligosaccharides of glycoproteins by high performance capillary electrophoresis. *Electrophoresis* **13**: 359–366

**Thoma S, Hecht U, Kippers A, Botella J, De Vries S, Somerville C** (1994) Tissue-specific expression of a gene encoding a cell wall-localized lipid transfer protein from *Arabidopsis*. *Plant Physiol* **105**: 35–45

**Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H** (2007) Solid-phase extraction of *N*-linked glycopeptides. *Nat Protoc* **2**: 334–339

**Triguero A, Cabrera G, Cremata JA, Yuen CT, Wheeler J, Ramirez NI** (2005) Plant-derived mouse IgG monoclonal antibody fused to KDEL endoplasmic reticulum-retention signal is *N*-glycosylated homogeneously throughout the plant with mostly high-mannose-type *N*-glycans. *Plant Biotechnol J* **3**: 449–457

**Trombetta ES** (2003) The contribution of *N*-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis. *Glycobiology* **13**: 77R–91R

**Uemura S, Kurose T, Suzuki T, Yoshida S, Ito M, Saito M, Horiuchi M, Inagaki F, Igarashi Y, Inokuchi J** (2006) Substitution of the *N*-glycan function in glycosyltransferases by specific amino acids: ST3Gal-V as a model enzyme. *Glycobiology* **16**: 258

**Varki A** (1999) Essentials of glycobiology. Cold Spring Harbor Laboratory Pr

**Wo ZG, Oswald RE** (1994) Transmembrane topology of two kainate receptor subunits revealed by *N*-glycosylation. Proc Natl Acad Sci U S A **91**: 7154

**Yoshida Y** (2003) A novel role for *N*-glycans in the ERAD system. J Biochem **134**: 183–190

**Yusa A, Kitajima K, Habuchi O** (2005) *N*-linked oligosaccharides are required to produce and stabilize the active form of chondroitin 4-sulphotransferase-1. Biochem J **388**: 115

**Yusa A, Kitajima K, Habuchi O** (2006) *N*-linked oligosaccharides on chondroitin 6-sulfotransferase-1 are required for production of the active enzyme, Golgi localization, and sulfotransferase activity toward keratan sulfate. J Biol Chem **281**: 20393

**Zhang H, Loriaux P, Eng J, Campbell D, Keller A, Moss P, Bonneau R, Zhang N, Zhou Y, Wollscheid B** (2006) UniPep-a database for human *N*-linked glycosites: a resource for biomarker discovery. Genome Biol **7**: R73

**Zhang Y, Giboulot A, Zivy M, Valot B, Jamet E, Albenne C** (2010) Combining various strategies to increase the coverage of the plant cell wall glycoproteome. Phytochemistry. doi: 10.1016/j.phytochem.2010.10.019

**Zhang Y, Huang L, Wang Z** (2008) Techniques for staining glycoprotein on gel electrophoresis or electroblotting. Prog Chem **20**: 1158–1164

## Supplementary data

### *Plant materials and growth conditions*

Seeds of *Arabidopsis thaliana* lines were sown on 9 cm 0.8% Daishin agar in Petri dishes and placed in a cold room at 4°C for 2 days in the dark to promote uniform germination. Plants were grown in soil in a greenhouse with a 16h day/ 8h dark cycle at a temperature of 22°C.

### *Cell specific complementation of cgl1-1 mutants*

*Arabidopsis* GnTI and the RBCS promoter were PCR amplified using Platinum Pfx polymerase (Invitrogen) and primers flanking the coding region. The primers used for cloning the GnTI were GnTI-F (5'-ACATATGGCGAGGATCTCGTGTGAC TTG-3') and GnTI-R (5'-ACTCGAG TCAGGAATTTCGAATCCCAAGCTGC-3') which contain a *Bgl*III and *Xho*I site respectively. The primers used for cloning the RBCS promoter were RBCS-F (5'-AA GCTTGACGATATATTACAGGAAAAATCT-3') and RBCS-R (5'-CATATGTTCTT CTTCTTCTTCTTTTGC-3') which contain a *Hind*III and *Sma*I site respectively. The epidermal LTP promoter has been cloned and characterized before (Thoma et al., 1994). The *Arabidopsis* GnTI was ligated together with either the epidermal LTP promoter or the mesophyll RBCS promoter into the binary pBinplus vector. The constructs containing the LTP1-GnTI or RBCS-GnTI cassettes were transferred to *Agrobacterium tumefaciens* AGL-0 by triparental mating using the *E. Coli* pRK2013 helper plasmid. *Arabidopsis* mutant plants were transformed by immersion in the bacteria suspension culture. Seeds from transformed plants were selected by growing on plates containing MS powder (4.4 g/L), sucrose (10 g/L) and Daishin agar (8 g/L) with kanamycin (50 µg/mL) as selective marker. Seeds were germinated and grown on selective medium for 1-2 weeks (25 °C long day conditions). The putative transformants were transferred to soil and grown using standard conditions (16 hours day/8 hours dark).

### *SDS-PAGE and immunoblotting*

Plants were grown in the condition as described in the supplement. Flowers were collected at developmental stage 12 and siliques were harvested at the length of around 1.5 cm. Plant material was ground in liquid nitrogen, resuspended in 10 µL phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) per mg of plant material and centrifuged. An aliquot of the supernatant (approximately 5 µg soluble protein) was immediately mixed with SDS-polyacrylamide gel electrophoresis (PAGE) loading buffer, denatured at 95°C for 5 min and subjected to SDS-PAGE (8 or 12.5%) under reducing conditions. Western blotting was performed using PVDF membranes, blocked with 5% (w/v) non-fat dry milk in Tris-buffered saline (TBS, 20 mM Tris-HCl, pH 7.6, 137 mM NaCl) with 0.05% Tween20. The membranes were probed with either anti-HRP (1:2000; Sigma) or anti-GFP (1:5000; Roche). Detection of bound primary antibodies was performed with BCIP/NBT or after

incubation with goat anti-rabbit antibodies. The faint staining with anti-HRP of the *cgll-1* is attributed to background staining. It may also be caused by the restore of GnTI activity in the *cgll-1* mutant when conditions favour underglycosylation. In this case the mutant GnTI is no longer retained into the ER and can reach the Golgi apparatus where, despite the mutation, it exhibits low level of activity (Frank et al., 2008).

# Chapter 3

## ***N*-glycan occupancy of *Arabidopsis* *N*-glycoproteins**

Wei Song <sup>a,b</sup>, Remco A. Mentink <sup>c</sup>, Maurice G.L. Henquet <sup>b</sup>, Jan H.G. Cordewener <sup>b</sup>,  
Aalt D.J. van Dijk <sup>b</sup>, Dirk Bosch <sup>b</sup>, Harro Bouwmeester <sup>a</sup>,  
Antoine H.P. America <sup>b</sup>, Alexander R. van der Krol <sup>a,\*</sup>

<sup>a</sup> Laboratory of Plant Physiology, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands

<sup>b</sup> Plant Research International, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands

<sup>c</sup> Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences and University  
Medical Center, Uppsalalaan 8, 3584 CT,  
Utrecht, The Netherlands

Published in: J Proteomics. 2013 Nov 20;93:343-55

## Abstract

Most secreted proteins in eukaryotes are modified on the amino acid consensus sequence N-P!-S/T by an *N*-glycan through the process of *N*-glycosylation. The *N*-glycans on glycoproteins are processed in the endoplasmic reticulum (ER) to different mannose-type *N*-glycans or, when the protein passes through the Golgi apparatus, to different complex glycan forms. Here we describe the capturing of *N*-glycopeptides from a trypsin digest of total protein extracts of *Arabidopsis* plants and release of these captured peptides following Peptide *N*-glycosidase (PNGase) treatment for analysis of *N*-glycan site-occupancy. The mixture of peptides released as a consequence of the PNGase treatment was analysed by two dimensional nano-LC-MS. As the PNGase treatment of glycopeptides results in the deamidation of the asparagine (N) in the N-P!-S/T site of the released peptide, this asparagine (N) to aspartic acid (D) conversion is used as a glycosylation ‘signature’. The efficiency of PNGase F and PNGase A in peptide release are discussed. The identification of proteins with a single glycopeptide was limited by the used search algorithm but could be improved using a reference database including deamidated peptide sequences. Additional stringency settings were used for filtering results to minimize false discovery. This resulted in identification of 330 glycopeptides on 173 glycoproteins from *Arabidopsis*, of which 28 putative glycoproteins, that were previously not annotated as secreted protein in The Arabidopsis Information Resource database (TAIR). Furthermore, the identified glycosylation site occupancy helped to determine the correct topology for membrane proteins. A quantitative comparison of peptide signal was made between *Arabidopsis* Col-0 and complex-glycan-less (*cgl1-1*) mutant *Arabidopsis* from three replicate leaf samples using a label-free MS peak comparison. As an example, the identified membrane protein SKU5 (AT4G12420) showed differential glycopeptide intensity ratios between *Arabidopsis* Col-0 and *cgl1-1* indicating heterogeneous glycan modification on single protein.

## Keywords

*N*-glycosylation; *N*-glycoproteomics; *Arabidopsis*; LC-MS; MS/MS; Data Independent Acquisition (MS<sup>E</sup>)

## Abbreviations

ER, Endoplasmatic Reticulum; *cgl*, complex glycan less *Arabidopsis* mutant; OST, oligo-saccharyl-transferase; PNGase, peptide-*N*-glycosidase; PAGE, polyacrylamide gel electrophoresis; LC-MS/MS, liquid chromatography tandem mass spectrometry; UPLC, Ultra performance liquid chromatography; ACN, acetonitrile; DDA, data dependent acquisition; DIA (MS<sup>E</sup>), data independent acquisition; Q-TOF, quadrupole time of flight; GO, Gene Ontology; TAIR: The Arabidopsis Information Resource database;



## Introduction

Glycosylation is one of the most common and essential post translational modifications of proteins in eukaryotic cells (Lee et al., 2005). In eukaryotes proteins may be modified by *N*-glycans, *O*-glycans or a glycosyl phosphatidyl inositol (GPI) lipid anchor (Spiro, 2002). Here we are mainly concerned with *N*-glycosylation which takes place on proteins secreted into the ER. In eukaryotes *N*-glycans are involved in protein folding, protein folding quality control, polar protein localization and stability, ligand binding, endocytosis, immune recognition, inflammation and pathogenicity, cell signalling and cell motility (Varki, 1993; Parodi, 2000; Schollen et al., 2005; Sun et al., 2005). The importance of *N*-glycan modifications in mammalian cells is exemplified by the strong congenital disorders resulting from mutations in *N*-glycan processing (Grunewald et al., 2002; Schollen et al., 2005). In contrast, in plants the structure of mannose type glycans in the ER seems to be non-essential as for example in the *alg3-2* mutant, where no clear phenotype can be observed (Henquet et al., 2008). Also the complex glycan less (*cgl1-1*) mutant of *Arabidopsis*, which completely lacks complex *N*-glycans has no obvious growth phenotype and only shows some altered salt stress tolerance (Kang et al., 2008). The lack of a clear phenotype from *N*-glycosylation mutants may be the reason that this process is less well studied in plants. However, techniques that can determine N-P!-S/T *N*-glycan site occupancy of proteins and follow these under different (a)biotic stress conditions or through development may eventually allow for wider elucidation of their biological function in plants. Moreover, because *N*-glycoproteins are part of the secretome, they make an interesting sub-class of the proteome for characterization and targeted functionality studies in developmental and (a)biotic stress response studies. In order to study the role of *N*-glycosylation and *N*-glycan occupancy on plant glycoproteins we optimized the previously described procedure for isolation of glycoproteins from yeast or mammalian cells and we applied it to capture and analyse glycopeptides from plants (Tian et al., 2007).

### ***Complex N-Glycan modifications in plants hinder glycoproteomics***

The core *N*-glycosylation pathway is evolutionary strongly conserved in all eukaryotes. Upon import into the ER, secreted proteins are glycosylated by the OST complex on the *N*-glycosylation consensus sequence site N-P!-S/T, where x can be any amino acid except proline. *N*-glycosylation starts with transfer of a glycosylated high mannose *N*-glycan (Rayon et al., 1998) to the asparagine side chain. In the ER the mannose glycans are slowly trimmed as part of a protein folding quality control mechanism (Liebminger et al., 2010). This processing in the ER results in different high mannose type *N*-glycans for ER resident proteins, while glycoproteins that are transported to and through the Golgi apparatus are further modified by Golgi resident glycosyltransferases and glycosidases into complex type *N*-glycans (Rayon et al., 1998). In general, the distinction between mannose and complex type *N*-glycans on a glycoprotein may therefore give information on its sub-cellular location history. Plant specific complex *N*-glycans contain  $\alpha$ -1,3 fucose attached to the glycan core and  $\beta$ -1,2 xylose attached

to the central mannose residue (Lerouge et al., 1998). The plant specific  $\alpha$ -1,3 fucose core modification prevents cleavage by PNGase F, an enzyme broadly used for deglycosylating glycopeptides. Procedures developed for yeast or mammalian glycoproteomics involving the usage of PNGase F are therefore severely limited in plant glycoproteomics. However, for full glycoproteomics in *Arabidopsis* we can make a use of the glycosylation mutant *cgll-1* (von Schaewen et al., 1993; Strasser et al., 2005). This mutant lacks a functional *N*-acetylglucosaminyltransferase 1 (GnT1) gene, which is essential for the formation of complex glycans. Consequently, the glycoproteins in *cgll-1* are all of the mannose type, which can be cleaved by PNGase F. Under normal growth conditions this *cgll-1* mutation has no visible effect on the growth phenotype of the *Arabidopsis* plants.

### ***Glycoproteomics in plants***

The presence of *N*-glycans on plant proteins has only been proven incidentally, usually by a de-glycosylation assay which results in increased mobility of the target protein in SDS-PAGE. However, such assays do not give direct information on glycosylation-site occupancy. Glycoproteins or glycopeptides may be studied using lectins (sugar binding proteins with either a broad or narrow specificity binding of glycan structures) or antibody affinity capturing (Chang and Chang, 1986; Varki, 1999; Hagglund et al., 2004; Hirabayashi, 2004; Qiu and Regnier, 2005; Uematsu et al., 2005; Yang and Hancock, 2005; Wang et al., 2006; Hagglund et al., 2007; Kaji et al., 2007; Kubota et al., 2008; Pan et al., 2011; Song et al., 2011). Recently, a lectin based *N*-glycoprotein enrichment approach was developed by Zielinska *et al.* (Zielinska et al., 2010) and applied to mammalian cells by using the filter aided sample preparation (FASP) method for *N*-glycopeptide capture. This method was also applied in *Arabidopsis* by Zielinska resulting in 2186 glycosylation sites mapped (Zielinska et al., 2012). A more general approach is chemical crosslinking of glycopeptides through their carbohydrates to a resin (Zhang et al., 2003; Tian et al., 2007; Zhang, 2007). For this, the carbohydrates in glycopeptides are chemically activated into aldehydes, which can be covalently coupled to hydrazide-derivatized resin beads. After washing away the non-bound peptides, the conjugated glycopeptides can subsequently be released by a peptide-*N*-glycosidase like PNGase F. Here we developed and optimized this procedure to capture and analyse glycopeptides from plants. Because plant specific complex *N*-glycans with a fucose modified core structure are not cleaved by PNGase F, we tested and compared the efficiency of glycopeptide release by both PNGase F and PNGase A (cleaves all types of asparagine bound *N*-glycans). Results showed that PNGase A was not effective in our experimental protocol, only PNGase F could be applied for glycopeptide release. This limits glycoproteomics of *Arabidopsis* Col-0 to proteins located in the ER or proteins targeted to subcellular compartments that bypass the Golgi apparatus. However, we also performed glycoproteomics on protein extracts of the *Arabidopsis cgll-1* mutant containing only mannose type glycan, which should allow for release by PNGase F of all glycopeptides bound to hydrazide. Results were limited by peptide identification protocols but could be improved by adjustment of reference protein database. Further,

identification of *N*-glycopeptides enabled mapping the topology of membrane proteins, prediction of heterogeneous glycosylation, protein localization and function.

## Materials and methods

### *Plant material and growth conditions*

For leaf material from mature rosette plants, *Arabidopsis thaliana* Col-0 and *cgl1-1* mutant were grown in a climate chamber for 35 days under long day conditions (16 hrs daylight/ 8 hrs darkness) before harvest. All rosette leaves from a plant were harvested and leaves from 10 plants were pooled. For seedling samples, seeds were germinated on 1% agarose (Sigma) in ½ MS medium (Duchefa Biochemie), cold incubated for 3 days at 4°C and subsequently grown under long day conditions at 22°C for 7 days. For harvesting, seedlings and plates were frozen in liquid nitrogen and seedlings were scraped off the plates. One sample contained the pooled material of ~500 seedlings.

### *Protein extraction*

Total protein was extracted by adding 750 µL of extraction buffer (8 M urea, 0.05% SDS, 5 mM DTT, 25 mM EDTA, 0.1 M Tris-HCl; pH8.0) to 30 mg of in liquid nitrogen grinded plant material (leaves or seedlings). After centrifugation (10 min at 20,000g) the proteins in the supernatant were precipitated by 15% (w/v) of pre-cold TCA (trichloroacetic acid). Sample was kept on ice for 30 min, followed by centrifugation at 4°C (10 min at 20,000g). The protein pellet was washed two times with 1 mL of 100% pre-cold acetone. After air-drying, the protein pellet was resuspended in 150 µL extraction buffer by sonication. Iodoacetamide (IAA) was added to a final concentration of 12 mM and incubated in the dark for 30 min. After adding 3 volumes of 0.1 M Tris-HCl (pH 8.0) buffer, the proteins were digested with trypsin (Sigma) (trypsin:protein=1:50) by overnight incubation at 37°C. Digestion was terminated by adding 0.1% TFA (trifluoroacetic acid). Before digestion, the protein concentration of the samples was determined by Bradford protein assay.

### *Capturing glycopeptides to hydrazide beads*

The digested sample was loaded on a 1 mL C18 SPE column (Supelco) and the peptides were eluted with 1 mL of 60% acetonitrile (ACN), 0.1% TFA. Subsequently, 10 mM NaIO<sub>4</sub> was added to the sample and incubated for one hour in the dark at 4°C without shaking. An additional C18 SPE purification step was applied to remove the NaIO<sub>4</sub>. Speedvac was applied to reduce the volume of the elution till 200 µL. 25 mg of hydrazide resin beads (Bio-rad) (50 µL of 50% slurry) were added to the 200 µL sample in 60% ACN, 0.1% TFA and incubated overnight at 37°C with gently shaking. To remove the un-coupled non-glycosylated peptides, the beads were washed three times with 1 mL of DMF, three times with 1 mL of deionized water, and finally three times with 1 mL of PNGase buffer (for PNGase F using 50 mM sodium phosphate buffer, pH

7.5; for PNGase A using 10 mM sodium acetate buffer, pH 5.0).

### ***Release of peptides from hydrazide beads using PNGase F or PNGase A***

To release the peptides from the beads, 25 mg resin beads was incubated with 3  $\mu\text{L}$  of PNGase F (500 U  $\mu\text{L}^{-1}$ , New England Biolabs) or 10  $\mu\text{L}$  of PNGase A (0.05 mU  $\mu\text{L}^{-1}$ , Roche) in 50  $\mu\text{L}$  PNGase buffer overnight at 37°C with gently shaking. After centrifugation at 2,500g for 5 min, the supernatant was collected and the beads were washed three times with the PNGase buffer. The combined supernatants were loaded on a C18 SPE column and the peptides were eluted with 60% ACN, 0.1% TFA. The purified peptides were dried by vacuum centrifugation and dissolved in 20  $\mu\text{L}$  of 0.1 M ammonium formate (pH 10) for 2D-LC-MS analysis.

### ***Peptide analysis using 2D-LC-MS/MS and 2D-LC-MS<sup>E</sup>***

For high-resolution separation of the complex tryptic (glyco)peptide samples a nanoAcquity 2-D UPLC system (Waters Corporation, Manchester, UK) was used employing orthogonal reversed phase separation at high and low pH, respectively. With this 2-D set up, the pool of peptides was eluted from the first dimension XBridge C18 trap column (in 20 mM ammonium formate pH 10) using a discontinuous step gradient of 12%, 15%, 18%, 20%, 25%, 35% and 65% ACN. For the second dimension an acidic ACN gradient was applied using a BEH C18 column (75  $\mu\text{m}$  x 25 cm, Waters, UK) and a 65 min linear gradient from 3 to 40% ACN (in 0.1 % FA) at 200 nL/min. The eluting peptides were on-line injected into a Synapt G1 Q-TOF MS instrument (Waters Corporation, Manchester, UK) using a nanospray device coupled to the second dimension column output. The Synapt MS was operated in positive mode with [Glu<sup>1</sup>] fibrinopeptide B (1 pmol/ $\mu\text{L}$ ; Sigma) as reference (lock mass) sampled every 30 s. LC-MS data were collected with the Synapt operating in either MS/MS or MS<sup>E</sup> mode for data-dependent acquisition (DDA) or data-independent acquisition (DIA) using low (6 eV) and elevated (ramp from 15 to 35 eV) collision energy every 0.6 s over a 140-1900 m/z range, respectively. DDA was performed by peptide fragmentation on the three most intense multiply charged ions that were detected in the MS survey scan (0.6 s) over a 300-1400 m/z range and a dynamic exclusion window of 60 s with an automatically adjusted collision energy based on the observed precursor m/z and charge state. Confidence level of the glycoprotein identification was determined based on the following features: (1) hydrazide bonding; (2) comparison of measured masses and predicted masses; (3) peptide fragmentation pattern; (4) AA-sequence covering the consensus N-P!-S/T motif; (5) Deamidation of N to D as a signature for PNGase treatment; (6) literature confirmation.

### ***Data base search***

LC-MS/MS and MS<sup>E</sup> data were processed using ProteinLynx™ Global Server software (PLGS version 2.5, Waters Corporation) and the resulting list of masses containing all

the fragment information was searched against the TAIR protein sequence database downloaded from The Arabidopsis Information Resource (TAIR, [www.arabidopsis.org](http://www.arabidopsis.org)). Peptide identification was done by matching the detected peptide fragmentation spectrum to that of predicted peptide spectra using the *Arabidopsis* TAIR protein sequence database as input in PLGS. For matching MS<sup>E</sup> data to predicted peptide spectra the following settings were used: minimum fragment ion matches per peptide: four, per protein: seven, minimum one peptide per protein match and a false discovery threshold of four percent. The false discovery rate was determined automatically in PLGS by searching the randomized TAIR protein sequence database. Peptide and fragment tolerance was set to automatic: the software determines tolerance thresholds based on the actual resolution and stability in the MS data (Geromanos et al., 2009; Li et al., 2009). In practice this resulted in mass tolerances below 10 ppm for precursor and below 30 ppm for fragment ions. Carbamido-methylation of cysteine was set as fixed modification, and oxidation of methionine (M) and deamidation on asparagine (N) as (enriched) variable modification. The deamidation of N can be used as a signature for the determination of glycosylation site on peptides which were released by PNGase F (PNGase F cleavage between the peptide backbone and the N-glycan results in an N to D conversion at the N-P!-S/T glycosylation consensus site). For DDA analysis the peptide tolerance was set to 30 ppm and a fragment tolerance of 0.05 Da. Carbamido-methylation (Cys) was used as fixed modification and deamidation (NQ) and oxidation (M) as variable modifications. The AutoMod option was applied as secondary search to the database search results with a maximum of one missed trypsin cleavage and non-specific secondary digest reagent were chosen. Finally, the DDA and MS<sup>E</sup> outputs were merged in Excel. Gene ontology annotation was done by using the tools from TAIR (<http://www.arabidopsis.org/tools/bulk/go/index.jsp>).

### ***Construction of a modified Arabidopsis protein reference database***

A limitation of the software for protein identification was observed, which is caused partly by the N to D conversion (similar to a deamidation) with 1 Da increase from the PNGase treatment. To improve the protein identification, a custom constructed *Arabidopsis* protein database was built. In the constructed database, the whole TAIR proteome which contains putative glycosylation sites was used. To find putative glycosylation sites (pattern N[ACEFGHIPLMNQRSTVWY][ST]) and putative ER resident proteins (patterns KDEL>, RDEL>, HDEL> and KKxx>) the program ps\_scan was used (Gattiker et al., 2002). A single asparagine (N) on the N-P!-S/T motif from the putative glycosylation site was modified to aspartic acid (D) in the modified database. If the protein contained more than one putative glycosylation site a combination of two modifications per protein was entered in all possible combinations. Each modified sequence was given a new accession number, causing a dramatically increased complexity of the modified database, which increased the searching time of the software for protein identification. Therefore, only single and double asparagine modification was employed and combined in our modified database.

### ***Protein quantification in Progenesis***

Protein quantification from raw MS<sup>E</sup> data was done in Progenesis LC-MS version 4.0 (Nonlinear Dynamics, UK) using lock-spray and dead time correction as mass calibration. Progenesis performed extraction of ion features such as mass, charge, intensity, retention time, etc, after retention time alignment and peak detection of all the samples. Peptide intensity was normalized by the algorithm in Progenesis. Further, peptide identification from ProteinLynx were linked to matching features in Progenesis for coupling identification to quantification. Eventually, the identified peptides with normalized intensity and other features from Progenesis were exported to Excel for further data analysis (e.g. intensity ratio analysis, see Fig. 5B). Three experimental replicates (separate extracts) from *cgli-1* and *Arabidopsis* Col-0 leaf samples were used for protein quantification.

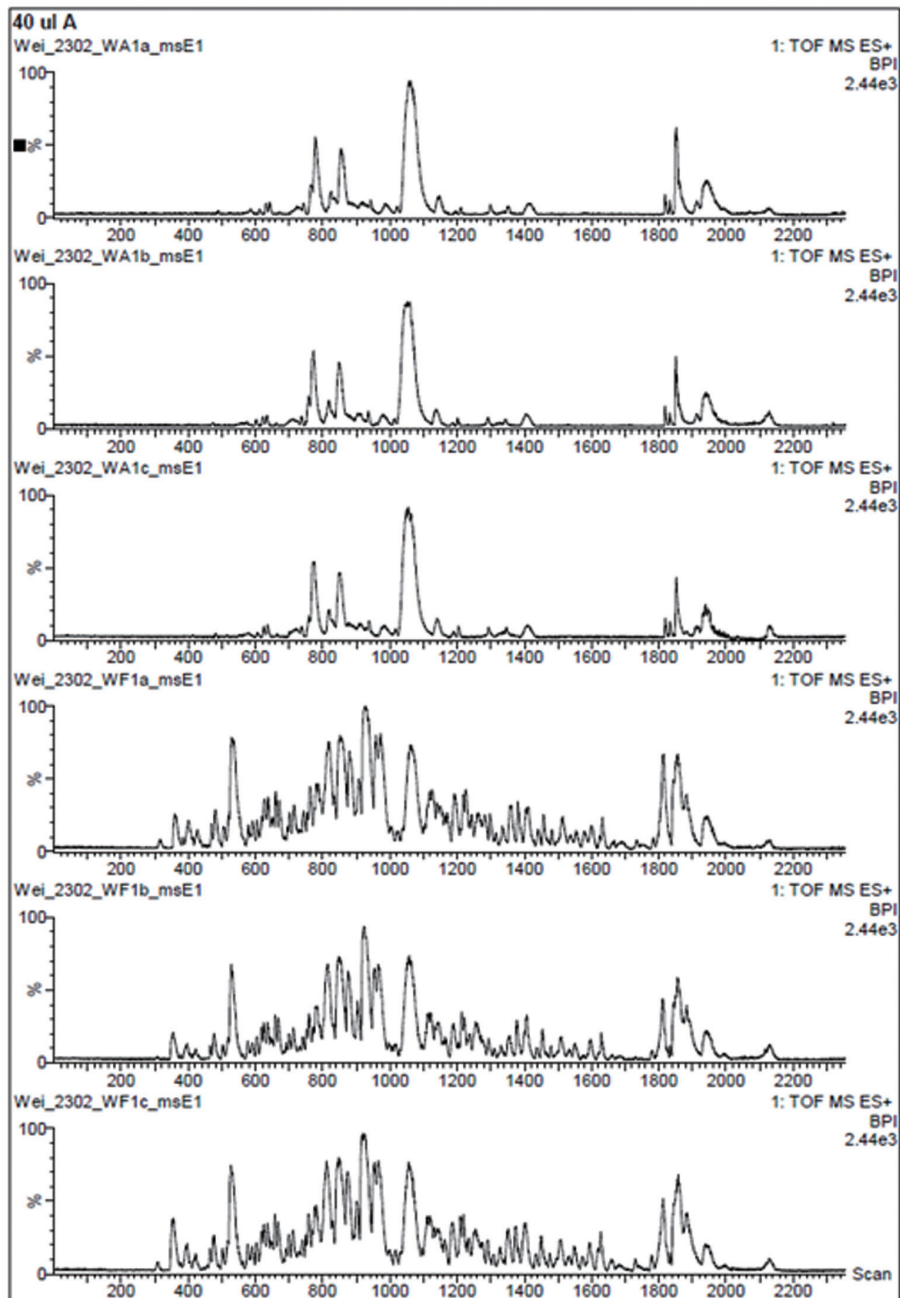
## **Results**

### ***Evaluation of PNGase treated glycopeptide release***

3

Proteins were extracted from *Arabidopsis* Col-0 leaves and digested with trypsin, followed by activation of the glycans on glycopeptides by NaIO<sub>4</sub> for coupling to the hydrazide resin beads (see methods). This procedure results in coupling of all peptides containing oxidized sugars (both *N*-glycopeptides and *O*-glycopeptides) to the hydrazide resin. After extensive washing of the resin to remove unbound peptides, the release of the bound *N*-glycopeptides with PNGase F or PNGase A was tested. In this procedure only PNGase A (cleaves all types of asparagine bound *N*-glycans) can release all peptides bound by *N*-glycans, while PNGase F (does not cleave *N*-glycans modified with fucose) will only release peptides bound by mannose type *N*-glycans. Because the commercial PNGase A is only available at 0.05 mU/μL, only 0.5 mU of PNGase A were used for 25μL of resin with bound peptides. The more restrictive PNGase F is available at 500 U/μL and 1500 U of PNGase F were used for release of glycopeptides from 25μL of resin. Although 1 U is defined differently for the two enzymes by the two suppliers, results show that the signal (base peak chromatogram (2<sup>nd</sup> LC) of fraction 1 (1<sup>st</sup> LC) from 2D-LC-MS<sup>E</sup>) from peptides released by PNGase A was much lower than that from PNGase F (Fig. 1 and also Supplementary Fig. S1A and B) due to the enormous difference in units of the two enzymes applied in the experiment. In repeated experiments the yield of peptides released by PNGase A was always much lower than that by PNGase F. Apparently, the high levels of PNGase A needed for peptide release combined with the low concentration of available PNGase A do not allow for our type of proteomics grade experiments. Therefore, in subsequent experiments only the proteomics compatible PNGase F was used. For protein extracts of *Arabidopsis* Col-0 this means that only peptides that are bound to the hydrazide resin with ‘mannose’ type glycans will be released. However, for release of all captured *N*-glycopeptides, allowing for full glycoproteome analysis, we used the *cgli-1* mutant of *Arabidopsis*. This mutant is not capable of making complex *N*-glycans and consequently all glycoproteins in this

mutant only contain ‘mannose’ type *N*-glycans, which all can be cleaved by PNGase F.



**Figure 1.** Comparison of PNGase A and PNGase F base peak chromatogram (2<sup>nd</sup> LC) of fraction 1 (1<sup>st</sup> LC) from 2D LC-MS<sup>E</sup> chromatograms of three replicate *Arabidopsis* Col-0 extracts released with PNGase A (top 3) and PNGase F (bottom 3).

### ***Limited protein identification matching to Arabidopsis standard protein sequence database***

Proteins were extracted from leaf or seedlings from the *Arabidopsis* mutant *cgll-1* and digested by trypsin. After activation of the sugar groups, glycopeptides were captured on hydrazide beads. Non bound peptides were washed away and covalently linked *N*-glycopeptides were released by PNGase F treatment (see methods). Released peptides were analysed by 2D-LC-MS and peptide fragmentation spectra were obtained both in the DDA and MS<sup>E</sup> mode. Table 1 shows the summary of total peptide numbers that were detected from 19 leaf samples and 11 seedlings samples respectively from *Arabidopsis cgll-1* mutant. DDA analysis with ProteinLynx Global Server software (PLGS) yielded in total 1194 detected peptide masses that matched 190 proteins from the TAIR protein sequence database. In contrast, analysis of the MS<sup>E</sup> spectra resulted in only 641 peptide matches. This lower percentage of identified peptides from MS<sup>E</sup> spectra was not in line with the quality of the MS<sup>E</sup> spectra, suggesting that PLGS may have a problem identifying individual peptides from the mixed spectral input of MS<sup>E</sup>. Some of the detected peptides masses that matched peptides in the database did not have the N to D glycosylation signature suggesting contamination with non-glycopeptides. However, analysis of the wash fluids showed that after three washes of resin with bound glycopeptides, no peptides could be detected in the wash fluid (data not show). Results therefore indicate that non-glycopeptides seem to be trapped to the resin by the coupled glycopeptides and that these contaminants were only released upon the PNGase F treatment. For this reason we used the presence of the *N*-glycosylation consensus site in the identified peptide sequence and the N to D conversion as two strict filters to distinguish between glycopeptides and putative contaminants (Table 1).

**Table 1.** Identified *Arabidopsis N*-glycoproteome. Cumulative *Arabidopsis N*-glycoproteome identifications from leaf and seedling samples with DDA and MS<sup>E</sup> using normal TAIR database or modified TAIR database containing all possible combinations of N to D conversions of putative glycopeptides. Confirmed glycoproteins fit criteria of sequence covering N-P!-S/T site, N to D conversion and being part of annotated secretome. Data was obtained from 19 leaf and 11 seedling samples from *Arabidopsis* and combined.

TAIR10	Total peptides	Peptides with deamidation	Peptides with N-P!-S/T	Peptides with deamidation and N-P!-S/T	Total proteins	Confirmed glycoproteins	proteins with single identified glycopeptide	proteins with multiple identified glycopeptides
DDA	1194	338 (28.3%)	374 (31.3%)	235 (19.7%)	576	190 (33.0%)	115 (20.0%)	75 (13.0%)
DDA (N to D)	1107	N/A	513 (46.3%)	N/A	479	252 (52.6%)	142 (29.6%)	110 (23.0%)
MS <sup>E</sup>	641	30 (4.7%)	57 (8.9%)	6 (0.9%)	90	3 (3.3%)	1 (1.1%)	2 (2.2%)
MS <sup>E</sup> (N to D)	1254	N/A	470 (37.5%)	N/A	101	88 (87.1%)	13 (12.9%)	75(74.3%)



### ***Improved peptide identification by matching to modified protein sequence database***

Because of the relative low number of peptide identifications in PLGS from MS<sup>E</sup> spectra, we tested whether the 1Da modification of a glycopeptide was affecting the number of identifications, when used as a variable modification while searching the standard TAIR protein sequence database, or when incorporated into the sequence database beforehand. A new database of *Arabidopsis* protein sequences was made in which all putative glycosylation sites (containing an N-P!-S/T consensus site) were present both in the unmodified form as well as with an N to D converted form. For proteins with multiple *N*-glycosylation sites all possible combinations of single and double N to D conversions were included (giving each combination a new accession number,). When spectral data from DDA and MS<sup>E</sup> mode were matched to this modified database (TAIR with N to D conversion), the number of matching peptides for DDA spectra decreased from 1194 to 1107. However, the number of confirmed glycoproteins increased from 190 to 252 (for DDA). For the MS<sup>E</sup> spectra the number of matching peptide masses increased from 641 to 1254, while the number of confirmed glycoproteins increased from 3 to 87. The glycoproteins identified from MS<sup>E</sup> mode in PLGS showed a bias towards glycoproteins with multiple *N*-glycosylation sites (75 out of 87 for MS<sup>E</sup> versus 110 out of 252 for DDA). Apparently, the search algorithm was less efficient for identification of proteins with single glycopeptides using the variable modification (N to D), especially for the MS<sup>E</sup> peptide identification. And the algorithm may rely strongly on identification of multiple peptide hits from a single protein. The scoring algorithm gives a lower score to a peptide match with a (variable) modification than to a non-modified peptide (Li et al., 2009). This is an aspect of the search algorithm scoring that strongly affects the identification rate here, as all our glycopeptides contain such a (variable) modification. This is exemplified by the strong difference in the number of identified peptides in the modified versus the unmodified database, especially those containing the consensus site (513 vs. 374 for DDA, 470 vs. 57 for MS<sup>E</sup>).

### ***Quality control and FDR of glycopeptide identification***

The glycopeptide identification in this procedure heavily relies on the presence of the N to D glycosylation signature. Therefore it was of importance to determine the reliability of deamidation as signature for identification of all glycopeptides released from the hydrazide beads. Different control experiments were used to get insight into the level of spontaneous deamidation that may occur during the different sample preparation steps. Recent studies by Palmisano *et al.* (Palmisano et al., 2012) give evidence of deamidation on asparagine of non-glycan linked N-P!-S/T sites on *E.coli* proteins by PNGase F/A treatment. In our experiments such PNGase F activity would lead to false positives. We tested this potential false positive rate under our own experimental conditions by using proteins isolated from *E.coli* (which does not contain *N*-linked glycoproteins) and being processed in the same way as the protein fractions from plants. Trypsin digested

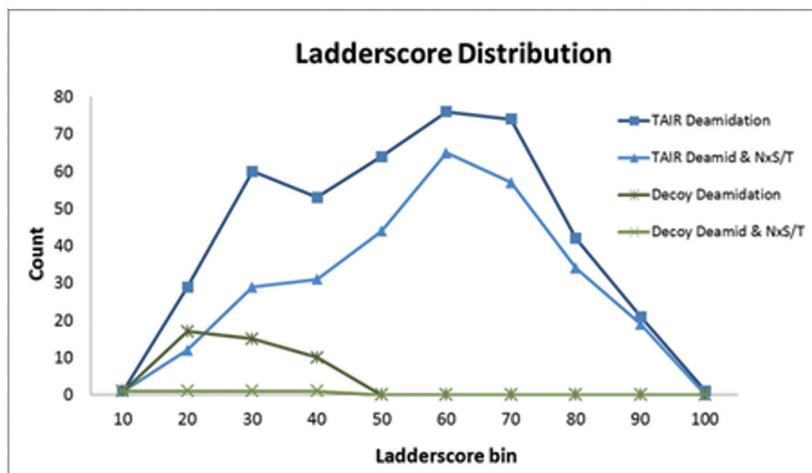
proteins were treated with periodate or PNGase F separately and treated and untreated peptide samples were compared for difference in level of deamidated peptides. Analysis of the *E.coli* peptide spectra showed a detected deamidation level of approximately 8% in the untreated peptide fraction, while both the sodium periodate treatment (used to activate glycans for binding to hydrazide) and the conditions during PNGase treatment did not result in an increase in the level of peptide deamidation (Table 2). The detected level of deamidation could be due to true deamidated peptides, but could also be the result from wrong selection of the parent ion mass (e.g. due to selection of the first C13 isotope peak as parent mass during MS/MS). However, when the criteria of deamidation is combined with N to D conversion of an N-P!-S/T site, the level of *E.coli* matching peptides dropped to below 0.8%. Combined, the results showed that the glycopeptide enrichment procedure did not increase peptide deamidation, while the level of detected deamidation (which could be interpreted as false discovery rate) is estimated to be below 1% when sequence confirmation of N-P!-S/T glycosylation consensus is taken into account at the peptide level. Deamidation of the N-P!-S/T glycosylation consensus site is therefore a reliable signature to identify glycopeptides that were coupled to hydrazide beads and were released by PNGase F. The matching of peptide masses to a given protein database is a function of the quality of the spectra that is obtained for each peptide. In Fig. 2 we plotted the number of peptides in the enriched glycopeptide fraction that showed a match to the database, as function of the spectral quality as given by the ladder score, a spectrum matching quality metric provided by PLGS. The graph shows the results for matching to the conventional *Arabidopsis* TAIR protein sequence database. To determine false positive detection, the same peptide spectral data was matched to a decoy protein database. This decoy protein database was constructed from the normal protein data by inverting and randomizing the protein sequences, until the numbers of N-P!-S/T sites in both databases were identical. When the matching to the *Arabidopsis* TAIR protein sequence database is compared to matching the decoy protein database, results show that for a spectral ladder score higher than 50 only 15 peptides match with the decoy dataset. For the ladder score larger than 50, 175 peptides were identified in *Arabidopsis* TAIR protein sequence database and no matches in the decoy database when also the deamidation and N-P!-S/T criteria was applied (Table 3).

**Table 2.** The number and percentage of identified peptides and proteins from *E.coli* treated with or without PNGase F or NaIO<sub>4</sub> filtered on criteria of N-P!-S/T site and/or deamidation.

Treatment	Peptides				Protein			
	Total	+N-P!-S/T (% of tot)	+deamid (% of tot)	+N-P!-S/T +deamid (% of tot)	total	+N-P!-S/T (% of tot)	+deamid (% of tot)	+N-P!-S/T +deamid (% of tot)
Control	6850	397 (5.8%)	562 (8.2%)	53 (0.8%)	457	167 (36.5%)	228 (49.9%)	44 (9.6%)
PNGase F	5345	285 (4.2%)	229 (4.3%)	31 (0.6%)	383	132 (34.5%)	137 (35.8%)	27 (7.0%)
NaIO <sub>4</sub>	4565	244 (5.3%)	255 (5.6%)	31 (0.7%)	311	111 (35.7%)	135 (43.4%)	28 (9.0%)

**Table 3.** Number of identified N-linked glycopeptides from *Arabidopsis* (leaf, seeding) with ladder score higher than 50 on TAIR normal database or decoy database at different filter levels.

Ladder score > 20	TAIR10	Decoy
Peptides tot	1025	269
Peptides + deamidation	391 (38.1%)	25 (9.3%)
Peptides + deamidation + N-P!-S/T	279 (27.2%)	2 (0.7%)



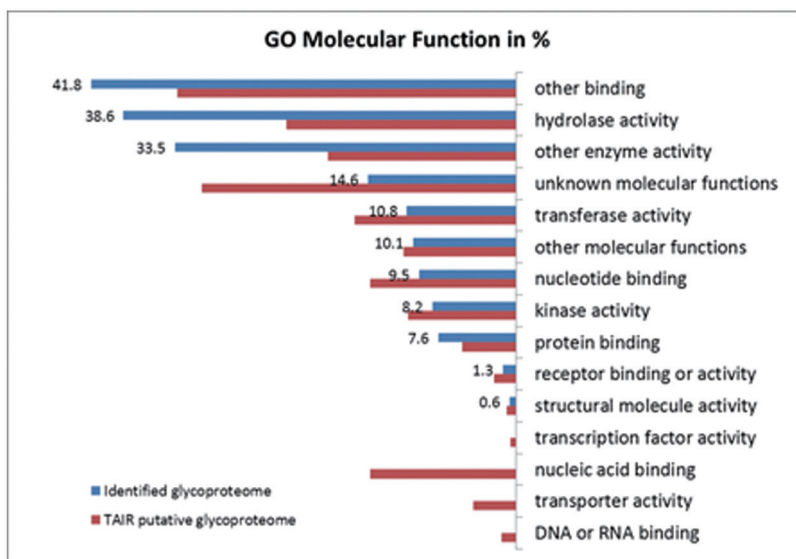
**Figure 2.** Ladder score distribution versus peptide numbers when searching against TAIR protein sequence database, applying different criteria either with deamidation, or deamidation and N-P!-S/T consensus sequence.

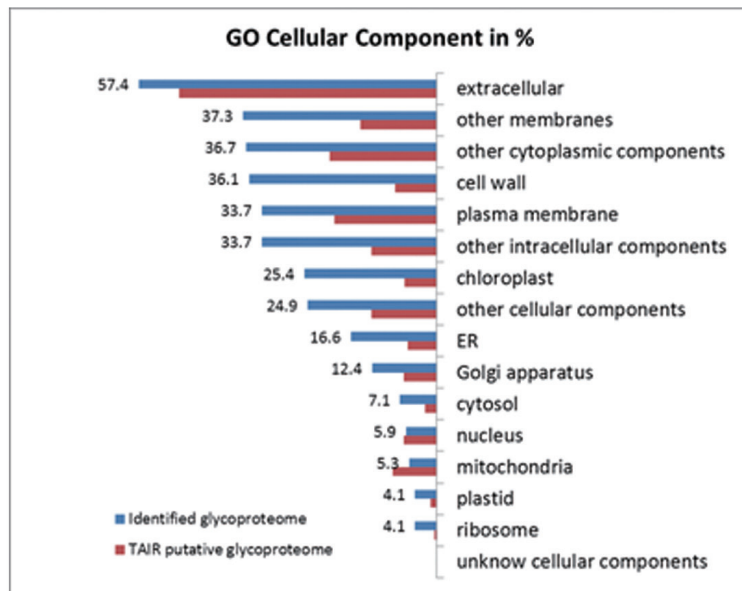
### *The Arabidopsis N-glycoproteome*

In order to obtain the highest number of identified glycopeptides (while meeting the most strict selection criteria) the glycoproteomics results from multiple *Arabidopsis* Col-0 and *cgl1-1* mutant samples were combined (Col-0, *cgl1-1*, DDA, MS<sup>E</sup>, unmodified database, modified database). The 3793 identified peptides belonged to 817 identified proteins (Table S1), of which only 146 were annotated as confirmed glycoproteins after application of our most stringent filter settings (presence of secretion signal and N-P!-S/T site with deamidation of the N-P!-S/T site) (Table S2). Of the 817 identified proteins 28 proteins were identified from peptides which do have an N-P!-S/T site which was also deamidated, but which were not part of the annotated TAIR secreted proteome. Inspection of these proteins show that 11 are predicted to have one or more

trans-membrane domains (TMD), 10 proteins were predicted to have ER or plasma membrane localization according to SLocX (<http://mapman.mpimp-golm.mpg.de/general/slocx/>), and four proteins were predicted to have ER or extracellular location according to TargetP (<http://www.cbs.dtu.dk/services/TargetP/>). This data suggest that these are secreted proteins. Therefore, all these 28 proteins were added to the list of identified and confirmed glycoproteins (Table S2). Proteins which could be identified by this method were mainly receptor kinase, glucosidase, Clavata1 like proteins, peroxidase, phosphatase and glycosyl hydrolase. The list of glycoproteins was sorted according to Gene Ontology (GO) molecular function or GO cellular component and compared to the GO annotation of the full secreted proteome (Fig. 3). Results show that the majority of molecular functions (Fig. 3A) of the identified glycoproteins are hydrolases, other enzyme and other binding activity, transferase activity and kinase receptor binding, similar to the GO molecular function distribution of the putative full secreted *N*-glycoproteome in *Arabidopsis* as described by Song *et al.* (Song *et al.*, 2011). GO annotation of cellular component (Fig. 3B) shows that the putative cellular localization of the majority of identified proteins matches to what is expected of proteins in the secretion pathway, but the distribution over cellular components is different from that of the full glycoproteome (Fig. 3B). This may be due to the small set of identified glycoproteins and redundancy in cellular component annotation for a single protein.

3

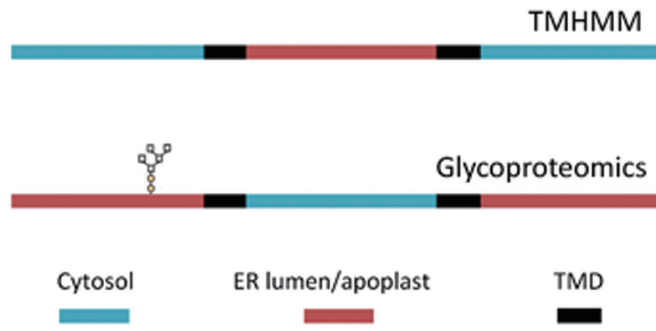




**Figure 3.** GO annotation of all 173 confirmed identified glycoproteins comparing with the whole putative *N*-glycoproteome from TAIR. (Fig. 3A) GO annotation of molecular functions. (Fig. 3B) GO annotation of cellular components.

### *Alternative topology for three membrane proteins identified*

Of the 173 confirmed glycoproteins 67 proteins were membrane proteins with one or more membrane spanning domains. Because the *N*-glycan is attached to the protein in the ER lumen, the *N*-glycan always faces ER lumen, or extracellular space (when located in the plasma membrane). *N*-glycan mapping can therefore be used to verify the predicted topology of membrane proteins. The protein topology based on identification of glycan site occupancy was compared with the topology predicted by TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>). Of 64 from the 67 membrane proteins the topology according to the *N*-glycan site occupancy matched the topology predicted by TMHMM. For three glycoproteins the topology was opposite to that predicted by TMHMM (Table 4). As example, Fig. 4 shows a simplified illustration of the topology of DnaJ domain protein (AT1G61770) as predicted by TMHMM and as predicted from the *N*-glycan position. As our result showed evidence of glycosylation of asparagine at position AA 88 (glycan in Fig. 4 indicates the glycosylation site). Based on the topology of the membrane proteins confirmed by our glycoproteomics experiments two potential N-P!-S/T glycosylation sites actually map to protein domains that map to the cytosolic face of the membrane protein and thus are not true *N*-glycosylation target sites (in AT2G23200 and AT5G14030 C-terminal domain, see Table 4). The information of the topology of a membrane protein provides important insights in the functionality of the different protein domains in relation to the intra- versus extra-cellular location.



**Figure 4.** An example of new topology found. Topology predicted by TMHMM (top) and topology as predicted from the *N*-glycoproteomics (down). The glycan cartoon shows the relative position of glycosylation occupancy. ER = endoplasmic reticulum; TMD = transmembrane domain.

**Table 4.** Three glycoproteins for which the determined topology is different from the TMHMM prediction; the numbers indicate the AA sequence position. The number in brackets shows how many glycosylation sites were mapped/predicted at that domain. Glycosylation site occupancy indicates the number of detected glycosylation sites versus the predicted glycosylation sites from NetNGlyc in the protein.

Gene ID	Gene description	Confirmed topology (mapped/predicted glyco-site)					Glyco-site occupancy Mapped/predicted
		Cytosol	TM	Extra cellular	TM	Cytosol	
AT2G23200	Protein kinase	1–9	10–27	28–401 (1/10)	402–424	425–834 (0/1)	1/11
AT5G14030	Translocon-associated protein beta (TRAPB) family protein	1–9	10–32	33–157 (1/1)	158–180	181–195 (0/1)	1/2
AT1G61770	Chaperone DnaJ domain protein	1–9	10–32	33–121 (1/2); 233–300	122–139; 210–232	140–209	1/2

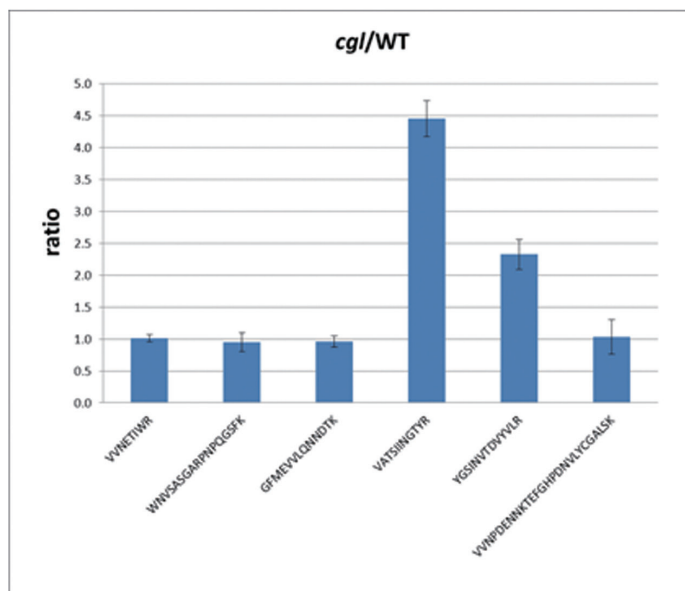
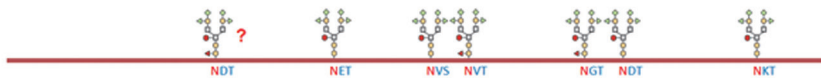
### ***Different glycans on same glycoprotein***

In protein samples from *Arabidopsis* Col-0, some of the *N*-glycopeptides will contain complex *N*-glycans which cannot be released by PNGase F, while in protein samples from the *cgl1-1* mutant *Arabidopsis* all captured *N*-glycopeptides can be cleaved and released from the resin by PNGase F because all glycans in *cgl1-1* are high mannose type. Therefore, signals of PNGase F released glycopeptides which are higher in *cgl1-1* protein extracts compared to Col-0 protein extracts are indicative of a complex *N*-glycan at that position in the Col-0 plant. Similarly, peptide masses with similar signal intensity in Col-0 and *cgl1-1* protein extracts indicate the presence of a similar mannose type *N*-glycan in both mutant and Col-0 plant. Peptide intensity signals were compared between replicate extracts of Col-0 versus *cgl1-1* mutant leaf samples. Intensity ratios were calculated on the average normalized peak intensities per peptide (Table S3). Indeed, some of the identified glycopeptides give a higher signal in *cgl1-1* than Col-0 protein samples. Moreover, peptides belonging to the same glycoprotein with multiple glycosylation sites, may show different ratio's for signal intensity in *cgl1-1*/Col-0 samples, indicating that not all *N*-glycan structures on the same protein are the same and may vary from mannose to complex type *N*-glycans. Besides, the signal intensity in *cgl1-1*/Col-0 samples may be affected by digestion efficiency, binding to the hydrazide beads and PNGase treatment-release efficiency. An example is the identified protein SKU5 (see Fig. 5 and Table S3), which was detected in three replicate leaf samples of both *cgl1-1* and *Arabidopsis* Col-0 after release by PNGase F treatment. SKU5 is distantly related to multiple-copper oxidases, ascorbate oxidase and laccase involved in directional root growth (Sedbrook et al., 2002). Complementary evidence of SKU5 glycosylation comes from SDS-PAGE mobility shift assays by Sedbrook (Sedbrook et al., 2002). They showed a shift of SKU5 mobility in SDS-PAGE after PNGase F treatment, suggestive of glycans without core modification by fucose. They also showed that treatment of SKU5 with Endo H does not lead to a mobility shift on SDS-PAGE, indicating that the PNGase F cleavable *N*-glycans contain the multi-antennary structure which cannot be cleaved by Endo H (Trimble and Tarentino, 1991). In our study, seven out of thirteen putative glycosylation sites on SKU5 were identified (highlighted N-P!-S/T in Fig. 5A). Fig. 5B shows the quantitative peak intensity ratio of the six glycopeptides that were detected in *cgl1-1* and Col-0 samples. Four glycopeptides were present at ratio close to one (highlighted in grey in Fig. 5A), indicating equal binding and equal release by PNGase F in *cgl1-1* and Col-0 samples. However, for two peptides the ratio was higher than one, suggesting reduced release by PNGase F in Col-0 protein samples. This indicates that at these positions the glycans on the glycopeptides have the core of the *N*-glycan modified by fucose for part of the SKU5 protein population in Col-0 plants (highlighted in yellow in Fig. 5A). One glycopeptide in SKU5 (YNDTLVADGIDFETITVHPGK, see Table S2) was only identified in *cgl1-1* mutant samples, as it was not matched to peak intensity data we do not have quantitative information for this peptide (highlighted in dark yellow in Fig. 5A). Combining the results from Sedbrook and ours, the SKU5 *N*-glycan occupancy can be predicted as the cartoon showed in the bottom of Fig. 5A. The heterogeneous glycan modification on

SKU5 could be related to the protein tertiary structure, which may bury some of the *N*-glycans within the protein and thus make them not accessible for fucosyl transferase enzymes in the Golgi apparatus. Alternatively, heterogeneous *N*-glycan modification on the same protein may be indicative of pools of the same protein at different subcellular locations (e.g. an ER localized pool and a pool which has been in the Golgi apparatus)

## AT4G12420: SKU5

```
MDLFKILLLVFFVNISFCFAADPYSFYNFEVSYITASPLGVPQQVIAINGKFFPGPTINVTTNENLVVNVNRKLDE
GLLLHWNGIQRRVSWQDGLGTNCPIPPKWNWTYEFQVKDQIGSFFYPSLSLHFQRASGGFGSFVVNPRAIIPVP
FSTPDGDITVTIGDWYIRNHTALRKALDDGKDLGMPDGLINGKGPYRYNDTLVADGIDFETITVHPGKTYRLRV
SNVGISTSLNFRIOGHNLVLAESEGSYTVQQNYTSLDIHVGQSYSPLVTMDQNASSDYIIVASARVVNETIWRRRV
TGVGILKYTNSKGKAGQLPPGPQDEFDKTFSMNQARSIRWNVSASGARPNPQGSFKYGSINVTDVYVLRNMPPV
TISGKRRTTLNGISFKNPSTPIRLADKLVKVDVYKLDFPKRPLTGPAKVATSIINGTYRGFMEVVLQNNDTKMQS
YHMSGYAFFVVGMDYGEWTENSRGTYNKWDGIARSTIQVYPGAWSAILISLDNPGAWNLRLENLDSWYLGQETYY
RVVNPDENNKTEFGHPDNVLYCGALSKLQKQKVSSASKSIGTSLSMVVMALVMMMLLQH
```



**Figure 5.** *N*-glycan occupancy of SKU5. Fig. 5A: Amino acid sequence of the SKU5 protein with putative *N*-glycosylation sites. Glycosylation sites as predicted by NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) are highlighted with N in red. Peptides with consensus sequences highlighted in grey (*N*-glycans without core fucose modification) or yellow (*N*-glycan with core fucose modification) were detected in our experiments. Peptide with potential core fucose (complex glycan) is highlighted in dark yellow; Bottom: predicted *N*-glycan occupancy of SKU5 based on our and Sedbrook (Sedbrook et al., 2002) results. Fig. 5B: Intensity ratio of six from seven



identified glycopeptides of SKU5 by PNGase F treatment from *Arabidopsis* Col-0 and *cgll-1* samples. Error bar shows the standard deviation of the *cgll-1*/Col-0 intensity ratio among 3 replicates. Four peptides show ratio close to 1 (highlighted in grey in Fig. 5A) and two show ratio >1 (highlighted in yellow in Fig. 5A).

## Discussion

### *Glycopeptide identification*

We used hydrazide selective coupling to beads to enrich glycopeptides from the *Arabidopsis* glycoproteome. PNGase F treatment was used to release the captured peptides. Contrary to mammalian glycoproteins, this procedure has limitations for plant glycoproteins due to the plant specific glycan core modification to fucose on most complex glycans, which cannot be cleaved by PNGase F. Tests with PNGase A (from Roche, MERCK, and EUROPA), which can cleave complex glycan structures, showed that at present this enzyme is not concentrated enough to use in our proteomics experiments. (The PNGase A contains 10,000,000 times lower units of enzymatic activity per microliter, compared to PNGase F, although the units are defined differently by both suppliers.) In all cases by using a practical amount of units the PNGase A elution contained much less peptides compared to PNGase F. Therefore, for alternative assessment of the *Arabidopsis* glycoproteome we used the *cgll-1* mutant, which cannot make complex *N*-glycans, in combination with PNGase F for peptide release. The *cgll-1* mutant has a defective GnT1 protein. The GnT1 activity is required to add fucose or xylose to the core mannose residue. Because this enzyme is located in Golgi apparatus, while glycan coupling to the protein occurs in the ER, it is expected that the actual glyco-site occupancy has not changed in the *cgll-1* mutant, only the downstream processing of the glycan structure. Thus the *cgll-1* plant provides evidence for the occupancy of a glycosylation site just as in the Col-0 plant. Any quantitative difference in detection of the same (glyco)peptide in Col-0 versus *cgll-1* can (to the best of our knowledge) only be interpreted as a difference in downstream processing of the *N*-glycan. In the quantitative comparison of peptide intensity signals between Col-0 and *cgll-1* for the SKU5 protein, four peptides display equal intensity in both samples, while two peptides display a higher signal in the *cgll-1* mutant (see Fig. 5B and Table S3). The latter indicates that for these 2 peptides the glycan form has changed in the Col-0 plant, while the other four have not been processed.

As most glycoproteins only render a single glycopeptide, protein identification was not straight forward. Database search algorithms often rely on matching of multiple peptides per protein. Different methods were used for peptide identification. Most of the glycoproteins were identified by data dependent acquisition (DDA) MS/MS mode. In contrast, in data independent acquisition (MS<sup>E</sup>) mode, there was only limited peptide/protein identification, both with the normal or the modified database which included N to D modifications used as reference (Table 1). In the MS<sup>E</sup> mode, fragmentation

of multiple peptides occurs simultaneously, resulting in mixed fragment ion spectra. Although the quality of these spectra was excellent, the database search produced only low amounts of identified proteins. In the MS<sup>E</sup> mode the algorithm presumes that a protein would render multiple peptides: peptides are assigned to a particular protein with higher confidence, when multiple peptides of the same protein have been detected (Geromanos et al., 2009; Li et al., 2009). This actually results in a bias of peptide identifications from the MS<sup>E</sup> spectra towards proteins from which multiple glycopeptides were detected in the same run and under-representation of glycoproteins with only single glycosylation site (Table 1). Apparently the MS/MS search algorithm for DDA data has less bias in this situation, as more proteins have been detected with single (glyco)peptide.

3

Recently, Zielinska *et al.* mapped 2186 *N*-glycosylation sites in *Arabidopsis* (Zielinska et al., 2012) by using the *N*-glyco-FASP method. In total, 984 out of 1290 proteins were identified with high confidence. In Supplementary Table S1 and S2, we marked the protein overlap between our results and those of Zielinska's. In addition, venn diagrams showing the overlap among total or high confidence identified proteins from us and Zielinska and the secretome from TAIR are displayed in Supplementary Fig. S2A and B. Fig. S2A shows that in both data sets there is a high (background) level of non-secreted proteins, approximately 65% (528 non-secreted vs 817 total) from our results, while 31% (401 out of 1290) from Zielinska's data. Fig. S2B shows that after applying the criteria of ladder score >50, presence of N-P!-S/T site and deamidation as confidence filters, the background detection in our data drops to 16% (28 out of 173) and 27% (267 out of 984) from Zielinska respectively. While in both procedures a considerable amount of non-secreted proteins were detected, the importance of using confidence criteria (consensus site, deamidation, and secretion signal) is demonstrated. We identified approximately five times less glycoproteins than Zielinska (145 versus 717). This may be a reflection of a lower detection sensitivity (MS equipment), or also a lower identification efficiency (software). However, there are 42 proteins uniquely identified by our method. Therefore, both approaches can complement each other to maximise the *N*-glycoprotein identification from *Arabidopsis*.

### ***Contaminants or special secreted proteins?***

Twenty eight proteins were identified from peptide sequences with deamidation of the N-P!-S/T site that were not listed in the TAIR secreted protein list. Our results suggest that these proteins have entered the ER and were modified by *N*-glycans. However, we also identified peptides which do not contain an N-P!-S/T sequence and which do not have the hallmarks of secreted proteins, suggesting that despite intensive washing of the resin with bound glycopeptides, the release with PNGase F apparently still yielded non-glycopeptides. We propose that these peptides were somehow trapped on the resin by the coupled glycopeptides and were therefore co-released with the bound glycopeptides upon PNGase F treatment. For instance, mitochondrial and chloroplast proteins were identified by multiple peptides with high spectrum scores, in multiple

samples, indicating that these proteins are truly eluted from the hydrazide resin by PNGase F treatment. However, after applying the stringent filters for secreted protein, N-P!-S/T, deamidation and ladder score >50, all these proteins were removed from the list, indicating that the combination of these criteria is stringent enough to distinguish contaminant identifications from glyco-protein identifications.

Our control experiments indicate that the level of spontaneous deamidation under the experimental conditions is low (Table 2). Nevertheless, the spontaneous deamidation indeed occurs by chemical and enzymatic treatment during sample preparation, and it is difficult to discriminate between the true and false positive deamidation, as described by Palmisano *et al.* (Palmisano *et al.*, 2012). Alternatively, remaining low levels of deamidation without PNGase treatment could be the result of false peak selection: in longer peptides the intensity of the first C13 peak is higher than that of the C12 peak, in which case in the DDA mode the acquisition software may select the C13 peak instead of the C12 peak as precursor ion mass. In such case the mass of the (selected) precursor ion will have 1Da increase and may be wrongly assigned as being modified. Therefore, stringent methods and filters need to be applied to keep the false positive discovery rate as low as possible as discussed before.

### ***Perspectives and Future plans***

The *Arabidopsis* TAIR protein sequence database provides a list of 5310 putative secreted proteins based on signalP and TargetP ([ftp://ftp.arabidopsis.org/home/tair/Proteins/Properties/TargetP\\_analysis.TAIR](ftp://ftp.arabidopsis.org/home/tair/Proteins/Properties/TargetP_analysis.TAIR)), of which 3597 proteins actually contain one or more N-P!-S/T consensus glycosylation sites. Of these potential 3597 glycoproteins only 173 proteins were confirmed by 330 glycopeptides identified in the experiments described in this paper. Clearly at present the technology is mostly limited by the peptide identification and great advances could be made with improved detection sensitivity and improvement in the software used to interpret MS<sup>E</sup> spectra in this context. The data generated in these experiments can be used to enhance predictors for *N*-glycosylation site occupancy, subcellular localization and topology of membrane proteins. Glycoproteomics as described in this paper could form an alternative to membrane proteomics, as many of the important proteins that interact with extracellular signals during plant development or extracellular signals from invading pathogens are located on the plasma membrane and are glycosylated. However, to optimally use glycoproteomics in such context it would be most useful to use a mutant background which prevents formation of complex *N*-glycans like the *cgl1-1* mutant in *Arabidopsis*. In future research we will apply this technology to perform extensive comparison of the *N*-glycoproteome (as can be released by PNGase F) of *Arabidopsis* Col-0 and *cgl1-1* mutant, and comparison of Col-0 and another glycosylation mutant *alg3-2* (Henquet *et al.*, 2008).

## **Acknowledgements**

This project was carried out within the research program of the Centre of BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

## References

- Chang CS, Chang KP** (1986) Monoclonal antibody affinity purification of a *Leishmania* membrane glycoprotein and its inhibition of leishmania-macrophage binding. *Proc Natl Acad Sci U S A* **83**: 100–104
- Gattiker A, Gasteiger E, Bairoch A** (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl Bioinformatics* **1**: 107–108
- Geromanos SJ, Vissers JPC, Silva JC, Dorschel CA, Li GZ, Gorenstein M V, Bateman RH, Langridge JI** (2009) The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**: 1683–1695
- Grunewald S, Matthijs G, Jaeken J** (2002) Congenital disorders of glycosylation: a review. *Pediatr Res* **52**: 618–624
- Hagglund P, Bunkenborg J, Elortza F, Jensen ON, Roepstorff P** (2004) A new strategy for identification of *N*-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation. *J Proteome Res* **3**: 556–566
- Hagglund P, Matthiesen R, Elortza F, Hojrup P, Roepstorff P, Jensen ON, Bunkenborg J** (2007) An enzymatic deglycosylation scheme enabling identification of core fucosylated *N*-glycans and *O*-glycosylation site mapping of human plasma proteins. *J Proteome Res* **6**: 3021–3031
- Henquet M, Lehle L, Schreuder M, Rouwendal G, Molthoff J, Helsper J, Van Der Krol S, Bosch D** (2008) Identification of the gene encoding the  $\alpha$ 1,3-mannosyltransferase (ALG3) in *Arabidopsis* and characterization of downstream *N*-glycan processing. *Plant Cell Online* **20**: 1652
- Hirabayashi J** (2004) Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconj J* **21**: 35–40
- Kaji H, Yamauchi Y, Takahashi N, Isobe T** (2007) Mass spectrometric identification of *N*-linked glycopeptides using lectin-mediated affinity capture and glycosylation site-specific stable isotope tagging. *Nat Protoc* **1**: 3019–3027
- Kang JS, Frank J, Kang CH, Kajiura H, Vikram M, Ueda A, Kim S, Bahk JD, Triplett B, Fujiyama K, et al** (2008) Salt tolerance of *Arabidopsis thaliana* requires maturation of *N*-glycosylated proteins in the Golgi apparatus (vol 105, pg 5933, 2008). *Proc Natl Acad Sci U S A* **105**: 7893

**Kubota K, Sato Y, Suzuki Y, Goto-Inoue N, Toda T, Suzuki M, Hisanaga S, Suzuki A, Endo T** (2008) Analysis of glycopeptides using lectin affinity chromatography with MALDI-TOF mass spectrometry. *Anal Chem* **80**: 3693–3698

**Lee RT, Lauc G, Lee YC** (2005) Glycoproteomics: protein modifications for versatile functions. Meeting on glycoproteomics. *EMBO Rep* **6**: 1018–1022

**Lerouge P, Cabanes-Macheteau M, Rayon C, Fischette-Laine AC, Gomord V, Faye L** (1998) *N*-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol Biol* **38**: 31–48

**Li GZ, Vissers JPC, Silva JC, Golick D, Gorenstein M V, Geromanos SJ** (2009) Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**: 1696–1719

**Liebming E, Veit C, Mach L, Strasser R** (2010) Mannose trimming reactions in the early stages of the *N*-glycan processing pathway. *Plant Signal Behav* **5**: 476–478

**Palmisano G, Melo-Braga MN, Engholm-Keller K, Parker BL, Larsen MR** (2012) Chemical deamidation: a common pitfall in large-scale *N*-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res* **11**: 1949–1957

**Pan S, Chen R, Aebersold R, Brentnall TA** (2011) Mass spectrometry based glycoproteomics--from a proteomics perspective. *Mol Cell Proteomics* **10**: R110003251

**Parodi AJ** (2000) Role of *N*-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem J* **348 Pt 1**: 1–13

**Qiu R, Regnier FE** (2005) Use of multidimensional lectin affinity chromatography in differential glycoproteomics. *Anal Chem* **77**: 2802–2809

**Rayon C, Lerouge P, Faye L** (1998) The protein *N*-glycosylation in plants. *J Exp Bot* **49**: 1463–1472

**von Schaewen A, Sturm A, O'Neill J, Chrispeels MJ, Vonschaewen A, Sturm A, O'Neill J, Chrispeels MJ** (1993) Isolation of a mutant *Arabidopsis* plant that lacks *N*-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex *N*-linked glycans. *Plant Physiol* **102**: 1109–18

**Schollen E, Grunewald S, Keldermans L, Albrecht B, Korner C, Matthijs G** (2005) CDG-Id caused by homozygosity for an ALG3 mutation due to segmental maternal isodisomy UPD3(q21.3-qter). *Eur J Med Genet* **48**: 153–158

**Sedbrook JC, Carroll KL, Hung KF, Masson PH, Somerville CR** (2002) The *Arabidopsis* SKU5 gene encodes an extracellular glycosyl phosphatidylinositol-anchored glycoprotein involved in directional root growth. *Plant Cell* **14**: 1635–1648

**Song W, Henquet MGL, Mentink RA, van Dijk AJ, Cordewener JHG, Bosch D, America AHP, Van der Krol AR** (2011) N-glycoproteomics in plants: Perspectives and challenges. *J Proteomics* **74**: 1463–1474

**Spiro RG** (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**: 43R–56R

**Strasser R, Stadlmann J, Svoboda B, Altmann F, Glossl J, Mach L, Glössl J, Mach L, Glossl J, Mach L** (2005) Molecular basis of N-acetylglucosaminyltransferase I deficiency in *Arabidopsis thaliana* plants lacking complex N-glycans. *Biochem J* **387**: 385–391

**Sun L, Eklund EA, Chung WK, Wang C, Cohen J, Freeze HH** (2005) Congenital disorder of glycosylation id presenting with hyperinsulinemic hypoglycemia and islet cell hyperplasia. *J Clin Endocrinol Metab* **90**: 4371–4375

**Tian Y, Zhou Y, Elliott S, Aebersold R, Zhang H** (2007) Solid-phase extraction of N-linked glycopeptides. *Nat Protoc* **2**: 334–339

**Trimble RB, Tarentino AL** (1991) Identification of distinct endoglycosidase (endo) activities in *Flavobacterium meningosepticum*: endo F<sub>1</sub>, endo F<sub>2</sub>, and endo F<sub>3</sub>. Endo F<sub>1</sub> and endo H hydrolyze only high mannose and hybrid glycans. *J Biol Chem* **266**: 1646–1651

**Uematsu R, Furukawa J, Nakagawa H, Shinohara Y, Deguchi K, Monde K, Nishimura SI** (2005) High throughput quantitative glycomics and glycoform-focused proteomics of murine dermis and epidermis. *Mol Cell Proteomics* **4**: 1977–1989

**Varki A** (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**: 97–130

**Varki A** (1999) *Essentials of glycobiology*. Cold Spring Harbor Laboratory Pr

**Wang Y, Wu S, Hancock WS** (2006) Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap—Fourier transform mass spectrometry. *Glycobiology* **16**: 514

**Yang ZP, Hancock WS** (2005) Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography. *J Chromatogr A* **1070**: 57–64

**Zhang H** (2007) Glycoproteomics using chemical immobilization. *Curr Protoc Protein Sci* Chapter 24: Unit 24 3

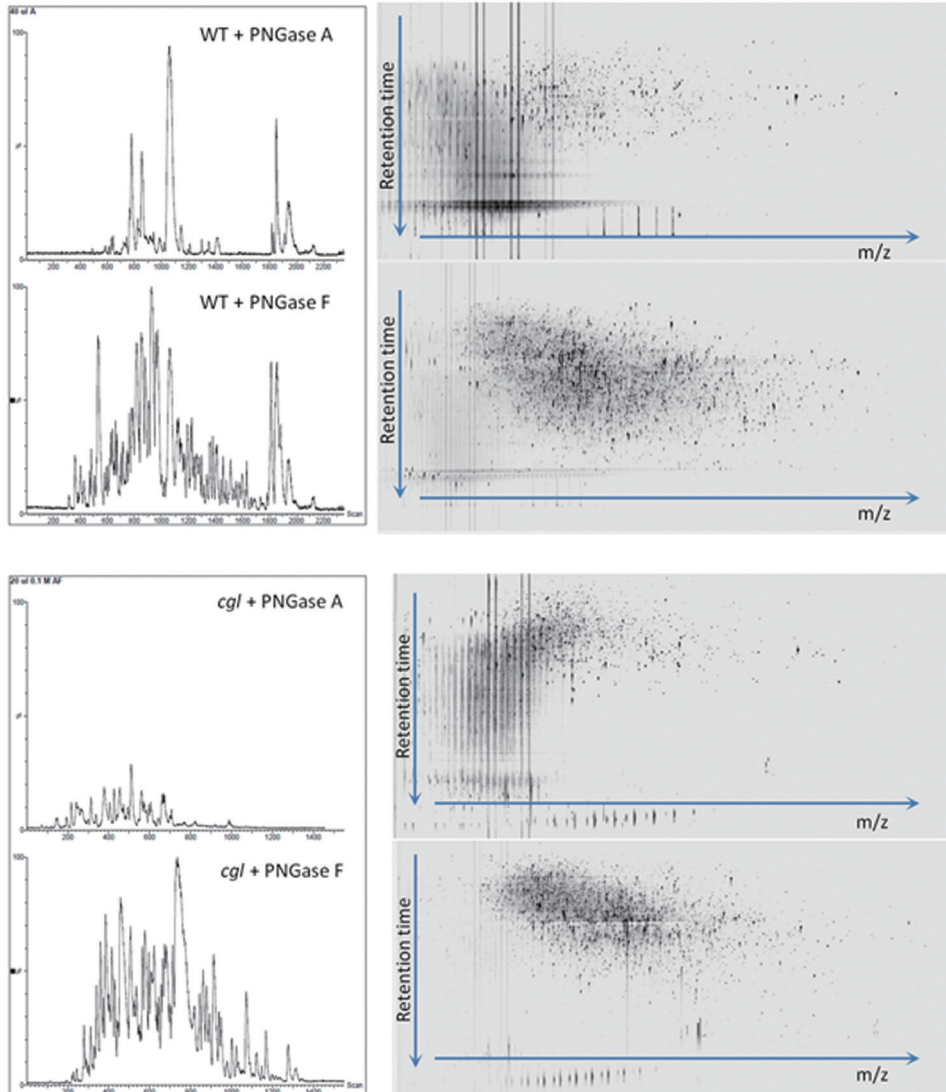
**Zhang H, Li X-JJ, Martin DB, Aebersold R** (2003) Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **21**: 660–6

**Zielinska DF, Gnad F, Schropp K, Wisniewski JR, Mann M** (2012) Mapping *N*-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. *Mol Cell* **46**: 542–548

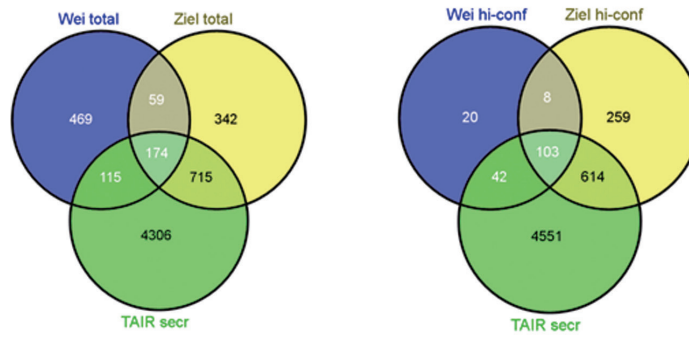
**Zielinska DF, Gnad F, Wisniewski JR, Mann M, Wiśniewski JR, Mann M** (2010) Precision Mapping of an In Vivo *N*-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* **141**: 897–907



## Supplementary Data



**Figure S1.** Comparison of base peak chromatograms (left) and their 2D visualization (right) from PNGase A and PNGase F released *Arabidopsis* Col-0 (Fig. S1A) or *cgl* (Fig. S1B) extracts (of fraction F1 from 2D-LC-MS<sup>E</sup>).



**Figure S2.** (Fig. S2A) Venn diagram shows the overlap among total identified proteins from us and Zielinska and the whole secretome from TAIR; (Fig. S2B) Venn diagram shows the overlap among the high confidence identified proteins (filtered on criteria: presence and deamidation of N-P!-S/T) from us and Zielinska and the whole secretome from TAIR.

## Supplementary Data

Supplementary data to this article can be found online at:

[https://1drv.ms/f/s!Asszhx\\_to9U8hxXix8NpPWFDJNdA](https://1drv.ms/f/s!Asszhx_to9U8hxXix8NpPWFDJNdA)

or by scanning the QR code:



**Table S1.** Total list of identified proteins from *Arabidopsis* leaf and seedling samples with DDA and MS<sup>E</sup> using TAIR normal and modified database containing all the peptides, not filtered but with listed N-P!-S/T consensus site and potential modifications. Proteins were marked with the presence of trans-membrane domain (TMD), signal peptide (SignalP) or GPI-anchored protein (GPI-AP). Overlaps of total identified proteins from us and Zielinska were marked. A confidence filter on the DDA and MS<sup>E</sup> should be applied in order to provide a lower false positive detection rate. Based on previous experience we apply a lower threshold of 50 for the ladder score in DDA and a minimal score of 6 for MS<sup>E</sup>, in order to filter for high confidence individual peptide-spectrum matches.

**Table S2.** High confidence list of identified N-glycoproteins from *Arabidopsis* leaf and seedling samples with DDA and MS<sup>E</sup> using TAIR normal and modified database; On DDA, peptides were filtered by ladder score > 50, deamidation on N as modification or D for N as substitution and N-P!-S/T consensus site; On MS<sup>E</sup>, peptides were filtered by score > 6 and D-P!-S/T by using TAIR modified database. Proteins were marked for the presence of TMD, SignalP or GPI-AP as in Table S1. Also overlap of high confidence identified proteins from us and Zielinska (reported as a separate table in (Zielinska et al., 2012)) was marked.

**Table S3.** Peak intensity data of Fig. 5 from six SKU5 peptides. For the identified peptides the peak areas were quantified in three replicate extracts from *Arabidopsis* Col-0 and *cgl1-1* mutant. Normalized peak area is displayed as average plus standard deviation over three replicates.



# Chapter 4

## **Comparative proteomics of *Arabidopsis* plants with different susceptibility to *Phytophthora brassicae* identifies Cysteine-rich protein kinase 37 as resistance component**

Wei Song <sup>a, b</sup>, Jan H.G. Cordewener <sup>b</sup>, Harro Bouwmeester <sup>a</sup>,  
Alexander R. van der Krol <sup>a</sup>, Antoine H.P. America <sup>b, \*</sup>

<sup>a</sup>Laboratory of Plant Physiology, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>b</sup>Plant Research International, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

This work was carried out in collaboration with  
Klaas Bouwmeester <sup>c, d</sup> and Francine Govers <sup>c</sup>

<sup>c</sup>Laboratory of Phytopathology, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>d</sup>Plant-Microbe Interactions, Department of Biology, Utrecht University,  
Utrecht, The Netherlands

## Abstract

*Arabidopsis* Col-0 is resistant to infection by *Phytophthora brassicae* isolate strain HH. In contrast, transgenic *Arabidopsis* Col-0 expressing the *Phytophthora infestans* RXLR effector *ipiO* (35S-*ipiO1*) and Col-0 with a T-DNA insertion knockout in *LecRK-I.9* gene (*lecrk-I.9*) both show a gain of susceptibility to *P. brassicae*. Here we investigated whether the shared phenotype of the 35S-*ipiO1* and *lecrk-I.9* lines relates to a consistent altered plasma membrane (PM) protein composition using label-free comparative proteomics. Two-dimensional LC-MS of tryptic digests of PM enriched protein fractions from *Arabidopsis* Col-0, 35S-*ipiO1* and *lecrk-I.9* lines resulted in 2151 identified proteins of which 613 are putative PM-proteins either with a predicted transmembrane domain (TMD) or glycosylphosphatidylinositol (GPI) anchor. Quantitative analysis of the peptide profiles revealed 99 peptide features with more than two-fold difference in peptide mass intensity between Col-0 and the 35S-*ipiO1* and *lecrk-I.9* lines. Because none of these peptides were derived from proteins identified in ProteinLynx™ Global Server (PLGS), MS-Fit was used as alternative protein identification method, using strict filter criteria to reduce false discovery rate. This allowed us to link 25 of the 99 peptide features to 17 plasma membrane proteins. One of these proteins is Cysteine-rich protein kinase 37 (CRK37) for which the peptide signals were higher in *Arabidopsis* Col-0 than in 35S-*ipiO1* and *lecrk-I.9*. In contrast, the transcript level of *CRK37* was not significantly different between Col-0 and 35S-*ipiO1* and *lecrk-I.9*, suggesting that *CRK37* protein level is regulated post-transcriptionally. Subsequently, the role of *CRK37* in resistance to *P. brassicae* was proven by showing that an *Arabidopsis crk37* T-DNA insertion mutant displays a gain of susceptibility to *P. brassicae* infection. In conclusion, although label-free comparative proteomics is not without limitations, it is a powerful tool to complement transcriptomics approaches in plant-pathogen interaction studies.

## 4

## Keywords

Legume-like lectin receptor kinase (LecRK), RXLR effector, Plasma membrane proteomics, *Arabidopsis*, *Phytophthora brassicae*, label-free comparative proteomics, LC-MS, plant-pathogen interactions

## Abbreviations

PM, Plasma membrane; UPLC, ultra-performance liquid chromatography; ESI, electrospray ionization; Q-TOF, quadrupole time-of-flight; DDA, data dependent acquisition; DIA, data independent acquisition

## Introduction

*Arabidopsis* Col-0 is resistant against the oomycete pathogen *Phytophthora brassicae* HH, but can be genetically modified to gain susceptibility (Roetschi et al., 2001). This allows the use of *Arabidopsis* as a model to study aspects of the *Phytophthora*-hostplant interaction. The cell wall (CW) plasma-membrane (PM) continuum plays an important role in the first line of defense against invading pathogens in plants. Legume-like lectin receptor kinases (LecRKs) are a subclass of receptor-like kinases (RLKs), with an extracellular lectin domain. There are 45 LecRKs in *Arabidopsis*, some of which have been shown to have a role in plant-pathogen interactions (Bouwmeester et al., 2011b; Desclos-Theveniau et al., 2012; Singh et al., 2012; Wang et al., 2016). LecRK-I.9 of *Arabidopsis* was shown to be involved in stabilizing the CW-PM continuum in *Arabidopsis* (Mellersh and Heath, 2001; Gouget et al., 2006). Indeed, disruption of LecRK-I.9 by a T-DNA insertion resulted in a gain of susceptibility to infection by *P. brassicae* (Bouwmeester et al., 2011b). LecRK-I.9 was also recently identified as an extracellular ATP (eATP) receptor (Choi et al., 2014). Peptides containing the Arg-Gly-Asp (RGD) motif bind to LecRK-I.9 and are involved in mediating cell adhesion in mammalian cells and contact between the cell wall and plasma-membrane in plants (Dsouza et al., 1991; Takagi, 2004; Bouwmeester et al., 2011a). The RGD motif is also present in the *Phytophthora infestans* effector protein IPI-O and *Arabidopsis* Col-0 plants expressing *ipiO* (35S-*ipiO1*) show a similar gain of susceptibility to infection by *P. brassicae* as the *lecrk-I.9* insertion mutant (Bouwmeester et al., 2011b). Because of the contrasting susceptibility to *P. brassicae* infection between *Arabidopsis* Col-0 and the *lecrk-I.9* and 35S-*ipiO1* plants and because both LecRK-I.9 and IPI-O act at or near the PM of the plant cell, we suspected that the different response to *P. brassicae* may relate to a different state of the PM proteome of Col-0 and *lecrk-I.9* and 35S-*ipiO1* plants. Differences in PM proteins may be identified by comparative PM proteomics and differentially expressed proteins are potential candidates for having a role in the defense against the oomycete pathogen. Thus, comparative PM proteomics of these three lines may lead to the identification of factors involved in plant-pathogen interactions. To achieve this we applied label-free proteomics on protein fractions enriched for PM proteins isolated from mature, uninfected Col-0, *lecrk-I.9* or 35S-*ipiO1* *Arabidopsis* leaves. Despite limitations in linking differential peptide mass features to proteins, eventually 17 PM proteins with differential peptide abundance were identified. One of these proteins was the cysteine-rich protein kinase CRK37, which was chosen for further validation studies because of its specific gene expression and protein abundance profile and because of a potential signaling role of kinases in plant-pathogen interactions. An *Arabidopsis* mutant with a T-DNA insertion in *CRK37* gene indeed showed a gain of susceptibility for infection by *P. brassicae*, indicating a role for CRK37 in the defense against *P. brassicae*. Results show the potential of comparative proteomics to identify new factors involved in plant-pathogen interactions.

## Materials and methods

### *Plant material and growth conditions*

The *Arabidopsis* T-DNA insertion mutant *lecrk-1.9* (SALK\_042209) and transgenic line expressing *ipiO1* (35S-*ipiO1*) were previously described by Bouwmeester *et al.* (Bouwmeester *et al.*, 2011b). *Arabidopsis* T-DNA insertion mutant *crk37* (SALK\_004923) was obtained from the European Arabidopsis Stock Center (<http://arabidopsis.info/>). This mutant line was previously checked by PCR for T-DNA insertion (Bourdais *et al.*, 2015). *Arabidopsis thaliana* seeds were sown on 9 cm 0.8% Daishin agar (Sigma) in Petri dishes and cold (4°C) treated in the dark for 2 days to promote uniform germination. Five days after germination seedlings were transferred to soil and plants were grown for 35 days in a climate chamber at 21°C with 75% relative humidity (RH) and a 12h photoperiod.

### *Plasma membrane enrichment and plasma membrane protein isolation*

For each genotype three biological replicates of 30 gram rosettes were harvested, pooled and frozen in liquid nitrogen. PM isolation was carried out on these three biological replicates, by homogenizing the 30 gram of frozen, pulverized leaf material with 50 ml of ice-cold homogenization buffer (0.3 M sucrose, 50 mM 3-(N-morpholino) propanesulfonic acid (MOPS)-KOH, 5 mM Na-EDTA, 5 mM ascorbic acid, 5 mM dithiothreitol (DTT), 0.55% insoluble polyvinylpyrrolidone (PVPP), 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 7.5) using a kitchen blender (3 × 20 sec strokes). The homogenate was filtered through 0.2 mm pore size Nylon and further clarified by centrifugation at 10,000 g for 30 min. Subsequently the supernatant was transferred into a ultracentrifuge tube (SS-34) and centrifuged at 100,000 g for 1 hr at 4°C (JA20, Beckman Ultra-centrifuge). The microsomal pellet was re-suspended in 2.5 mL re-suspension medium (0.3 M sucrose, 50 mM MOPS-KOH, 5 mM Na-EDTA, 1 mM DTT, pH 7.5) and three consecutive two-phase extractions were performed as described by Alexandersson *et al.* (Alexandersson *et al.*, 2004) to enrich for plasma membranes (PM). Therefore, 2.25 g of microsomal fraction was added to 6.75 g ATPS (6.2% Dextran T500, 6.2% PEG 3350, 2 mM KCl, 0.33 M sucrose), mixed well and centrifuged at 2,000 g for 10 min in order to allow separation of the Dextran and PEG phase. In the upper PEG phase plasma membranes will be enriched, and the lower Dextran phase will be enriched in intracellular membranes. The upper phase was added to a fresh pretreated lower phase of ATPS, while the lower phase was added to a fresh PEG top-layer. After centrifugation the two-phase extraction was repeated once more and the combined upper phases were added to 50 ml of resuspension buffer (0.3 M sucrose, 5mM potassium phosphate, 0.1 mM Na-EDTA, 1mM DTT, pH7.8). After ultra-centrifugation at 100,000g at 4°C for 1 hr the PM pellet was resuspended in 100 µL resuspension buffer for SDS-PAGE analysis. For peptide analysis the PM pellet was washed three times with chloroform/methanol/MQ (1:4:5) for removal of lipids. Eventually, the protein pellets were dried and dissolved in 50 µL protein dissolving



buffer (0.1 M Tris-HCl (pH 8.0), 8 M urea and 5 mM DTT). After alkylation by iodoacetamide (IAA), 4 volumes 0.1 M ammonium bicarbonate buffer (pH 8.3) was added to dilute urea to 2 M for protein digestion. Protein concentration was measured by Bradford assay (Bradford, 1976). Protein was digested with Porcine, TPCK treated, trypsin (proteomics grade, Sigma T6567) (trypsin:protein=1:50) by overnight incubation at 37°C. Digestion was terminated by adding trifluoroacetic acid (TFA) to a final 0.1% v/v. After digestion, a reverse phase C18 solid phase extraction (SPE) column (Supelco) was applied for peptide purification. Peptides were eluted with 1 mL of 60% acetonitrile (ACN), 0.1% TFA. The purified peptides were dried by vacuum centrifugation and dissolved in 20 µL of 0.1 M ammonium formate (pH 10) for 2D nano LC-MS analysis.

### ***Pathogen growth and infection assays***

*P. brassicae* strain HH was grown at 18°C on 10% V8-juice agar plates, as described by Bouwmeester *et al.* (Bouwmeester *et al.*, 2011b). Inoculation was performed by placing plugs of young mycelium (Ø 5 mm) on the abaxial leaf surface. Inoculated plants were kept in trays covered with lids to maintain a high humidity and placed in the dark, and placed in a growth chamber with a 16 h photoperiod at 18°C and a RH of 75%. The first day the trays were kept in the dark. After two days the mycelial plugs were removed from plants to stop the facilitation of nutrition from the agar medium. Disease symptoms were scored three days after inoculation. Infection efficiencies (IEs) were calculated as percentages successful infected leaves relative to the total number of inoculations.

### ***Peptide analysis using 2D nano LC-MS/MS and 2D nano LC-MS<sup>E</sup>***

For high-resolution separation of the complex PM protein digests a nano ACQUITY™ 2D UPLC system (Waters Corporation, Manchester, UK) was used employing orthogonal reversed phase separation at high and low pH, respectively. With this 2D set up, the pool of peptides was eluted from the first dimension XBridge™ C18 trap column (in 20 mM ammonium formate pH 10) using a discontinuous step gradient of seven fractions 12%, 15%, 18%, 20%, 25%, 35% and 65% ACN. For the second dimension we used a BEH C18 column (75 µm × 25 cm, Waters, UK) eluting with 65 min linear gradient from 3 to 40% ACN (in 0.1 % FA) at 200 nL/min. The eluting peptides were on-line injected into a SYNAPT™ G1 Q-TOF MS instrument (Waters Corporation, Manchester, UK) using a nanospray device coupled to the second dimension column output. The SYNAPT MS was operated in positive mode with [Glu<sup>1</sup>] fibrinopeptide B (1 pmol/µL; Sigma) as reference (lock mass) sampled every 30s. Accurate LC-MS data were collected with the SYNAPT operating in either the MS/MS or MS<sup>E</sup> mode for data-dependent acquisition (DDA) or data-independent acquisition (DIA) using low (6 eV) and elevated (ramp from 15 to 35 eV) energy spectra every 0.6 s over a 140-1900 m/z range, respectively. DDA was performed by peptide fragmentation on the three most intense multiply charged ions that were detected in the MS survey scan (0.6 s) over a

300-1400 m/z range and a dynamic exclusion window of 60 s with an automatically adjusted collision energy based on the observed precursor m/z and charge state.

### ***RNA isolation and RNA-seq analysis***

Total RNA was isolated from 4-week old *Arabidopsis* leaves using a NucleoSpin RNA Plant kit (Macherey-Nagel). RNA was checked for concentration and quality using a Qubit fluorometer. Libraries were constructed for each individual RNA sample using standard construction protocols. RNA sequencing (RNA-seq) was performed on an Illumina HiSeq 2500 sequencing system. Samples were assigned to one parallel run that included 9 samples (3 biological replicates per *Arabidopsis* line). RNA sequencing and data analysis was performed at Bioinformatics facility of Plant Research International (Wageningen UR). Reads were expressed in units of Transcripts per Million (TPM), which were normalized by calculating the Z-score.

### ***Database search for protein identification***

LC-MS/MS and MS<sup>E</sup> data were processed using ProteinLynx™ Global Server software (PLGS version 2.5, Waters Corporation) and the resulting list of masses containing all the fragment ion information was searched against the TAIR protein sequence database downloaded from The Arabidopsis Information Resource (TAIR, <https://www.arabidopsis.org/>). For matching MS<sup>E</sup> data to predicted peptide spectra the following settings were used: minimum fragment ion matches per peptide: four, per protein: seven, minimum one peptide per protein match and a false discovery threshold of four percent. The false discovery rate (FDR) was determined automatically in PLGS by searching the randomized TAIR protein sequence database. Peptide and fragment tolerance was set to automatic: the software determines tolerance thresholds based on the actual resolution and stability in the MS data (Geromanos et al., 2009; Li et al., 2009). In practice this resulted in mass tolerances below 10 ppm for precursor and below 30 ppm for fragment ions. Carbamido-methylation of cysteine was set as fixed modification, and oxidation of methionine (M) and deamidation on asparagine (N) as (enriched) variable modification.

For DDA analysis the peptide tolerance was set to 30 ppm and a fragment tolerance of 0.05 Da. Carbamidomethylation (Cys) was used as fixed modification and Deamidation (NQ), and Oxidation (M) as variable modifications. The AutoMod option was applied as secondary search to the database search results with a maximum of one missed trypsin cleavage and non-specific secondary digest reagent were chosen. Finally, the DDA and MS<sup>E</sup> protein identifications were merged in Excel. Gene Ontology (GO) annotation was performed by using the tools from TAIR (<https://www.arabidopsis.org/>) (Berardini et al., 2004). Further filters and tools were applied to remove the non-membrane proteins such as presence of trans-membrane domain (TMD) (TMHMM: <http://www.cbs.dtu.dk/services/TMHMM/>), or GPI-anchor (GPI-SOM: <http://gpi.unibe.ch/>). Protein domain information was manually checked with Pfam (<http://pfam.xfam.org/>) and SMART (<http://smart.embl-heidelberg.de/>).

### ***Protein quantification in Progenesis LC-MS™***

Protein quantification from raw MS<sup>E</sup> data was done in Progenesis LC-MS™ version 4.0 (Nonlinear Dynamics, UK) using lock-spray and dead time correction as mass calibration. Progenesis performed extraction of ion features such as mass, charge, intensity, retention time, etc, after retention time alignment and peak detection of all the samples. Total peptide mass peak intensities were normalized among different samples by the algorithm in Progenesis. Further, statistical analysis was done in Progenesis with all the matched features from different samples and experiments. Features were clustered, grouped and tagged according to fold change, similarity and *P*-value for significance. Features which showed significantly differential peptide mass intensity among samples were selected for matching peptide/protein identification from PLGS. The non-matched differential features can be exported as inclusion list to run additional dedicated DDA for protein identification. Eventually, identified proteins, which showed significant differential peptide mass intensity patterns among samples were exported into Excel for further analysis.

### ***Additional protein identification by MS-Fit***

Mass features which showed significant differential peptide mass intensity profiles among samples were matched to protein identifications from PLGS in Progenesis as described above. Eventually, the non-matched features were searched with the help of MS-Fit (<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard>). MS-Fit only uses precursor ion (MS1) information for matching mass features to peptide sequences without taking into account the fragment ion information from MS/MS (MS2) data. We first clustered differential features in Progenesis, so that features with similar peptide mass intensity profiles were loaded as a group into MS-Fit search. To reduce the FDR a minimum of three peptides matching per protein was used as a filter in MS-Fit. Used database: SwissProt 2013-6-27; taxonomy: *Arabidopsis*; constant modifications: carbamidomethyl (C); possible modifications: oxidation (M); mass tolerance: 20 ppm. For proteins identified by MS-Fit, further filters and tools were applied to remove the non-membrane proteins such as presence of trans-membrane domain (TMHMM: <http://www.cbs.dtu.dk/services/TMHMM/>), or GPI-anchor (GPI-SOM: <http://gpi.unibe.ch/>). Protein domain information was manually checked with Pfam (<http://pfam.xfam.org/>) and SMART (<http://smart.embl-heidelberg.de/>).

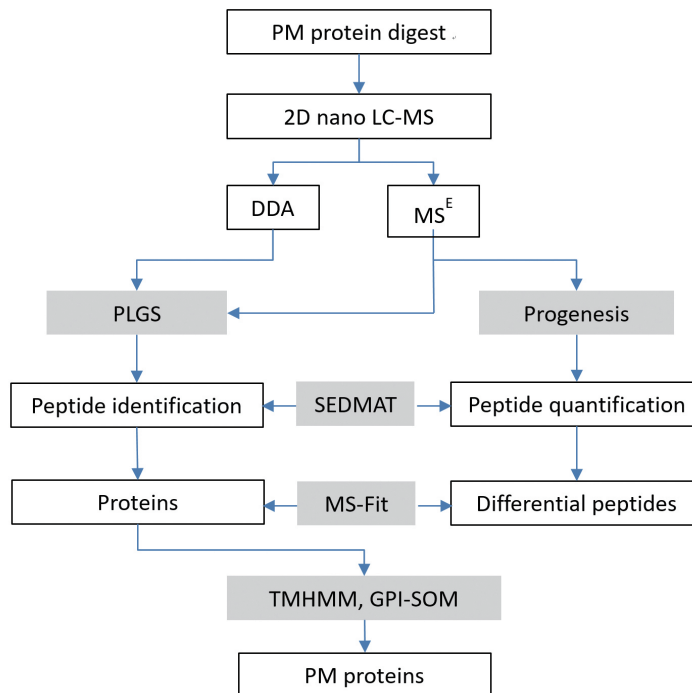
## Results

### *Identification of proteins in PM enriched fractions from Arabidopsis Col-0, lecrk-I.9 and 35S-ipiO1 lines*

Uninfected Col-0, *lecrk-I.9* and *35S-ipiO1* plants do not display any obvious growth phenotype, but the different response to *P. brassicae* infection could be due to a difference in the PM protein composition in *lecrk-I.9* and *35S-ipiO1*. Comparison of the PM proteome of these two different transgenic lines with that of Col-0 could help us understand how *lecrk-I.9* and *35S-ipiO1* plants gain susceptibility to infection by *P. brassicae*. However, PM proteins are often under-represented in proteomics studies due to their hydrophobicity and low abundance. Enrichment of PM vesicles using aqueous two-phase partitioning may improve detection of PM proteins (Alexandersson et al., 2004). For the isolation of PM proteins, microsomes were isolated by ultracentrifugation (see methods). This microsomal pellet contains PM, ER and chloroplast membranes and the different membrane fractions were separated by density centrifugation on a Dextran/PEG step gradient, from which the PM-enriched microsome fraction was isolated (Alexandersson et al., 2004).

Because we use label free comparative proteomics we used three replicate samples for each genotype, to estimate variations in protein yield introduced by the PM isolation procedure. During the formation of microsomes, cytosolic proteins may be captured inside the microsomes. Such contaminating cytosolic proteins may be removed through a Brij 58 treatment (Alexandersson et al., 2004), however, because in our hands this detergent strongly interfered with downstream LC-MS analysis, this wash step was omitted. Instead, from all the proteins identified by LC-MS/MS the predicted sub-cellular localization based on signal peptide or GPI-anchor site was used to filter out non-membrane proteins. The proteins extracted from the nine PM-enriched microsome samples (3 genotypes  $\times$  3 biological replicates) were digested with trypsin, after which peptides were purified by SPE for analysis by 2D nano-LC-MS. For optimal identification and quantification of peptides, each sample was eluted in seven steps from the first reverse phase (RP) column after which each of the seven eluted fractions was analyzed with a long gradient on the second RP column. All samples were run in MS<sup>E</sup> mode, yielding 21 data sets (7 LC fractions from each sample  $\times$  3 replicates) per genotype. In addition, replicate PM samples were pooled and analyzed in DDA mode, with the aim to improve protein identification (see Materials and methods). Fig. 1 shows an overview of the workflow that was used for peptide processing, identification and semi-quantification. ProteinLynx Global Server software (PLGS) was used for protein identifications using the LC-MS<sup>E</sup> data. The identified peptides were linked to quantitated peptide mass features extracted from the same sets of LC-MS<sup>E</sup> data with Progenesis software and in-house developed software called SEDMAT ([https://toolshed.g2.bx.psu.edu/view/pieterlukasse/prims\\_proteomics/6107b74eeb11](https://toolshed.g2.bx.psu.edu/view/pieterlukasse/prims_proteomics/6107b74eeb11)). The search program MS-Fit was used to help identify differential peptide features, which were not identified by PLGS. By searching the MS<sup>E</sup> and DDA fragmentation spectra against the

*Arabidopsis thaliana* protein database, using PLGS as a search engine, in total 2151 proteins could be identified in the nine peptide samples. From these, 249 proteins were identified from DDA spectra, which escaped identification in MS<sup>E</sup>, while MS<sup>E</sup> yielded 832 proteins not found by DDA. Hence, both MS methods were complementary and together produced a more comprehensive list of identified proteins in the enriched PM fractions (Supplementary Table S1). The GO annotation of subcellular localization by TAIR (<https://www.arabidopsis.org/tools/bulk/go/>) (Berardini et al., 2004) suggested that 45% of the 2151 identified proteins are PM-localized (Fig. S1A-B). Because only 13% of the total *Arabidopsis* proteome is annotated by TAIR to have PM localization, this confirms enrichment for PM proteins in our samples. The non-PM proteins were filtered from the list of 2151 identified proteins by selecting only those proteins with a minimum of one trans-membrane domain as predicted by TMHMM (Krogh et al., 2001) and/or a potential GPI-anchor as predicted by GPI-SOM (Fankhauser and Maser, 2005). This resulted in a list of 613 putative membrane proteins: 564 proteins with one or more TMD and 49 proteins with a putative GPI-anchor attachment site. In this list of 613 membrane proteins there were 26 LecRK proteins out of the 45 LecRKs annotated in *Arabidopsis* (see Table S2).

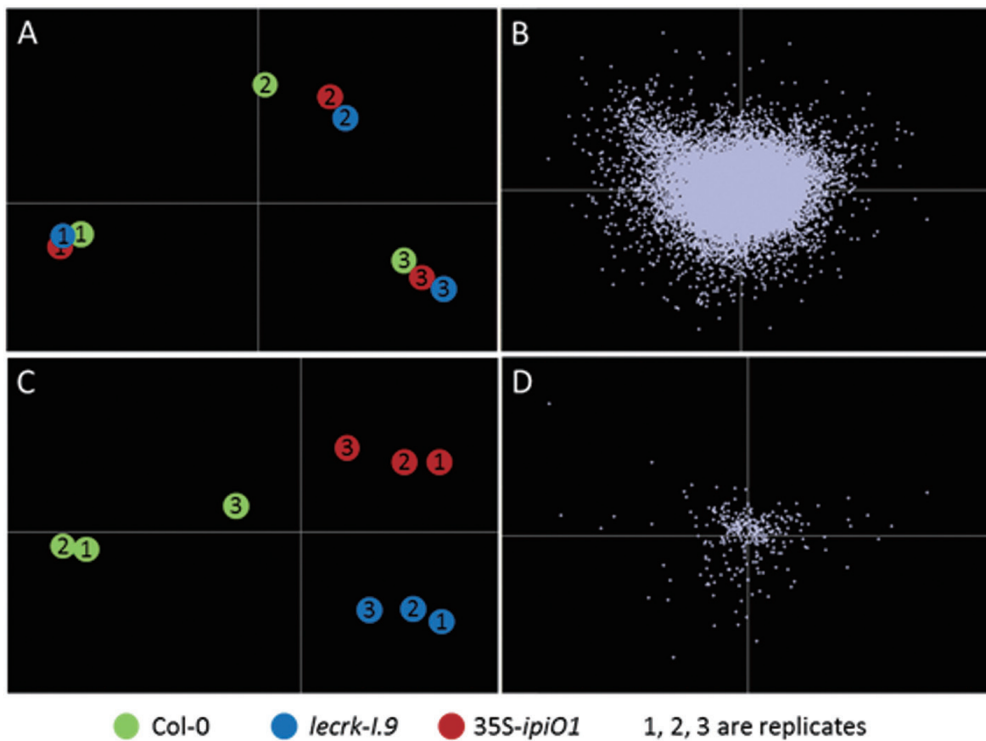


**Figure 1.** General proteomics workflow used in this study. Arrows indicate the order of data processing. Tools for protein identification, quantification and annotation are shown in gray-shaded boxes.

### ***Differential peptide mass intensity of PM proteins in Arabidopsis Col-0, lecrk-1.9 and 35S-ipiO1 lines***

For quantitative comparison of peptide mass intensities in the peptide samples isolated from the different genotypes, the peak intensities from the 2D nano LC-MS<sup>E</sup> acquired data were analyzed with Progenesis. This software allows for alignment of LC-MS peptide signals between multiple samples, in order to screen for differentially expressed proteins/peptides. Alignment of each of the seven corresponding fractions obtained from the 2D nano LC-MS<sup>E</sup> from the three genotypes (in triplicate) resulted in a total of 76890 peptide features. We note that because of multiple charge states of peptides in ESI-MS one peptide can generate multiple features with the same RT in the ion intensity map generated in Progenesis. Therefore, the number of features exceeds the actual number of peptides. Eventually, these 76890 features could be linked to 12326 identified peptides in PLGS (note that this data set was not yet filtered for non-PM proteins). Analysis of these 12326 identified features from the nine samples by Principal Component Analysis (PCA) did not show separation of samples by genotype (Fig. 2A-B). This could indicate that there are only minor differences in the PM proteome between the three genotypes, but also that there is a relatively high level of contaminating non-membrane proteins present. Subsequently, the set of 76890 features was filtered for significant two-fold differences between the average peptide intensity of two genotypes (ANOVA  $P < 0.05$ ). This filtering proved to be very stringent, resulting in a reduction to only 99 differential features. The PCA plot based on these 99 remaining features now showed good clustering of the replicate samples and a separation between the three genotypes (Fig. 2C-D). The 99 differentially expressed features were manually checked in Progenesis for correct alignment and differential expression. Fig. 3 shows an overview of the 99 differentially expressed features distributed over the 6 possible expression patterns. Unfortunately, none of the 99 features could be identified directly by matching PLGS protein ID to Progenesis using SEDMAT software. Therefore, we used MS-Fit (Clauser et al., 1999) for peptide/protein identification of these 99 differential features. To enhance reliability of protein identification by MS-Fit we used the strict criterion of at least three differential features (peptides) linking to the same protein. In this way, 25 from the 99 differential features could be linked to a total of 17 PM proteins. This set of 17 candidate PM proteins cluster into two expression classes (Fig. 3), for the other four expression patterns the limited number of peptide features could not be linked to a known protein. Analysis of gene expression data from The Bio-Analytic Resource for Plant Biology (BAR) expression browser ([http://bar.utoronto.ca/ntools/cgi-bin/ntools\\_expression\\_angler.cgi](http://bar.utoronto.ca/ntools/cgi-bin/ntools_expression_angler.cgi)) (Toufighi et al., 2005) indicated that most of these 17 genes are responsive to *P. infestans* as shown in Fig. S2. Since *lecrk-1.9* and *35S-ipiO1* share the susceptibility to *P. brassicae* infection, they may show similar changes in the PM proteome. Therefore, differential PM proteins with a low peptide abundance in Col-0 and high abundance in the two transgenic lines and, on the other hand, with high abundance in Col-0 and low abundance in the two transgenic lines are of most interest. Sixteen of the identified PM proteins show low expression in Col-0 and high expression in both transgenic lines (LIW: low in Col-0), while one PM protein (cysteine-rich

protein kinase 37; CRK37) shows high expression in Col-0 and low expression in both transgenic lines (HIW: high in Col-0) (Fig. 3). Therefore, all 17 proteins that passed the filtering criteria are potential candidates that may have causal relationship to the difference in susceptibility to *P. brassicae* infection.



**Figure 2.** PCA plot of peptide intensities of 2D nano LC-MS<sup>E</sup> data from PM enriched protein fractions. PM enriched proteins were isolated from three replicates of *Arabidopsis* Col-0, *lecrk-1.9* and *35S-ipiO1* and tryptic digests analyzed by 2D nano LC-MS<sup>E</sup>. This yielded 76890 features after peak detection and alignment in Progenesis, of which 12326 features could be identified in PLGS. A-B: PCA plots of total 76890 features; C-D: PCA plots of a subset of 310 features selected based on significant difference with Col-0 ( $P < 0.05$ ).

	Col-0	<i>lecrk-1.9</i>	<i>35S- ipiO1</i>	Differential features	Identified PM protein features	Identified PM proteins
<b>LIW</b>	Green	Red	Red	45	22	16
<b>LIK</b>	Red	Green	Red	21	0	0
<b>LIO</b>	Red	Red	Green	0	0	0
<b>HIW</b>	Red	Green	Green	24	3	1
<b>HIK</b>	Green	Red	Green	1	0	0
<b>HIO</b>	Green	Green	Red	8	0	0
<b>Total</b>				99	25	17

**Figure 3.** Classification of differential mass features (data obtained from Progenesis). The 99 differentially expressed features were classified in 6 different expression profiles: LIW (low in Col-0 and high in both *lecrk-1.9* and *35S-ipiO1* lines); LIK (low in *lecrk-1.9* and high in Col-0 and *35S-ipiO1*); LIO (low in *35S-ipiO1* and high in Col-0 and *lecrk-1.9*); HIW (high in Col-0 and low in both *lecrk-1.9* and *35S-ipiO1* lines); HIK (high in *lecrk-1.9* and low in Col-0 and *35S-ipiO1*); HIO (high in *35S-ipiO1* and low in Col-0 and *lecrk-1.9*). Of the 99 differentially expressed features 25 could be assigned to 17 PM proteins, using MS-Fit to link m/z values to protein identification. Results were filtered for PM proteins based on the presence of TMD or GPI. (Color scheme: green = low peptide abundance, red = high peptide abundance).

## 4

### ***Transcription of the 17 PM-proteins identified from Arabidopsis Col-0, lecrk-1.9 and 35S-ipiO1***

The same leaf material that was used for PM-proteomics was used for extraction of mRNA to study expression of the genes encoding the candidate proteins. The transcript profile over the three genotypes of these 17 candidate genes was compared with the encoded protein levels (Fig. 4). For three out of the 17 genes the transcription profile partly matched the protein abundance profile. For these genes the transcript level was both low in *Arabidopsis* Col-0 and *35S-ipiO1* and enhanced in *lecrk-1.9*, while protein abundance tended to be low in Col-0 and high in both phenocopy lines, indicating a consistent discrepancy between transcription and translation levels in the three genotypes. The transcript profile of the gene encoding CRK37 (AT4G04500) was most different from the protein expression pattern: transcript levels were low in Col-0 and *35S-ipiO1* and high in *lecrk-1.9*, while protein levels were lower in *35S-ipiO1* and *lecrk-1.9* compared to Col-0 (Fig. 4).

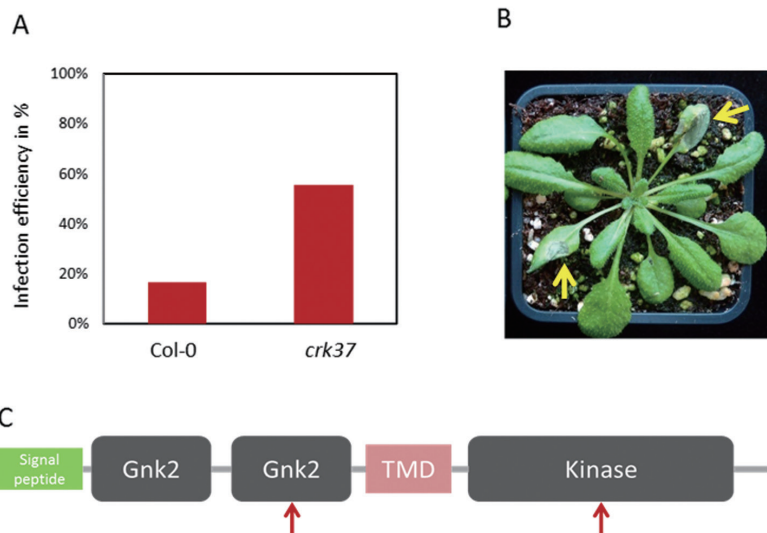


Accession	Proteome			Transcriptome		
	Col-0	35S- <i>ipiO1</i>	<i>lecrk-1.9</i>	Col-0	35S- <i>ipiO1</i>	<i>lecrk-1.9</i>
*AT1G02145	345,2	3801,9	4131,8	14,8	13,5	13,5
AT2G33640	133,4	374,4	425,8	0,4	0,5	0,7
*AT3G47790	252,1	2992,3	3097,3	0,0	0,0	0,0
*AT5G61700	46,0	202,9	161,6	0,0	0,0	0,0
AT1G16860	216,5	568,8	656,5	6,6	8,3	8,2
*AT1G56140	452,7	1211,1	1144,6	3,2	3,0	4,9
*AT2G39350	275,6	852,0	881,6	0,8	0,6	1,0
*AT3G60330	57,4	193,8	136,4	0,2	0,2	0,2
*AT2G19210	270,3	3049,8	3099,2	0,0	0,0	0,0
*AT4G23700	136,4	345,7	406,9	3,2	1,6	1,6
*AT5G49150	276,7	3130,7	3142,5	0,3	0,3	0,3
AT4G28890	247,4	3041,7	3096,6	0,0	0,0	0,0
*AT1G61490	301,0	954,1	1072,9	1,5	1,3	2,1
*AT1G69870	49,9	555,9	526,2	91,8	69,7	85,1
*AT3G47740	78,4	368,0	385,1	0,0	0,0	0,0
*AT2G37940	247,4	3041,7	3096,6	82,6	62,6	61,5
*AT4G04500	627,2	129,2	97,1	3,7	3,2	4,4

**Figure 4.** Comparison of proteome and transcriptome expression data of 17 identified PM proteins. Left: protein intensity heatmap of 17 proteins with differential peptide abundance between *Arabidopsis* Col-0 and 35S-*ipiO1* and/or *lecrk-1.9*, as determined with comparative proteomics. Right: transcript levels of the protein encoding genes as determined by RNAseq. The intensity of mRNAs and proteins were normalized against total signal intensity in transcriptome and proteome data respectively. Genes that show a different pattern for protein and mRNA abundance across the samples are marked with an asterisk.

#### *Arabidopsis crk37* mutant shows gain of susceptibility to *P. brassicae*

Because of the unique RNA and protein expression profile of *CRK37* in the set of 17 differentially expressed proteins and the potential role of kinases in plant-pathogen signaling, *CRK37* was selected as candidate for further functional analysis. The low level of *CRK37* in both 35S-*ipiO1* and *lecrk1.9* coincides with gain of susceptibility to *P. brassicae* and therefore we tested whether a *CRK37* mutant shows a similar phenotype. The *crk37* *Arabidopsis* T-DNA insertion mutant (SALK\_004923) was obtained from the SALK T-DNA insertion collection (Bourdais et al., 2015). Gene expression analysis confirmed *crk37* as knock-out mutant of the *CRK37* gene (data not shown). The *crk37* mutant was analyzed for response to *P. brassicae* in a leaf infection assay (Fig. 5A-B). The *crk37* mutant plant did indeed display a gain of susceptibility to *P. brassicae* compared to Col-0. Future analysis will have to show whether lack of *CRK37* in Col-0 also impairs the CW-PM continuum, similar to the effect of absence of *LecRK-I.9* and overexpression of *ipiO* as demonstrated by Bouwmeester *et al.* (Bouwmeester et al., 2011b).



**Figure 5.** Infection of *Arabidopsis crk37* by *P. brassicae* and CRK37 protein domain architecture. **A:** Infection efficiencies of *Phytophthora brassicae* HH on *Arabidopsis* Col-0 and *crk37* leaves at 4 days post inoculation. Percentages indicate the number of lesions on the inoculated leaves (18 leaves per line). **B:** Phenotype of *crk37* mutant upon *P. brassicae* infection; lesions are indicated with yellow arrows. **C:** Illustration of CRK37 protein domain architecture. Gnk2: Ginkbilobin-2 antifungal domain; Red arrows indicates location of peptide sequences identified by MS-Fit.

## Discussion

### *Differential proteins identified in Arabidopsis Col-0 and lecrk-1.9 and 35S-ipiO1 plants*

Here a label-free comparative proteomics approach was used to detect differences in PM protein composition of *Arabidopsis* leaves of Col-0 plants and *lecrk-1.9* and *35S-ipiO1* plants, which both show a gain of susceptibility to *P. brassicae*. An enrichment of PM vesicles using aqueous two-phase partitioning was used (Alexandersson et al., 2004) in which we omitted the Brij 58 washing step, which is normally used to remove contaminating cytosolic proteins. This washing step was omitted because of greatly reduced protein yield and interference of Brij 58 with LC-MS. To solve this, stringent filtering (based on predicted targeting information) was applied to the identified proteins in order to remove none-PM proteins. Thanks to this filtering, even without the Brij 58 wash step, 45% of the identified proteins in our PM fractions were annotated to be localized on the PM according to TAIR (Berardini et al., 2004), which is a clear increase compared to 13% PM-proteins in the total *Arabidopsis* proteome as deduced from the

TAIR GO annotation as shown in Fig. S1. In addition, contaminating proteins will show variable abundance between replicate samples and therefore are removed by filtering on consistent protein levels in replicate samples. Another major problem we encountered in the comparative data analysis is the matching of differential features to peptide and protein identification. To solve this, MS-Fit was employed. MS-Fit correlates Mass Spectrometry data (parent masses and charge only, not fragment masses) with a protein in the TAIR sequence database which fits that data the best. However, this method faces the same drawback as regular proteomics approaches; it cannot distinguish closely similar proteins, if the identified peptides belong to conserved protein regions. Therefore, it can be expected that peptides originating from homologous proteins will show a similar peptide(feature) abundance pattern in Progenesis. To overcome this, we used the stringent criteria that at least three feature/peptide masses should be detected with a similar expression profile in a narrow mass tolerance window. In this way 17 PM proteins were identified from those 99 features. The biological function of those protein candidates must be validated through knockout or overexpression and analysis of the resulting phenotype. Analysis of the transcript levels of these 17 genes over the three genotypes showed there were only subtle differences in gene expression. Gene expression analysis would hence not have been sufficient to link these genes to a role in plant-pathogen interaction. One of the candidate proteins (CRK37) that displayed a higher protein level in Col-0 than in the two transgenic lines, was tested for a function in the response of *Arabidopsis* to *P. brassicae* infection.

### ***CRK37 is required as a disease resistance component to *P. brassicae****

*CRK37* encodes a cysteine-rich receptor kinase with an extracellular putative receptor domain and a cytosolic kinase domain (Fig. 5C). The *lecrk-I.9* mutant has a low abundance of CRK37, while CRK37 transcription is actually upregulated in *lecrk-I.9* compared to Col-0 (Fig. 4). This suggests that the LecRK-I.9 and CRK37 proteins may interact in plants, leading to stabilization of CRK37. CRK37 is predicted to be *N*-glycosylated by the NetNGlyc server (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and it could be that the lectin binding motif of LecRK-I.9 interacts with the *N*-glycans on CRK37. Because CRK37 level is also reduced in the *35S-ipiO1* line and IPI-O was shown to interact with LecRK-I.9, this raises the question whether the interaction of IPI-O with LecRK-I.9 causes destabilization of LecRK-I.9 or interferes with the putative interaction of LecRK-I.9 and CRK37, possibly resulting in destabilization of the CRK37 protein.

A role for CRK37 in the response to *P. brassicae* was subsequently demonstrated in the *Arabidopsis crk37* mutant, which displayed a gain of susceptibility to infection by *P. brassicae*. This matches the phenotypes of the *lecrk-I.9* and *35S-ipiO1* plants (susceptibility to *P. brassicae* and reduced CRK37 protein levels). Recently, a study of CRKs on their function in the response to oxidative stress in reactive oxygen species (ROS) signaling, was demonstrated by Bourdais *et al.* (Bourdais *et al.*, 2015), who suggested that CRKs might play an important role in a wide range of plant-abiotic/

biotic stress interactions. We note that in the proteomics experiment proteins were selected on the basis of at least two-fold difference between Col-0 and either one of the transgenic lines, while the transcript levels of the corresponding genes differed only between 1.3 to 1.6-fold between Col-0 and the transgenic lines. Because usually a cut-off value of at least two-fold difference is used to identify significant differences in gene expression, the transcriptomics study alone would not have identified CRK37 as a candidate disease resistance component.

### ***PM proteomics identifies more LRR RLKs than previous reports in Arabidopsis***

Compared to other publications on plant PM proteomics (Mitra et al., 2009), we identified PM proteins with higher confidence in peptide score and ranking. This can be attributed to the gain in sensitivity of peptide detection by optimal peptide separation, first in seven steps using RP chromatography at high pH, followed by RP chromatography for each of the seven fraction at low pH. This increased LC resolution of peptides prior to ionization and MS analysis resulted in a considerable improvement in the detection of low abundant peptides. This is illustrated by (1) the high number of identified LRR RLKs and (2) the high number of identified LecRKs (Table S2). LRR RLKs are the most well-studied RLKs in relation to plant development, pathogen resistance and hormone perception (Shiu and Bleecker, 2001; Goff and Ramonell, 2007; Nishimura and Dangl, 2010; Vaid et al., 2013). In previous PM proteomics studies of LRR RLKs only 30 LRR RLKs were identified in *Arabidopsis* (Mitra et al., 2009), while in total 82 LRR RLKs were identified in the present study (Table S3). However, LecRK-I.9 was not among the LecRK proteins identified by proteomics, suggesting that protein levels of this receptor kinase may be very low.

## 4

### **Acknowledgements**

This project was carried out within the research program of the Centre of BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

## References

**Alexandersson E, Saalbach G, Larsson C, Kjellbom P** (2004) *Arabidopsis* plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol* **45**: 1543–56

**Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang PF, Mueller LA, Yoon J, Doyle A, Lander G, et al** (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiol* **135**: 745–755

**Bourdais G, Burdiak P, Gauthier A, Nitsch L, Salojarvi J, Rayapuram C, Idanheimo N, Hunter K, Kimura S, Merilo E, et al** (2015) Large-scale phenomics identifies primary and fine-tuning roles for CRKs in responses related to oxidative stress. *Plos Genet.* doi: ARTN e100537310.1371/journal.pgen.1005373

**Bouwmeester K, Meijer HJG, Govers F** (2011a) At the frontier; RXLR effectors crossing the *Phytophthora*-host interface. *Front Plant Sci.* doi: Artn 7510.3389/Fpls.2011.00075

**Bouwmeester K, de Sain M, Weide R, Gouget A, Klamer S, Canut H, Govers F** (2011b) The lectin receptor kinase LecRK-I.9 is a novel *Phytophthora* resistance component and a potential host target for a RXLR effector. *Plos Pathog.* doi: ARTN e100132710.1371/journal.ppat.1001327

**Bradford MM** (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–254

**Choi J, Tanaka K, Cao YR, Qi Y, Qiu J, Liang Y, Lee SY, Stacey G** (2014) Identification of a plant receptor for extracellular ATP. *Science* (80- ) **343**: 290–294

**Clauser KR, Baker P, Burlingame AL** (1999) Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS MS and database searching. *Anal Chem* **71**: 2871–2882

**Desclos-Theveniau M, Arnaud D, Huang TY, Lin GJ, Chen WY, Lin YC, Zimmerli L** (2012) The *Arabidopsis* lectin receptor kinase LecRK-V.5 represses stomatal immunity induced by *Pseudomonas syringae* pv. *tomato* DC3000. *Plos Pathog* **8**: e1002513

**Dsouza SE, Ginsberg MH, Plow EF** (1991) Arginyl-glycyl-aspartic acid (RGD) - a cell-adhesion motif. *Trends Biochem Sci* **16**: 246–250

**Fankhauser N, Maser P** (2005) Identification of GPI anchor attachment signals by a

Kohonen self-organizing map. *Bioinformatics* **21**: 1846–1852

**Geromanos SJ, Vissers JPC, Silva JC, Dorschel CA, Li GZ, Gorenstein M V, Bateman RH, Langridge JI** (2009) The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependent LC-MS/MS. *Proteomics* **9**: 1683–1695

**Goff KE, Ramonell KM** (2007) The role and regulation of receptor-like kinases in plant defense. *Gene Regul Syst Bio* **1**: 167–175

**Gouget A, Senchou V, Govers F, Sanson A, Barre A, Rouge P, Pont-Lezica RP, Canut H** (2006) Lectin receptor kinases participate in protein-protein interactions to mediate plasma membrane-cell wall adhesions in *Arabidopsis*. *Plant Physiol* **140**: 81–90

**Krogh A, Larsson B, von Heijne G, Sonnhammer ELL** (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* **305**: 567–580

**Li GZ, Vissers JPC, Silva JC, Golick D, Gorenstein M V, Geromanos SJ** (2009) Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**: 1696–1719

**Mellersh DG, Heath MC** (2001) Plasma membrane-cell wall adhesion is required for expression of plant defense responses during fungal penetration. *Plant Cell* **13**: 413–424

**Mitra SK, Walters BT, Clouse SD, Goshe MB** (2009) An efficient organic solvent based extraction method for the proteomic analysis of *Arabidopsis* plasma membranes. *J Proteome Res* **8**: 2752–2767

**Nishimura MT, Dangl JL** (2010) *Arabidopsis* and the plant immune system. *Plant J* **61**: 1053–1066

**Roetschi A, Si-Ammour A, Belbahri L, Mauch F, Mauch-Mani B** (2001) Characterization of an *Arabidopsis-Phytophthora* pathosystem: resistance requires a functional PAD2 gene and is independent of salicylic acid, ethylene and jasmonic acid signalling. *Plant J* **28**: 293–305

**Shiu SH, Bleecker AB** (2001) Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE* **2001**: re22

**Singh P, Kuo YC, Mishra S, Tsai CH, Chien CC, Chen CW, Desclos-Theveniau M, Chu PW, Schulze B, Chinchilla D, et al** (2012) The lectin receptor kinase-VI.2 is

required for priming and positively regulates *Arabidopsis* pattern-triggered immunity. *Plant Cell* **24**: 1256–1270

**Takagi J** (2004) Structural basis for ligand recognition by RGD (Arg-Gly-Asp)-dependent integrins. *Biochem Soc Trans* **32**: 403–406

**Toufighi K, Brady SM, Austin R, Ly E, Provard NJ** (2005) The botany array resource: e-Northern, expression angling, and promoter analyses. *Plant J* **43**: 153–163

**Vaid N, Macovei A, Tuteja N** (2013) Knights in action: lectin receptor-like kinases in plant development and stress responses. *Mol Plant* **6**: 1405–1418

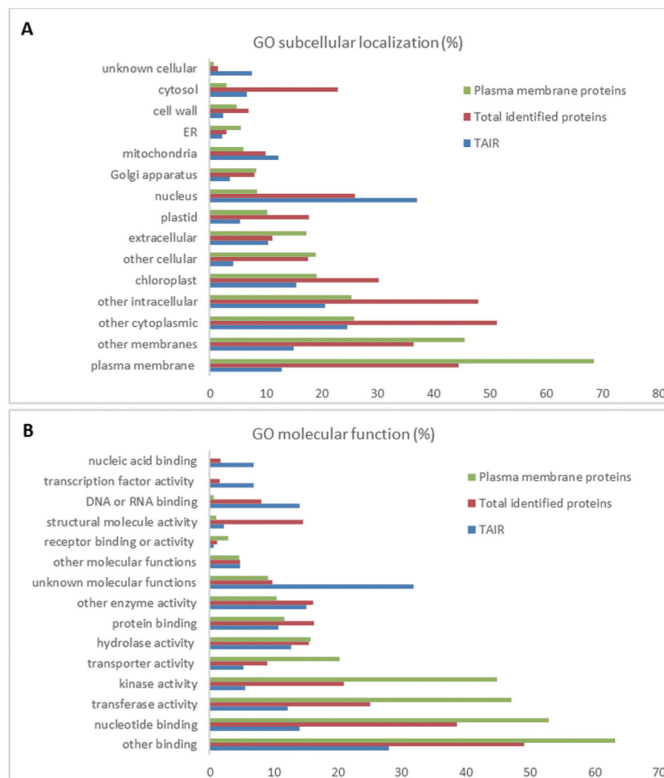
**Wang Y, Nsibo DL, Juhar HM, Govers F, Bouwmeester K** (2016) Ectopic expression of *Arabidopsis* L-type lectin receptor kinase genes LecRK-I.9 and LecRK-IX.1 in *Nicotiana benthamiana* confers *Phytophthora* resistance. *Plant Cell Rep* **35**: 845–855

## Supplementary Data

Supplementary data to this article can be found online at:

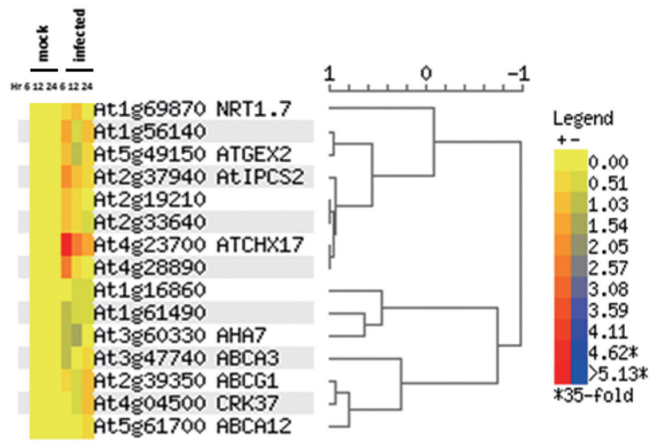
[https://1drv.ms/f/s!Asszhx\\_to9U8hxXix8NpPWFDJNdA](https://1drv.ms/f/s!Asszhx_to9U8hxXix8NpPWFDJNdA)

or by scanning the QR code:



**Figure S1.** GO annotation of subcellular localization (A) and molecular function (B) of identified PM proteins (613), total identified proteins (2151), and the whole proteome of *Arabidopsis thaliana*. GO annotation of subcellular localization shows that plasma membrane proteins were enriched (based on their subcellular localization annotation) in the total identified proteins (2151) and the putatively identified PM proteins (613), compared with the whole *Arabidopsis* proteome from TAIR. GO annotation of protein molecular function shows that binding and kinase activity (typical plasma membrane protein functions) were enriched in the total identified proteins and PM proteins compared with the whole *Arabidopsis* proteome from TAIR.





**Figure S2.** Gene expression of 17 candidates under biotic stress in the BAR expression browser. Heatmap on the left indicates the level of gene expression under *Phytophthora infestans* treatment at different time points. Genes are clustered according to the similarity of their expression pattern under the different treatments. The heatmap shows that expression of most of the genes increases with time in response to *P. infestans* treatment.

**Table S2.** LecRKs identified by proteomics in plasma membrane enriched fractions of *Arabidopsis* Col-0, *lecrk-I.9* and *35S-ipiO1*. LecRK-I.9 is highlighted (yellow) because it was knocked-out in the *lecrk-I.9* line. Numbers are the sum of identified peptides, in brackets are the numbers of unique identified peptides for the corresponding LecRK.

Name	Accession	Col-0	<i>lecrk-I.9</i>	<i>35S-ipiO1</i>	Name	Accession	Col-0	<i>lecrk-I.9</i>	<i>35S-ipiO1</i>
LecRK-I.1	AT3G45330	32			LecRK-V.5	AT3G59700			
LecRK-I.2	AT3G45390			12	LecRK-V.6	AT3G59730			
LecRK-I.3	AT3G45410	13	10		LecRK-V.7	AT3G59740			
LecRK-I.4	AT3G45420	39 (17)	15 (3)	23 (2)	LecRK-V.8	AT3G59750			
LecRK-I.5	AT3G45430	36	11		LecRK-V.9	AT4G29050			
LecRK-I.6	AT3G45440				LecRK-VI.1	AT3G08870	51 (8)	69 (16)	34 (5)
LecRK-I.7	AT5G60270				LecRK-VI.2	AT5G01540	62 (16)	62 (19)	31 (7)
LecRK-I.8	AT5G60280				LecRK-VI.3	AT5G01550	62 (20)	54 (11)	54 (19)
LecRK-I.9	AT5G60300				LecRK-VI.4	AT5G01560	49 (4)	47 (8)	59 (18)
LecRK-I.10	AT5G60310	6			LecRK-VII.1	AT4G04960	1		22
LecRK-I.11	AT5G60320	10	19		LecRK-VII.2	AT4G28350			
LecRK-II.1	AT5G59260	69 (10)	80 (23)	40 (5)	LecRK-VIII.1	AT3G53380	49 (16)	52 (11)	38 (6)
LecRK-II.2	AT5G59270	36 (7)	30 (15)	33 (11)	LecRK-VIII.2	AT5G03140	14 (3)	40 (18)	33 (10)
LecRK-III.1	AT2G29220				LecRK-IX.1	AT5G10530	56 (21)	38 (10)	20 (5)
LecRK-III.2	AT2G29250				LecRK-IX.2	AT5G65600	17 (6)	62 (21)	46 (11)
LecRK-IV.1	AT2G37710	173 (11)	144 (13)	151 (9)	LecRK-S.1	AT1G15530	28 (5)	46 (8)	59 (5)
LecRK-IV.2	AT3G53810	67 (3)	61 (12)	59 (11)	LecRK-S.2	AT2G32800	56 (17)	76 (13)	48 (19)
LecRK-IV.3	AT4G02410	65 (2)	75 (13)	72 (9)	LecRK-S.3	AT3G46760			
LecRK-IV.4	AT4G02420	102 (2)	105 (7)	113 (16)	LecRK-S.4	AT3G55550			
LecRK-V.1	AT1G70110				LecRK-S.5	AT5G06740	92 (23)	82 (21)	50 (8)
LecRK-V.2	AT1G70130				LecRK-S.6	AT5G42120			
LecRK-V.3	AT2G43690				LecRK-S.7	AT5G55830	27 (7)	55 (10)	44 (14)
LecRK-V.4	AT2G43700								

**Table S3.** LRR-RLKs identified by 2D nano LC-MS, using a combination of DDA and MS<sup>E</sup> protein identification.

AGI number	Protein	TMDs	GPI	SignalP	No. of peptides	Mitra <i>et al.</i>
AT1G05700	Leucine-rich repeat transmembrane protein kinase	2		Y	21	
AT1G06840	Leucine-rich repeat protein kinase family protein	2		Y	39	
AT1G07550	Leucine-rich repeat protein kinase family protein	1		Y	1	
AT1G07560	Leucine-rich repeat protein kinase family protein	1		Y	35	
AT1G07650	Leucine-rich repeat transmembrane protein kinase	2		Y	70	
AT1G27190	Leucine-rich repeat protein kinase family protein	1		Y	26	Y
AT1G29720	Leucine-rich repeat transmembrane protein kinase	2		Y	41	
AT1G29730	Leucine-rich repeat transmembrane protein kinase		Y	Y	27	
AT1G29740	Leucine-rich repeat transmembrane protein kinase	2		Y	29	
AT1G31420	Leucine-rich repeat transmembrane protein kinase	1		Y	1	
AT1G35710	Leucine-rich repeat transmembrane protein kinase	2		Y	37	

AT1G49100	Leucine-rich repeat protein kinase family protein	2	Y	40	
AT1G50610	Leucine-rich repeat protein kinase family protein	2		10	
AT1G51800	Leucine-rich repeat transmembrane protein kinase	1	Y	50	
AT1G51805	Leucine-rich repeat protein kinase family protein	1	Y	74	Y
AT1G51810	Leucine-rich repeat protein kinase family protein	1		38	
AT1G51820	Leucine-rich repeat protein kinase family protein	1	Y	50	
AT1G51830	Leucine-rich repeat protein kinase family protein	1		24	
AT1G51850	Leucine-rich repeat protein kinase family protein	1	Y	49	
AT1G51860	Leucine-rich repeat protein kinase family protein	1	Y	45	
AT1G51890	Leucine-rich repeat protein kinase family protein	1	Y	51	
AT1G51910	Leucine-rich repeat protein kinase family protein	1	Y	37	
AT1G53420	Leucine-rich repeat transmembrane protein kinase	2	Y	70	
AT1G53430	Leucine-rich repeat transmembrane protein kinase	1	Y	94	Y
AT1G53440	Leucine-rich repeat transmembrane protein kinase	2	Y	87	Y
AT1G56120	Leucine-rich repeat transmembrane protein kinase	1		24	
AT1G56130	Leucine-rich repeat transmembrane protein kinase	2	Y	20	
AT1G56140	Leucine-rich repeat transmembrane protein kinase	2		14	
AT1G56145	Leucine-rich repeat transmembrane protein kinase	2	Y	1	
AT1G66830	Leucine-rich repeat protein kinase family protein	2	Y	1	
AT1G67720	Leucine-rich repeat protein kinase family protein	1	Y	16	
AT1G69990	Leucine-rich repeat protein kinase family protein	1		21	
AT1G79620	Leucine-rich repeat protein kinase family protein	2	Y	49	
AT2G01210	Leucine-rich repeat protein kinase family protein		Y	23	
AT2G01820	Leucine-rich repeat protein kinase family protein	1	Y	1	Y
AT2G04300	Leucine-rich repeat protein kinase family protein	2	Y	39	
AT2G14440	Leucine-rich repeat protein kinase family protein	1	Y	41	
AT2G14510	Leucine-rich repeat protein kinase family protein	1	Y	41	
AT2G19210	Leucine-rich repeat transmembrane protein kinase	2	Y	22	
AT2G19230	Leucine-rich repeat transmembrane protein kinase	2	Y	1	
AT2G24230	Leucine-rich repeat protein kinase family protein	1	Y	24	
AT2G26730	Leucine-rich repeat protein kinase family protein	1	Y	37	Y
AT2G28960	Leucine-rich repeat protein kinase family protein	2	Y	61	
AT2G28970	Leucine-rich repeat protein kinase family protein	1	Y	19	
AT2G28990	Leucine-rich repeat protein kinase family protein	1	Y	29	
AT2G29000	Leucine-rich repeat protein kinase family protein	2	Y	29	Y
AT2G31880	Leucine rich repeat transmembrane protein	2	Y	48	
AT2G35620	Leucine-rich repeat receptor kinase protein	1	Y	9	
AT2G37050	Leucine-rich repeat protein kinase family protein	1	Y	17	Y
AT3G02880	Leucine-rich repeat protein kinase family protein	2	Y	58	Y
AT3G08680	Leucine-rich repeat protein kinase family protein	2	Y	22	
AT3G14840	Leucine-rich repeat transmembrane protein kinase	2	Y	93	Y
AT3G21340	Leucine-rich repeat protein kinase family protein	2	Y	41	
AT3G23750	Leucine-rich repeat protein kinase family protein	2	Y	40	Y

AT3G28040	Leucine-rich repeat protein kinase family protein	1	Y	21	
AT3G28450	Leucine-rich repeat protein kinase family protein	2	Y	47	Y
AT3G46330	maternal effect embryo arrest 39 (MEE39)	2	Y	1	
AT3G46350	Leucine-rich repeat protein kinase family protein	2	Y	1	
AT3G46370	Leucine-rich repeat protein kinase family protein	1		11	
AT3G46400	Leucine-rich repeat protein kinase family protein	2	Y	1	
AT3G49670	CLAVATA1-related receptor kinase-like protein	1	Y	29	Y
AT3G53590	Leucine-rich repeat protein kinase family protein			25	
AT4G08850	Leucine-rich repeat receptor-like protein kinase	2		93	Y
AT4G28650	Leucine-rich repeat transmembrane protein kinase		Y	2	
AT4G29990	Leucine-rich repeat transmembrane protein kinase	1	Y	1	
AT5G01950	Leucine-rich repeat protein kinase family protein	1	Y	27	
AT5G10020	Leucine-rich repeat protein kinase family protein	1	Y	73	
AT5G10290	Leucine-rich repeat transmembrane protein kinase	1	Y	44	
AT5G16590	Leucine-rich repeat protein kinase family protein	2	Y	62	Y
AT5G16900	Leucine-rich repeat protein kinase family protein	1	Y	33	
AT5G37450	Leucine-rich repeat protein kinase family protein	1	Y	2	
AT5G46330	Leucine-rich repeat serine/threonine protein kinase	1	Y	2	
AT5G49760	Leucine-rich repeat protein kinase family protein	2	Y	63	Y
AT5G49770	Leucine-rich repeat protein kinase family protein	2	Y	54	
AT5G49780	Leucine-rich repeat protein kinase family protein	1		47	
AT5G59650	Leucine-rich repeat protein kinase family protein	1	Y	43	
AT5G59660	Leucine-rich repeat protein kinase family protein	1		60	
AT5G59670	Leucine-rich repeat protein kinase family protein	1	Y	49	
AT5G59680	Leucine-rich repeat protein kinase family protein	1	Y	50	
AT5G62710	Leucine-rich repeat protein kinase family protein	2	Y	8	
AT5G65240	Leucine-rich repeat protein kinase family protein	1	Y	48	
AT5G65700	CLAVATA1-related receptor kinase-like protein	1		43	

# Chapter 5

## **Plasma membrane proteomics of tomato MsK8 suspension cells infected with *Phytophthora infestans***

Wei Song <sup>a,b</sup>, Nina Lintermans <sup>c</sup>, Jan H. G. Cordewener <sup>b</sup>, Harro Bouwmeester <sup>a</sup>,  
Alexander R. van der Krol <sup>a</sup>, Antoine H.P. America <sup>b,\*</sup>

<sup>a</sup>Laboratory of Plant Physiology, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>b</sup>Plant Research International, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>c</sup>Faculteit Bio-ingenieurswetenschappen, Katholieke Universiteit,  
Leuven, Belgium

This work was carried out in collaboration with  
Klaas Bouwmeester <sup>d,e</sup> and Francine Govers <sup>d</sup>

<sup>d</sup>Laboratory of Phytopathology, Wageningen University and Research Centre,  
Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>e</sup>Plant-Microbe Interactions, Department of Biology, Utrecht University,  
Utrecht, The Netherlands

## Abstract

In plants, the plasma membrane (PM) forms the barrier between the external and internal cellular environment and proteins at the PM help regulate many essential processes like growth, development and adaption to the environment, including abiotic and biotic stresses. In order to be able to deal with pathogens, plants have evolved receptors to recognize so-called pathogen associated molecular patterns (PAMPs). In turn, the recognition by plant receptors of these PAMPs may be suppressed by effector proteins secreted by the pathogen. Much of the signaling related to PAMPs or effector detection takes place by proteins at the PM and here we used cultured tomato MsK8 suspension cells infected by *Phytophthora infestans*, as a model to study the role of the PM proteome in the *Phytophthora*-tomato interaction. Tomato MsK8 suspension cells were either challenged with *Phytophthora infestans* zoospores (spo) or the supernatant of a zoospore suspension (sup). The PM was isolated from the treated cells, proteins were extracted and label-free comparative proteomics was used for protein identification and quantification. A combination of data dependent acquisition (DDA) and data independent acquisition (DIA or MS<sup>E</sup>) mass spectrometry resulted in the identification of 1698 proteins (1101 from tomato and 597 from *P. infestans*). From these 1698 proteins, a total of 1365 proteins could be quantified by the Single Experiment Data Matching Tool from the Galaxy Proteomics toolbox (<https://usegalaxyp.org/>). In total, 287 tomato proteins showed a more than 1.5-fold change between control and treated samples. Among these, 73 proteins contain a trans-membrane domains, while five proteins contain the consensus sequence of a GPI-anchor addition site. Finally, ten proteins were selected for further study, based on the presence of domains found in proteins with a putative role in plant-pathogen interactions. The *Arabidopsis* homologs of these ten candidate tomato proteins were used for rapid validation studies.

## 5

## Keywords

Plasma membrane (PM), MsK8 suspension cells, label-free comparative proteomics, UPLC, mass spectrometry;

## Abbreviations

PM, Plasma Membrane; UPLC, ultra-performance liquid chromatography; ESI, electrospray ionization; QTOF, quadrupole time of flight;

## Introduction

The oomycete *Phytophthora infestans* is a pathogen which causes important agricultural problems in crops like potato and tomato and understanding how this pathogen is recognised by the plant to trigger immune responses may eventually help combat its effects. Plant resistance against pathogens is continuously evolving. For example, plants have evolved pattern recognition mechanisms using Pattern Recognition Receptors (PRRs) to perceive molecules derived from microbes. Perception of these pathogen-associated molecular patterns (PAMPs) results in a defense response called PAMP triggered immunity (PTI) (Boller and Felix, 2009; Macho and Zipfel, 2014). Such immunity may be suppressed by effector proteins secreted by the pathogen, which in turn may result in the evolution of effector-triggered immunity (ETI) in the plant (Win et al., 2012). Most microbial plant pathogens enter the extracellular space (also called the apoplast) between the cell wall and the cellular plasma membrane (PM) of plant cells to initiate infection, making the PM the first interface in the crosstalk between host and pathogen. Multiple receptors involved in pathogen recognition and immune responses are therefore located on the PM. Activated perception of pathogens by such PM located receptor proteins often affects their stability. Therefore, plant pathogen infection may be associated with alterations in the protein content in and around the PM. Investigation of changes in these PM proteins could thus help identify factors involved in pathogen recognition and the subsequent defense responses. Indeed it has been shown that proteins at the PM are important determinants for biotic stress resistance as they trigger a first biotic stress response (Macho and Zipfel, 2014).

Here we set out to identify proteins that may play a role in the recognition and defense response of tomato to *Phytophthora*. Identification of such proteins may lead to a better understanding of how the pathogen is recognized and how the plant activates specific defense responses. Eventually, such insights may provide the knowledge required for conventional or biotechnological breeding strategies aimed at improved plant resistance. Because the proteins in and on the PM are only a small fraction of the total protein content of cells, PM proteins are often under-represented in global proteomics studies. Analysis of the PM associated proteome can therefore greatly benefit from an enrichment of the PM fraction. To achieve this, plant cells can be extracted in specific buffers in which the different membranes of the cell form small lipid vesicles. Because the membrane lipid content of different cellular subcompartments differs, such lipid vesicles may be separated from each other using density centrifugation. When using the right conditions for this density centrifugation specific enrichment of PM derived vesicles is possible (Alexandersson et al., 2004). Initially we tried such PM enrichment for potato leaf material with and without infection with *P. infestans*. However, analysis of the proteins in the lipid fractions isolated in this way showed that, despite different variations in the procedure, no enrichment of PM proteins was obtained. As alternative to the potato-*P. infestans* interaction PM proteome study, we switched to tomato MsK8 suspension cells (Felix et al., 1991). The MsK8 suspension cells can, like tomato leaves, be infected by *P. infestans* (Felix et al., 1991), but has some advantages: (1) uniform

and synchronous infection compared with infection in leaves; (2) the MsK8 suspension cells are more uniform, with less different cell types compared with leaves; (3) for the tomato MsK8 suspension cells the PM enrichment procedure could be optimized (in contrast to potato leaf material).

During infection, zoospores of *P. infestans* secrete effector proteins, which can decrease the plant host defense response and thus help establish a successful infection. The host may counteract the effect of the pathogen effector proteins by ETI. Therefore, in order to distinguish between changes in PM proteome potentially related to PTI or ETI, the PM proteome was characterized for tomato MsK8 cells challenged by zoospores (PM protein changes related to PTI + ETI) and zoospore supernatant (PM protein changes mostly related to PTI). For PM proteomics, cells were harvested 16 hours after the start of the treatments, after which PM fractions and subsequently the PM proteins were extracted for Label-free comparative proteomics.

Using this approach, 1101 tomato proteins could be identified in all samples combined. From these, 1047 proteins were responding to *P. infestans* zoospore infection, of which 540 proteins also responded to treatments with supernatant of *P. infestans* zoospore cultures. For ten candidate proteins the homologs in *Arabidopsis* were identified and used for rapid follow up validation studies.



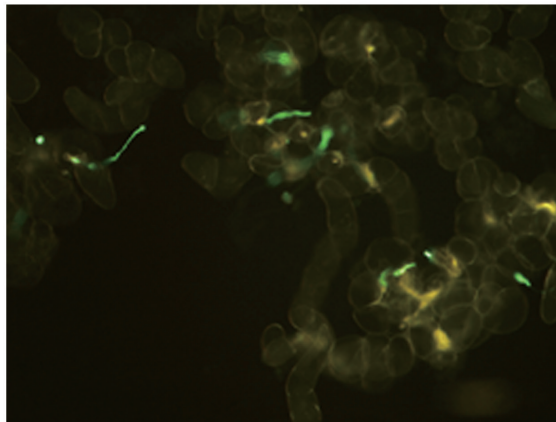
## Materials and methods

### *P. infestans* growth and inoculum preparation

*P. infestans* transformant 14-3-GFP was grown on rye sucrose agar at 18°C (Du et al., 2015). Zoospores were harvested as described previously by Champouret *et al.* (Champouret et al., 2009). Zoospore density was determined using a haemocytometer, and adjusted to an inoculum  $4 \times 10^6$  zoospores/mL. Supernatant fractions of zoospore suspensions were obtained by centrifugation at 4000 rpm and collection of supernatant.

### Culturing and inoculation of tomato MsK8 suspension cells

MsK8 suspension cells (Koornneef et al., 1987) were cultured in MS and vitamins medium (pH 5.7) supplemented with 3% (w/v) sucrose, 1 mg/L 2,4-Dichlorophenoxyacetic acid (2,4-D) and 0.1 mg/L kinetin at 24°C in the dark and shaking at 125 rpm. Fresh sub-cultures were initiated every week. Five days old cultures were used for inoculation with *P. infestans* zoospores. In total 50 ml of MsK8 suspension cells were inoculated with an equal volume of zoospore inoculum ( $4 \times 10^6$  zoospores/mL), and incubated with shaking (~ 80 rpm) at 18°C in the dark. MsK8 suspension cells were harvested 16 hours post inoculation. GFP fluorescence was used to monitor hyphae and appressoria to evaluate infection efficiency (Fig. 1). Medium was removed by fast vacuum filtration and remaining tomato MsK8 suspension cells were frozen in liquid nitrogen in small clumps (ca. 0.5 cm diameter) for easier homogenization.



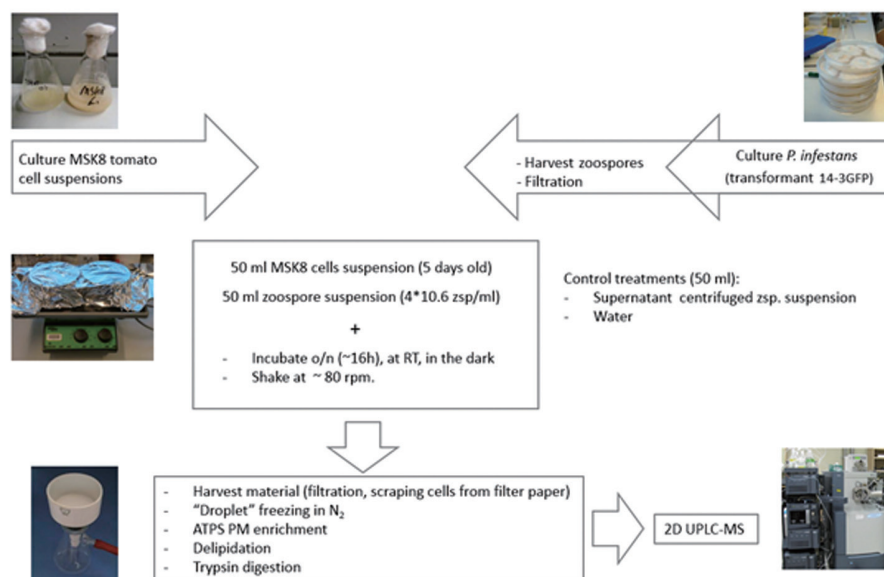
**Figure 1.** Tomato MsK8 suspension cells infected with *P. infestans* 14-3-GFP after 16hrs. GFP fluorescence can be observed in hyphae and appressoria penetrating MsK8 suspension cells. (Picture courtesy of K. Bouwmeester)

### ***Plasma membrane enrichment and PM protein isolation***

For enrichment of the plasma-membrane fraction, 50g liquid nitrogen frozen, pulverized cell fine powders were homogenized with 50 ml of ice-cold homogenization buffer (0.3 M sucrose, 50 mM 3-(N-morpholino) propanesulfonic acid (MOPS)-KOH, 5 mM Na-EDTA, 5 mM ascorbic acid, 5 mM dithiothreitol (DTT), 0.55% insoluble polyvinylpyrrolidone (PVPP), 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 7.5). The homogenate was filtered through 0.2 mm pore size Nylon and further clarified by centrifugation at 10,000 g for 30 min. Subsequently the supernatant was transferred into a ultracentrifuge tube (SS-34) and centrifuged at 100,000 g for 1 hr at 4°C (JA20, Beckman Ultra-centrifuge). The microsomal pellet was re-suspended in 2.5 mL re-suspension medium (0.3 M sucrose, 50 mM MOPS-KOH, 5 mM Na-EDTA, 1 mM DTT, pH 7.5 ) and three consecutive two-phase extractions were performed as described by Alexandersson *et al.* (Alexandersson *et al.*, 2004) to enrich for plasma membranes (PM). Therefore, 2.25 g of microsomal fraction was added to 6.75 g ATPS (6.4% Dextran T500, 6.4% PEG 3350, 2 mM KCl, 0.33 M sucrose), mixed well and centrifuged at 2,000 g for 10 min in order to allow separation of the Dextran and PEG phase. In the upper PEG phase, plasma membranes will be enriched, and the lower Dextran phase will be enriched in intracellular membranes. The upper phase was added to a fresh pretreated lower phase of ATPS, while the lower phase was added to a fresh PEG top-layer. After centrifugation the two-phase extraction was repeated two more times (three times ATPS were applied as discussed in Results and discussion) and the combined upper phases were added to 50 ml of resuspension buffer (0.3 M sucrose, 5mM potassium phosphate, 0.1 mM Na-EDTA, 1mM DTT, pH7.8). The triple fractionated PM fraction was pelleted by ultra-centrifugation at 100,000g for 1 hour. The PM pellet was re-suspended in 100  $\mu$ L sucrose buffer (0.3M sucrose, 5mM potassium phosphate (1:10 K<sub>2</sub>HPO<sub>4</sub>:KH<sub>2</sub>PO<sub>4</sub>, pH7.8), 0.1mM Na-EDTA, 1mM DTT).

## 5

For proteome analysis the PM pellet was washed three times with chloroform/methanol/MQ (1:4:5) for removal of lipids (Mirza *et al.*, 2007). Eventually, the protein pellets were dried and dissolved in 50  $\mu$ L protein dissolving buffer (0.1 M Tris-HCl (pH 8.0), 8 M urea and 5 mM DTT). After alkylation by iodoacetamide (IAA), 4 volumes 0.1 M ammonium bicarbonate buffer (pH 8.3) was added to dilute urea to 2 M for protein digestion. Protein concentration was measured by Bradford assay (Bradford, 1976). Protein was digested with Porcine, TPCK treated, trypsin (proteomics grade, Sigma T6567) (trypsin:protein=1:50) by overnight incubation at 37°C. Digestion was terminated by adding trifluoroacetic acid (TFA) to a final 0.1% v/v. After digestion, a reverse phase C18 solid phase extraction (SPE) column (Supelco) was applied for peptide purification. Peptides were eluted with 1 mL of 60% acetonitrile (ACN), 0.1% TFA. The purified peptides were dried by vacuum centrifugation and dissolved in 20  $\mu$ L of 0.1 M ammonium formate (pH 10) for 2D nano LC-MS analysis. In summary, the whole workflow including harvesting of zoospores and plasma membrane protein isolation is shown in Fig. 2:



**Figure 2.** Workflow plasma membrane protein sample preparation (Picture courtesy of K. Bouwmeester).

### *Peptide analysis using 2D nano LC-MS/MS and 2D nano LC-MSE*

For high-resolution separation of the complex plasma membrane peptide samples a nano ACQUITY™ 2D UPLC system (Waters Corporation, Manchester, UK) was used employing orthogonal reversed phase separation at high and low pH, respectively. With this 2-D set up, the pool of peptides was eluted from the first dimension XBridge™ C18 trap column (in 20 mM ammonium formate pH 10) using a discontinuous 7-step gradient of 12%, 15%, 18%, 20%, 25%, 35% and 65% ACN. For the second dimension an acidic ACN gradient was applied using a BEH C<sub>18</sub> column (75 μm × 25 cm, Waters, UK) and a 65 min linear gradient from 3 to 40% ACN (in 0.1 % FA) at 200 nL/min. The eluting peptides were on-line injected into a SYNAPT™ G1 Q-TOF MS instrument (Waters Corporation, Manchester, UK) using a nanospray device coupled to the second dimension column output. The SYNAPT MS was operated in positive mode with [Glu<sup>1</sup>] fibrinopeptide B (1 pmol/μL; Sigma) as reference (lock mass) sampled every 30s. Accurate LC-MS data were collected with the SYNAPT operating in either the MS/MS mode for data-dependent acquisition (DDA) or MS<sup>E</sup> for data-independent acquisition (DIA) using low (6 eV) and elevated (ramp from 15 to 35 eV) energy spectra every 0.6 s over a 400-1400 and 140-1900 m/z range, respectively. DDA was performed by peptide fragmentation on the three most intense multiply charged ions that were detected in the MS survey scan (0.6 s) over a 300-1400 m/z range and a dynamic exclusion window of 60 s with an automatically adjusted collision energy based on the observed precursor m/z and charge state.

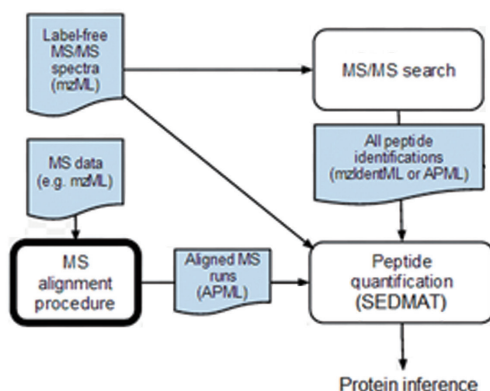
### ***Database search for protein identification***

LC-MS/MS and LC-MS<sup>E</sup> data were processed using ProteinLynx™ Global Server software (PLGS version 2.5, Waters Corporation) and the resulting list of masses containing all the fragment information was searched against the a combined non-redundant database, containing tomato protein sequence database downloaded from the *Sol Genomics Network* (<http://solgenomics.net/>) (iTAG2.3 from 20110503), *Phytophthora* database (DB t30-4-1 + gene index GI 070908 + Cogeme database Translated) and contaminants (keratins + trypsin + enolase from yeast). Peptide identification was done by matching the detected peptide fragmentation spectrum to that of predicted peptide spectra using the tomato protein sequence database as input in PLGS. For matching MS<sup>E</sup> data to predicted peptide spectra the following settings were used: minimum fragment ion matches per peptide: four, per protein: seven, minimum one peptide per protein match and a false discovery threshold of four percent. The false discovery rate was determined automatically in PLGS by searching the randomized tomato protein sequence database. Peptide and fragment tolerance was set to automatic: the software determines tolerance thresholds based on the actual resolution and stability in the MS data (Geromanos et al., 2009; Li et al., 2009). In practice this resulted in mass tolerances below 10 ppm for precursor and below 30 ppm for fragment ions. Carbamido-methylation of cysteine was set as fixed modification, and oxidation of methionine (M) and deamidation on asparagine (N) as variable modification. For DDA analysis the peptide tolerance was set to 30 ppm and a fragment tolerance of 0.05Da. Carbamidomethylation (Cys) was used as fixed modification and Deamidation (NQ), and Oxidation (M) as variable modifications. The AutoMod option was applied as secondary search to the database search results with a maximum of one missed trypsin cleavage and non-specific secondary digest reagent were chosen. Finally, the DDA and MS<sup>E</sup> outputs were merged in Excel.

## **5**

### ***Peptide/Protein Identification***

The proteomics procedure for protein identification and quantification is summarized Fig. 3. In brief, digested peptides were identified by PLGS either with DDA or MS<sup>E</sup> dataset. Protein quantification was done with MS<sup>E</sup> dataset in Progenesis LC-MS™ version 4.0 (Nonlinear Dynamics, UK). Quantification in Progenesis resulted in a list of masses (features) with differential abundance which can be further linked with protein identification from PLGS or Mascot, or our “in house tool” SEDMAT, as incorporated into Galaxy Proteomics toolbox (Lukasse and America, 2014) (with updated optimized parameters and additional filter setup), which provides improved confidence for the protein identifications matching to quantification.



**Figure 3.** Pipeline for label-free data in which peptide identification and peptide quantification are done in separate MS/MS and MS runs respectively. The “Peptide quantification” step in this case is SEDMAT which couples MS/MS peptide identifications to quantified MS features based on mass and retention time similarity.

### *Peptide/Protein quantification*

Protein quantification from raw MS<sup>E</sup> data was done in Progenesis LC-MS<sup>TM</sup> version 4.0 (Nonlinear Dynamics, UK) using lock-spray and dead time correction as mass calibration. Progenesis performed extraction of ion features such as mass, charge, intensity, retention time, etc, after retention time alignment and peak detection of all the samples. Total peptide mass peak intensities were normalized among different samples by the algorithm in Progenesis. A minimum one unique peptide (feature) per protein was employed with high peptide score and ranking for protein quantification. Protein intensity was calculated with sum of all unique peptides and shared peptides ion abundance. Further, statistical analysis was done in Progenesis with all the matched features from different samples and experiments. Features were clustered, grouped and labelled (tagged by different colors) according to fold change, similarity and *P*-value for significance. Specific peptide features which displayed significantly differential abundance among samples were selected for matching peptide/protein identification from PLGS. The non-matched specific features were exported as inclusion list to run additional dedicated DDA for protein identification. Subsequently, the peptide fragmentation spectra were exported as .pkl file and submitted to Mascot for protein identification and protein identification read back into Progenesis. Eventually, identified proteins which show significant differential abundance condition among samples were exported into MS Excel for further analysis.

Peptide mass features which showed a differential abundance profile could only partly be linked to protein identification in Progenesis with limited hits. For those

non-matching masses, an in-house built tool (implemented as the PRIMS package in Galaxy (available from Galaxy toolshed (Lukasse and America, 2014)) named SEDMAT (single experiment data matching tool) can be used. SEDMAT is a tool meant for peptide quantification and has an algorithm that assigns to each quantified MS feature its respective MS/MS identification based on mass and retention time similarity. The SEDMAT algorithm is also able to detect the retention time shifts between the MS and the MS/MS runs and correct for these automatically. In Galaxy, all the proteins identified by PLGS and features quantified/non-quantified by Progenesis were loaded and used for linking identification to quantification. SEDMAT will do the alignment for the features with its own algorithm, also for those already quantitatively identified features in Progenesis, then try to link the quantified peptide features to their protein identifications. Finally, after filtering by peptide scoring and ranking, all the quantified and identified proteins were combined into a single output file, in which each quantitatively identified peptide feature was annotated according to SEDMAT, Progenesis, or both.

### ***In silico candidate protein function study***

Candidate protein domains were obtained from SMART bioinformatics tool (<http://smart.embl-heidelberg.de/>) (Schultz et al., 1998; Letunic et al., 2015). *Arabidopsis* homologs of tomato candidate proteins were retrieved from TAIR BLAST (Huala et al., 2001) and the protein description was provided based on the gene annotation (GO: <http://www.geneontology.org/>) (Ashburner et al., 2000; Gene Ontology Consortium, 2015). *Arabidopsis* homolog gene expression data under different stress condition was retrieved from the Bio-Analytic Resource for Plant Biology (BAR) under *Arabidopsis* eFP browser (<http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi>) (Winter et al., 2007).

## 5

## **Results and discussion**

### ***Optimization of plasma membrane fractionation of plant materials***

Here we set out to use PM proteomics to study changes in the proteome of plants with or without infection by *P. infestans*. The PM enrichment procedure is based on two principles: separation of PM derived vesicles from other membrane vesicles by centrifugation in an optimized Aqueous Two-Phase (PEG/Dextran) Partitioning System (ATPS) and the number of repeats of the ATPS (Alexandersson et al., 2004). For the isolation of PM fractions we first needed to determine the optimal two-phase partitioning conditions for the plant material initially used in this study, i.e. potato leaves infected or not infected by *P. infestans*. Different percentages of PEG/Dextran (5.8%, 6.0%, 6.2%, 6.4%, 7.0%) were tested for enrichment of membrane vesicles isolated from potato leaf material (Fig. S1). The relative yield of PM proteins in the various PM fractions was evaluated by relative quantitation of identified known PM proteins using LC-MS (data not shown). None of the tested two-phase partitioning conditions resulted in a satisfactory enrichment of PM proteins from potato leaf.

Therefore, as alternative to PM proteomics on potato leaves, we tested PM isolation from tomato MsK8 suspension cells, which are also susceptible to infection by *P. infestans* (Felix et al., 1991). Different percentages of PEG/Dextran and numbers of repeats of the two-phase partitioning were tested for optimal fractionation of PM vesicles from MsK8 suspension cells. This showed that the total protein yield of the final PM fraction was very sensitive to the percentage of PEG/Dextran. For instance, the total yield of proteins varied more than two-fold between the use of 6.4% and 6.2% PEG/Dextran. To validate the enrichment of PM proteins, protein fractions were analyzed both by SDS-PAGE (Fig. S2) and by LC-MS. As an example, Table S1 lists a set of proteins identified in both 6.4% and 6.2% PEG/Dextran PM protein fractions analyzed by LC-MS, which are higher in 6.4% and of which approx. 50% are predicted as PM proteins. The SDS-PAGE gel (Fig. S2) shows several high molecular weight protein bands in the PM fraction, which are absent from the soluble protein fraction. Combined, these results suggest an enrichment of PM proteins using the 6.4% PEG/Dextran step gradient compared to the 6.2%. Although repeating ATPS application substantially reduced total protein yield. Eventually three cycles of ATPS application were used to balance PM enrichment and yield. During the PM isolation small vesicles are formed, which may capture contaminant soluble proteins. A treatment of the final PM fraction with Brij 58 detergent results in vesicle inversion, thereby releasing captured soluble protein contaminants (Johansson et al., 1995). However, in our hands the Brij 58 treatment also resulted in a major reduction of membrane protein yield from the PM fractions of the MsK8 suspension cells. Therefore, we opted to omit the Brij 58 treatment. Instead, we used *in silico* filtering of LC-MS identified proteins by selecting proteins with a predicted transmembrane domain and/or GPI-anchor.

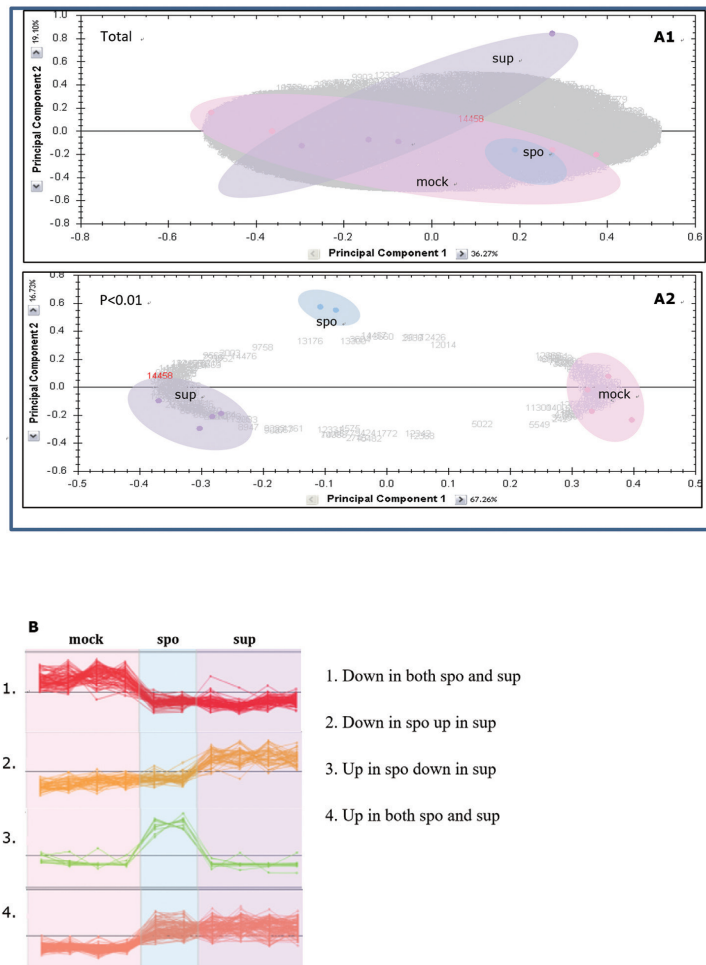
### ***Plasma membrane proteins identified in MsK8 suspension cells***

After optimization of the PM fractionation procedure, MsK8 suspension cells were either mock-inoculated, treated with zoospore supernatant or inoculated with a zoospore suspension (at  $4 \times 10^6$  spores/mL). For each treatment four independent replicate experiments were performed and cells from each of the treatments were harvested at 16 hours after inoculation for isolation of the PM fractions and extraction of proteins (see Material and methods). For each sample, 100  $\mu$ g of protein was digested with trypsin and peptides were purified using C18 SPE. Peptide analysis was carried out on 2D nano LC-MS using a seven step gradient (Fig. S3) in the first dimension (see Material and methods). Peptide identification was done using DDA and MS<sup>E</sup> by matching measured peptide spectra with predicted spectra from tomato proteins or *P. infestans* proteins. Overall, 94189 features were detected from DDA and MS<sup>E</sup> in all 12 samples combined. From these only 15754 features (17%) could be linked to 1698 proteins from either the tomato or *P. infestans* database. The complete list of proteins identified is shown in Table S2. From the 1698 identified proteins, 1101 proteins (12622 features) were from tomato and 597 proteins (3132 features) from *P. infestans*.

### ***Quantitative peptide feature analysis***

For quantitative analysis of the peptides, the MS<sup>E</sup> data were loaded into Progenesis software. Progenesis performs peak detection and quantitation of the multiple isotope peaks per peptide, which are integrated and reported together as a single feature. For each of the seven fractions obtained by 2D nano-LC the samples had to be aligned 4 × mock treated control (mock-samples), 4 × infected with zoospore (spo-samples) and 4 × infected with zoospore supernatant (sup-samples). After applying the retention time shift correction, all peptide features from the seven analyzed fractions were combined in Progenesis, which allows comparison between the different samples (replicates/treatments) and statistical analysis. Results showed that in the total data set, two of the four spo-samples were deviating too much from the other samples, due to too low total protein loading, resulting in a very reduced number of detected features. Therefore, these two deviating spo-samples were omitted for further analysis. The peak intensities of the remaining samples were normalized using total peak intensity. Analysis of the samples by PCA (Fig. 4A) shows that the two remaining spo-samples show little variation (example shown is fraction 5 of the 2D nano-LC), while the replicate samples of mock-infected and sup-treated tomato cells display a higher level of variance. This was an unexpected result since we assumed the highest variation in the spo-samples because of possible variation in infection efficiency by *P. infestans* zoospores. Fig. 4A2 shows the PCA from the same data set, but now filtered in ANOVA for *P*-value < 0.01 for features in replicate samples. The PCA shows clustering of the replicate samples in different positions in the PCA plot, indicating consistent difference in the PM proteome of mock, spo- and sup-samples. Clustering of the quantitative features detected in all seven fractions in Progenesis resulted in four major protein abundance profiles over the sample sets: (1) down in both spo and sup compared to mock samples; (2) down in spo and up in sup samples compared to mock samples; (3) up in spo and down in sup samples compared to mock samples and (4) up in both spo and sup samples compared to mock samples. As example the sorted expression profiles of proteins detected in fraction 5 is shown (Fig. 4B). The combined data contain features from proteins derived from tomato and *P. infestans* (in spo-samples and possibly also in sup-samples). Features with expression profile 3 (up in spo-samples) could therefore be mostly from *P. infestans* proteins. Features with expression profile 1 and 4 are from peptides/proteins with similar response in sup- and spo-samples. Curiously, there were no major group of features/proteins with an expression profile of ‘only low in spo-samples’ nor an expression profile with ‘only low in sup-samples’.





**Figure 4.** A: PCA of multiple alignment datasets of LC-MS runs (fraction 5) of the three treatments. Spo: infected with zoospore ( $2\times$ ); Sup: infected with zoospore supernatant ( $4\times$ ); Mock: mock treated control ( $4\times$ ) (A1: total features, A2: selected features with  $P$ -value  $< 0.01$ ); B: selected features can be clustered into four abundance profiles.

### *Linking feature quantification to protein identification*

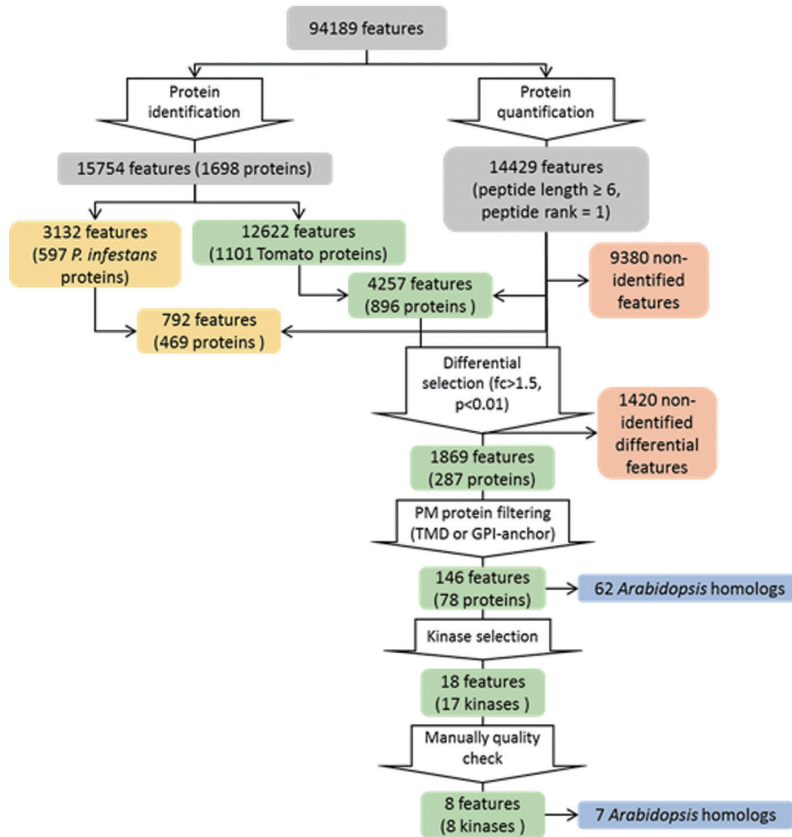
In total, 14429 features from all LC-MS measurements could be quantified by Progenesis and/or SEDMAT). From these, 4257 quantified features linked to 896 tomato proteins, while 792 linked to 469 proteins from *Phytophthora*. From the 896 tomato proteins (4257 features), 287 proteins (1869 features) showed a significant

differential abundance between mock and one or two of the treated samples (minimum 1.5 fold change with  $P$ -value  $< 0.01$ ). This set of 287 identified and quantified proteins contains 78 putative PM proteins, based on filtering for predicted TMD and GPI-anchor site (Table 1). From these 78 proteins, 73 proteins had at least one or more trans-membrane domain, while five proteins had a putative GPI-anchor attachment AA sequence. For a quick putative functional inventory of the 78 candidate tomato proteins, we searched for the nearest *Arabidopsis* homologs. In total 62 *Arabidopsis* homologs of the 78 tomato proteins were found, because a number of different tomato proteins had the same *Arabidopsis* homolog (Table 1). The overall step-wise selection is summarized in Fig. 5, in which the numbers of identified or quantified peptide features or proteins along with each selection step are listed.

**Table 1.** 78 tomato plasma membrane proteins with differential abundance profile between the mock-, spo- and sup-samples. The 78 tomato plasma membrane proteins with differential abundance profile in the MsK8 suspension cell (mock), MsK8 suspension cell infected with *P. infestans* zoospore (spo) and MsK8 suspension cell treated by *P. infestans* zoospore supernatant (sup) samples are grouped in four abundance profiles. The closest *Arabidopsis* homologs are shown. Colors of the cells are based on protein functional annotation: grey: unknown protein, orange: cell wall biosynthesis related, yellow: kinases, light blue: water transport/stress related, green: transporter or translocating activity related, brown: protein degradation related, dark blue: oxidase activity related. \* biotic stress responses in *Arabidopsis* are based on transcript data information in the TAIR gene locus site: N: not detected in experiments using different biotic stresses; Y: transcriptional response to one or more biotic stresses. N/A: no close *Arabidopsis* homolog.

Tomato protein	<i>Arabidopsis</i> homolog	Transcr biotic stress *	Tomato gene description	Abundance profile	
Solyc08g029230.2	AT5G43100	N	Eukaryotic aspartyl protease family protein;	(1) down in both spo and sup	
Solyc02g080500.2	AT1G61900	N	Unknown protein		
Solyc01g103680.2	AT4G24330	Y	Unknown protein		
Solyc05g054440.2	AT3G55430	Y	O-Glycosyl hydrolases family 17 protein		
Solyc12g056580.1	AT2G21770	N	Cellulose synthase, related to CESA6.		
Solyc11g010290.1	AT5G64290	Y	Dicarboxylate transport 2.1 (DIT2.1)		
Solyc02g089440.2	AT3G02350	N	Encodes a protein with putative galacturonosyltransferase activity.		
Solyc03g033320.2	N/A		Prolyl 4-hydroxylase alpha subunit-like protein		
Solyc07g006800.1	AT2G15480	Y	UDP-glucosyl transferase 73B5 (UGT73B5) , quercetin 3-O-glucosyltransferase activity. Active in redox regulation		
Solyc02g091590.2	AT2G29000	N	Leucine-rich repeat protein kinase family protein		
Solyc03g121050.2	AT1G51940	Y	Encodes a LysM-containing receptor-like kinase 3		
Solyc10g084120.1	AT3G53420	Y	A member of the plasma membrane intrinsic protein subfamily PIP2.		
Solyc04g063230.2	N/A		Dehydration-responsive family protein		
Solyc11g062130.1	N/A		Mitochondrial ADP/ATP carrier proteins		
Solyc11g062190.1	N/A		Mitochondrial ADP/ATP carrier proteins		
Solyc01g106570.2	AT2G14520	N	Unknown protein		(2) down in spo and up in sup
Solyc01g088500.2	AT3G17350	N	Unknown protein		
Solyc03g111560.2	AT2G36850	Y	Encodes GSL8, a member of the Glucan Synthase-Like (GSL)		
Solyc03g121230.2	AT4G29990	N	Barely any meristem3, Leucine-rich repeat transmembrane protein kinase		
Solyc01g080770.2	AT4G20270	Y	Encodes a CLAVATA1-related receptor kinase-like protein		
Solyc01g103530.2	AT4G23180	Y	Encodes a receptor-like protein kinase		
Solyc02g080070.2	AT4G23180	Y	Encodes a receptor-like protein kinase		
Solyc04g039730.2	AT1G60800	Y	NSP-interacting kinase 3 (NIK3)		
Solyc05g005140.2	AT1G60800	Y	NSP-interacting kinase 3 (NIK3)		
Solyc04g006930.2	AT1G68690	N	Encodes a member of the proline-rich extensin-like receptor kinase (PERK) family		
Solyc06g063360.2	AT1G78980	N	STRUBBELIG-receptor family 5 (SRF5)		
Solyc04g008310.1	AT4G34131	Y	UDP-glucosyl transferase 73B3 (UGT73B3) , quercetin 3-O-glucosyltransferase activity. Active in redox regulation		
Solyc01g111660.2	AT4G35100	N	A member of the plasma membrane intrinsic protein PIP		
Solyc10g055630.1	AT4G35100	N	A member of the plasma membrane intrinsic protein PIP		
Solyc08g023440.2	AT1G30360	N	Early-responsive to dehydration 4 (ERD4)		
Solyc12g008360.1	AT5G19690	Y	Encodes an oligosaccharyl transferase involved response to high salt		
Solyc12g017450.1	AT1G16860	N	Ubiquitin-specific protease family C19-related protein	(3) up in spo and down in sup	
Solyc11g022540.1	AT1G10950	N	Encodes an <i>Arabidopsis</i> Transmembrane nine (TMN) protein		
Solyc12g088720.1	AT4G34110	Y	Putative poly-A binding protein		
Solyc06g065830.1	AT5G35160	N	Endomembrane protein 70 protein family		
Solyc09g009150.1	AT5G03700	N	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain		
Solyc04g078750.2	AT1G45688	N	Unknown protein		

Solyc03g078640.1	AT4G39840	N	Unknown protein
Solyc04g080370.2	AT1G47330	N	Unknown protein
Solyc11g008960.1	AT5G10290	N	Leucine-rich repeat transmembrane kinase family protein
Solyc09g083210.2	AT2G33580	Y	Encodes a LysM-containing receptor-like kinase 5
Solyc02g081070.2	AT4G22130	Y	STRUBBELIG-receptor family 8 (SRF8)
Solyc10g074940.1	AT5G04930	Y	Encodes a putative aminophospholipid translocase (p-type ATPase)
Solyc03g098730.1	AT1G17860	Y	Kunitz family trypsin and protease inhibitor protein
Solyc12g006200.1	AT4G12700	N	Unknown protein
Solyc07g045210.1	AT2G04280	N	Unknown protein
Solyc03g116590.2	AT5G62200	Y	Unknown protein
Solyc07g049180.2	AT3G21630	Y	LysM receptor-like kinase1
Solyc10g044740.1	AT3G21630	Y	LysM receptor-like kinase1
Solyc01g108840.2	AT1G66150	Y	Receptor-like transmembrane kinase I (TMK1)
Solyc04g015600.2	AT2G41820	N	Leucine-rich repeat kinase family protein
Solyc08g081940.2	AT4G23740	N	Leucine-rich repeat kinase family protein
Solyc10g054560.1			
Solyc10g054570.1	AT2G16510	Y	ATPase, F0/V0 complex, subunit C protein
Solyc10g054590.1			
Solyc01g111980.2	AT1G47670	N	Transmembrane amino acid transporter family protein
Solyc02g084360.2			
Solyc04g081090.2	AT1G19910	N	Vacuolar H <sup>+</sup> -pumping ATPase 16 kDa proteolipid (ava-p2)
Solyc09g075820.2	AT5G26340	Y	Encodes a protein with high affinity, hexose-specific/H <sup>+</sup> symporter activity
Solyc05g054800.1	AT3G12740	N	Physically interacts with ALA3
Solyc09g090730.1	AT4G13510	Y	Encodes a plasma membrane localized ammonium transporter
Solyc01g105450.2			
Solyc05g051530.2	AT1G17840	Y	ATP-BINDING CASSETTE G11
Solyc11g065360.1	AT3G21090	N	ATP-BINDING CASSETTE G15: export plant cuticular waxes together with ABC-G11
Solyc02g079220.2	AT1G11260	Y	Encodes a H <sup>+</sup> /hexose cotransporter.
Solyc06g075570.1			
Solyc11g072800.1	AT5G60010	N	Ferric reductase-like transmembrane component family protein
Solyc02g088390.2	AT5G15350	Y	Early nodulin-like protein 17 (ENODL17)
Solyc01g099620.2	AT1G09090	N	NADPH-oxidase AtRbohB plays a role in seed after-ripening
Solyc03g117980.2			
Solyc06g068680.2	AT5G47910	Y	NADPH/respiratory burst oxidase homologue D (RbohD)
Solyc08g081690.2	AT1G64060	Y	Respiratory burst oxidase homologue F; in <i>Arabidopsis thaliana</i> AtRbohF interacts with AtRbohD gene to fine tune the spatial control of ROI production and hypersensitive response to cell in and around infection site.
Solyc02g036480.1	AT5G36970	Y	NDR1/HIN1-like protein, expression induced during incompatible response to a pathogen
Solyc01g006950.2	AT3G11820	Y	Encodes a syntaxin localized at the plasma membrane (SYR1)
Solyc02g069150.2			
Solyc09g061620.2	AT1G04760	N	Member of Synaptobrevin -like protein family
Solyc09g091610.2			
Solyc03g123920.2	AT1G78880	N	Ubiquitin-specific protease family C19-related protein



**Figure 5.** Step-wise selection of protein candidates from tomato and *Arabidopsis* homologs. Note that for quantification the software selects a minimum of one unique peptide/feature per protein, while protein identification is based on multiple peptides/features. TMD: transmembrane domain; GPI-anchor: glycosphosphatidylinositol anchor.

### ***Tomato proteins with similar expression in spo- and sup-samples***

The final selection of 78 PM proteins mainly function in transferase, kinase, transporter or binding activities (Fig. S5; Table 1) and can be grouped according to four major protein abundance profiles (Table 1). Of 1420 differential features the corresponding protein could not be identified, and thus there are at least 100 additional proteins with potential differential abundance which remain unidentified. Therefore we should be cautious by giving too much meaning to protein functional differences between groups, but nevertheless there are some striking differences in protein functions between the four protein-abundance-profile groups.

For example, protein abundance profiles 1 (high in mock, low in spo and sup samples) and 4 (low in mock and high in spo and sup samples) contain proteins with a similar response in spo- and sup- samples, indicating that these respond to factors common to both treatments. Most Leucine-rich repeat receptor kinases (LRR-RLKs) fall in these two groups (Table 1; highlighted in yellow), suggesting that these are kinases that respond to factors both generated by *P. infestans* infection and factors present in *P. infestans* culture supernatant.

Six proteins with functions related to carbohydrate and cell wall metabolism show reduced abundance in condition 1 (down in both spo and sup). This suggests that factors from *P. infestans* infection and supernatant may have the potential to suppress these activities, which could help establish a successful infection. The cell wall is the first barrier that pathogens need to destroy for a successful infection with their degrading enzymes. And plants have developed a mechanism to sense the pathogen and cell wall integrity (Bellincampi et al., 2014).

## 5

Multiple proteins with high abundance in spo and sup samples function in transport over the PM membrane (highlighted in green in Table 1). Two of these are ABC transporter proteins are homologous to the *Arabidopsis* ABC-G11 and ABC-G15 which are involved in transport of waxes to the cuticle (McFarlane et al., 2010). The cuticle provides a physical barrier for plant protection against pathogens (Serrano et al., 2014) and ABC transporters in plants have been implicated in plant defense responses (Yazaki, 2006; Krattinger et al., 2009; Kang et al., 2011).

Another group of proteins showing up-regulation in both spo- and sup-samples are annotated as oxidases. For example, one of these shows homology to respiratory burst oxidase homologue D (RbohD), which plays a key role in plant immunity. The Rbohs are involved in generating ROS triggered by PAMPs during the hypersensitive response (HR) (Torres et al., 2006). *Arabidopsis* RbohD (AtRbohD) was initially described as a key component of plant defense (Torres et al., 2002). Our results suggest that a tomato Rboh is induced both by *P. infestans* zoospores and by supernatant. Overall, based on comparison with homologs in *Arabidopsis*, most identified tomato PM protein with similar abundance profile in spo- and sup-samples may indeed play a role in the plant pathogen interaction.

### ***Tomato proteins only up in sup- and spo-samples***

Proteins abundance profiles 2 and 3 show a unique abundance induction for proteins in sup-samples (profile 2) or in spo-samples (profile 3). There are number of receptor kinases that are involved in the regulation of plant growth and development, that show a uniquely high abundance in the sup-samples, such as CLAVATA1-related receptor kinase and STRUBBELIG-receptor protein. STRUBBELIG has been reported to be involved in plant-pathogen interactions (Eyüboğlu et al., 2007). Our study indicates that a STRUBBELIG-like receptor encoding gene is also induced by *P. infestans* zoospore supernatant in tomato, but not by infection with *P. infestans* zoospores. Another group of proteins showing high abundance specific in the sup-treatment (profile 2) are water stress related (highlighted in light blue), such as PLASMA MEMBRANE INTRINSIC PROTEIN (PIP) and EARLY-RESPONSIVE TO DEHYDRATION 4 (ERD4). This suggests that some secreted factors in *P. infestans* zoospore supernatant alone can already cause an osmotic stress response in the tomato cells. And one tomato protein shows homology to a UDP glycosyltransferases from *Arabidopsis* which has been implemented in the redox response to *Pseudomonas syringae* pv. *tomato* (Simon et al., 2014).

### ***Manual validation of eight selected kinases***

The multiple steps in LC-MS data processing may result in false positives if for instance samples from different treatments are misaligned. Manual validation of all 78 candidate proteins would be too labor intensive, but for eight protein kinases we validated the quality of identification and quantification by manually checking peptide spectra alignment and peptide abundance profiles in Progenesis (Fig. S4; Table 1).

As shown in Table 2, half of these protein kinases were identified with multiple peptide hits, while four were only identified with one peptide. As described in the Materials and methods, a protein can be quantified with a minimum of one unique peptide. After manual peptide quality checking, only one out of eight proteins turned out to be identified using two confident (manually confirmed) peptides, while the others were identified on the basis of one confident peptide. This indicates that either Progenesis or SEDMAT makes mistake for peptide peak alignment and feature quantitation which is resulting in false positives, due to the vast sample complexity. However, with the manual quality check, these false positives can be removed.

For the 8 validated tomato kinases, the protein abundance profile in tomato was compared to the transcript profile of their closest *Arabidopsis* homologs. For this purpose we used the *Arabidopsis* homolog gene expression response to *P. infestans* or *Pseudomonas syringae* infection and to different elicitor treatments from selected *Arabidopsis* transcriptome databases. In addition, because multiple of the identified tomato proteins function in responses to water stress (either salt stress, drought or osmotic stress), we also searched for the expression responses of the *Arabidopsis* homologs under water stress (Table 3) (<http://bar.utoronto.ca/efp/cgi-bin/efpweb.cgi>)

(Winter et al., 2007).

Table 3 shows that for most tomato proteins that belong to a single expression class, the transcriptional response to (a)biotic stress of the *Arabidopsis* homologs show similar patterns. For two of the proteins with opposite response (but similar in spo- and sup-samples), the expression of the two *Arabidopsis* homologs was also opposite and mostly similar under biotic and abiotic stresses. Both of these tomato proteins show homology to *Arabidopsis* proteins functioning in chitin responses: one is homologous to LysM-containing Receptor Kinase 3 (encoded by *Arabidopsis* AT1G51940), the other is homologous to Chitin Elicitor Receptor Kinase1 (encoded by *Arabidopsis* AT3G21630). In *Arabidopsis* both of these proteins have been shown to be involved in responses to chitin elicitor treatments. LysM-containing Receptor Kinase 3 has been shown to regulate a cross talk between biotic response to chitin and ABA (Paparella et al., 2014) and is down regulated in response to chitin and osmotic stress.

The expression of AT1G51940 and AT3G21630 in *Arabidopsis* fits well with the expression profile of the tomato proteins in spo- and sup-samples (down for the protein with homology to LysM-containing Receptor Kinase 3 and up for the protein with homology to Chitin Elicitor Receptor Kinase1). However, we realize that caution should be taken in interpreting this comparison. First of all, the comparison is tomato proteins versus *Arabidopsis* mRNA; second, many of the responses at the protein level may not occur at the mRNA level (or vice versa), and third, also the time frame of 16 hour treatment used in the MsK8 suspension cells may be different from the time frame used for transcript profiling in *Arabidopsis* plants. Moreover, biotic defense may be different in *Arabidopsis thaliana* and *Solanaceae*. However, because validation experiments in tomato are more time-consuming than in *Arabidopsis*, the comparison with *Arabidopsis* genes may help select the most promising tomato candidate genes for more effective validation studies.

## 5

**Table 2.** Eight tomato candidate proteins manually validated in Progenesis. Manually validated tomato proteins and their seven homologs from *Arabidopsis*. Proteins are clustered according to four expression profiles in the tomato samples. \* indicates that the same peptide was used for protein quantification.



Conditions	<i>Solanaceae</i> accession	No. peptides for protein identification	No. peptides for protein quantification	No. confidence peptide	Description <i>Arabidopsis</i> homolog
(1) down in both spo and sup	Solyc03g121050.2	8	2	1	LysM-containing receptor-like kinase3 Induction of chitin-responsive genes by chitin treatment is not blocked in the mutant
(2) down in spo up in sup	Solyc01g080770.2	24	2	1*	BARELY ANY MERISTEM 3 (BAM3); Clavata 1 related receptor
	Solyc01g103530.2	22	2	1*	
(3) up in spo down in sup	Solyc11g008960.1	13	1	1	Leucine-rich repeat transmembrane protein kinase, implicated in mediating Plasma Membrane-Cell Wall Adhesions in <i>Arabidopsis</i> .
	Solyc09g083210.2	8	1	1	LysM-containing receptor-like kinase5, major chitin receptor and forms a chitin-induced complex with related kinase CERK1.
(4) up in both spo and sup	Solyc01g108840.2	13	1	1	Transmembrane kinase 1 (TMK1), TMK1 functions in growth regulation in plants
	Solyc04g015600.2	10	1	1	PXC3, PXY/TDR-CORRELATED, Leucine-rich repeat kinase family protein: regulator of secondary cell wall biosynthesis
	Solyc07g049180.2	22	2	2	LysM-containing receptor-like kinase1 Chitin elicitor receptor kinase 1 (CERK1), CERK1 phosphorylates LIK1, a LLR-RLK that is involved in innate immunity

**Table 3.** Comparison of eight tomato kinases with transcriptome data of their seven *Arabidopsis* homologs. *Arabidopsis* gene expression data are from Bio-Analytic Resource for Plant Biology (BAR). Heatmap of transcription data shows down regulation in blue and up regulation in red.

Stress		Condition/Gene		(1) down in both spo and sup		(2) down in spo up in sup		(3) up in spo down in sup		(4) up in both spo and sup		
		AT1G51940	AT4G20270	AT5G10290	AT2G33580	AT1G66150	AT2G41820	AT3G21630				
Abiotic	Osmotic (300mM Mannitol)	0.19 (3hr)	0.34 (12hr)	0.81 (1hr)	5.47 (24hr)	0.3 (3hr)	0.42 (3hr)	1.95 (3hr)				
	Salt (150mM NaCl)	1.89 (3hr)	1.11 (0.5hr)	1.22 (0.5hr)	2.25 (24hr)	0.54 (3hr)	1.22 (1hr)	1.28 (6hr)				
	Drought	0.52 (0.25hr)	1.53 (0.25hr)	1.48 (0.25hr)	3.57 (1hr)	0.45 (0.5hr)	1.35 (0.25hr)	1.33 (0.5hr)				
Biotic	<i>P. syringae</i> (vir)	0.15 (24hr)	1.62 (2hr)	1.87 (24hr)	1.72 (2hr)	0.54 (2hr)	0.07 (24hr)	1.75 (2hr)				
	<i>P. syringae</i> (avr)	0.43 (6hr)	1.59 (2hr)	2.33 (24hr)	2.05 (6hr)	0.58 (24hr)	0.38 (24hr)	1.62 (2hr)				
	Elicitor (1µM Flg22)	0.43 (4hr)	0.6 (1hr)	2.04 (4hr)	2.04 (4hr)	0.54 (1hr)	0.71 (4hr)	2.61 (1hr)				
	Elicitor (10µM HrpZ)	0.35 (4hr)	0.7 (1hr)	1.39 (4hr)	2.53 (4hr)	0.6 (1hr)	0.47 (1hr)	2.86 (1hr)				
	Elicitor (1µM GST-NPP1)	0.47 (4hr)	0.88 (4hr)	1.25 (4hr)	1.62 (1hr)	0.54 (4hr)	1.62 (1hr)	2.12 (1hr)				
	<i>P. infestans</i> (10 <sup>6</sup> spo/mL)	0.39 (6hr)	0.62 (24hr)	1.35 (6hr)	2.44 (24hr)	0.44 (6hr)	0.51 (6hr)	1.85 (6hr)				

## Acknowledgements

This project was carried out within the research program of the Centre of BioSystems Genomics (CBGS) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

## References

**Alexandersson E, Saalbach G, Larsson C, Kjellbom P** (2004) *Arabidopsis* plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol* **45**: 1543–56

**Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al** (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–9

**Bellincampi D, Cervone F, Lionetti V** (2014) Plant cell wall dynamics and wall-related susceptibility in plant-pathogen interactions. *Front Plant Sci* **5**: 228

**Boller T, Felix G** (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu Rev Plant Biol* **60**: 379–406

**Bradford MM** (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**: 248–254

**Champouret N, Bouwmeester K, Rietman H, van der Lee T, Maliepaard C, Heupink A, van de Vondervoort PJI, Jacobsen E, Visser RGF, van der Vossen EAG, et al** (2009) *Phytophthora infestans* isolates lacking class I *ipiO* variants are virulent on *Rpi-blb1* potato. *Mol Plant Microbe Interact* **22**: 1535–45

**Du Y, Mpina MH, Birch PRJ, Bouwmeester K, Govers F** (2015) *Phytophthora infestans* RXLR effector AVR1 interacts with exocyst component Sec5 to manipulate plant immunity. *Plant Physiol* **169**: pp.01169.2015

**Eyüboğlu B, Pfister K, Haberer G, Chevalier D, Fuchs A, Mayer KF, Schneitz K, Walker J, Zhang R, Becraft P, et al** (2007) Molecular characterisation of the *STRUBBELIG-RECEPTOR FAMILY* of genes encoding putative leucine-rich repeat receptor-like kinases in *Arabidopsis thaliana*. *BMC Plant Biol* **7**: 16

**Felix G, Grosskopf DG, Regenass M, Basse CW, Boller T** (1991) Elicitor-induced ethylene biosynthesis in tomato cells: characterization and use as a bioassay for elicitor action. *Plant Physiol* **97**: 19–25

**Gene Ontology Consortium** (2015) Gene ontology consortium: going forward. *Nucleic Acids Res* **43**: D1049–56

**Geromanos SJ, Vissers JPC, Silva JC, Dorschel CA, Li GZ, Gorenstein M V, Bateman RH, Langridge JI** (2009) The detection, correlation, and comparison of

peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**: 1683–1695

**Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, et al** (2001) The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res* **29**: 102–5

**Johansson F, Olbe M, Sommarin M, Larsson C** (1995) Brij 58, a polyoxyethylene acyl ether, creates membrane vesicles of uniform sidedness. A new tool to obtain inside-out (cytoplasmic side-out) plasma membrane vesicles. *Plant J* **7**: 165–73

**Kang J, Park J, Choi H, Burla B, Kretschmar T, Lee Y, Martinoia E** (2011) Plant ABC transporters. *Arabidopsis Book* **9**: e0153

**Koornneef M, Hanhart CJ, Martinelli L** (1987) A genetic analysis of cell culture traits in tomato. *Theor Appl Genet* **74**: 633–41

**Krattinger SG, Lagudah ES, Spielmeier W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B** (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. *Science* **323**: 1360–3

**Letunic I, Doerks T, Bork P** (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* **43**: D257–60

**Li GZ, Vissers JPC, Silva JC, Golick D, Gorenstein M V, Geromanos SJ** (2009) Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**: 1696–1719

**Lukasse PNJ, America AHP** (2014) Protein inference using peptide quantification patterns.

**Macho AP, Zipfel C** (2014) Plant PRRs and the activation of innate immune signaling. *Mol Cell* **54**: 263–72

**McFarlane HE, Shin JJH, Bird DA, Samuels AL** (2010) *Arabidopsis* ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. *Plant Cell* **22**: 3066–75

**Mirza SP, Halligan BD, Greene AS, Olivier M** (2007) Improved method for the analysis of membrane proteins by mass spectrometry. *Physiol Genomics* **30**: 89–94

**Paparella C, Savatin DV, Marti L, De Lorenzo G, Ferrari S** (2014) The *Arabidopsis* LYSIN MOTIF-CONTAINING RECEPTOR-LIKE KINASE3 regulates the cross talk between immunity and abscisic acid responses. *Plant Physiol* **165**: 262–76

**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

**Schultz J, Milpetz F, Bork P, Ponting CP** (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**: 5857–64

**Serrano M, Coluccia F, Torres M, L'Haridon F, Métraux J-P** (2014) The cuticle and plant defense to pathogens. *Front Plant Sci* **5**: 274

**Simon C, Langlois-Meurinne M, Didierlaurent L, Chaouch S, Bellvert F, Massoud K, Garmier M, Thareau V, Comte G, Noctor G, et al** (2014) The secondary metabolism glycosyltransferases UGT73B3 and UGT73B5 are components of redox status in resistance of *Arabidopsis* to *Pseudomonas syringae* pv. *tomato*. *Plant Cell Environ* **37**: 1114–29

**Torres MA, Dangl JL, Jones JDG** (2002) *Arabidopsis* gp91<sup>phox</sup> homologues *AtrbohD* and *AtrbohF* are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proc Natl Acad Sci* **99**: 517–522

**torres MA, Jones JDG, Dangl JL** (2006) Reactive oxygen species signaling in response to pathogens. *PLANT Physiol* **141**: 373–378

**Win J, Chaparro-Garcia A, Belhaj K, Saunders DGO, Yoshida K, Dong S, Schornack S, Zipfel C, Robatzek S, Hogenhout SA, et al** (2012) Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb Symp Quant Biol* **77**: 235–47

**Winter D, Vinegar B, Nahal H, Ammar R, Wilson G V., Provart NJ** (2007) An “electronic fluorescent pictograph” browser for exploring and analyzing large-scale biological data sets. *PLoS One* **2**: e718

**Yazaki K** (2006) ABC transporters involved in the transport of plant secondary metabolites. *FEBS Lett* **580**: 1183–1191

## Supplementary Data

Supplementary data to this article can be found online at:

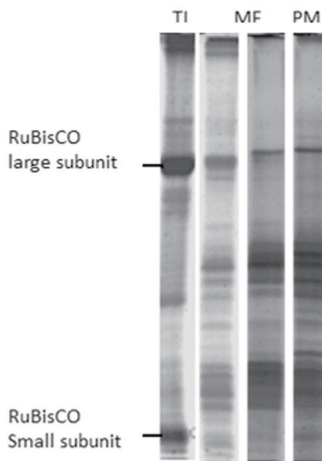
[https://1drv.ms/f/s!Asszhx\\_to9U8hxXix8NpPWFDJNdA](https://1drv.ms/f/s!Asszhx_to9U8hxXix8NpPWFDJNdA)

or by scanning the QR code:

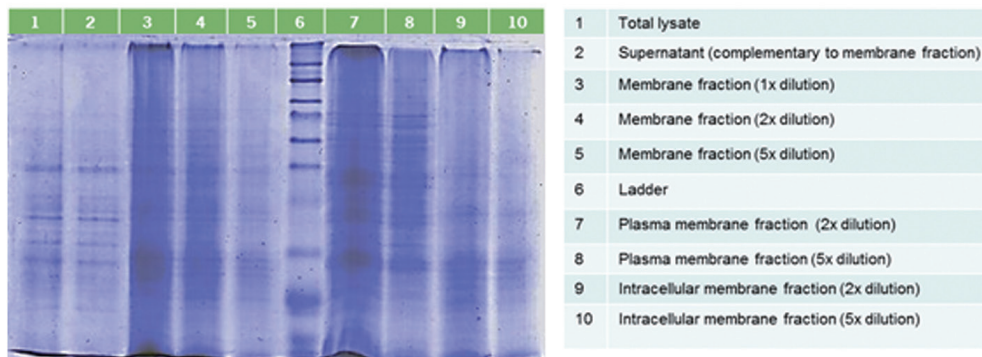


### *Optimization of protocol for PM enrichment from potato leaves*

The original project goal was to enrich plasma membrane proteins from potato leaf, and study the PM protein profile difference between *P. infestans* infected versus non-infected plants. For PM enrichment, the original method from *Arabidopsis* had to be modified and optimized to make it adaptable for potato leaf since ATPS PM enrichment is based on the size, density and polarity difference of the membrane system as described by Santoni *et al.* (Santoni *et al.*, 2000). Based on this, we tested different concentration combinations of PEG/Dextran, in combination with different application times for two phase partitioning steps. Besides, the concentration of KCl and pH are also playing an important role for PM enrichment (data not shown). However, due to the limited time, not much was done with the KCl concentration and pH adjustment. For PEG/Dextran, concentrations of 5.8%, 6.0%, 6.2%, 6.4% and 7.0% were tested in combination with different ATPS steps. However, none of these combinations, which had worked for *Arabidopsis*, resulted in PM enrichment (data not shown). Fig. S1 shows the SDS-PAGE of proteins from different subcellular compartments isolated from Bintje in 6.2% PEG/Dextran as an example. From the SDS-PAGE, there is no significant protein composition difference between PM and IM fractions indicating insufficient PM enrichment. This might be due to the different membrane characteristics between *Arabidopsis* and potato leaves. Possibly, the high concentration of ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) is influencing the PM enrichment in Bintje, since the relative abundance of RuBisCO is much higher than in *Arabidopsis*. And RuBisCO is quite often interacting with plasma membrane after sample homogenization, resulting in remaining RuBisCO protein in PM fractions even after enrichment and washing. In LC-MS analysis, RuBisCO peptides will suppress the ionization of the PM peptides, which makes the PM protein identification more difficult.



**Figure S1.** SDS-PAGE of protein isolation from potato Bintje fresh leaves in 6.2% as an example. TL: total lysate; MF: membrane fraction; PM: plasma membrane; IM: intracellular membrane; the protein loading amounts are different, therefore no quantitative conclusion can be drawn.



**Figure S2.** SDS PAGE of the PM proteins from MsK8 suspension cells. Lane 1: Total lysate; Lane 2: supernatant (complementary to total membrane fraction); Lane 3-5: Total membrane fraction (original, 2 times, 5 times dilution); Lane 7-8: plasma membrane fraction (2 times, 5 times dilution from original extract); Lane 9-10: intracellular membrane fraction (2 times, 5 times dilution from original extract), Highlighted in red: difference in the protein composition between plasma membrane and intracellular membrane fractions.

### ***PM quality and quantity optimization***

In order to obtain highly purified PM, additional two phase separation might be applied. However, the additional ATPS application upper phase has to be combined and washed to maximize the retrieval of the PM fraction. The more ATPS steps the more washing steps are required. Considering the time and ultra-centrifuge machine cost, we decide to limit to three times ATPS application. After ATPS, in general the Brij 58 washing step must be applied to remove the cytosolic soluble proteins which were sealed in the membrane vesicle during isolation. However, as described in chapter 4 for *Arabidopsis* samples, Brij 58 washing resulted in dramatic protein losses. Therefore, in order to maximize the coverage of protein identification in order to be able to reveal the subtle difference between mock, spo- and sup-samples, the Brij 58 washing step was omitted. Instead we made use of the advantage of the high resolution 2D nano UPLC for resolving complex protein mixtures and used bioinformatics software for further filtering and protein candidates selection.

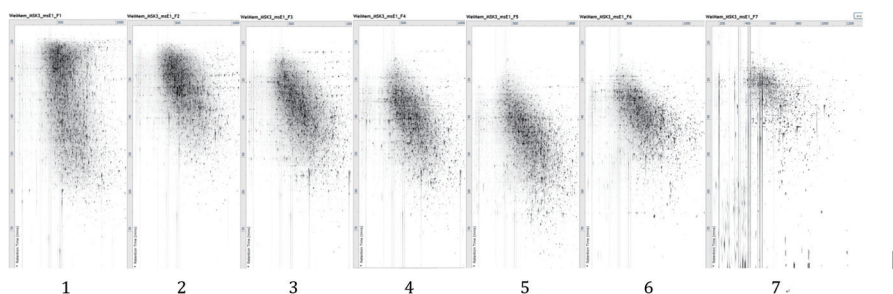
**Table S1.** PM protein enriched in 6.4% PEG/Dextran in comparison with 6.2% PEG/Dextran from Progenesis (with fold change  $\geq 2$ ,  $P$ -value  $< 0.01$ ). Proteins with asterisk were predicted as plasma membrane proteins with either transmembrane domain or GPI-anchor.

Sol accession	Max FC	Description
Solyc03g118200.2.1 *	8.8	Calcium-dependent phospholipid-binding copine-like protein
Solyc06g007120.2.1	6.4	Unknown protein
Solyc08g079090.2.1 *	2.9	Monocopper oxidase-like protein sku5
Solyc11g039650.1.1	2.6	Dynamamin-like protein
Solyc06g071050.2.1 *	2.6	Hypersensitive-induced response protein
Solyc08g041710.2.1	2.5	Unknown protein
Solyc01g112080.2.1 *	2.5	Domain-containing GPI-anchored protein
Solyc08g062910.2.1	2.5	Elongation factor ef-2
Solyc11g064890.1.1 *	2.4	Kinase family protein
Solyc02g071350.2.1	2.4	Multidrug pheromone mdr abc transporter family
Solyc07g063740.2.1	2.4	Uncharacterized protein loc100305941 precursor
Solyc07g053540.1.1 *	2.4	Fasciclin-like arabinogalactan protein 9
Solyc05g056580.2.1	2.4	Probable methyltransferase pmt26-like
Solyc06g063330.2.1	2.3	V-type proton Atpase catalytic subunit a
Solyc03g083970.2.1	2.3	Bcl-2-associated athanogene 7
Solyc04g077330.2.1	2.3	Unknown protein
Solyc03g113220.2.1 *	2.3	Hypersensitive-induced response protein
Solyc10g080940.1.1	2.3	Tubulin beta-4 chain-like
Solyc07g045440.1.1 *	2.2	Fasciclin-like arabinogalactan protein

Solyc02g067470.2.1 *	2.2	Hypersensitive-induced response protein
Solyc04g076810.2.1	2.2	Unknown protein
Solyc11g069430.1.1 *	2.2	Aquaporin pip2-7
Solyc07g065120.2.1 *	2.1	Glycerophosphoryl diester phosphodiesterase family protein
Solyc10g081980.1.1 *	2.1	Hin1-like protein
Solyc03g118760.2.1	2.1	Tubulin beta-1 chain-like
Solyc07g063480.2.1 *	2.1	Receptor-like protein kinase theseus 1-like
Solyc05g052510.2.1	2.1	Clathrin heavy chain 1-like
Solyc03g123740.2.1 *	2	Leucine-rich repeat transmembrane protein kinase
Solyc01g091530.2.1 *	2	Fasciclin-like arabinogalactan protein

### *Contamination and false positive discovery*

Contamination is always the case in proteomics. The source of contamination can be from buffers used for protein extraction, unwanted proteins of high abundance (such as RuBisCO) or human keratin during sample preparation. Critical proteomics working condition was considered to keep the contamination to a minimum. Furthermore, to enhance the signal of low abundant target PM protein, and to reduce the overall protein complexity, we employed online fractionation on 2D nano UPLC for peptide separation using seven fractions and the 2D representation of chromatograms from each fraction is shown in Fig. S3. Subsequently, in Progenesis, peaks were further clustered and selected based on their charge, which helped to remove contaminants such as secondary metabolites. However, from Fig. S3 it is clear that even with seven fractions pre-separation, the remaining peptide mixtures were still very complex, hampering protein identification in PLGS and quantification in Progenesis.

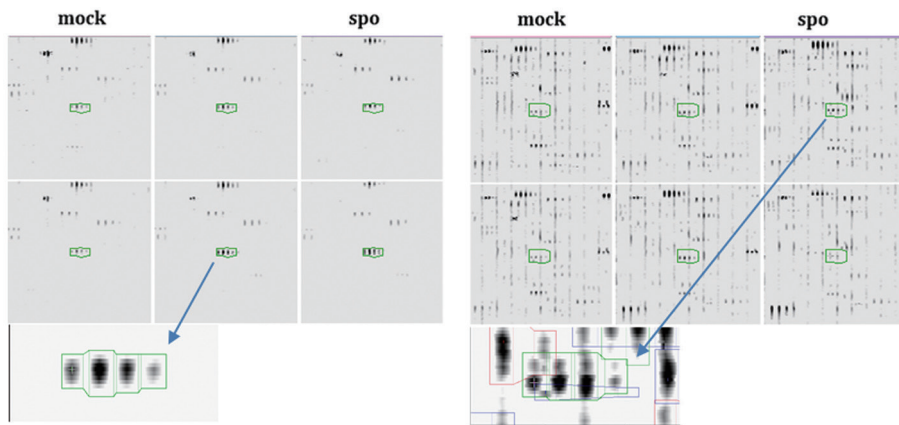


**Figure S3.** 2D representation of chromatograms of MS<sup>E</sup> from seven fractions (left to right) of mock sample replicate 3. Vertical axis represents the retention time and horizontal represents the m/z.

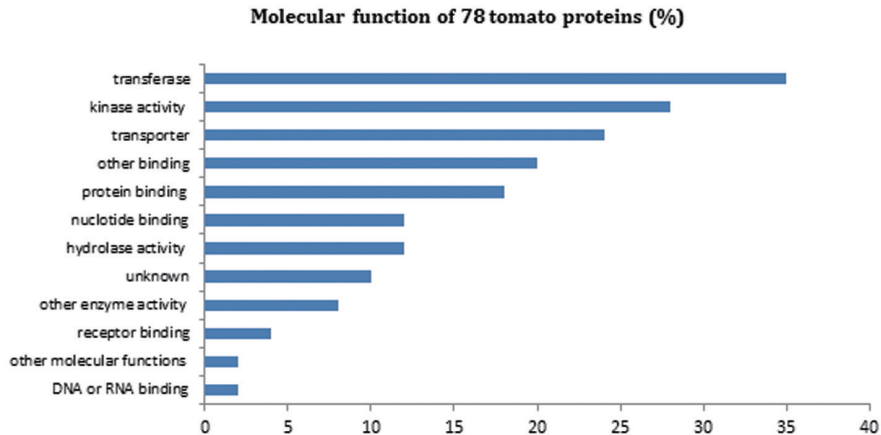
Eventually, the final output from PLGS protein identification was filtered by stringent criteria, such as peptide score, peptide length, number of ions per peptide, number of peptides per protein, trans-membrane domain presence, signal peptides (option), kinase domain (option), protein annotation, publication, etc. to reduce contaminants to a minimum.



In addition, as mentioned in the Results, Progenesis takes into account the noise (background) peak feature around the target caused by wrong alignment when there is a lot of peptides co-eluting at the same retention time, resulting in false positive protein quantitation. Fig. S4 shows a correctly aligned sample (left) and a wrongly aligned sample (right). It is quite obvious that there are many co-eluting peaks at the same retention time for the right sample, which makes it more difficult for Progenesis to distinguish the individual peaks for a correct peak detection. This miss-aligned feature (peptide) will cause the wrong protein abundance calculation, resulting in wrong protein quantification. To overcome this, we have to adjust the selected area manually (green rectangle) to allow for correct quantification of each potentially interesting peptide, to make sure the quantified peptide differential abundance profile is trustworthy.



**Figure S4.** The correct (left) and wrong (right) peptide feature quantification by Progenesis. The zoomed area shows the details of the specific feature detection for quantitation.



**Figure S5.** GO annotation of 78 tomato PM proteins.



# Chapter 6

## General discussion

Wei Song

## 1. Introduction

In this thesis, we set out to develop proteomics technologies to study *N*-glycosylation and proteins in the plasma membrane of plants. To reach these goals, we developed and optimized the protocol for enrichment of *N*-glycopeptides derived from *N*-glycoproteins and we developed and optimized a protocol for the enrichment of plasma membrane (PM) proteins from different plant species. In both approaches, subsequent peptide analysis was done by one- or two-dimensional LC, followed by Mass Spectroscopy. During this work we faced several challenges that took time to solve but have not always been discussed in detail in the previous chapters. In this discussion chapter I therefore first discuss some of these challenges in enrichment proteomics we faced as well as some potential alternatives or solutions. After their development, both enrichment proteomics protocols were used to study two interesting biological problems: the still not well-understood biological function of *N*-glycan modification of plant proteins and the role of PM proteins in plant pathogen interactions. Therefore I end this chapter with a brief discussion of the latest developments in these two fields in relation to (glyco)proteomics.

## 2. Challenges in enrichment proteomics and potential solutions

### 2.1 Filtering out (potential) contaminant proteins

Gel-based proteomics has limitations in the dynamic range of proteins that can be detected and it often suffers from insufficient separation of proteins (Ong and Pandey, 2001). For this reason, LC-MS based proteomics has been developed. Indeed, the direct peptide analysis from a total protein trypsin digest by high resolution LC-MS makes it possible to identify more proteins than usually obtained from gel-based approaches to proteomics. However, also for direct peptide analysis from total protein extract by LC-MS, the detection of low abundant peptides (and thus the identification of low abundant proteins) still remains a challenge. Especially from green plant tissue, the protein profile may be dominated by proteins like RuBisCO and other proteins from the photosynthesis systems which makes any low abundant protein easily escape detection. The solution to this problem is the use of specific protein enrichment procedures. However, even upon enrichment of low abundant proteins the samples are still never free of contamination by high abundant proteins. Therefore, if the aim of the experiment is to obtain information on a specific class of proteins (e.g. *N*-glycoproteins or PM proteins), filtering is required after analysis, using specific features of the proteins of interest. For our glycoproteomics (Chapter 3) we needed to remove contaminants by filtering for secreted proteins with an N-P!-S/T glycosylation site, while for PM proteomics (Chapter 4 and 5) we had to filter for proteins with a predicted transmembrane domain (TMD) or glycoposphatidylinositol (GPI) lipid anchor site. However, although such filtering decreases false discovery, it also comes with a drawback: even though there is no indication of *N*-glycosylation on other sites on proteins than N-P!-S/T, filtering for this

consensus sequence in identified proteins in glycoproteomics also precludes potential discovery of *N*-glycosylation sites deviating from this consensus amino acid (AA) site. A potential solution to the filtering problem for glycoproteomics is a new development in glycoproteomics making it possible to directly analyze the *N*-glycopeptide for both the AA sequence and *N*-glycan structure (Xu et al., 2016).

Also the filtering for PM proteins comes with a drawback: potential interesting proteins with a role in plant pathogen interactions may indirectly associate with the PM by protein-protein interactions. Such proteins do not have a TMD or lipid anchor and would thus be lost during our filtering procedure. For PM proteomics the solution could be the application of the Brij 58 wash step, which removes enclosed soluble proteins but which should retain non-membrane proteins with tight interaction with PM proteins. Below we discuss the problems encountered in PM enrichment for plant materials and why the Brij 58 wash step in this procedure was omitted.

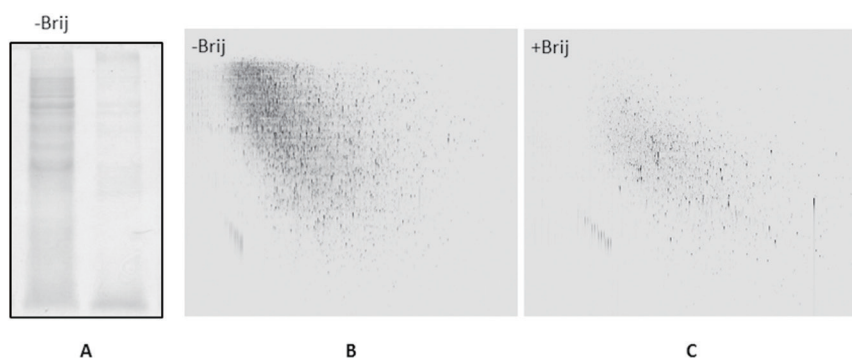
## 2.2 The use of Brij 58 for plasma membrane enrichment

PM proteins are usually underrepresented in total protein extract proteomics both due to their low abundance in the overall protein pool, and because of their embedding in a hydrophobic lipid environment, which makes them more difficult to solubilize for efficient trypsin digestion. For both our work on *Arabidopsis* (Chapter 4), potato and tomato (Chapter 5) and on *Nicotiana benthamiana* (Bouwmeester et al., 2014a), we enriched PM proteins using an aqueous two phase partitioning system (ATPS) (Alexandersson et al., 2004). This method is based on the fact that membranes form vesicles during extraction and that membrane vesicles of different subcellular compartments can be separated based on differences in their density. The purity of the fractions can be assessed by measuring marker proteins from the different subcellular compartments (PM, mitochondria, chloroplast, cytosol), as we did by peptide LC-MS, or by determining the ATPase activity which is characteristic for different types of membranes.

For each experiment with different plant species or tissues, the conditions for vesicle density separation had to be optimized by testing different combinations of PEG and Dextran. This could either indicate that the PM protein/lipid composition is different for the different plant materials tested. However, it could also be due to differences in the cell wall and cuticle in the different samples, which causes differences in vesicle formation, abundance and densities. Eventually I was able to get a satisfactory PM enrichment for *Arabidopsis* leaves and for tomato suspension cells. For potato leaves, however, we did not manage to obtain a satisfactory enrichment of the PM vesicle fraction.

During the membrane vesicle formation upon extraction, soluble proteins are enclosed into the vesicles. Brij 58 is a light detergent, which transiently opens the vesicles during which contaminating soluble proteins enclosed in the vesicles may leak out (Johansson

et al., 1995). However, from my observation, each Brij 58 wash step dramatically decreased PM protein yield as determined by protein analysis on SDS-PAGE (Fig. 1A) and as determined by LC-MS (Chapter 4 and 5). I spent a lot of time trying to optimize the Brij 58 wash step, however the results were never satisfactory, in contrast to other PM proteomics studies. However, although the Brij 58 washing step has been applied in multiple PM protein studies, in each of these cases still a large number of cytosolic soluble proteins were identified in the PM enriched fractions, indicating that also with the Brij 58 wash step potential contaminants are not always efficiently removed (Alexandersson et al., 2004; Cao et al., 2006; Cheng et al., 2009; Komatsu et al., 2009; Mitra et al., 2009; Nouri and Komatsu, 2010; Lund and Fuglsang, 2012). Moreover, we also note that in none of the papers on PM proteomics a quality check by peptide LC-MS of the PM fraction before and after a Brij 58 wash is described. In conclusion, the use of Brij 58 washing depends on the biological question to be addressed. Because in some of our samples I expected only very subtle changes in PM protein content (e.g. *Arabidopsis* Col-0 versus *lecrk-1.9* mutant and *35S-ipiO1* plants without infection; Chapter 4) I did not want to end up with the most abundant PM proteins and therefore decided to omit the Brij 58 washing step and use rigorous filtering for TMD or lipid anchor site to select relevant proteins.



**Figure 1.** A: SDS-PAGE of plasma membrane proteins from *Arabidopsis* Col-0 without (-) or with (+) Brij 58 washing. B-C: 2D representation of chromatogram of Col-0 plasma membrane protein from MS<sup>E</sup> dataset without (-) or with (+) Brij 58 washing.

## 6

### 2.3 Label-free proteomics versus labeled proteomics

In all our analyses we used label-free ‘shotgun’ liquid chromatography coupled to mass spectrometry (LC-MS) for the analysis of peptide mixtures derived from specific enriched protein fractions. In this LC-MS approach, the whole protein fraction is digested by trypsin into peptides, which are subsequently separated by strong cation exchange (SCX) chromatography, followed by high resolution MS. Overall, this procedure tends to be highly technically reproducible and have a higher dynamic range than gel-based approaches (Domon and Aebersold, 2006). However, biological plant

samples (or the isolation of proteins from plant materials) show intrinsic variation in protein content and for detection of reliable differences between samples from different treatments multiple biological replicates are needed. When samples are analyzed in multiple LC-MS runs (from replicate samples of control and treated samples) they need to be aligned for quantitative comparison of the same peptide features. In the label free proteomics analysis that I used this means that between samples and experiments independent LC-MS runs had to be aligned in order to compare sets of control and treated samples. Potentially, alignment problems can be solved by co-analysis of two different samples with different protein/peptide labeling. Indeed, for MS based protein quantification, many labeling techniques can be employed. Some of these involve labeling in situ in living cells, before protein sample preparation (Ong et al., 2002; Hung et al., 2016). Alternatively, proteins or peptides may be differentially labeled after protein isolation from different samples. These labels include isotope-coded affinity tags (ICAT) (Gygi et al., 1999), tandem mass tags (TMT) (Thompson et al., 2003), chemically synthesized peptide standards (Gerber et al., 2003), stable isotope labeled peptides (Kuhn et al., 2004), and more recently, isobaric tags for relative and absolute quantification (iTRAQ) (Ross et al., 2004; Ow et al., 2008).

With differential mass-labeling of proteins/peptides from two different samples the samples can be mixed after labeling and be analyzed in a single LC-MS run. In such a single run, identical peptides from the same protein in each of the two samples can be distinguished from each other based on their (tagged) mass difference. The relative mass intensity signal for each of these labeled peptide gives a direct value of the relative abundance of the corresponding protein in the two samples. Therefore, when the same peptide from different samples can be distinguished from each other in the same LC-MS run, the problem of direct comparison is solved. However, this mostly works for two-by-two comparison of samples and not for three or more samples. Moreover, most of those labeling techniques also have limitations, such as complex sample preparation, low reproducibility, incomplete labeling of low abundant peptides and sometimes create additional problems in peptide identification (Patel et al., 2009). In contrast, label-free proteomics allows for easier protein identification and high-throughput sample analysis. However, label-free proteomics does require more sample replicates, to reliably detect quantitative differences between samples (Zhang et al., 2009). In my study, both *N*-glycoproteomics and PM proteomics required multiple steps for sample preparation, which may introduce additional variation between samples, especially when not all samples can be prepared on a single day. Indeed, without filtering, my samples from experiments in Chapter 3 and 4 tended to cluster by preparation day rather than by treatment. Such variation in the samples eventually may also create problems for the software used for feature detection, quantitation and comparison between independent LC-MS runs. Therefore I used manual peptide feature alignment complementary to the auto-mode alignment algorithm in Progenesis and used multiple filters for a step-wise quantitative feature selection. At least for the case of eight PM proteins from the tomato MsK8 suspension cell culture experiment with *P. infestans* I could confirm that the automated selection, identification and quantification done in Progenesis was correct

(Chapter 5). However, most problematic turned out to be linking quantified peptide features to protein identification and many quantified features remained unidentified (see below). Recently, MaxLFQ was developed, a new intensity determination and normalization procedure for peptide features and which is freely available (Cox et al., 2014). Such new software development should in the future allow for more accurate and robust proteome-wide quantification using label-free proteomics.

## 2.4 Peptide quantification linked to protein identification

In our study, protein quantification from peptide features was mainly done in Progenesis. Each peptide is represented by multiple features in each LC-MS run (multiple isotope masses, different adducts or charges), and for quantification these multiple signals have to be reduced to a single, representative quantitative signal. For this, Progenesis first aligns all multiple features from a single peptide from each individual run and subsequently calculates for each individual peptide a single ion abundance as representative feature for each peptide in the multiple samples. Subsequently, all the quantified peptide features are linked to protein identification, based on their fragmentation spectra. The protein abundance in a sample is calculated by summation of all unique peptide signals linked to a protein.

However, many quantified features from Progenesis cannot be linked to their protein identification. This may be mainly caused by misalignment of these features in Progenesis, which makes it difficult to link them to feature identification. Such misalignment especially happens when samples are too complex, causing co-elution of multiple peptides, which makes it difficult for the software to distinguish each peptide. Especially when protein differences are very subtle between samples as for the experiment in Chapter 4, this caused problems to link the few quantified differential features between samples to protein identification. In Chapter 4, after applying all the critical filtering steps, I ended up with 99 quantitative features showing differential peptide abundance between resistant *Arabidopsis* Col-0 and two susceptible phenocopy lines. None of these features could be linked to protein identification, but we still managed to assign some protein identification using the web-based tool MS-Fit, which only uses the accurate mass of a peptide to link it to a protein. To make this more reliable I selected for the criterion that at least three peptides should match to the same protein, meaning that proteins for which only two or one peptide feature was quantified, the protein identity could not be determined.

In Chapter 5 I used the new software SEDMAT which improves the alignment by Progenesis with in-house built algorithms. SEDMAT re-aligns all the features across different runs, also with the one already correctly matched by Progenesis. Subsequently, it tries to match the quantified peptide features to the protein identification using different scoring and ranking algorithms. Eventually, in the integrated output of SEDMAT and Progenesis, every quantified feature is annotated either as non-identified, identified by SEDMAT, identified by Progenesis, or identified by both. With the help



of the SEDMAT software, the number of identified and quantified proteins increased approximately five-fold (for the tomato cell culture PM proteomics experiment: 16740 by SEDMAT, 2999 by Progenesis, 2970 overlap from both). It indicates that in-house development of specialized custom software can be a great complement to the commercial software, taking into account the customer's specific research demand.

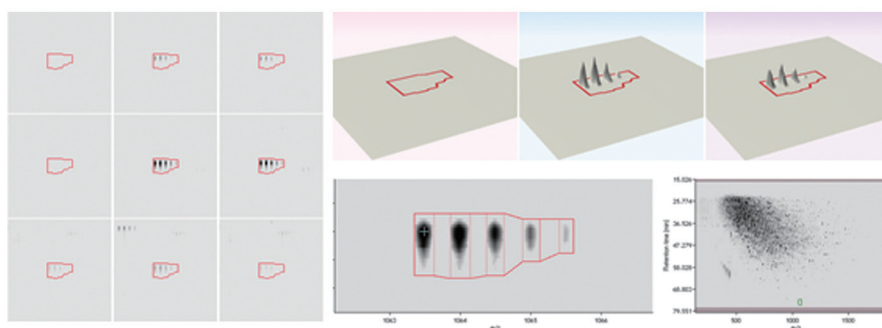
## 2.5 New alternative strategies for *N*-glycoproteomics

In Chapter 3 we describe the enrichment of glycopeptides using coupling to hydrazide beads and release by PNGase F. Using this procedure bypasses a common problem in glycoproteomics caused by the heterogeneity of *N*-glycans on a glycoprotein. Due to maturation of the *N*-glycan, which depends on subcellular localization of the glycoprotein (ER or Golgi), a single glycan attachment site on a protein may contain different *N*-glycan forms. After tryptic digestion of a glycoprotein this results in multiple mass signals from a single glycopeptide, which makes identification of the glycopeptide very difficult. Quantification of these different masses of a single glycopeptide provides information on how much of the same protein is still in the ER, and how much has passed through the Golgi. In our glycoproteomics approach (Chapter 3) potential subcellular localization information is deleted, as the *N*-glycan is removed by the PNGase F treatment. However, novel analytical techniques by which intact glycopeptides (containing the *N*-glycan) can be analyzed both for AA sequence information and for *N*-glycan composition would allow for such additional subcellular localization information to be extracted. Moreover, such novel analytical techniques may also allow for better quantification of *N*-glycan maturation and or turnover, which could help further elucidate *N*-glycan function for glycoproteins (Xu et al., 2016). Such novel analytical techniques would benefit from whole glycoprotein enrichment, rather than glycopeptide enrichment. The most common used approach for isolation of intact glycoproteins is based on the carbohydrate binding capacity of lectins. However, lectins have a bias towards mannose type *N*-glycans, resulting in underrepresentation of glycoproteins with complex *N*-glycans, even when multiple types of lectins are used (Zielinska et al., 2010; Zielinska et al., 2012; Xu et al., 2016).

## 2.6 Linking single glyco-peptides to protein identification

Label-free proteomics allows for high-throughput protein identification. However, the software algorithm that links peptide features to protein identification relies on multiple peptide hits per protein for reliable protein identification. In glycoproteomics this introduces a problem for the identification of many glycoproteins, because many glycoproteins only have as single glycosylated site, resulting in a single glycopeptide from such a protein. After release from hydrazide beads of the captured glycopeptide by PNGase F this renders only a single (deamidated) peptide to be analyzed by LC-MS (Chapter 4). In the conventional peptide-to-protein identification algorithm, such single peptide feature may be linked to the correct protein, but will still be filtered out by the software because the criterion of multiple peptides matching to the same protein

is not matched. It was only after I noticed that all identified glycoproteins have a bias towards having three or more *N*-glycosylation sites that I became aware of this build in selection criterion. It shows that non-standard application of commercial software with hidden algorithms may lead to unexpected mistakes. Thus, for glycoproteins with a single *N*-glycosylation site we adjusted the parameter settings of the software for single peptides, but this also increased the chance of false protein matches. This was solved by filtering for proteins/peptides with a consensus *N*-glycan attachment site, deamidation signature and presence of a secretion signal. Furthermore, for the final selected protein candidates, we will manually check their spectra quality, alignment, peptide abundance pattern, peptide score, and peptide ranking information, to ensure their reliability for further biological validation (Fig. 2).



**Figure 2.** Example of manual verification of feature alignment and quantitation in Progenesis.

## 2.7 Additional quantified proteins from alternative or in-house developed software

In our study, we made use of high resolution 2D nano UPLC followed by SYNAPT™ Q-TOF mass spectrometry for protein separation and detection. We used ProteinLynx™ Global Server (PLGS) for protein identification and Progenesis LC-MS™ for protein quantification. However, to our knowledge, only a small portion of overall peptide features can be finally quantitatively identified. As shown in Chapter 4, we could not link quantified features showing differential ion abundance to their protein identifications in Progenesis. Those differential features are, however, the most interesting and valuable output from the proteomics. Therefore, we made use of an alternative, web-based, tool, MS-Fit, to use those features to identify the corresponding protein. The algorithm of MS-Fit will introduce more false positives because it only uses the MS data without taking into account peptide fragmentation data from MS/MS. To minimize false discovery, I applied stringent filtering for the output of MS-Fit. An additional problem is that MS-Fit relies on multiple hits per protein. Therefore, for my *N*-glycoprotein data (Chapter 3), I could not use it to improve the overall quantitatively identified *N*-glycoproteins, due to the fact that many glycoproteins

are singly glycosylated. Later on, in Chapter 5, a similar bottleneck for the tomato MsK8 suspension cells was encountered and the overall quantitatively identified features were limited when using Progenesis. Fortunately, the in-house developed tool SEDMAT was finalized within the Galaxy toolshed. With the help of SEDMAT in total I quantitatively identified 16769 features from MsK8 suspension cells, of which only 2999 were identified by Progenesis, and 16740 by SEDMAT (with 2970 overlapping with Progenesis). Since SEDMAT was developed and finalized in the last year of my PhD, with the limited time frame I could not re-process my earlier data to get more quantitatively identified glycoproteins from Chapter 3 and plasma membrane proteins from *Arabidopsis* in Chapter 4, but it would be valuable if this is still done by others. Overall, open access tools and custom in-house developed tools can be a great complement of commercially available tools helping using quantified features to identify the corresponding protein, especially for researchers with specific proteomics demands such as studying specific post-translational modifications (PTMs) or proteins in response to plant-pathogen interactions.

### 3. Applications of enrichment-proteomics

During my thesis work most time was spend on developing the methods for protein enrichment and especially on data analysis using different software packages. Unfortunately, this left limited time to work on biological applications of the developed technologies. However, some biologically interesting results were obtained in setting up the technologies. Below I briefly discuss the biological relevant results in this thesis work, what can be done to get more out of the obtained data and how results were used by others to enhance their biological research.

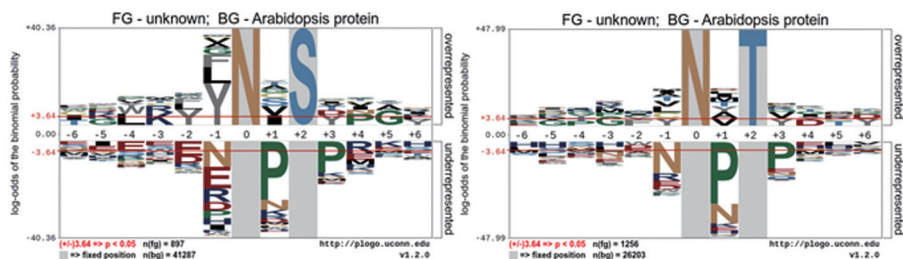
#### 3.1 Using *N*-glycoproteomics data to improve *N*-glycan occupancy prediction tools

Glycosylation is one of the most abundant protein post-translational modifications in eukaryotes. However, the only role of *N*-glycosylation which is well understood is their role in protein folding. Proteins that enter the secretory pathway are imported into the ER during translation of the mRNA while the AA string is mostly in an unfolded state. Upon entering the lumen of the ER such secreted proteins need to fold correctly during which *N*-glycosylation plays an important role. Trimming of the *N*-glycan in the ER can be a sign of prolonged misfolding and may trigger a misfolded protein response to remove the misfolded protein. As such *N*-glycans function in protein folding quality control. In mammalian cells *N*-glycosylation of proteins has also been shown to be important for protein localization and stability, ligand binding, endocytosis and for immune responses (Varki, 1993; Parodi, 2000; Schollen et al., 2005; Sun et al., 2005). The important role of *N*-glycosylation in mammals is also clear from the strong phenotype of many *N*-glycosylation mutants in mammals. In contrast, many

*N*-glycosylation mutants in plants have no or only very weak phenotypes, making it difficult to study the role of *N*-glycosylation in plants. Recent study shows that glycoprotein levels in soybean may change upon abiotic stress like flooding (Mustafa and Komatsu, 2014). However these studies did not reveal whether up and down regulation of glycoprotein levels followed a transcriptional regulation or whether the *N*-glycans themselves are involved in changing protein abundance under stress conditions.

Understanding the role of *N*-glycosylation starts with an inventory of the proteins which are modified by *N*-glycans, to which our research has made a large contribution (Chapter 3). With the large data set of *N*-glycan occupancy on glycoproteins we can re-evaluate the consensus AA sequence which predicts *N*-glycan attachment on glycoproteins. The consensus sequence of *N*-glycosylation in plants is known as N-P!-S/T (Varki et al., 2009). However, amino acids before N and after S/T might also follow some typical pattern specific for *N*-glycosylation. We loaded all the experimental confirmed *N*-glycosylated peptides sequences obtained in the glycoproteomics studies described in Chapter 3 to build a sequence logo (<http://plogo.uconn.edu/>) which presents the frequency of six amino acids around the N-P!-S/T consensus site (Fig. 3). In this representation, the overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino acid at that position. The upper part of the logo shows the most likely AAs, while the lower part shows the most un-likely AAs. The results indicate that tyrosine (Y) and isoleucine (I) are most often present before and after the asparagine to which the *N*-glycan is attached, especially for the consensus sequence N-P!-S/T. Attachment of the hydrophilic Glycan between these hydrophobic AAs would force a directional folding in the AA chain (Lu et al., 2012). There are web server tools that predict *N*-glycosylation site occupancy on glycoproteins using different algorithms or artificial neural networks that examine the AA sequence context of N-P!-S/T glycosylation consensus sites. However, most of these prediction tools are based on glycoproteomics information from non-plant glycoproteins (Gupta, 2002; Chuang et al., 2012). Therefore, our results may be incorporated into such prediction tools in order to improve the *N*-glycosylation site occupancy prediction for plant proteins.

## 6



**Figure 3.** Sequence logo is built based on all the identified glycopeptides. The upper part of each logo showing the most likely AA, and lower part showing the most unlikely AA.

### 3.2 Using *N*-glycans for cell specific protein tagging

The glycoproteomics technology I developed in my thesis was also used by me and others in combination with cell specific *N*-glycan tagging. In cell specific *N*-glycotagging we make use of the *cgl1-1* glycosylation mutant, which lacks proteins with complex *N*-glycans and from which therefore all glycopeptides can be released by PNGase F after capturing of glycopeptides on hydrazide beads. By cell specific complementation of the *cgl1-1* mutant through tissue specific expression of the wild type *CGL1* gene, complex *N*-glycosylation will be restored in the complemented tissue, and these proteins will be missing in our glycoproteomics analysis. Such cell specific complementation was done by expressing *CGL1* either under the epidermal specific promoter of the *Arabidopsis LTP1* gene and the mesophyll specific promoter of the *Arabidopsis RBCS* gene. The results from these cell specific glycoproteomics indicate indeed a difference in protein abundance for the two complemented *cgl* lines compared to Col-0 (Table 1). In the *pLTP1::GNT1 cgl1-1* line 53 proteins showed a higher abundance compared to Col-0, suggesting that these could be proteins that are mainly active in non-epidermal cells. Similarly, in the *pRBCS::GNT1 cgl1-1* line 29 proteins showed a higher abundance, indicating that these may be mainly present in the epidermal cells. Unexpectedly, we also found proteins that show a lower abundance in the *pLTP1::GNT1 cgl1-1* lines (11 proteins) and *pRBCS::GNT1 cgl1-1* lines (6 proteins) compared to that in Col-0. This may indicate that glycoprotein levels in one tissue influence the glycoprotein levels in another tissue. However, the results from this cell specific *N*-glycan tagging experiment need further validation.

**Table 1.** Glycoprotein abundance in *pLTP1::GNT1* and *pRBCS::GNT1* complemented *cgl1-1* mutant compared to Col-0. up: protein abundance higher vs Col-0; down: protein abundance lower vs Col-0.

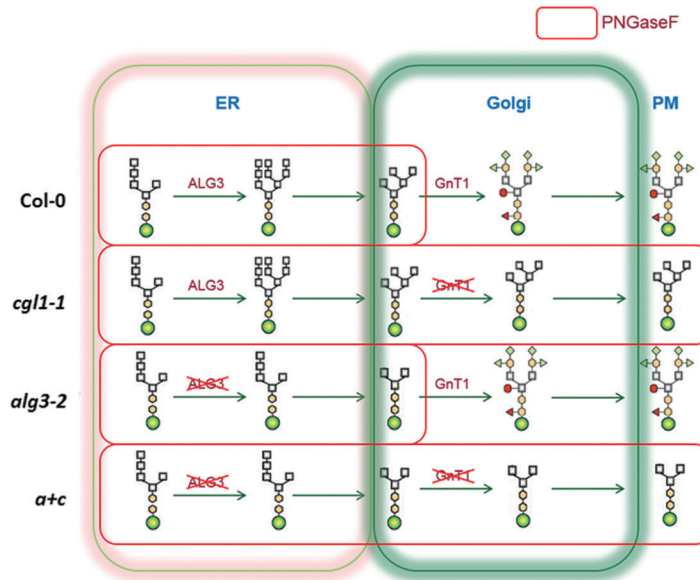
LTP1-GNT1 <i>cgl1-1</i> vs Col-0	LTP1-GNT1 <i>cgl1-1</i> vs Col-0	RBCS-GNT1 <i>cgl1-1</i> vs Col-0	RBCS-GNT1 <i>cgl1-1</i> vs Col-0
up	down	up	down
53	11	29	6

Cell specific complementation in *cgl1-1* mutant background was also done using a pathogen inducible promoter of *Arabidopsis DOWNY MILDEW RESISTANT 6* (DMR6) fused to the coding sequence of *GNT1*. In such transgenic *cgl* line, complex *N*-glycosylation will only be restored in the downy mildew infected cells containing downy mildew *Hyaloperonospora arabidopsidis* (*Hpa*) haustoria, but not in neighboring cells. In this case the proteins with complex *N*-glycans were isolated from infected cells by immunoprecipitation (IP), using antibodies specific against plant-type complex *N*-glycans. Subsequently, label-free comparative proteomics was employed for protein identification and quantification. In this study, the property of anti-horseradish peroxidase (HRP) antibodies binding with plant complex glycan was utilized to enrich complex glycoproteins by immunoprecipitation. Complex glycoproteins released from the antibodies were digested with trypsin and peptides were analyzed by LC-MS. In

this analysis the peptides containing the complex *N*-glycan are actually not detected. The identified proteins were filtered for putative glycoproteins (secreted proteins with N-P!-S/T consensus site). This resulted in the identification of 18 candidate disease-related complex *N*-glycosylated proteins, several of which seem to have a potential role in susceptibility of *Arabidopsis* to *Hpa*. Three of the 18 identified glycoproteins were also detected in the glycoproteomics experiment described in Chapter 3 (Song et al., 2013).

### 3.3 Using *N*-glycoproteomics to study the *Arabidopsis alg3-2* glycosylation mutant

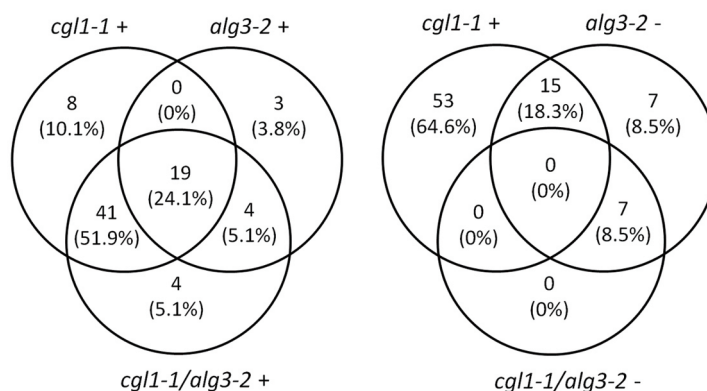
In my thesis I have described the application of *N*-glycoproteomics to make a full inventory of the *N*-glycoproteome of *Arabidopsis* using the *cgl1-1* mutant, which lacks complex *N*-glycans. We also performed *N*-glycoproteomics on another glycosylation mutant, *alg3-2*, with a mutation in the gene *ASPARAGINE-LINKED GLYCOSYLATION 3* (*ALG3*). *ALG3* encodes a mannosyltransferase which forms high mannose type *N*-glycans in the ER (see Fig. 4). In addition, we made double mutants of *cgl1-1* and *alg3-2* and *N*-glycoproteomics was used to compare the non-complex *N*-glycoproteome of single and double glycosylation mutants with that of Col-0. Comparison of the *N*-glycoproteome of Col-0 and *alg3-2* can give indications whether *N*-glycan structure is involved in secreted protein stability in the ER. Similarly, comparison of the *N*-glycoproteome of *cgl1-1* with the *cgl1-1/alg3-2* double mutant can give indications whether the *N*-glycan structure on proteins in the Golgi and beyond have an effect on protein stability. Preliminary results of these studies indicate that 26 proteins are more abundant, while 29 proteins are less abundant in *alg3-2* compared to Col-0 (Table 2 and Fig. 5). Interestingly, 19 of the 26 proteins that are more abundant in *alg3-2* are also more abundant in *cgl1-1*, suggesting that these are proteins, which normally have complex *N*-glycans. This could indicate that for these proteins the aberrant *N*-glycan structure as result of the *alg3-2* mutation, makes them less efficient substrates for complex *N*-glycan modification. Of the 29 proteins with lower abundance in *alg3-2*, 14 proteins have higher abundance in *cgl1-1*, indicating that these are proteins with complex *N*-glycan structure, but for which protein stability may be reduced in the *alg3-2* background. Combined, these preliminary data give a first evidence for the importance of *N*-glycan structure for the stability of plant glycoproteins. Moreover, these data provide a clear molecular phenotype for these glycosylation mutants, which lack a clear growth phenotype.



**Figure 4.** Comparing glycoproteomics in *N*-glycosylation mutants with glycoproteomics in *Arabidopsis* Col-0. *cgl1-1*: complex glycan less 1-1, lack of enzyme activity of GnTI in the Golgi apparatus; *alg3-2*: asparagine-linked glycosylation 3-2, lack of ALG3 enzyme activity that forms high mannose type glycans in the ER; *a+c*: double mutant of *cgl1-1* and *alg3-2*. Red area indicates the location where *N*-glycans can be cleaved by PNGase F and from which *N*-glycoproteins are analyzed using our enrichment procedure using release from hydrazide beads.

**Table 2.** Glycoprotein abundance in *cgl1-1*, *alg3-2* and *cgl1-1/alg3-2* compared to in Col-0. Up: protein abundance higher vs Col-0; down: protein abundance lower vs Col-0.

<i>cgl1-1</i> vs Col-0	<i>cgl1-1</i> vs Col-0	<i>alg3-2</i> vs Col-0	<i>alg3-2</i> vs Col-0	<i>cgl1-1/alg3-2</i> vs Col-0	<i>cgl1-1/alg3-2</i> vs Col-0
up (+)	down (-)	up (+)	down (-)	up (+)	down (-)
68	1	26	29	68	7



**Figure 5.** Inventory of glycoprotein differences between *Arabidopsis Col-0*, *cgl1-1*, *alg3-2* and *cgl1-1/alg3-2* double mutant plants. (+): protein abundance higher vs Col-0; (-): protein abundance lower vs Col-0.

#### 4. Plasma membrane enrichment to confirm PM localization of LecRK receptor protein

The procedure of PM enrichment, which was developed for different plant tissue types, was also used to confirm PM localization of the target protein *Arabidopsis LecRK-I.9* in *N. benthamiana*. In a study carried out by Bouwmeester *et al.* (Bouwmeester *et al.*, 2014b) the function of the *Arabidopsis LecRK-I.9* gene was investigated in a cross species experiment by transiently expressing the *Arabidopsis LecRK-I.9* in *N. benthamiana* leaves, which were subsequently challenged with *P. infestans*. In these experiments, our PM enrichment procedure helped confirm the PM localization of the *Arabidopsis LecRK-I.9* protein in *N. benthamiana* leaves, as the LecRK-I.9 protein was enriched and mainly detected on immunoblots in the PM enriched fraction and not in other membrane fractions. PM proteomics also provided experimental proof of protein localization for other researchers studying PM proteins in response to different biotic stresses (Raffaele *et al.*, 2009; Keinath *et al.*, 2010; Qi *et al.*, 2011; Elmore *et al.*, 2012) or different abiotic stresses such as salt (Nohzadeh Malakshah *et al.*, 2007), flooding stress (Komatsu *et al.*, 2009) and osmotic stress (Nouri and Komatsu, 2010) or during cold acclimation (Li *et al.*, 2012; Takahashi *et al.*, 2013). In all these studies PM enrichment was also based on the ATPS that we used for PM proteomics.

## 6

Finally, in my thesis I have shown that enrichment technology provides unique opportunities to focus on low abundant proteins from specific subcellular localization or with post-translational modification and that analysis by label-free comparative proteomics provides a powerful tool for high-throughput protein identification and quantification. However, there is still room for improvement, both for sample processing (e.g. sample purification) as well as for data processing (e.g. peptide feature alignment, quantitative differential feature identification). Moreover, proteomics results only get meaning after validation in follow-up experiments, which is very time-consuming especially for crop species.



## References

**Alexandersson E, Saalbach G, Larsson C, Kjellbom P** (2004) *Arabidopsis* plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. *Plant Cell Physiol* **45**: 1543–56

**Bouwmeester K, Han M, Blanco-Portales R, Song W, Weide R, Guo L-Y, van der Vossen EAG, Govers F** (2014a) The *Arabidopsis* lectin receptor kinase LecRK-I.9 enhances resistance to *Phytophthora infestans* in Solanaceous plants. *Plant Biotechnol J* **12**: 10–16

**Bouwmeester K, Han M, Blanco-Portales R, Song W, Weide R, Guo L-Y, van der Vossen EAG, Govers F** (2014b) The *Arabidopsis* lectin receptor kinase LecRK-I.9 enhances resistance to *Phytophthora infestans* in Solanaceous plants. *Plant Biotechnol J* **12**: 10–16

**Cao R, Li X, Liu Z, Peng X, Hu W, Wang X, Chen P, Xie J, Liang S** (2006) Integration of a two-phase partition method into proteomics research on rat liver plasma membrane proteins. *J Proteome Res* **5**: 634–642

**Cheng Y, Qi Y, Zhu Q, Chen X, Wang N, Zhao X, Chen H, Cui X, Xu L, Zhang W** (2009) New changes in the plasma-membrane-associated proteome of rice roots under salt stress. *Proteomics* **9**: 3100–14

**Chuang G-Y, Boyington JC, Joyce MG, Zhu J, Nabel GJ, Kwong PD, Georgiev I** (2012) Computational prediction of *N*-linked glycosylation incorporating structural properties and patterns. *Bioinformatics* **28**: 2249–55

**Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M** (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**: 2513–2526

**Elmore JM, Liu J, Smith B, Phinney B, Coaker G** (2012) Quantitative proteomics reveals dynamic changes in the plasma membrane during *Arabidopsis* immune signaling. *Mol Cell Proteomics*. doi: 10.1074/mcp.M111.014555

**Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP** (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **100**: 6940–5

**Gupta R** (2002) Prediction of glycosylation across the human proteome and the correlation to protein function 1 Introduction 2 Methods. *Proteins* **322**: 310–322

**Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R** (1999) Quantitative

analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**: 994–999

**Hung V, Udeshi ND, Lam SS, Loh KH, Cox KJ, Pedram K, Carr SA, Ting AY** (2016) Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. *Nat Protoc* **11**: 456–75

**Johansson F, Olbe M, Sommarin M, Larsson C** (1995) Brij 58, a polyoxyethylene acyl ether, creates membrane vesicles of uniform sidedness. A new tool to obtain inside-out (cytoplasmic side-out) plasma membrane vesicles. *Plant J* **7**: 165–73

**Keinath NF, Kierszniowska S, Lorek J, Bourdais G, Kessler SA, Shimosato-Asano H, Grossniklaus U, Schulze WX, Robatzek S, Panstruga R** (2010) PAMP (Pathogen-associated Molecular Pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity. *J Biol Chem* **285**: 39140–39149

**Komatsu S, Wada T, Abalea Y, Nouri M-ZZ, Nanjo Y, Nakayama N, Shimamura S, Yamamoto R, Nakamura T, Furukawa K, et al** (2009) Analysis of plasma membrane proteome in soybean and application to flooding stress response. *J Proteome Res* **8**: 4487–4499

**Kuhn E, Wu J, Karl J, Liao H, Zolg W, Guild B** (2004) Quantification of C-reactive protein in the serum of patients with rheumatoid arthritis using multiple reaction monitoring mass spectrometry and  $^{13}\text{C}$ -labeled peptide standards. *Proteomics* **4**: 1175–86

**Li B, Takahashi D, Kawamura Y, Uemura M** (2012) Comparison of plasma membrane proteomic changes of *Arabidopsis* suspension-cultured cells (T87 Line) after cold and ABA treatment in association with freezing tolerance development. *Plant Cell Physiol* **53**: 543–554

**Lu D, Yang C, Liu Z** (2012) How hydrophobicity and the glycosylation site of glycans affect protein folding and stability: a molecular dynamics simulation. *J Phys Chem B* **116**: 390–400

**Lund A, Fuglsang AT** (2012) Purification of plant plasma membranes by two-phase partitioning and measurement of  $\text{H}^+$  pumping. *Methods Mol Biol* **913**: 217–23

**Mitra SK, Walters BT, Clouse SD, Goshe MB** (2009) An efficient organic solvent based extraction method for the proteomic analysis of *Arabidopsis* plasma membranes. *J Proteome Res* **8**: 2752–2767

**Mustafa G, Komatsu S** (2014) Quantitative proteomics reveals the effect of protein

glycosylation in soybean root under flooding stress. *Front Plant Sci* **5**: 627

**Nohzadeh Malakshah S, Habibi Rezaei M, Heidari M, Salekdeh GH, Malakshah SN, Rezaei MH, Heidari M, Salekdeh GH** (2007) Proteomics reveals new salt responsive proteins associated with rice plasma membrane. *Biosci Biotechnol Biochem* **71**: 2144–2154

**Nouri MZ, Komatsu S** (2010) Comparative analysis of soybean plasma membrane proteins under osmotic stress using gel-based and LC MS/MS-based proteomics approaches. *Proteomics* **10**: 1930–1945

**Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M** (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**: 376–386

**Ow SY, Cardona T, Taton A, Magnuson A, Lindblad P, Stensjö K, Wright PC** (2008) Quantitative shotgun proteomics of enriched heterocysts from *Nostoc* sp. PCC 7120 using 8-plex isobaric peptide tags. *J Proteome Res* **7**: 1615–28

**Parodi AJ** (2000) Role of *N*-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem J* **348 Pt 1**: 1–13

**Patel VJ, Thalassinos K, Slade SE, Connolly JB, Crombie A, Murrell JC, Scrivens JH** (2009) A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* **8**: 3752–3759

**Qi YP, Tsuda K, Nguyen L V, Wang X, Lin JS, Murphy AS, Glazebrook J, Thordal-Christensen H, Katagiri F** (2011) Physical association of *Arabidopsis* hypersensitive induced reaction proteins (HIRs) with the immune receptor RPS2. *J Biol Chem* **286**: 31297–31307

**Raffaele S, Bayer E, Lafarge D, Cluzet S, Retana SG, Boubekeur T, Leborgne-Castel N, Carde JP, Lherminier J, Noirot E, et al** (2009) Remorin, a *solanaceae* protein resident in membrane rafts and plasmodesmata, impairs *potato virus X* movement. *Plant Cell* **21**: 1541–1555

**Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, et al** (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**: 1154–69

**Schollen E, Grunewald S, Keldermans L, Albrecht B, Korner C, Matthijs G** (2005) CDG-Id caused by homozygosity for an ALG3 mutation due to segmental maternal isodisomy UPD3(q21.3-qter). *Eur J Med Genet* **48**: 153–158

**Song W, Mentink RA, Henquet MGL, Cordewener JHG, Van Dijk ADJ, Bosch D, America AHP, Van der Krol AR** (2013) *N*-glycan occupancy of *Arabidopsis* *N*-glycoproteins. *J Proteomics* **93**: 343–355

**Sun L, Eklund EA, Chung WK, Wang C, Cohen J, Freeze HH** (2005) Congenital disorder of glycosylation id presenting with hyperinsulinemic hypoglycemia and islet cell hyperplasia. *J Clin Endocrinol Metab* **90**: 4371–4375

**Takahashi D, Li B, Nakayama T, Kawamura Y, Uemura M** (2013) Plant plasma membrane proteomics for improving cold tolerance. *Front Plant Sci*. doi: Artn 9010.3389/Fpls.2013.00090

**Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C** (2003) Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS (vol 15, pg 1895, 2003). *Anal Chem* **75**: 4942

**Varki A** (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**: 97–130

**Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, Etzler ME** (2009) *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press

**Xu SL, Medzihradzky KF, Wang ZY, Burlingame AL, Chalkley RJ** (2016) *N*-glycopeptide profiling in *Arabidopsis* inflorescence. *Mol Cell Proteomics* **15**: 2048–2054

**Zhang GA, Fenyo D, Neubert TA** (2009) Evaluation of the variation in sample preparation for comparative proteomics using stable isotope labeling by amino acids in cell culture. *J Proteome Res* **8**: 1285–1292

**Zielinska DF, Gnad F, Schropp K, Wisniewski JR, Mann M** (2012) Mapping *N*-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol Cell* **46**: 542–548

**Zielinska DF, Gnad F, Wisniewski JR, Mann M, Wiśniewski JR, Mann M** (2010) Precision mapping of an *in vivo* *N*-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**: 897–907

# Summary



This thesis is based on two technology projects from the Centre for BioSystems Genomics (CBSG), entitled ‘Comparative proteomics on Plant Pathogen interactions through enrichment of the *N*-glycoproteome and tagged-glycoproteome’ (TD7) and ‘Plasma Membrane proteomics for Plant Pathogen interactions’ (TD5). In the former project we developed the protocol for isolation, identification and quantification of *N*-linked glycoproteins from plants and used it to obtain a comprehensive inventory of glycan-occupancy of *Arabidopsis* glycoproteins. In the second project, a protocol for the enrichment of plasma membrane (PM) fraction from plant material was developed and applied to study the role of the PM proteome in the interaction of plants with the plant pathogen *Phytophthora infestans*. Combined these activities have resulted in a thesis devoted to technical developments in label-free comparative enrichment proteomics, with validation in a number of different biological systems.

In **Chapter 1** the background of these two enrichment proteomics approaches is briefly described. In this chapter I also briefly introduce the biology of the glycosylation mutants used in our glycoproteomics studies and the biology behind the two types of PM proteomics experiments that were performed on *Arabidopsis* genotypes with contrasting susceptibility to *Phytophthora brassicae* and on tomato suspension cells in response to infection by *P. infestans*.

*N*-glycosylation is one of the important protein modifications that occur in eukaryotes and plays an important role in protein folding, stability, localization and cell signaling. The importance of *N*-glycoproteins in mammalian organisms is well studied, however, in plants, much less is known. **Chapter 2** is an overview of recent developments in *N*-glycoproteomics in plants. In this chapter I discuss the different approaches to plant *N*-glycoproteomics and how these studies can be used to address important biological questions. I also discuss the technical challenges in studying *N*-glycoproteomics in plants and bring up some potential solutions and opportunities.

*N*-glycosylation in plants is different from that in mammals, which has consequences for the approach to *N*-glycoprotein/peptide enrichment. In **Chapter 3**, I describe how we modified the protocol for glycopeptide enrichment (as originally set up for samples from mammalian cells) for plant samples and how we had to make use of the *Arabidopsis* complex-glycan-less (*cgl1-1*) glycosylation mutant to allow full capture of the *N*-glycoproteome. In this work we identified 330 glycopeptides, belonging to 173 glycoproteins from *Arabidopsis* among which were 28 novel glycoproteins. A special problem encountered in this proteomics approach was the fact that common algorithms for protein identification are based on multiple peptide features, which especially for glycoproteins with single *N*-glycan group poses a problem.

Enrichment of the PM proteins may be useful to study proteome changes in the PM in relation to plant pathogen infection and resistance responses. In **Chapter 4**, I describe how PM enrichment was used to study the PM proteome of three *Arabidopsis* genotypes. *Arabidopsis* Col-0 is resistant to infection by *P. brassicae* isolate strain

HH. In contrast, transgenic *Arabidopsis* Col-0 expressing the effector *ipiO* (*35S-ipiO1*) and Col-0 with a T-DNA insertion in *LecRK-I.9* gene (*lecrk-I.9*) both show a gain of susceptibility to *P. brassicae*. Because we suspected that the differences in susceptibility are due to constitutive differences in PM proteins, we performed label-free comparative proteomics on PM protein composition in these lines, without pathogen infection. PM enrichment was performed using an aqueous two-phase partitioning system (ATPS). In the processing of the samples we encountered problems with PM fractionation and with soluble protein contamination. However, after protein identification, contaminating proteins were filtered out using bioinformatics by selecting proteins with predicted transmembrane domain (TMD) or glycosylphosphatidylinositol (GPI) anchor site. From the 2151 proteins that I identified in these samples, 613 are putative PM proteins of which 99 peptide features displayed significant quantitative differences between Col-0 and the *35S-ipiO1* and *lecrk-I.9* lines. Another major problem in label-free quantitative proteomics is the link between identified proteins and quantitative peptide features. Therefore, MS-Fit was employed to help link 25 out of the 99 differential features to 17 differential PM proteins. From these, CYSTEINE-RICH PROTEIN KINASE 37 (CRK37) is the only protein showing a peptide intensity that is higher in *Arabidopsis* Col-0 than in *35S-ipiO1* and *lecrk-I.9*. Subsequent validation studies showed indeed that an *Arabidopsis crk37* T-DNA insertion mutant displays a gain of susceptibility to *P. brassicae* infection. Interestingly, the transcript level of *CRK37* is not significantly different between Col-0 and the two phenocopy lines showing that proteomics is a valuable addition to transcriptomics studies for unraveling plant-pathogen interactions.

PM enrichment proteomics was also used to investigate the PM protein composition changes in the interaction between plants and *P. infestans*. In **Chapter 5**, initially enrichment of PM fractions was attempted from potato, but this was not successful. As an alternative, the interaction between tomato MsK8 suspension cells and *P. infestans* was used as a model system. Different PEG/Dextran concentrations and number of ATPSs were tested for optimal PM enrichment from these tomato cells, without too much loss in protein yield. The optimized PM protein enrichment procedure for MsK8 cells was subsequently applied to mock treated cells and cells treated with *P. infestans* zoospores or zoospore culture supernatant. Eventually 290 tomato protein were identified with significant peptide mass intensity difference ratio between control and treated cells. Among these were 17 receptor kinases, eight of which were manually verified for identity and differential abundance. Comparison of the identified tomato proteins with *Arabidopsis* homologs suggests that these proteins are indeed involved in biotic stress responses.

The challenges encountered during the different enrichment proteomics experiments in this thesis are explained and discussed in **Chapter 6** and potential solutions for the future are suggested. In this chapter I also compare the proteins identified by glycoproteomics (**Chapter 2**) and PM proteomics (**Chapter 3**) in *Arabidopsis*. The labor intensive procedures of enrichment proteomics and processing of the resulting data set did not leave enough time for validation of all *Arabidopsis* and tomato

## Summary

---

candidate proteins that were identified. Future validation should expose the full potential of the proteomics studies that I undertook and possibly result in additional candidates that can be used to protect crops against pathogens.



# Acknowledgements



Time really flies! Already eight years in the Netherlands and ten years in Europe. I am really glad I spent so many years in the Netherlands, at Wageningen University, along with all of you from the Laboratory of Plant Physiology, my Chinese fellows and my Dutch friends. So many happy memories, so many laughs, it made me feel Wageningen was my second hometown.

First I would like to thank CBSG for allowing two CBSG hotel proteomics projects to join to fund my PhD project.

There are so many people during those years I would like to give my sincere thanks. Forgive me that I can't mention all your precious names here, but I will remember all of you in my heart forever.

First of all, I would like to give my deepest gratitude to Sander. You are the most brilliant scientist I ever met and the great mentor for me in both science and life. I learned so many science related things from you. You are so patient and kind, considerate and caring to all your students. You are really like a father to us all. I don't know if the others also think so, but I do. Besides science, you also taught me how to deal with life's daily pressures, how to do time management, how to plan things well and how to work more efficiently. Also, one of the most important things I learned from you is how to enjoy life ;-). I very much enjoyed our yearly PhD boat trips, which I could even join during the time when I was working in Germany. You are a versatile person with many hobbies like painting, singing, swimming, diving and boating. I remember when I just started my PhD, you invited all your new PhDs to a concert of Beethoven symphony No. 9, in which you participated in the choir. I'm also a super classical music fan, and enjoyed the concert very much. It was a first time for me sitting in the front row during a live concert, watching my supervisor sing; very special. For both science and life, you have always encouraged and inspired me, no matter when and where. Your work will shine through my whole life.

Twan, you are a great scientist as well and you are my great mentor and teacher. You brought me into the proteomics field and taught me how to use the most high-tech LC-MS machines when I just started my PhD, always patient and kind. I remembered the many hours we spent together in the PC room in PRI for data analysis. You are a busy person, handling so many different projects. But you always tried your best to make time for my project and tried to give it high priority. You work so hard, making long hours. I remembered how many times we rushed out of the building, just before closing at 10:00 PM. After that you still had to drive back home to Utrecht. When I heard you had sleep problems because of work pressure I wished I could do more for you, not just recommending some yoga or Tai Chi to relax more. When I heard the tea I brought from China did help a bit, I felt much better. Over all these years I deeply appreciated your great efforts and knowledge. Without you, I could not have made a single step in the proteomics field. You are my life time mentor and friend!

Jan, I have to say, without your help nothing was going to happen for my PhD. How many samples you handled for me, always be very patient and focused on the work. I have to say that you work like a Chinese, dedicated and so hard, even during weekends. Because so many samples need to be processed, your job of keeping the LC-MS machine running is crucial for many projects, for which you sacrifice a lot of your spare time. You are a person who takes his responsibilities very serious. No matter how many samples I gave to you, you always managed to process them in time and get the most out of it. You are also expert in cycling, which makes you fit and keeps you in good condition all the time. Keep riding and keep fit! Maybe one day I can see you in Giro d' Italia ;-)

Harro, I want to express my sincere gratitude to you for all your great help and efforts. Being the most busy person in PPH, you have to take care of so many things, but you still give your best to spend time supervising PhDs, helping correct their thesis and in the meantime apply for some new grants. I remember the many rounds back and forth during manuscript revision, during which you remained patient and highly efficient. Many times I received your email with feedback late at night or just after your holidays. I deeply appreciate that. From you, I see how scientific life can be, quite exciting and challenging, but also enjoyable. I wish you all the best working at the University of Amsterdam with a fresh start group. Take care, and do your best!

Marielle, I deeply appreciate all your help in the lab. Without you, some experiments would have been a big headache for me. You keep the lab running smoothly and provide countless help to me and others. For us, you are the steward of the PPH lab, always willing to help. Here I represent all the people in our lab; we salute you!

Klaas and Francine, I give my deeply gratitude to both of you for helping me with all the pathogen related issues. Not only that, but you also taught me how to be a good scientist and researcher. I learned a lot by joining your group meeting. I appreciate your patience with me as naive person in the phytopathology field. You always encouraged and inspired me on how to improve and be critical. I enjoyed so much being a temporary member of the Laboratory of Phytopathology, as you treated me as one of your students, also inviting me for many dinners at home. We had many nice chats and discussions for which I would like to say thank you very much. Phytopathology is an exciting and challenging field where proteomics techniques can be applied, and I hope we can keep in touch for future collaborations.

Dmitry, you are good friend and also as a project collaborator. You were always energetic and work hard. And you are a good thinker, I remembered that when we first met you already had tons of questions ready for me. And later on in our lab you were unstoppably working and questioning, haha, such a busy time with you. And you were always kind, warmhearted and humorous, it was great to have you as collaborator in our lab. After studying in the Netherlands, our friendship continued in the Max Planck Institute of Plant Breeding in Cologne, both as post-doc. And I had a chance to work

---

in your lab for a short period of time, again with so many nice chats and laughs. I wish you all the best for your scientific carrier and maybe one day we will meet again somewhere in the world; I look forward to it!

Remco, you were a great MSc student and we worked hard in the same lab when I was not a PhD candidate yet. We had many nice chats and liquid nitrogen/dry ice plays, which I remember well, so many laughs. After your MSc, you started your PhD in Utrecht, while I started mine in Wageningen, but the world is so small: we met again in the same Max Planck Institute of Plant Breeding in Cologne as post-docs. I am sure you will make a lot of new friends there and I wish you all the best for your scientific carrier.

Maurice, as former PhD student of Sander, you provided great help at the beginning of my PhD. Your thesis worked as a tutorial and made it much easier for me to catch up all the aspects of glycobiology. You are always happy and smiling, which radiates onto the whole lab and the people around you.

I would like to also thank Aalt Jan for helping making the modified glycoproteome database and Pieter for programming the tools in Galaxy. Special thanks also to my external supervisor Dirk, to provide great ideas for glycoproteomics approaches at the beginning of my PhD.

Thanks to all PPH staff members for taking care of all the PhDs (Harro, Henk, Sander, Dick, Wilco, Leónie, Carolien, Richard, Iris, Francel, Jacqueline, Diaan), and Rina and Margaret for taking care all the tedious administration work, Dick for chairing the Monday morning seminars, Wilco for the journal club and Leo for Friday beer session, and the great PPH activity committee members. Also I would like to express my sincere thanks to all the PPH international colleagues (former and current), without you my PhD life would have been much less colorful and enjoyable. Special thanks to Esmer, Beatriz, Giovanni, Arman, Umidjon, Kristyna, Bas, Monica, Thierry, Natalia M., Lydia, Renake, Emilie, Elise, Sang, Alexandre, Gonda, Karen, Manus, Nafiseh, Cecilia, Desalegn, Julio, Jimmy and Ya-fen, Anderson, Deborah, Phuong, Lotte, Johanna, Wouter, Benyamin, Neli, Tila, Linda, Bart, Aaron, Natalia C., Aldana, Rik, Charles, Ana, Frank, Anna, Thamara, Margriet, Padraic, Ralph, Manickam, Thomas, and many others.

I would also like to give my special thanks to all my Chinese friends, former and current, for many years friendship in sports, gaming, traveling, dinners, and parties together. Special thanks to Wei Liu, Qing Liu, Ting Yang, Bing Bai, Jun He, Yanxia Zhang, Yuanyuan Zhang, Yunmeng Zhang, Yanting Wang, Junwei Liu, Hanzi He, Bo Wang, Xi Cheng, Jianhua Zhang, Yanli Wang, Feng Zhu, Xu Cheng, Tingting Xiao, Huchen Li, Guiling Ren, Tian Zeng, Defeng Shen, Jieyu Liu, Fengjiao Bu, Wenkun Zhou, Jianyong An, Xi Wan, Zhen Wei, Tao Zhao, Jiao Long, Pingping Huang, Lemeng Dong, Shuhang Wang, Chenlei Hua, Lisha Zhang, Weicong Qi, Xi Chen, Zhao Zhang,

Chunzhao Zhao, Hui Li, Yin Song, Xiaoqian Shi, Yan Wang, Yu Du, Miao Han, Guozhi Bi, Juan Du, Wei Qin, Chunxu Song, Yang Yu, Jingmeng Wang, Mingtian Yao, Xuan Xu, Kaile Sun, Zheng Zheng, Suxian Zhu, Liping Gao, Ke Lin, Songlin Xie, Xianwen Ji, Ningwen Zhang, Ying Zhang, Xingfeng Huang, Xiaomin Tang, and many others.

I also want to give my sincere gratitude to my Dutch friends in Wageningen, special thanks to my Karate Sensei: Sydney, Wijnand, Peter and Bart. To my house owner: Fely, Hans, Florene. To my family and sisters in Netherlands: Yoko, Jan, Mariko, Junko. You made me feel that Wageningen is more than just a city for study but a home.

致我亲爱的父母：

感谢父母对我一贯的信任和支持，是你们的爱鼓励着我不断前行，在我最困难无助的时候给予我力量和信心；是你们的爱激励着我一次又一次地超越自己，突破极限，克服种种艰难险阻，千言万语无以表达我对你们的爱，待儿子回国来报；也感谢亲人们在我身处异国他乡之时对我父母无微不至的关照。

Wei Song  
October 19, 2016  
Wageningen, the Netherlands



# Curriculum Vitae



## Curriculum Vitae



Wei Song was born on October 28, 1979 in Taiyuan, Shanxi, China. In 1998 he started his bachelor degree in olericulture at the department of Horticulture Science at Shanxi Agricultural University in Taigu, Shanxi, China. During the four year study he showed his interest in plant biochemistry. In 2002, after obtaining his bachelor degree, he continued with a master in plant biochemistry at the

department of Life Science at the same university. His study focused on elucidating the mechanism of botanical pesticide, in a project studying the effects of Marigold extract against *Fusarium Oxysporum*. During his master study, he was inspired by a lecture given by a Professor visiting from Wageningen, which lead to the decision to go to Wageningen University for further personal development. In 2006, he came to his dream pilgrimage university – Wageningen University. However, in the beginning life was not that easy, as there were no PhD positions available. However, he was given the chance to work as guest researcher at Plant Research International WUR under supervision of Prof. dr. Leo van Griensven, studying the immunomodulatory properties of medicinal mushroom extract. After three year hard work, he successfully published four articles, of which two as first author. After his research work in PRI, in 2009, he started his PhD at department of Plant Physiology, Wageningen UR, involved in two CBSG TD (technique development) projects: one in glycoproteomics and one in plasma membrane proteomics, with applications in plant-pathogen interactions. The PhD was under combined supervision by Dr. Sander van der Krol and Dr. Twan America and the work in plant-pathogen interactions was in collaboration with Prof. dr. Francine Govers and Dr. Klass Bouwmeester. During his PhD study he successfully published four articles, two as first author and hopefully there will be opportunity to publish the results from the last two experimental chapters as well. Before officially finishing his PhD, he got a scholarship in 2014 to work as postdoc for two year post-doc at the Max Planck Institute (MPI) of Plant Breeding in Cologne, Germany. The post-doc research was focused on characterization of acetyltransferases in *Arabidopsis*. In 2016 he returned to Wageningen to finalize his PhD thesis.



## Publications

### *Publications related to this thesis*

**Song, W.**; Mentink, R.; Henquet, M.G.L.; Cordewener, J.H.G.; Dijk, A.D.J., van; Bosch, H.J.; America, A.H.P.; Krol, A.R., van der. (2013) *N*-glycan occupancy of *Arabidopsis* *N*-glycoproteins. *Journal of Proteomics* 93: 343-355.

**Song, W.**; Henquet, M.G.L.; Mentink, R.; Dijk, A.D.J., van; Cordewener, J.H.G.; Bosch, H.J.; America, A.H.P.; Krol, A.R., van der. (2011) *N*-glycoproteomics in plants: perspectives and challenges. *Journal of Proteomics* 74(8): 1463-1474.

Bouwmeester, Klaas; Han, Miao; Blanco-Portales, Rosario; **Song, Wei**; Weide, Rob; Guo, LiYun; van der Vossen, Edwin; Govers, Francine. (2013) The *Arabidopsis* lectin receptor kinase LecRK-I.9 enhances resistance to *Phytophthora infestans* in Solanaceous plants. *Plant Biotechnology Journal* 12(1): 10-16.

Grace Armijo<sup>1</sup>, Paula Salinas, Mariela Inés Monteoliva, Consuelo García, Aldo Seguel, **Wei Song**, Alexander Ronald van der Krol, María Elena Álvarez and Loreto Holuigue. (2013) A Salicylic acid-induced lectin-like protein plays a positive role in the effector-triggered immunity response of *Arabidopsis thaliana* to *Pseudomonas syringae* AvrRpm1. *Molecular Plant-Microbe Interactions* 36(12): 1395-1406.

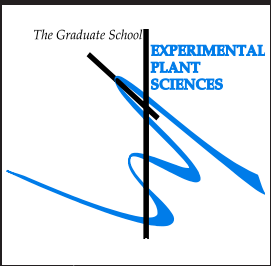
### *Other publications*

**Song, W.**; Griensven, L.J.L.D. van. (2008) Pro-and antioxidation properties of medicinal mushroom extracts. *International Journal of Medicinal Mushroom* 10(4): 315-324.

**Song, W.**; Helsper, J.P.F.G.; Griensven, L.J.L.D. van. (2008) Phenolic compounds present in medicinal mushroom extracts generate reactive oxygen species in human cells in vitro. *International Journal of Medicinal Mushroom* 10(1): 1-13.

Shnyreva, A.V.; **Song, W.**; Griensven, L.J.L.D. van. (2010) Extracts of medicinal mushrooms *agaricus bisporus* and *phellinus linteus* induce proapoptotic effects in the human leukemia cell line K562. *International Journal of Medicinal Mushroom* 12(2): 167-175.

Julia J Volman; Johannes P.F.G. Helsper; **Song Wei**; Johan J.P. Baars; Leo J.L.D. van Griensven; Anton S.M. Sonnenberg; Ronald P. Mensink; Jogchum Plat. (2009) Effects of mushroom-derived  $\beta$ -glucan rich polysaccharide extracts on nitric oxide production by bone marrow-derived macrophages and nuclear factor-kB transactivation in Caco-2 reporter cells: Can effects be explained by structure? *Molecular Nutrition & Food Research* 54(2): 268-276.

Education Statement of the Graduate School Experimental Plant Sciences	
<b>Issued to:</b> Wei Song <b>Date:</b> 19 October 2016 <b>Group:</b> Laboratory of Plant Physiology <b>University:</b> Wageningen University & Research	
1) Start-up phase	date
▶ First presentation of your project	
Title: Glycoproteomics in plants	Jan 18, 2010
▶ Writing or rewriting a project proposal	
▶ Writing a review or book chapter	
Review: N-glycoproteomics in Plants: Perspective and Challenges, Journal of Proteomics 74 (2011), 1463-1474. DOI: 10.1016/j.jprot.2011.05.007	Dec 16, 2010
▶ MSc courses	
▶ Laboratory use of isotopes	
Subtotal Start-up Phase	7.5 credits*
2) Scientific Exposure	date
▶ EPS PhD student days	
EPS PhD student day (Leiden, NL)	Feb 26, 2009
EPS PhD student day (Utrecht, NL)	Jun 01, 2010
EPS PhD retreat (Norwich, UK)	Aug 14-17, 2012
EPS PhD student day (Amsterdam)	Nov 30, 2012
▶ EPS theme symposia	
EPS theme 2 'Interactions between Plants and Biotic Agents', Utrecht University	Jan 22, 2009
EPS theme 1 'Developmental Biology of Plants', University of Leiden	Jan 30, 2009
EPS theme 2 'Interactions between Plants and Biotic Agents', Utrecht University	Jan 15, 2010
EPS theme 2 'Interactions between Plants and Biotic Agents', University of Amsterdam	Feb 03, 2011 Feb 10, 2011
EPS theme 3 'Metabolism and Adaptation', Wageningen University	Feb 10, 2012
EPS theme 2 'Interactions between Plants and Biotic Agents', Wageningen University	
▶ Lunteren days and other National Platforms	Apr 06-07, 2009
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 19-20, 2010
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 04-05, 2011
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	Apr 02-03, 2012
Annual meeting 'Experimental Plant Sciences', Lunteren, NL	
▶ Seminars (series), workshops and symposia	2009-2013
Invited seminars (attended 2x per year)	
▶ Seminar plus	
▶ International symposia and congresses	Oct 15-16, 2009
EUPP COST meeting (Nitra, Slovakia)	Oct 18-20, 2010
EUPP COST meeting (Luxembourg)	Sep 9-13, 2012
11th HUPPO World Congress (Boston, US)	
▶ Presentations	
Thematic meeting PRI Bioscience (2 X Talk)	2009-2012
Cluster meeting PRI (4 X Talk)	2009-2012
Poster (EUPP COST meeting Nitra)	Oct 15-16, 2009

Poster (CBSG summit TD5&7)	Mar 15, 2010
Lunteren meeting (Talk)	Apr 19-20, 2010
Poster (EUPP COST meeting Luxembourg)	Oct 18-20, 2010
CBSG summit TD7 (Talk)	Oct 21, 2010
Poster (CBSG summit TD5&7)	Jan 31, 2011
CBSG summit TD5&7 (Talk)	Oct 20, 2011
CBSG Pathogenomics TD5&7 (Talk)	Oct 25, 2012
CBSG Proteomics hotel TD5&7 (Talk)	Jan 18, 2013
▶ IAB interview	
Meeting with a member of the International Advisory Board of EPS	Feb 17, 2011
▶ Excursions	
Keygene excursion	Jan 26, 2012
NPC PhD day (visit DSM)	Jan 17, 2013
Rijk Zwaan excursion	Sep 27, 2013
Subtotal Scientific Exposure	29.5 credits*
<b>3) In-Depth Studies</b>	<b>date</b>
▶ EPS courses or other PhD courses	
Bioinformatics (EPS PhD course)	Mar 15-19, 2010
Proteomics (EPS PhD course)	Apr 21-23, 2010
4th European Summer School 'Proteomics basics: High Troughput Data Analysis and Statistics (EUPP COST) (Italy)	Aug 01-07, 2010
10th HUPO satellite workshop by ISB (Geneva)	Aug 29-Sep 02, 2011
▶ Journal club	
Literature discussions at PRI and Plant Physiology	2009-2012
▶ Individual research training	
Plasma membrane isolation from Arabidopsis (Lund University, Sweden)	Nov-Dec, 2009
Subtotal In-Depth Studies	11.7 credits*
<b>4) Personal development</b>	<b>date</b>
▶ Skill training courses	
EPS Expectation career day, Wageningen, NL	Nov 19, 2010
EPS Expectation career day, Wageningen, NL	Nov 18, 2011
EPS Expectation career day, Wageningen, NL	Feb 01, 2013
BCF event (Amsterdam RAI)	May 20, 2010
BCF event (Amsterdam RAI)	May 26, 2011
BCF event (Amsterdam RAI)	May 24, 2012
CBSG matchmaking event	Oct 18, 2012
NPC successful conferencing	Oct 31, 2012
Make more business with your research	Apr 23, 2013
▶ Organisation of PhD students day, course or conference	
▶ Membership of Board, Committee or PhD council	
Subtotal Personal Development	2.5 credits*
<b>TOTAL NUMBER OF CREDIT POINTS*</b>	<b>51.2</b>
Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits	
* A credit represents a normative study load of 28 hours of study.	

This work was performed at the Laboratory of Plant Physiology, Wageningen University, with financial support from the Centre of BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

**Cover design and layout:** Wenbing Li & Yu Zhang

**Printed by:** Proefschriftmaken.nl || Uitgeverij BOXPress