

Proteins make the difference



In collaboration with Plant Research International

**Dr. J.P.F.G. Helsper
Dr. A.H.P. America
Dr. J.H.G. Cordewener
Dr. I.M. van der Meer**

**Prof. Dr. H.J. Bosch
Dr. M.G.L. Henquet
Dr. A.R. van der Krol**

Introduction

Given your interest in producing proteins and/or glycoproteins with various functionalities at a commercial scale and the expertise and infrastructure at Plant Research International (PRI) in glycomics and proteomics, we are potential collaborators in the development of novel proteins and glycoproteins.

Proteomics analyses add to characterization at the genomics level in that it gives direct confirmation of how native or newly introduced genes are expressed. Expression includes their modes of post-translational modifications, which are of high relevance for biological activity and stability. Since glycoproteins are often the ultimate metabolic end product of genetic expression proteomics analyses provide information which will remain inaccessible for regular genomics and transcriptomics strategies.

In the recent past, PRI has developed strategies to produce plants and plant cell cultures from various origins and transformed with different gene constructs. Characterization of these plant materials requires that the analysis for quantity and quality of the (glyco)proteins is performed at a (semi) high throughput scale.

This survey gives a detailed description of the state of development and potential for high-throughput performance of the strategies in which PRI can fulfill a major role. This survey is divided in four sections:

Gene constructs

I) Gene constructs for glycoprotein production

Chemical analyses

II) Proteomics strategies

III) Proteins containing post-translational modifications

IV) Glycan and glucan structure analyses

I) Gene constructs at PRI for glycoprotein production

At PRI gene constructs of animal, microbiological or plant origin have been developed which show expression in plant systems. These gene constructs may be applied to modify and/or control the glycan structure of glycoproteins. Gene constructs encoding for glycosyltransferases, which extend glycan chains, and glycosidases, which trim existing glycans to the appropriate structure, are available in this gene collection. In addition, timing of transferase and glycosidase activities may be controlled by directing the localisation of each enzyme activity in the biosynthetic process (proximal/distal Golgi apparatus or ER) via the protein structure of the transmembrane domain.

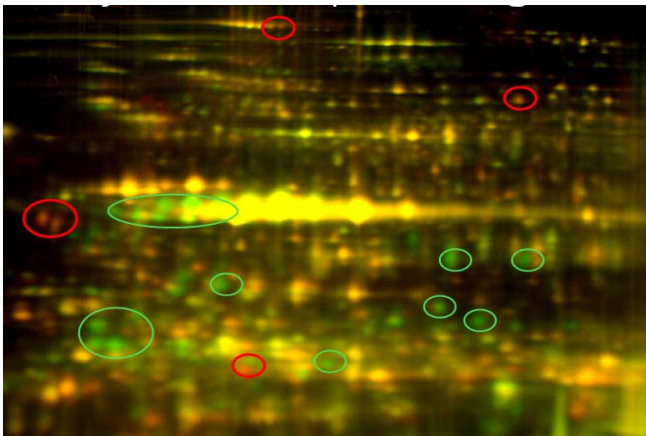
Below a survey of the enzyme activities for which these gene constructs encode:

- Arabidopsis GnT-I (N-acetylglucosaminyl transferase I)
- Arabidopsis GnT-II (catalytic domain)
- Humane GnT-I
- Humane GnT-III
- Humane GnT-IV
- *Tetraodon nigroviridis* α 1,3-fucosyltransferase (Lewis X)
- *Tetraodon nigroviridis* α 1,2-mannosidase (Man-I)
- Arabidopsis Man-II catalytic domain
- Humane Man-II
- *Spodoptera frugiperda* Man-III
- *Beta vulgaris* α 1,4-fucosyltransferase (Lewis A)
- Humane and bovine β 1,4-galactosyltransferase I (β 1,4-GalT1)
- Rat α 2,6-sialyltransferase
- Bovine α 2,6-sialyltransferase
- Rat β 1,3-glucuronyltransferase

II) Proteomics technologies

2D electroforesis usig fluorophores

In 2D-electrophoresis proteins, covalently labeled with fluorophores, are separated using isoelectric focussing (IEF) (1st dimension) and SDS-PAGE (2nd dimension). Prior to 2D-electrophoresis proteins are covalently labeled with fluorophores, which differ between samples. On the basis of distinctive emission wavelengths from these fluorophores, proteins from 3 samples can be quantitatively compared in a single 2D gel.

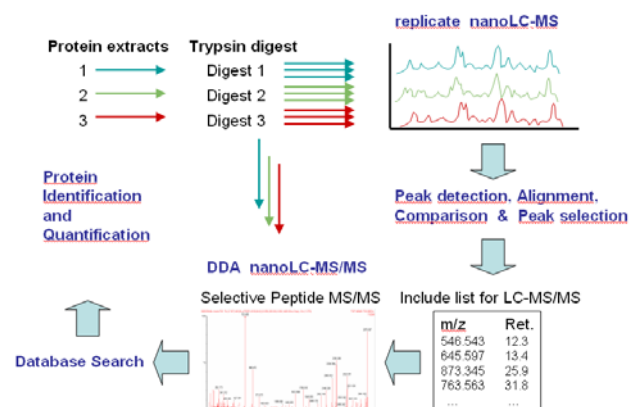


The adjacent figure shows a mixture of 2 protein samples. Proteins, unique for one of two samples, are coloured red or green, respectively. Yellow spots represent proteins shared between both sources.

In comparison to earlier 2-D strategies accuracy in quantification and matching between protein patterns are drastically enhanced. Dependent on sample complexity 1500-2500 protein spots can be detected per sample. Ten gels can be run simultaneously, so that in a single experiment 20 samples (or 10 duplicates) can be analysed. The accuracy in quantification allows distinction of 1.5- to 2-fold (and more) in protein abundance.

Quantitative Comparative LCMS

The method of choice for comparison of protein composition between multiple samples is LCMS of their tryptic peptides. As for 2D-electrophoresis, LCMS allows detection of 1.5- to 2-fold differences in abundance. In each LC-run up to 10,000 peptides are investigated, dependent on the complexity of the original protein mixture which may contain up to 100 proteins. In the LCMS-approach a protein mixture can be analysed without prior separation by gel electrophoresis. The peptide mixtures are separated using reversed phase chromatography (RP-C18), linked on-line to a mass spectrometer. For very complex protein mixtures two-dimensional LC-MS may be applied allowing detection of 1000 or more proteins.



Protein identification and quantification using LC-MSMS

Proteins can be identified by sequencing trypsin-derived peptides with LC-MSMS. In this strategy individual peptides are separated in the mass spectrometer in the MS-mode on the basis of molecular mass and charge (m/z -value). In the MSMS-mode each individual peptide is fragmented further to smaller peptides with sequential loss of amino acids in the MSMS-mode.

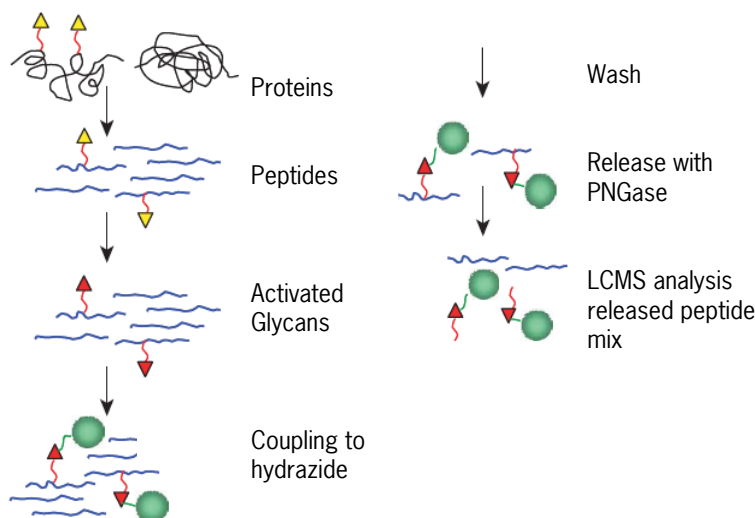


Using LC mass spectrometry in the MSMS mode individual proteins can also be quantified on the basis of three peptides with the highest response factor by comparison with standard proteins. Dedicated software can match peptide fragmentation patterns to identify the native protein and calculate their abundancies.

III) Proteins containing post-translational modifications

Bead-mediated glycoproteomics

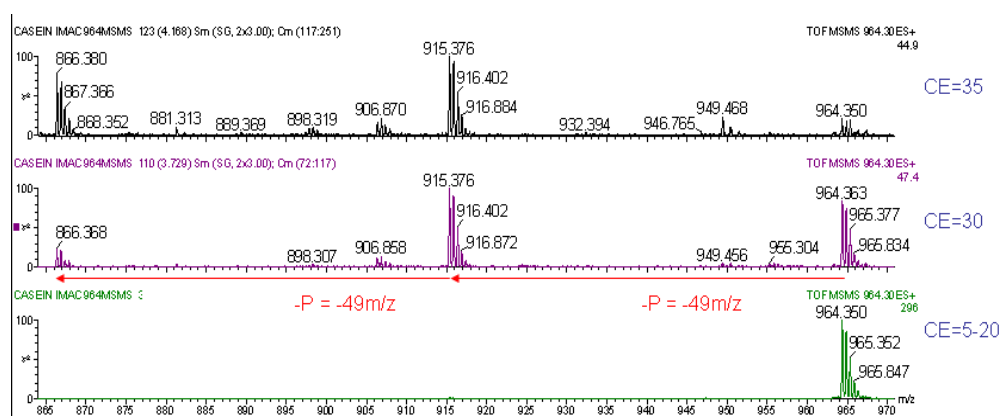
The concept of bead-mediated enrichment of glycopeptides from a total peptide mixture aims to greatly reduce the complexity of the peptide pool that needs to be analyzed, which enables direct analysis by LCMS. It also facilitates high throughput analysis (HTP) of samples, which enables a more detailed analysis of changes in the proteome, e.g. the effects of multiple environmental conditions or treatments resulting in (a)biotic stress. When this strategy is combined with cell specific glycan tagging, it enables cell specific glycoproteomics. For cell specific glycan-tagging the *Arabidopsis cgl* mutant is used, which is defective in the enzyme GnTI, and therefore has no glycoproteins containing complex N-glycans. Complex N-glycan processing is restored in specific cell types of this mutant and glycoproteins/peptides with complex glycans can be specifically eluted and analyzed by LCMS.



Glycopeptides are enriched from the entire proteome on the basis of their capacity to form a covalent linkage to hydrazide residues, which are immobilized on agarose beads. The conjugated glycopeptides can be released by a peptide-N-glycosidase. The eluted peptide mixture can then be analyzed by LCMS.

Assessment of glycoforms and location on the peptide backbone

Post-translational modifications (PTMs), like phosphorylation and glycosylation, can be identified while still attached to the peptide by LC-MSMS. Using this approach insight can be gained on the various glycoforms, i.e. isoforms of the same protein with varying (or no) N-glycan structures attached. Location assessment of the modification is often, but not always, possible and depends on the stability of the modification relative to that of the fragmented peptide. PTMs can be detected in a dedicated, targeted approach, where the LC-MS method can detect modified peptides in a complex mixture. Alternatively, a non-targeted approach can detect (also unexpected) PTMs in peptides (proteins) with known amino acid sequence.



The figure above shows a shift in m/z-value (X-axis) from 964.4 to 915.4 due to release of a phosphate residue at collision energy (CE) = 30. At a higher CE of 35 also the peptide backbone is fragmented thus revealing the amino acid sequence and the amino acid on which this post-translational modification is located.

However, in-source fragmentation, in particular of the glycan part of the molecule, may underestimate some of the glycoforms and overestimate others.

Assessment of glycoforms using MALDI-TOF

Once the location of the glycan on the protein backbone is known, MALDI-TOF analyses may be performed at high throughput rate of the glycopeptide structures. This approach is not impaired by in-source fragmentation and forms a good complement of the LCMSMS approach described above to achieve a more accurate quantification of glycoforms.

IV) Glycan and glucan structure analyses

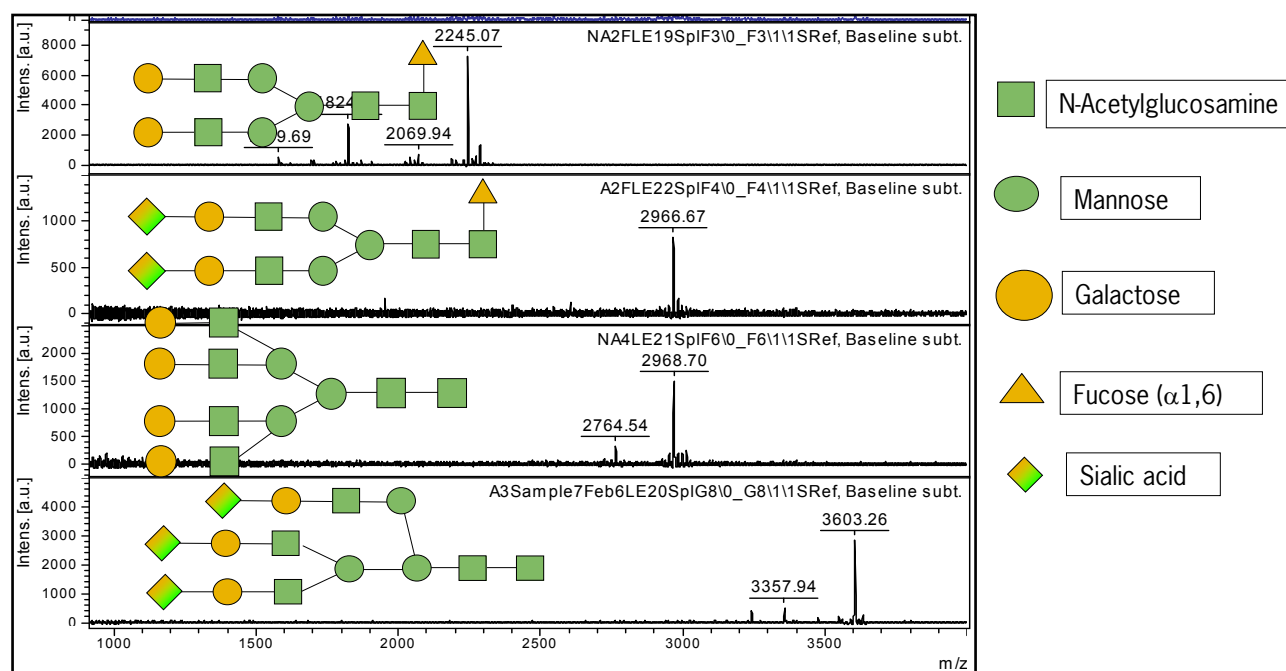
Monosaccharide composition, degree of polymerization

The monosaccharide composition of oligo- and polysaccharides forms the basis for their biological activities. For example, galactose in N-glycans is essential for the anticarcinogenic activity of glycoproteins and protection (capping) of galactose residues by sialic acid prevents their degradation in the blood stream, while fucose and xylose in N-glycans can make the glycoprotein allergenic.

Monosaccharide analysis includes acid hydrolysis followed by separation of monosaccharide species on a Dionex HPLC facility, equipped with pulsed amperometric electrochemical detection (PAED). When applied to non-hydrolysed samples the approach provides information on degree of polymerization. Given the selectivity of PAED, sample preparation does not require extensive purification protocols and thus also facilitates high-throughput approaches and enables analysis of 100 samples per day.

Maldi-TOF analyses on N- (and O-) glycans

Maldi-TOF analysis of N-glycans, after enzymatic detachment from the glycoprotein, is the state-of-the-art method for their identification and used in quality control for pharmaceutical applications. Maldi-TOF analyses are semiquantitative allowing relative comparisons between samples, but not absolute quantification. Maldi-TOF will give direct information on molecular masses from which in most cases complete chemical identification will be possible, although isomers cannot be distinguished in first dimension MS or TOF. Distinction between isomers can often be accomplished by MSMS or MSⁿ. As complementary strategies monosaccharide analysis and glycosidic linkage analysis (see below) can fill up the gaps of information left by Maldi-TOF, e.g. distinction of galactose from mannose, which are both hexoses and therefore not distinguished by Maldi-TOF.



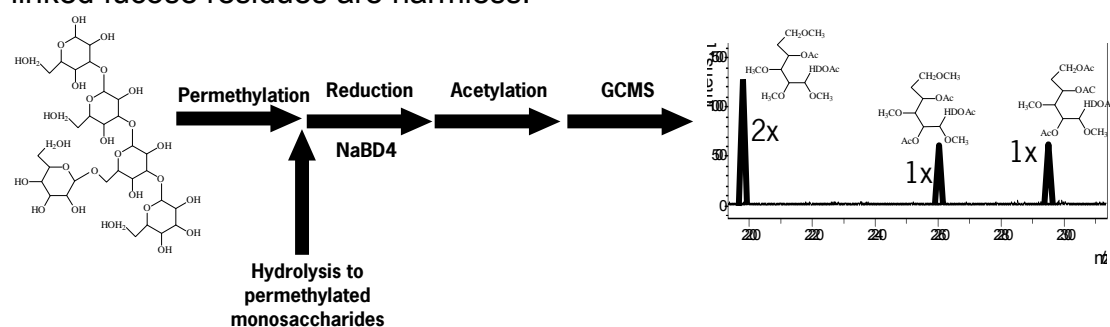
Maldi-TOF pattern of sialylated and/or fucosylated N-glycans after permethylation

High-throughput Maldi-TOF analysis is possible at a rate of about 300 samples per day as far as generation of Maldi-TOF spectra is concerned. The threshold here is clean-up of the samples (at present about 25 per day) and interpretation of the spectra. The former can be further robotised. Rate of interpretation of the spectra is strongly dependent on their complexity and the level of information density required.

For O-glycans the hydrolysis procedure leading to deproteinated glycan structures still has to be implemented at PRI. The basic procedure reported in literature is a chemical, alkaline hydrolysis as compared to enzymatic hydrolysis for N-glycans. Alkaline hydrolysis has been described to be problematic due to the possibility of artifacts.

Glycosidic linkage analyses

Glycosidic linkages are relevant for biological activities such as immunogenic, anticarcinogenic and immunomodulatory activities, clearance rate from the blood stream, cell signalling processes, etcetera. For example α -1,3 linked fucose residues, as occurring in plants, make glycoproteins allergenic and should thus be absent for pharmaceutical application, while animal-type α -1,6 linked fucose residues are harmless.



The state-of-the-art method for glycosidic linkage analysis in N- and O-glycans and other oligo- and polysaccharides is permethylation followed by hydrolysis, reduction to alditols, acetylation and GCMS of the resulting PMAAs (partially methylated alditol acetates). GCMS of PMAAs is a routine protocol at PRI. In this strategy, all free OH-groups in an oligo- or polysaccharide, which are not involved in a glycosidic linkage, are first methylated. Then acid hydrolysis and reduction with sodium borodeuteride will convert all linkage-involved oxygens into OH-groups which are subsequently acetylated to PMAAs. The mixture of these PMAAs can be quantitatively analysed by GCMS. These GCMS patterns provide the information for elucidation of the glycosidic linkages and branching level in the original oligo- or polysaccharide.

This strategy has been developed for structural analysis of N-glycans and branched glucans. The latter are relevant for controlling serum cholesterol and glucose levels and immunomodulatory activities of fibres from fungi and higher plants, e.g oat. It also allows determination of the degree of capping of galactosyl residues by sialic acid which determines their stability in the blood stream.