DETERMINATION OF INAUTHENTIC PROTEIN GLYCOSYLATION IN TRANSGENIC PLANTS

Project Ref. No G03029

Funded for 2 years work at Cambridge and 18 months at Rothamsted over a period of 3 years

Project Lead Contractor:
Professor Peter Shewry.
Rothamsted Research,
Centre for Crop Genetic Improvement.
West Common,
Harpenden,
AL5 2JQ

Dr. Paul Dupree:
University of Cambridge,
Department of Biochemistry,
Hopkins Building,
Downing Site,
Cambridge,
CB2 1QW

Authors of Report at Rothamsted Research:
Professor Peter Shewry.
Dr. Alison Lovegrove.
Dr. Justin Marsh.

Authors of Report at University of Cambridge:
Dr. Paul Dupree.
Dr. Elaine Stephens.
Dr. Theodora Tryfona.

Cambridge Start Date: 1st October, 2008.
Rothamsted Start Date: 1st April, 2009.
Cambridge Finish Date: 31st December, 2010.

Note: As a result of an agreed change in staff working hours at Cambridge, i.e. Dr. Tryfona became 50% on the FSA project from 1st July, 2009. This then had an effect on the Rothamsted time-line of work, such that Rothamsted finished experimental work on 4th February, 2010.

Executive Summary

The main aims of project GO3029 was the development of methods to identify and characterise *N*- and *O*-linked glycans on plant proteins to determine whether proteins expressed in transgenic plants are inauthentically glycosylated.

Methods developed and evaluated at the University of Cambridge were successfully transferred to Rothamsted Research where they were applied to transgenic and non-transgenic plants. These included the transgenic pea lines expressing the bean α-amylase inhibitor (Prescott et al. 2005) which had caused public concern when they were used in animal feeding studies and found to produce an altered immune response, compared to the native bean α-amylase inhibitor.

The *N*-glycosylation methods developed at the University of Cambridge were reduced to a Standard Operating Procedure (SOP) at Rothamsted Research suitable for routine use in industry and by regulatory authorities (Appendix 1).

Analysis of *O*-linked glycans was much more technically challenging however, excellent progress was made. The methods developed at the University of Cambridge were evaluated using arabinogalactan peptide (AGP) isolated from bread wheat. These methods were also successfully transferred to Rothamsted Research where they were used to analyse AGP isolated from transgenic bread wheat.

Methods for analysing *N*-glycans in transgenic plants were based on current protocols already employed by the University of Cambridge for identifying *N*-glycosylation in insect cells (Stephens et al. 2004; Aurikko et al. 2005; Maslen et al. 2006). Briefly, protein mixtures (or purified glycoproteins) were precipitated, resolubilised and digested with proteases. The *N*-glycans are then released from glycopeptides using peptide *N*-glycosidase A. The released glycans were purified from the complex peptide mixture using a reverse-phase cartridge and the glycans were desalted. The native *N*-glycan mixtures were then analysed by MALDI-ToF-MS in reflectron mode using 2,5-dihydroxybenzoic acid as the matrix. The method was tested in Arabidopsis and cauliflower.

These methods were then used at Rothamsted Research to analyse purified phaseolin from *Phaseolus vulgaris* and various other bean species as well as the α-amylase inhibitor from native bean and the transgenic pea lines from the Australian laboratory (Prescott et al. 2005).

It was found that the method developed at Cambridge was very sensitive. Three new structures were identified in Arabidopsis that had not previously been reported and one new

structure from cauliflower was also detected. The robustness of the method was confirmed when it was successfully transferred to Rothamsted Research. The sensitivity was also tested and it was found that even at the level of single seed extraction the same glycans could be identified. The method was applied to purified proteins (phaseolin and α-amylase inhibitor) as well as to crude protein preparations from cauliflower, peas and beans. The glycan profiles from various species of bean were determined the method and showed that whilst the same glycans were detected, the relative amounts of the glycans varied between species. The method included MALDI-ToF mass spectrometry and was therefore not strictly quantitative but the relative levels of the glycans appear to be indicative of a high degree of natural variation of glycan profile between species. When comparing the α-amylase inhibitor from native and transgenic origin variation in glycan profiles was again observed. (Statistical analysis is on-going in order to determine if canonical variates analysis (CVA) is able to separate the data points). Until these analyses are completed it is not possible to conclude whether the difference in glycan profiles between transgenic and native α-amylase inhibitors proteins are outside the natural range of variation, but we can clearly distinguish the two profiles.

O-linked glycosylation was investigated, initially using GFP-reporter constructs expressed in transgenic tobacco (Tan et al. 2004). Following hydrolysis with barium hydroxide sugar analysis was carried out using HPAEC-PAD. This gave the expected peaks, in accordance with the structure reported (Tan et al. 2004). Initial mass spectrometric analysis of the complex mixture did not yield the expected oligosaccharides mainly due to the great complexity of the polysaccharide. Therefore, normal phase-liquid chromatography (NP-LC) in conjunction with MALDI-ToF-MS were employed in order to reduce the complexity of the sample and aid the structural analysis. Pre-fractionation with NP-LC confirmed the great heterogeneity of the tobacco sample and attempts to further simplify the sample with the use of AGP specific hydrolytic enzymes were made. Enzyme hydrolysis and consequent fractionation with NP-LC greatly reduced the complexity of the polysaccharide sample and showed that our methodology could be used successfully for the detection of *O*-linked glycans. The method was evaluated using wheat flour AGP (Fincher et al. 1974). Briefly, glycoprotein samples were hydrolysed with TFA and their monosaccharide content was analysed by HPAEC-PAD, this allowed the identification of compositional profile (quantitative and qualitative) of the AGP sample. For the structural characterisation of the AGP side chains, the glycoproteins were digested with exo-β-(1→3)-galactanase (Tsumuraya *et al.* 1990) and the oligosaccharides produced were purified by reverse-phase cartridge and

cation exchange clean-up. Accordingly the purified glycans were per-methylated and analysed by MALDI-ToF-MS.

Following evaluation of the method at Cambridge with AGP from bread wheat AGP the method was transferred to Rothamsted Research. At Rothamsted the method was applied to an AGP preparation from bread wheat and the same oligosaccharides identified, verifying the robustness of the method. AGP isolated from durum and einkorn wheats was also analysed and a similar set of oligosaccharides identified. Finally, the method was applied to control and transgenic bread wheat lines, expressing extra copies of AGP. As with *N*-glycan analysis**,** the *O*-linked glycan methods developed worked well but statistical analysis will be required to determine whether the differences are of significant when compared with the range of natural variation.

Glossary

MADLI -TOF MS	Matrix assisted desorption laser ionisation-Time of flight mass spectrometry.
N-glycosylation	Glycans attached on the amide nitrogen of asparagines residues.
O-glycosylation	Glycans attached to the OH groups of hydroxyproline.
Per-methylation	All hydrogen atoms on the hydroxyl groups of carbohydrates are replaced by methyl groups.
HPAEC-PAD	High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection.
PA20	Carbo Pac PA20 strong anion exchange column, that provides high-resolution separations of mono- and disaccharides.
PA200	CarboPac PA200 strong anion exchange column, that provides high-resolution separations of oligosaccharides.
Normal Phase Liquid Chromatography	Polar stationary phase and less polar mobile phase.
CID	Collision induced dissociation
PNGase A	Peptide N-glycosidase A
GlcNAc	N-acetylglucosamine

Gal	Galactose
Man	Mannose
Ara	Arabinose
Fuc	Fucose
Xyl	Xylose
Rha	Rhamnose
GlcA	Glucuronic acid
AGP	Arabinogalactan peptide
Hyp	Hydroxyproline
PACE	Polysaccharide Analysis using Carbohydrate gel Electrophoresis.

Aims and Objectives

In order to achieve acceptable levels of stable accumulation of recombinant proteins in transgenic plants it is usual to direct them to the secretory system, using specific signals to ensure that they either accumulate directly within the lumen of the endoplasmic reticulum (Molvig et al, 1997, Hagan et al, 2003, Conrad and Fiedler1998,) or are transported via the Golgi apparatus to the vacuole. This can certainly result in high levels of accumulation (over 5% total seed protein), but can also result in inauthentic processing, due to the action of enzyme systems present in the ER, the Golgi or the vacuole. Although a range of protein modifications occur in plants the two most important in relation to protein accumulation in transgenic plants are glycosylation and proteolysis. This is because inauthentic processing in heterologous host systems may lead to changes in the structure and properties of the recombinant proteins which may affect their immunogenicity and hence potentially also their allergenicity. Crops expressing heterologous proteins are being developed for improved nutrition, stress response and pathogen resistance. In addition, they are very attractive hosts for the production of recombinant proteins of pharmaceutical interest, as they are inexpensive and versatile systems and provide rapid and economic scale up. However, inauthentic processing of proteins is a major limitation of plant expressions systems.

Glycosylation is the major type of post-translational processing occurring in plants with two major types. *N*-glycosylation occurs on the amide nitrogen of asparagine residues with the initial glycosylation occurring in the ER and further modifications and trimming occurring in the Golgi and vacuole. It is possible to predict potential *N*-glycosylation sites based on sequence context but these may not be recognised *in planta*. Similarly, even when proteins are expressed in non-transgenic plants the extent to which these sites are recognised may vary (eg. in phaseolin of bean (Bollini *et al.* 1983)). *N*-glycosylation may also occur when proteins are expressed in heterologous hosts at sites which are not recognised in the species of origin, as reported recently for the bean α-amylase inhibitor expressed in pea seeds (Prescott *et al.* 2005). Nevertheless the structures of *N*-linked glycans tend to be conserved between species and proteins, meaning that it is possible to adapt established methodologies to develop routine methods to identify the presence of *N*-linked glycans and determine their structures.

The occurrence, extent and pattern of inauthentic glycosylation are important as glycan groups are known to be highly immunogenic (and potentially also allergenic). Thus, although Prescott *et al* (2005) showed that the α-amylase inhibitors expressed in pea

(recombinant) and bean (authentic) differed in several respects, they suggested that differences in glycosylation patterns may have accounted for the fact that the recombinant form but not the authentic form was highly immunogenic. Similarly, *N*-glycosylated forms of the cereal α-amylase inhibitors appear to be more active in bakers' asthma than their unglycosylated counterparts (Sanchez-Monge *et al.* 1992; van Ree *et al.* 2000). The mugwort pollen allergen Art v 1 contains unusual *O*-linked glycans on Hyp (Leonard *et al.* 2005), which may contribute to the allergenicity. There is also evidence that inauthentic *O*-glycosylation can occur when proteins are expressed in heterologous hosts. For example, it has been reported that expression of the normally non glycosylated sweet potato sporamin in tobacco leaf cells, leads to arabinogalactan addition (Matsuoka *et al.* 1995).

The overall aim of this project was to develop methods to identify and characterise *N*- and *O*-linked glycans on plant proteins, in order to determine whether transgenically expressed proteins are inauthentically glycosylated and to reduce these methods to standard operating procedures for routine use in industry. Methods for the analysis of *N*-glycosylation were already established but it was necessary to evaluate them and modify those which were most appropriate for routine application. Development of methods for the analysis of *O*-linked glycans was much more challenging due to the great complexity of arabinogalactan peptides (AGPs). In fact, AGPs may be the most complex macromolecules in nature and so far no complete structure has been solved for any AGP (Majewska-Sawka *et al.* 2000; Gaspar *et al.* 2001).

Objective 01 (Cambridge)

Establishment of methods to routinely determine patterns of protein N-glycosylation in transgenic plants using protein mixtures.

Objective 02 (Cambridge)

Establishment of methods for the analysis of O-glycosylation in transgenic plants.

Objective O3 (Rothamsted)

Evaluation of method for N-linked glycan analysis on "native" and recombinant proteins

Objective 04 (Rothamsted)

Evaluation of methods to study O-glycosylation of wheat arabinogalactan peptide.

Objective 05 (Rothamsted)

SOP for N-glycans (appendix 1)

Experimental Procedures

Protein acetone precipitation

Glycoprotein sample (about 50 µg of protein) is suspended into cold (-20ºC) acetone volume five times that of the protein sample to be precipitated and after thorough mixing is incubated for 24 h at -20ºC. Accordingly, the glycoprotein sample is centrifuged for 5 min at 13,000-15,000 x g, the supernatant is decanted and disposed making sure not to dislodge the protein pellet. Acetone is allowed to evaporate at room temperature for 30 min. Once dry, the buffer for the downstream process is added and is mixed thoroughly to dissolve protein pellet.

N-glycan release from plant glycoproteins

Glycoprotein sample is dissolved in 100 µl of pepsin digestion buffer (5% formic acid, pH 3) and 28 µL of pepsin solution were added. The mixture was incubated at 37ºC for 24 h and the enzyme was inactivated by incubation at 100ºC for 5 min and the glycopeptide sample was lyophilized. Accordingly, the lyophilized samples were dissolved in 100 µL of PNGase A digestion buffer (50 mM ammonium acetate, pH 5) and 4 µL of the enzyme were added. The mixture was incubated at 37ºC for 16 h, followed by lyophilization.

If immobilized pepsin is used then the glycoprotein sample is dissolved in 100 µl of pepsin digestion buffer (100 mM ammonium acetate pH 4.5, adjust pH with 5% acetic acid) and 28 µL immobilized pepsin (Poroszyme®) are added. Incubate at room temperature for 24 h with constant mixing and remove pepsin beads by centrifugation for 5 min at room temperature at 10,000 rpm (with the use of *Nanosep* spin cartridges). The flow-through is kept and add 4 µL PNGase A are added. The mixture is incubated at 37ºC for 16 h followed by lyophilization.

Exo-β-1,3-galactanase digestion of arabinogalactan polysaccharides

Arabinogalactan peptide preparations (1 mg) were digested with exo-β-(1→3)-galactanase (Tsumuraya et al. 1990) (EC 3.2.1.145; 16 mU) in 20 mM ammonium acetate buffer (pH 4.6, 100 µl) at 37 ˚C for 24 h. The enzyme was inactivated for 5 min at 100 ˚C and the sample was dried in a rotor evaporator.

AGP oligosaccharide sample desalting and clean-up

Following the enzyme digestions and prior to per-methylation released peptides and enzymes were removed using reverse-phase Sep-Pak C_{18} cartridges (WATERS) as follows: prior to use, the columns were washed with 5 ml 99.8% MeOH (HPLC grade) ensuring that no air bubbles appear in the Sep-Pak cartridge, followed by a wash with 5 ml of 5% (v/v) acetic acid (Fisher). The samples were dissolved in 0.5 ml of 5% (v/v) acetic acid and were loaded onto the cartridge, avoiding the appearance of air bubbles. Accordingly, the arabinogalactan polysaccharides were eluted with 5 ml of 5% (v/v) acetic acid and were lyophilised. Dry arabinogalactan polysaccharides were desalted using Dowex beads (50 x 8, H^+ form, 50-100 mesh) as follows: prior to use Dowex beads were equilibrated by washing 3 times with 4 M HCl, followed by repetitive washes with water until the pH of the water wash measured by paper pH is slightly below 7. Accordingly, beads were washed 3 times with 5% (v/v) acetic acid and were stored in the 5% (v/v) acetic acid solution for several months at room temperature. A Pasteur pipette plugged at the tapered end with a small amount of glass wool, was packed with Dowex beads and the column was washed with 2 volumes of 5% (v/v) acetic acid (approx. 4 ml). Dry samples were dissolved in 0.5 ml of 5% (v/v) acetic acid and were loaded onto the column, arabinogalactan polysaccharides were eluted with 4 ml of 5% (v/v) acetic acid. Consequently, purified polysaccharides were lyophilised.

Per-methylation of arabinogalactan polysaccharides

Per-methylation of glycans was performed using the method described by Ciucanu and Kerek (Ciucanu et al. 1984). Glycans were lyophilised in a glass tube and in order to avoid undermethylation, samples were kept in the lyophiliser until the time of *per*-methylation. In a dry mortar a few NaOH pellets were placed and about 3 ml of dry DMSO (Romil) were added. The NaOH pellets were ground until a slurry was formed and 0.5-1 ml of the slurry was added to each sample tube. Accordingly 1 ml of methyl iodide (Fluka) was added and the glass tubes were sealed with a screw cap. The samples were mixed vigorously for 10 min and the reaction was quenched with the addition of 1 ml water. Consequently, 2 ml of chloroform were added and the samples were mixed well and allowed to settle into two layers. The upper aqueous phase was removed by a Pasteur pipette and was discarded, while the lower phase was washed several times with water. Samples were then dried under a gentle N_2 stream and dry samples were resuspended in 100 µl MeOH and were kept at room temperature for MALDI-ToF-MS analysis.

Purification of* per-*methylated N-glycans

Per-methylated *N*-glycan samples were dissolved in 200 µL of 35% acetonitrile and salt was removed using reverse-phase Sep-Pak C_{18} cartridges (WATERS) as follows: prior to use, the cartridges were washed with 5 ml of HPLC grade water followed by 5 ml acetonitrile, 5 ml of ethanol and 10 ml of HPLC grade water ensuring that no air bubbles appear in the Sep-Pak cartridge. The samples were loaded onto the cartridge and glycans were eluted with 3 ml of 50% (v/v) acetonitrile and were lyophilised.

MALDI-ToF/ToF-MS/MS

Methylated MeOH dissolved samples (5 µl) were mixed 5 µl of 2,5-DHB matrix (10 mg/ml dissolved in 50% MeOH) and 1 µl of the mixture was spotted on a MALDI target plate and was rapidly dried in a vacuum dessicator in order to produce small crystals for easy spectral acquisition. Samples were analysed by MALDI-ToF/ToF-MS/MS (4700 proteomics analyser, Applied Biosystems, Foster City, CA, USA) as previously described.(Maslen et al. 2007) This MALDI tandem amss spectrometer uses a 200 Hz frequency-triple Nd-YAG laser operating at a wavelength of 355 nm. Averages of 2000 laser shots/spectrum were used to obtain MS spectra and glycan molecular ions $[M+Na]^+$ were identified. High energy MALDI-CID spectra were acquired with an average 10,000 laser shots/spectrum, using a high collision energy (1 kV). The polysaccharide ions were allowed to collide in the CID cell with argon at a pressure of $2x10^{-6}$ Torr.

Mild acid hydrolysis and HPAEC-PAD using Dionex PA20 for monosaccharide analysis

Samples were hydrolysed in 2M TFA for 1h at 120 ˚C. TFA was removed by evaporation under vacuum. Monosaccharide residues were then stored at -20°C prior to analysis by HPAEC-PAD. The stationary phase used was a PA20 column (Dionex, CA, USA) with eluent flow at a rate of 0.5 ml min^{-1}. The elution program consisted of an initial isocratic elution in 2 mM sodium hydroxide from 0 to 20 min, followed by a linear gradient up to 800 mM sodium hydroxide from 20 to 40 min, followed by an isocratic wash step in 200 mM sodium hydroxide from 40 to 50 min. The column was then equilibrated for 10 min in 2 mM sodium hydroxide prior to the next injection. A standard mixture of 20 µl of a mixture of fucose, rhamnose, arabinose, galactose, glucose, xylose, mannose, galacturonic acid, glucuronic acid was injected, with each sugar at a concentration of 25 µM.

Alkaline hydrolysis and HPAEC-PAD using Dionex PA200 for oligosaccharide analysis

Dry samples were transferred into 0.5 ml plastic screw cap tubes (Sarstedt, Germany) and hydrolysed in 600 µl of saturated barium hydroxide (0.22 M) for 18 h at 105°C in a heating block. The solution was then cooled on ice, and the pH adjusted to 7 by addition of ice cold 18 M sulphuric acid, producing a precipitate of barium sulphate. The barium sulphate precipitate was pelleted by centrifugation in a micro-centrifuge at 13, 000 *g* and the supernatant was dialysed against water using a dialysis membrane (500 Da cut-off) for 48 h. Samples were lyophilised and oligosaccharides were studied by HPAEC-PAD. Chromatography was on a PA200 column (Dionex, CA, USA) at a flow rate of 0.5 ml min^{-1}. The buffers used for the elution of oligosaccharides were 100 mM NaOH (Solvent A) and 100 mM NaOH with 1 M NaOAc (Solvent B). The elution program consisted of an isocratic wash with 100% Solvent A from 0 to 5 min, followed by a linear gradient up to 40% Solvent B from 5 to 60 min, followed by a wash step in 100% Solvent B from 60 to 85 min. The column was equilibrated for 20 min in 100% Solent A prior to the next injection.

Capillary Normal Phase (NP)-LC coupled off line to MALDI-ToF/ToF tandem mass spectrometry (MS/MS) for the structural elucidation of Hyp-polysaccharides.

Hyp-polysaccharides, dissolved in 80% acetonitrile, were loaded onto an amide-80 column (LC Packings/Dionex) and eluted with increasing aqueous concentration. The gradient generated flowed at 3 µl min^{-1} and Solvent A was 50 mM ammonium formate adjusted to pH 4.4 with formic acid, while Solvent B was 20% solvent A in acetonitrile. The NP column was equilibrated at 5% Solvent A and the gradient was initiated 5 min after injection and increased linearly to 70% Solvent A over 75 min, followed by a wash step in 90% Solvent A from 80 to 90 min. The column was equilibrated for 10 min with 5% Solvent A prior to the next injection. The column effluent was passed directly to a Probot sample fractionation system (LC Packings/Dionex), which spotted onto a MALDI target plate at 30 sec intervals. Accordingly, sample spots are overlaid with 60 nl of 2,5-DHB (in 50% methanol) and mass spectra were recorded on a 4700 Proteomics Analyzer (Applied Biosystems).

Phaseolin isolation.

Mature *Phaseolus vulgaris* cotyledons were homogenised in Tris/HCl pH 7.8 containing 1mM EDTA and 12% sucrose. The homogenate centrifuged for 90 minutes at 95,000xg and the supernatant passed in aliquots through a Sephadex G50 column equilibrated with 25mM Na-phosphate pH 7.0 containing 25mM NaCl. The cloudy excluded fractions from the column were loaded on a DEAE-sepharose column equilibrated with phosphate buffer.

Following extensive washing of the column the proteins of interest will be eluted with a salt gradient of 25-400mM in 25mM phosphate buffer, pH 7.0.The main proteins peaks (280nm) were combined dialysed against water overnight and lyophilized.

α-αmylase inhibitor isolation.

Cotyledons were soaked overnight in water and ground in water with a polytron. Proteins were left to extract, with stirring for two hours at room temperature in water (1:5 w/v). The extract was spun at 15,000xg for 40 minutes at 4 ^{o}C and pH of the supernatant brought to pH 4.0 by slow addition of 1*N* HCl, to precipitate proteins. The suspension was heated to 60^{o}C, cooled rapidly and spun again (as above). The supernatant was then neutralised to pH 7 with 2 *N* NaOH. The proteins of interest were fractionated by precipitation with 95% (v/v) ethanol to a final concentration of 40% (v/v), added over a period of 30 minutes. Following stirring for 3hrs at 4^{o}C the mixture was spun as previously and the precipitate discarded. The retained supernatant was adjusted to 80% (v/v) ethanol, stirred for a further 3hrs as before and centrifuged as above. The precipitate was dissolved in water and dialysed against water overnight at 4^{o}C and then lyophylised. The lyophilised material was dissolved in 0.02M potassium phosphate, 0.1M NaCl, pH 7.6 and applied to a pre-equilibrated DEAE column. Following washing of the column with the same buffer, the proteins of interest were eluted with 0.02M potassium phosphate, 0.15M NaCl, pH 7.6, and protein elution monitored at 280nm. Eluted fractions were pooled and dialysed against 0.02M sodium acetate pH 4.0 and applied to a CM-sepharose column equilibrated with the same buffer.

Proteins were eluted with a linear gradient of 0-0.4M NaCl in the same buffer. Eluted fraction were dialysed against water and lyophilised.

Isolation and Purification of Water-Extractable Arabinogalactan-Peptide from wheat

All steps were carried out at room temperature unless otherwise indicated.

200g of air-dry flour were transferred into stainless steel bowls and heated in a drying oven (130°C, 90 min) to inactivate endogenous enzymes. The flour was extracted with refluxing, with 80% ethanol for 30 minutes x2. The ethanolic extract was discarded and the ethanol-insoluble residue extracted with deionized water (1:5 w/v) for 60 minutes. The extract was then centrifuged (3,000x*g*, 15 min) and the aqueous supernatant heated to 90°C for 15 minutes to precipitate soluble proteins.

Residual starch was hydrolyzed by addition of 0.4 mL (7000 units) of heat-stable α-amylase solution, kept at 90°C for 30 minutes and then cooled to room temperature and centrifuged as above.

The mixture was then treatment with amyloglucosidase (120 units, 60°C, 12 h), cooled to room temperature and centrifuged as above.

The supernatant was then dialysed for 48 h, 4°C (against water).

The water extractable arabinoxylan (WE-AX) and water extractable arabinogalactan peptide (WE-AGP) were precipitated by stepwise addition of aliquots of ethanol (95%) to a final concentration of 80%, the mixtures stirred for 30 minutes and then kept at 4°C overnight. The mixture was then centrifuged at 10,000x*g* for 30 minutes 4°C.

The precipitates obtained were dissolved by shaking in 300 mL of deionized water for 60 minutes at room temperature.

To separate WE-AX from WE-AGP, ethanol was added to a final concentration of 65% (v/v) as above ie the mixtures were stirred for 30 min, kept at 4°C overnight and then centrifuged (10,000x*g*, 30 minutes, 4°C).

The supernatant containing the AGP's was evaporated at 50^{o}C (under reduced vacuum) and the residue dissolved in water.

Contaminating arabinoxylan (AX) was removed by incubation for 15hours at 40^{o}C with recombinant xylanase (a gift from Harry Gilbert).

AX hydrolysis products were removed by dialysis 24h, 4^{o}C and the xylanase treatment was then repeated again (i.e. 15hrs at 40^{o}C, followed by dialysis 24hrs, 4^{o}C)

Enzymes were deactivated by heating to 100^{o}c for 10minutes and AGP precipitated with 95% ethanol to a final concentration of 80% dissolving the residue in water and freeze drying.

Results and Discussion

Objective 01 (Cambridge)

Establishment of methods to routinely determine patterns of protein* N-*glycosylation in transgenic plants using protein mixtures

Task 01/a Test methods developed on mixtures of plant proteins.
Task 01/b Test method developed on a range of food crop species.

Evaluation of published methods for the mass spectrometric analysis of *N*-linked glycans

Initially, experiments were carried out on Arabidopsis membrane glycoproteins using methods for *N*-glycan release that use peptide *N*-glycosidase A. This is one of the most widely used enzymes for the analysis of plant glycans because it can cleave structures containing α-(1→3)-linked core fucose, which is a common substituent on plant glycoproteins. Following this, the less popular method of chemical cleavage called hydrazinolysis was also evaluated on glycoprotein standards and Arabidopsis membrane proteins. The most successful method from these evaluations was then validated using a related species (cauliflower, *Brassica oleracea* var. *botrytis*) from the Brassicaceae.

- **Enzymatic release of *N*-glycans using PNGase A**

Several methods, using different proteases prior to release of *N*-glycans by *N*-glycosidase A, were investigated on glycoproteins precipitated from Arabidopsis membranes isolated from callus tissue. Briefly, the precipitated protein/glycoprotein mixture was dissolved and separate aliquots were digested with different enzymes (pronase, trypsin or pepsin). These different proteases were evaluated in order to establish the best conditions for the efficient and comprehensive release of Arabidopsis oligosaccharides from the glycopeptides released by PNGase A. Released *N*-glycans were purified (from peptides and salt by C18 cartridge clean-up and cation exchange, respectively) and analysed by MALDI-ToF-MS either in their native or *per*-methylated form. Pronase, trypsin and pepsin are proteases with different specificities, hence it was anticipated that different glycan profiles might be obtained when used in combination with PNGase A. Our results indicate that pepsin treatment yielded a more complete glycan profile compared to trypsin or pronase treatment. Pepsin digestion allowed the detection of seven different glycan species as opposed to just two using trypsin and one using pronase (**Figure 1**).

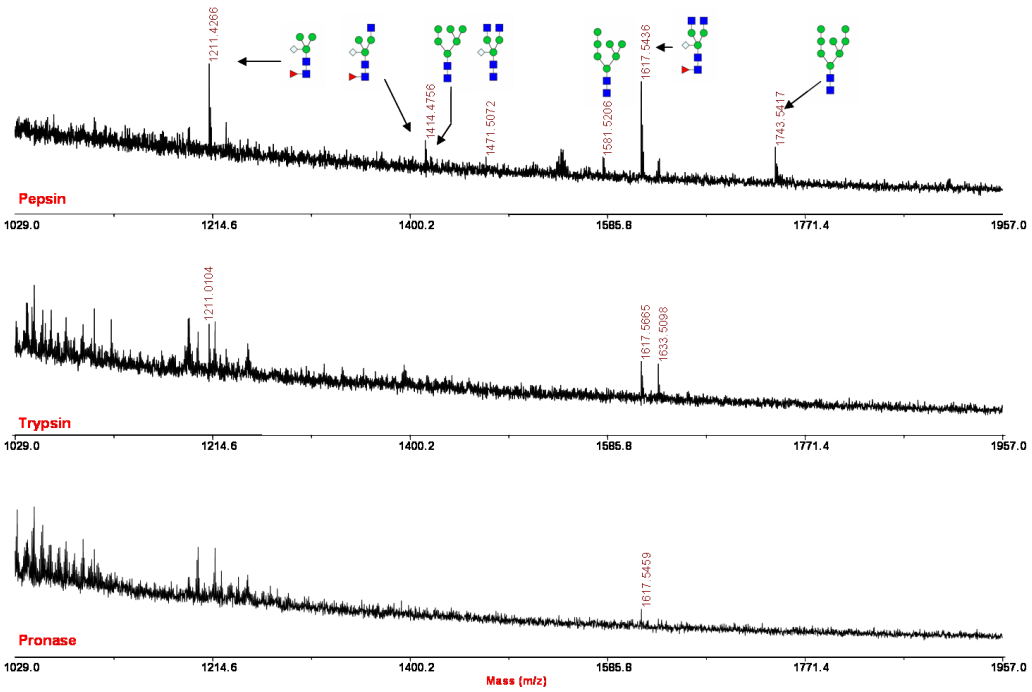

Figure 1. Comparison of spectra obtained from enzymatic treatment of *Arabidopsis* glycoproteins with pepsin, trypsin and pronase (glycans were analysed in their native form). Pepsin treatment yielded a more complete profile compared to trypsin or pronase treatment.

Per-methylation is a classical approach to increase the hydrophobicity of oligosaccharides to allow better detection with MALDI-ToF-MS (Ciucanu *et al.* 1984). With this method, all hydrogen atoms on the hydroxyl groups of the carbohydrates are replaced by methyl groups resulting in increased hydrophobicity. Derivatisation of *N*-glycans with this method yielded in increased signal strength and also allowed the detection of a greater range of *N*-glycans. Interestingly, this approach is shown to be very sensitive allowing the detection of three glycan structures novel for *Arabidopsis* (**Figure 2**).(Maslen et al. 2006)

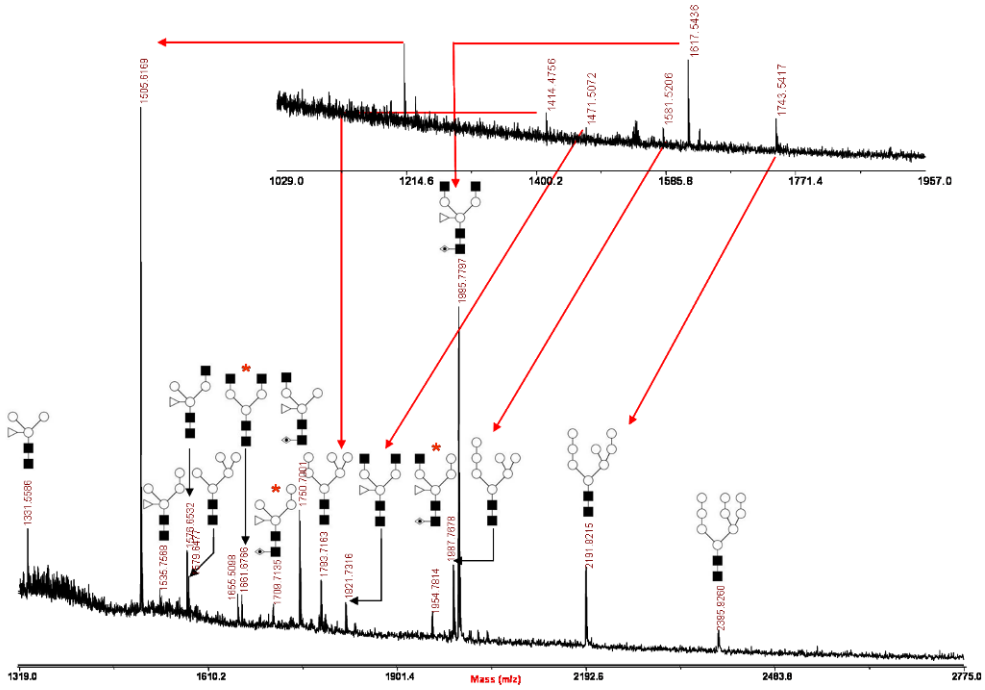

Figure 2. *Per*-methylated pepsin digested *Arabidopsis* samples compared to glycan profile of pepsin digested *Arabidopsis* samples in their native form (inset). *Per*-methylation allowed the detection of many more glycan species. In fact, with this approach all glycans reported for *Arabidopsis* were detected. In addition, three new structures never seen before in *Arabidopsis* (marked with red *) were detected.

- **Chemical release of *N*-glycans**

Chemical release of *N*-glycans is less widely employed than enzymatic release described above. The most common chemical procedure, called hydrazinolysis (Takasaki *et al.* 1982), was investigated to compare the glycans released using this and the enzymatic approach . The precise mechanism by which hydrazine cleaves glycans from glycoproteins is not completely understood, a proposed mechanism is shown in **Figure 3**. According to this model hydrazine cleaves the glycosylamide linkage and releases an intact glycan from the asparagine attachment site on the polypeptide chain. The terminal GlcNAc forms a covalent bond with hydrazine and as a side reaction all *N*-acetyl sites on all amino sugars are hydrolysed under these conditions. Therefore, after hydrazinolysis, re-*N*-acetylation of free amino sugars with acetic anhydride is needed in order to convert the amino sugars back to their original acetylated state. Because some GlcNAc residues might not lose hydrazine after re-*N*-acetylation, mild acid hydrolysis will achieve complete regeneration of the reducing end of glycans.

For the chemical release of *N*-glycans we used the Prozyme Glyko hydrazinolysis kit. The main limitations of hydrazinolysis are: a) hydrazine and acetic anhydride utilised by this method, are extremely toxic chemicals and the use of anhydrous hydrazine is forbidden in the UK; b) Anhydrous hydrazine is also very hydroscopic and thus all experiments should be conducted in very dry conditions; c) This method requires protein samples to be free from salts, metal ions and detergents which requires a lengthy dialysis step prior to analysis. The method is also time-consuming because long drying and incubation steps are required. d) Finally, it is possible that *O*-linked glycans may also be released.

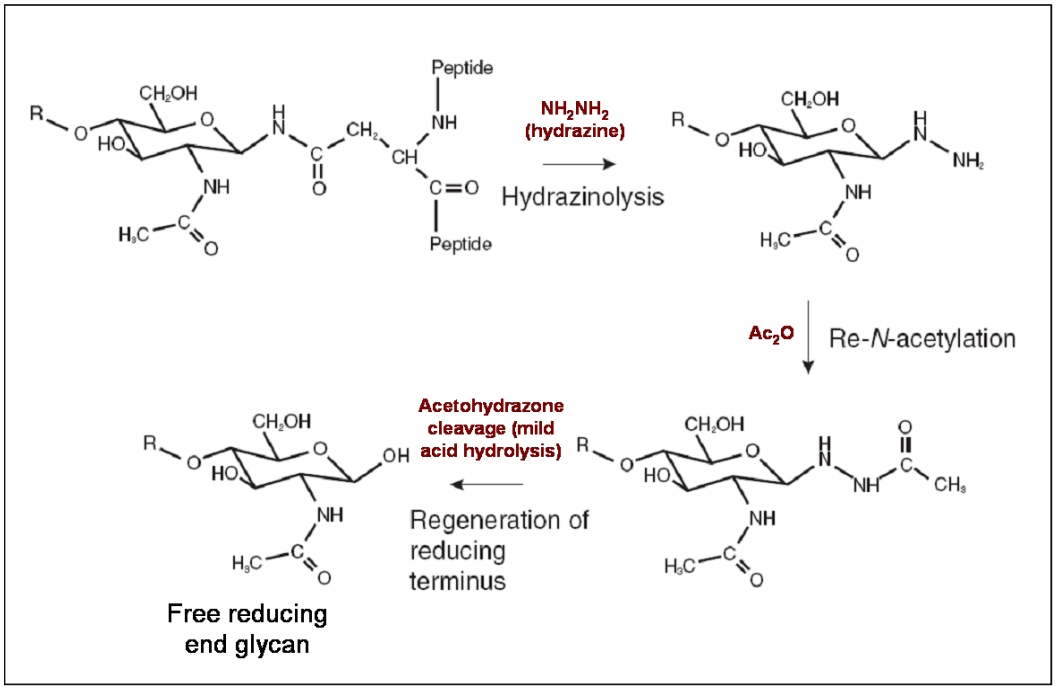

Figure 3. Proposed mechanism by which hydrazine cleaves glycans from glycoproteins.

Several attempts were made to chemically release the *N*-linked glycans from glycoproteins (including glycoprotein standards) via hydrazinolysis using the kit from Glyko. However, all attempts resulted in the absence of glycan signals in all MALDI-ToF-MS spectra. Consequently, it was concluded that the kit was not reliable and attempts to source anhydrous hydrazine from other UK suppliers failed. Consequently, the release of *N*-glycans from plant glycoproteins using the chemical and enzymatic approaches could not be directly compared. However, previous studies (using hydrazinolysis) on *N*-glycans from Arabidopsis glycoproteins purified from cell-culture (Fujiyama *et al.* 2007) showed the release of glycan structures that were easily identified using the pepsin/PNGase A method described above.

Evaluation of method on other *Brassicas*

In order to validate the enzymatic method (using pepsin/PNGase A) for the analysis of *N*-linked glycans, glycoproteins were isolated from edible parts of cauliflower and their *N*-glycan profile was obtained (**Figure 4**). With our approach we were able to detect all *N*-glycans reported for cauliflower (Wilson *et al.* 2001), suggesting that our method is very robust yielding reproducible results. In addition, we detected a new glycan structure never reported in cauliflower before highlighting the robustness and sensitivity of our approach, which allows the accurate analysis of *N*-glycan profiles from plant glycoproteins.

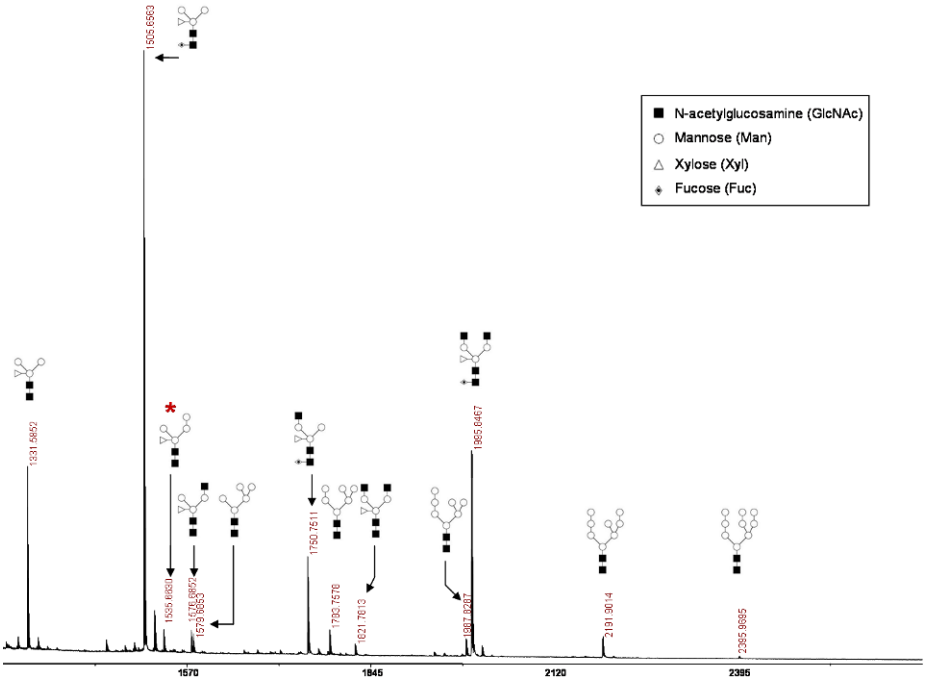

Figure 4. *N*-glycan profile (*per*-methylated form) obtained from cauliflower pepsin digested glycoproteins. All glycans reported for cauliflower were detected, while a new structure never seen in cauliflower before was also identified (marked with *).

Figure 5 shows a flowchart of our protocol for the mass spectrometric analysis of plant *N*-linked glycans. In addition, we further reduced the time required for the analysis of *N*-glycosylation by using immobilised pepsin. This enzyme is immobilised on beads which, at the completion of digestion, can be removed via simple centrifugation with the use of spin cartridges. Furthermore, the same buffer employed for this immobilised pepsin digestion can be used for PNGase A, eliminating the need for drying between the two different enzymatic digestion steps (**Figure 6**).

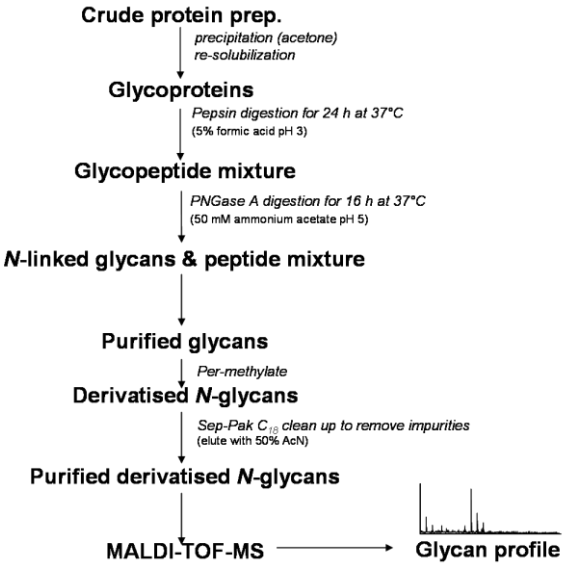

Figure 5. Protocol developed for the detailed and accurate analysis of plant *N*-linked glycans.

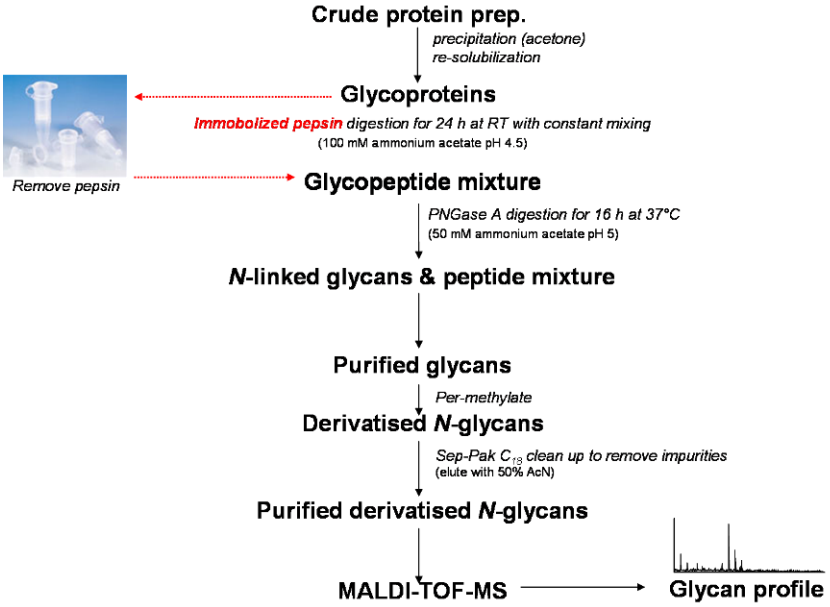

Figure 6. The use of immobilized pepsin and its removal after digestion with spin cartridges reduces the process time.

Milestone 01/01 Establish method for characterisation of *N*-linked glycans:

Deliverable 01/01 complete

Objective 02 (Cambridge)

Establishment of methods for the analysis of O-glycosylation in transgenic plants

GFP-glycan reporter constructs expressed in transgenic tobacco were used for initial method development because this material is available in large quantities and has comparatively small size glycans which have been relatively well characterised by NMR (Tan *et al*. 2004). The expression product of these GFP-glycan reporter constructs is an $(Ala\text{-}Hyp)_{51}$ repeat, in which all Hyp-residues are modified with an arabinogalactan polysaccharide (called AP51) (Tan *et al*. 2004). In order to produce glycans linked to hydroxyproline (Hyp) alone, the protein was first hydrolysed in saturated barium hydroxide. This cleaves the peptide backbone to give amino acids and releases the intact glycans, still attached to the Hyp residues. Preliminary experiments on this material included sugar analysis by HPAEC with pulsed amperometric detection. The HPEAC-PAD chromatogram gave peaks which corresponded to arabinose (Ara), galactose (Gal), rhamnose (Rha) and glucuronic acid (GlcA) residues (**Figure 7**), which is in accordance with the arabinogalactan structure reported by Tan *et al*. 2004.

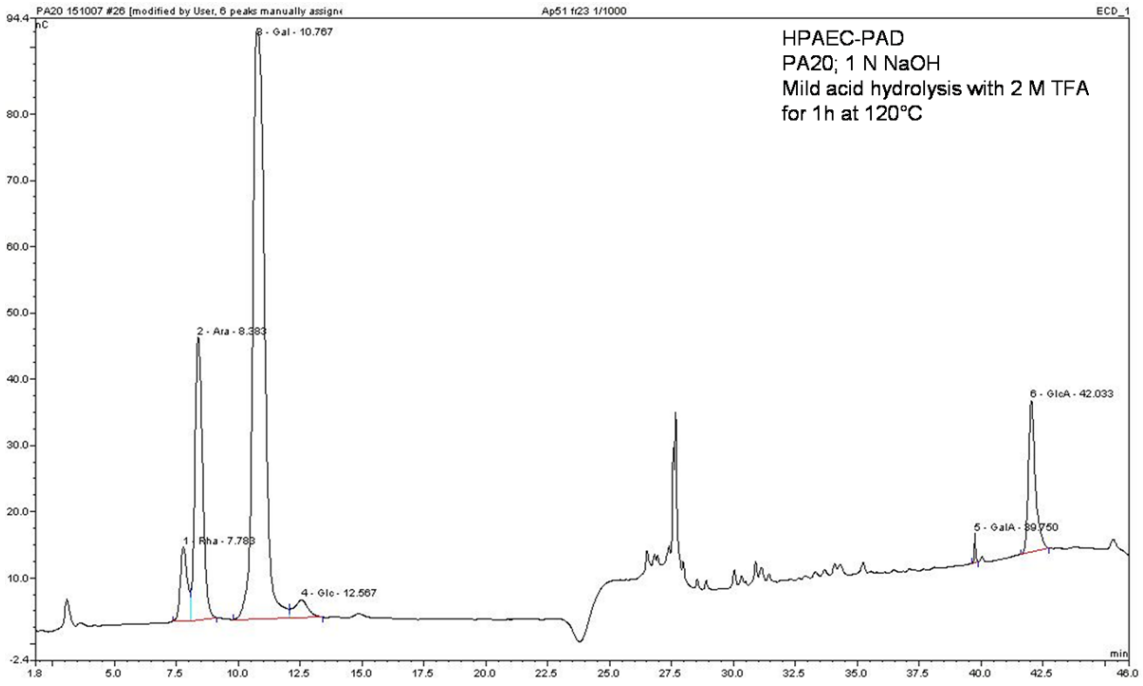

Figure 7. Sugar analysis of the AP51 polysaccharide by HPAEC-PAD using a Dionex PA20 column. AP51 contains rhamnose (Rha), arabinose (Ara), galactose (Gal) and glucuronic acid (GlcA) residues.

HPAEC-PAD was also used to profile the glycans present in the AP51 polysaccharide. However, a different column (PA 200) was employed that is capable of separating intact oligosaccharides based on their size, composition and charge properties (**Figure 8**). A broad peak is observed in the chromatogram but the sample was too heterogeneous to be resolved

by chromatography. Nevertheless, the elution position of these oligosaccharides indicates that the structures are likely to be larger and/or more highly negatively charged than the *N*-glycans from fetuin (major glycoprotein in fetal calf serum, (Green *et al.* 1988)) which elute much earlier (lower trace).

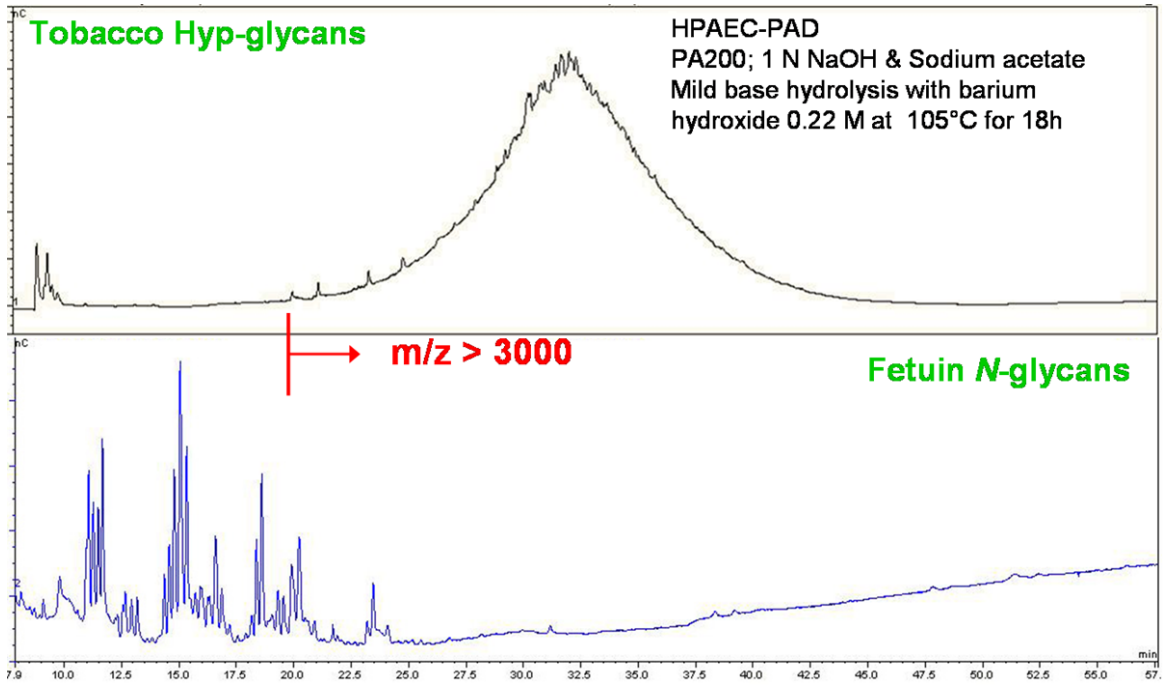

Figure 8: HPAEC-PAD analysis of AP51 polysaccharide using a Dionex PA200 column. These data indicate that the Hyp-linked glycans are extremely heterogeneous.

Initial mass spectrometric experiments involved the direct analysis of this complex mixture by MALDI-ToF-MS. Unfortunately, no signals for arabinogalactan oligosaccharides could be observed, even after attempts to desalt the sample with clean-up cartridges. This suggested that the mixture was too complex for direct analysis by MALDI-ToF-MS, and a pre-fractionation step was considered necessary. Consequently, the complex mixture of Hyp-polysaccharides was analysed by Normal Phase-Liquid Chromatography (NP-LC) in conjunction with off-line MALDI-ToF-MS. NP-LC was carried out using a capillary Dionex amide-80 column. Here, the hydrophilic AP51 Hyp-polysaccharides were retained on the column by polar interactions and were eluted with an increasing aqueous gradient (**Figure 9**). The effluent from the normal phase column was collected using an automated fractionation system (Probot, LC Packings/Dionex). Sample fractions were collected directly onto the MALDI target every 30 seconds between 30 and 98 minutes (see **Figure 9**). The sample spots were overlaid with 2,5-dihydroxybenzoic acid (DHB) matrix and were analysed off-line via

MALDI-ToF-MS. Several MALDI spectra contained signals that differed by masses corresponding to sugar residues. A representative MALDI spectrum acquired from one such fraction is shown in **Figure 10**, where many of the signals could be assigned to Hyp-linked oligosaccharides. These data show that the AP51 Hyp-linked glycans from tobacco are extremely heterogeneous and pre-fractionation by normal-phase chromatography is necessary for them to be observed by mass spectrometry.

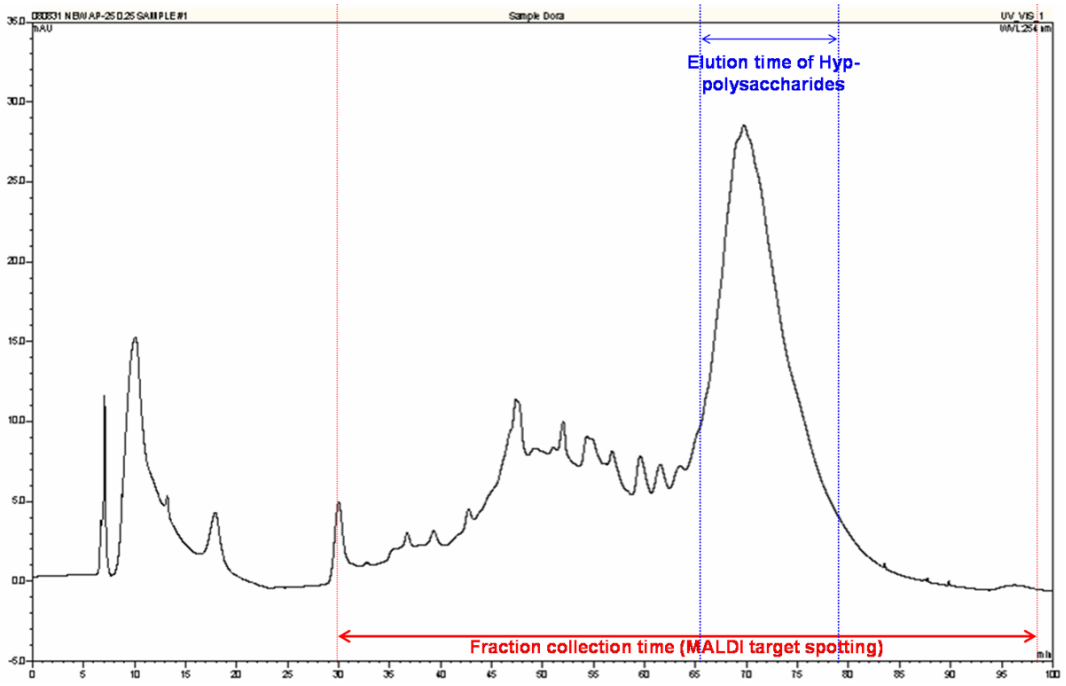

Figure 9: UV chromatogram (at 254 nm) obtained by Normal Phase-Liquid Chromatography (NP-LC) of AP51 polysaccharide. The chromatogram shows the elution time of Hyp-polysaccharides.

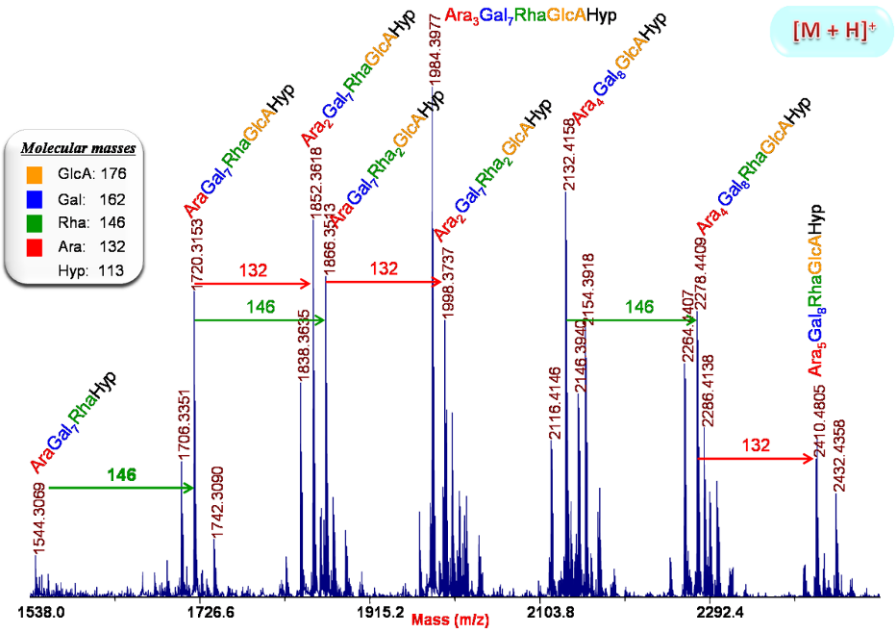

Figure 10: MALDI mass spectrum showing signals $[M+H]^+$ for native glycans from AP51 sample. The mass differences between peaks correspond to sugar masses (arabinose: 132 m/z, galactose: 162 m/z, rhamnose: 146 m/z and glucuronic acid: 176 m/z). Possible Hyp-glycan compositions are shown.

Additional exoglycosidase digestions were carried out in order to facilitate the detailed characterisation of the Hyp-linked glycan 'core'. The set of AGP specific hydrolyzing enzymes which were available to us was as follows:

- Native exo-β-1,3-galactanase (Tsumuraya et al. 1990): specifically cleaves β-1,3-galactosidic linkages and can bypass branching points, hence liberating β-1,6-linked galactosyl side chains of branched AGPs as various oligomers, ranging in degree of polymerisation from 1 up to at least 20.
- Recombinant endo-β-1,6-galactanase (Kotake et al. 2004): this enzyme hydrolyzes β-1,6-galacto oligomers with a degree of polymerization higher than 3, producing galactose and galactobiose fragments. It requires the previous action of arabino*furano*sidase.
- Recombinant β-glucuronidase (Haque et al. 2005): it acts upon Me-β-glucuronic acid or β-glucuronic acid residues in long β-1,6-linked galactosyl side chains of the AGPs.
- Recombinant α-L-arabino*furano*sidase (Hata et al. 1992): attacks both α(1→3) and α(1→5) arabino*furano*syl residues but the latter is much less susceptible to the enzyme.

Subsequent digestion of AP51 Hyp-polysaccharides with arabino*furano*sidase and analysis by Normal Phase-Liquid Chromatography (NP-LC) in conjunction with off-line MALDI-ToF-MS (as previously described) revealed that the enzyme not only can act upon the polysaccharide but also that the heterogeneity of the polysaccharide is considerably reduced (**Figure 11**).

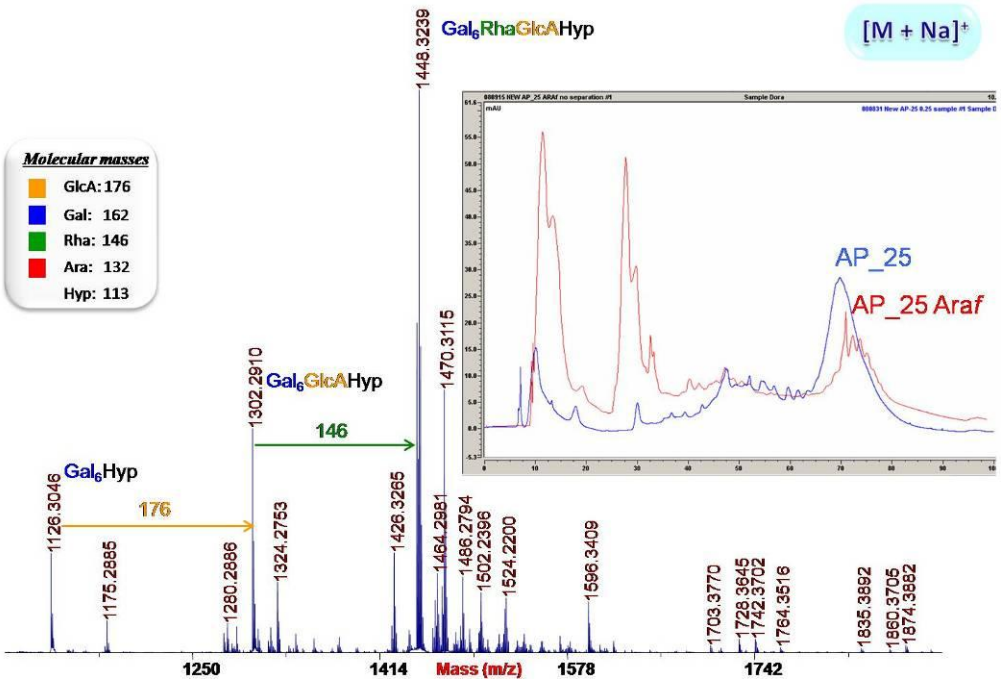

Figure 11: MALDI mass spectrum showing signals $[M+Na]^+$ for native arabino*furano*sidase digested glycans from AP51 sample. The mass differences between peaks correspond to sugar masses (rhamnose: 146 m/z and glucuronic acid: 176 m/z). Possible Hyp-glycan compositions are shown. The inset is showing the elution times for the crude AP51 AGP sample (blue) and for the arabino*furano*sidase digested AP51 AGP sample (red).

Attempts to further characterise these structures by MS/MS however, were not successful as signals from overlapping structural isomers and/or from neighbouring oligomer structures did not allow a reliable interpretation of the spectra (data not shown). In conclusion, the AP51 polysaccharide from transgenic tobacco was a rather complex mixture of glycans which were much larger in size than it was originally anticipated. However those preliminary experiments showed that our methodology can be used successfully for the detection of *O*-linked glycans. The wheat arabinogalactan peptide (AGP) was therefore selected to evaluate the methodology for the analysis of *O*-glycosylation in transgenic plants, due to its small peptide size and the relatively small complexity of the polysaccharide (Tan *et al*. 2003).

Deliverable 01/02 complete.

Evaluation of method on wheat arabinogalactan peptide (AGP)

The non-starch polysaccharides of wheat flour, consist mainly of arabinoxylans (AX) and water extractable arabinogalactan (WE-AG) peptides (Mares *et al.* 1973), and influence bread making properties of the wheat (Courtin *et al.* 2002). There is very little structural variation

of WE-AGP of flour from different wheat varieties or within one wheat variety (Loosveld *et al.* 1998). The major wheat endosperm AGP has been estimated to be present at a level of 0.27%-0.38% of the flour dry weight. It has an average mass of 22 kDa (Fincher *et al.* 2004), and is typically composed of 15 amino acids with Hyp (hydroxyproline), Ala (alanine), Ser (serine) and Thr (threonine) the most abundant. The polysaccharide portion of the wheat AGP constitutes over 92% of the molecular mass. The first reports about the structure of the wheat flour AGP polysaccharide date back to 1974 and are based on NMR spectrometry and methylation analyses.(Fincher *et al.* 1974) According to those analyses, the polysaccharide consists of a β-(1→6)-galactopyranosyl backbone, substituted with single α-L-arabinofuranosyl residues at C(O)3 position of the β-(1→6)-galactose backbone, and no terminal galactose was found. Later NMR and methylation studies elaborated this model by suggesting that the β-(1→6)-galactopyranosyl backbone can also be substituted with a single β-galactopyranosyl residue, which is in turn substituted in its C(O)3 position with a single α-L-arabinofuranosyl residue (Van den Bulck *et al.* 2005) (**Figure 12**).

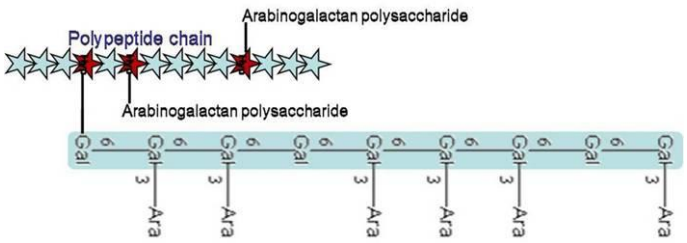

Figure 12: Structure of wheat arabinogalactan peptide based on data of Fincher GB, Sawer WH and Stone BA (1974).

However, since the most frequently reported structure for AG consists of a β-(1→3)-galactopyranosyl backbone with β-(1→6)-galactopyranosyl side chains, the proposed structure of wheat AGP is unusual. Moreover, Tan *et al.* (2004) have reported a structure of the arabinogalactan from a repetitive Ala-Hyp expressed in tobacco, which consisted of a Hyp-linked (1→3)- β-galactan pentasaccharide backbone with a single 1,6-linked β-galactose "kink". This backbone had single 1,6-linked β-galactosyl side chains substituted at position C(O)3 with arabinofuranosyl residues and at C(O)6 with glucuronic acid or rhamnosyl glucuronic acid (Tan *et al.* 2004).

One-dimensional and two-dimensional ^{1}H and ^{13}C NMR spectroscopy to characterise arabinogalactan proteins from stigmas and styles of *Nicotiana alata* described a highly

branched backbone of 3-, 6- and 3,6-linked galactopyranose residues with terminal galactopyranosyl and arabinofuranosyl residues (Gane *et al.* 1995). In a similar manner, Haque *et al.* (2005) proposed a structure for radish root arabinogalactan polysaccharide, according to which the backbone consists of (1→3)-linked β-galactosyl residues with (1→6)-linked β-galactose side chains (Haque *et al.* 2005). According to that model the side chains may be further substituted at position C(O)3 with arabinose residues and can be carrying glucuronic acid or 4-Me-glucuronic acid at the non-reducing terminals linked at C(O)6 position.Based on those reports we believe that it is more likely that the wheat arabinogalactan polysaccharide consists of β-1,3-galactose backbone with β-1,6-galactose side chains which are further modified with other auxiliary sugars. Monosaccharide analysis of crude wheat AGP revealed that arabinose, galactose and glucuronic are the main sugars components of the polysaccharide (data not shown). This has been the first time that the presence of glucuronic acid on wheat arabinogalactan polysaccharide is reported. In addition, the arabinose to galactose ratio was 0.5 which is in accordance with other reports regarding wheat grain arabinogalactan polysaccharides (Van den Bulck *et al.* 2005). In order to test our hypothesis, crude wheat arabinogalactan protein was digested with exo-β-1,3-galactanase (the enzyme was a kind donation from Prof. Yoichi Tsumuraya, Japan). Since the enzyme specifically cleaves β-1,3-galactosidic linkages and can bypass branching points, it will liberate any β-1,6-linked galactosyl side chains of branched wheat AGPs as various oligomers. Should the Fincher *et al*. model be correct, no side chain fragments should be detected.

Our method briefly was as follows: 1 mg of crude wheat glycoprotein mixture was digested with 6mU of AGP specific exo-β-1,3-galactanase for 24h and 37˚C in 20mM ammonium acetate buffer pH 4.6. Any released β-1,6-galactosyl residues were purified (from peptide and salt by C18 cartridge clean-up and cation exchange, respectively) and analysed by MALDI-ToF-MS in their *per*-methylated form (see appendix for detailed methods). Part of the digested sample was also subjected to PACE (Polysaccharide Analysis using Carbohydrate gel Electrophoresis) (Goubet et al. 2002), a method developed in our laboratory for the detailed analysis of sugar oligomers. **Figure 13a** shows that wheat arabinogalactan polysaccharide consists of a β-1,3-galactose backbone which is susceptible to exo-β-1,3-galactanase. The degree of polymerisation (DP) of the released β-1,6-linked galactose side chains ranges from 1 to at least 11. In order to evaluate the effect of previous arabino*furano*sidase digestion on exo-β-1,3-galactanase treated wheat AGP, digestion with 6

mU of arabino*furano*sidase in 50 mM ammonium acetate buffer pH 5.0, preceded the exo-β-1,3-galactanase digestion and the released oligosaccharide fragments were analysed by PACE (**Figure 13a**). Our results demonstrate clearly that pre-treatment with arabino*furano*sidase enhances the exo-β-(1→3) galactanase activity. The DP of the released fragments is now ranging from 1 up to at least 21. Aliquots of the enzyme digested samples were also *per*-methylated and analysed by MALDI-ToF-MS (**Figure 13b and c**). Our data suggests that the side chains of wheat AGP consist predominantly of galactose and arabinose (**Figure 13b**) and also that, arabino*furano*sidase is capable of removing a large portion of auxiliary arabinose residues (**Figure 13c**). In addition, we were able to detect MS peaks corresponding to oligosaccharides carrying glucuronic acid residues (m/z 691.5 and 855.6).

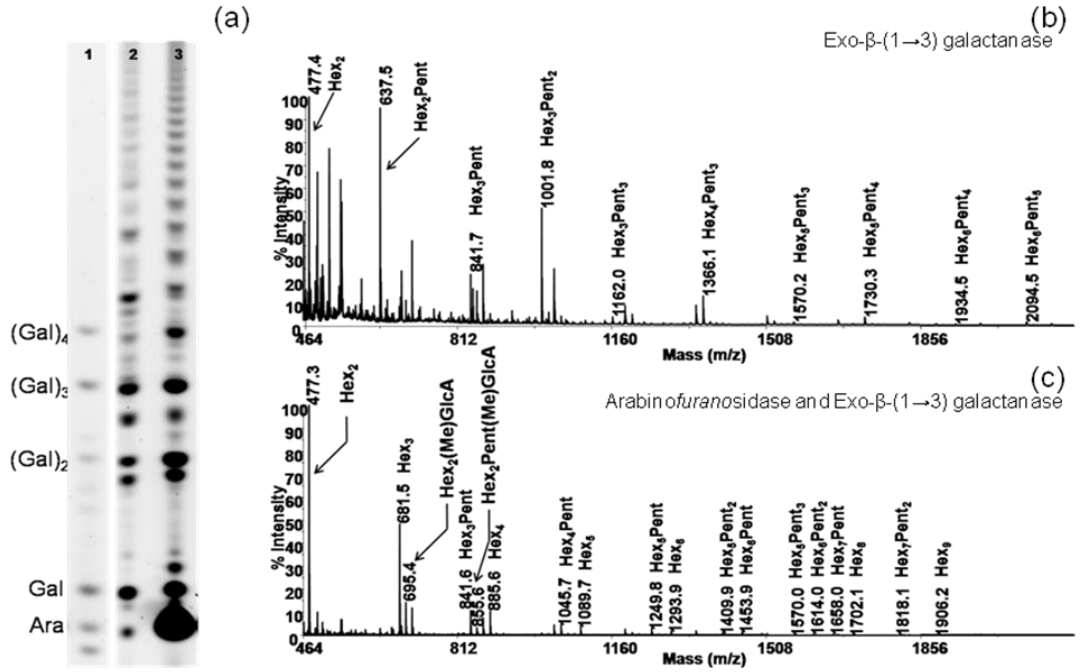

Figure 13: (a) Oligosaccharides released by AGP specific enzymes were derivatised with 8-aminonaphthalene-1,3,6-trisulphuric acid and separated by polyacrylamide gel electrophoresis (PACE): 1, arabinose and galactan standards; 2, Cadenza wheat AGP digested with exo-β-(1→3) galactanase; 3, Cadenza wheat AGP treated with arabino*furano*sidase followed by exo-β-(1→3) galactanase digestion. (b) and (c) MALDI-ToF MS spectra showing signals of the *per*-methylated oligosaccharides released by exo-β-(1→3) galactanase or arabino*furano*sidase and exo-β-(1→3) galactanase, respectively. Peak corresponding to m/z 1001.8 on spectrum (b) was selected for MS/MS analysis.

Consequent MS/MS analysis of the oligosaccharide fragments released by exo-β-(1→3) galactanase demonstrated that these fragments consist of 1,6-linked galactose chains which are modified by single arabinose residues at position C3 (**Figure 14**). In order to verify the

presence of glucuronic acid on the released oligosaccharide fragments, the samples were subjected to glucuronidase digestion followed by PACE and MALDI-ToF-MS analysis as described above. PACE analysis shows an alteration on the band pattern after glucuronidase digestion, indicative of the presence of glucuronic acid in the sample (**Figure 15a**). The MALDI-ToF-MS spectra verified those findings (data not shown). In addition, further structural characterisation by MS/MS showed that glucuronic acid residues are linked on the C(O)6 position of the 1,6-linked galactose side chains, quite possibly on their nonreducing ends (**Figure 16b**). In order to establish whether glucuronic acid was present in its methylated form, we performed deutero-methylation, a classical approach to distinguish between methylated and non methylated forms of uronic acids. This approach revealed that no methylated glucuronic acid residues were present in the sample (data not shown).

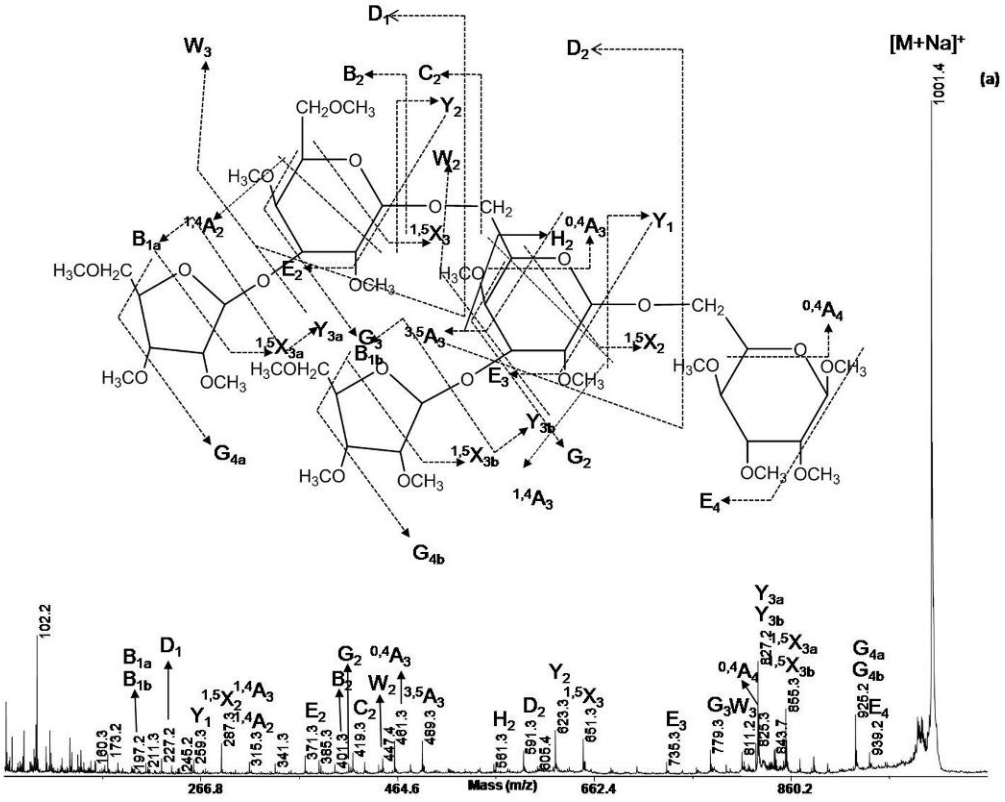

Figure 14: MALDI-CID of Hex_3Pent_2 oligosaccharide released by exo-β-(1→3) galactanase of Cadenza wheat arabinogalactan glycoprotein. Glycosidic and cross-rind fragments are identified according to Domon and Costello nomenclature (Domon et al. 1988). Hexose (Gal) residues are 1→6 linked while single pentose (Ara) residues are linked to the C3 position of galactose.

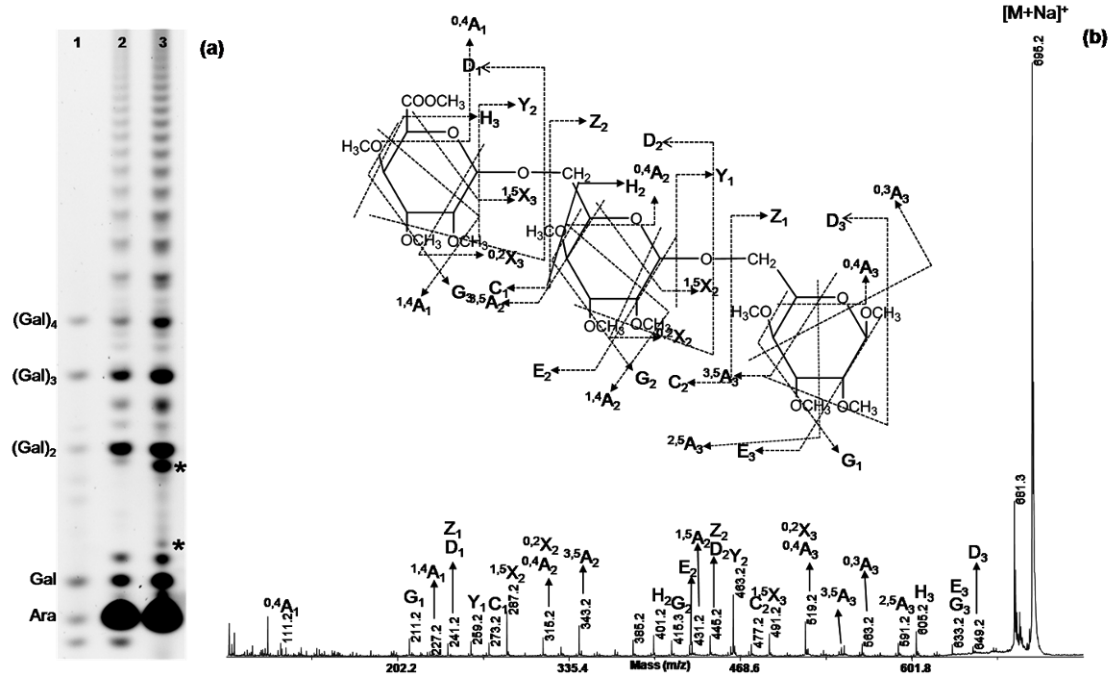

Figure 15: (a) Glucuronidase digestion is altering PACE profile suggesting that glucuronic acid is present on Cadenza wheat arabinogalactan polysaccharide. Oligosaccharides released by AGP specific enzymes were derivatised with 8-aminonaphthalene-1,3,6-trisulphuric acid and separated by polyacrylamide gel electrophoresis: 1, Arabinose and galactan standards; 2, Cadenza wheat AGP digested with arabinofuranosidase, exo-β-(1→3) galactanase and glucuronidase; 3, combined arabino*furano*sidase (Ara*f*) and exo-β-(1→3) galactanase digestion. Asterisk (*) shows oligosaccharides sensitive to glucuronidase. (b) MALDI-CID of Gal_2GlcA oligosaccharide released by combined digestion of arabino*furano*sidase and exo-β-(1→3) galactanase of Cadenza wheat arabinogalactan glycoprotein, reveals that the glucuronic acid residue is linked to the C6 position of galactose. Glycosidic and cross-ring fragments are identified according to Domon and Costello nomenclature (Domon et al. 1988).

Further combinations of exoglycosidase digestions followed by PACE and MALDI-ToF-MS analyses provided a good insight into the wheat arabinogalactan polysaccharide structure. In particular, our approach for the analysis of *O*-linked glycosylation in plants allowed not only the detection of *O*-linked glycans but also enabled us to propose a model for the structure of those complex molecules in wheat AGP. Specifically, our data suggests that wheat AGP consists of a short β-1,3-linked galactose backbone which is substituted at C(O)6 positions with long β-1,6-linked galactose side chains. These side chains are in turn modified by 1,3-linked arabinose residues, while a few short galactose side chains are carrying at the C(O)6 position, quite possibly on their non-reducing end, glucuronic acid residues. These findings are in accordance with the Haque *et al.* proposed structure for the radish root AG polysaccharide. In conclusion, our approach is quite sensitive allowing the analysis of *O*-linked glycan profiles from plant glycoproteins.

The overall aim of this project was to develop methods to identify and characterise *N*- and *O*-linked glycans on plant proteins, in order to determine whether transgenically expressed proteins are inauthentically glycosylated. **Figure 16** shows a flowchart of our protocol for the mass spectrometric analysis of plant *O*-linked glycans. According to this, with simple monosaccharide analysis, possible compositional differences (both qualitative as well as quantitative) on transgenic plants could be detected. Accordingly, treatment of the glycans with exo-β-1,3-galactanase and consecutive analysis of the *per*-methylated sugars with MALDI-ToF-MS will allow the identification of major structural modifications. In the future, PACE could also be used as a fingerprinting method to reveal structural and/or quantitative differences.

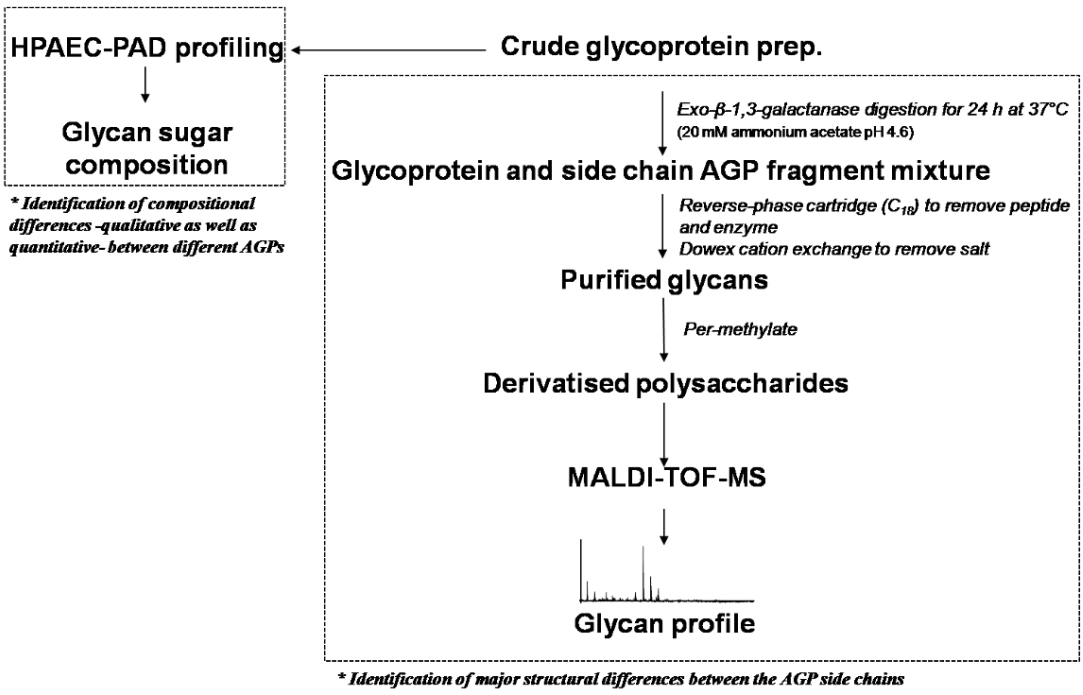

Figure 16. Protocol developed for the analysis of *O*-linked glycans.

Deliverable 04/01 complete.

Objective O3 (Rothamsted)

Evaluation of method for N-linked glycan analysis on "native" and recombinant proteins

Methods developed at Cambridge to be evaluated on two types of material:-

A. Non-GM

The major seed storage protein of *Phaseolus* is a 7S globulin called phaseolin (Hall et al, 1999). This protein contains two putative *N*-glycosylation sites, Asn^{252} and Asn^{341} (Sturm et al, 1987a) but only one of these may be glycosylated (Bollini et al, 1983,; Sturm et al, 1987a, Sturm et al, 1987b,). Furthermore, the structure of the glycan chains has been determined, showing that molecules have either Man_7 $(GlcNA_c)_2$ attached to Asn^{252} and Man_9 $(GlcNA_c)_2$ attached to Asn^{341}, or xylose-Man_3 $(GlcNA_c)_2$ attached to Asn^{252} with no glycosylation of Asn^{341} (Sturm et al, 1987b).

This system therefore provides an ideal model to test new methods which will allow the determination of variation in glycan structure between individual protein chains and between individual plants and cultivars.

Phaseolin was isolated according to the method of Sturm et al, 1987b (see Appendix L for a more detailed description of the isolation method). The chromatogram from the G50 size exclusion stage of purification is shown in **Fig. 17** and the protein of interest arrowed, in this case from *Phaseolin vulgaris* cv Masterpiece. Following size exclusion the indicated fractions were applied to a DEAE ion exchange column and phaseolin eluted with a linear NaCl gradient (**Fig.17**)

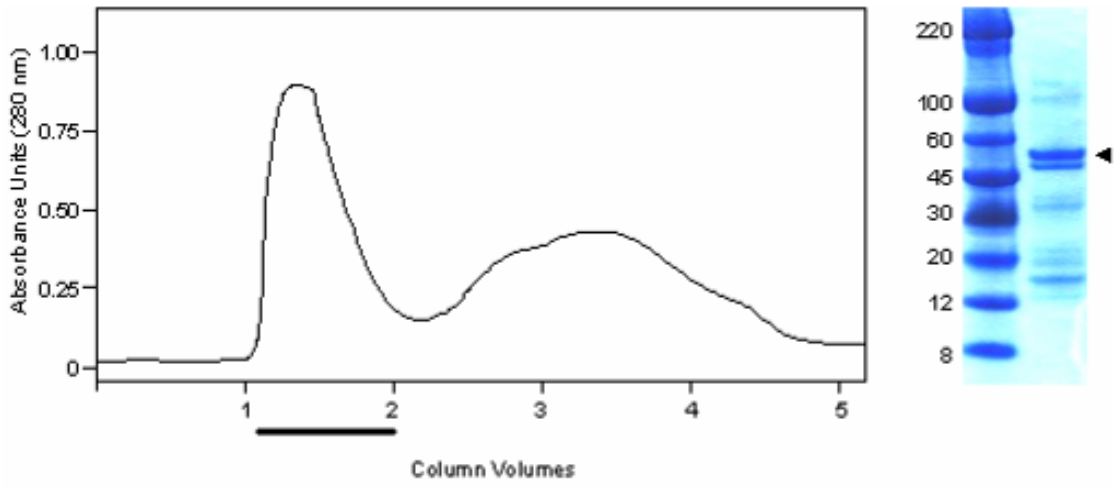

Figure 17: Purification of phaseolin. (a) Size exclusion chromatography of crude bean extract (*Phaseolus vulgaris* cv. Masterpiece) on a G50 Sephadex column. The excluded volume indicated by the black bar was subsequently loaded onto a DEAE cellulose column (see Figure 18). (b) SDS-PAGE of excluded fraction. Phaseolin is indicated by black arrow.

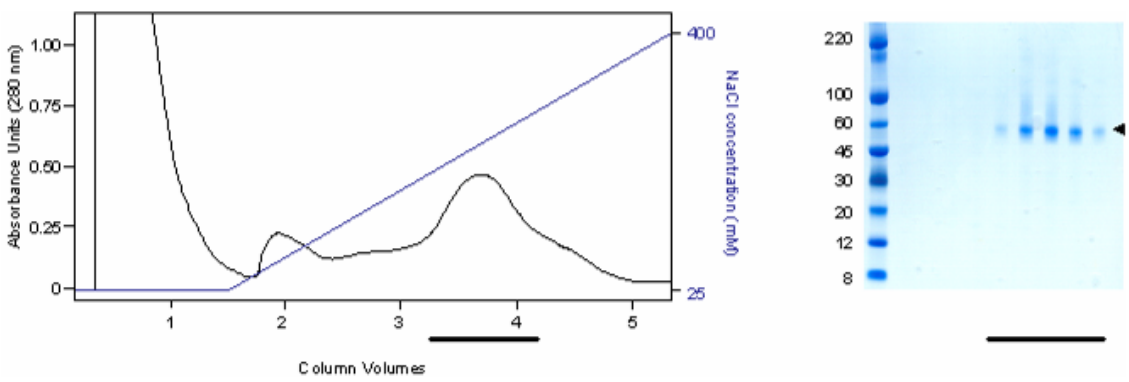

Figure 18: Purification of phaseolin. (a) DEAE chromatography of bean protein from G50 column. The black bar indicates the fractions run on SDS-PAGE. (b) SDS-PAGE of DEAE column fractions. Phaseolin is indicated by black arrow.

The purified phaseolin was then lyophilised and then analysed using the methodology developed at Cambridge for *N*-glycan isolation and purification (Appendix A-H)

Methods developed in Objective 01 for analysis of *N*-glycosylation were applied to the purified phaseolin protein isolated from individual plants and cultivars of *Phaseolus vulgaris* and the mass spectra for the *N*-glycans isolated are shown in the following figures **(Fig 19)**

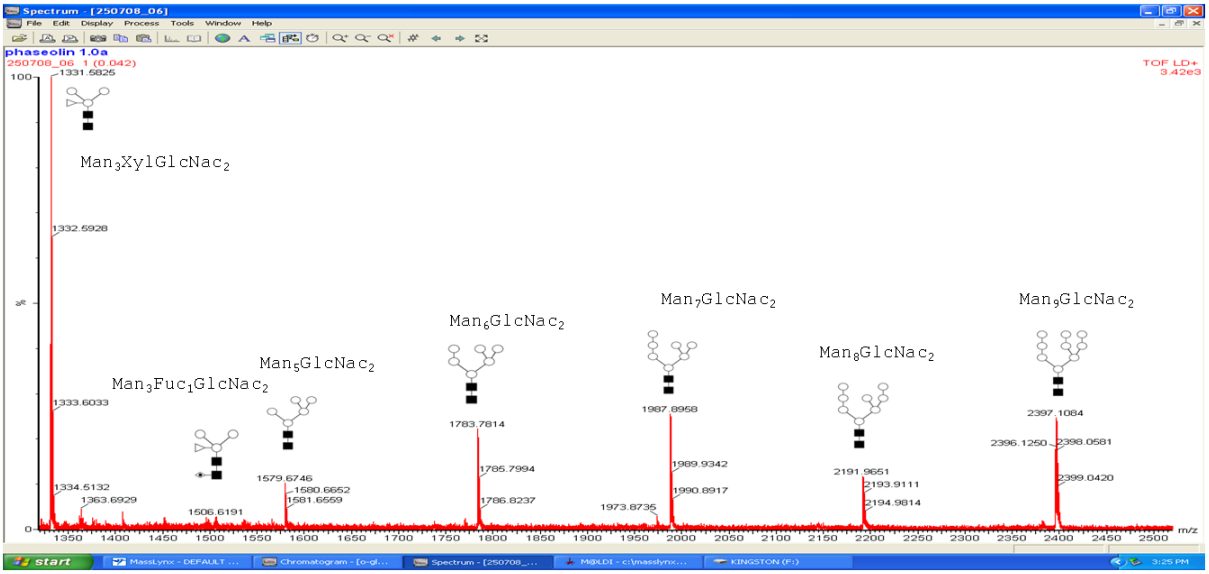

Figure 19. MALDI mass spectrum of glycans from phaseolin isolated from *Phaseolus vulgaris* cv. Tendergreen. Glycoproteins were digested with pepsin and *per*-methylated, as described in the methods developed at Cambridge.

The original 1987 paper of Sturm *et al.*, identified $Man_3XylGlcNac_2$, $Man_7GlcNac_2$ and $Man_9GlcNac_2$. Thus the identification of Man 5, 6 and 8 demonstrate the increased sensitivity of the methods developed at Cambridge.

Additionally, as a positive control used throughout, a cauliflower extract was isolated at Rothamsted **Fig 20**) according to the methods developed in Cambridge and treated in parallel with all subsequent samples, as a positive control, to check that the *N*-glycan isolation and purification steps were proceeding as expected.

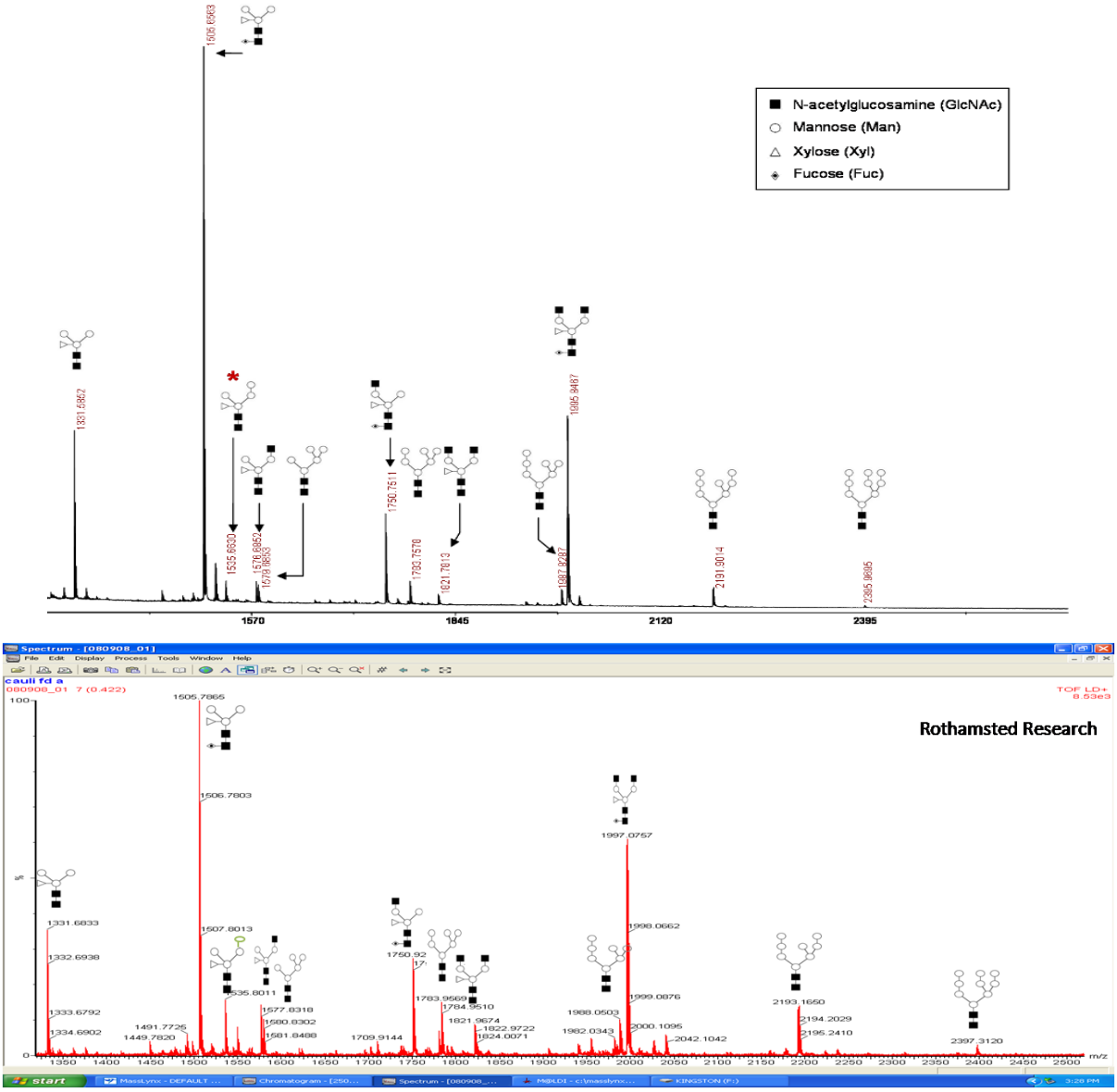

Figure 20: Comparison of glycan spectra obtained from cauliflower samples isolated at Cambridge and Rothamsted.

It can be clearly seen that all the glycans identified in the cauliflower preparation by Cambridge, were all isolated and assigned in the cauliflower preparation made at Rothamsted. Again, this indicates the robustness of this methodology.

Two different cultivars of *Phaseolus vulgaris* beans were also analysed using the methods established at Cambridge cv. Masterpiece (a dwarf bean) and cv. Blue Lake (a climbing bean). As with the cv. Tendergreen initially analysed, $Man_3XylGlcNac_2$, $Man_5GlcNac_2$, $Man_6GlcNac_2$, $Man_7GlcNac_2$, $Man_8GlcNac_2$ and $Man_9GlcNac_2$ were identified **(Fig 21).**

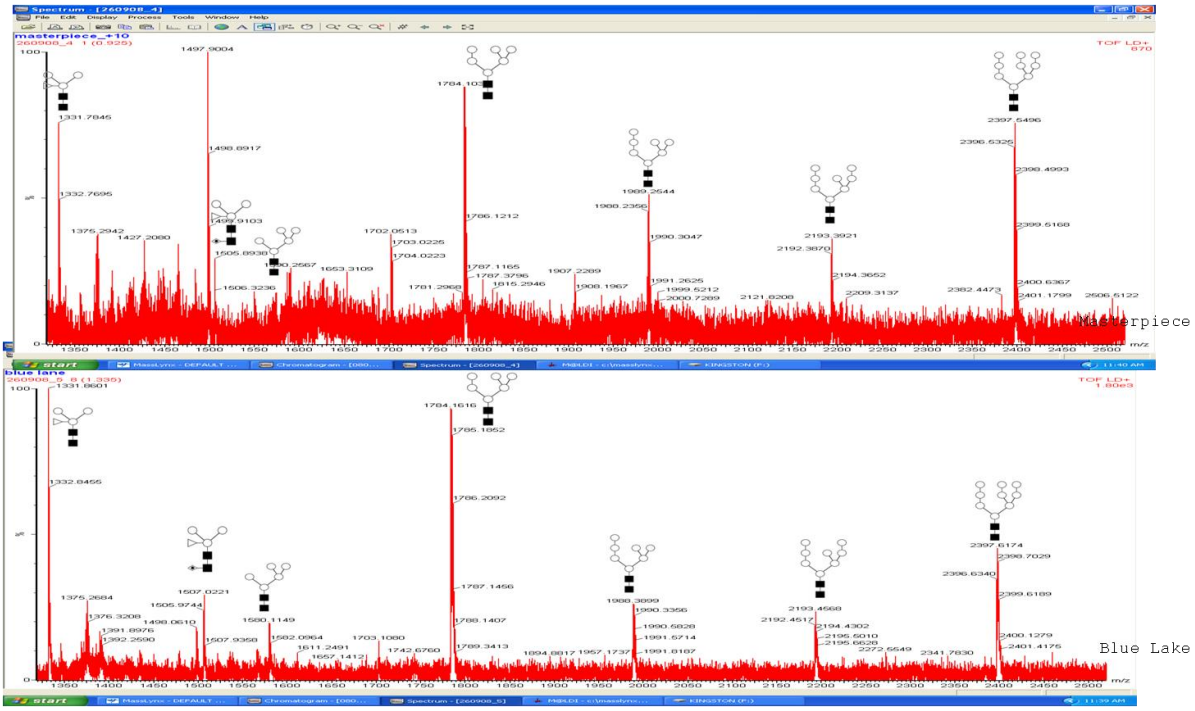

Figure 21: Comparison of glycan spectra of phaseolin isolated from *Phaseolus vulgaris* cultivars Masterpiece and Blue Lake. Glycoproteins were digested with pepsin and *per*-methylated, as described in the methods developed at Cambridge.

Based upon the fresh weight of starting material it is calculated that the spectra obtained were equivalent to glycans from less than 0.3 of a single seed.

Phaseolin was isolated from single seeds of three cultivars of *Phaseolus vulgaris* (Tendergreen, Masterpiece and Blue Lake) to compare the *N*-glycans patterns in a single seed. This showed that the methods developed can d be used at the single seed level as all the glycans identified in the larger preparations are observed (Figure 22a-c). In general, the glycan spectra at the single seed level, were similar for each of the cultivars but it should be noted to obtain spectra of single seeds that did not contain some hexose polymer units, which were not removed by the clean-up steps in the method was difficult. These hexose polymers make the spectra more complicated, although they are easily distinguished (by mass) from the

glycans of interest The hexose polymers were not as clear in the larger-scale preparations probably because the signal to noise ratio was much greater.

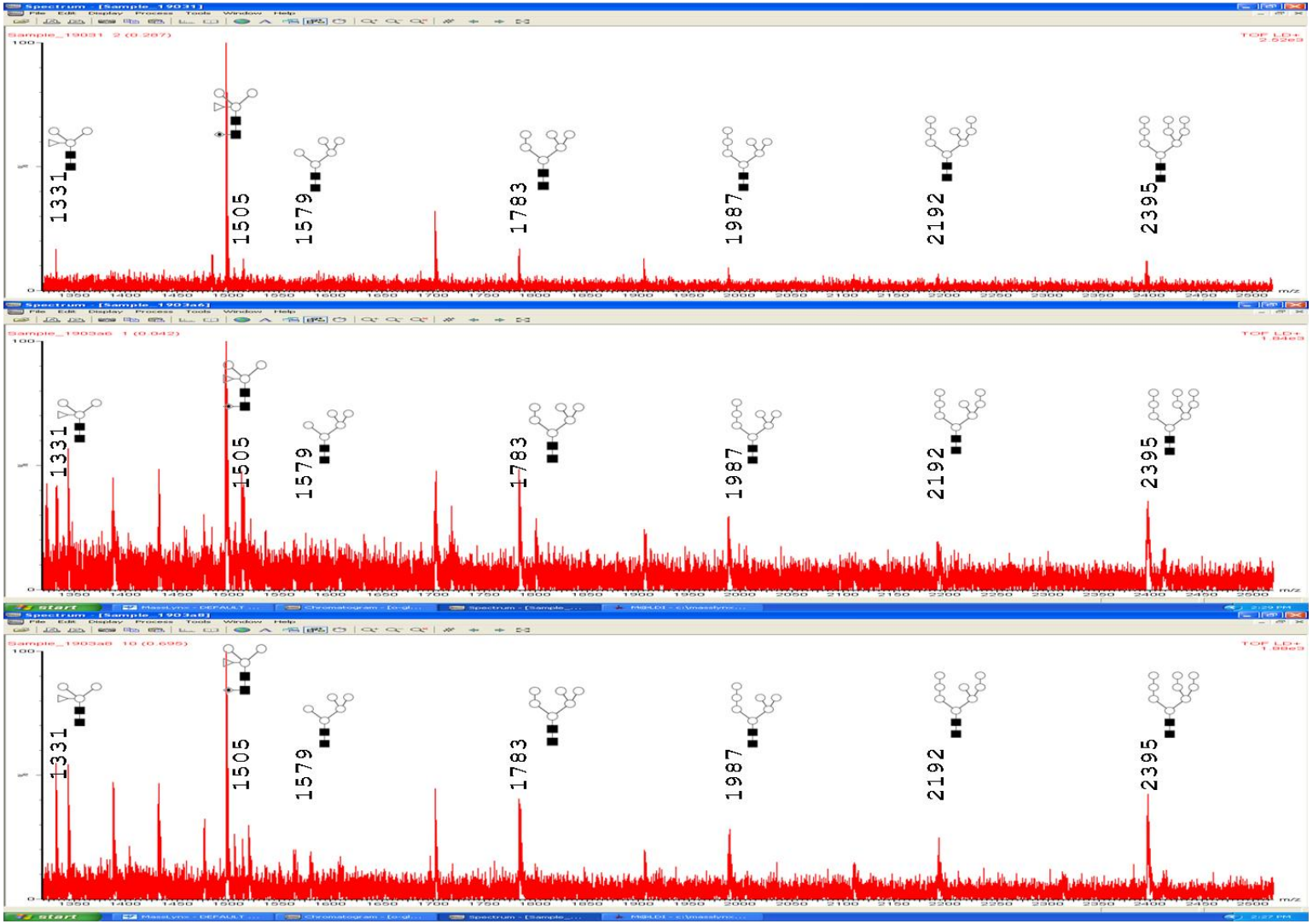

Figure 22a ***N*-glycan spectra of phaseolin extracted from single seeds (1, 2 and 3) from *Phaseolus vulgaris*, cv. Tendergreen.**

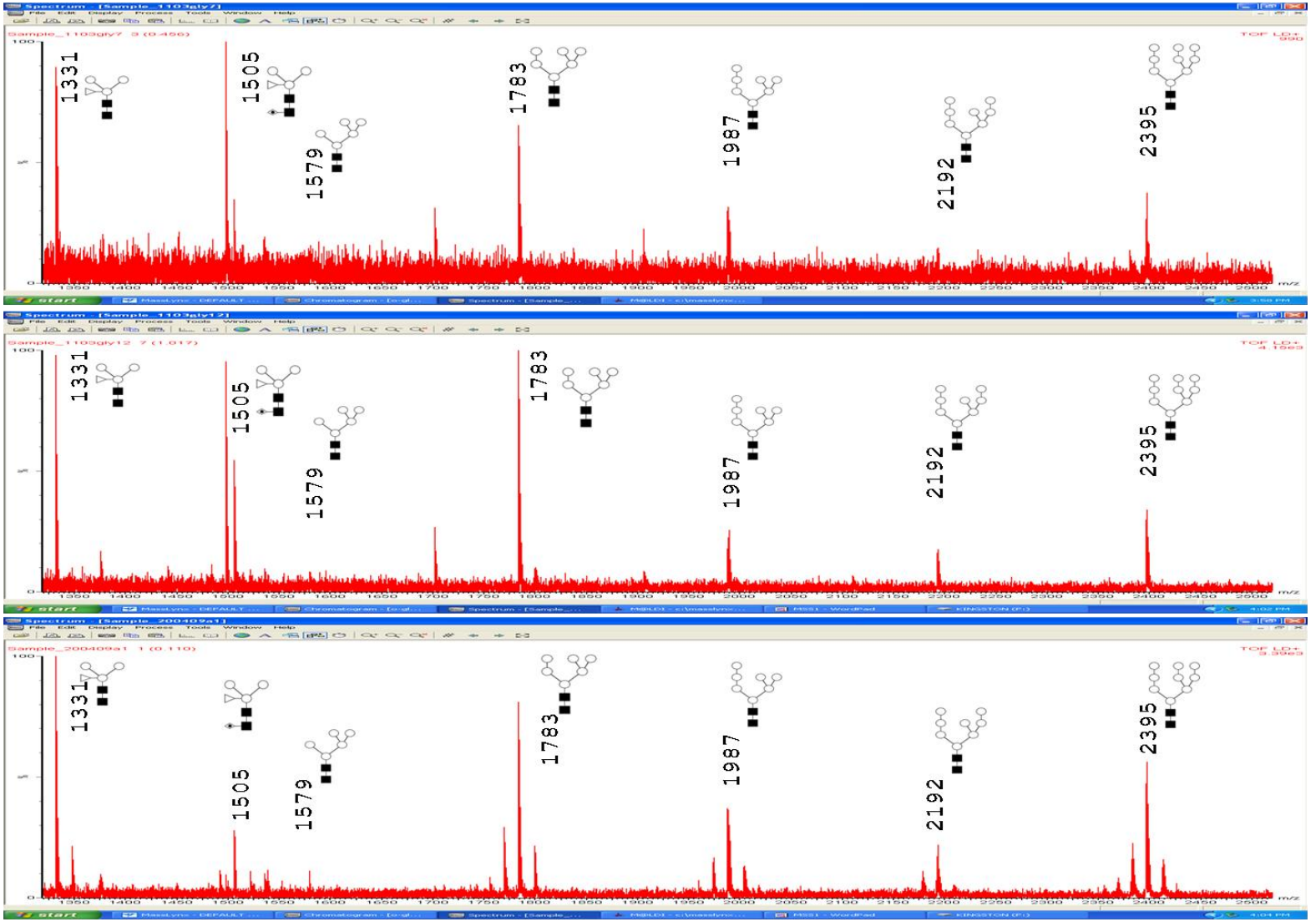

Figure 22b ***N*-glycan spectra of phaseolin extracted from single seeds (1, 2 and 3) from *Phaseolus vulgaris*, cv. Masterpiece.**

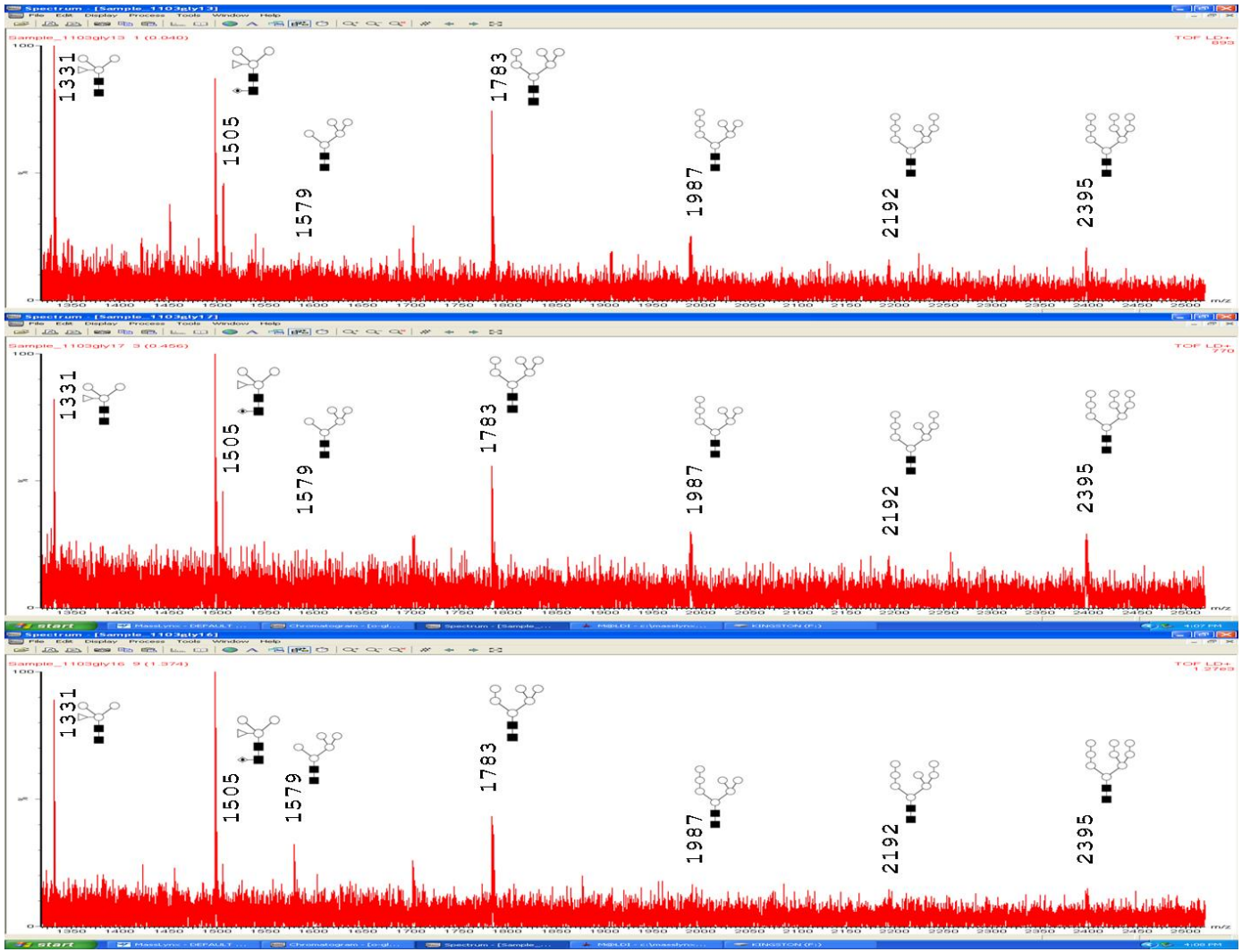

Figure22c ***N*-glycan spectra of phaseolin extracted from single seeds (1, 2 and 3) from *Phaseolus vulgaris*, cv. Blue Lake.**

Figure 22a, b and c. *N*-glycan spectra of phaseolin extracted from single seeds (1, 2 and 3) from three cultivars of *Phaseolus vulgaris*, cvs. Tendergreen (a), Masterpiece (b) and Blue Lake (c)

Milestone 03/01 Evaluation of methods to study *N*-linked glycosylation in natural phaseolin (Deliverable 02/01). Complete.
Task 03/a Test methods to determine *N*-glycosylation in Non-GM material. Complete.

In addition to the milestones set out in the project we analysed whether there were differences in the glycosylation pattern in different bean cultivars, as a measure of the natural variation in *N*-glycosylation. It can clearly be seen that all the glycans observed in *Phaseolus vulgaris* are also present in the Lima and Tepary beans. However, although the three cultivars of *Phaseolus vulgaris* have similar spectra, the spectra from Lima and Tepary beans are different in terms of the relative ratios of the various glycans. MALDI-Tof mass spectrometry is not quantitative, but the results indicate that natural variation in the *N*-glycosylation pattern occurs in phaseolin from various bean species. This is being analysed at Rothamsted with statistical methods, but is beyond the scope of the current project.

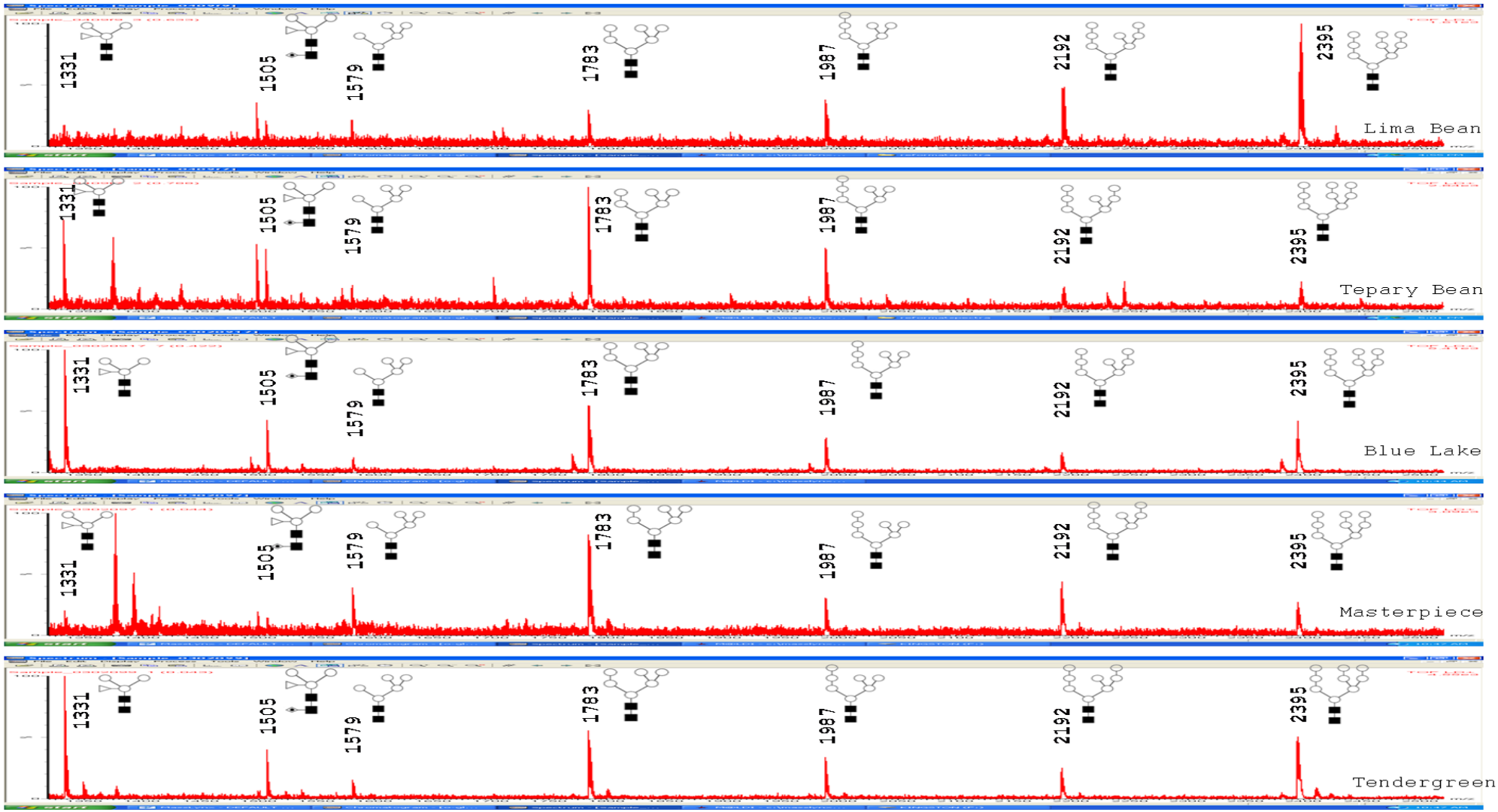

Figure 23. *N*-glycan spectra of phaseolin extracted from three bean species, *Phaseolus lunatus* (Lima bean), *Phaseolus acutifolius* (Tepary bean) and *Phaseolus vulgaris* (Common bean) cvs. Blue Lake, Masterpiece and Tendergreen.

B. Recombinant bean α-amylase inhibitor expressed in transgenic peas.

We were provided with the transgenic pea line expressing the bean α-amylase inhibitor (as reported by Prescott et al, 2005) by Dr TJ Higgins (CSIRO, Canberra), including the null segregant control line which was not used in the published study. These have been multiplied in our containment glasshouse to provide the starting material required.

We initially made several attempts at isolating the α-amylase inhibitor according to the method of Moreno and Chrispeels, 1989, but the yields were very low. We therefore developed a purification scheme based on Powers and Whitaker, 1977.

As with phaseolin, comparisons were made between proteins from individual seeds and plants to identify variation in glycan spectra of the α-amylase inhibitor.

Chromatograms and gels showing the purification of the α-amylase inhibitor from *Phaseolus vulgaris* cv Tendergreen and from Transgenic *Pisum sativum* cv Mukta are shown below (**Figs. 24** and **25**). Following extraction and ethanol precipitation the proteins of interest were separated sequentially on anion exchange and cation exchange columns. The yield of α-amylase inhibitor appears to be much greater using this purification scheme than was the affinity chromatography scheme tried initially. These extracts were then analysed using the *N*-glycan isolation and purification methodology used successfully for phaseolin.

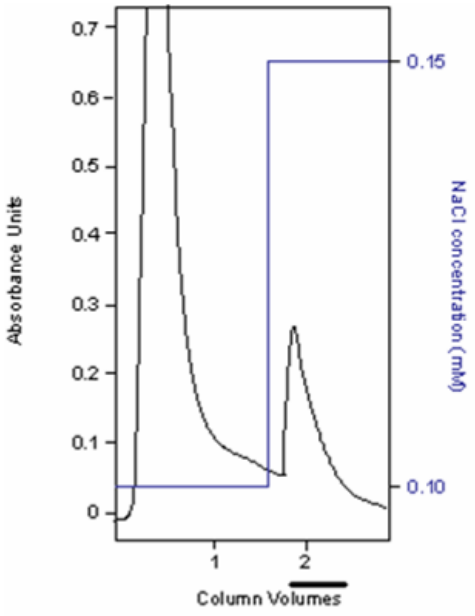

Figure 24: Purification of alpha-amylase inhibitor. DEAE chromatography of crude pea-transgenic extract (Pisum sativum cv. Mukta). The black bar indicates the fractions loaded onto a CM-sepharose column (see Figure 25).

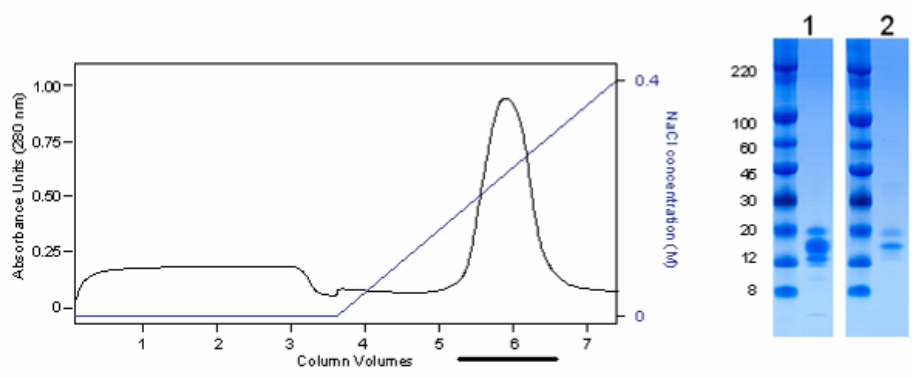

Figure 25: Purification of alpha-amylase inhibitor. (a) CM chromatography of pea-transgenic protein from DEAE column. The black bar indicates the fractions run on SDS-PAGE. (b) SDS-PAGE of CM column fractions. (1) Transgenic pea expressing bean alpha-amylase inhibitor cv. Mukta. (2) Bean cv. Tendergreen alpha-amylase inhibitor.

Milestone 03/02 Evaluation of methods on bean α-amylase expressed in transgenic pea.

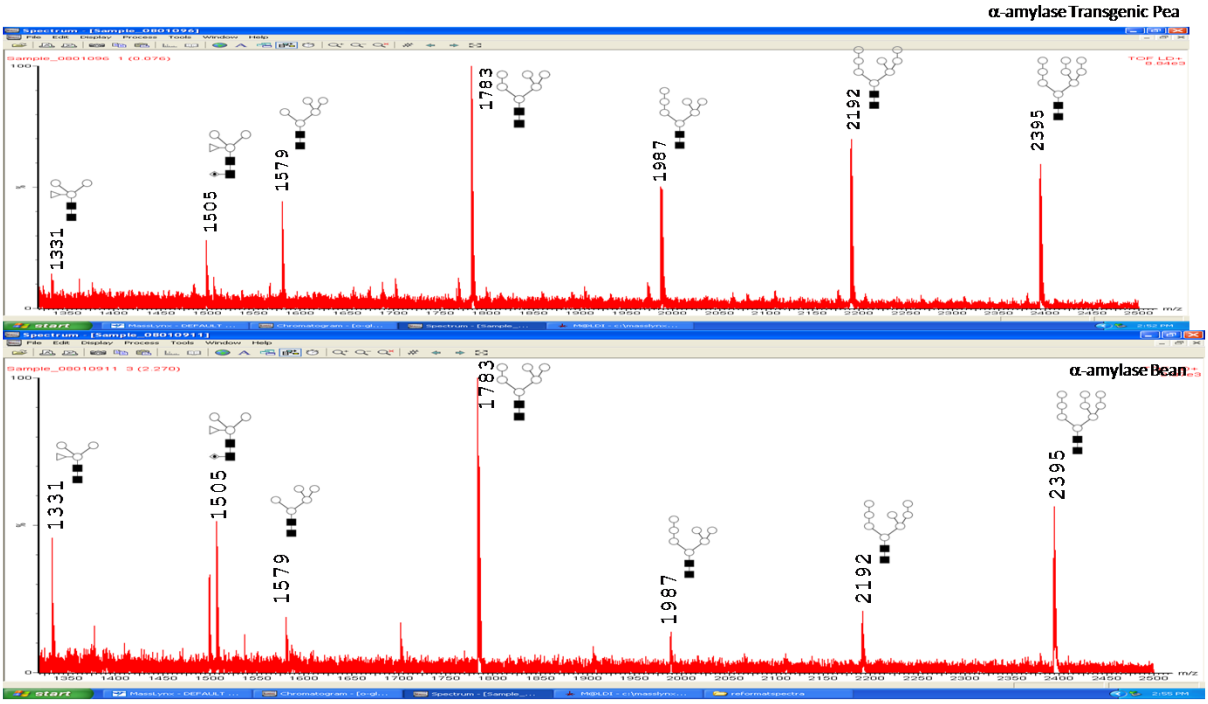

Figure 26. ***N*-glycan spectra of α-amylase inhibitor extracted from transgenic pea and native bean.**

It can be seen that the α-amylase inhibitor from transgenic pea and bean extracts contain the same glycans, but as with the phaseolin extracts from different bean species, the ratios of the various glycans do differ.

To assess the level of natural variation of glycosylation in α-amylase inhibitor we purified α-amylase inhibitor from single seeds from the transgenic pea line and the native bean (Figure 27a and b).

Figure 27a

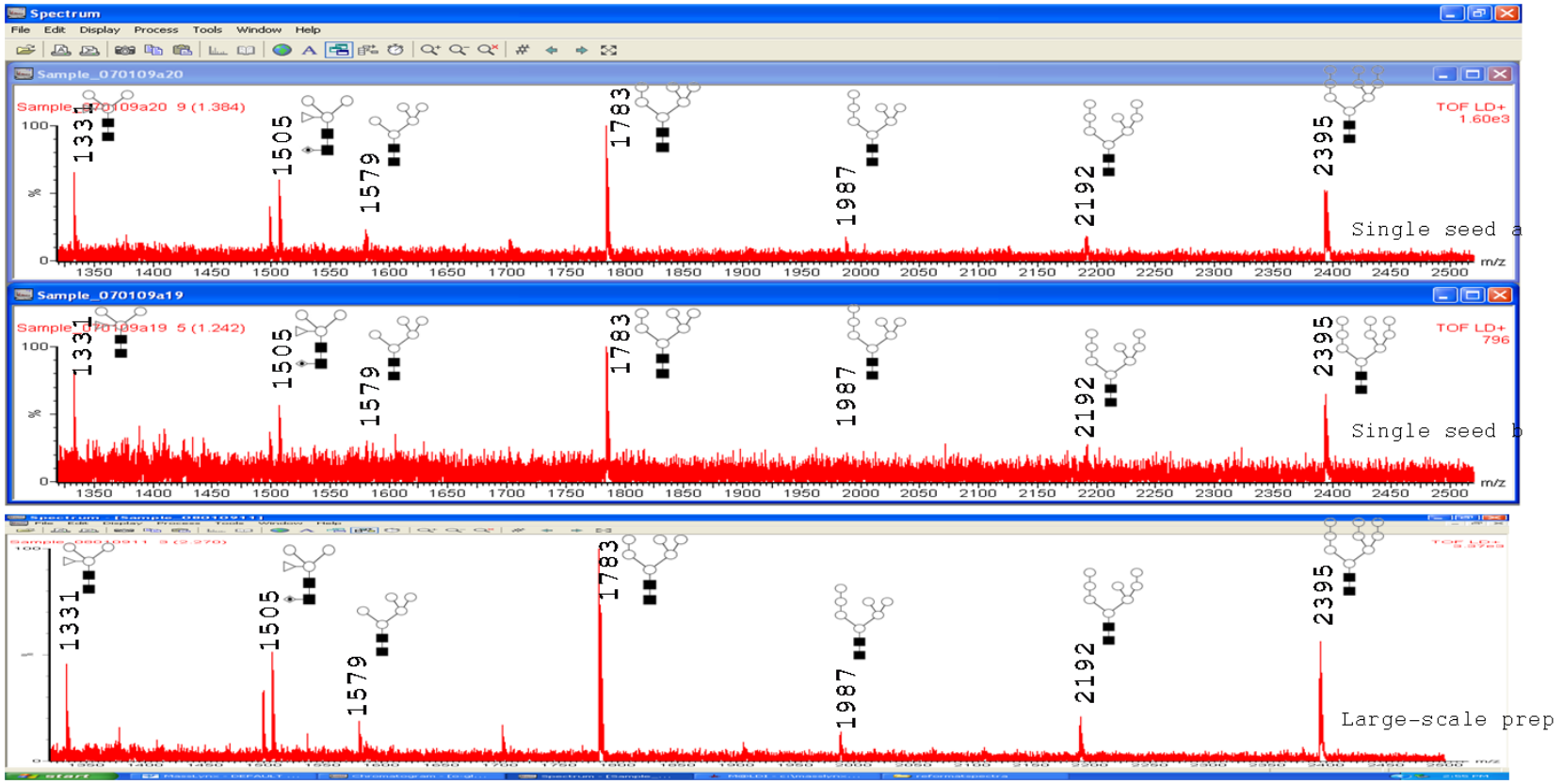
Spectrum
File Edit Display Process Tools Window Help
Sample_070109a20
Sample_070109a20 9 (1.384)
TOF LD+
1.60e3
1331
1505
1579
1783
1987
2192
2395
Single seed a
m/z
Sample_070109a19
Sample_070109a19 5 (1.242)
TOF LD+
796
Single seed b
Large-scale prep

Figure27b

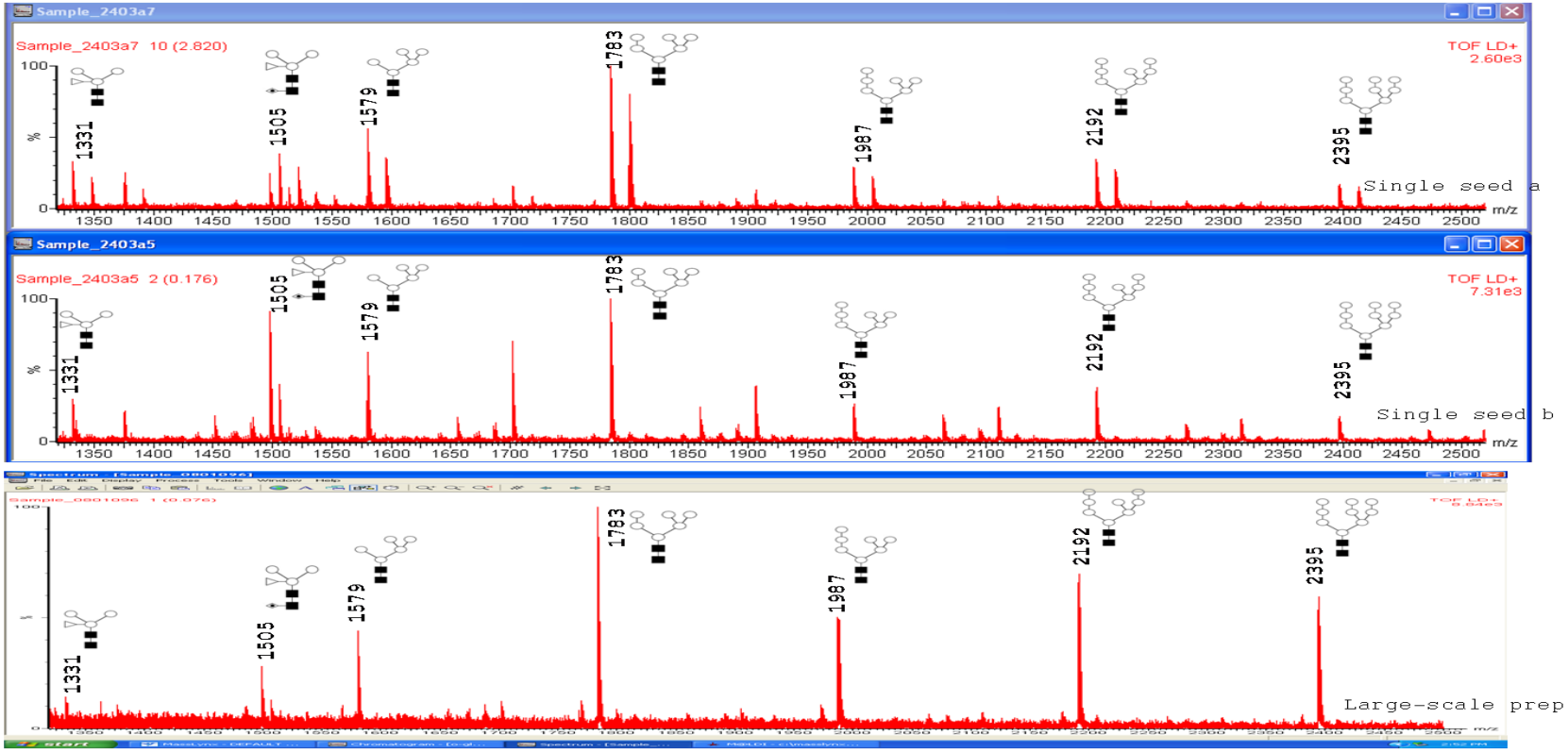

Figure 27a and b. *N*-glycan spectra of α-amylase inhibitor extracted from single seeds or large scale preparations of native bean (a) or transgenic pea (b). Upper two spectra in each panel are single seed extractions. Lower panels are large-scale preparations.

The glycan spectra from single seeds of bean are similar to those obtained from large-scale preparations, showing the robustness and sensitivity of the method developed in Cambridge. The spectra of the single seed preparations of α-amylase inhibitor from the transgenic pea are also similar but there is a higher level of contaminating hexose polymer in the spectra (as was observed in the case of phaseolin isolated from single seeds). This would indicate that there is little variation in glycan spectra at the single seed level. Thus, we have shown that the method is applicable to analysis of glycans from single seeds, and that the spectra are comparable to those obtained from large-scale-preparations.

Deliverable 03/01. Complete.
Task 03/b Test methods to determine *N*-glycosylation in GM material. Complete.

Additionally, we then compared at the glycan pattern from α-amylase inhibitor isolated from various bean species (Figure 28).

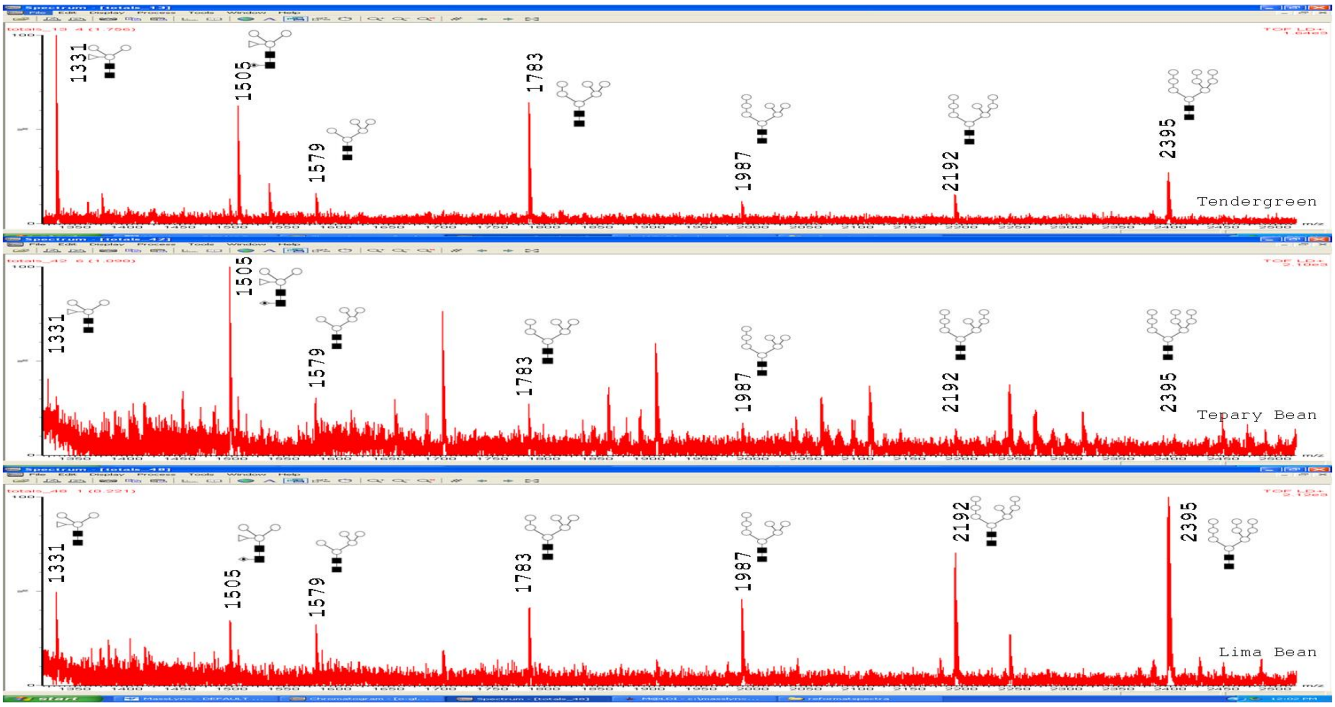

Figure 28. *N*-glycan spectra of α-amylase inhibitor extracted from three bean species, *Phaseolus vulgaris* (Common bean) cv. Tendergreen, *Phaseolus acutifolius* (Tepary bean) and *Phaseolus lunatus* (Lima bean).

As with phaseolin, the same glycans were identified but the relative ratios varied between the bean species.

In order to provide further information on the differences in glycan spectra of the transgenic and native α-amylase inhibitors we analysed the glycan spectra of total water and salt extracts from pea and bean. The major storage protein of bean, phaseolin is glycosylated. However, the storage proteins of pea are not glycosylated. We hypothesized, therefore that the reason that the glycan ratios were different in pea compared to bean was that the pea simply had a 'lower capacity' than the bean for glycosylation. Water- soluble protein extracts were made from pea and bean to extract the α-amylase inhibitor and 1M NaCl extracts, to also extract the major storage proteins, and the glycan spectra compared.

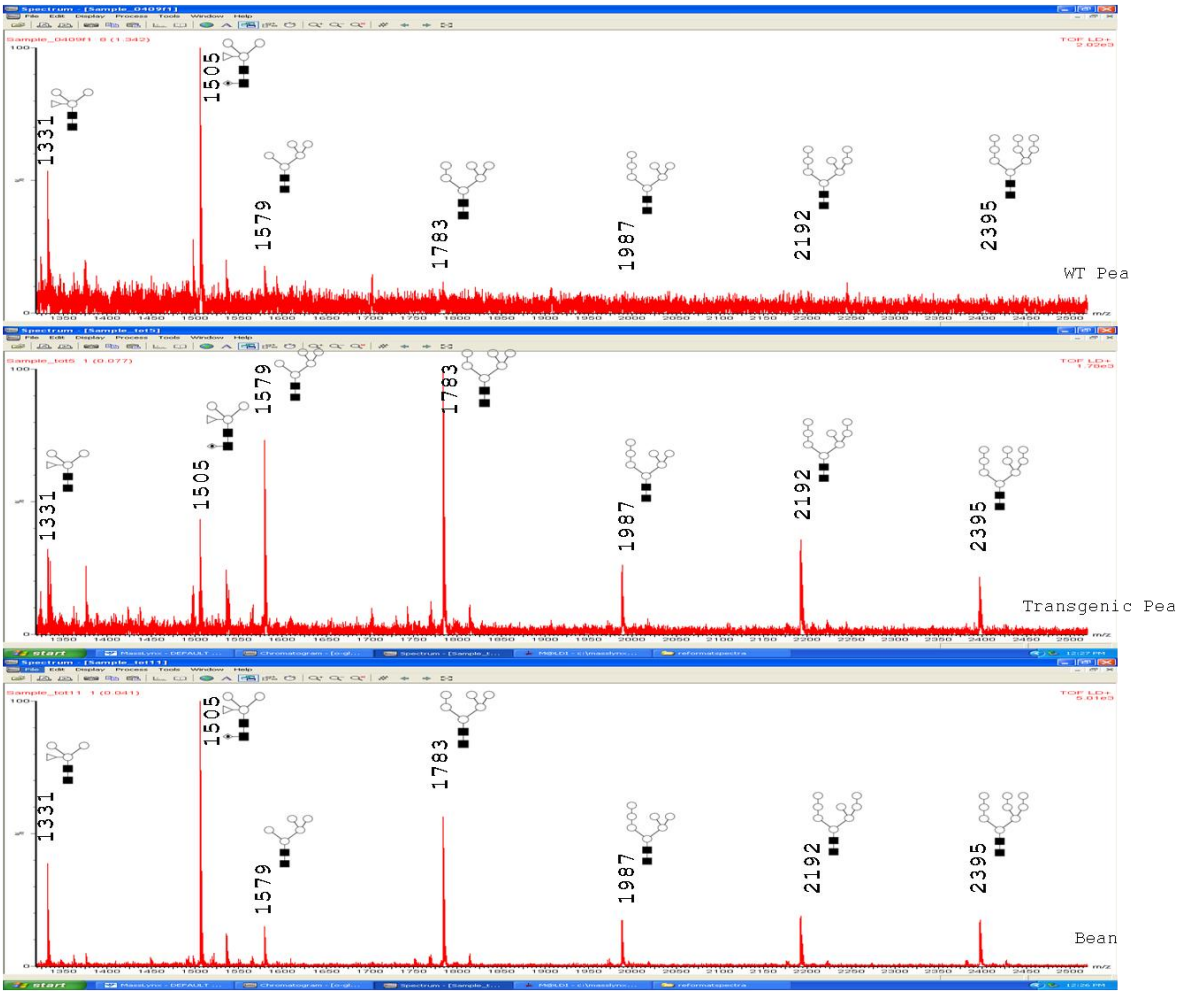

Figure 29. *N*-glycan spectra following water extraction of proteins from wild-type and transgenic *Pisum sativum* and from *Phaseolus vulgaris.*

The glycan patterns of the transgenic pea and the bean extracts were similar to those obtained for the purified α-amylase inhibitor proteins, whereas wild-type pea does not contain endogenous α-amylase inhibitor and few glycans are observed in the spectra. The NaCl extract shows more glycans in the wild-type pea (Figure 30), compared to the water extract (Figure 29) but the transgenic pea and the bean spectra are similar.

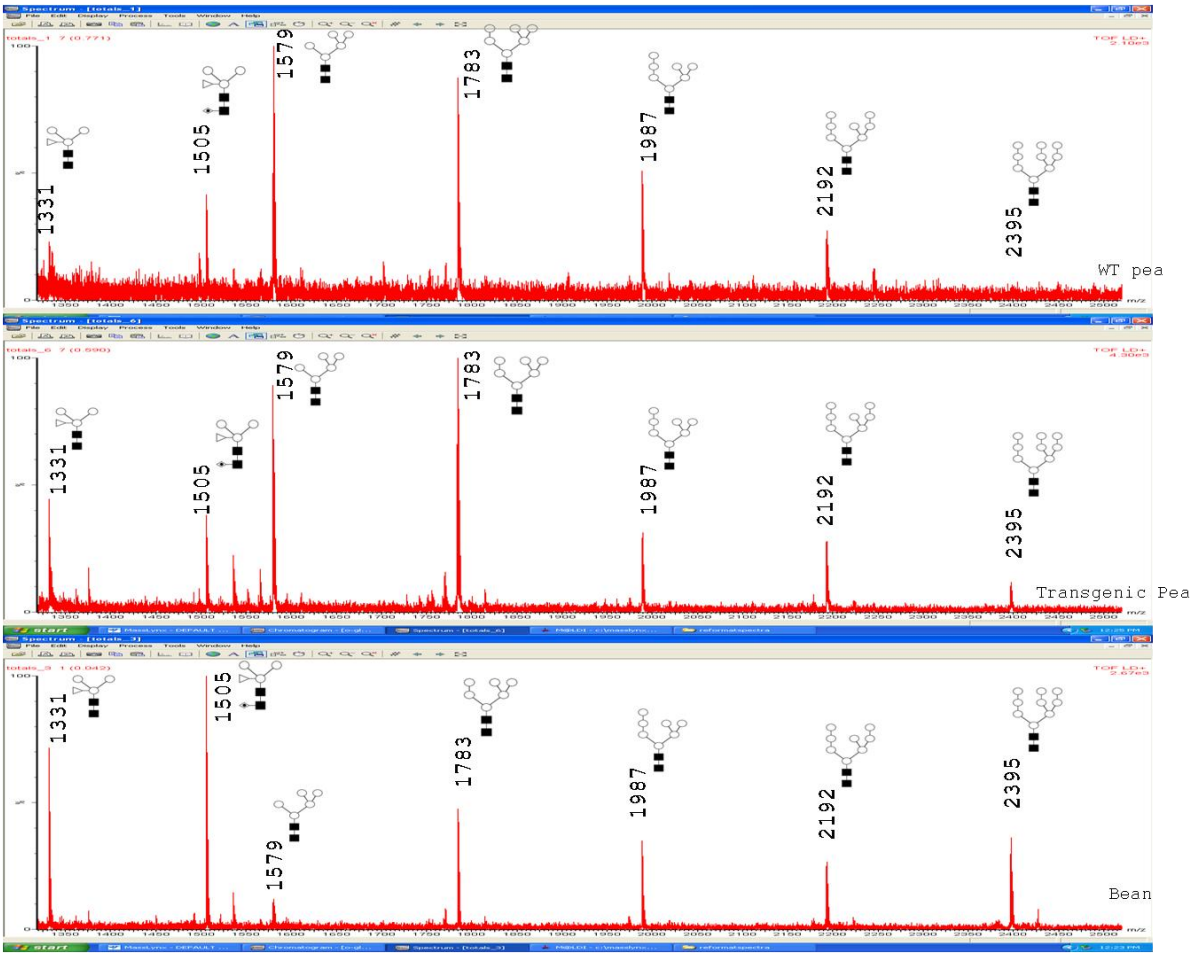

Figure 30. *N*-glycan spectra following 1M NaCl extraction of proteins from wild-type and transgenic *Pisum sativum* and from *Phaseolus vulgaris.*

Comparison of the glycan profiles shows that high mannose glycans are almost absent in the salt extracts from pea compared to bean, and that the mannose-5 glycan ratio is much higher in the pea compared to the bean. Thus, it is clear that the glycan profiles of endogenous pea and bean proteins are different. It is therefore possible that the differences observed in the ratio of glycans in the transgenic α-amylase inhibitor, compared to the native α-amylase inhibitor in bean, reflect differences in the endogenous glycosylation machinery in the species.

In conclusion, the method developed at Cambridge and transferred to Rothamsted is very robust and has been developed as a Standard Operating Procedure (SOP) which is attached as Appendix 1.

Deliverable 05/01. Complete.

Task 05/a SOP for analysis of *N*-glycans: Complete.

SOP appended to this report as Appendix 1.

Objective 04: Evaluation of methods to study *O*-glycosylation of wheat arabinogalactan peptide.

Milestone 04/02: Evaluation of methods to study O-linked glycosylation on arabinogalactan from wheat flour.

As part of project objective 02 Rothamsted provided Cambridge with wheat arabinogalactan peptide, isolated by the methods of Looveld et al, 1997 and Van den Bulck et al, 2005. In order to test the new methods being developed at Cambridge for identification and characterisation of *O*-linked glycans, these methods were transferred to Rothamsted for analysis of AGP extracted from Cadenza (hexaploid bread wheat), durum (tetraploid pasta wheat) and einkorn (diploid) wheat, as well as from the control and two transgenic wheat lines expressing additional copies of the gene encoding AGP produced at Rothamsted.

(a)

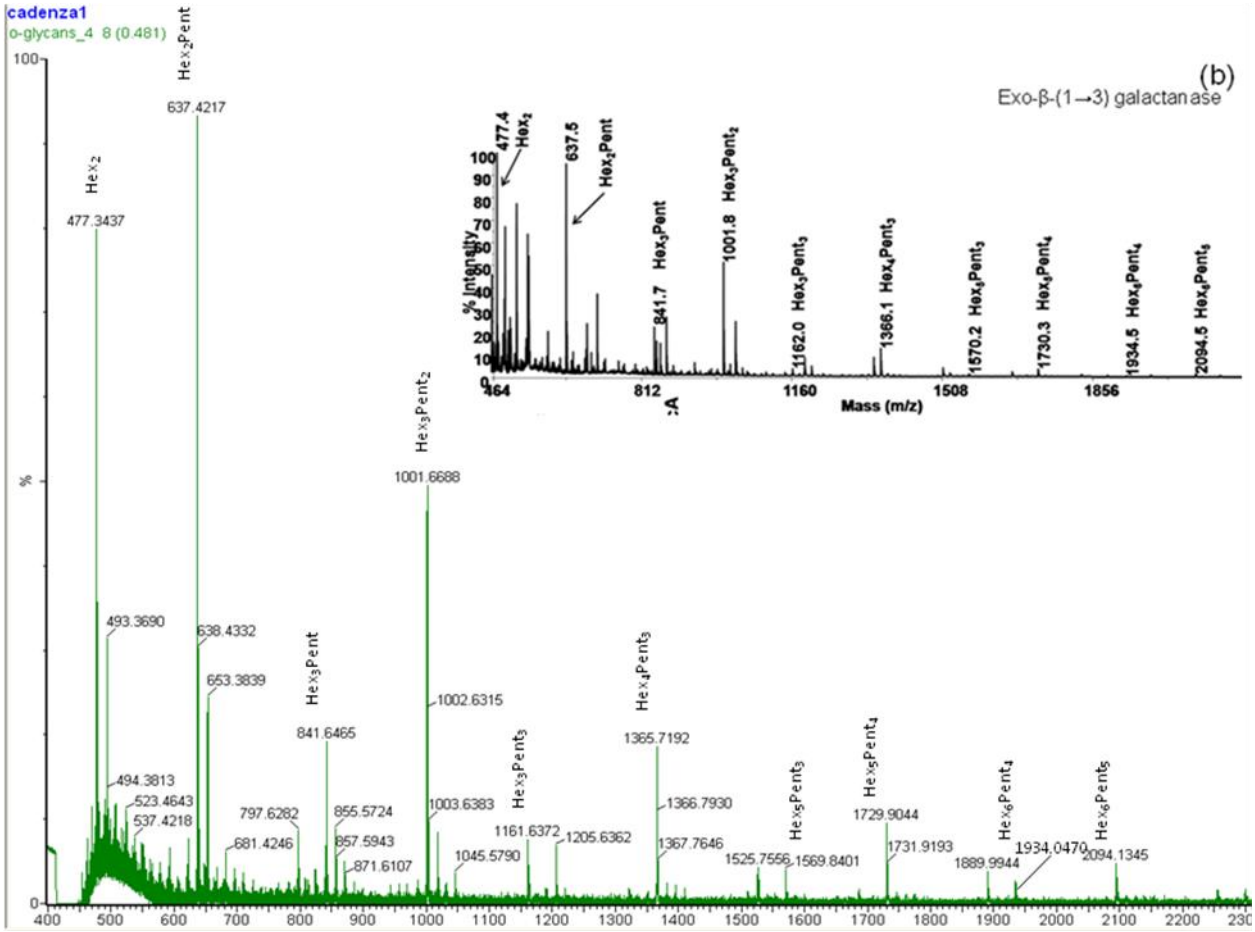

Figure 31a and b. MALDI-ToF MS spectra showing signals of the *per*-methylated oligosaccharides released by exo-β-(1→3) galactanase from AGP isolated from Cadenza bread wheat, (a) spectra obtained at Rothamsted (b) spectra obtained at Cambridge (from Figure 13).

The methods developed in Cambridge for the analysis of *O*-linked glucans were transferred to Rothamsted and applied to AGP purified from the various wheat lines. Figure 31 shows that the method was very robust. A very similar spectrum was obtained at Rothamsted for the AGP from the bread wheat Cadenza, as that obtained in Cambridge (see insert (b)). We then applied this method to AGP isolated from durum and einkorn wheats (Figure 32), and to two transgenic wheat lines expressing extra copies of the gene encoding AGP (Figure 33).

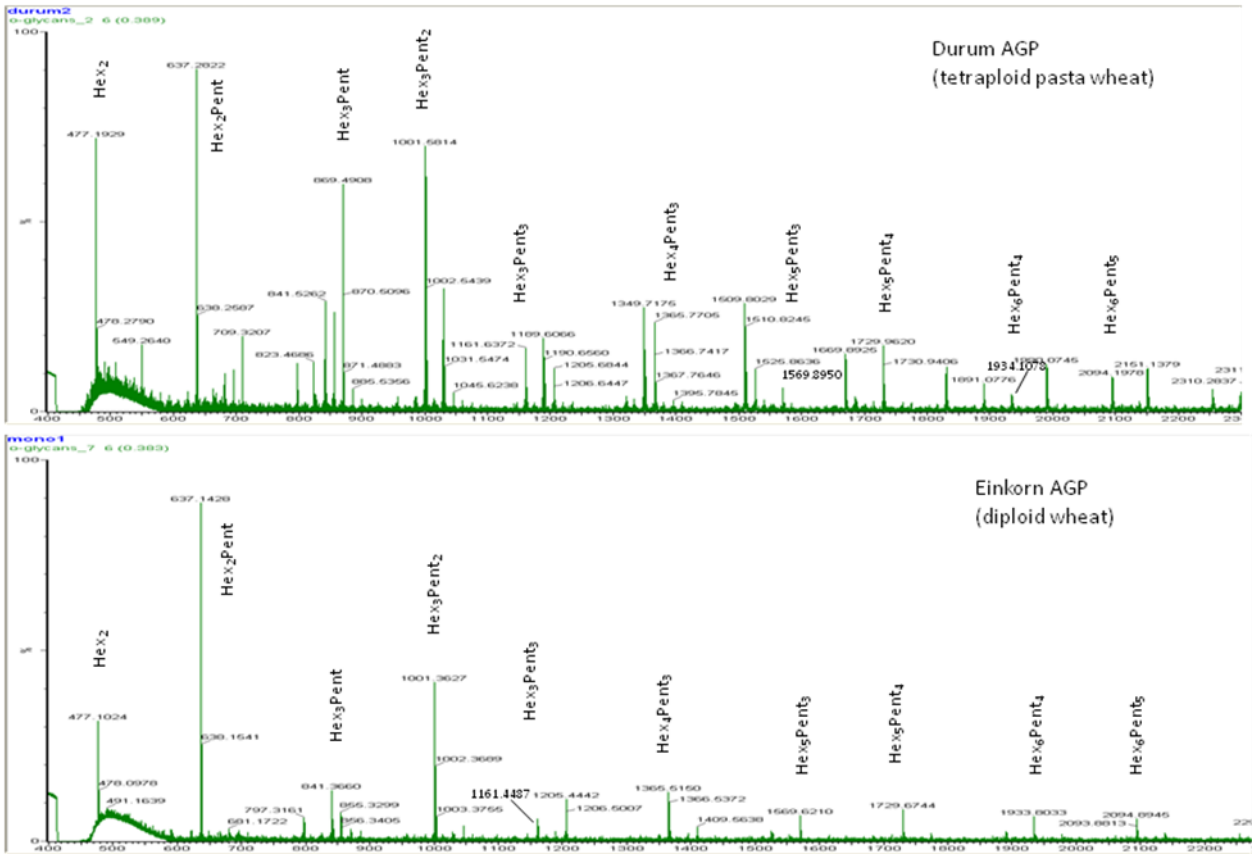

Figure 32. MALDI-ToF MS spectra showing signals of the *per*-methylated oligosaccharides released by exo-β-(1→3) galactanase from AGP isolated from Durum and Einkorn wheats

The oligosaccharide pattern was similar to that obtained for Cadenza AGP, but the AGP isolated from durum wheat appears to have a higher ratio of long oligosaccharides compared to AGP isolated from einkorn wheat.

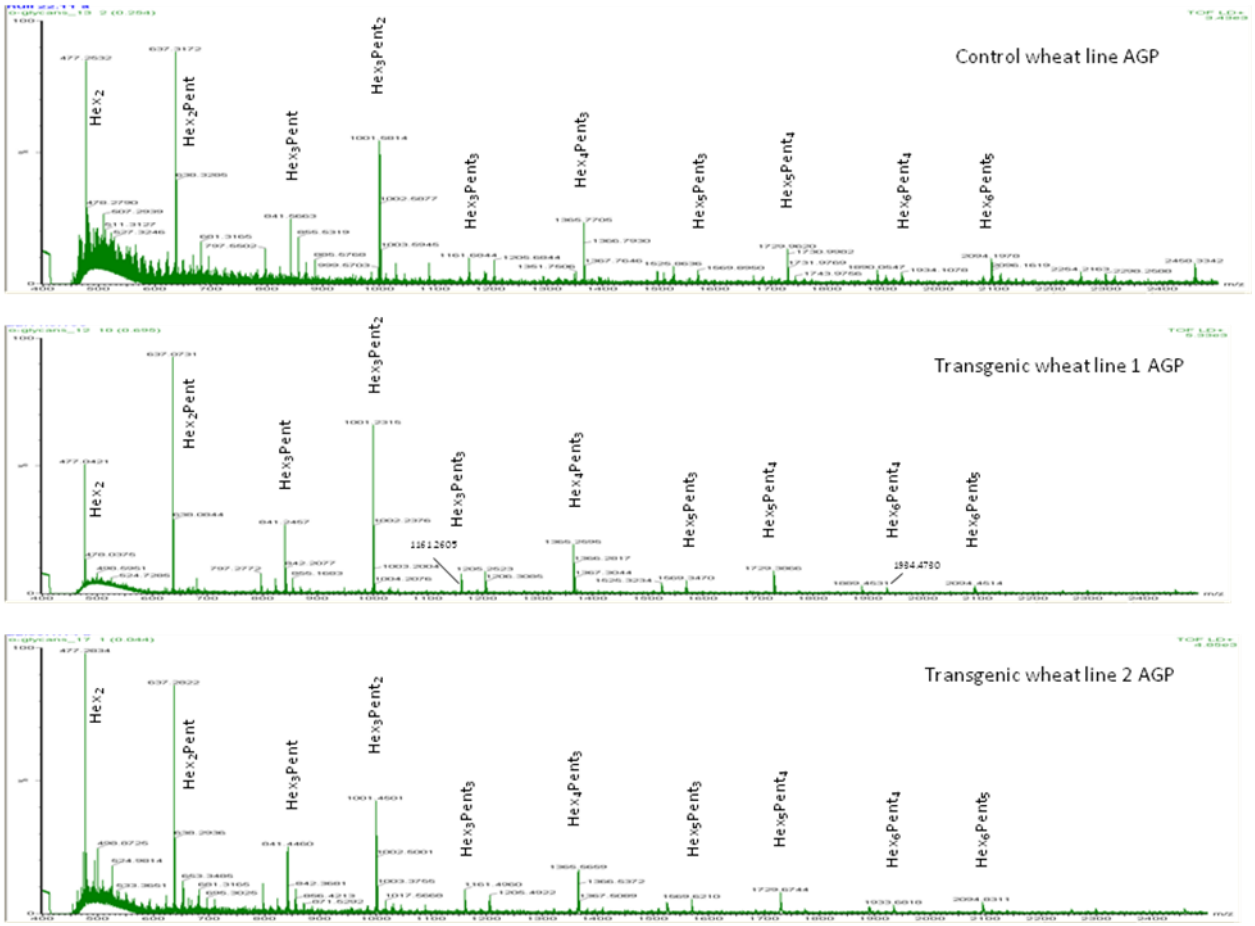

Figure 33. MALDI-ToF MS spectra showing signals of the *per*-methylated oligosaccharides released by exo-β-(1→3) galactanase from AGP isolated from two transgenic lines and one control line.

Again, the method worked well, and all the oligosacchrides were identified using the Cambridge method. However, the ratios of the oligosaccharides differed between the extracts. For example, the Hex_2:Hex_2Pent_2 ratios, were different in each extract. Transgenic wheat line 1 AGP seems to have a lower Hex_2:Hex_2Pent ratio than either the control line or transgenic line 2 suggesting that perhaps there are less unsubstituted oligosaccharides in that sample. The control line also appears to have higher levels of longer oligosaccharide fragments compared to transgenic lines 1 and 2.

Deliverable 04/01 complete.

Task 04/a Test methods to determine *O*-glycosylation in Non-GM material (AGP). Complete.
Task 04/b Test methods to determine *O*-glycosylation in GM material (AGP). Complete.

Conclusions

The first objective of project G03029 was to develop methods to identify and characterise *N*-glycans on plant proteins in order to determine whether proteins expressed in transgenic plants are inauthentically glycosylated. The proposed strategy involved the use of enzymatic and chemical methods for the release of *N*-linked glycans and the evaluation of the developed method on other *Brassica* species. Our work showed that hydrazinolysis could not be used for the analysis of *N*-linked glycans as it did not yield detectable amounts of *N*-glycans. In addition the chemicals used toxic, presenting a great health and safety hazard. By contrast, pepsin digestion of Arabidopsis glycoproteins with consecutive release of *N*-glycans with PNGase A digestion followed by *per*-methylation of free *N*-linked glycans, proved to be very robust and sensitive leading to the detection of glycan species not previously reported for Arabidopsis. The reproducibility of the method was tested on cauliflower glycoproteins, where it allowed the detection of a new *N*-glycan species which was not known to be present in cauliflower glycoproteins from the florets of this plant. Another advantage of this method is that accurate results can be obtained within a short period of time (about six to seven days) and this period can be even shorter (about five days when used in a standard lab at full speed) if immobilized pepsin is utilised.

The *N*-glycan methods developed in Cambridge have been applied at Rothamsted (Objective 03) to native phaseolin from three different bean cultivars and three species of bean. It has also used to analyse crude extracts from pea and cauliflower and native α-amylase inhibitor from three bean species has proven robust, sensitive (down to the single seed level) and reproducible. The method has also been applied to recombinant α-amylase inhibitor protein from transgenic pea. The method has been reduced to an SOP (Objective 05) and may become a valuable tool in future risk assessment strategies (see Appendix 1).

Objective 02 of the G03029 project was to develop methods for the analysis of *O*-glycosylation in transgenic plants. GFP-glycan reporter constructs expressed in tobacco were initially used due to the reported small size of the glycans and the fact that they have been relatively well characterised. Preliminary HPAEC-PAD data suggest that these glycans contain arabinose, galactose, rhamnose and glucuronic acid modifications. In addition, our data indicate that AP51 polysaccharides are extremely heterogeneous, and are likely to be

larger than *N*-glycans from fetuin. Capillary Normal Phase-Liquid Chromatography coupled off-line with MALDI-ToF-mass spectrometry was used in order to reduce the complexity of the above samples. This has provided the first MS data on the structure of *O*-linked glycans. Attempts to further characterise these structures by MS/MS however, were not successful as signals from overlapping structural isomers and/or from neighbouring oligomer structures did not allow a confident interpretation of the spectra. The arabinogalactan polysaccharide from transgenic tobacco was a rather complex mixture of glycans, much larger in size than originally anticipated. However those preliminary experiments revealed that our methodology could be used successfully for the detection of *O*-linked glycans. Our approach involves monosaccharide analysis for the detection of possible compositional differences (both qualitative as well as quantitative) in transgenic plants and consecutive treatment of the glycans with exo-β-1,3-galactanase followed by MALDI-ToF-MS analysis of the *per*-methyalted sugars, for the identification of major structural modifications. The reproducibility of the method was tested on wheat arabinogalactan peptide (AGP), where it allowed not only the detection of *O*-linked glycans but also enabled us to propose a model for the structure of those complex molecules in wheat AGP. The method was successfully transferred to Rothamsted (Objective 04) where it permitted the detection of *O*-linked glycans in AGP preparations isolated from hexploid, tetraploid and diploid wheat lines, as well as from transgenic wheat expressing extra copies of the gene encoding AGP.

Acknowledgements:

This work was funded by the Foods Standards Agency.

References

Aurikko, J. P., *et al.* (2005). "Characterization of Symmetric Complexes of Nerve Growth Factor and the Ectodomain of the Pan-neurotrophin Receptor, p75NTR." J. Biol. Chem. **280**(39): 33453-33460.

Bollini, R., *et al.* (1983). "In vivo and in vitro processing of seed reserve protein in the endoplasmic reticulum: evidence for two glycosylation steps." J. Cell Biol. **96**(4): 999-1007.

Ciucanu, I., *et al.* (1984). "A simple and rapid method for the permethylation of carbohydrates." Carbohydrate Research **131**(2): 209-217.

Courtin, C. M., *et al.* (2002). "Arabinoxylans and Endoxylanases in Wheat Flour Bread-making." Journal of Cereal Science **35**(3): 225-243.

Domon, B., *et al.* (1988). "A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates." Glycoconjugate Journal **5**(4): 397-409.

Fincher, G. B., *et al.* (1974). "Chemical and physical properties of an arabinogalactan-peptide from wheat endosperm." Biochem. J. **139**(3): 535-545.

Fincher, G. B., *et al.* (2004). Encyclopedia of grain science, Elsevier Academic Press.

Fujiyama, K., *et al.* (2007). "*N*-linked glycan structures of glycoproteins in suspension-cultured *Arabidopsis thaliana* MM2d cells." Plant Biotechnology **24**: 255-259.

Gane, A. M., *et al.* (1995). "Structural analysis of the carbohydrate moiety of arabinogalactan-proteins from stigmas and styles of Nicotiana alata." Carbohydrate Research **277**(1): 67-85.

Gaspar, Y., *et al.* (2001). "The complex structures of arabinogalactan-proteins and the journey towards understanding function." Plant Molecular Biology **47**(1): 161-176.

Goubet, F., *et al.* (2002). "Polysaccharide Analysis Using Carbohydrate Gel Electrophoresis: A Method to Study Plant Cell Wall Polysaccharides and Polysaccharide Hydrolases." Analytical Biochemistry **300**(1): 53-68.

Green, E. D., *et al.* (1988). "The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy." J. Biol. Chem. **263**(34): 18253-18268.

Hall, T.C., *et al*. (1999). "Phaseolin: its Past, Properties, Regulation and Future" *in* Seed Proteins pp 209-240 Eds.Shewry, P.R. and Cassey, R. Kluwer Academic Publishers.

Haque, M. A., *et al.* (2005). "Mode of Action of β-Glucuronidase from Aspergillus niger on the Sugar Chains of Arabinogalactan-Protein." Bioscience, Biotechnology, and Biochemistry **69**(11): 2170-2177.

Hata, K., *et al.* (1992). "{alpha}-L-Arabinofuranosidase from Radish (Raphanus sativus L.) Seeds." Plant Physiology **100**(1): 388-396.

Kotake, T., *et al.* (2004). "Molecular cloning and expression in Escherichia coli of a Trichoderma viride endo-beta-(1-->6)-galactanase gene." Biochem. J. **377**(3): 749-755.

Leonard, R., *et al.* (2005). "Two novel types of O-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding." Journal of Biological Chemistry **280**(9): 7932-7940.

Loosveld, A. M. A., *et al.* (1998). "Structural variation and levels of wtaer-extractable arabinogalactan-peptide in European wheat flour." Cereal Chemistry **75**: 815-819.

Majewska-Sawka, A., *et al.* (2000). "The Multiple Roles of Arabinogalactan Proteins in Plant Development." Plant Physiol. **122**(1): 3-10.

Mares, D. J., *et al.* (1973). "Studies on wheat endosperm. 1. Chemical composition and ultrastructure of the cell walls." Aust J Biol Sci **26**: 793-812.

Maslen, S., *et al.* (2006). "Differentiation of Isomeric N-Glycan Structures by Normal-Phase Liquid Chromatography-MALDI-TOF/TOF Tandem Mass Spectrometry." Anal. Chem. **78**(24): 8491-8498.

Maslen, S. L., *et al.* (2007). "Structure elucidation of arabinoxylan isomers by normal phase HPLC-MALDI-TOF/TOF-MS/MS." Carbohydrate Research **342**(5): 724-735.

Matsuoka, K., *et al.* (1995). "O-glycosylation of a precursor to a sweet potato vacuolar protein, sporamin, expressed in tobacco cells." The Plant Journal **8**(6): 877-889.

Moreno, J. and Chrispeels M.J. (1989)."A lectin gene encodes the alpha-amylase inhibitor of the common bean." *PNAS USA* **86** (20):7885-7889.

Powers, J.R. and Whitaker J.R (1977)."Purification and some physical and chemical properties of red kidney bean (Phaseolus vulgaris) α-amylase inhibitor."J.Food Biochem **1** (3):217-238.

Prescott, V. E., *et al.* (2005). "Transgenic Expression of Bean α-Amylase Inhibitor in Peas Results in Altered Structure and Immunogenicity." J. Agric. Food Chem. **53**(23): 9023-9030.

Sanchez-Monge, R., *et al.* (1992). "Wheat and barley allergens associated with baker's asthma. Glycosylated subunits of the alpha-amylase-inhibitor family have enhanced IgE-binding capacity." Biochem J **281**(Pt 2): 401-405.

Stephens, E., *et al.* (2004). "The N-linked oligosaccharides of aminopeptidase N from *Manduca sexta* - Site localization and identification of novel N-glycan structures." European Journal of Biochemistry **271**(21): 4241-4258.

Sturm ,A.J., *et al*, 1987a "Structure, Positions, and Biosynthesis of the High mannose and the Complex Oligosaccharide Side Chains of the bean Storage Protein Phaseolin." J. Biol. Chem. **262** (28): 13392-13403.

Sturm ,A.J., *et al*, 1987b "Subcellular Location of Glycosidases and Glucosyltransferases Involved in the Processing of *N*-Linked Oligosaccharides. " Plant Physiology **85** (3):741-745

Takasaki, S., *et al.* (1982). Hydrazinolysis of asparagine-linked sugar chains to produce free oligosaccharides. Methods in Enzymology, Academic Press. **Volume 83:** 263-268.

Tan, L., *et al.* (2003). "Glycosylation Motifs That Direct Arabinogalactan Addition to Arabinogalactan-Proteins." Plant Physiol. **132**(3): 1362-1369.

Tan, L., *et al.* (2004). "Structure of a Hydroxyproline (Hyp)-Arabinogalactan Polysaccharide from Repetitive Ala-Hyp Expressed in Transgenic *Nicotiana tabacum*." J. Biol. Chem. **279**(13): 13156-13165.

Tsumuraya, Y., *et al.* (1990). "Purification of an exo-beta-(1----3)-D-galactanase of Irpex lacteus (Polyporus tulipiferae) and its action on arabinogalactan-proteins." Journal of Biological Chemistry **265**(13): 7207-7215.

Van den Bulck, K., *et al.* (2005). "Isolation of cereal arabinogalactan-peptides and structural comparison of their carbohydrate and peptide moieties." Journal of Cereal Science **41**(1): 59-67.

van Ree, R., *et al.* (2000). "beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens." Journal of Biological Chemistry **275**(15): 11451-11458.

Wilson, I. B. H., *et al.* (2001). "Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core {{alpha}}1,3-linked fucose and xylose substitutions." Glycobiology **11**(4): 261-274.

Appendix

Glycan legend key: