

Combined N-Glycome and N-Glycoproteome Analysis of the *Lotus japonicus* Seed Globulin Fraction Shows Conservation of Protein Structure and Glycosylation in Legumes

Svend Dam,^{†,‡} Morten Thaysen-Andersen,[‡] Eva Stenkjær,[†] Andrea Lorentzen,[§] Peter Roepstorff,[§] Nicolle H. Packer,[‡] and Jens Stougaard^{*,†}

[†]Centre for Carbohydrate Recognition and Signalling, Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10, DK-8000 Aarhus C, Denmark

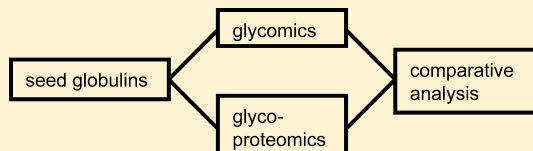
[‡]Department of Chemistry and Biomolecular Sciences, Macquarie University, NSW, 2109, Australia

[§]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Supporting Information

ABSTRACT: Legume food allergy, such as allergy toward peanuts and soybeans, is a health issue predicted to worsen as dietary advice recommends higher intake of legume-based foods. *Lotus japonicus* (*Lotus*) is an established legume plant model system for studies of symbiotic and pathogenic microbial interactions and, due to its well characterized genotype/phenotype and easily manipulated genome, may also be suitable for studies of legume food allergy. Here we present a comprehensive study of the *Lotus* N-glycoproteome. The global and site-specific N-glycan structures of *Lotus* seed globulins were analyzed using mass spectrometry-based glycomics and glycoproteomics techniques. In total, 19 N-glycan structures comprising high mannose (~20%), pauci-mannosidic (~40%), and complex forms (~40%) were determined. The pauci-mannosidic and complex N-glycans contained high amounts of the typical plant determinants β -1,2-xylose and α -1,3-fucose. Two abundant *Lotus* seed N-glycoproteins were site-specifically profiled; a predicted lectin containing two fully occupied N-glycosylation sites carried predominantly pauci-mannosidic structures in different distributions. In contrast, *Lotus* convicilin storage protein 2 (LCP2) carried exclusively high mannose N-glycans similar to its homologue, Ara h 1, which is the major allergen in peanut. *In silico* investigation confirmed that peanut Ara h 1 and *Lotus* LCP2 are highly similar at the primary and higher protein structure levels. Hence, we suggest that *Lotus* has the potential to serve as a model system for studying the role of seed proteins and their glycosylation in food allergy.

KEYWORDS: Ara h 1, food allergy, globulin, glycomics, glycoproteomics, *Lotus japonicus*, N-glycosylation, peanut



■ INTRODUCTION

N-Linked glycosylation alters the physicochemical properties of modified proteins and the N-glycans are increasingly recognized for modulating the protein function and for showing functions independent of the protein carrier.^{1–5} N-Glycosylation is initiated in the secretory pathway where a common precursor N-glycan can be attached to an asparagine residue in the Asn-X-Ser/Thr motif (X $\neq$ Pro).⁶ Although, many similarities exist between the protein N-glycosylation of plants and mammals, in particular in the initial glycan processing in the endoplasmic reticulum, significant differences have also been documented. Most apparent is the N-glycan maturation in the Golgi apparatus.^{7,8} Some of the main structural differences are: (1) presence of xylose residues and absence of sialic acid residues in plant N-glycans, (2) presence of α -1,3 core-fucosylation in plant N-glycans opposed to α -1,6 core-fucosylation in mammalian N-glycans (which make plant N-glycans insensitive to the common deglycosylation enzyme PNGase F), and (3) absence of the bisecting β -1,4-N-acetylglucosamine (GlcNAc) in plants.⁹ The enzymatic processing in the Golgi apparatus usually leads to a

heterogeneous N-glycan decoration of each glycosylation site. Following processing, mature plant glycoproteins are incorporated into vesicles and transported to their intended destinations, mainly organelles (e.g., lysosome and chloroplasts), membranes (e.g., cell membrane), and extracellular matrix.

Glycosylation of plant proteins has been suggested to play a role in food allergy, and it is well-known that legume seeds, such as peanut and soybean as well as seeds from species such as hazelnut, walnut, and almond, upon ingestion can trigger an allergic response in a subset of the human population.¹⁰ Mature legume seeds contain mainly globulins, which can be divided into three protein families: vicilin, convicilin, and legumin.¹¹ All of these are insoluble in water but soluble in high-salt solvents. Several seed globulins are known to be allergens.^{12,13} The general assumption is that legume food allergy is caused by aberrant proteolytic digestion in the gastrointestinal tract where proteolytic enzymes are located.^{14–16} Partially degraded peptides

Received: March 13, 2013

of more than six to eight amino acid residues can cross into the blood circulation with their epitopes exposed on MCH II receptors on antigen-presenting B-cells. Following initial exposure to such an allergen, specific IgE antibodies will be produced that maintain a background level over time. This molecular memory functionality enables humans to trigger a strong response upon subsequent allergen exposures resulting in histamine release and allergic symptoms (i.e., heat, pain, swelling, redness, and itchiness).¹⁰

Allergen Ara h 1, clone P41B (recommended Uniprot name, Accession number P43238, hereafter called Ara h 1), a peanut globulin from the vicilin super family, is one of the best-characterized allergens. In the vast majority of people suffering from a peanut allergy, the immune system triggers a response against one or several protein epitopes from Ara h 1.¹⁷ Identified epitopes are located throughout the protein sequence with no preference for a specific area of the protein.^{17–19} The three-dimensional protein structure of Ara h 1 indicates that several epitopes are located at the protein surface which is in good agreement with the fact that Ara h 1 is resistant to both proteolysis and heat.^{15,16}

Generally, the proteins from the convicilin and legumin super families are rarely modified with N-glycosylation whereas vicilin globulins are frequently glycosylated.²⁰ Several of the glycosylated legume allergens contain plant specific glycan determinants, that is, α -1,3-core-fucose and β -1,2-xylose residues. It has been reported that IgE antibodies are produced partly against these glycan determinants.^{21–25} This was supported by gene silencing studies where the reduction of α -1,3-fucose and β -1,2-xylose levels in plant glycopeptides decreased the immunogenic response.^{26–28} However, other studies reported that the plant-specific N-glycan determinants are not involved in triggering the allergic response.²⁹

The single N-glycosylation site of the allergenic vicilin globulin, Ara h 1, is exposed on the protein surface. Synthetic peptides confirmed that the polypeptide chain around this site is a partial epitope for IgE.^{16,18} However, not all individuals are allergic and the naturally occurring peptide sequence of peanut Ara h 1 is N-glycosylated.³⁰ Thus, it is possible that the plant N-glycans, which sterically cover a relatively large volume close to the polypeptide backbone due to α -1,3-fucose and β -1,2-xylose residues of the core glycan structure, could mask the proteolytic cleavage sites leading to incomplete degradation into allergenic peptides. At present, the exact relationship between the immunogenic/allergic response and plant protein glycosylation is not known.

With the increasing incidence of legume allergy and the higher consumption of legumes in recent years,^{31–33} it is of importance to define the molecular mechanisms causing legume allergy. As a first step in this direction, we report an analysis of the protein glycosylation of the seed globulins of the model legume *Lotus japonicus*. We suggest that the available genome information and tractable genetics of *Lotus* provides a useful system for comparative studies of plant proteins and their glycosylation in legume food allergy.

MATERIALS AND METHODS

Plant Material, Protein Extraction, and 2-D PAGE

Globulins from mature *Lotus japonicus* Gifu B-129 seeds (kindly provided by Prof. Peter Gresshoff, University of Queensland, Australia) were extracted and separated using 2-D PAGE as previously described.³⁴ In brief, seeds were ground and dissolved

in water and the supernatant was discarded. The pellet was subsequently dissolved in 1 M NaCl and 0.1 M Tris, pH 8.0), centrifuged, and the globulins were precipitated with 10% TFA. The protein pellet was washed twice in 80% acetone and dried.

Site-specific N-glycosylation analysis was performed on the 2-D gel separated globulin glycoproteins. In contrast, the unfractionated globulins were used for the *Lotus* seed globulin identification and N-glycome profiling.

In-solution N-Glycan Release

The dried seed globulins were solubilized in a buffer (2 M urea and 25 mM Tris, pH 7.5) and reduced with 10 mM DTT for 45 min at 56 °C followed by 20 mM iodoacetamide alkylation for 40 min at room temperature in the dark. The sample was diluted to 1 M urea and digested in-solution with 2% w/w of modified trypsin (Promega) overnight at 37 °C, dried, and redissolved in 10 μ L of H₂O. Then the sample was heated at 100 °C for 7 min, dried, and redissolved in 50 mM citrate/phosphate buffer at pH 5.0 containing 18 mU/mL PNGase A and incubated at 37 °C overnight. To isolate the released N-glycans, a homemade C-18 column was made³⁵ and equilibrated with aqueous 0.1% TFA. The flow through of the applied sample containing the N-glycans was collected and pooled with three column washes using aqueous 0.1% TFA. Then, 20 μ L of 100 mM ammonium acetate at pH 5.0 was added for 1 h and the sample was dried. Reduction, desalting, and carbon cleanup was performed as previously described.³⁶ Prior to MS, the dried N-glycans were dissolved in water.

Protein Identifications from the Globulin Fraction

The dried seed globulins were solubilized in either 2 or 7 M urea containing 25 mM Tris, pH 7.5 followed by reduction, alkylation, and in-solution digestion as described above (see Supplemental Figure 1 for the workflow, Supporting Information). The two samples were divided into four aliquots and two of these were deglycosylated with PNGase A. All samples were subjected to a second round of digestion using 2% w/w trypsin overnight at 37 °C. The four aliquots were split in half and desalting was performed for one-half using homemade C18 columns and eluted twice using 2 $\times$ 15 μ L (80% acetonitrile in aqueous 0.1% TFA). The other half was left untouched. All eight samples were then dried down and redissolved in aqueous 0.1% FA for MS.

Deglycosylation, In-gel Digestion, and Glycopeptide Enrichment of Globulins

Protein spots were excised from the 2-D gels and a subset was deglycosylated in gel using PNGase F (New England BioLabs) to verify N-glycosylation. Briefly, the spots were excised and transferred to siliconized tubes and destained at 37 °C for 2 $\times$ 30 min. The samples were dried and 10 μ L of PNGase F in 50 mM NH₄HCO₃ (pH 8.5) was added and incubated at 37 °C for 30 min. Then, 20 μ L of H₂O was added and the samples were incubated overnight at 37 °C, sonicated, and the supernatant was discarded.

Glycosylated and deglycosylated samples were in-gel digested using 6 ng/ μ L trypsin or Asp-N for 1 h at 37 °C (Calbiochem, excision grade) as previously described.³⁷ Subsequently, samples were dried.

Then, a subset of the glycosylated samples were redissolved in 10 μ L (80% acetonitrile in aqueous 2% FA) and loaded onto a homemade ZIC-hydrophilic interaction liquid chromatography (HILIC) (SeQuant) micro column equilibrated with 80% acetonitrile in aqueous 2% FA. The column was washed twice with 80% acetonitrile in aqueous 2% FA and the peptides were

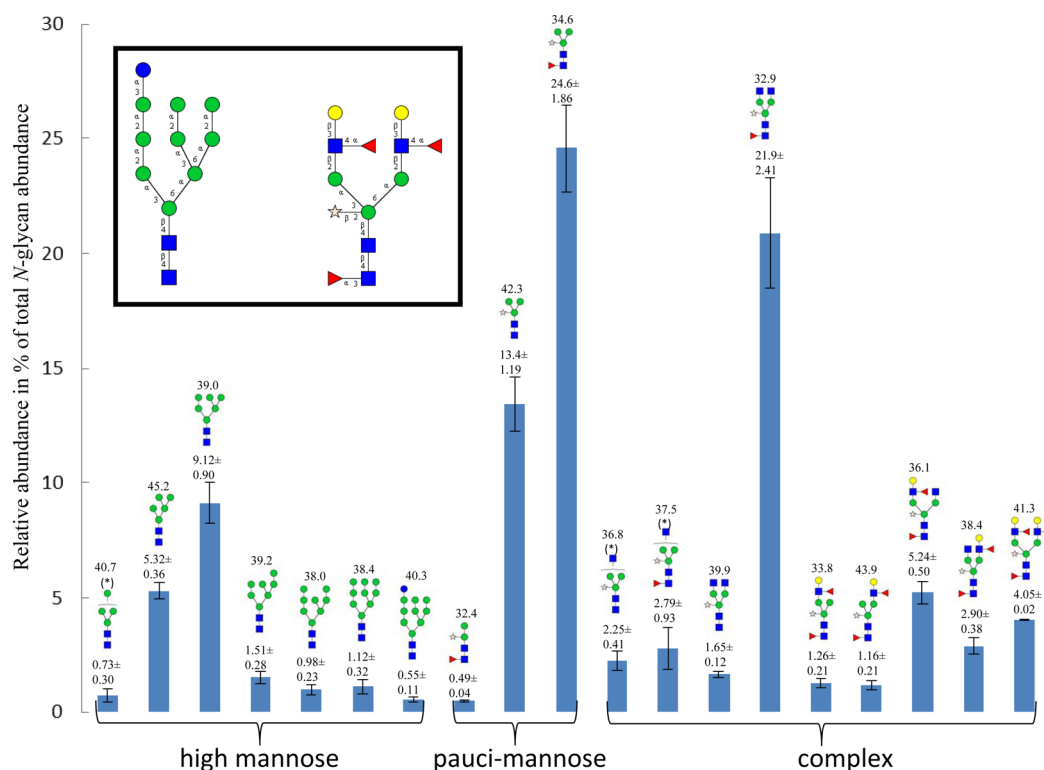


Figure 1. Structure and relative abundance of N-glycans released from *Lotus* seed globulins. In total, 19 different N-glycans were determined from *Lotus* seed globulins. The relative abundance (in molar %) and standard deviation from three biological replicates are indicated below each structure. The elution time (min) from the PGC column is indicated above each structure. The N-glycans are grouped into high mannose, pauci-mannosidic, and complex structures. Using the current knowledge of plant N-glycosylation, the monosaccharide linkages are shown for the two N-glycan extremes in the glycosylation pathway (insert) and are applicable for all 19 N-glycans. * indicates that MS/MS spectra could not be used to determine the arm linkage (i.e., α -1,3 or α -1,6 mannose arm). Consortium of Functional Glycomics (CFG) nomenclature has been used: Blue square, N-acetylglucosamine; red triangle, fucose; white star, xylose; green circle, mannose; yellow circle, galactose; blue circle, glucose.

eluted using 0.8 μ L of 2% FA directly on the MALDI target or dried.

Mass Spectrometry Analyses

For globulin identification, the peptide mixtures were analyzed by electrospray ionization (ESI)–MS/MS in positive polarity mode using a HCT 3-D ion trap (Bruker Daltonics, Newark, DE) coupled to an Ultimate 3000 LC (Dionex). The peptides were separated using a C18 column (300 μ m ID, 10 cm, 3 μ m particles, and pore size at 300 Å, SGE, Australia) equipped with 0.5 μ m peek filter (Upchurch, Oak Harbor, WA). Solvent A (aqueous 0.1% (v/v) FA) was used for equilibration. The flow rate was kept constant at 5 μ L/min. After the sample was loaded onto the column, a gradient (0.8%/min slope) up to 50% solvent B (0.1% FA in acetonitrile) was applied. The column was washed in 80% (v/v) solvent B for 10 min and re-equilibrated in solvent A for 15 min. LC–MS/MS was performed as a full MS scan (m/z 400–2200) followed by a data dependent fragmentation of the two most abundant precursor ions using CID fragmentation. For protein identification, a Mascot (version: 2.3.02) search was performed against the in-house *Lotus* genome version 2.5 (37971 sequences; 9782974 residues) available from <http://www.kazusa.or.jp/lotus/> with the following settings: Carbamidomethylation of Cys was considered a fixed modification, whereas oxidation of Met and deamidation of Asn and Gln were variable modifications. The peptide tolerance was set to 0.6 Da and two missed cleavages were accepted. The MS/MS tolerance was set to 0.6 Da and acceptable charge states were 1+, 2+, and 3+. The significance of protein identification was <0.01 and ion score for

peptide assignment was 0.005. The confident protein identifications were pooled from the eight differently prepared samples and shown in Supplemental Tables 1 and 2, Supporting Information.

Released N-glycans were separated using a Hypercarb porous graphitized carbon (PGC) column (320 μ m ID, 10 cm, and 5 μ m particle size, Thermo Scientific) on an Agilent 1100 capillary LC (Agilent Technologies, Santa Clara, CA) linked to Agilent MSD three-dimensional ion-trap XCT Plus mass spectrometer in negative polarity mode with a constant flow rate of 2 μ L/min. The column was equilibrated in aqueous 10 mM NH_4HCO_3 and a gradient of 2–16% (v/v) 10 mM NH_4HCO_3 in acetonitrile was applied for 45 min, followed by a gradient from 16 to 45% over 20 min before the column was washed in 45% 10 mM NH_4HCO_3 in acetonitrile for 6 min and re-equilibrated in aqueous 10 mM NH_4HCO_3 . MS full scan events (m/z 300–2200) were followed by data-dependent MS/MS scans of the three most intensive precursor ions using CID fragmentation.

For site-specific N-glycan analysis, glycopeptides were enriched using HILIC micro columns as described above and applied to a MALDI target plate where glycopeptides were mixed with 0.5 μ L of 20 mg/mL 2,5-dihydroxybenzoic acid in 70% acetonitrile/0.1% (v/v) TFA. MS and MS/MS spectra of relevant precursor ions were recorded using a 4800 Plus MALDI TOF (AB SCIEX) mass spectrometer, operated in reflector, positive ion mode, and using an acceleration voltage of 20 kV. Laser intensity and number of laser exposures were varied to optimize the spectral quality. The mass range was set to 700–5000 Da. For all MALDI-TOF/TOF MS/MS fragmentations, air

was used as collision gas. The default calibration was used for the samples eluted with 2,5-dihydroxybenzoic acid whereas an external calibration of tryptic lactoglobulin mixed with α -cyano-4-hydroxycinnamic acid was used for samples eluted with α -cyano-4-hydroxycinnamic acid. Data was exported as a text file using Data Explorer (version 4.6) and each spectrum was baseline corrected, smoothed, labeled, and then manually analyzed using M/Z (Genomic Solutions) software. MS/MS spectra of glycosylated peptides were manually annotated.

N-Glycan and N-Glycopeptide Characterization and Quantitation

For the N-glycome analysis, the structural assignment of the individual N-glycans was based on molecular mass and MS/MS fragmentation using the current knowledge of N-glycans determined in plants.⁹ Mass spectra were annotated manually using the DataAnalysis v4.0 (Bruker Daltonics) software. The relative peak area of the extracted ion chromatograms (EICs), corresponding to all observed charge states of the individual N-glycans against the total EIC peak area from negative ion ESI-MS for all N-glycans, was used as a measure for the relative quantitative distribution of N-glycans.

N-glycopeptides were characterized manually based on molecular mass and MS/MS fragmentation. Relative signal intensities from positive ion MALDI MS was used for quantification of N-glycopeptides. Both MS based quantification methods assume that the differently N-glycosylated released glycan and peptide variants ionize equally well in ESI and MALDI MS, respectively, which has been shown to be an accurate assumption, in particular for populations of neutral (nonsialic acid containing) glycans and glycopeptides.^{38–41} To determine the occupancy the peak intensity was compared for the glycosylated and deglycosylated (of +1 Da mass due to Asn to Asp conversion) peptides from untreated and PNGase F treated peptide mixtures (Supplemental Figure 2, Supporting Information). This has been shown to be an accurate approach to establish site occupancy.^{42,43}

Evaluating the Protein Structure Similarity between *Lotus* LCP2 and Peanut Ara h 1

To evaluate the similarity between LCP2 and Ara h 1, polypeptide sequence alignments were made. This was performed using CLUSTALW (<http://www.genome.jp/>). Homologous proteins from other species were included. Sequence identity/similarities between the polypeptide sequences were calculated using BLASTP (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>).

The higher protein structural level of LCP2 and Ara h 1 was also compared. The defined N-glycosylation was included in this comparison. As such, Man₅GlcNAc₂, which was found to occupy the single N-glycosylation site of both LCP2 (this study) and Ara h 1,³⁰ was modeled onto both glycoproteins using GlyProt (<http://www.glycosciences.de>). Default settings were used. The following PDB structures were used as input in GlyProt: Ara h 1: PDB-code 3S7I and LCP2 (accession number CAR78998). The soybean conglycinin: PDB code: 1UIK has the highest primary sequence homology to LCP2, and thus was used as the template for LCP2 structure modeling. Finally, the modeled PDB structures (protein and N-glycans) were visualized with space-fill atomic presentation using RasMol Ver 2.7.5 (RasWin Molecular Graphics).

RESULTS AND DISCUSSION

Global N-Glycan Profiling of *Lotus* Seed Globulins

Food allergy is an increasing problem and legume seed proteins frequently cause allergic responses.^{12,13,31–33} As a first step in

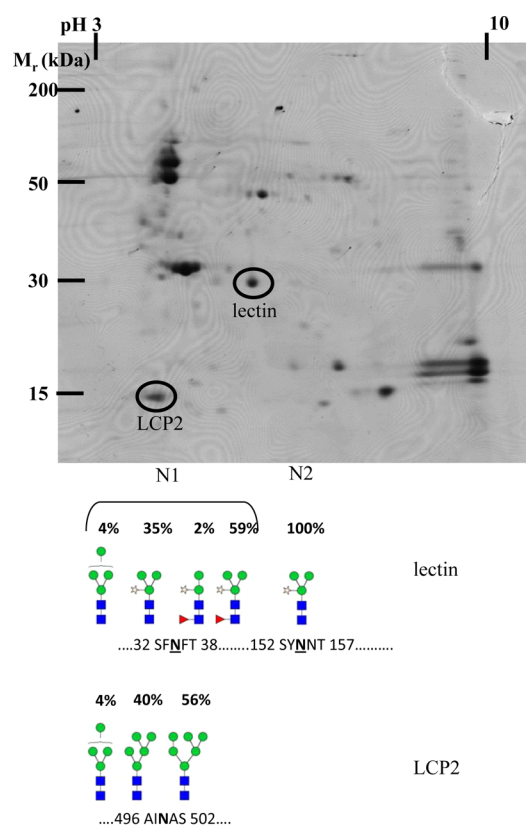


Figure 2. Relative percent molar abundances of protein-specific N-glycans for 2D-gel separated *Lotus* seed globulins. For LCP2, three high mannose N-glycans were observed for the single N-glycosylation site. Numbers above the structures indicate the relative molar occupancy (in %). The predicted lectin (chr5.LjT43D06.200.r2.d) is occupied by four N-glycans at the Asn-35 site (N1) and one N-glycan at the Asn-154 site (N2). MALDI-TOF MS was used to identify the glycopeptides. The pH gradient (nonlinear) and M_r are indicated. CFG nomenclature: Blue square, N-acetylglucosamine; red triangle, fucose; white star, xylose; green circle, mannose.

determining whether glycosylation of legume proteins has a role in food allergy the N-glycome of the *Lotus* seed globulins including the abundant LCP2, the homologue of peanut Ara h 1, was determined. The N-glycans were released in-solution using PNGase A followed by characterization and relative quantification using PGC-LC-MS/MS. The monosaccharide composition, topology and detailed architecture of the N-glycans were determined from MS/MS spectra and the PGC-LC eluting profile. In total, 19 N-glycan structures were identified using the current knowledge of N-glycans determined in plants (Figure 1). For validation purposes, the presence and relative abundances of high mannose and noncore-fucosylated N-glycan subset were additionally compared and confirmed employing PNGase F deglycosylation (Supplemental Figure 3, Supporting Information).³⁶ For the two complex N-glycan monosaccharide compositions, Gal₁Man₃Xyl₁Fuc₂GlcNAc₃ and Gal₁Man₃Xyl₁Fuc₂GlcNAc₄, two isomeric structures were identified. No N-glycan isobaric variants were observed for the

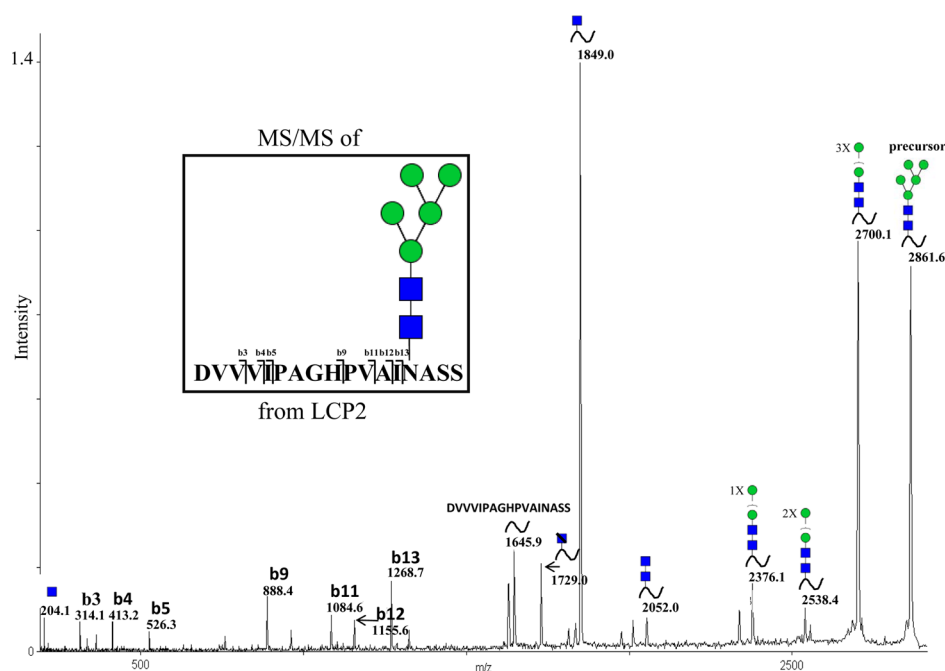


Figure 3. Fragmentation of LCP2 N-glycopeptide carrying $\text{Man}_5\text{GlcNAc}_2$. Fragment spectrum of m/z 2863.0 (precursor), which corresponds to the LCP2 N-glycopeptide with $\text{Man}_5\text{GlcNAc}_2$ attached to the NAS motif (inset). Peptide b-ions and glycopeptide fragments are indicated. CFG nomenclature: Blue square, *N*-acetylglucosamine; blue square with black line, 0,2X-ring cleavage of the innermost *N*-acetylglucosamine;⁶⁰ green circle, mannose.

other monosaccharide compositions. The molar distribution of the N-glycan types was approximately 20% high mannose, 40% pauci-mannosidic, and 40% complex structures. No hybrid structures were identified. In comparison, the global N-glycan profiles of total protein from other legume seeds (i.e., peanut, soybean, pea, and mung bean) have been reported to contain a greater proportion of high mannose structures comprising up to more than 90% of the total N-glycans.⁴⁴

The three most abundant N-glycan compositions on the *Lotus* seed globulins were $\text{GlcNAc}_2\text{Man}_3\text{Xyl}_1\text{Fuc}_1\text{GlcNAc}_2$, $\text{Man}_3\text{Xyl}_1\text{Fuc}_1\text{GlcNAc}_2$, and $\text{Man}_3\text{Xyl}_1\text{GlcNAc}_2$, which are similar to the abundant N-glycans in leaf tissues of other plant species.⁴⁵ All pauci-mannosidic and complex structures from the *Lotus* seed globulins contained the plant-specific β 1,2-xylose and 80% of these were α -1,3-core-fucosylated whereas the high mannose N-glycans were not fucosylated. In the context of food allergy, the high content of the plant-specific α -1,3-fucose and β -1,2-xylose determinants in the *Lotus* globulin fraction may be relevant for triggering allergic reactions either by recognition of the plant specific N-glycans by the immune system^{21–25} and/or by altering the proteolytic degradation of proteins in the gastrointestinal tract.

Site-specific N-Glycan Profiles of LCP2

Considering the N-glycome heterogeneity of *Lotus* seed globulins, site-specific profiling is essential to link the observed N-glycan structures to individual proteins and eventually to function. Three *Lotus* seed globulins containing N-glycosylation motifs have previously been identified using 2-D gels.³⁴ Two of these are predicted to be secreted and, thus, putative glycoproteins, that is, *Lotus* convicilin storage protein 2 (LCP2) and a predicted lectin (chr5.LjT43D06.200.r2.d) containing one and two putative N-glycosylation sites, respectively.

To determine the N-glycosylation of LCP2 we used a previously validated method.⁴⁶ The 2-D gel separated protein (Figure 2) was proteolytically digested, and a fraction of the resulting Asp-N generated peptides were deglycosylated. N-glycosylation of LCP2 was confirmed by comparing the MALDI-TOF MS spectra of the untreated and the deglycosylated peptides, Supplemental Figure 2A, Supporting Information. Signals corresponding to deglycosylated peptides (m/z 1646.8 = predicted mass +1 Da due to deamidated Asn) covering the single N-glycosylation site of LCP2 were observed in the deglycosylated sample and the peptide sequences were confirmed by MS/MS (data not shown). The lack of the nonglycosylated peptides in the untreated sample (m/z 1645.8) indicated full N-glycan occupancy of the LCP2 N-glycosylation site. However, it cannot be excluded that non-glycosylated LCP2 has a different electrophoretic mobility and, thus, is located at different positions in the 2-D gel.

For site-specific N-glycan characterization, the Asp-N generated glycopeptides were enriched using hydrophilic interaction liquid chromatography (HILIC) prior to MALDI-TOF MS. Supplemental Figure 4 (Supporting Information) shows the glycopeptide profiles of LCP2. MS/MS fragment ions of the most abundant glycopeptides precursors confirmed their peptide sequences and the monosaccharide compositions (see Figure 3 for an example showing the MS/MS of the high mannose glycopeptide from LCP2). Three N-glycans were found to occupy the LCP2 N-glycosylation site, that is, $\text{Man}_4\text{GlcNAc}_2$ (~4%), $\text{Man}_5\text{GlcNAc}_2$ (~40%), and $\text{Man}_6\text{GlcNAc}_2$ (~56%). The quantifications were determined from peak intensities in Supplemental Figure 4, Supporting Information. See Figure 2 and Table 1 for an overview of identified N-glycans. The molar ratio of these high mannose structures of LCP2 (as well as the predicted lectin, see below) and the global N-glycan analysis of *Lotus* seed globulins is in good agreement. In both analyses, $\text{Man}_6\text{GlcNAc}_2$ is the most abundant

Table 1. Overview of Observed N-Glycopeptides^a

Lotus protein	Digestion enzyme	N-glycosylated peptide sequence	Experimental m/z	N-glycan structure
lectin	trypsin	VSF N FTK (32-38)	1851.1	
lectin	trypsin	VSF N FTK (32-38)	1867.0	
lectin	trypsin	VSF N FTK (32-38)	1897.2	
lectin	trypsin	VSF N FTK (32-38)	2013.1	
lectin	trypsin	VSF N FTKFTDDGSLILQGDAK (32-52)	3328.0	
lectin	trypsin	VSF N FTKFTDDGSLILQGDAK (32-52)	3473.9	
lectin	trypsin	NGYNQFVAVEFDSY N NTR (140-157)	3162.8	
lectin	Asp-N	VSF N FTKFTDDGSLILQG (32-49)	3013.1	
lectin	Asp-N	VSF N FTKFTDDGSLILQG (32-49)	3159.2	
lectin	Asp-N	DSY N NTR (151-157)	1893.6	
LCP2	Asp-N	DVVVIPAGHPVAI N ASS (486-502)	2700.9	
LCP2	Asp-N	DVVVIPAGHPVAI N ASS (486-502)	2863.0	
LCP2	Asp-N	DVVVIPAGHPVAI N ASS (486-502)	3025.1	

(*) Verified with MS/MS

^aThe N-glycopeptides identified with tryptic or Asp-N digestion of LCP2 and the predicted lectin. The glycopeptide MS spectra are shown in Supplemental Figure 4, Supporting Information. The N-glycan structures with a * were verified with MS/MS fragmentation; the remaining were deduced from molecular mass only. Blue square, N-acetylglucosamine; red triangle, fucose; white star, xylose; green circle, mannose.

high mannose structure. In comparison, the five Ara h 1 N-glycans identified from peanuts were reported to have a similar distribution of these three high mannose N-glycans together with plant specific β -1,2-xylose, that is, Man₆GlcNAc₂ (19%), Man₅GlcNAc₂ (27%), Man₄GlcNAc₂ (5%), Man₄Xyl₁GlcNAc₂ (33%), and Man₃Xyl₁GlcNAc₂ (16%),³⁰ which was not observed in LCP2. The N-glycan composition identified from a French bean vicilin containing two N-glycosylation sites is also mainly high-mannose N-glycosylated.⁴⁷ Here, Man₇GlcNAc₂ and Man₃Xyl₁GlcNAc₂ were reported to be the most abundant N-glycans on one of the glycosylation sites (Asn-228) and Man₉GlcNAc₂ and Man₈GlcNAc₂ were the most abundant N-glycans on the other site (Asn-317).

Homology Modeling Suggests LCP2 as a Potential Allergen

Food allergy, where some of the major allergens are legume seed globulins, is a universal problem with 1–5% of the total human population displaying clinical symptoms.^{48,49} Interestingly, the legume seed globulins seem to be highly conserved in protein sequences between species judging from the high similarity between the Lotus LCP2 and the known globulin allergens from other legumes, that is, mung bean, lupin, lentil, and peanut (Figure 4). All five proteins have a high protein sequence similarity around the single conserved N-glycosylation site (where several of the Ara h 1 allergen epitopes are located) and, together with the similar high mannose glycoproteins,

reinforce the potential for LCP2 to be an allergen.^{17–19} To further support this, *in-silico* digestion of LCP2 and Ara h 1, based on cleavage sites of proteases present in the gastrointestinal tract (i.e., pepsin, trypsin, and chymotrypsin) produces similar peptides (data not shown). LCP2 and Ara h 1 protein sequences are predominantly degraded into single or few amino acid residues by the *in-silico* digestion using the combination of the three proteases. Few peptides of more than 8 amino acid residues were found for both proteins with complete digestion, which is well above the minimum peptide epitope size required for recognition by the immune system. However, these potentially undigested regions were not overlapping with the identified immunogenic epitopes for Ara h 1^{17–19} and, thus, cannot explain why Ara h 1 is an allergen.

Given the high allergenicity and the sequence similarity of Ara h 1 (to LCP2) and the fact that most immunogenic epitopes are located at the surface of the Ara h 1 protein,¹⁶ it is relevant to study the three-dimensional structure of LCP2 in terms of epitope presentation around the N-glycosylation site. As no such structure currently is available, we investigated the three-dimensional structure using the soybean conglycinin (PDB code: 1UIK), which is the closest homologue to LCP2, as a template. The LCP2 homology model and Ara h 1 (PDB code: 3S7I¹⁶) is overall very similar (Figure 5a). In addition, the accessibility of the Asn residue in the glycosylation motif and the local structure of the polypeptide backbone around the N-

complex N-glycan structures using our recent correlation between structural glycosylation site accessibility and N-glycan processing,⁵⁰ which is supported by this data. Taken together, Ara h 1 and LCP2 are similar at the primary and higher protein structure levels and are similarly N-glycosylated. Thus, we suggest that LCP2 is a potential allergen similar to the peanut Ara h 1, and thus that *Lotus* is a potential model system to investigate legume allergies such as peanut allergy. However, further studies using, for example, human allergy tests are needed to confirm whether LCP2 or other *Lotus* seed globulins actually trigger allergic responses in humans.

In addition to LCP2, another putatively N-glycosylated protein, a predicted lectin, was separated and identified on the 2-D gels (Figure 2). This protein was confirmed to have two fully occupied N-glycosylation sites (Table 1, Supplemental Figures 2B, 4B and 4C, Supporting Information) for the analyzed spot. The trypsin or Asp-N derived glycopeptides were separated and analyzed and showed similar site-specific N-glycan profiles; however, a higher number of tryptic glycopeptides were identified (Supplemental Figures 4B and 4C, Supporting Information). In total, four N-glycan structures were observed on the Asn-35 N-glycosylation site (see Supplemental Figure 5 for examples showing the MS/MS of glycopeptides from the predicted lectin, Supporting Information). These include the three pauci-mannosidic N-glycans identified in the global N-glycome of *Lotus* seed globulins and $\text{Man}_4\text{GlcNAc}_2$. In contrast, only $\text{Man}_3\text{Xyl}_1\text{GlcNAc}_2$ was found to decorate the Asn-154 N-glycosylation site. In comparison, a bean cotyledon lectin has similarly been reported to contain high mannose and pauci-mannosidic structures.⁵¹

Interestingly, none of the complex N-glycans identified from the global N-glycome of *Lotus* seed globulins could be found on LCP2 or on the predicted lectin. These two proteins only account for a minor part of the total *Lotus* seed globulin fraction (Figure 2) and, thus, less abundant proteins or proteins which are not separated using 2-D gels might be N-glycosylated with complex structures. Hence, we used an LC-MS/MS based approach to identify other putative glycoproteins from an unfractionated seed globulin fraction, which may be contributing to the complex type N-glycans in the total pool. In total, 12 proteins were identified from eight differently prepared globulin samples, Supplemental Tables 1 and 2, Supporting Information (see Supplemental Figure 1 for experimental setup, Supporting Information). This included LCP2, the predicted lectin, and some other secreted non N-glycosylated common globulins (LCP1, LLP2, LLP3, LLP5) observed previously from mature *Lotus* seeds.^{34,52} Interestingly, one additional secreted protein was identified: a predicted globulin peptidase (chr1.CM0544.570.r2.d). This protein contains a highly surface exposed N-glycosylation motif as estimated from a homology model based on the predicted three-dimensional structure of a close homologue from *Glycine max* (PDB code: 3AUP, 80% sequence similarity).^{50,53} A similar peptidase from barley has been shown to be N-glycosylated.⁵⁴ Taken together, the *Lotus* predicted peptidase is an excellent candidate to carry the missing complex N-glycan structures. Unfortunately, we were unable to

glycosylation site seem to be similarly exposed for the two proteins (Figure 5b). Man₅GlcNAc₂, which is identified as present on this site in both LCP2 and Ara h 1,³⁰ is here modeled onto the protein structures. The similar N-glycan decoration of LCP2 and Ara h 1 is in good agreement with similar surface accessibilities of their N-glycosylation sites. Both LCP2 and Ara h 1 would be predicted to have incompletely processed N-glycans, that is, intermediate between high mannose and fully processed

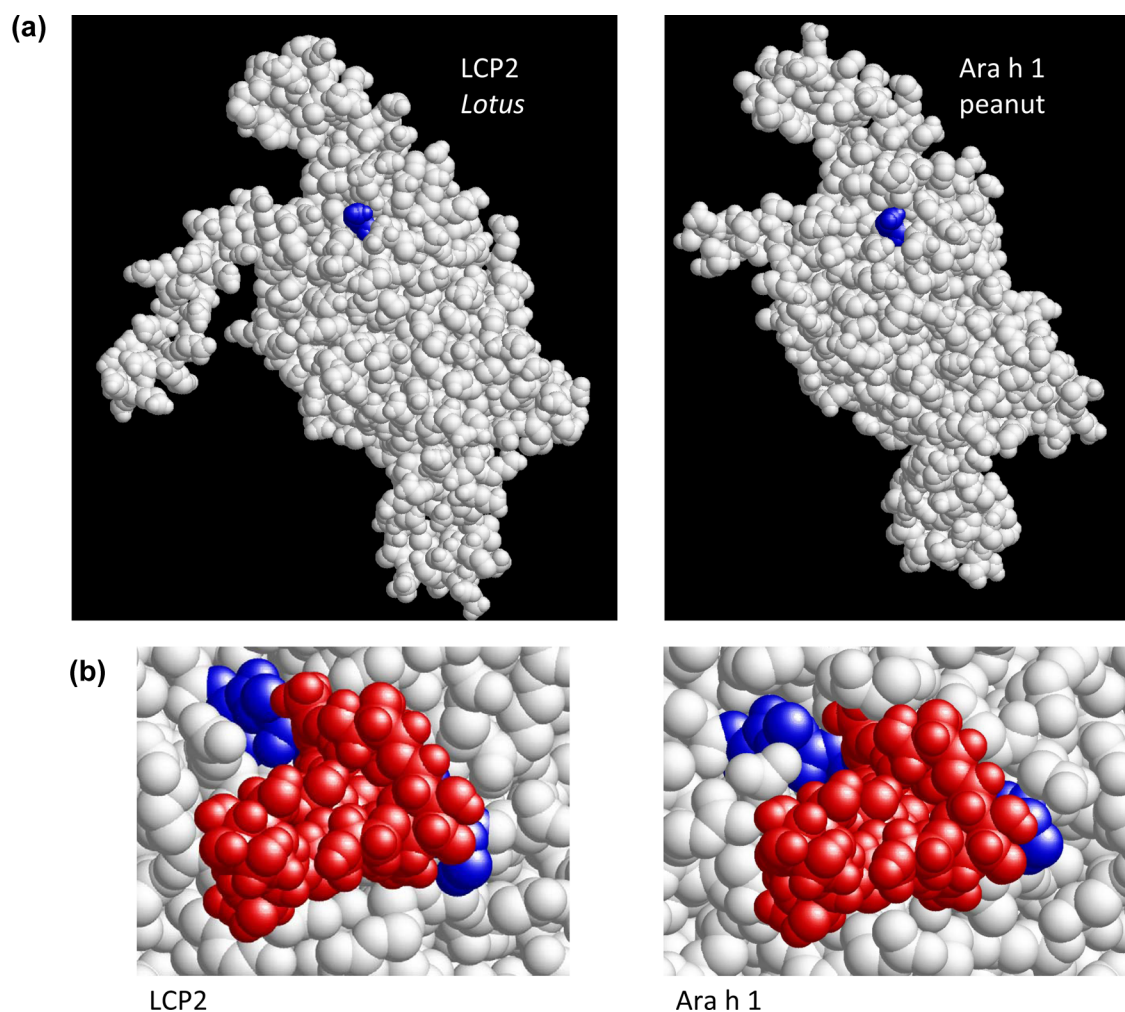


Figure 5. Comparison of *Lotus* LCP2 (homology model) and peanut Ara h 1 tertiary protein structure. The LCP2 homology model is based on the determined structure of the homologous soybean conglycinin (PDB code: 1UIK). For Ara h 1, a previously determined structure was used (PDB code: 3S7I).¹⁶ (a) Monomeric structure of LCP2 and Ara h 1, where the Asn in the N-glycosylation site is in blue. (b) Zoom of the N-glycosylation site where the Val-Ala-Ile-Asn-Ala-Ser-Ser amino acid sequence is in blue and the determined N-glycan for both proteins, Man₅GlcNAc₂, is modeled in red.

confirm this further. In addition, a predicted glucosidase (LjSGA_063286.1) was identified from the *Lotus* seed globulin preparations. Using blastp this *Lotus* glucosidase gene is only partial annotated. However, glucosidase proteins are normally N-glycosylated⁵⁵ and, thus, the *Lotus* glucosidase, which contains a N-glycan motif, is also a candidate to carry N-glycosylation.

This study, and the absence of complete correlation between the global glycans released from the proteins and the peptides found to be glycosylated, stresses the importance of complementing the N-glycome with site-specific N-glycan analysis to obtain an accurate N-glycosylation analysis even for simple protein mixtures such as the *Lotus* seed globulin fraction.

CONCLUSION

This is the first study to describe protein N-glycosylation in *Lotus*. Here we have focused on the global and protein-specific N-glycosylation of the seed globulins. The seed globulin fraction was chosen as a subproteome since it accounts for a large proportion of the total *Lotus* seed proteome and since the globulin homologues in other legumes are known allergens. Heterogeneous mixtures of plant-specific N-glycans were observed on *Lotus* seed globulins. These were similar in structure to N-glycans from other crop legumes. This N-glycome was

carried by a surprisingly small fraction of the *Lotus* seed globulins. The two glycoproteins characterized in detail, a putative lectin and LCP2, interestingly had completely different glycosylation profiles. Pauci-mannose structures with xylose and core fucose decorated the lectin (Asn-154 carrying only a single structure), while only high mannose structures were found on LCP2. The *Lotus* LCP2 was closely related to the major peanut allergen Ara h 1 at the primary and higher protein structural levels as well as at the glycosylation level. Studies in a model legume like *Lotus* might in this context serve to link individual glycans or glycan substructures (e.g., α -1,3-fucose or β -1,2-xylose determinants) to immunogenic responses causing allergy in humans. The diploid *Lotus* is an established model system,⁵⁶ the genome is sequenced, and a large population of insertion mutants, generated with a LORE1 retrotransposon, is available.^{57–59} These resources and the easier genetic manipulations compared to peanut plants provide an advantage for the analysis of cellular processes and the role of protein modifications. Thus, the *Lotus* mutants available^{57,58} and the homology between *Lotus* proteins and peanut allergens shown in this study provide an opportunity for a more systematic study to correlate food allergy with protein N-glycosylation. Ultimately, this can be the basis for the breeding of

less allergenic crops and for production of less allergenic biopharmaceutical glycoproteins.

■ ASSOCIATED CONTENT

● Supporting Information

Five supplemental figures and two supplemental tables. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +45 8715 5504. Fax: +45 8612 3178. E-mail: stougaard@mb.au.dk

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Peter Gresshoff (University of Queensland, Australia) for providing wild type *Lotus* seeds and Dr. Pia H. Jensen (University of Southern Denmark, Denmark) for a fruitful discussion of N-glycans in general. This work was supported by the Danish National Research Foundation grant no. DNRF79 and the Danish Council for Independent Research | Technology and Production Sciences (FTP).

■ REFERENCES

- (1) Sumer-Bayraktar, Z.; Kolarich, D.; Campbell, M. P.; Ali, S.; Packer, N. H.; Thaysen-Andersen, M. N-glycans modulate the function of human corticosteroid-binding globulin. *Mol. Cell. Proteomics* **2011**, *10*, M111.009100–1–M111.009100–14.
- (2) Anthony, R. M.; Ravetch, J. V. A Novel role for the IgG Fc glycan: the anti-inflammatory activity of sialylated IgG Fcs. *J. Clin. Immunol.* **2010**, *30*, S9–14.
- (3) Ohtsubo, K.; Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **2006**, *126*, 855–867.
- (4) Haltiwanger, R. S.; Lowe, J. B. Role of glycosylation in development. *Annu. Rev. Biochem.* **2004**, *73*, 491–537.
- (5) Royle, L.; Roos, A.; Harvey, D. J.; Wormald, M. R.; van Gijlswijk-Janssen, D.; Redwan, e. M.; Wilson, I. A.; Daha, M. R.; Dwek, R. A.; Rudd, P. M. Secretory IgA N- and O-glycans provide a link between the innate and adaptive immune systems. *J. Biol. Chem.* **2003**, *278*, 20140–20153.
- (6) Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **1983**, *209*, 331–336.
- (7) Altmann, F.; Fabini, G.; Ahorn, H.; Wilson, I. B. Genetic model organisms in the study of N-glycans. *Biochimie* **2001**, *83*, 703–712.
- (8) Wilson, I. B. Glycosylation of proteins in plants and invertebrates. *Curr. Opin. Struct. Biol.* **2002**, *12*, 569–577.
- (9) Faye, L.; Boulaflous, A.; Benchabane, M.; Gomord, V.; Michaud, D. Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine* **2005**, *23*, 1770–1778.
- (10) Verma, A. K.; Kumar, S.; Das, M.; Dwivedi, P. D. A comprehensive review of legume allergy. *Clin. Rev. Allergy Immunol.* **2012**, DOI: 10.1007/s12016-012-8310-6.
- (11) Natarajan, S. S.; Xu, C.; Bae, H.; Caperna, T. J.; Garrett, W. M. Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *J. Agric. Food Chem.* **2006**, *54*, 3114–3120.
- (12) Jiang, S.; Wang, S.; Sun, Y.; Zhou, Z.; Wang, G. Molecular characterization of major allergens Ara h 1, 2, 3 in peanut seed. *Plant Cell Rep.* **2011**, *30*, 1135–1143.
- (13) Kroghsbo, S.; Bogh, K. L.; Rigby, N. M.; Mills, E. N.; Rogers, A.; Madsen, C. B. Sensitization with 7S globulins from peanut, hazelnut, soy

or pea induces IgE with different biological activities which are modified by soy tolerance. *Int. Arch. Allergy Immunol.* **2011**, *155*, 212–224.

(14) Astwood, J. D.; Leach, J. N.; Fuchs, R. L. Stability of food allergens to digestion *in vitro*. *Nat. Biotechnol.* **1996**, *14*, 1269–1273.

(15) Maleki, S. J.; Kopper, R. A.; Shin, D. S.; Park, C. W.; Compadre, C. M.; Sampson, H.; Burks, A. W.; Bannon, G. A. Structure of the major peanut allergen Ara h 1 may protect IgE-binding epitopes from degradation. *J. Immunol.* **2000**, *164*, 5844–5849.

(16) Chruszcz, M.; Maleki, S. J.; Majorek, K. A.; Demas, M.; Bublin, M.; Solberg, R.; Hurlburt, B. K.; Ruan, S.; Mattison, C. P.; Breiteneder, H.; Minor, W. Structural and immunologic characterization of Ara h 1, a major peanut allergen. *J. Biol. Chem.* **2011**, *286*, 39318–39327.

(17) Burks, A. W.; Shin, D.; Cockrell, G.; Stanley, J. S.; Helm, R. M.; Bannon, G. A. Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur. J. Biochem.* **1997**, *245*, 334–339.

(18) Cong, Y.; Lou, F.; Xue, W.; Li, L.; Chena, M. Characterisation of the IgE-binding immunodominant epitopes on Ara h 1. *Food Agric. Immunol.* **2008**, *19*, 175–185.

(19) Shinmoto, H.; Takeda, M.; Matsuo, Y.; Naganawa, Y.; Tomita, S.; Takano-Ishikawa, Y. Epitope analysis of peanut allergen Ara h 1 with human monoclonal IgM antibody clone #86. *Hum. Antibodies* **2010**, *19*, 101–105.

(20) Duranti, M.; Guerrieri, N.; Takahashi, T.; Cerletti, P. The legumin-like storage protein of *Lupinus albus* seeds. *Phytochemistry* **1988**, *27*, 15–23.

(21) Batanero, E.; Crespo, J. F.; Monsalve, R. I.; Martin-Esteban, M.; Villalba, M.; Rodriguez, R. IgE-binding and histamine-release capabilities of the main carbohydrate component isolated from the major allergen of olive tree pollen, Ole e 1. *J. Allergy Clin. Immunol.* **1999**, *103*, 147–153.

(22) Mari, A.; Iacovacci, P.; Afferni, C.; Barletta, B.; Tinghino, R.; Di Felice, G.; Pini, C. Specific IgE to cross-reactive carbohydrate determinants strongly affect the *in vitro* diagnosis of allergic diseases. *J. Allergy Clin. Immunol.* **1999**, *103*, 1005–1011.

(23) Hiemori, M.; Bando, N.; Ogawa, T.; Shimada, H.; Tsuji, H.; Yamanishi, R.; Terao, J. Occurrence of IgE antibody-recognizing N-linked glycan moiety of a soybean allergen, Gly m bd 28K. *Int. Arch. Allergy Immunol.* **2000**, *122*, 238–245.

(24) Iacovacci, P.; Afferni, C.; Butteroni, C.; Pironi, L.; Puggioni, E. M.; Orlandi, A.; Barletta, B.; Tinghino, R.; Ariano, R.; Panzani, R. C.; Di Felice, G.; Pini, C. Comparison between the native glycosylated and the recombinant Cup a 1 allergen: role of carbohydrates in the histamine release from basophils. *Clin. Exp. Allergy* **2002**, *32*, 1620–1627.

(25) Bencurova, M.; Hemmer, W.; Focke-Tejkl, M.; Wilson, I. B.; Altmann, F. Specificity of IgG and IgE antibodies against plant and insect glycoprotein glycans determined with artificial glycoforms of human transferrin. *Glycobiology* **2004**, *14*, 457–466.

(26) Sourrouille, C.; Marquet-Blouin, E.; D'Aoust, M. A.; Kiefer-Meyer, M. C.; Seveno, M.; Pagny-Salehabadi, S.; Bardor, M.; Durambur, G.; Lerouge, P.; Vezina, L.; Gomord, V. Down-regulated expression of plant-specific glycoepitopes in alfalfa. *Plant. Biotechnol. J.* **2008**, *6*, 702–721.

(27) Matsuo, K.; Matsumura, T. Deletion of fucose residues in plant N-glycans by repression of the *GDP-Mannose 4,6-Dehydratase* gene using virus-induced gene silencing and RNA interference. *Plant. Biotechnol. J.* **2011**, *9*, 264–281.

(28) Kaulfürst-Soboll, H.; Rips, S.; Koiwa, H.; Kajiura, H.; Fujiyama, K.; von Schaewen, A. Reduced immunogenicity of *Arabidopsis* Hg11 mutant N-glycans caused by altered accessibility of xylose and core fucose epitopes. *J. Biol. Chem.* **2011**, *286*, 22955–22964.

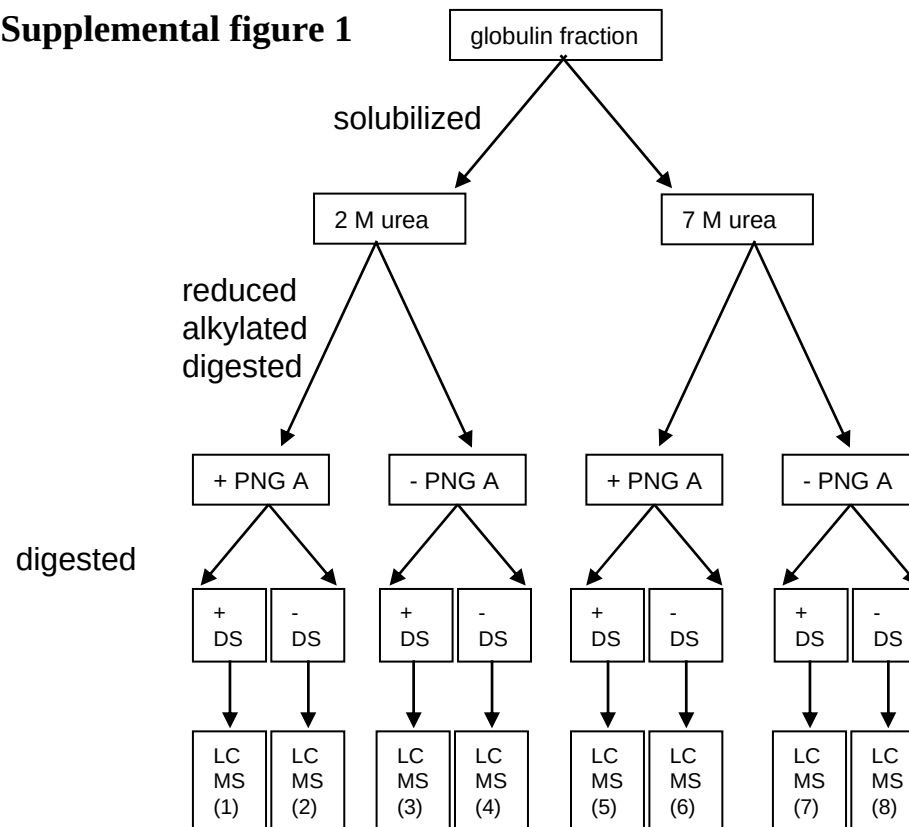
(29) Altmann, F. The role of protein glycosylation in allergy. *Int. Arch. Allergy Immunol.* **2007**, *142*, 99–115.

(30) Kolarich, D.; Altmann, F. N-glycan analysis by matrix-assisted laser desorption/ionization mass spectrometry of electrophoretically separated nonmammalian proteins: application to peanut allergen Ara h 1 and olive pollen allergen Ole e 1. *Anal. Biochem.* **2000**, *285*, 64–75.

(31) Branum, A. M.; Lukacs, S. L. Food allergy among children in the United States. *Pediatrics* **2009**, *124*, 1549–1555.

- (32) Sicherer, S. H.; Munoz-Furlong, A.; Sampson, H. A. Prevalence of peanut and tree nut allergy in the United States determined by means of a random digit dial telephone survey: a 5-year follow-up study. *J. Allergy Clin. Immunol.* **2003**, *112*, 1203–1207.
- (33) Grundy, J.; Matthews, S.; Bateman, B.; Dean, T.; Arshad, S. H. Rising prevalence of allergy to peanut in children: data from 2 sequential cohorts. *J. Allergy Clin. Immunol.* **2002**, *110*, 784–789.
- (34) Dam, S.; Laursen, B. S.; Ørnfelt, J. H.; Jochimsen, B.; Staerfeldt, H. H.; Friis, C.; Nielsen, K.; Goffard, N.; Besenbacher, S.; Krusell, L.; Sato, S.; Tabata, S.; Thøgersen, I. B.; Enghild, J. J.; Stougaard, J. The proteome of seed development in the model legume *Lotus japonicus*. *Plant Physiol.* **2009**, *149*, 1325–1340.
- (35) Gobom, J.; Nordhoff, E.; Mirgorodskaya, E.; Ekman, R.; Roepstorff, P. Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.* **1999**, *34*, 105–116.
- (36) Jensen, P. H.; Karlsson, N. G.; Kolarich, D.; Packer, N. H. Structural analysis of N- and O-glycans released from glycoproteins. *Nat. Protoc.* **2012**, *7*, 1299–1310.
- (37) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **1996**, *68*, 850–858.
- (38) Thaysen-Andersen, M.; Mysling, S.; Højrup, P. Site-specific glycoproteomics of N-linked glycopeptides using MALDI-TOF MS: strong correlation between signal strength and glycoform quantities. *Anal. Chem.* **2009**, *81*, 3933–3943.
- (39) Harmon, B. J.; Gu, X.; Wang, D. I. Rapid monitoring of site-specific glycosylation microheterogeneity of recombinant human interferon-gamma. *Anal. Chem.* **1996**, *68*, 1465–1473.
- (40) Gil, G. C.; Kim, Y. G.; Kim, B. G. A relative and absolute quantification of neutral N-linked oligosaccharides using modification with carboxymethyl trimethylammonium hydrazide and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **2008**, *379*, 45–59.
- (41) Harvey, D. J. Quantitative aspects of the matrix-assisted laser desorption mass spectrometry of complex oligosaccharides. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 614–619.
- (42) Zhang, S.; Sherwood, R. W.; Yang, Y.; Fish, T.; Chen, W.; McCardle, J. A.; Jones, R. M.; Yusibov, V.; May, E. R.; Rose, J. K.; Thannhauser, T. W. Comparative characterization of the glycosylation profiles of an influenza hemagglutinin produced in plant and insect hosts. *Proteomics* **2012**, *12*, 1269–1288.
- (43) Stavenhagen, K.; Hinnenburg, H.; Thaysen-Andersen, M.; Hartmann, L.; Silva, D. V.; Fuchser, J.; Kaspar, S.; Rapp, E.; Seiberger, P. H.; Kolarich, D. Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *J. Mass Spectrom.* **2013**, *48*, 627–639.
- (44) Wilson, I. B.; Zeleny, R.; Kolarich, D.; Staudacher, E.; Stroop, C. J.; Kamerling, J. P.; Altmann, F. Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core alpha1,3-linked fucose and xylose substitutions. *Glycobiology* **2001**, *11*, 261–274.
- (45) Nagels, B.; Santens, F.; Weterings, K.; Van Damme, E. J.; Callewaert, N. Improved sample preparation for CE-LIF Analysis of plant N-glycans. *Electrophoresis* **2011**, *32*, 3482–3490.
- (46) Broghammer, A.; Krusell, L.; Blaise, M.; Sauer, J.; Sullivan, J. T.; Maolanon, N.; Vinther, M.; Lorentzen, A.; Madsen, E. B.; Jensen, K. J.; Roepstorff, P.; Thirup, S.; Ronson, C. W.; Thygesen, M. B.; Stougaard, J. Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 13859–13864.
- (47) Kimura, A.; Tandang-Silvas, M. R.; Fukuda, T.; Cabanos, C.; Takegawa, Y.; Amano, M.; Nishimura, S.; Matsumura, Y.; Utsumi, S.; Maruyama, N. Carbohydrate moieties contribute significantly to the physicochemical properties of french bean 7S globulin phaseolin. *J. Agric. Food Chem.* **2010**, *58*, 2923–2930.
- (48) Rona, R. J.; Keil, T.; Summers, C.; Gislason, D.; Zuidmeer, L.; Sodergren, E.; Sigurdardottir, S. T.; Lindner, T.; Goldhahn, K.; Dahlstrom, J.; McBride, D.; Madsen, C. The prevalence of food allergy: a meta-analysis. *J. Allergy Clin. Immunol.* **2007**, *120*, 638–646.
- (49) http://www.foodallergens.info/Facts/How_Many.html (last accessed June 20, 2013).
- (50) Thaysen-Andersen, M.; Packer, N. H. Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. *Glycobiology* **2012**, *22*, 1440–1452.
- (51) Bardor, M.; Loutelier-Bourhis, C.; Marvin, L.; Cabanes-Macheteau, M.; Lange, C.; Lerouge, P.; Faye, L. Analysis of plant glycoproteins by matrix-assisted laser desorption ionisation mass spectrometry: application to the N-glycosylation analysis of bean phytohemagglutinin. *Plant Physiol. Biochem.* **1999**, *37*, 319–325.
- (52) Nautrup-Pedersen, G.; Dam, S.; Laursen, B. S.; Siegmundfeldt, A. L.; Nielsen, K.; Goffard, N.; Staerfeldt, H. H.; Friis, C.; Sato, S.; Tabata, S.; Lorentzen, A.; Roepstorff, P.; Stougaard, J. Proteome analysis of pod and seed development in the model legume *Lotus japonicus*. *J. Proteome Res.* **2010**, *9*, 5715–5726.
- (53) Yoshizawa, T.; Shimizu, T.; Yamabe, M.; Taichi, M.; Nishiuchi, Y.; Shichijo, N.; Unzai, S.; Hirano, H.; Sato, M.; Hashimoto, H. Crystal structure of basic 7S globulin, a xyloglucan-specific endo-beta-1,4-glucanase inhibitor protein-like protein from soybean lacking inhibitory activity against endo-beta-glucanase. *FEBS J.* **2011**, *278*, 1944–1954.
- (54) Cambra, I.; Martinez, M.; Dader, B.; Gonzalez-Melendi, P.; Gandullo, J.; Santamaria, M. E.; Diaz, I. A cathepsin F-like peptidase involved in barley grain protein mobilization, HvPap-1, is modulated by its own propeptide and by cystatins. *J. Exp. Bot.* **2012**, *63*, 4615–4629.
- (55) Cicek, M.; Esen, A. Expression of soluble and catalytically active plant (monocot) beta-glucosidases in *E. Coli*. *Biotechnol. Bioeng.* **1999**, *63*, 392–400.
- (56) Handberg, K.; Stougaard, J. *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J.* **1992**, *2*, 487–496.
- (57) Urbanski, D.; Malolepszy, A.; Stougaard, J.; Andersen, S. Genome-wide *LORE1* retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *Plant J.* **2011**, *69*, 731–741.
- (58) Fukai, E.; Soyano, T.; Umehara, Y.; Nakayama, S.; Hirakawa, H.; Tabata, S.; Sato, S.; Hayashi, M. Establishment of a *Lotus japonicus* gene tagging population using the exon-targeting endogenous retrotransposon *LORE1*. *Plant J.* **2012**, *69*, 720–730.
- (59) Sato, S.; Nakamura, Y.; Kaneko, T.; Asamizu, E.; Kato, T.; Nakao, M.; Sasamoto, S.; Watanabe, A.; Ono, A.; Kawashima, K.; Fujishiro, T.; Katoh, M.; Kohara, M.; Kishida, Y.; Minami, C.; Nakayama, S.; Nakazaki, N.; Shimizu, Y.; Shinpo, S.; Takahashi, C.; Wada, T.; Yamada, M.; Ohmido, N.; Hayashi, M.; Fukui, K.; Baba, T.; Nakamichi, T.; Mori, H.; Tabata, S. Genome structure of the legume, *Lotus japonicus*. *DNA Res.* **2008**, *15* (4), 227–239.
- (60) Wührer, M.; Catalina, M. I.; Deelder, A. M.; Hokke, C. H. Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2007**, *849*, 115–128.

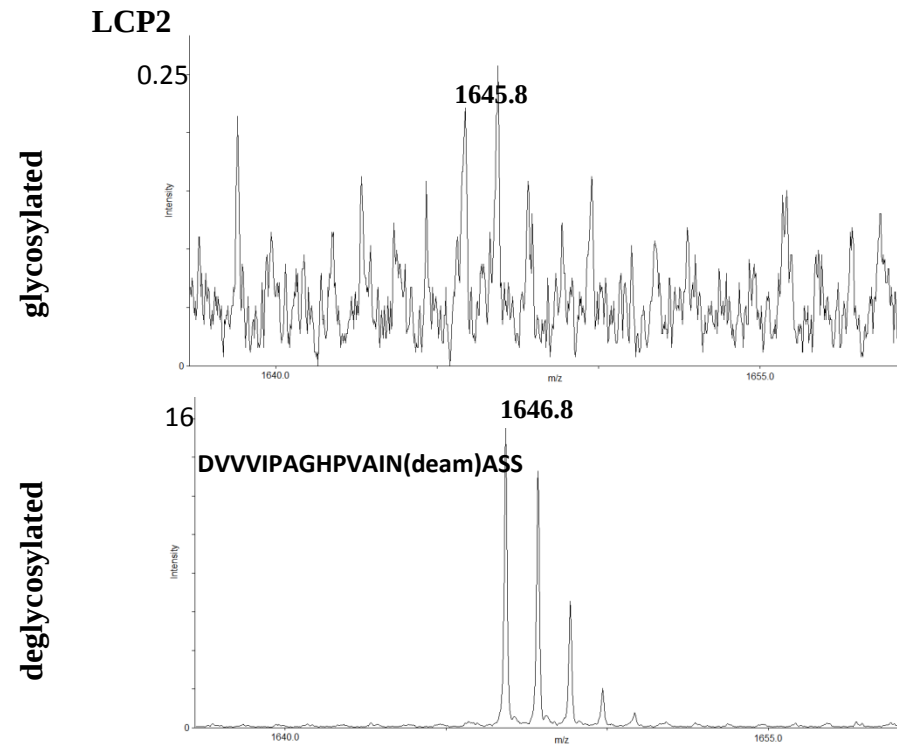
Supplemental figure 1



Supplemental figure 1: Workflow for protein identification in the *Lotus* seed globulin fraction. The globulin fraction was solubilized in either 2 or 7 M urea, in the next step samples were split and *N*-glycan released using PNGase A, in half of the samples, subsequently half of these samples were desalted. The eight different samples for LC-MS are marked 1-8 which corresponds to the protein identification list given in Supplemental table 2. For a further description of the procedure, see materials and methods.

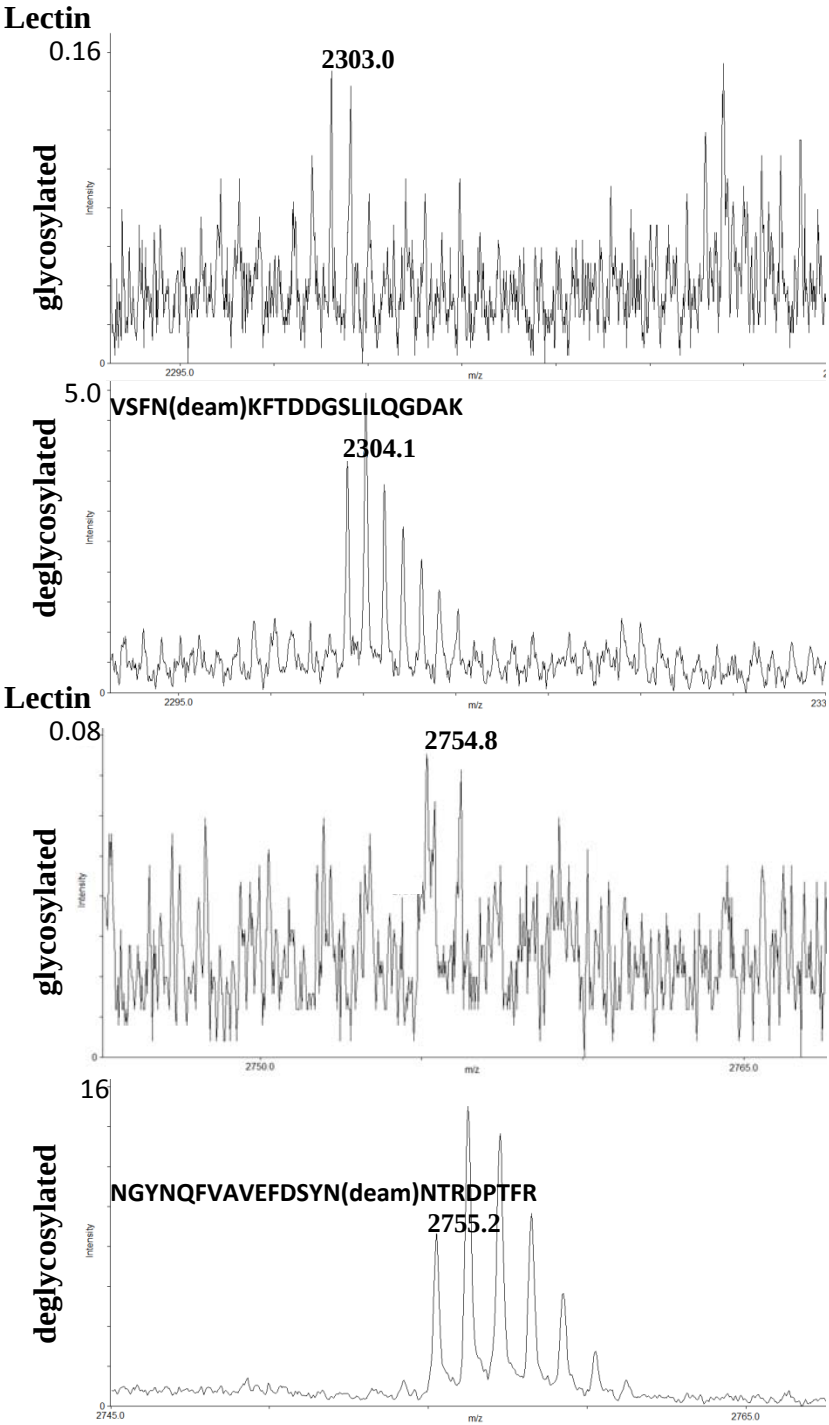
PNG A; PNGase A, DS; desalting.

Supplemental figure 2A

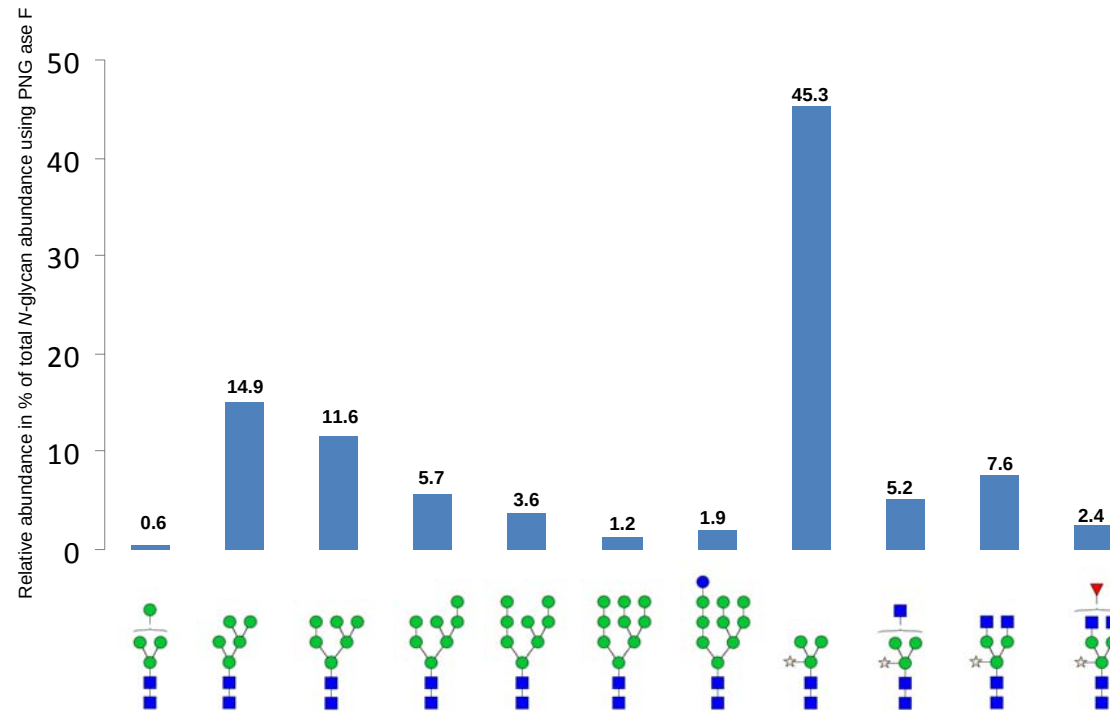


Supplemental figure 2: Informative MS spectra of glycosylated and deglycosylated LCP2 and predicted lectin. A and B) The three ions at m/z 1646.8, 2304.1, and 2755.2 in the deglycosylated samples corresponding to peptides with an *N*-glycosylation motif with deaminated Asn (peptide sequences verified by MS/MS), whereas for the three glycosylated samples only noise is found in that m/z range. LCP2 was in-gel digested with Asp-N and the predicted lectin was *in-gel* digested with trypsin.

Supplemental figure 2B



Supplemental figure 3

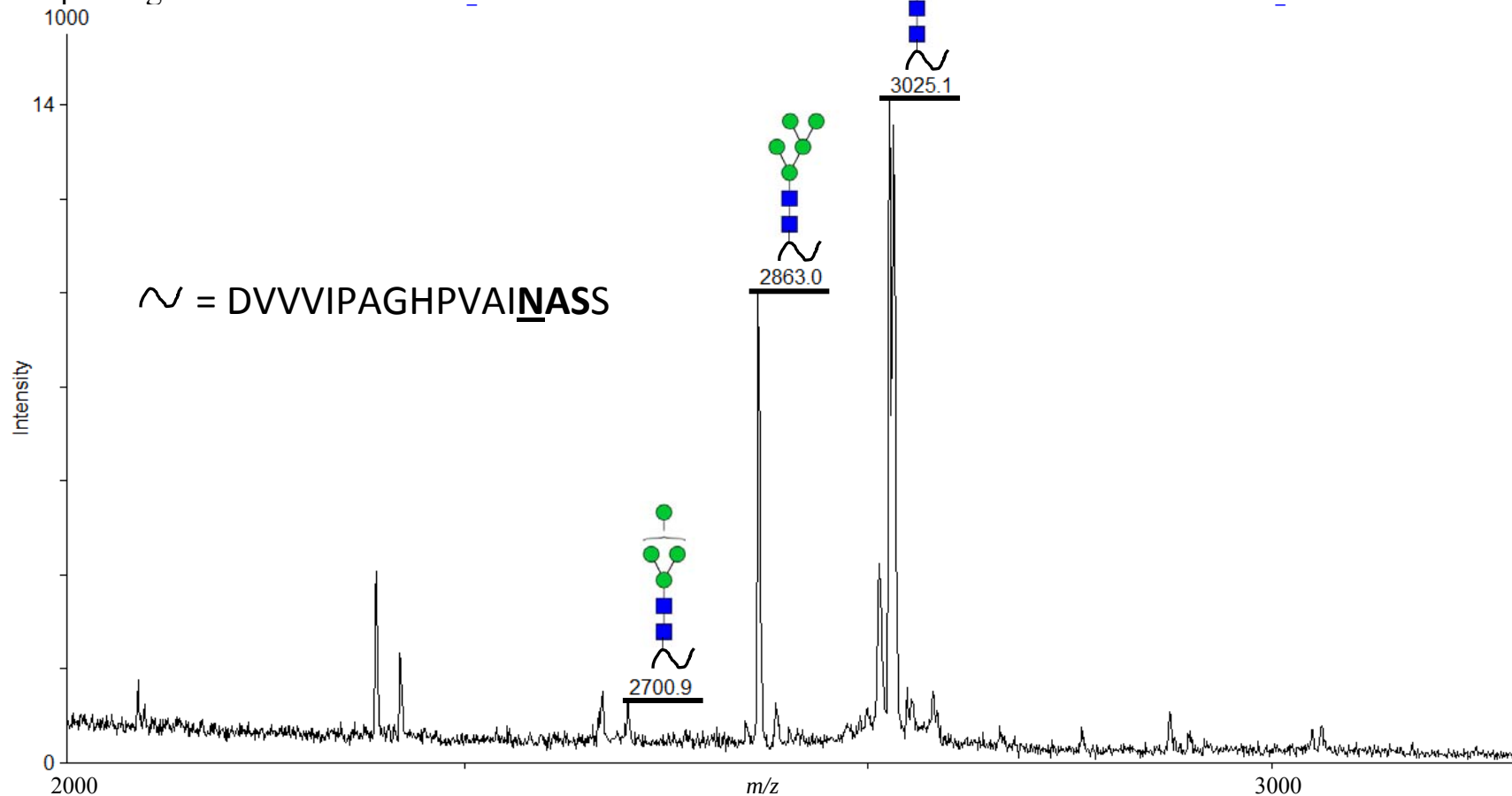


Supplemental figure 3: Structure and relative abundance of N-glycans released from *Lotus* seed globulins using PNGase F.

The relative abundance (in molar %) of the 11 different N-glycans identified are indicated for each structure.

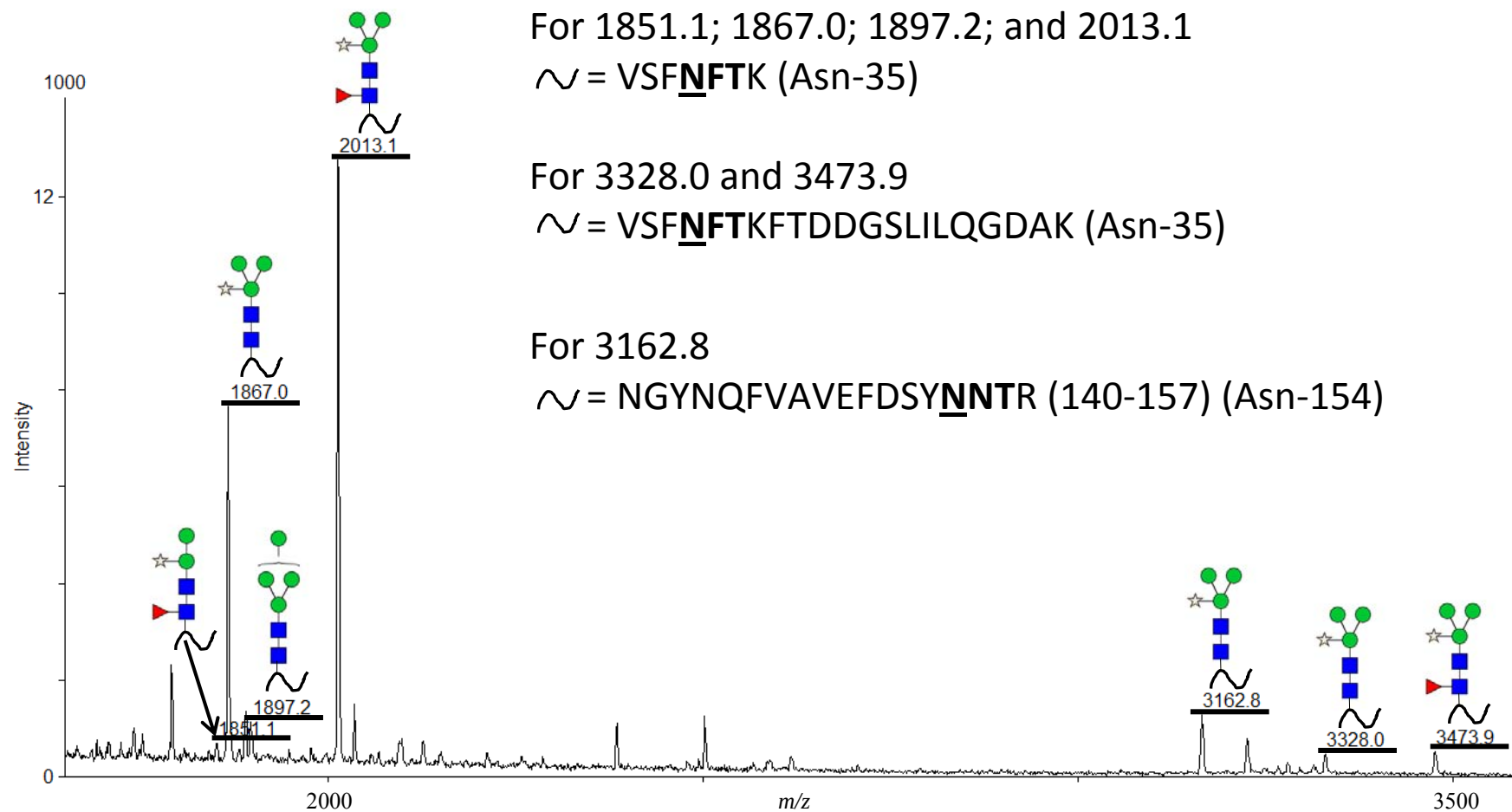
CFG nomenclature has been used: Blue square; N-acetylglucosamine, red triangle; fucose, white star; xylose, green circle; mannose, yellow circle; galactose, blue circle; glucose.

Supplemental figure 4A
Asp-N digestion of LCP2

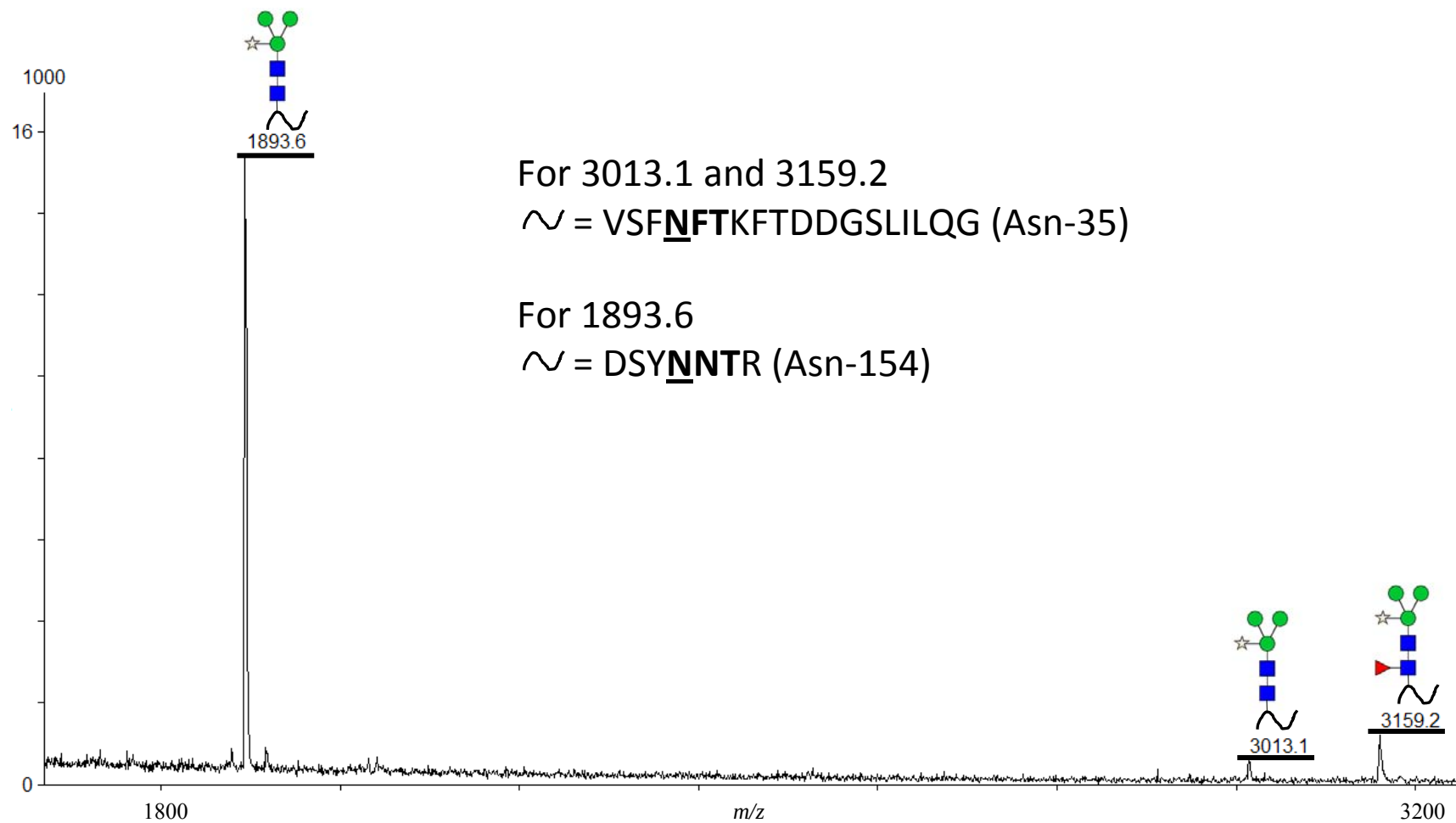


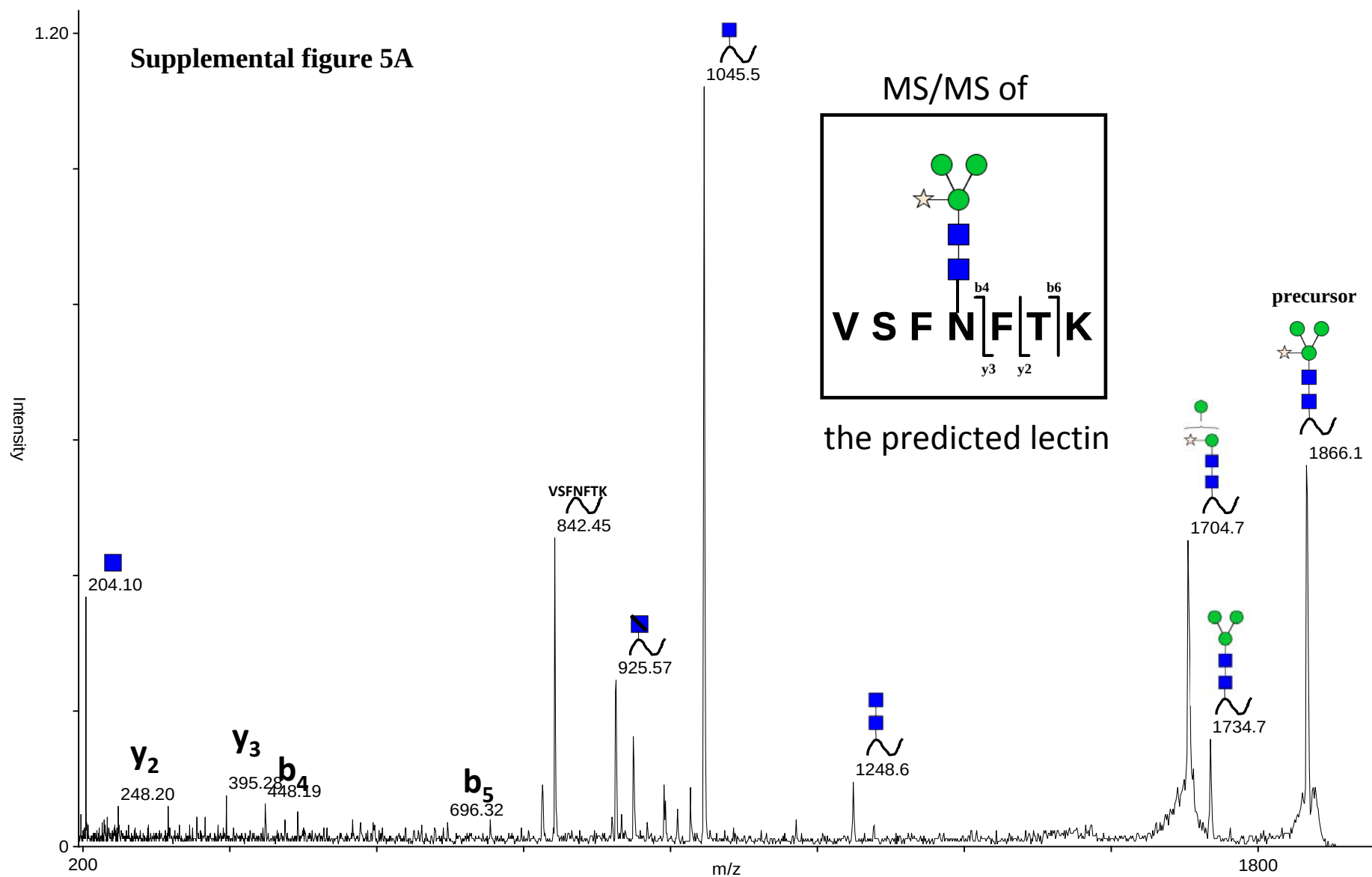
Supplemental figure 4: MS spectra of HILIC enriched trypsin or Asp-N derived glycopeptides from LCP2 and the predicted lectin. A, B, and C) m/z values underlined corresponding to glycopeptide signals. The *N*-glycan structure corresponding to each signal is shown and the peak intensity is used for relative *N*-glycan occupancy for each glycosylation site shown in figure 5. Blue square; *N*-acetylglucosamine, red triangle; fucose, white star; xylose, green circle; mannose.

Supplemental figure 4B
tryptic digestion of lectin



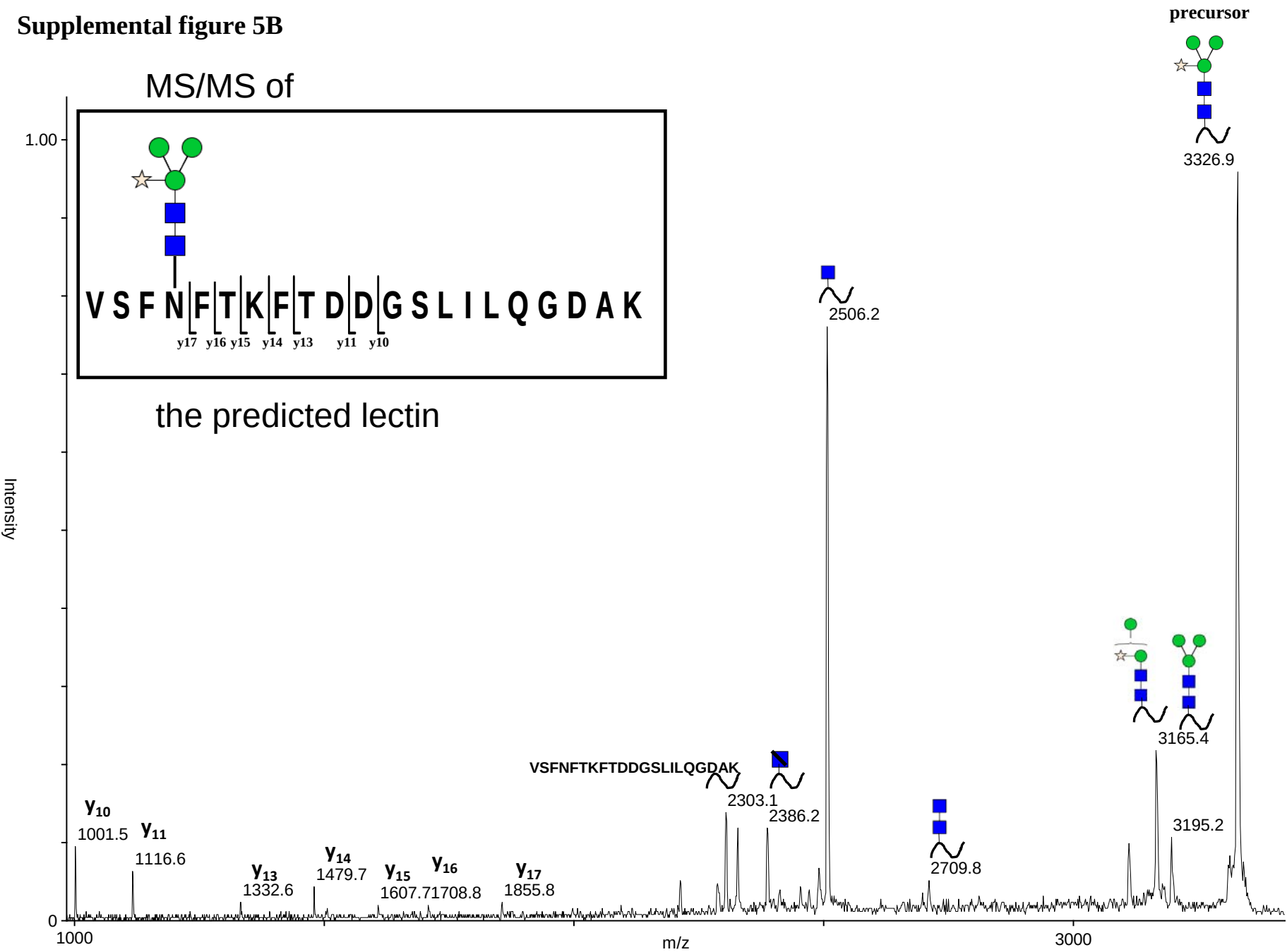
Supplemental figure 4C
Asp-N digestion of lectin





Supplemental figure 5: Fragmentation of the lectin *N*-glycopeptides carrying $\text{Man}_3\text{Xyl}_1\text{GlcNAc}_2$. Fragment spectra of m/z 3326.9 and 1866.1 (precursors), which correspond to the predicted lectin *N*-glycopeptides with $\text{Man}_3\text{Xyl}_1\text{GlcNAc}_2$ attached to the NFT motif (insert). Peptide b- and y-ions together with glycopeptide fragments are indicated. CFG nomenclature: Blue square; *N*-acetylglucosamine, blue square with black line; 0,2X-ring cleavage of the innermost *N*-acetylglucosamine, green circle; mannose, star; xylose.

Supplemental figure 5B



Supplemental table1

Accession	N-glycan motifs	Transmembrane region	Signal peptide	Localized
chr5.CM0200.930.r2.m (LCP2)	1	no	yes	secreted
chr5.LjT43D06.200.r2.d (predicted lectin)	2	no	yes	secreted
chr5.CM0200.910.r2.m (LCP1)	0	no	yes	secreted
chr5.LjT15N12.20.r2.m (LLP3)	0	no	yes	secreted
chr5.LjT15N12.10.r2.m (LLP2)	0	no	yes	secreted
chr1.CM0295.280.r2.d	0	no	yes	secreted
chr1.CM0295.230.r2.m	0	no	yes	secreted
LjSGA_063286.1 (glucosidase)	1	no	no	other
chr1.CM0544.570.r2.d	1	no	yes	secreted
chr1.CM0295.240.r2.d (LLP5)	0	no	yes	secreted
chr3.CM0241.200.r2.d	9	no	no	other
chr5.LjT43D06.110.r2.d (psedu lectin)	3	no	no	other

Supplemental table 1: Identified proteins from the globulin fraction. In total, 12 proteins were identified from eight different globulin fraction preparations (see Supplemental table 2 for detailed information of the protein identifications from the eight samples). N-glycan motifs are Asn-X-Ser/Thr where X $\neq$ Pro. For transmembrane region prediction, the freely available TMHMM software (<http://www.cbs.dtu.dk/services/TMHMM/>) was used. For signal peptide prediction and localization of the protein, the freely available TargetP 1.1 software (<http://www.cbs.dtu.dk/services/TargetP/>) was used where the plant option was selected. The *Lotus* accessions correspond to the 2.5 version of the genome downloaded from <http://www.kazusa.or.jp/lotus/>. The accession numbers in bold correspond to the suggested proteins carrying complex N-glycans. LLP; *Lotus* legume storage protein, LCP; *Lotus* convicilin storage protein.