JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Determination of the *N*-Glycosylation Patterns of Seed Proteins: Applications To Determine the Authenticity and Substantial Equivalence of Genetically Modified (GM) Crops

Justin T. Marsh,[†] Theodora Tryfona,[‡,§] Stephen J. Powers,[†] Elaine Stephens,[§,#] Paul Dupree,[‡] Peter R. Shewry,[*,†] and Alison Lovegrove[†]

[†]Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, United Kingdom

[‡]Department of Biochemistry, School of Biological Sciences, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, United Kingdom

[§]Department of Chemistry, School of Physical Sciences, University Chemical Laboratory, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Supporting Information

ABSTRACT: Methods have been developed to determine the *N*-glycosylation pattern of proteins at the single-seed level in two different biological systems. These were the well-characterized and widely consumed storage protein phaseolin from several species of *Phaseolus* (bean) and the α-amylase inhibitor from the same *Phaseolus* species expressed transgenically in pea. The *N*-glycosylation pattern of the α-amylase inhibitor expressed transgenically in pea was different from that of the inhibitor present in common bean (*P. vulgaris*), the species of origin of the gene. However, multivariate analysis showed that the differences in *N*-glycan patterns between the α-amylase inhibitors from common bean and pea were less than those between the inhibitors from common bean and two related bean species, lima bean (*Phaseolus lunatus*) and tepary bean (*Phaseolus acutifolius*).

KEYWORDS: *N*-glycosylation, MALDI-ToF-MS, inauthentic glycosylation, genetic modification, allergenicity

INTRODUCTION

Glycosylation is the major type of post-translational modification occurring in plants, with two main types existing: *N*-linked and *O*-linked glycosylation. *N*-Glycosylation occurs on the amide nitrogen of asparagine residues, and possible *N*-glycosylation sites can be predicted on the basis of sequence context.[1] On the other hand, *O*-linked glycans are mainly attached to the hydroxyl group of hydroxyproline (Hyp), and possible glycosylation sites can be predicted according to the Hyp contiguity hypothesis.[2] Whereas *O*-glycosylation is poorly understood in plants, *N*-glycosylated plant proteins have been widely studied, including the glycosylated forms of abundant legume seed proteins that are widely consumed as part of the human diet (for example, bean phaseolin, reviewed in ref 3). *N*-Glycosylation of plant proteins may also contribute to their allergenicity.[4] This is particularly well-characterized for the cereal α-amylase inhibitors, which are major allergens in bakers' asthma.[5] A number of allergenic subunits have been identified in flour of wheat and barley, with the highest levels of IgE binding being observed with three quantitatively minor *N*-glycosylated forms.[6]

Further interest in the role of glycosylation in allergenicity has resulted from the transgenic expression of a bean α-amylase inhibitor in pea seeds.[7] This protein, which is not related to the cereal α-amylase inhibitors discussed above, accumulated in a modified form when expressed in transgenic pea, differing from the authentic form present in bean in the pattern of *N*-glycosylation. Furthermore, the authors showed that consumption of the modified form of the protein present in peas, but not the native form, predisposed mice to antigen-specific $CD^{4+}Th_2$-type inflammation and that consumption with other proteins promoted immunological cross-priming and elicited specific immunoreactivity to these proteins. This paper highlighted the need to develop methods to monitor glycosylation patterns of plant proteins and to assess the relevance of these patterns to allergenicity.

We have therefore developed methods to determine the *N*-glycosylation of plant proteins at the single-seed level and applied them to two biological systems. These methods were initially applied to phaseolin of *Phaseolus vulgaris* in which one or two potential glycosylation sites are known and the glycans identified.[8] The methods were then applied to the α-amylase inhibitors in the transgenic pea lines described in ref 7 and to appropriate control lines of pea and bean.

MATERIALS AND METHODS

Seeds. Seeds of *Phaseolus vulgaris* (common bean) cv. Tendergreen were obtained from Sutton's Seeds (Paignton, Devon, U.K.), of cvs. Masterpiece and Blue Lake from Unwin's Seeds (Huntingdon, Cambridgeshire, U.K.), and of *Phaseolus lunatus* (lima bean) and *Phaseolus acutifolius* (tepary bean) from Two Willies Nursery (Lucedale, MS). Transgenic and control seeds of *Pisum sativum* (pea) cv. Mukta were provided by Dr. T. J. Higgins (CSIRO Division of Plant Industry, Canberra, Australia). Transgenic and control lines of pea were

Received: March 17, 2011
Revised: July 18, 2011
Accepted: July 25, 2011
Published: July 25, 2011

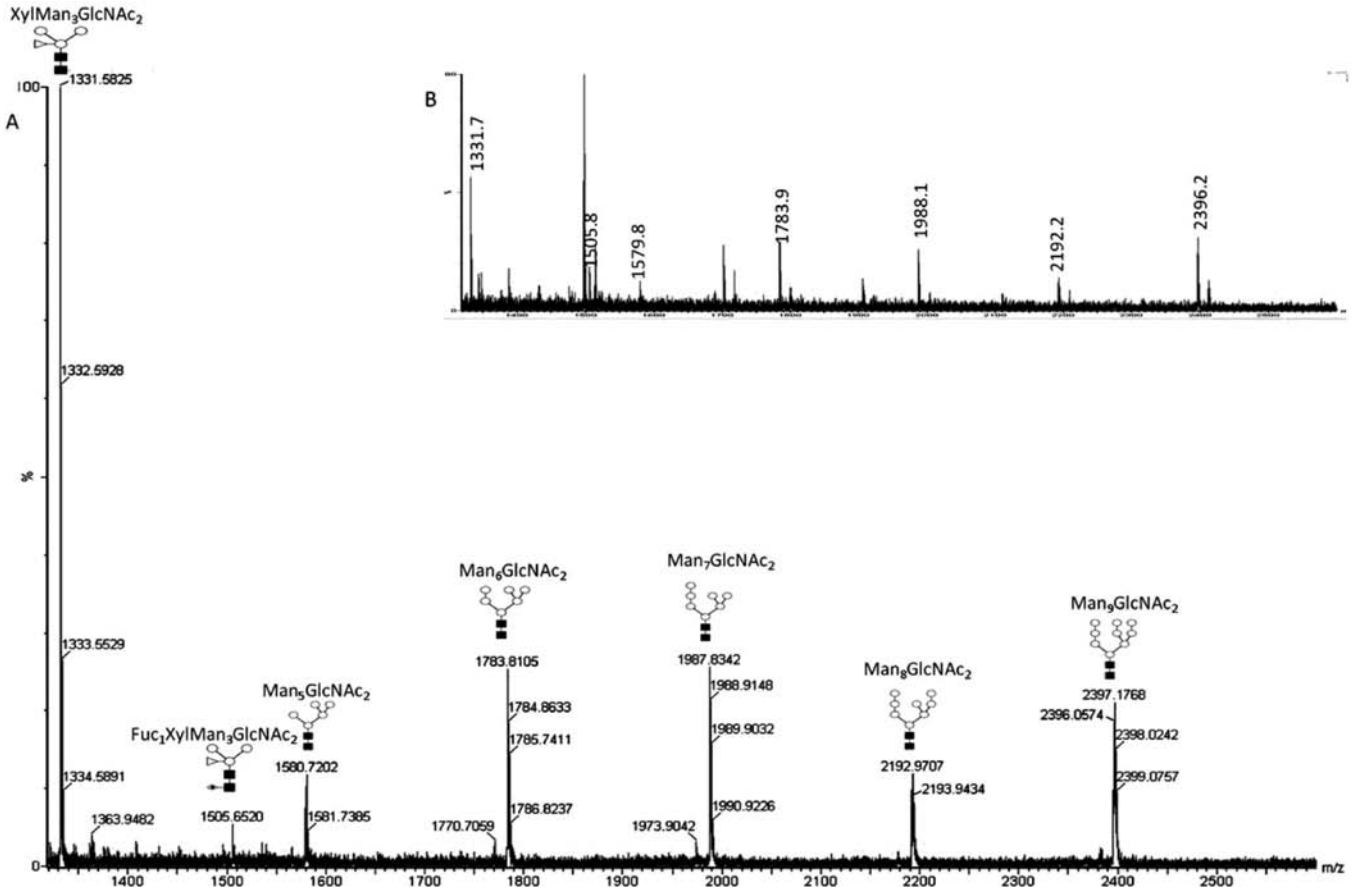

Figure 1. MALDI-ToF-MS spectrum of permethylated *N*-glycans isolated from phaseolin: (A) cv. Tendergreen; (inset B) MS spectrum from phaseolin isolated from a single seed of *Phaseolus vulgaris* cv. Tendergreen. The glycan structure is indicated by the cartoon and the name given (see the Supporting Information, Tables S1 and S2, for details). Masses of the glycan $[M + Na]^+$ ions are given.

multiplied in a containment glasshouse at Rothamsted Research (15/18 °C night/day, with a 16 h photoperiod). Seeds of all bean and pea species were selected at random to provide biologically replicated material. At least two technical replicates of each biological replicate were analyzed.

Preparation and Analysis of Proteins. Total protein fractions were isolated from hand-dissected cotyledons by grinding with a mortar and pestle with water or 1 M NaCl (1:5 w/v) followed by stirring for 2 h at room temperature. The extract was then filtered through two layers of muslin and centrifuged at 15000*g* for 40 min at 4 °C, and the supernatants were dialyzed against water overnight and lyophilized.

Phaseolin was isolated according to the method of ref 9. The post-anion exchange fractions were dialyzed against water overnight and lyophilized. α-Amylase inhibitor was isolated according to the method of ref 10. The final post-cation exchange fractions were dialyzed and lyophilized.

N-Glycans were released from the glycopeptides by PNGase A digestion. The glycopeptides were generated by dissolving the protein sample in 50% (v/v) formic acid to which was added pepsin solution. The mixture was incubated for 24 h at 37 °C. Following digestion, the enzyme was inactivated by boiling for 5 min and the glycopeptides were recovered by lyophilization. The lyophilized glycopeptide mixture was dissolved in 50 mM ammonium acetate, pH 5. Four microliters of PNGase A was added and the mixture incubated at 37 °C for 16 h. After lyophilization, the sample was redissolved in 0.5 mL of 5% (v/v) acetic acid and the glycans were purified by passage through a Sep-Pak C18 cartridge. Briefly, the cartridge was wetted with 100% methanol and equilibrated with 5% (v/v) acetic acid. The sample was loaded and eluted with 5% (v/v) acetic acid, leaving nonpolar peptides behind. The eluted glycans were then lyophilized and further purified by passage through Dowex 50WX8-100 resin. The Dowex resin was washed repeatedly with 4 M HCl and then with water until the pH had returned to that of the water. The Dowex resin was then washed three times in 5% (v/v) acetic acid and packed into a Pasteur pipet, with glass wool as a frit. The lyophilized sample was dissolved in 100 μL of 5% (v/v) acetic acid and eluted with the same solvent, removing salt and polar peptides. The eluted glycans were lyophilized. Permethylation of glycans was performed using the NaOH slurry method described by ref 11 using 1 mL of methyl iodide (Fluka). The derivatized glycans were redissolved in 200 μL of 35% (v/v) acetonitrile and passed through a Sep-Pak C18 cartridge. The glycans were eluted with 50% (v/v) acetonitrile and freeze-dried. The permethylated glycan samples (5 μL) were mixed with 5 μL of 2,5-dihydroxybenzoic acid (2,5-DHB; 10 mg/mL dissolved in 50% MeOH), and 1 μL was spotted in duplicate on the MALDI target and allowed to air-dry.

MALDI-ToF-MS was carried out using a Micromass MALDI-LR mass spectrometer (Waters, Manchester, U.K.) using a standard peptide mass fingerprinting method and mass acquisition between m/z 1200 and 3500. (The laser firing rate was 5 Hz, 40 random aims per spot, 10 shots per spectrum, 10 spectra per scan, 10 scans combined, 10% adaptive background subtracted, smoothed (Savitzky−Golay), and centroided to obtain the MS spectra.) The MALDI-MS was tuned to 10000 fwhm and calibrated with a tryptic digest of ADH following the manufacturer's instructions. Glycan adduct ions $[M + Na]^+$ were assigned.

Canonical Variates Analysis (CVA). For the application of multivariate statistical analysis, the data were converted to .pkl files

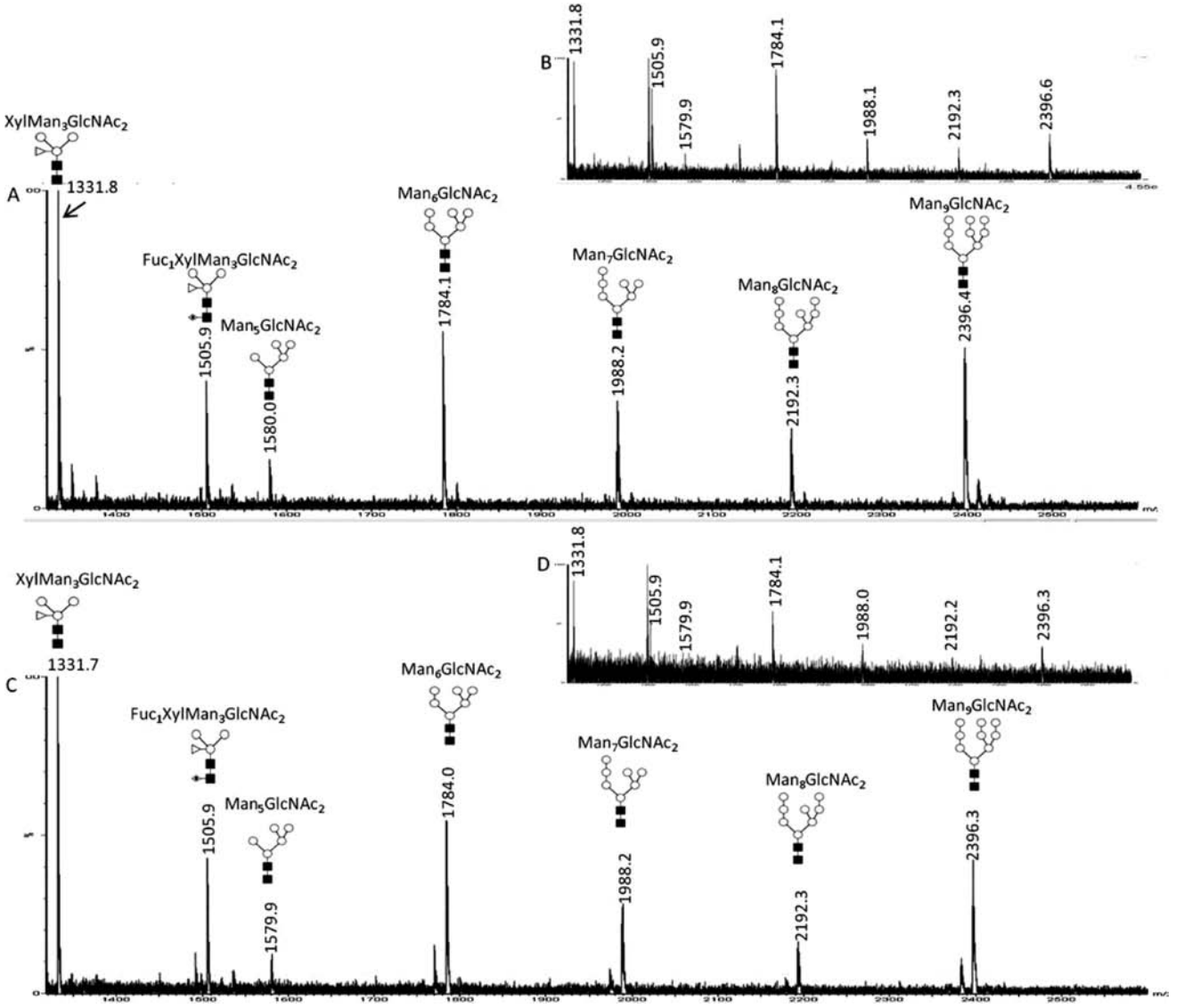

Figure 2. MALDI-ToF-MS spectrum of permethylated *N*-glycans from phaseolin isolated from *Phaseolus vulgaris* cv. Masterpiece (A) and cv. Blue Lake (C); (insets) MS specta from phaseolin isolated from a single seed of *P. vulgaris* cv. Masterpiece (B) and *P. vulgaris* cv. Blue Lake (D). The glycan structure is indicated by the cartoon and the name given (see the Supporting Information, Tables S1 and S2, for details). Masses of the glycan $[M + Na]^+$ ions are given.

to produce a peak mass−intensity list and the masses of the seven *N*-glycans extracted. CVA (see, for example, ref 12) was used to maximize the ratio of the between species variation to the within-species variation, thus performing a discrimination between all species. Mathematically, the direction in the seven-dimensional, given seven observed *N*-glycans that performs this task most ably is a derived linear combination of the seven variates. This linear combination is termed the first canonical variate (CV). The second and subsequent CVs maximize the variation (discrimination) subject to the constraint of orthogonality (in terms of variance) with respect to all previous CVs. The fewest number of canonical variates is retained that take up the most variation in the data and hence make the most discrimination. The data are then visualized on the new dimensions, by plotting the canonical variate scores for each sample. The mean of canonical variate scores in each dimension for each species is also plotted. By making the assumption of a multivariate Normal distribution for the data, which is a reasonable assumption for the data on the natural logarithm ($\log_e$) scale, 95% confidence circles can be placed around the canonical variate (species) means. The radius of these circles is $(\sqrt{\chi^2_{2,0.05}})/\sqrt{n}$, where n is the biological replication (in this case being between 1 and 4) and where $\chi^2_{2,0.05} = 5.99$ is the upper 5% point of a χ-squared distribution on 2 degrees of freedom. Nonoverlapping confidence circles give evidence of significant differences between species at the 5% level of significance. The GenStat (2009, 12th ed.; Lawes Agricultural Trust (Rothamsted Research), VSN International Ltd., Hemel Hempstead, U.K.) statistical system was used for the analysis.

RESULTS AND DISCUSSION

Purification and Analysis of Phaseolin. We wished to develop methods to determine the substantial equivalence of protein glycosylation patterns in transgenic and conventionally bred food crops, focusing on methods that could be routinely applied to large numbers of samples at the single-seed or organ level. We therefore initially evaluated the methods on a well-characterized glycosylated food protein, the 7S globulin

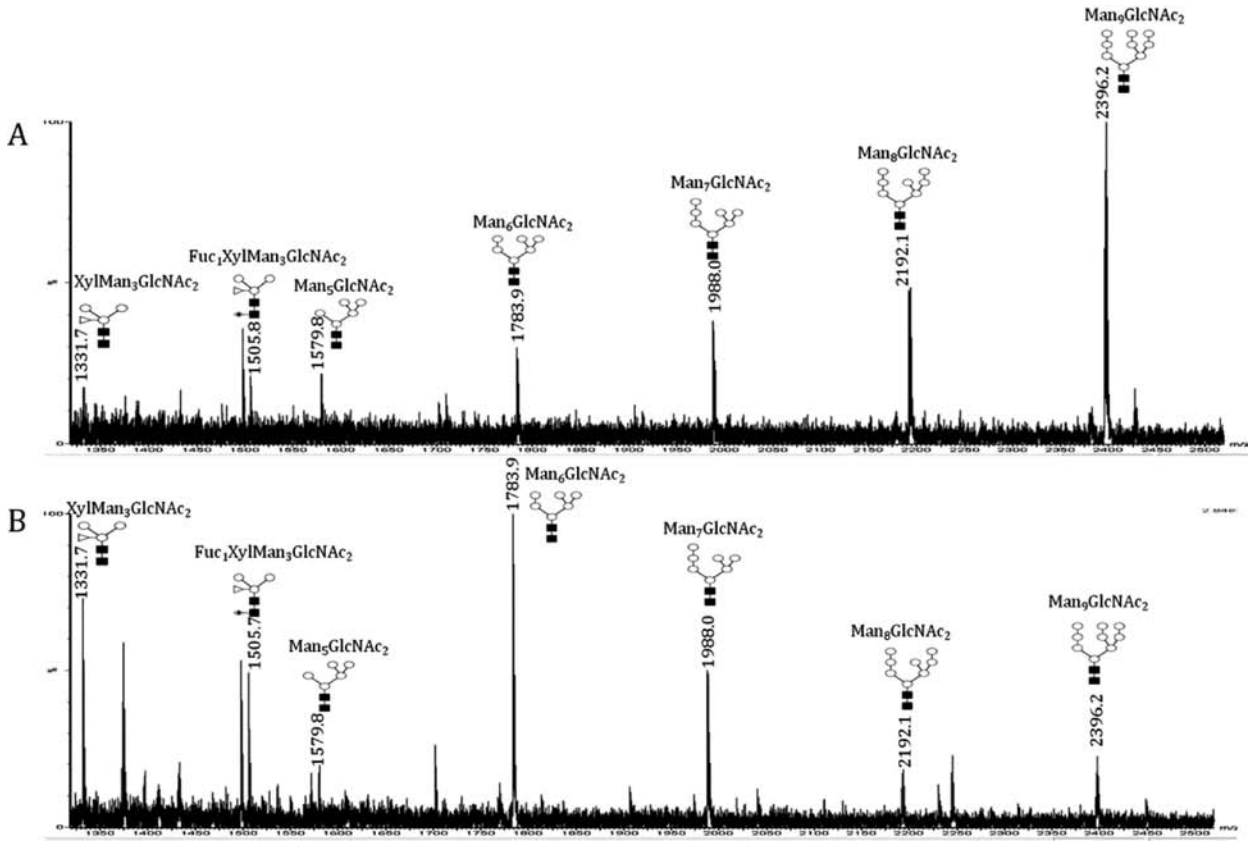

Figure 3. MALDI-ToF-MS spectra of permethylated *N*-glycans isolated from phaseolin from two different bean species. *N*-Glycan spectra of phaseolin were extracted from two bean species, *Phaseolus lunatus* (A) (lima bean) and *Phaseolus acutifolius* (B) (tepary bean). The glycan structure is indicated by the cartoon and the name given (see the Supporting Information, Tables S1 and S2, for details). Masses of the glycan $[M + Na]^+$ ions are given.

storage protein phaseolin from common bean (*Phaseolus vulgaris*) (reviewed by ref 3).

Phaseolin was prepared from milled seed of the common bean (*Phaseolus vulgaris*) cv. Tendergreen by size exclusion chromatography using Sephadex G50 followed by ion-exchange chromatography using DEAE-Sepharose. SDS-PAGE showed major bands of M_r approximately 50000, corresponding to the subunits of the trimeric phaseolin protein (see Hall et al.[3]) (Supporting Information, Figure S1). MALDI-ToF-MS of the glycan chains released by digestion with *N*-glycosidase A revealed seven glycan peaks ranging in *m*/*z* from 1331 to 2396. The structures of these were assigned as shown in Figure 1A, based upon Glycomod (ExPaSy) (http://www.expasy.ch/tools/glycomod/) and GlycanBuilder (http://www.eurocarbdb.org/applications/structure-tools) software tools (see the Supporting Information, Table S1, for masses of glycans and a key to the structures shown in all figures). Similar patterns of glycans were observed when phaseolin from two other cultivars of bean (Masterpiece (Figure 2A) and Blue Lake (Figure 2C)) was analyzed, although $Fuc_1XylMan_3(GlcNAc)_2$ (*m*/*z* 1505) was more prominent in these samples.

Similarly, analyses of phaseolin preparations from two related cultivated species of bean, lima bean (*Phaseolus lunatus*) (also called butter bean in the United Kingdom and United States) and tepary bean (*Phaseolus acutifolius*) (which is grown in the southern United States and Central America), identified the same structures as in phaseolin from common bean (Figure 3A (lima bean) and B (tepary bean)).

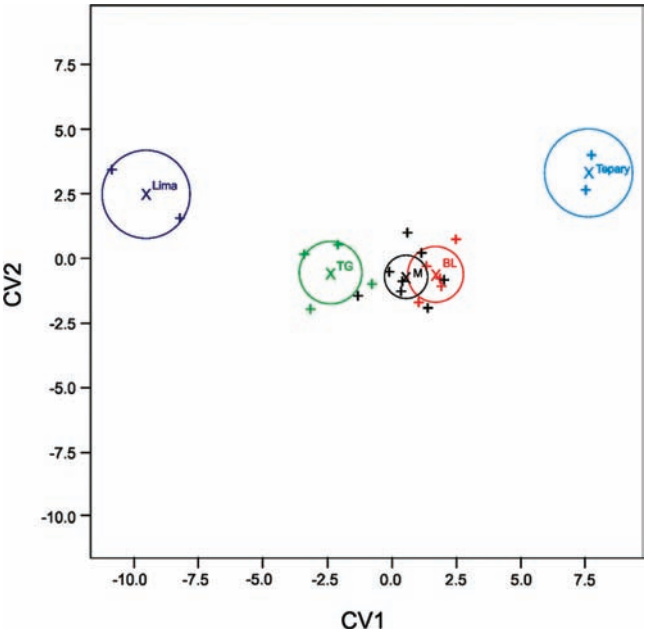

Figure 4. Canonical variates analysis plot of variation in intensity of *N*-glycans isolated from phaseolin from three cultivars of *Phaseolus vulgaris* and two other bean species: Tendergreen (TG, green); Masterpiece (M, black); Blue Lake (BL, red); *Phaseolus lunatus* (lima bean, blue); *Phaseolus acutifolius* (tepary bean, cyan). The + symbols are the CV scores for all of the replicates, and the × symbols are the means of these scores for each bean or pea sample. The circles represent 95% confidence limits.

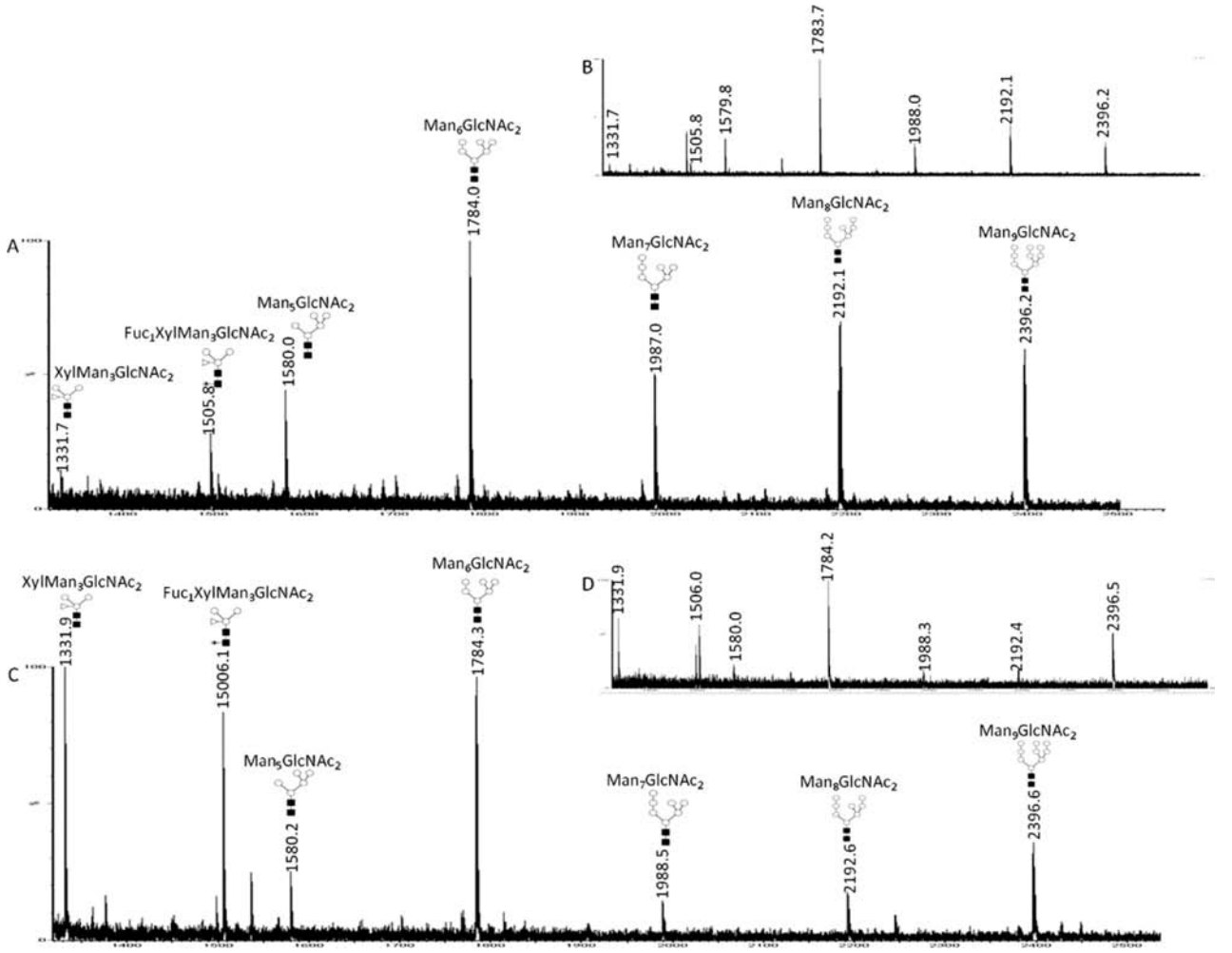

Figure 5. MALDI-ToF-MS spectra of permethylated *N*-glycans isolated from the α-amylase inhibitor extracted from (A) transgenic pea (*Pisum sativum* cv. Mukta) and (C) native bean (*Phaseolus vulgaris* cv. Tendergreen); (insets) α-amylase inhibitor extracted from single seeds of (B) transgenic pea and (D) native bean.

Individual phaseolin polypeptides are known to differ in their *N*-glycosylation patterns, with either one or two sites being recognized.[8] *N*-Linked plant glycans fall into two major categories, high mannose and complex glycans (although hybrid and paucimannosidic glycan structures do also exist), and ref 9 demonstrated some phaseolin polypeptides had high mannose type glycans, $Man_7(GlcNAc)_2$ and $Man_9(GlcNAc)_2$, attached at sites Asn^{252} and Asn^{341}, respectively. However, when the complex glycan $XylMan_3(GlcNAc)_2$ was present on Asn^{252}, the site at Asn^{341} was not occupied.

Our analysis of phaseolin did not distinguish between the glycosylation patterns of individual phaseolin polypeptides but did identify all of the glycans that had previously been reported for phaseolin in ref 9. In addition to the glycans identified in ref 9, we were able to identify $Man_5(GlcNAc)_2$, $Man_6(GlcNAc)_2$, $Man_8(GlcNAc)_2$, $Man_9(GlcNAc)_2$, and Fuc_1 $XylMan_3(GlcNAc)_2$. The same glycans were identified in all three cultivars of *Phaseolus vulgaris* (Tendergreen, Masterpiece, and Blue Lake) and in two other bean species (*Phaseolus lunatus* (lima bean) and *Phaseolus acutifolius* (tepary bean)). However, the proportions of the various glycans differed between the species. For example, in *P. lunatus* there were relatively lower levels of the complex $XylMan_3(GlcNAc)_2$ (m/z 1331) and high-mannose $Man_6(GlcNAc)_2$ (m/z 1783) glycans, but relatively higher proportions of the high-mannose glycan $Man_9(GlcNAc)_2$ (m/z 2395) compared to *Phaseolus vulgaris* cultivars. In *Phaseolus acutifolius* the relative proportions of $XylMan_3$-$(GlcNAc)_2$ (m/z 1331) and $Man_6(GlcNAc)_2$ (m/z 1783) were more similar to the *Phaseolus vulgaris* cultivars, but the high-mannose *N*-glycans, $Man_8(GlcNAc)_2$ (m/z 2192) and $Man_9(GlcNAc)_2$ (m/z 2395), were reduced.

It is possible that the extent of glycosylation varies between cultivars or even between individual seeds within a cultivar, depending on the environment or other factors (for example, the position in the pod could affect the supply of carbon into the developing seeds). We therefore applied the analytical procedure developed for milled samples to single seeds, omitting the size exclusion step of the purification to prevent protein dilution and losses. Typical mass spectra of glycans from single seeds of each of the three bean cultivars are shown as insets: in Figure1B (Tendergreen single seed) and in Figure 2B (Masterpiece single seed) and D (Blue Lake single seed). Despite the higher baseline, it is possible to identify the same glycan patterns as observed in the larger protein preparations.

It is essential to develop objective statistical methods to determine substantial equivalence rather than make subjective comparisons of MS patterns. We therefore used multivariate

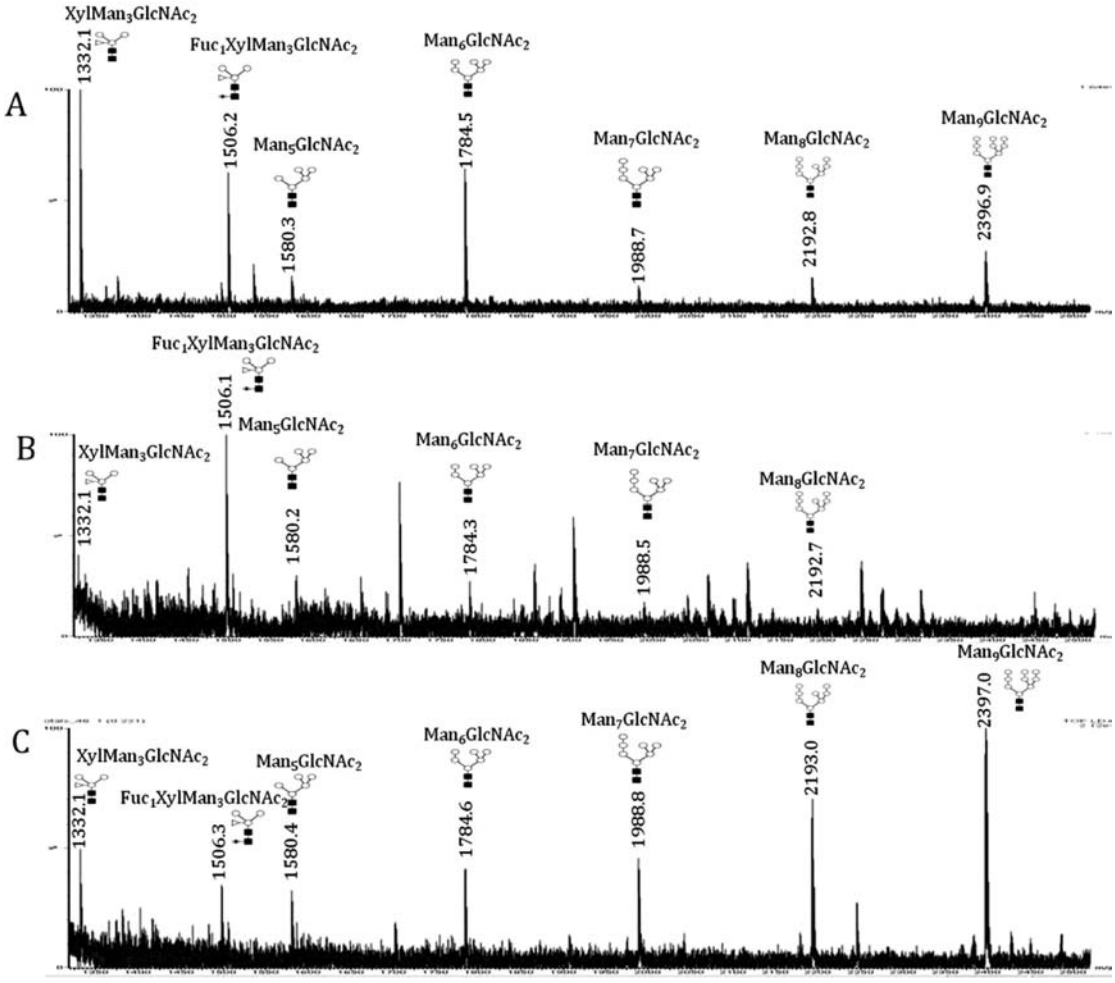

Figure 6. MALDI-ToF-MS spectrum of permethylated *N*-glycans isolated from α-amylase inhibitor extracted from three bean species: (A) *Phaseolus vulgaris* (common bean). Masses of the glycan $[M + Na]^+$ ions are given; (B) *Phaseolus acutifolius* (tepary bean); (C) *Phaseolus lunatus* (lima bean).

statistical analysis methods to compare the mass spectra on the basis of the fact that the relative peak heights of the glycans provide semiquantitative estimates of their relative amounts.

Duplicate MALDI spectra were determined for extracts of biological replicate single seeds. The peak areas were centroided and 30% background subtracted from all spectra using Protein Links Global Server 4.1 (Waters). The intensities of the peaks corresponding to the glycans (m/z 1331, 1505, 1579, 1783, 1987, 2192, and 2395) were then compared using CVA (Figure 4); 97.4% of the variation observed was accounted for by CV1 and CV2, with CV1 accounting for 86.6% and CV2 for 10.8%. The cultivars Masterpiece (black data points) and Blue Lake (red data points) are the most similar, but can be distinguished from *Phaseolus vulgaris* cv. Tendergreen (green data points). However, the most striking differences are between the bean species (*Phaseolus lunatus*, *Phaseolus acutifolius*, and *Phaseolus vulgaris*), which are greater than the differences between the cultivars of *Phaseolus vulgaris*.

This analysis confirmed the utility of the method for determination of the substantial equivalence of glycosylated seed proteins. It was therefore used to compare the native and transgenically expressed forms of the bean α-amylase inhibitor

Analysis of the Bean α-Amylase Inhibitor from Bean and Transgenic Pea. The bean α-amylase inhibitor was initially prepared from seeds of transgenic pea by ion-exchange chromatography on DEAE Sepharose (Supporting Information, Figure S2A) followed by CM Sepharose (Supporting Information, Figure S2B). SDS-PAGE showed one major band and several minor bands of M_r between about 12000 and 20000 (transgenic *Pisum sativum*, Supporting Information, Figure S2B, track 1). The same procedure was then applied to bean cv. Tendergreen, resulting in a fraction with a similar band pattern on SDS-PAGE (native *Phaseolus vulgaris*, Supporting Information, Figure S2B, track 2). Similar patterns of bands on SDS-PAGE have been reported in ref 13 for the native protein from bean and in refs 14 15, and 7 for the protein expressed in transgenic pea.

The preparations from bean and pea were finally separated by gel filtration chromatography on Superdex 75 and peak fractions separated by SDS-PAGE (Supporting Information, Figure S2C). Groups of bands (indicated by arrows in Supporting Information, Figure S2D) were transferred from a 15% Tricine SDS-PAGE gel to a PVDF membrane for N-terminal sequence analysis. The mature α-amylase inhibitor of bean is post-translationally processed to give major α- and β-chains of masses 11646 and 17319 Da, respectively, with minor forms containing alternative glycans.[13,16−18] The higher M_r group of bands present in the bean preparation (Supporting Information, Figure S3D) gave a single major N-terminal sequence AXETXF (where X is

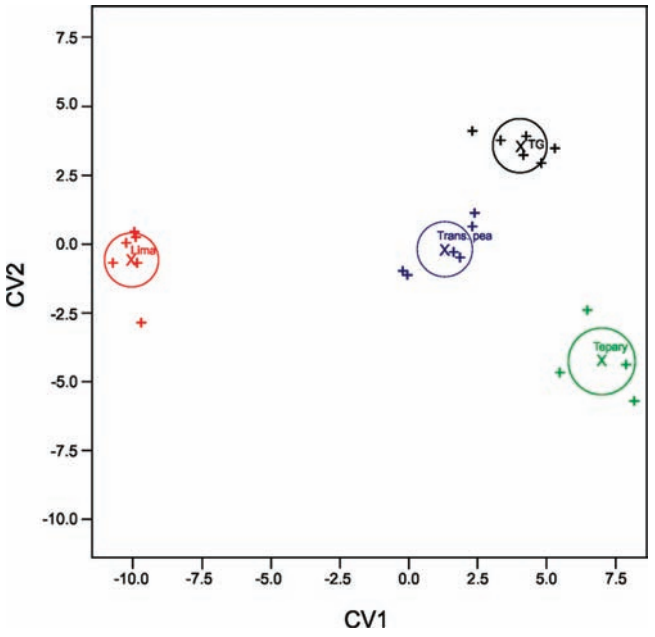

Figure 7. Canonical variates analysis plot of variation in intensity of *N*-glycans isolated from α-amylase inhibitor extracted from transgenic pea and three bean species: *Pisum sativum* cv. Mukta (trans pea, blue); *Phaseolus lunatus* (lima bean, red); *Phaseolus acutifolius* (tepary bean, (green); *Phaseolus vulgaris* cv. Tendergreen (common bean, TG, black). The + symbols are the CV scores for all of the replicates, and the × symbols are the means of these scores, for each bean or pea sample. The circles represent 95% confidence limits.

unidentified), corresponding to the α-chain of the α-amylase inhibitor (sequence ATETSF).[13] In contrast, the three groups of lower M_r bands from the bean and pea preparations (Supporting Information, Figure S2D) each gave mixtures of two major sequence types. These were ATETSFIIDAFXKTNL and SAVGLDFVLVPVQPES and corresponded to the α-chain and β-chains, respectively, of the processed bean α-amylase inhibitor.[13]

The patterns of glycans released by digestion with *N*-glycosidase A and analyzed by MALDI-ToF-MS are shown in Figure 5A, C. Although the two preparations exhibit the same range of glycans, there are clear differences in their relative proportions; for example, the high-mannose glycan (m/z 2395) appears to be proportionally greater in the native bean compared to the transgenic pea when compared to the other glycans present and the $Man_5(GlcNAc)_2$ (m/z 1579) relatively lower in the native bean α-amylase inhibitor than in the transgenic pea α-amylase inhibitor. Prescott et al.[7] assigned the glycans by alignment of their linear MALDI spectra to data published previously in ref 15. They suggested that the most abundant glycan present in the α-amylase inhibitor in the transgenic pea was $Man_7(GlcNAc)_2$ but that it was the least abundant in the native bean α-amylase inhibitor. They also suggested that the $Man_9(GlcNAc)_2$ glycan was more abundant in the native bean α-amylase inhibitor than in that present in the transgenic pea.

The current study was able to identify the same *N*-glycans present in the α-amylase inhibitors in the native bean and the transgenic pea, which were the high-mannose series $Man_{5-9}(GlcNAc)_2$, and additionally the complex glycans $XylMan_3(GlcNAc)_2$ (m/z 1331) and $Fuc_1XylMan_3(GlcNAc)_2$ (m/z 1505). It also confirmed that $Man_9GlcNAc_2$ (m/z 2395) is relatively more abundant in the native bean α-amylase inhibitor compared to that in the transgenic pea, but did not find $Man_7(GlcNAc)_2$ (m/z 1983) to be the most abundant glycan in the transgenic pea. The relative proportions of $XylMan_3(GlcNAc)_2$ (m/z 1331), $Fuc_1XylMan_3(GlcNAc)_2$ (m/z 1505), and $Man_5(GlcNAc)_2$ (m/z 1579) also differed in the α-amylase expressed in the transgenic pea compared to the native bean.

To determine whether the differences observed were consistent at the single-seed level, we repeated the experiment on preparations from replicate single seeds of transgenic pea (Figure 5B) and single-seed extractions of bean (Figure 5D). The same pattern was observed at the single-seed level, although there was some increase in the background compared to that obtained for the large-scale preparations, probably due to omission of the gel filtration step. The differences in glycan pattern between the two species are clearly apparent, with little variation observed between the single seeds within each species. To examine these differences further, we isolated the α-amylase inhibitor from two other bean species, lima (*Phaseolus lunatus*) and tepary (*Phaseolus acutifolius*) beans (Figure 6). Although this showed differences in the relative proportions of the glycans present in common bean (A), tepary bean (B), and lima bean (C), the same glycans (the high-mannose series and two complex glycans) are present in all of the bean species. CVA showed that CV1 and CV2 together accounted for 95.0% of the variation (Figure 7). There was a clear separation of all species, but the α-amylase inhibitors from transgenic pea and common bean are more similar to each other (although clearly separated), whereas the glycan patterns of the α-amylase inhibitors from lima and tepary beans are the most different.

This analysis therefore confirmed that the single-seed method was appropriate for determining the substantial equivalence of transgenically and natively expressed forms of glycoprotein.

Do the Differences between Pea and Bean Relate to Wider Differences in Protein Glycosylation between the Two Species? Although pea and bean are closely related legume species, they differ in the extent to which the major globulin storage proteins are glycosylated. Two storage protein fractions are present in pea, the 11S "legumin-type" globulins and the 7S "vicilin-type". Of these, the 11S globulins are not glycosylated, whereas a single site is glycosylated in some, but not all, 7S globulin subunits.[18,19] In contrast, *Phaseolus* contains only 7S storage globulins (phaseolin), which are always glycosylated on one and sometimes on two sites.[3] It can be hypothesized, therefore, that these differences may be associated with differences in the activity and specificity of the protein glycosylation machinery in the two species.

To explore this relationship, we determined the glycosylation patterns of total protein preparations made with water (which would contain the α-amylase inhibitor, Figure 8A−C) and 1.0 M NaCl (which contains the 7S and 11S storage globulins, Figure 8D−F) from the control and transgenic lines of pea and from bean cv. Tendergreen.

The water-soluble fraction from the control line of pea contained only the two complex glycans $XylMan_3(GlcNAc)_2$ (m/z 1331) and $Fuc_1XylMan_3(GlcNAc)_2$ (m/z 1505) and the high-mannose $Man_5(GlcNAc)_2$ (m/z 1579) and $Man_6(GlcNAc)_2$ (m/z 1783) (Figure 8A), demonstrating the presence of a few water-extractable glyans. In contrast, the pattern of glycans observed in the water extract from the transgenic pea line was similar to that of the α-amylase inhibitor purified from this species (compare Figure 8B with Figure 5A). This shows that glycosylation of the intrinsic water-soluble pea proteins was very

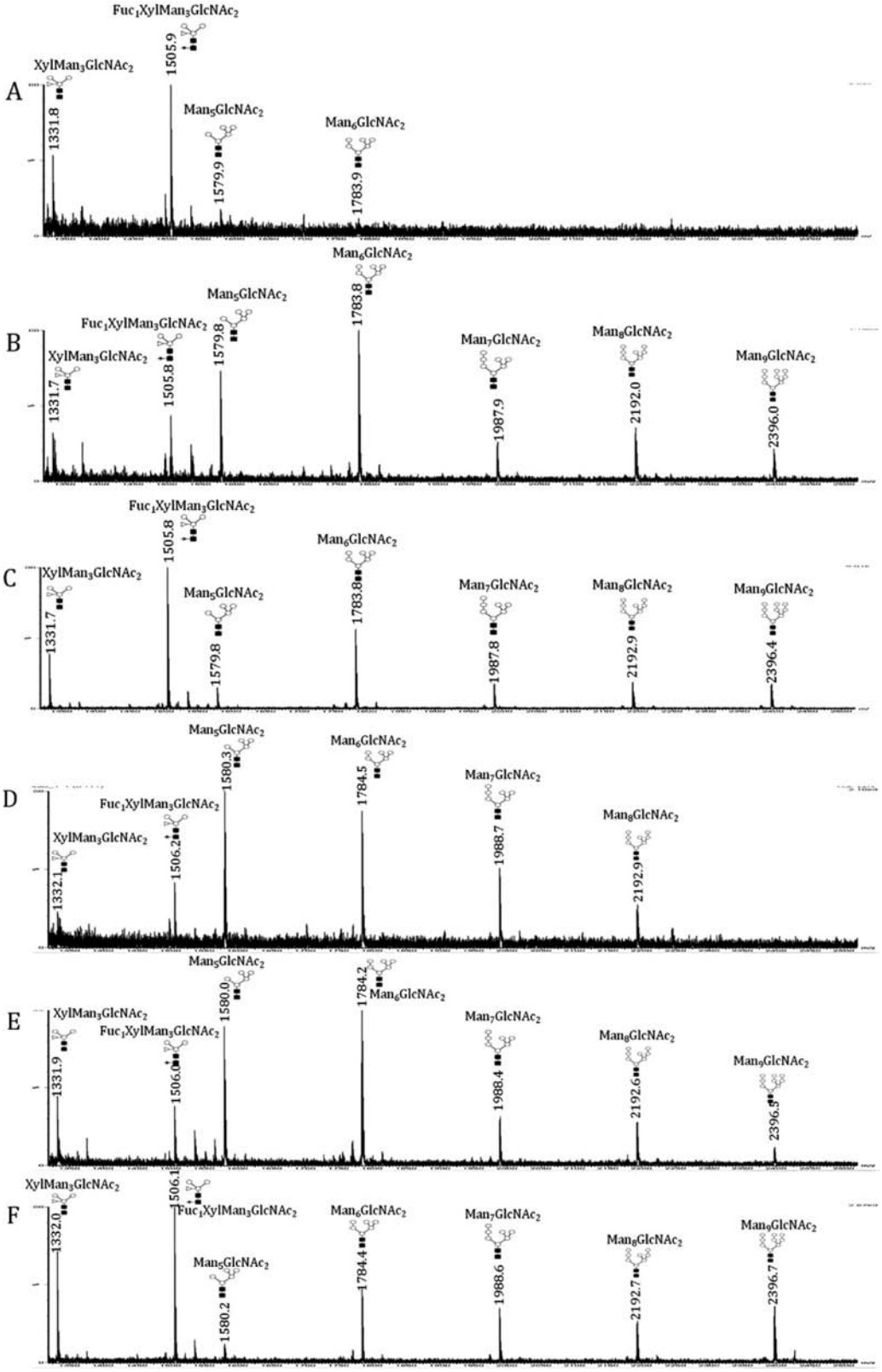

Figure 8. MALDI-ToF-MS spectrum of permethylated *N*-glycans isolated from total water or salt extracts from wild-type and transgenic *Pisum sativum* and native *Phaseolus vulgaris*: *N*-glycan spectra following water extraction of proteins from wild-type (A) and transgenic *Pisum sativum* (B) and native *Phaseolus vulgaris* (C); panels D−F show the *N*-glycan spectra following 1 M NaCl extraction of proteins from wild-type (D) and transgenic *Pisum sativum* (E) and from native *Phaseolus vulgaris* (F). Masses of the glycan $[M + Na]^+$ ions are given.

dx.doi.org/10.1021/jf2010854 |*J. Agric. Food Chem.* 2011, 59, 8779–8788

limited and that crude aqueous extracts could be used to characterize the glycans present on the α-amylase inhibitor expressed in the transgenic line. Analysis of the total water-soluble proteins from bean also showed a range of glycans similar to those present in the α-amylase inhibitor preparations (compare Figure 8C with Figure 5C).

The glycan pattern of the salt-soluble fraction from bean (Figure 8F) was similar to that of the purified phaseolin preparation (Figures 1 and 2), which is consistent with phaseolin being the major glycosylated protein in bean seeds. However, the pattern in the control pea line differed, with no high-mannose $Man_9(GlcNAc)_2$ (m/z 2395) being observed. The ratio of $Man_5(GlcNAc)_2$ (m/z 1579) to $Man_3Xyl(GlcNAc)_2$ (m/z 1331) appeared to be greater in both the control pea and the transgenic pea compared to the native bean, as it was when α-amylase inhibitors were extracted from the bean and transgenic pea (compare Figure 8D−F with Figure 5).

Taken together, the mass spectra obtained from the water and salt extracts from bean and pea indicate that the activity of the glycosylation machinery in the two species does differ. However, it is not known whether this relates to different specific activities of the modifying enzymes involved in glycosylation in the two species or to other factors.

In summary, we have developed routine methods for the analysis of glycan structures of glycoproteins from seeds of widely consumed legumes (peas and beans) and shown that these can be applied to single seeds with high reproducibility and patterns similar to those observed with larger samples of milled seeds. CVA also allowed the patterns of protein glycosylation to be compared objectively rather than subjectively by visually comparing the mass spectra.

This method was initially used to determine the glycosylation pattern of a characterized food glycoprotein, phaseolin of common bean, showing that the patterns of glycosylation were essentially identical in different cultivars but differed in two related food crops, tepary bean and lima bean. The method may therefore applicable to determination of the authenticity of samples of these species. Differences in relative peak heights may also be used to indicate the presence of mixtures of species in seed samples, but because MALDI-MS is only semiquantitative, the precise proportions of the individual species cannot be calculated.

The method was then used to determine the substantial equivalence of the bean α-amylase inhibitor expressed in transgenic pea with the native protein present in bean itself. Although the patterns of the two proteins clearly differed, the differences were smaller than those between the inhibitors present in bean and related *Phaseolus* species (lima and tepary beans). Hence, the transgenically expressed and native bean inhibitors can be considered to be "substantially equivalent" within the wider context of variation within bean species that are consumed by humans.

■ ASSOCIATED CONTENT

Supporting Information. Additional table and figures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +44 (0)1582 763133. Fax: +44 (0)1582 763010. E-mail: peter.shewry@rothamsted.ac.uk.

Present Addresses

#Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K.

Funding Sources

This work was funded by the Foods Standards Agency G03029. Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom.

■ ACKNOWLEDGMENT

We are grateful to Dr. T. J. Higgins (CSIRO, Canberra, Australia) for providing transgenic and control lines of pea.

■ ABBREVIATIONS USED

CVA, canonical variates analysis; GlcNAc, *N*-acetylglucosamine; Man, mannose; Xyl, xylose; Fuc, fucose; MALDI-ToF-MS, matrix-assisted laser desorption time-of-flight mass spectrometry; SDS-PAGE, sodium dodecyl sulfate−polyacrylamide gel elerctrophoresis.

■ REFERENCES

(1) Yan, B.; Zhang, W.; Ding, J.; Gao, P. Sequence pattern for the occurrence of N-glycosylation in proteins. *J. Protein Chem.* **1999**, *18*, 511–521.

(2) Tan, L.; Leykam, J. F.; Kieliszewski, M. J. Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. *Plant Physiol.* **2003**, *132*, 1362–1369.

(3) Hall, T. C.; Chandrasekharan, C. B.; Li, G. Phaseolin: its past, properties, regulation and future. In *Seed Proteins*; Shewry, P. R., Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999; pp 209−240.

(4) Wilson, I. B. H.; Zeleny, R.; Kolarich, D.; Staudacher, E.; Stroop, C. J. M.; Kamerling, J. P.; Altmann, F. Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core α1,3-linked fucose and xylose substitutions. *Glycobiology* **2001**, *11*, 261–274.

(5) Sanchez-Monge, R.; Gomez, L.; Barber, D.; Lopez-Otin, C.; Armentia, A.; Salcedo, G. Wheat and barley allergens associated with baker's asthma. Glycosylated subunits of the α-amylase-inhibitor family have enhanced IgE-binding capacity. *Biochem. J.* **1992**, *281*, 401–405.

(6) Salcedo, G.; Sanchez-Monge, R.; Garcia-Casado, G.; Armentia, A.; Gomez, L.; Barber, D. The cereal α-amylase/trypsin inhibitor family associated with bakers' asthma and food allergy. In *Plant Food Allergens*; Mills, E. N. C., Shewry, P. R., Eds.; Blackwell Publishing: Oxford, U.K., 2004; pp70−86.

(7) Prescott, V. E.; Campbell, P. M.; Moore, A.; Mattes, J.; Rothenberg, M. E.; Foster, P. S.; Higgins, T. J. V.; Hogan, S. P. Transgenic expression of bean α-amylase inhibitor in peas results in altered structure and immunogenicity. *J. Agric. Food Chem.* **2005**, *53*, 9023–9030.

(8) Bollini, R.; Vitale, A.; Chrispeels, M. J. In vivo and in vitro processing of seed reserve protein in the endoplasmic reticulum: evidence for two glycosylation steps. *J. Cell Biol.* **1983**, *96*, 999–1007.

(9) Sturm, A.; Van Kuik, J. A.; Vliegenthart, J. F. G.; Chrispeels, M. J. Structure, position and biosynthesis of the high mannose and the complex oligosaccharide side chains of the bean storage protein phaseolin. *J. Biol. Chem.* **1987b**, *262*, 13392–13403.

(10) Powers, J. R.; Whitaker, J. R. Purification and chemical properties of red kidney bean (*Phaseolus vulgaris*) α-amylase inhibitor. *J. Food Biochem.* **1977**, *1*, 217–238.

(11) Ciucanu, I.; Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **1984**, *131*, 209–217.

(12) Krzanowski, W. J. *Principles of Multivariate Analysis: A User's Perspective*; Clarendon Press: Oxford, U.K., 1988.

(13) Moreno, J.; Chrispeels, M. J. A lectin gene encodes the α-amylase inhibitor of the common bean. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7885–7889.

(14) Altabella, T.; Chrispeels, M. J. Tobacco plants transformed with the bean α*ai* gene express an inhibitor of insect α-amylase in their seeds. *Plant Physiol.* **1990**, *93*, 805–810.

(15) Shade, R. E.; Schroeder, H. E.; Pueyo, J. J.; Tabe, L. M.; Murdock, L. L.; Higgins, T. J. V.; Chrispeels, M. J. Transgenic pea seeds expressing the α-amylase inhibitor of the common bean are resistant to bruchid beetles. *Bio/Technology* **1994**, *12*, 793–796.

(16) Young, N. M.; Thibault, P.; Watson, D. C.; Chrispeels, M. J. Post-translational processing of two α-amylase inhibitors and an arcelin from the common bean, *Phaseolus vulgaris*. *FEBS Lett.* **1999**, *446*, 203–206.

(17) Yamaguchi, H.; Funaoka, H.; Iwamoto, H. Structures of sugar chains of the subunits of an α-amylase inhibitor from *Phaseolus vulgaris* white kidney beans. *J. Biochem.* **1992**, *111*, 388–395.

(18) Sawada, S.; Takeda, Y.; Tashiro, M. Primary structures of α- and *β*-subunits of α-amylase inhibitors from seeds of three cultivars of *Phaseolus* beans. *J. Protein Chem.* **2002**, *21*, 9–17.

(19) Casey, R.; Domoney, C. Pea globulins. In *Seed Proteins*; Shewry, P. R., Casey, R., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999; pp 171−208.