

Research Article

Partial Characterization of a Vicilin-Like Glycoprotein from Seeds of Flowering Tobacco (*Nicotiana sylvestris*)

Jared Q. Gerlach,^{1,2} Veer P. Bhavanandan,¹ Paul A. Haynes,^{3,4,5} and Lokesh Joshi^{1,2}

¹ Center for Glycosciences and Technology, Biodesign Institute at Arizona State University, 1001 S McAllister Avenue, Tempe, AZ 85287, USA

² National Centre for Biomedical Engineering Science, National University of Ireland, Galway, Ireland

³ Bio5 Institute for Collaborative Bioresearch, The University of Arizona, Tucson, AZ 85721, USA

⁴ Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde NSW 2109, Australia

⁵ Australian Proteome Analysis Facility, Macquarie University, North Ryde NSW 2109, Australia

Correspondence should be addressed to Lokesh Joshi, lokesh.joshi@nuigalway.ie

Received 5 February 2009; Accepted 31 March 2009

Recommended by Andrew Wood

A vicilin-like glycoprotein from the seeds of *Nicotiana sylvestris*, flowering tobacco, has been identified using nanoLC/ESI-MS/MS. Sequences from a fragment of protein demonstrated homology with vicilins from other members of the Solanaceae family, notably potato (*Solanum demissum*). Reducing and nonreducing SDS-PAGE analyses of the identified protein indicated that fragments resulting from in situ proteolytic processing are joined by *intrachain* disulphide bonds. Staining with Con A lectin was specifically inhibited by mannose suggested the presence of *N*-linked glycosylation which was confirmed by carbohydrate compositional analysis of PVDF-bound protein subunits. HPAEC-PAD analysis of the monosaccharides released from the glycoprotein by acid hydrolysis revealed glucosamine and mannose. *N*-acetylglucosamine termination of attached oligosaccharides was further verified by inhibitable WGA lectin staining. Immunostaining of PVDF-bound *N. sylvestris* proteins with antibodies against *G. max* total protein demonstrated cross-staining at masses corresponding to fragments from the proteolytically processed protein subunits.

Copyright © 2009 Jared Q. Gerlach et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Solanaceae encompasses many plants that are nutritionally and/or economically important including potatoes, tomatoes, eggplant (all *Solanum* spp.) and tobacco (*Nicotiana* spp.). Species of the genus *Nicotiana* have become leading proof-of-concept laboratory platforms for plant-based biopharmaceuticals [1–6] and have gained popularity as recombinant hosts through the development of successful transformation protocols including viral and bacterial methods [7, 8]. In 1990, tobacco was stably transformed with a gene encoding *Streptococcus mutans* protein antigen A, demonstrating that plants could be useful as vaccine-producing organisms [7]. Fully functional recombinant antibodies, which may find eventual use as therapeutics, have also been produced in tobacco plants [9, 10]. Seed deposition of recombinant proteins in tobacco species has also been demonstrated [9, 11, 12].

Seeds represent a highly desirable depository for recombinant protein products due to their excellent accumulation and stability [9, 11, 13]. Endogenous storage proteins are highly concentrated in seeds and constitute one of the largest caches of non-animal protein in nature. Storage proteins are systematically degraded during germination to ensure a constant supply of nitrogen and carbon-rich intermediates useful for synthesis of new materials [14, 15].

Previously, Sano and Kawashima [16] partially characterized the abundant 11S (also called legumin) protein component from *N. tabacum* seeds. As is the case with many nonleguminous, dicotyledonous (dicot) plants, the seed protein content of *Nicotiana* spp. is composed largely of 11S proteins [16]. However, dicot species may also produce large amounts of 7S (sometimes 8S) globulins more commonly referred to as vicilins.

Vicilin-type globulins may comprise up to 80% of protein found in some legume seeds [15, 17], but usually

constitute a significantly smaller percentage in nonlegumes. Typically, vicilins are sparsely glycosylated trimeric clusters of combined molecular mass between 140 and 190 kDa [18]. Under physiological conditions, vicilin complexes associate without interchain disulphide bonds. However, at higher concentrations, and in the presence of magnesium and calcium ions, they may also form complexes with greater numbers of subunits [15, 19]. Many vicilins undergo proteolytic modification, or “nicking,” and this processing creates a high degree of heterogeneity in the subunit population [12, 20, 21]. Vicilins of adzuki bean (*Vigna angularis*) have been shown to have trends in thermostability and surface hydrophobicity attributed to posttranslational proteolytic processing [21].

Vicilin subunits contain two cupin domains. Each cupin domain is a 6-stranded, short beta-barrel structure common to several subfamilies of the globulin superfamily of proteins, including the 11S storage proteins, to which they are closely related (Figure 1) [15, 22, 23]. Characterized vicilin glycoproteins from dicots soy (*Glycine max*) [24], pumpkin (*Cucurbita maxima*) [25], and carrots (*Daucus carota*) [26] have also been identified as allergens. Vicilin-like proteins from monocotyledonous species with characteristics similar to dicot 7S vicilins have been identified in wheat (*Triticum* spp.), barley (*Hordeum* spp.), and oat (*Avena* spp.) [27, 28].

While Higgins et al. [12] noted the presence of endogenous vicilin in their report of recombinant pea (*Pisum sativum*) vicilin expression in *N. tabacum* seeds, based on cross-reactivity of an antibody fraction made against the pea protein, no characterization of the endogenous protein was carried out at that time. This is the first report, to our knowledge, describing an endogenous, glycosylated, vicilin-like protein from plants of the *Nicotiana* genus.

2. Materials and Methods

2.1. Soluble Protein Extraction. Seeds from *Nicotiana sylvestris* were purchased from Johnny's Seeds (Winslow, Me, USA). All procedures performed at 4°C unless specified otherwise. Seeds (28 g) were ground to a uniform fine powder in a coffee grinder, and defatted in 100 mL cold acetone with rapid stirring for 1 hour. The acetone slurry was filtered, and the residue washed with one additional volume of fresh acetone and dried completely under vacuum. 2 g portions of the dry delipidated seed powder were first extracted using 20 mL double-distilled water (ddH₂O) with gentle shaking for 30 minutes. Albumin-containing supernatant was removed following centrifugation at 4,000 × g for 15 minutes. To extract the total soluble globulin protein (TSG) the pellet was suspended in 20 mL of extraction buffer containing 50 mM Tris-HCl, 200 mM NaCl, and 0.1 mM PMSE, pH 8.0 and the slurry continuously inverted for 1 hour. The majority of the insoluble content was removed by centrifugation for 15 minutes at 4,000 × g. Supernatant, in 1.5 mL aliquots, was further centrifuged (14,000 × g, 30 minutes) to yield TSG. Extracts from dried uncooked seeds of *G. max* and *C. maxima* were prepared using the same extraction buffer and general process but were subjected to neither delipidation nor separate extraction of

albumins prior to globulin extraction. Protein content was estimated by Bradford method [29] (absorbance at 595 nm) using a Spectramax spectrophotometer (Molecular Devices, Sunnyvale, Calif, USA).

2.2. Size-Exclusion Chromatography. Approximately 2.5 mg TSG prefiltered through 0.22 µm acetate membrane (Millipore, Billerica, Mass, USA) was subjected to fast protein liquid chromatography (FPLC) using a Superdex 200 10/30 size exclusion column (GE Health Sciences, Uppsala, Sweden). Isocratic elution was performed at a flow rate of 250 µL/min using 50 mM Tris-HCl, 200 mM NaCl, pH 8.0. The eluant was monitored by absorption at 280 nm, collected in fractions, and pooled appropriately. Elution calibration was performed using goat IgG (Mr 150 kDa), human transferrin (Mr 81 kDa), and human myoglobin (Mr 39 kDa).

2.3. SDS-PAGE and Enrichment of 7S Proteins. Crude and postsize exclusion TSG was subjected to SDS-PAGE analysis under reducing (R-SDS-PAGE) and nonreducing (NR-SDS-PAGE) conditions. Enrichment of 7S proteins was carried out by first separating pooled postsize exclusion protein fractions under nonreducing conditions in 12% Bis-Tris polyacrylamide gel (Invitrogen, Carlsbad, Calif, USA) followed by direct excision of the nonreduced vicilin (Mr ~48 kDa) band. The excised gel containing the 7S protein was macerated to a fine paste in a microcentrifuge tube. The protein was then re-extracted from the paste with the addition of 40 µL of 5 × R-SDS-PAGE sample buffer containing β-mercaptoethanol (β-ME) followed by agitation at 65°C for 30 minutes. An additional 20% v/v of 1 × reducing sample buffer was added, the contents vortexed briefly, and the gel pieces sedimented by centrifugation at 5000 × g for 1 minute. Supernatant containing the reduced protein was loaded onto 4–12% Bis-Tris polyacrylamide gel and separated in MES running buffer (Invitrogen).

2.4. Protein Sequence Identification. Protein samples were run on 4–15% Tris-HCl gels (Bio-Rad) under reducing conditions after completion of size exclusion chromatography. Gels were stained with Coomassie Brilliant Blue (CBB) in a fixative solution containing 0.05% (w/v) CBB, 30% methanol, and 10% acetic acid and partially destained in 30% methanol, 10% acetic acid for protein visualization. Bands of interest were excised and completely destained, macerated in deionized water and subjected to exhaustive trypsin digestion. Peptides were extracted from the gel pieces with 5 % formic acid/50 % acetonitrile, evaporated to dryness in a Speedvac vacuum centrifuge (Savant, Farmingdale, NY), and reconstituted in 10 µL of 1% formic acid.

A microbore HPLC system (Surveyor, Thermo, San Jose, Calif, USA) was modified to operate at capillary flow rates using a simple T-piece flow-splitter. Columns (8 cm × 100 µm I.D.) were prepared by packing 100 Å, 5 µm Zorbax C18 resin at 500 psi into columns with integrated electrospray tips made from fused silica, pulled to a 5 µm tip using a laser puller (Sutter Instrument Co., Novato, Calif,

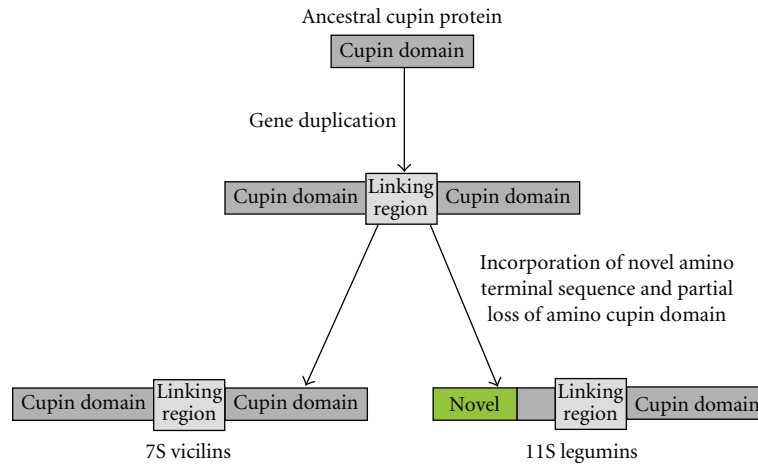


FIGURE 1: Proposed evolutionary link between 7S proteins (vicilins) and 11S proteins (legumins) based on the repeating structure of the cupin domain genes. Diagram modeled after Gibbs et al. [23].

USA). Electrospray voltage of 1.8 kV was applied using a gold electrode via a liquid junction upstream of the column. Samples were introduced onto the analytical column using a Surveyor autosampler (Thermo, San Jose, Calif, USA). The HPLC column eluent was eluted directly into the electrospray ionization source of a Thermo LCQ Deca ion trap mass spectrometer.

Peptides were eluted in a gradient using 5% acetonitrile, 0.1% formic acid (buffer A) and 90% acetonitrile, 0.1% formic acid (buffer B), at a flow rate of 500 nL/min. Following an initial wash with buffer A for 10 minutes, peptides were eluted with a linear gradient from 0–50% buffer B over a 25-minute interval, followed by 50–98% B over 5 minutes and a 5-minute wash at 98% B. Automated peak recognition, dynamic exclusion, and daughter ion scanning of the top three most intense ions were performed using the Xcalibur software as previously described [30]: (i) full mass survey scan 400–1500 amu, (ii) MS/MS of most abundant ion from survey scan, (iii) MS/MS of 2nd most abundant ion from survey scan, (iv) MS/MS of 3rd most abundant ion from survey scan. Other instrument parameters included: collision energy 39%, activation Q 0.25, activation time 30 milliseconds, isolation width 2.0 amu, dynamic exclusion enabled with repeat count 2, duration 0.5 minute, exclusion duration 5 minutes, exclusion mass width low 1.5 amu, high 1.5 amu.

2.5. Database Searching, Result Filtering, and Validation. MS/MS data were analyzed using SEQUEST run under Bioworks 3.1 (Thermo, San Jose, Calif, USA) [31, 32]. All spectra were searched against the NCBI nonredundant protein sequence database (downloaded August 2007). Search parameters included peptide and fragment mass tolerance of 2 Da and 0.2 Da, respectively, modification of cysteine with 57 amu (iodoacetamidation), and differential modification of methionine with 16 amu (oxidation).

Initial criteria for a preliminary positive peptide identification for a doubly charged peptide were a correlation

factor (Xcorr) greater than 2.5, a delta cross-correlation factor (dCn) greater than 0.1, a minimum of one tryptic peptide terminus, and a high preliminary scoring [33]. For triply charged peptides the correlation factor threshold was set at 3.5, and for singly charged peptides the threshold was set at 1.8. False positive rates for protein identification were assessed using reversed database searching [34, 35]. Imposing a minimum of two peptides per protein on each nanoLC-MS/MS data set [36], no proteins fitting the criteria were detected in the reversed database searches, indicating a protein identification confidence level of greater than 99% [37].

Peptide sequences identified from the NCBI non-redundant protein sequence database were used with the BLAST algorithm [38, 39] to find homologues. Homologues with greatest identity and greatest amount of characterization were considered. FASTA amino acid sequences were aligned using the ClustalW alignment tool (<http://align.genome.jp/>). Isoelectric point and mass estimation were performed using the ExPASy online algorithm (http://www.expasy.ch/tools/pi_tool.html).

2.6. Transfer of Proteins to PVDF and Lectin Staining. 15 µg of protein per well was loaded onto 4–15% Tris-HCl gels (Bio-Rad) and electrophoresed for approximately 1 hour at 50 mA (starting potential 150 V). Proteins were transferred to sequencing grade polyvinylidene difluoride (PVDF) 0.45 µm membrane (Millipore) in a semidry transfer apparatus (Bio-Rad) at 1.5 mA/cm² for 2 hours at a maximum of 15 V. Prior to transfer, PVDF was washed twice in 100% methanol and rinsed in transfer buffer. Blots were used directly after transfer or washed in ddH₂O and dried for storage at room temperature for later use.

Concanavalin A (Con A lectin) conjugated to alkaline phosphatase (EY Laboratories, San Mateo, Calif, USA) was used to probe for N-linked oligosaccharides. PVDF membranes with bound proteins from seeds of *N. sylvestris* were washed in 20 mM Tris-HCl, 100 mM NaCl, 0.05%

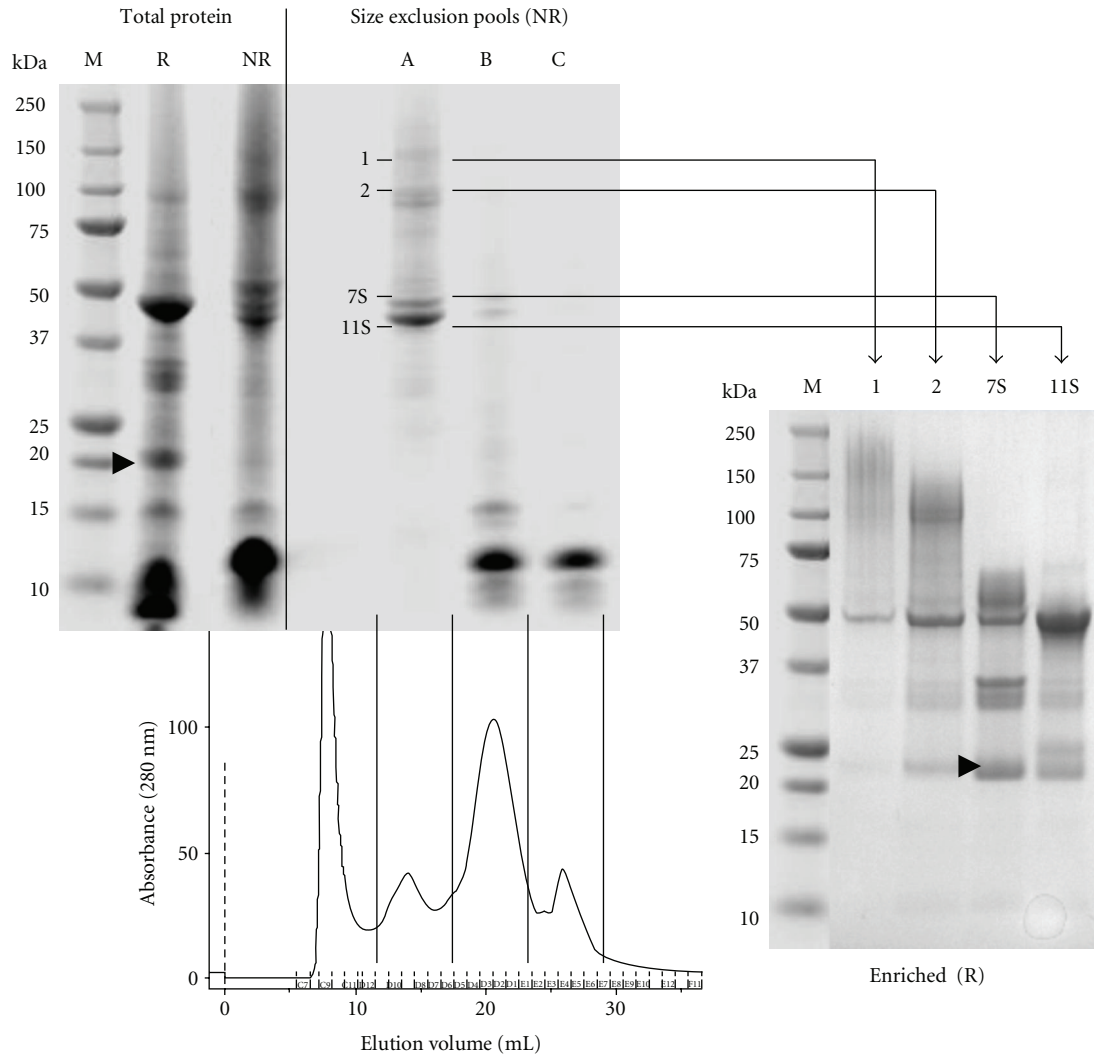


FIGURE 2: SDS-PAGE of *N. sylvestris* total protein under reducing (R) and nonreducing (NR) conditions. Size exclusion pools A, B, and C under nonreducing conditions showing relative positions of possible trimeric (1) and dimeric complexes (2) and 7S and 11S monomers. Enriched fractions prepared under reducing conditions showing separation of disulphide-dependent components, lanes 1, 2, 7S, and 11S. All gels 4–12% Bis-Tris. Protein was visualized by CBB staining. Arrowheads indicate the position of band excised for mass spectrometric analysis.

Tween 20, pH 7.4 (TBST), and then blocked for 1 hour in 1% porcine gelatin (Sigma-Aldrich, St. Louis, Mo, USA) in TBST. After washing twice (all washes 10 minutes) in TBST and once in lectin buffer (TBST plus 1 mM concentrations of MgCl_2 , MnCl_2 , and CaCl_2), Con A in lectin buffer ($20\text{ }\mu\text{g}$ in 15 mL) was added to the blot and incubated for 1 hour at room temperature with gentle agitation. A control blot was incubated in an identical manner with the addition of 10 mM mannose (Sigma-Aldrich) in the lectin buffer. After incubation in lectin solution, the blot was washed twice with TBST and once with ddH_2O . Lectin binding was visualized colorimetrically by reaction of BCIP/NBT (Sigma-Aldrich) and was stopped by addition of water. Alkaline phosphatase conjugated WGA (EY Labs) was used with the same protocol with the exception of the inhibitory procedure which consisted of preincubation of the lectin in 100 mM

GlcNAc for 1 hour prior to and 10 mM GlcNAc during the membrane incubation. Blots were dried at room temperature in the absence of light prior to being scanned with white light and stored as digital files.

2.7. Immunostaining. Polyclonal antibody prepared against *G. max* total seed protein (S2519, Sigma-Aldrich) was used to probe PVDF-bound proteins for cross-reactivity. SDS-PAGE samples of *G. max*, *C. maxima*, and *N. sylvestris* proteins as well as control proteins. Protein samples were prepared with β -ME as mentioned previously and separated in 4–12% Bis-Tris gel (Invitrogen). Control samples consisting of human transferrin, bovine fetuin, bovine serum albumin (BSA), asialofetuin, peanut lectin (PNA), and ribonuclease B (all Sigma-Aldrich) were prepared in the same manner. After protein was transferred to PVDF, the membranes were

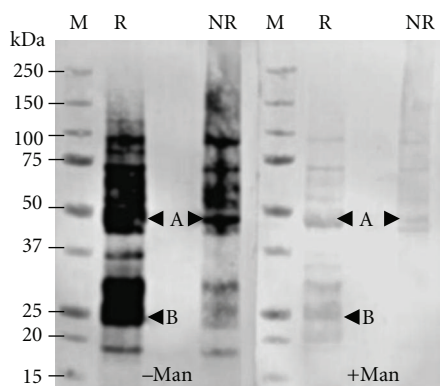


FIGURE 3: Con A staining without (–Man) and in presence of 10 mM mannose (+Man). Both blots show *N. sylvestris* seed protein after size exclusion chromatography with (R) and without (NR) β -ME. A and B indicate position of vicilin bands used for the TFA hydrolysis and subsequent HPAEC-PAD carbohydrate analysis. 4–15% Tris-HCl gel transferred to PVDF.

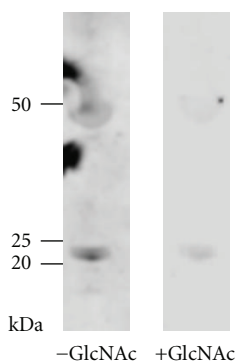


FIGURE 4: Protein from size exclusion pool A under reducing conditions stained with WGA (–GlcNAc). Inhibition through a combined 100 mM preprolonged and 10 mM prolonged incubation with GlcNAc is shown in comparison (+GlcNAc).

washed for 10 minutes in PBST (20 mM phosphate buffer, 140 mM NaCl, 20 mM KCl, 0.05% Tween 20, pH 7.4) prior to blocking for 30 minutes in 1% BSA + PBST. Antibody was used at 1 : 2000 dilution in PBST, and incubation time was 1 hour at room temperature with gentle shaking. Membranes were washed in PBST twice for 5 minutes prior to the addition of secondary antibody against rabbit IgG (A3812, Sigma-Aldrich), conjugated to alkaline phosphatase, used at a concentration of 1 : 10 000 in PBST. BCIP/alkaline phosphatase reaction was stopped by addition of water. The blot was rinsed twice for 30 seconds in ddH₂O and allowed to dry at room temperature in the dark prior to being scanned with white light and stored as a digital image.

2.8. Carbohydrate Analysis. Unstained protein bands were excised from PVDF membrane and hydrolyzed in sealed glass ampules containing 4 N trifluoroacetic acid (TFA) for 3 hours at 100°C. As a positive control for the hydrolysis conditions, bovine fetuin (Sigma-Aldrich) was also subjected to TFA. To correct for background, identical portions of PVDF

from buffer control lanes were subjected to TFA hydrolysis. Hydrolysates were neutralized by dropwise addition of 0.5 N sodium hydroxide and dried in a vacuum centrifuge (Labconco, Kansas City, Mo, USA). Dry carbohydrates were reconstituted in 100 μ L HPLC grade water prior to high-pH anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD, Dionex, Sunnyvale, Calif, USA). 20 μ L aliquots of the carbohydrate solutions were separated on a CarboPac PA10 column (4 $\times$ 250 mm, analytical) preceded by an amino trap column (4 $\times$ 50 mm) to minimize interference from protein degradation (Dionex, Sunnyvale, Calif, USA). Isocratic elution of 18 mM NaOH was performed at a flow rate of 1 mL/min over 18.5 minutes at a column temperature of 25°C. A standard panel consisting of fucose (Fuc), galactosamine (GalN), glucosamine (GlcN), galactose (Gal), glucose (Glc), and mannose (Man) was used as a reference.

3. Results

3.1. Protein Extraction and Visualization. 200 mM NaCl was used during the protein extraction due to the insignificant increase in solubility with 500 mM NaCl and to aid in downstream analysis and processing. Soluble globulin protein was readily extracted from *N. sylvestris* seeds at slightly alkaline pH, but solubility dropped sharply below pH 8.0. Although tris-(hydroxymethyl)-aminomethane (Tris) is a weak chelator, no additional measures were taken to sequester metal ions during protein extraction. Solubility reached a maximum of approximately 5 mg/mL at 4°C at which point aggregates of protein began to form and precipitate. The presence of free Mg²⁺ and Ca²⁺ ions may have been a limiting factor on the maximum total protein solubility [40]. R-SDS-PAGE of TSG produced one major band at approximately 48 kDa and several bands between ~31 and ~35 kDa and between ~20 and ~23 kDa as well as many smaller polypeptides which were poorly resolved (Figure 2, Total protein, lane R). A separation of protein aggregates >300 kDa by centrifugation in 300 kDa molecular weight cut-off (MWCO) spin filters resulted in a pellet which could be resolubilized to give a similar protein profile to the total extract (data not shown), leading to the conclusion that the globulins aggregate into poorly soluble, high molecular mass complexes when in high concentration as noted by Freitas et al. [14]. A portion of the protein also aggregated after heat denaturation with SDS and did not readily migrate within polyacrylamide gel, resulting in streaking and areas of poor resolution at relative masses greater than 75 kDa as shown in Figure 2 (Total protein, lanes R and NR). Molecular masses corresponding to complexes of proteins (~100 to 200 kDa) were visualized in the gel and were prevalent in the total protein mixture (Figure 2, Total protein, R and NR and in size exclusion pool A, bands 1 and 2).

3.2. Size Exclusion Chromatography and SDS-PAGE. Size exclusion chromatography was used as preparative separation method for TSG. Isocratic elution from size-exclusion media via FPLC produced three major peaks in the 280 nm absorbance curve (Figure 2, Size exclusion pools A, B, and C)

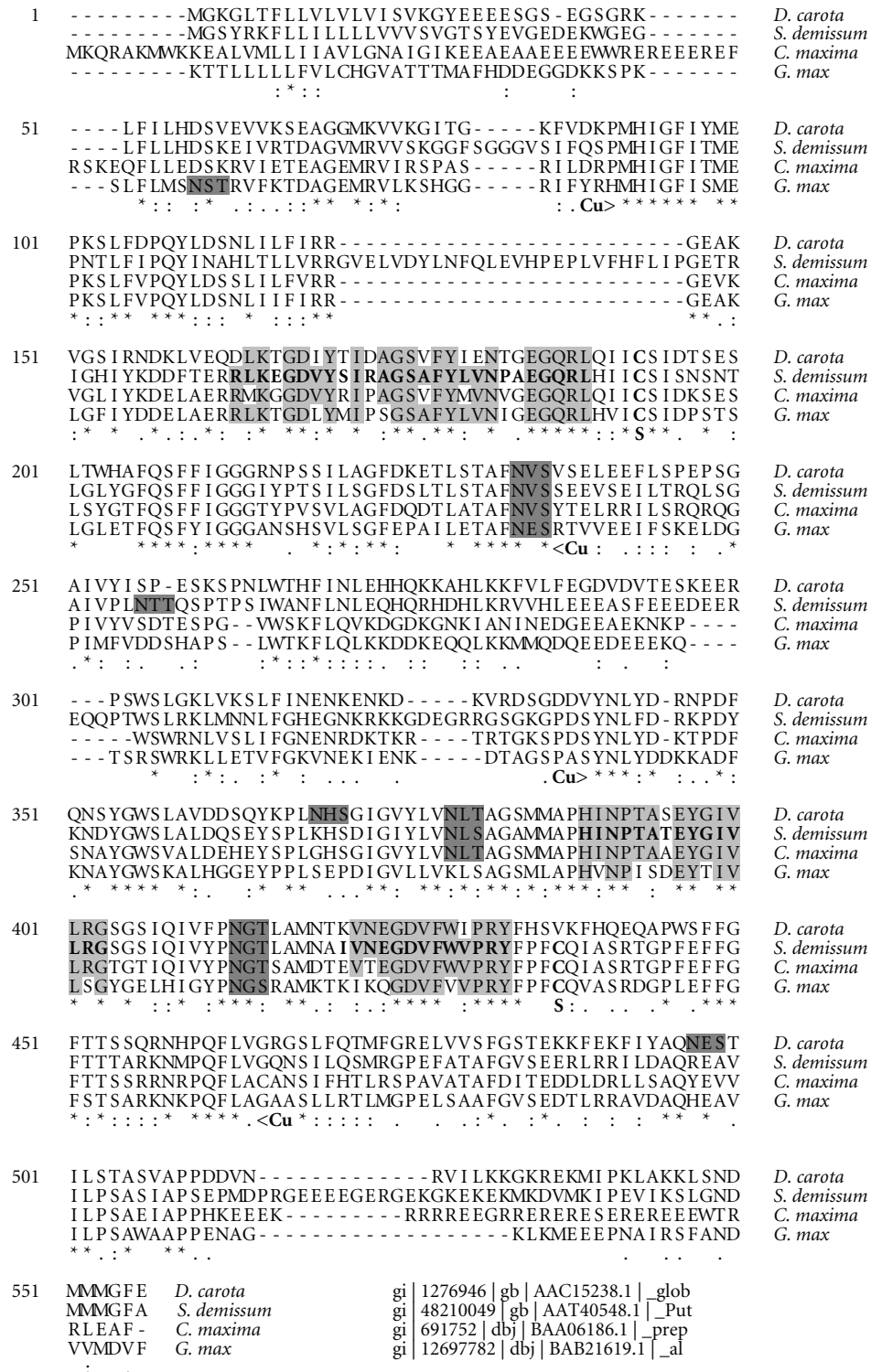


FIGURE 5: ClustalW alignment of sequences from *D. carota*, *S. demissum*, *C. maxima*, and *G. max*. Amino acids in bold print represent sequences obtained from putative vicilin of *N. sylvestris* where light gray sections indicate identical amino acids. Asterisks (*) under sequences indicate identically conserved residues while double dots (:) indicate conserved substitutions, and single dots (·) indicate semiconserved substitutions. Approximate beginning and ending of each cupin domain are indicated by *Cu* > or < *Cu*, respectively. Cysteine residues are shown in bold print and marked by *S*. *N*-glycosylation sites (*Asn-Xaa-Ser/Thr*) are marked in gray boxes.

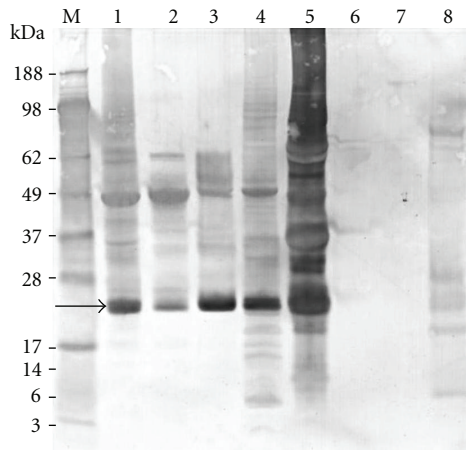


FIGURE 6: Cross-reactivity of anti-*G. max* IgG. Lane M, mass marker; lane 1, total soluble protein from *N. sylvestris*; lane 2, 11S-enriched fraction from *N. sylvestris*; lane 3, 7S (vicilin)-enriched fraction from *N. sylvestris*; lane 4, total soluble protein from *C. maxima*; lane 5, total soluble protein from *G. max*; lane 6, bovine serum albumin; lane 7, bovine fetuin; lane 8, mixed animal and plant protein control consisting of human transferrin (~75 kDa), bovine serum albumin (~62 kDa), asialofetuin (~50 kDa), peanut lectin (~25 kDa), and ribonuclease B (~20 kDa). Arrow indicates position of homologous allergens from *G. max* (23 kDa peptide from Bd 28 K) and *C. maxima* (MP27/MP32). 4–12% Bis-Tris gel transferred to PVDF.

with the largest percentage of proteins eluting together over a range of 140–160 kDa in pool A. Post-size-exclusion fractions of the protein demonstrated a similar NR-SDS-PAGE profile in comparison to total soluble protein (Figure 2, lane NR versus size exclusion pool A), although the number of major bands present under nonreducing conditions was significantly less in the post size-exclusion pool A. Nonreducing conditions produced a pair of distinct bands at ~44 kDa and ~48 kDa (Figure 2, Size exclusion pool A, bands 7S and 11S).

Based on the previous characterization of the 11S proteins from *N. tabacum* [16], it was surmised that the ~44 kDa band is 11S protein (Figure 2, Size exclusion pool A, band 11S).

Attempted anion exchange separation of intact 7S proteins resulted in poor resolution and considerable precipitation on the column (data not shown). Alternatively, for the purpose of this initial characterization, an enrichment step for the individual protein fractions was performed using a two-step SDS-PAGE procedure following size-exclusion chromatography. Individual bands were isolated from the post-size-exclusion NR-SDS-PAGE gel as indicated (Figure 2, Size exclusion pool A, bands 1, 2, 7S, and 11S), treated with β -ME, and electrophoresed a second time (Figure 2, Enriched).

Protein bands at ~100 kDa and ~150 kDa Mr which may have, respectively, represented di- and tri-meric complexes comprised of ~50 kDa structures are visible in both NR-SDS-PAGE (Figure 2, Size exclusion pool A, bands 1 and 2) and R-SDS-PAGE (Figure 2, Enriched, lanes 1 and 2) gels. It is

TABLE 1: Ion proportion and mass to charge (m/z) data from nanoLC/ESI-MS/MS for the corresponding sequences identified.

Ion proportion	$[m + H]^+$	Sequence
83.3	1180.85	R.LKEGDVYSIR.A
67.9	1580.69	R.AGSIFYLVNPAEGQR.L
75	1448.06	H.INPTATEYGIVLR.G
80	1318.73	I.VNEGDFVFWVPR.Y

unlikely, however, that these are complexes of 7S proteins based on the subsequent R-SDS-PAGE patterns (Figure 2, Enriched, lanes 1 and 2). R-SDS-PAGE of the excised and enriched 7S band (Figure 2, Size exclusion pool A, band 7S) displayed a large diffuse band above 50 kDa, a small but distinct band at ~48 kDa, and several fragments between 20 and 37 kDa (Figure 2, Enriched, lane 7S).

3.3. Sequence Identity and Comparison. Bands excised from R-SDS-PAGE gels were subjected to trypsin digestion and subsequent mass spectrometric analysis by nanoLC/ESI-MS-MS. Four sequence fragments were identified from a protein band of ~22 kDa (Figure 2, Total protein, lane R, arrow and Enriched, lane 7S, arrow) as shown in Table 1. The four masses were found to entirely match portions of the sequence from the putative vicilin protein from *Solanum demissum*. Of these, two were adjacent and the sequences overlapped by two residues. From the NCBI nonredundant protein database, there were no additional identical hits for the individual sequences obtained from MS/MS. A number of less homologous matches, including nonstorage proteins and functional proteins and precursors, were identified when each sequence component was compared to sequences from other members of Solanaceae (not shown).

3.4. Lectin Blotting and Carbohydrate Analysis. Concanavalin A (Con A) has a high affinity for mannose [41] found in *N*-linked oligosaccharide chains. Staining of *N. sylvestris* seed protein with Con A (Figure 3) was readily inhibited by the addition of 10 mM mannose to the incubation buffer. Monosaccharides hydrolyzed from PVDF-bound bands (unstained) after R-SDS-PAGE separation of *N. sylvestris* protein were analyzed by HPAEC-PAD. Quantities of each monosaccharide (picomoles) were then used to establish a ratio in relation to Man as shown in Table 2. Deacetylation of amino sugars occurs as a result of the TFA hydrolysis, and the resulting peaks corresponding to GlcN and GalN are representative of *N*-acetylglucosamine (GlcNAc) and *N*-acetylgalactosamine (GalNAc), respectively. The monosaccharide profile of the ~48 kDa mass and smaller proteolytically processed components (~20–24 kDa) showed the presence of differing glycosylation among these populations (Table 2). Glc, which is a ubiquitous contaminant, was also present and was not evaluated. No evidence of Fuc was present.

WGA lectin, which has been reported to bind to GlcNAc residues [42], was used to probe the post-size-exclusion pool A proteins. The pool A protein was separated

gb | BT013421.1 | *Solanum lycopersicum* clone 132059R, mRNA sequence

Length = 1748

Score = 157 bits (396), Expect(2) = 0

Identities = 109/143 (76%), positives = 112/143 (78%), gaps = 26/143 (18%)

Frame = +2

<i>S.dem</i>	1	MGSYRKFLllllllvvvsvGTSYEVGEDEKWGEGFLFLHDSKEIVRTDAGVMRvvskggf	60
		MGSYR+FLLI LLL VVVSVGTSYEV E+EKWGEGLFLHDSKEIVRTDAGVMRVVSKGGF	
<i>L.esc</i>	53	MGSYREFLLI LLLFVVVSVGTSYEV EEEEEKWGEGFLFLHDSKEIVRTDAGVMRVVSKGGF	232
<i>S.dem</i>	61	sgggvsIFQSPMHIGFITMEPNTLFI PQYINAHLTLLVRRGVELVDYLNFLQLEVHPEPLV	120
		GGGVSI FQSPMHIGFITMEPNTLFI PQY+NA LTL VRR	
<i>L.esc</i>	233	GGGVSI FQSPMHIGFITMEPNTLFI PQYMNALTLFVRR - - - - -	352
<i>S.dem</i>	121	FHFLIPGETRIGHIYKDDFTERR	143
		GETRIGHIYKDDFTERR	
<i>L.esc</i>	353	- - - - GETRIGHIYKDDFTERR	403

Identities = 370/396 (93%), positives = 383/396 (96%), gaps = 0/396 (0%)

Frame = +3

<i>S.dem</i>	140	TER RLKEGDVYSIRAGSAFYLVNPAEGQRL HIICSISNSNTLGLYGFQSF F IGGGIYPTS	199
		T L +GD+YSIRAGSAFYLVNPAEGQRLHIICSISNSN LGLYGFQSF F IGGGIYPTS	
<i>L.esc</i>	378	TRMILLKGDIIYSIRAGSAFYLVNPAEGQRLHIICSISNSN LGLYGFQSF F IGGGIYPTS	557
<i>S.dem</i>	200	ILSGFDSLTLSTAF NVS SEEVSEILTRQLSGAIVPL NTT QSPTPSIWANFLNLEQHQRHD	259
		ILSGFD+LTLSTAF NVS SEEVSEILTRQLSGAIVPL NTT QSPTPSIWANFLNLEQHQRHD	
<i>L.esc</i>	558	ILSGFDTLTLSTAF NVS SEEVSEILTRQLSGAIVPL NTT QSPTPSIWANFLNLEQHQRHD	737
<i>S.dem</i>	260	HLKRVVHLeeeasfeeedeereQQPTWSLRKLMNNLFGHEGNKRKKKGDEGRRGSGKGPDS	319
		HLKRVVHLEEEAS EEEDE REQPTWSLRK M NLFGEHGNKRKKKGDEGRRGSGKGPDS	
<i>L.esc</i>	738	HLKRVVHLEEEAS EEEDEGREQPTWSLRKFMINLFGHEGNKRKKKGDEGRRGSGKGPDS	917
<i>S.dem</i>	320	YNLFDRKPDYKNDYGWSLALDQSEYSPLKHSDIGIYLV NLS SAGAMMAP HINPTATEYGIV	379
		YNLFDRKPDYKNDYGWSLALD++EYSPLKHSDIG+YLV NLS SAGAMMAPHINPTATEYGIV	
<i>L.esc</i>	918	YNLFDRKPDYKNDYGWSLALDRTEYSPLKHSDIGVYLV NLS SAGAMMAPHINPTATEYGIV	1097
<i>S.dem</i>	380	LRGSGS IQIYVP NGT LAMNA IVNEGDVFWVPRY FPFCQIASRtgpfef f g f t t t ARKNMP	439
		LRGSGS IQIYVP NGT LAMNA+VNEGDVFWVPRYFPFCQIASRTGPF E F F G F T T T RKNMP	
<i>L.esc</i>	1098	LRGSGS IQIYVP NGT LAMNAVNEGDVFWVPRYFPFCQIASRTGPF E F F G F T T T GRKNMP	1277
<i>S.dem</i>	440	QFLVGQNSILQSMRGPEFATAFGVSEERLRRILDAQREAVILPSASIA PSEPMDPrg e e e	499
		QFLVGQNSILQSMRGP+FA AFGVSEERLRRILD+QREAVILPSASIA PSEPMDPRG E E E	
<i>L.esc</i>	1278	QFLVGQNSILQSMRGPQFAA AFGVSEERLRRILDSQREAVILPSASIA PSEPMDPRG E E E	1457
<i>S.dem</i>	500	egergekgk e k e kMKDVMKIPEVIKSLGNDMMMGFA	535
		EGERGEKGKEK+KMKDVMK+PEVIKSLG+D+MMG A	
<i>L.esc</i>	1458	EGERGEKGKEKKMKDVMKMP E V I K S L G D D I M M G L A	1565

FIGURE 7: Split comparison of *S. demissum* putative vicilin sequence with mRNA-derived *in silico* polypeptide from *S. lycopersicum* (tBLASTn). Top, comparison of the first 143 amino acids from *S. demissum* sequence to the amino end of the derived *S. lycopersicum* polypeptide from nucleotides 53 through 403. Bottom, comparison between *S. demissum* amino acids 140–535 and the cupin domain of the *in silico* translation of *S. lycopersicum* carboxy domain nucleotides 378 through 1565. Bold amino acids in grey blocks are those identified from *N. sylvestris*. Potential Asn-Xaa-Ser/Thr glycosylation sites are marked by grey box. Small letters represent regions subjected to repeat-motif filtering during BLAST execution.

emb | AJ617015.1 | *Nicotiana plumbaginifolia*, cDNA-AFLP fragment, clone Np168

Length = 181

Identities = 39/41 (95%), positives = 40/41 (97%), gaps = 0/41 (0%)

Frame = +3

<i>S.dem</i>	6	TATEYGIVLRGSGS IQIYVP NGT LAMNA IVNEGDVFWVPRY	46
		TATEYGIVLRGSGS IQIYVPNGT LAMN+ VNEGDVFWVPRY	
<i>N.plu</i>	3	TATEYGIVLRGSGS IQIYVP NGT LAMNSRVNEGDVFWVPRY	125

FIGURE 8: Alignment of *S. demissum* sequences with cDNA-derived fragment from *N. plumbaginifolia*. Bold amino acids in grey blocks represent fragments of peptide identified from *N. sylvestris*. Asn-Gly-Ser/Thr glycosylation site is marked by grey box.

TABLE 2: Compositional analysis of monosaccharides hydrolyzed from PVDF-bound components of post-size-exclusion *N. sylvestris* protein. Values represent picomole ratios for each monosaccharide relative to Man for each sample.

Sample	Man	GlcN	GalN	Gal
A	1.00	1.88	0.08	0.08
B	1.00	0.85	0.04	0.04

via R-SDS-PAGE and transferred to PVDF. Although the band at ~48 kDa was lightly stained by WGA, prominent staining of bands between 20 and 25 kDa was observed (Figure 4). Staining was inhibited by preincubating the lectin followed by sustained presence of 10 mM GlcNAc in the blot incubation, indicating a legitimate carbohydrate-mediated binding between the PVDF-bound glycoproteins and WGA (Figure 4).

Prediction of glycosylation for a portion of the *S. demissum* vicilin incorporating sequences obtained from *N. sylvestris*, shown in Figure 5, indicated a possible *N*-linked site (Asn-Gly-Ser/Thr). The lack of defined mass to charge (*m/z*) for a peptide sequence at this region in the *N. sylvestris* protein spectra was most likely due to the presence of the oligosaccharide during MS analysis. Further comparison of amino acid sequences from *G. max*, *C. maxima*, *D. carota*, and *L. esculentum* (from the mRNA), to the *S. demissum* peptide sequence confirms that the glycosylation point within the cupin domain at the carboxy end of each of these peptide chains is highly conserved (Figure 5).

3.5. IgG Cross-Reactivity. The hypothesis that sequence homology alone may be an indicator of antibody cross-reactivity was explored by immunoblotting. An IgG fraction prepared against *G. max* total seed protein was used to probe PVDF-bound extracts from *N. sylvestris* and *C. maxima*. Anti-*G. max* IgGs bind to proteins from *C. maxima* seeds and *N. sylvestris* (Figure 6). Staining of *N. sylvestris* protein is most intense at molecular masses between 20 and 25 kDa and also corresponds to the known vicilin proteins and fragments from *C. maxima* (MP27/MP32) and *G. max* (Bd 28 K and 23 kDa glycopeptides from that protein) (Figure 6, arrow), reinforcing the possibility that vicilin proteins from tobacco may be similar immunogenic agents. Further testing of *N. sylvestris* protein samples with sera from presensitized individuals was not undertaken in this study.

4. Discussion

A glycoprotein from the seeds of *N. sylvestris* has been identified that it undergoes proteolytic cleavage and exhibits homology with vicilin proteins from other members of Solanaceae as well as other plant families. The protein forms intrachain disulfide bonds, forms noncovalently bound complexes of high molecular mass, has asparagine-linked glycosylation and is recognized by antibodies raised against total seed protein from *G. max*.

One of the hallmarks of the 7S vicilin protein family is the lack of cysteine residues and resulting *interchain*

disulphide bonds. The putative vicilin protein identified from *N. sylvestris* does, however, form *intrachain* disulphide bonds which can be broken under reducing conditions. Observation of the presence of cysteines in the sequence of the *S. demissum* vicilin to which the *N. sylvestris* protein fragments share homology indicated the possibility of *interchain* disulphide bonding between residues near the middle and toward the carboxy terminus of the polypeptide chain as indicated in Figure 5 (marked S below alignment). It can be inferred from this that proteolytic activity must occur at a midpoint between these two cysteines as the addition of β -ME resulted in separation of polypeptides with masses less than those observed after the initial NR-SDS-PAGE separations. Vicilins from *C. maxima* and *G. max* share these cysteine residues at conserved positions, although it should be noted that the *D. carota* protein, which was reported to be isolated from tap root tissue instead of seeds, has only one cysteine residue within the cupin domain at the amino end and no cysteine residues in the carboxy cupin domain as shown in Figure 5.

Although vicilins from *D. carota* (AAC15238) and *G. max* (BAB21619) have two complete cupin regions and vicilin symmetry as a result of this, the vicilin from *S. demissum* (AAT40548) varies in that the amino-terminal cupin region is incomplete at the amino end and incorporates a 26-amino acid insertion. Although insertions such as this have been proposed [23] as the evolutionary mechanism that led to the emergence of the extant structure of 11S seed proteins (Figure 1), the *S. demissum* protein shares more significant identity with 7S proteins (as evidenced through sequence comparison, data not shown). A search for nucleotide homology (tBLASTn) limited to the family Solanaceae produced a nearly identical match for the carboxy and amino terminal gene sequences of vicilin mRNA from *S. lycopersicum* (Figure 7), and a cDNA fragment from *N. plumbaginifolia*, commonly referred to as “Tex-Mex” tobacco (Figure 8). *In silico* translation of the mRNA from *S. lycopersicum* (BT013421.1) resulted in a polypeptide containing one partial cupin domain. This domain, which is toward the carboxy terminus of the peptide, has 93% identity and 96% homology with the putative *S. demissum* vicilin (Figure 7). At the amino end, the first 143 amino acids of *S. demissum* vicilin gave an identity of 76% (109/143) and homology of 78% (112/143) when compared to the *S. lycopersicum* sequence. Comparison of the sequences also revealed the highly conserved *N*-linked glycosylation site found in each of the other vicilins examined here.

The theoretical isoelectric point (pI) of the complete *S. demissum* vicilin sequence is 5.9 as determined by using the ExPASy online algorithm. In comparison, vicilin-like proteins from *N. sylvestris* were poorly solubilized below pH 7.5 and nearly insoluble at pH 6. The rapid decrease in solubility below pH 8 corresponds well with the findings detailed in a proteome analysis of endosperm and embryos from *S. lycopersicum* in which vicilin-like proteins isolated from the endosperm of seeds were determined to have pI values ranging from 6.9 to 8.1 [43].

Plants have many advantages as hosts for recombinant technologies, including the possibility of high yield, freedom

from animal pathogen contamination, and low production cost overhead [7, 44]. It is plausible, however, that endogenous protein allergens may be coextracted along with recombinant proteins [18, 22]. The vicilin glycoproteins from soy beans, pumpkin seeds, and tap roots of carrots compared here have also been identified as allergens, and each of these proteins has a highly conserved peptide sequence (Figure 5). Previously, evidence has been presented to suggest that those individuals presensitized to vicilin proteins from one plant species may be affected by IgE recognition of vicilins derived from these and additional species [20, 45]. Although allergy is a function of IgE antibodies, cross-reactive IgG binding (Figure 6) demonstrates the possibility that antibodies of other classes may bind to epitopes present on *N. sylvestris* protein.

Considerable interest has been given to cross-reactivity of seed proteins in relation to the increasing occurrence of severe allergic reaction [22]. General areas of conserved peptide sequences comprising large domains across the cupin superfamily (e.g., the cupin domains themselves) appear to be similar enough to produce cross-reactivity [22]. Of particular interest is the cross-species similarity of the more carboxy-localized cupin domains (Figure 5). It is tempting to assume that homology alone is a suitable determinant of allergenicity and Ig cross-reactivity, and this has previously been addressed without a firm consensus having been established [18]. It is possible that antibody cross-reactivity in vivo is dependent upon a combination of conformational similarity and sequence homology [18, 22, 46] and not exclusively one or the other. Given the degree of homology between the putative vicilin we have identified from *N. sylvestris* and the other vicilins discussed here, cross-reactivity prediction on the basis of peptide sequence homology alone is reasonable. Our results demonstrating that antibodies against *G. max* total seed protein also bind to proteins from *C. maxima* and *N. sylvestris* are in agreement with the above supposition.

Based on compositional analysis of monosaccharides released by acid hydrolysis, glycosylation of the vicilin-like protein from *N. sylvestris* is not uniform among subunits and fragments (Figures 3 and 4, Table 2). High molecular mass bands (Figure 3, band A), representative of vicilin monomers not subjected to proteolytic modifications, showed a nearly two-fold presence of GlcN with respect to Man, while the ~22 kDa (Figure 3, band B) fragments separated under reducing conditions produced a GlcN : Man ratio closer to 1 as shown in Table 2. Interestingly, when exposed to anti-*G. max* IgG, the ~22 kDa fragments appeared to stain with greater intensity than the larger (~50 kDa) mass (Figure 6, arrow, lane 3). These bands also stained with Con A and WGA lectins (Figures 3 and 4). The vicilin Bd 28K from *G. max* was reported to have a dependence upon the N-linked glycans attached to that protein in the binding of serum IgE from patients with *G. max* allergy [47]. It was determined that the Asn41 residue of Bd 28K is glycosylated (this site is shown in Figure 5) and that IgE binding to the protein appeared to be greatly enhanced by the presence of the N-linked oligosaccharide. Later work focusing primarily on the carboxy region of Bd 28K and on its proteolytically derived

23 kDa fragments demonstrated that Asn349 is glycosylated and that sera from *G. max*-sensitive patients showed reduced binding to fragments produced in *E. coli* due to the absence of the oligosaccharide [24]. However, recombinant peanut proteins rAra h2 and rAra h3, also produced in *E. coli* and therefore lacking glycosylation, were found to be allergens comparable to their counterparts naturally produced in peanut seeds [48, 49]. From these contrasting findings, it may be inferred that epitope similarities between seed proteins which originate both from the amino acid sequence and carbohydrate modifications have the potential to produce an immune response in presensitized individuals.

Abbreviations

CBB:	Coomassie Brilliant Blue
FPLC:	Fast protein liquid chromatography
PMSF:	Phenylmethylsulfonylfluoride
HPAEC-PAD:	High-pH anion exchange chromatography with pulsed amperometric detection
NanoLC/ESI-MS/MS:	Nanoscale liquid chromatography-electrospray ionization tandem mass spectrometry
SDS-PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ig:	immunoglobulin

Acknowledgments

J. Q. Gerlach and L. Joshi would like to thank Amy D. Smith, Mostafa Sheykhnazari, Miti Shah, Vinay Nagaraj, Michelle Kilcoyne, and Sasha Daskolova for valuable discussions and analysis advice. P. A. Haynes would like to thank Vicki Chandler and Stuart Bourne for support. Funding for this research was provided by the Biodesign Institute, Arizona State University, the Wallace Research Foundation, and Bio5 Institute.

References

- [1] A. Sorrentino, S. Schillberg, R. Fischer, R. Porta, and L. Mariniello, "Molecular farming of human tissue transglutaminase in tobacco plants," *Amino Acids*, vol. 36, no. 4, pp. 765–772, 2009.
- [2] M. D. McLean, K. C. Almquist, Y. Niu, et al., "A human anti-*Pseudomonas aeruginosa* serotype O6ad immunoglobulin G1 expressed in transgenic tobacco is capable of recruiting immune system effector function in vitro," *Antimicrobial Agents and Chemotherapy*, vol. 51, no. 9, pp. 3322–3328, 2007.
- [3] K. C. Almquist, M. D. McLean, Y. Niu, et al., "Expression of an anti-botulinum toxin A neutralizing single-chain Fv recombinant antibody in transgenic tobacco," *Vaccine*, vol. 24, no. 12, pp. 2079–2086, 2006.
- [4] P. Galeffi, A. Lombardi, I. Pietraforte, et al., "Functional expression of a single-chain antibody to ErbB-2 in plants and cell-free systems," *Journal of Translational Medicine*, vol. 4, article 39, pp. 1–13, 2006.

- [5] S. Reggi, S. Marchetti, T. Patti, et al., "Recombinant human acid β -glucosidase stored in tobacco seed is stable, active and taken up by human fibroblasts," *Plant Molecular Biology*, vol. 57, no. 1, pp. 101–113, 2005.
- [6] B. Wagner, K. Hufnagl, C. Radauer, et al., "Expression of the B subunit of the heat-labile enterotoxin of *Escherichia coli* in tobacco mosaic virus-infected *Nicotiana benthamiana* plants and its characterization as mucosal immunogen and adjuvant," *Journal of Immunological Methods*, vol. 287, no. 1–2, pp. 203–215, 2004.
- [7] F. Sala, M. Manuela Rigano, A. Barbante, B. Basso, A. M. Walmsley, and S. Castiglione, "Vaccine antigen production in transgenic plants: strategies, gene constructs and perspectives," *Vaccine*, vol. 21, no. 7–8, pp. 803–808, 2003.
- [8] R. G. Birch, "Plant transformation: problems and strategies for practical application," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 48, no. 1, pp. 297–326, 1997.
- [9] S. Petruccioli, M. S. Otegui, F. Lareu, et al., "A KDEL-tagged monoclonal antibody is efficiently retained in the endoplasmic reticulum in leaves, but is both partially secreted and sorted to protein storage vacuoles in seeds," *Plant Biotechnology Journal*, vol. 4, no. 5, pp. 511–527, 2006.
- [10] V. Gomord, C. Sourrouille, A.-C. Fitchette, et al., "Production and glycosylation of plant-made pharmaceuticals: the antibodies as a challenge," *Plant Biotechnology Journal*, vol. 2, no. 2, pp. 83–100, 2004.
- [11] E. Stoger, J. K.-C. Ma, R. Fischer, and P. Christou, "Sowing the seeds of success: pharmaceutical proteins from plants," *Current Opinion in Biotechnology*, vol. 16, no. 2, pp. 167–173, 2005.
- [12] T. J. V. Higgins, E. J. Newbigin, D. Spencer, D. J. Llewellyn, and S. Craig, "The sequence of a pea vicilin gene and its expression in transgenic tobacco plants," *Plant Molecular Biology*, vol. 11, no. 5, pp. 683–695, 1988.
- [13] D. A. Goldstein and J. A. Thomas, "Biopharmaceuticals derived from genetically modified plants," *QJM*, vol. 97, no. 11, pp. 705–716, 2004.
- [14] R. L. Freitas, A. R. Teixeira, and R. B. Ferreira, "Vicilin-type globulins follow distinct patterns of degradation in different species of germinating legume seeds," *Food Chemistry*, vol. 102, no. 1, pp. 323–329, 2007.
- [15] P. R. Shewry, J. A. Napier, and A. S. Tatham, "Seed storage proteins: structures and biosynthesis," *Plant Cell*, vol. 7, no. 7, pp. 945–956, 1995.
- [16] M. Sano and N. Kawashima, "Isolation and partial characterization of the major seed protein from *Nicotiana tabacum*, and the accumulation during development," *Agricultural and Biological Chemistry*, vol. 47, no. 6, pp. 1305–1310, 1982.
- [17] T. Itoh, R. N. Garcia, M. Adachi, et al., "Structure of 8S α globulin, the major seed storage protein of mung bean," *Acta Crystallographica Section D*, vol. 62, no. 7, pp. 824–832, 2006.
- [18] J. D. Astwood, A. Silvanovich, and G. A. Bannon, "Vicilins: a case study in allergen pedigrees," *Journal of Allergy and Clinical Immunology*, vol. 110, no. 1, pp. 26–27, 2002.
- [19] R. B. Ferreira, R. L. Freitas, and A. R. Teixeira, "Self-aggregation of legume seed storage proteins inside the protein storage vacuoles is electrostatic in nature, rather than lectin-mediated," *FEBS Letters*, vol. 534, no. 1–3, pp. 106–110, 2003.
- [20] R. Sanchez-Monge, G. Lopez-Torrejón, C. Y. Pascual, J. Varela, M. Martín-Esteban, and G. Salcedo, "Vicilin and convicilin are potential major allergens from pea," *Clinical and Experimental Allergy*, vol. 34, no. 11, pp. 1747–1753, 2004.
- [21] T. Fukuda, K. Prak, M. Fujioka, N. Maruyama, and S. Utsumi, "Physicochemical properties of native adzuki bean (*Vigna angularis*) 7S globulin and the molecular cloning of its cDNA isoforms," *Journal of Agricultural and Food Chemistry*, vol. 55, no. 9, pp. 3667–3674, 2007.
- [22] C. Radauer and H. Breiteneder, "Evolutionary biology of plant food allergens," *Journal of Allergy and Clinical Immunology*, vol. 120, no. 3, pp. 518–525, 2007.
- [23] P. E. Gibbs, K. B. Strongin, and A. McPherson, "Evolution of legume seed storage proteins—a domain common to legumins and vicilins is duplicated in vicilins," *Molecular Biology and Evolution*, vol. 6, no. 6, pp. 614–623, 1989.
- [24] M. Hiemori, H. Ito, M. Kimoto, et al., "Identification of the 23-kDa peptide derived from the precursor of Gly m Bd 28 K, a major soybean allergen, as a new allergen," *Biochimica et Biophysica Acta*, vol. 1675, no. 1–3, pp. 174–183, 2004.
- [25] R. Fritsch, H. Ebner, D. Kraft, and C. Ebner, "Food allergy to pumpkinseed—characterization of allergens," *Allergy*, vol. 52, no. 3, pp. 335–337, 1997.
- [26] B. K. Ballmer-Weber, B. Wüthrich, A. Wangorsch, K. Fötisch, F. Altmann, and S. Vieths, "Carrot allergy: double-blinded, placebo-controlled food challenge and identification of allergens," *Journal of Allergy and Clinical Immunology*, vol. 108, no. 2, pp. 301–307, 2001.
- [27] P. R. Shewry and N. G. Halford, "Cereal seed storage proteins: structures, properties and role in grain utilization," *Journal of Experimental Botany*, vol. 53, no. 370, pp. 947–958, 2002.
- [28] K. Adeli and I. Altosaar, "Characterization of oat vicilin-like polypeptides," *Plant Physiology*, vol. 75, no. 1, pp. 225–227, 1984.
- [29] M. M. Bradford, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding," *Analytical Biochemistry*, vol. 72, no. 1–2, pp. 248–254, 1976.
- [30] B. Cooper, D. Eckert, N. L. Andon, J. R. Yates III, and P. A. Haynes, "Investigative proteomics: identification of an unknown plant virus from infected plants using mass spectrometry," *Journal of the American Society for Mass Spectrometry*, vol. 14, no. 7, pp. 736–741, 2003.
- [31] J. K. Eng, A. L. McCormack, and J. R. Yates III, "An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database," *Journal of the American Society for Mass Spectrometry*, vol. 5, no. 11, pp. 976–989, 1994.
- [32] J. R. Yates III, J. K. Eng, A. L. McCormack, and D. Schieltz, "Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database," *Analytical Chemistry*, vol. 67, no. 8, pp. 1426–1436, 1995.
- [33] N. L. Andon, D. Eckert, J. R. Yates III, and P. A. Haynes, "High-throughput functional affinity purification of mannose binding proteins from *Oryza sativa*," *Proteomics*, vol. 3, no. 7, pp. 1270–1278, 2003.
- [34] J. Peng, J. E. Elias, C. C. Thoreen, L. J. Licklider, and S. P. Gygi, "Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome," *Journal of Proteome Research*, vol. 2, no. 1, pp. 43–50, 2003.
- [35] P. A. Haynes, S. Miller, T. Radabaugh, et al., "The wildcat toolbox: a set of perl script utilities for use in peptide mass spectral database searching and proteomics experiments," *Journal of Biomolecular Techniques*, vol. 17, no. 2, pp. 97–102, 2006.
- [36] E. Durr, J. Yu, K. M. Krasinska, et al., "Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture," *Nature Biotechnology*, vol. 22, no. 8, pp. 985–992, 2004.

- [37] J. G. Rohrbough, L. Breci, N. Merchant, S. Miller, and P. A. Haynes, "Verification of single-peptide protein identifications by the application of complementary database search algorithms," *Journal of Biomolecular Techniques*, vol. 17, no. 5, pp. 327–332, 2006.
- [38] A. A. Schäffer, L. Aravind, T. L. Madden, et al., "Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements," *Nucleic Acids Research*, vol. 29, no. 14, pp. 2994–3005, 2001.
- [39] S. F. Altschul, T. L. Madden, A. A. Schäffer, et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.
- [40] R. B. Ferreira, E. Franco, and A. R. Teixeira, "Calcium- and magnesium-dependent aggregation of legume seed storage proteins," *Journal of Agricultural and Food Chemistry*, vol. 47, no. 8, pp. 3009–3015, 1999.
- [41] I. J. Goldstein, "Studies on the combining sites of concanavalin A," *Advances in Experimental Medicine and Biology*, vol. 55, pp. 35–53, 1975.
- [42] A. K. Allen, A. Neuberger, and N. Sharon, "The purification, composition and specificity of wheat-germ agglutinin," *Biochemical Journal*, vol. 131, no. 1, pp. 155–162, 1973.
- [43] I. S. Sheoran, D. J. H. Olson, A. R. S. Ross, and V. K. Sawhney, "Proteome analysis of embryo and endosperm from germinating tomato seeds," *Proteomics*, vol. 5, no. 14, pp. 3752–3764, 2005.
- [44] M. Bardor, G. Cabrera, P. M. Rudd, R. A. Dwek, J. A. Cremata, and P. Lerouge, "Analytical strategies to investigate plant N-glycan profiles in the context of plant-made pharmaceuticals," *Current Opinion in Structural Biology*, vol. 16, no. 5, pp. 576–583, 2006.
- [45] A. Barre, J.-P. Borges, and P. Rougé, "Molecular modelling of the major peanut allergen Ara h 1 and other homotrimeric allergens of the cupin superfamily: a structural basis for their IgE-binding cross-reactivity," *Biochimie*, vol. 87, no. 6, pp. 499–506, 2005.
- [46] R. C. Aalberse, "Assessment of allergen cross-reactivity," *Clinical and Molecular Allergy*, vol. 5, article 2, pp. 1–6, 2007.
- [47] M. Hiemori, N. Bando, T. Ogawa, et al., "Occurrence of IgE antibody-recognizing N-linked glycan moiety of a soybean allergen, Gly m Bd 28 K," *International Archives of Allergy and Immunology*, vol. 122, no. 4, pp. 238–245, 2000.
- [48] P. Rabjohn, E. M. Helm, J. S. Stanley, et al., "Molecular cloning and epitope analysis of the peanut allergen Ara h 3," *The Journal of Clinical Investigation*, vol. 103, no. 4, pp. 535–542, 1999.
- [49] M. P. de Leon, A. C. Drew, I. N. Glaspole, C. Suphioglu, J. M. Rolland, and R. E. O'Hehir, "Functional analysis of cross-reactive immunoglobulin E antibodies: peanut-specific immunoglobulin E sensitizes basophils to tree nut allergens," *Clinical and Experimental Allergy*, vol. 35, no. 8, pp. 1056–1064, 2005.