Plant Physiol. (1996) 111: 49–60

Identification of a New Pea Gene, *PsNlec1*, Encoding a Lectin-like Glycoprotein Isolated from the Symbiosomes of Root Nodules[1]

Igor V. Kardailsky[2], D. Janine Sherrier, and Nicholas J. Brewin*

Department of Genetics, John Innes Centre, Norwich NR4 7UH, United Kingdom

A 27-kD glycoprotein antigen recognized by monoclonal antibody MAC266 was purified from isolated symbiosomes derived from pea (*Pisum sativum*) root nodules containing *Rhizobium*. The N-terminal amino acid sequence was obtained, and the corresponding cDNA clone was isolated by a polymerase chain reaction-based strategy. The clone contained a single open reading frame, and the gene was termed *PsNlec1*. Phylogenetic analysis of 31 legume sequences showed that the PsNlec1 protein is related to the legume lectin family but belongs to a subgroup that is very different from pea seed lectin. Expression of the *PsNlec1* transcript was much stronger in nodules than in other parts of the plant. It was found in both infected and uninfected cells in the central tissue of the nodule and in the stele of the root near the attachment point of the nodule. When uninfected pea seedlings were grown on medium containing nitrate, weak transcription of *PsNlec1* was observed in the root system. The identification of PsNlec1 inside the symbiosome is consistent with the observation that legume lectins are generally vacuolar proteins that may serve as transient storage components.

Pea (*Pisum sativum*) root nodules are symbiotic nitrogen-fixing organs that develop as a result of invasion of the plant root by *Rhizobium leguminosarum* bv *viciae* (Brewin, 1991; Franssen et al., 1992). Inside the infected cells of the nodule, differentiated bacteria inhabit a vesicular compartment, which in many respects represents a newly formed organelle and is referred to as a symbiosome (Roth et al., 1988). As a way to analyze the plant-bacterial surface interaction, we have generated and characterized several monoclonal antibodies raised against components of the peribacteroid material (Brewin et al., 1985). Previously described antibody MAC266 recognizes carbohydrate epitopes associated with a range of glycoproteins inside the symbiosome compartment (Perotto et al., 1991). Although these antigens were not exclusive to symbiosome locations, they were much more abundant there than in microsomal fractions derived from nodule or root homogenates (Perotto et al., 1991). Here we describe the purification of one of these soluble glycoproteins with an apparent molecular mass of 27 kD. Having determined its N-terminal sequence, the corresponding cDNA clone was obtained using PCR-based approaches. The predicted protein encoded by the cDNA has significant sequence similarity to legume lectins, and the gene was named *PsNlec1* because it is strongly expressed in nodules.

Seed lectins generally accumulate in vacuoles, and seed vacuolar proteins have served as a model for the study of protein targeting in plants (Vitale and Chrispeels, 1992; Nakamura and Matsuoka, 1993). Similarly, the vegetative (i.e. the nonseed) lectins also have a vacuolar localization (Law and Tonder, 1992; Yoshida et al., 1994), although an extracellular localization has also been reported for a Gal-binding lectin from peanut nodules (VandenBosch et al., 1994). Several functions have been proposed for plant lectins, but a comprehensive explanation for their presence is still lacking (Sharon, 1993; Rini, 1995). The most favored role for seed lectins is plant defense and protein storage (Chrispeels and Raikhel, 1991). Carbohydrate binding has been proposed to be an anti-nutritional property to stop animals from foraging on seeds (Mirkov et al., 1994; Schroeder et al., 1995). There is also some evidence suggesting the involvement of lectins in early recognition events in legume-*Rhizobium* interactions (Smit et al., 1992).

Growth and expansion of plant cells is often accompanied by the development of a vacuole, but during nodule development, this process does not occur to any great extent in infected cells. The vacuole of the infected cell is smaller than in uninfected cells and is sometimes completely absent; instead, infected cells are packed with symbiosomes that proliferate and grow as the cell matures. The peribacteroid fluid contains enzyme activities that are normally found in vacuoles (Roth and Stacey, 1989). In view of these parallels between the vacuole and the symbiosome, it has been suggested that bacteria inhabit a lytic vesicle in the plant or that the symbiosome at least evolved from such a compartment of the endomembrane system (Mellor, 1989; Brewin, 1991). Our finding that one of the components of the symbiosome is encoded by a lectin-like gene further elaborates this hypothesis.

[1] This research was supported by the John Innes Foundation (I.V.K.), the European Community Copernicus Programme (I.V.K.), and the UK Biotechnology and Biological Research Council (D.J.S.).

[2] Present address: The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037.

* Corresponding author; e-mail brewin@bbsrc.ac.uk; fax 44-1603-456844.

Abbreviations: Blec, bud lectin; Nlec, nodule lectin; ORF, open reading frame; PBm, peribacteroid material; PSL, pea seed lectin; RACE, rapid amplification of cDNA ends; RT, reverse transcriptase.

MATERIALS AND METHODS

Microorganisms

Recombinant plasmids were propagated in *Escherichia coli* JM101. For inoculation of peas, *Rhizobium leguminosarum* bv *viciae* wild-type strain 3841 was used (Wood et al., 1989). Strain B661 is its lipopolysaccharide-defective derivative produced by Tn*5* transposon mutagenesis (Kannenberg et al., 1992). This strain gives rise to small, nonfixing nodules that senesce prematurely (Perotto et al., 1994).

Plant Material

The pea (*Pisum sativum*) var Wisconsin Perfection was used for most experiments described here, unless indicated otherwise. The Sprint2 line was kindly provided by Dr. I. Tikhonovich (All Russia Research Institute for Agricultural Microbiology, St. Petersburg, Russia) (Borisov et al., 1992), and the Fix$^-$ Sym13 pea mutant (Kneen et al., 1990), a derivative of Sparkle, was obtained from Dr. T. LaRue (Boyce Thompson Institute, Ithaca, NY). Peas were grown in gnotobiotic conditions in conical flasks on agar medium containing mineral salts but lacking a nitrogen source (Bradley et al., 1988). Five- to 7-d-old seedlings were inoculated after transfer to agar medium, and the age of nodules was counted from the moment of inoculation. Mature pink nodules developed in this system after 3 to 4 weeks, and occurred only on lateral roots: they were normally harvested 4 weeks after inoculation. Artificially induced nodule senescence (Paau and Cowles, 1979) was achieved by keeping 4-week-old plants in the dark for 36 h prior to harvesting. Induction of the *PsNlec1* gene in uninfected pea roots was studied in plants grown for 3 weeks under hydroponic conditions using Fahraeus medium (Borthakur et al., 1986) supplemented with 20 mM KNO_3.

Plant RNA Purification

Total RNA was prepared from plant tissues using the hot-phenol extraction procedure (de Vries et al., 1982). Poly(A)$^+$ RNA was prepared from the total RNA using paramagnetic oligo(dT) beads (Dynal, Merseyside, UK) following the manufacturer's recommendations. Typically, 100 μg of total RNA (as measured by the A_{260}) was used with a 100-μL suspension of beads to produce an estimated 1 to 5 μg of mRNA.

Preparation of Nodule Extracts

Fractionation of nodule homogenates on Suc step gradients was as previously described (Brewin et al., 1985). To release PBm from the pellet of isolated symbiosomes, this was vigorously resuspended in buffer lacking Suc and containing a cocktail of proteinase inhibitors (Boehringer Mannheim): 0.1 μg/mL (4-amidinophenyl)-methanesulfonylfluoride, 1.0 μg/mL bestatin, 10 μg/mL *N*-[*N*-(L-3-*trans*-carboxyoxiran-2-carbonyl)-L-leucyl]-agmatine, 1.0 μg/mL leupeptin, 2.0 μg/mL pepstatin, and 5.0 μg/mL phosphoramidon. Typically, nodules collected from 50 plants yielded 0.5 to 1.0 mL of PBm containing 0.5 mg/mL total protein.

Purification of MAC266 Antigens

For immunoaffinity purification, MAC266 rat monoclonal antibody was covalently coupled with dimethylpimelimidate to protein-G Sepharose (Sigma) following standard protocols (Harlow and Lane, 1988). To produce the matrix, 2 mL of protein-G Sepharose were incubated with 20 mL of the hybridoma tissue culture supernatant containing the immunoglobulins. The amount of PBm corresponding to 1 mg of total protein was incubated overnight at 4°C with 2 mL of the resin in 10 mL of TBS, 10 mM DTT, 0.05% Tween 20, 10 μg/mL leupeptin, and 100 μg/mL *N*-[*N*-(L-3-*trans*-carboxyoxiran-2-carbonyl)-L-leucyl]-agmatine and then washed with TBS and packed on a minicolumn. After the final wash with 10 mM sodium phosphate buffer, pH 6.8, the antigens were eluted with 0.1 M Gly solution, pH 2.5; eluted fractions were immediately neutralized with an equal volume of Tris-HCl, pH 8.8, containing proteinase inhibitors and concentrated on a Centricon 10 column (Pharmacia). The purified and concentrated MAC266 antigens were then separated by PAGE and blotted on a PVDF membrane, and individual bands were subjected to microsequencing. Fast protein liquid chromatography purification of the MAC266 antigens was accomplished on a MonoQ column (Pharmacia) by NaCl gradient elution using PBm as starting material. Eluted fractions were analyzed for the presence of the antigen in the dot assay; positive fractions were pooled; and individual glycoproteins were separated on the preparative PAGE, transferred to the PVDF membrane from the gel, and subjected to N-terminal microsequencing.

Recombinant DNA Manipulations

For the synthesis of the first strand of cDNA for PCR amplifications, 2 to 3 pmol of primer [oligo(dT)$_{15-18}$ for RT-PCR or RT oligonucleotide CTCGAGGATCCGCGGC-CGC(dT)$_{16}$ for RACE] were annealed to 1 to 2 μg of poly(A)$^+$ RNA from 4-week old-nodules in 15 μL of 60 mM KCl, 60 mM Tris-HCl, pH 8.3, by cooling from 90 to 45°C over 30 min. Ten microliters of 20 mM $MgCl_2$, 20 mM DTT, 60 mM KCl, 60 mM Tris-HCl, pH 8.3, 2 mM dNTP, and 5 units of avian myeloblastosis virus RT (Pharmacia) were added to the annealed mixture. The reaction proceeded for 30 min at 45°C.

For 3′ RACE, the first strand of cDNA was used as template for PCR with one of the gene-specific primers listed in Table I, and the oligonucleotide GCTCGAGGATC-CGCGGC (amplification primer) was used with *Xho*I/*Bam*HI/*Sac*II/*Not*I sites. cDNA was used at final dilution (1:1000) without further purification. Amplification reactions contained 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, 125 μM dNTP, 5 to 10 pmol of oligonucleotides in 50 μL of total volume. Amplifications were given a "hot start" and were performed for 30 to 35 cycles with *Taq* polymerase (Perkin-Elmer) at 94°C for 15 s, 50°C for 1 min, 72°C for 2 min, and a final extension for 10

Table I. *Design of oligonucleotides for amplification of the PsNlec1 cDNA from the N-terminal protein sequence*

Origin[a]	Sequence[b]
Fast protein liquid chromatography	T T A N S G V V L Q G D X S I V I F
Immunoaffinity	T T A N S G A T Y? Q G D A Q R
consensus	t T A N S G v t ? q G D a q I a v v s
	T T A N S G v a
oligo1	*Sal* I cgtcgACIACIGCIAATTCIGGIG CAG
oligo2	*Sal* I Q G D A Q I cgtcgaCAAGGIGATGCTCAAAT G C A G
Actual as in *PsNlec1*	T T A N S G V T F Q G D A Q I

[a] Twenty-seven-kilodalton MAC266-recognized glycoprotein was purified by either fast protein liquid chromatography or immunoaffinity chromatography. [b] Alternative residues/bases are shown under sequences. "I" in the oligonucleotide sequence stands for inosine.

min at 72°C. 3′ RACE products were digested with *Sal*I and *Bam*HI and subcloned and sequenced in pBluescript SK(+).

For primer extension, the antisense oligonucleotide TGGAGGAAATGGTGAAC (oligo3 in Figure 1) was ^{32}P-labeled using T4 DNA kinase, and reverse transcription was performed as described. Extension products were analyzed on a sequencing gel. The 5′ RACE protocol (Frohman et al., 1988) was used with the following modifications: reverse transcription was performed as in the primer-extension experiment, but with unlabeled oligo3. The reaction mixture was diluted into 100 μL of 0.3 N NaOH, 10 mM EDTA, incubated at 65°C for 1 h to hydrolyze RNA, and neutralized with 2 μL of 9 M HCl and 10 μL of 1 M Tris-HCl, pH 8.0. After phenol/chloroform extraction, excess primer was removed by gel-filtration chromatography on a S-400 spin minicolumn (Pharmacia) and eluted cDNA was precipitated with ethanol. For (dA)-tailing, the cDNA was incubated for 20 min at 37°C in 20 μL with 0.1 M dATP and 6 units of terminal deoxyribonucleotide transferase (Pharmacia) according to the manufacturer's recommendations. The tailing reaction was stopped by adding EDTA to 25 mM and heating to 65°C for 5 min. Oligo(dA)-tailed cDNA was purified on paramagnetic oligo(dT) beads in essentially the same way as poly(A)$^+$ RNA. For the second-strand synthesis, the reaction mixture contained 5 pmol of the RT oligonucleotide, dNTP to 12 μM final concentration, buffer containing 20 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 100 mM KCl, 0.15 mM β-NAD, 50 μg/mL BSA, 10 units of Klenow fragment of DNA polI, and tailed cDNA in 20 μL. It was incubated at 37°C for 1 h, and a 1:50 dilution of the second-strand synthesis mixture was used as template for PCR with the second gene-specific antisense oligonucleotide GTGACGCCAGAATTAGC (oligo4 in Fig. 1) and the 3′ RACE amplification primer.

Other recombinant DNA manipulations followed standard protocols described elsewhere (Maniatis et al., 1982; Sambrook et al., 1989). Dideoxynucleotide sequencing was performed manually with the Sequenase 2.0 kit (United States Biochemical) or on the Applied Biosystems automated sequencing machine with the PRIZM kit.

Phylogenetic Analysis of Legume Lectins

Legume lectin preproprotein sequences were aligned using the Pileup program of the Genetics Computer Group (Madison, WI) suite. The alignment was refined manually to maximize the match around the residues that form the carbohydrate-binding domain (Bourne et al., 1992; Sharon, 1993; Rini, 1995). Using the Phylip3.5 package (Felsenstein, 1993), the distance matrix was calculated based on classification of amino acids into chemical categories, and the phylogeny was estimated from distance matrix data under the "additive tree model" as implemented in the Fitch program. Topology of the tree was confirmed by "bootstrapping," and the outgroup root was arbitrarily placed at camel foot tree lectin, based on the accepted ancestral phylogenetic position of this legume (Polhill et al., 1981). The following lectin or lectin-like sequences were included (if no reference is given, see Sharon, 1990; Rini, 1995): Phasv, lectin precursor, chromosomal gene, *Phaseolus vulgaris*; Phasva, arcelin-1 seed protein precursor, *Phaseolus vulgaris*; Phaslun, lectin precursor, *Phaseolus lunatus*; ArachiG, agglutinin, *Arachis hypogaea*; Blec, vegetative lectin, *Pisum sativum* (Pak et al., 1992); Nlec, nodule lectin, *Pisum sativum* (this study); Lotust, lectin, *Lotus tetragonolobus*; CytisHO, lectin, anti-H(O), *Cytisus sessilifolius* (Konami et al., 1992a); UlexII, lectin II, *Ulex europaeus* (Konami et al., 1991a); Laburn, lectin, *Laburnum alpinum* (Konami et al., 1991b); Dlab, Man/Glc-specific lectin, *Dolichos lab lab* (Gowda et al., 1994); CytisII, 2-acetamido-2-deoxy-D-Gal-binding lectin II, *Cytisus scoparius* (Konami et al., 1992b); Robp, bark lectin, *Robinia pseudoacacia* (Yoshida et al., 1994); UlexI, lectin I, *Ulex europaeus* (Konami et al., 1991a); Lathoch, LoLI isolectin, *Lathyrus ochrus*; PSL, lectin precursor, *Pisum sativum*; Lens, lectin, *Lens culinaris*; Medict, lectin LEC1, *Medicago truncatula*; Lathsp, lectin, *Lathyrus sphaericus* (Richardson et al., 1987); Erythcor, lectin precursor, *Erythrina corallodendron*; ErythvarG, Gal-specific isolectin, *Erythrina variegata* (Yamaguchi et al., 1993); PhasvE, phytohemagglutinin chain E precursor, *Phaseolus vulgaris*; PhasvL, phytohemagglutinin chain L precursor, *Phaseolus*

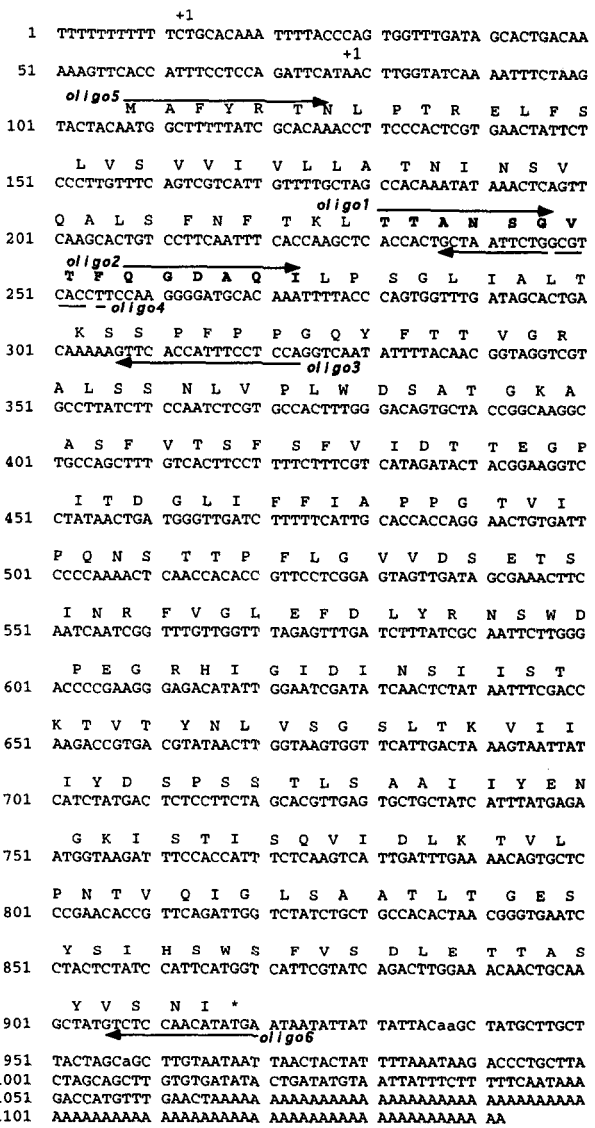

Figure 1. Sequence of the complete *PsNlec1* cDNA and of the ORF in single-letter code. The region corresponding to the N terminus of the 27-kD glycoprotein is shown in boldface in the protein translation. Arrows show the location and direction of oligonucleotides used to amplify the cDNA: *oligo1* and *oligo2* were derived from the N-terminal protein sequence (Table I) and were used in the 3′ RACE; *oligo3* and *oligo4* were based on the sequences of 3′ RACE products and used for primer extension and 5′ RACE. Two transcription start sites mapped in the primer extension (Fig. 2) are indicated as +1. *Oligo5* and *oligo6* amplified the fragment corresponding to the complete ORF found in the cDNA and spanning the junction of the 3′ and 5′ RACE products. The stretch of 11 Ts at the beginning of the sequence was artificially added as a result of 5′ RACE.

vulgaris; Db58, lectin DB58 precursor, *Dolichos biflorus*; Dbseed, lectin precursor, *Dolichos biflorus*; Glycin, lectin precursor, *Glycine max*; Onobry, lectin, *Onobrychis viciifolia* (Kouchalakos et al., 1984); ConA, concanavalin A precursor, *Canavalia ensiformis*; Bowrin, lectin BMA, *Bowringia mildbraedii* (Chawla et al., 1993); Diocle, lectin, *Dioclea lehmannii* (Perez et al., 1991); and Bauhi, lectin, *Bauhinia purpurea* (Kusui et al., 1991).

Northern Blotting and Quantification of the Gene Expression

The level of expression of *PsNlec1* in different tissues of the pea plant was assessed using northern-blot hybridization analysis with RNA preparations from young uninfected roots, mature 21-d-old nodules, apical buds, stem internodes, and leaves. Because of the high sequence similarity between *PsNlec1* and *Blec1*, some cross-hybridization was expected between the two mRNAs and their corresponding probes. To establish the identity of hybridizing bands, the same filter containing blotted RNA was probed with both genes successively and also with the ubiquitin probe to provide an internal control for loading. Approximately equal amounts of either total or poly(A)$^+$ RNA were loaded on the agarose-formaldehyde gels and transferred to nylon filters (Hybond N, Amersham) according to the manufacturer's instructions. Hybridization and reprobing with ^{32}P-labeled DNA insert from the clone containing the 3′ RACE *PsNlec1* DNA, pea bud lectin cDNA fragment (courtesy of M. Dobres, Drexel University, Philadelphia, PA), PSL (courtesy of R. Casey, John Innes Centre, Norwich, UK), or pea ubiquitin cDNA fragment (courtesy of F. Watts, University of Sussex, Brighton, UK) was performed in standard conditions, with [Na^+] = 0.2 M, 40% formamide at 42°C. The final wash was in [Na^+] = 0.016 M at 60°C. Filters were exposed with the PhosphoImager plates (Fujix, Stamford, CT). Image scanning, handling, and quantification of radioactivity stored in individual bands was done using the software provided with the scanner (BAS1000, Fujix). The level of the *PsNlec1* gene expression was measured as the ratio of the *PsNlec1* to ubiquitin signal for each individual band. Data from different filters were brought to the same relative scale using reference samples common to several filters. Normalization to the same scale assumed linear regression, and error was calculated as sample SD from regression for reference samples.

Immunogold Labeling

Nodules were fixed and embedded in LR White resin as described previously (VandenBosch et al., 1989). Sections 90 nm thick were collected on gold mesh grids for immunolabeling with monoclonal antibodies, and the secondary antibody was a conjugate with 10-nm gold particles (Amersham). After counterstaining of sections for 5 min in uranyl acetate and for 30 s in alkaline Pb citrate, grids were coated with a fine layer of carbon to strengthen the sections. Finally, sections were observed and photographed on a JEOL JEM-1200EA transmission electron microscope at 80 kV.

In Situ Hybridization

Median longitudinal sections of pea nodules were used that showed evidence of a persistent meristem at the tip. The leghemoglobin gene (*PsLb*) was used as a reference marker to identify the late zone II and the II-III interzone (Kardailsky et al., 1993) according to the accepted nomenclature for the indeterminate nodule zonation (Vasse et al., 1990). Developmental expression of many nodulin genes

has already been mapped along this axis (Nap and Bisseling, 1990; Franssen et al., 1992; Kardailsky et al., 1993). Digoxigenin-labeled riboprobes were synthesized using linearized plasmid that contained the 900-bp 3′ RACE product of *PsNlec1* subcloned in sense and antisense orientation under the T7 promoter of the pBluescript SK(+). The pea leghemoglobin clone was kindly provided by T. Bisseling (Agricultural University, Wageningen, The Netherlands). Labeling was done essentially as recommended by the supplier (Boehringer) with minor modifications. Nodules were fixed in 4% formaldehyde, dehydrated, embedded in wax, and sectioned following standard protocols (De Block and Debrouwer, 1993; Wilson et al., 1994). Slides with sections were hybridized with riboprobes and signal visualized as alkaline phosphatase activity. Sections were counterstained with Auramine O and photographed on a Zeiss Axiophot microscope.

RESULTS

Purification of MAC266 Antigens

The antigens recognized by monoclonal antibody MAC266 correspond to plant glycoproteins found in the symbiosome lumen (Perotto et al., 1991). Because antigen-antibody binding is sensitive to periodate oxidation, the relevant epitope is probably present on the carbohydrate decoration of these proteins. This limited the strategy for cloning the corresponding genes to purifying the antigens and determining the N-terminal protein sequences. One of these glycoproteins, with an apparent molecular mass of 27 kD, was purified using fast protein liquid chromatography and affinity chromatography, and yielded sufficient N-terminal sequence information to allow the design of degenerate oligonucleotides. Two versions of its N-terminal sequence are presented in Table I. Because of the low amounts of the material, some residues were apparently defined erroneously, leading to ambiguous sequence information. However, from two independent purifications of the same glycoprotein band, sufficient sequence information was obtained to design oligonucleotides for use as primers in a PCR-based cDNA cloning strategy (see Table I).

Amplification of the cDNA Encoding the 27-kD Glycoprotein

The sequence of the full-size cDNA encoding the 27-kD glycoprotein is presented in Figure 1. This was obtained in a series of PCR amplifications as described below. Initially, the 3′ RACE procedure (Frohman et al., 1988) was performed on nodule cDNA using oligonucleotides derived from the N-terminal sequence of the protein. To increase the specificity of amplification, it was carried out in two rounds: with oligo1 (Table I) in the first round using total nodule cDNA as template, then with oligo2 using a fraction of the product of the first round. This produced a 900-bp DNA fragment, which was subcloned, and the sequences of two independent subclones were determined. As expected, this sequence contained an ORF in phase with the one defined by oligonucleotides used for amplification, and ended with the poly(A)$^+$ tail. To obtain the part of the cDNA "upstream" from the region corresponding to oligo1 and oligo2, two antisense oligonucleotides (oligo3 and oligo4) were made, based on the sequence of the 3′ RACE product (Fig. 1). Primer extension with oligo3 is illustrated in Figure 2; it revealed two alternative cap sites in the mRNA with most of the messages starting at 246 bases from oligo3 and a minority of longer mRNAs starting at 302 bases. This 5′ region of mRNA was obtained after 5′ RACE amplification, which produced a range of DNA fragments from 200 to 300 bp with a maximum at approximately 230 bp (not shown). These products were cloned without size selection, and seven randomly selected subclones were sequenced. Their size distribution and DNA sequences confirmed that they corresponded to the 5′ region of mRNA. An ORF in the sequence of the 5′ region began with ATG in the appropriate consensus (Grunert and Jackson, 1994) and was continuous with the ORF in the 0.9-kb DNA produced by the 3′ RACE. Successful reverse transcription/amplification with oligo5 (which started at this ATG codon) and with the antisense oligo6 (which started at the putative translation termination codon) finally proved that both 3′ and 5′ RACE products indeed represented a continuous mRNA that was present in nodules. Sequences were determined for five independent subclones from this last amplification, and the consensus was found among the DNA sequences produced in all three amplifications (3′ RACE, 5′ RACE, and RT-PCR). The final sequence of the cDNA was assembled and the corresponding gene was named *PsNlec1* (*Pisum sativum* nodule lectin) because, as demonstrated below, it shows strong homology to the legume lectin gene family.

Homology of PsNlec1 to Other Lectins

The complete sequence of *PsNlec* cDNA is 1130 bp long (Fig. 2). The only long ORF in the sequence would encode a polypeptide of 270 amino acid residues with a calculated M_r of 30,000. The predicted PsNlec protein sequence

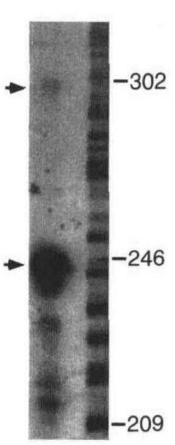

Figure 2. Primer extension for *PsNlec1* on nodule RNA. The labeled antisense oligo3 was used as primer in reverse transcription of nodule RNA. The size of extension products was determined on a sequencing gel. Two positions of the cap sites are marked with arrows. Markers are in bases.

showed significant similarity to the sequence of the legume lectin protein family, which has been very well studied (see reviews by Sharon, 1990, 1993; Rini, 1995). To analyze the relatedness of the PsNlec1 protein and identify its important structural features, it was compared with 31 published legume lectin sequences representing all four carbohydrate-specificity classes. The best match was with the sequence of the pea vegetative bud lectin Blec1 (Pak et al., 1992); the alignment of the two protein sequences is shown in Figure 3. Overall sequence similarity to other legume lectins is less strong, and there are only 37% of identical residues between PsNlec1 and PSL. Phylogenetic analysis of legume lectins produced the inferred phylogenetic tree based on their protein sequences, as shown in Figure 4.

Predicted Amino Acid Sequence for *PsNlec1*

The *Nlec1* ORF included the N-terminal sequence of the 27-kD glycoprotein originally purified from symbiosomes as the MAC266 antigen. This sequence started at Thr^{42}, which therefore corresponds to the beginning of the mature protein. Von Heijne rules for eukaryotic signal peptides (von Heijne, 1986) predict a cleavage site at Leu^{34} or at Thr^{26}. This indicates that PsNlec1 is synthesized as a preproprotein, and its signal peptide and propeptide are removed during posttranslational processing. The mature PsNlec1 that starts at Thr^{42} could have a calculated M_r of 24,400, whereas the glycoprotein migrates as approximately 27 kD in Laemmli gels. This difference may easily be accounted for by glycosylation of the mature protein. A possible N-linked glycosylation site is predicted at Asn^{94}. O-linked glycosylation sites are more difficult to predict from protein sequence, but a recently described system for identification of this type of glycosylation (Hansen et al., 1995) finds five Thr residues at positions 43, 111, 112, 129, and 136 that are likely to be O-glycosylated. (It should be noted, however, that this system was "trained" on mammalian glycoproteins, and the validity of its predictions for plant proteins is uncertain.)

```
Percent Similarity: 77.947   Percent Identity: 65.779

  1 MAFYRTNLPTRELFSLVSVVIVLLATNINSVQALSFNFTKLTTANSGVTF
    |::||    |:||:||||:::| |||||   :||||||.|:|.:|.::|:
  1 MGLYR....TKELLSLVSIMFVSLATNI...EALSFNFPKITPGNTAITL

 51 QGDAQILPSGLIALTKSSPFPPGQYFTTVGRALSSNLVPLWDSATGKAAS
    ||:|.||:.|::|||.| .:||.  |...|||| |. ||||||||..||
 44 QGNAKILANGVLALTNSTQIPPTTTFPSTGRALYSTPVPLWDSATGNVAS

101 FVTSFSFVIDTTEGPI.TDGLIFFIAPPGTVIPQNSTTPFLGVVDSETSI
    ||||||||| ...|.:  ||||:||||||:| ||.|| ..:|||||.|||
 94 FVTSFSFVILNPSGRVPTDGLVFFIAPPDTEIPNNSQSQYLGVVDSKTSI

150 NRFVGLEFDLYRNSWDPEGRHIGIDINSIISTKTVTYNLVSGSLTKVIII
    |||||||||| ||:|| ||||||||:|||||| ||:|||||||.||
144 NRFVGLEFDLYANSFDPYMRHIGIDINSLISTKTVRYNFVSGSLTKVTII

200 YDSPSSTLSAAIIYENGKISTISQVIDLKTVLPNTVQIGLSAATLTGESY
    |||||.||.|.|.||||.|||||| :|||.|||..| :|:||.. .: |.
194 YDSPSNTLTAVITYENGQISTISQNVDLKAVLPKDVSVGFSATSTIAVSH

250 SIHSWSFVSDLE.TTASYVSNI PsNlec1
    .||||||.|:|| ||:..||.:
244 NIHSWSFTSNLEATTGNIVSQV Blec1
```

Figure 3. Comparison of predicted amino acid sequences for PsNlec1 and bud lectin Blec1. Predicted locations of signal peptide cleavage sites are marked with arrows. The observed N-terminal sequence of the 27-kD glycoprotein is underlined; predicted Asn-linked glycosylation sites are in boldface. Similarity scores are calculated using the GAP program of the Genetics Computer Group suite.

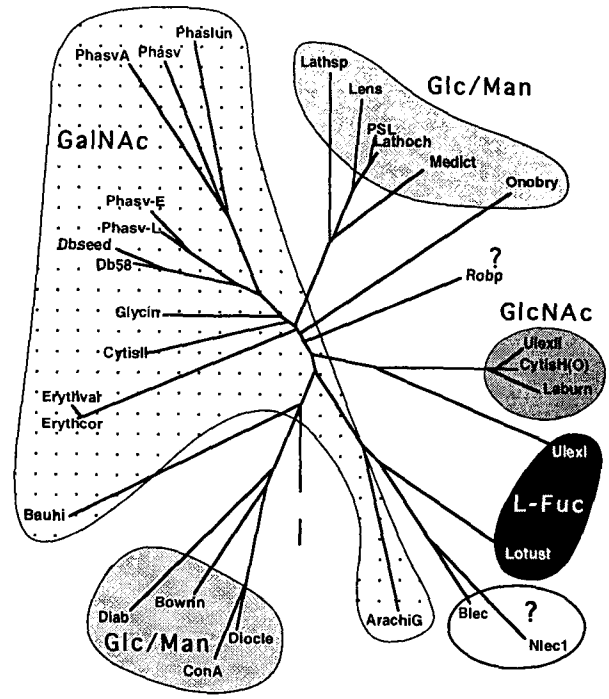

Figure 4. Inferred phylogenetic relationships between legume lectin genes. Lectins with known carbohydrate specificities as defined in competitive inhibition experiments are shaded: GalNAc, Gal/*N*-acetylgalactosamine; ?, specificity unknown. The outgroup root is shown by a dashed line. For abbreviations of lectin names, see "Materials and Methods."

Regulation of *PsNlec1* Gene Expression

According to the cDNA sequence, the predicted size of *PsNlec1* mRNA is 1050 to 1100 bases; the full-size *Blec1* cDNA is 950 bases long without the poly(A) tail (Mandaci and Dobres, 1993). The approximately 1.1-kb-long transcript detected by northern-blot hybridization with the *PsNlec1* probe was found only in nodule tissue, where it was expressed at a fairly high level. The blot shown in Figure 5A is overexposed to visualize the faint signal at approximately 950 bases detected with the *PsNlec1* probe in bud RNA. This band is smaller than the one seen in nodules, and this, together with the expected difference in the size of mRNAs, suggests that this signal was due to cross-hybridization with the *Blec1* mRNA. Reprobing of the same filter with *Blec1* probe (Fig. 5B) supported this conclusion. Use of the *PSL* probe showed that expression of this gene was largely confined to the seed. We could not detect any cross-hybridization between the *PSL* probe and nodule or bud RNA, confirming previous observations (Buffard et al., 1988) that expression of *PSL* was 4000 times less in the root than in the seed. Similarly, the *PsNlec1* probe gave no signal in seed RNA (results not shown).

The apparent nodule-specific expression of *PsNlec1* suggested that it should be classified as a nodulin gene. How-

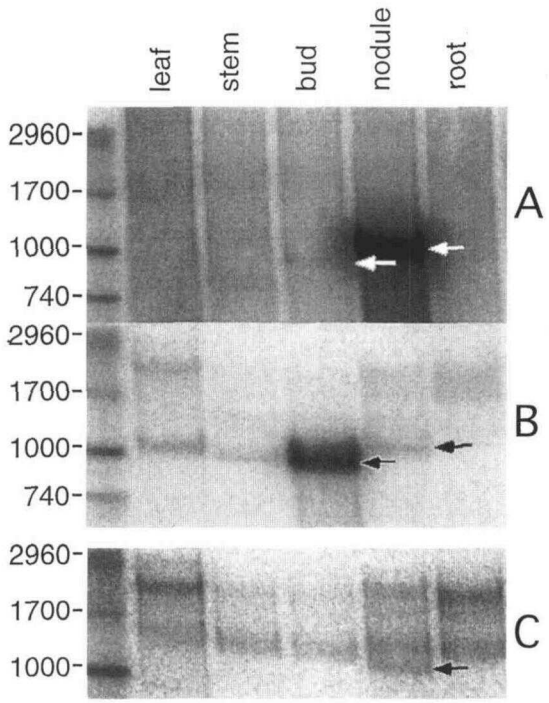

Figure 5. Nodule-specific expression of the *PsNlec1* gene. A northern blot with pea RNA was sequentially probed with *PsNlec1* (A), *Blec1* (B), and ubiquitin (C) probes. Arrows indicate mRNAs of the two lectin genes. The *PsNlec1* signal on C is carried over from the previous hybridization of the filter. Sizes of markers are in bases.

ever, in the in situ hybridization experiments described below, the *PsNlec1* signal was found in the stele of the root not only at the point of nodule attachment, but even at some distance from the nodule. A weak *PsNlec1* signal was also found in sterile pea roots grown on Fahraeus agar supplemented with KNO_3 (Fig. 6). This signal was detectable both on the northern blots (Fig. 6A) and, with more confidence, by PCR (Fig. 6B). On careful examination, it is not uncommon for nodulin genes or their close homologs to be found expressed in locations other than root nodules (Scheres et al., 1990; Brewin, 1991; Franssen et al., 1992).

To investigate regulation of the *PsNlec1* gene in the nodule, we quantified its expression at different stages of nodule development and in senescing or nonfixing nodules. Quantification data were obtained from nine northern blots, with the ubiquitin signal used as an internal control for RNA loading. The two smallest bands in the pea ubiquitin mRNA spectrum (approximately 1200 and 1300 bases) most accurately reflected the quantity of loaded RNA. The signal in these bands correlated with the quantity of total RNA loaded in a lane with an error of about 10% for nodules of different age and for uninfected roots. A gradual increase in the relative intensity of the *PsNlec1* hybridization signal occurred as nodules developed, but from 4 weeks to 2 months it remained constant (Fig. 6C). In senescing or nonfixing nodules the signal was still present, although at a lower level. The Sprint2 variety showed a considerably higher level of *PsNlec1* gene expression compared to Wisconsin Perfection. Finally, expression of the gene in nitrate-grown roots was less than 5% of that in the wild-type nodule.

Immunogold Localization of the MAC266 Antigens in Pea Nodule Tissue

In view of the significant difference between the level of *PsNlec1* gene expression in the Wisconsin Perfection and Sprint2 varieties, nodule tissue derived from Sprint2 was used for immunogold localizations with MAC266 antibody. Figure 7 shows that large inclusion bodies in the peribacteroid fluid of the symbiosome compartment were recognized by MAC266. This observation is consistent with the fact that the PsNlec1 glycoprotein was initially isolated from the symbiosome fraction of pea nodule homogenates, although it should be remembered that the MAC266 epitope is present on a range of glycoproteins and not

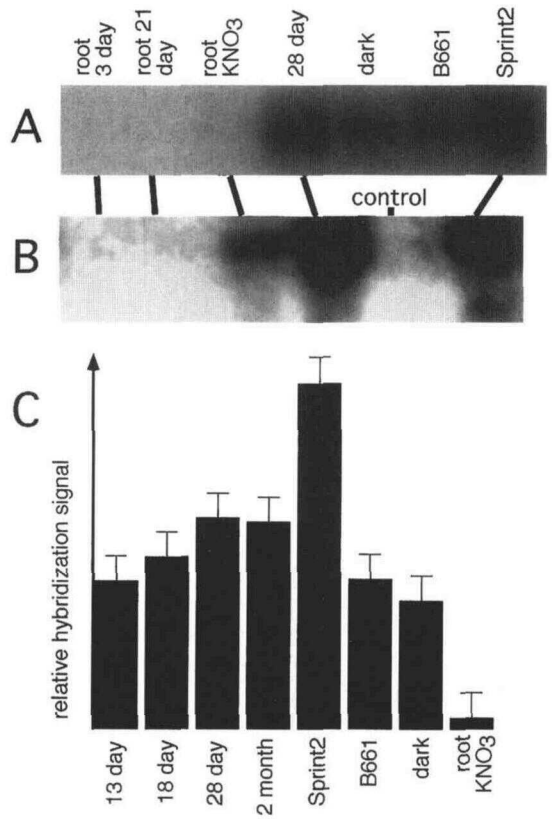

Figure 6. Regulation of *PsNlec1* expression in nodules and roots. Analyzed tissue samples from roots were as follows: 3- and 21-d-old uninfected roots, and 21-d-old roots grown in the presence of nitrate. Tissue samples from nodules were as follows: 13-, 18-, and 28-d-old and 2-month-old nodules of Wisconsin Perfection, 28-d-old nodules after 36 h in the dark, nodules formed by lipopolysaccharide *Rhizobium* mutant B661, and 28-d-old nodules from the pea variety Sprint2. A, Analysis by northern blot, hybridized with the *PsNlec1* probe. B, Analysis by RT-PCR, using oligo5 and oligo6 on cDNA made from indicated RNAs. Following gel electrophoresis, PCR products were transferred to the nylon sheet and hybridized to the *PsNlec1* probe. Control PCR reaction was without added template. C, Accumulated data from nine northern blots showing the relative intensity of the *PsNlec1* hybridization signal normalized to the ubiquitin hybridization signal.

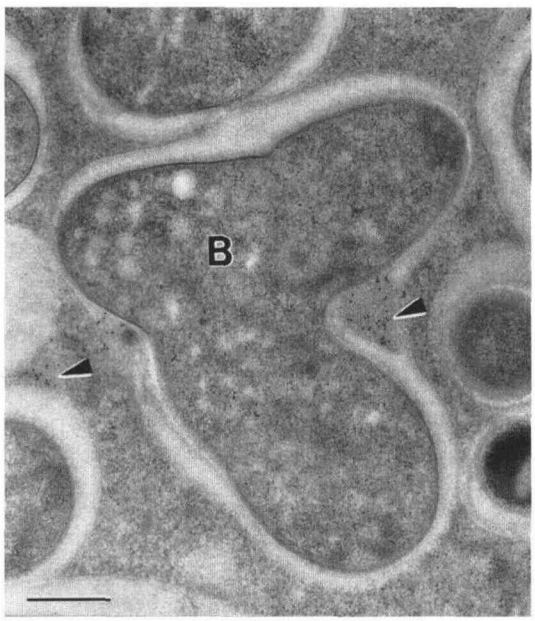

Figure 7. Immunogold localization of MAC266 antigens in an infected cell of a pea nodule. Glycoprotein antigens (including PsNlec1) are visible as inclusion bodies (arrowheads) in the peribacteroid fluid adjacent to the bacteroid (B). Bar = 0.5 μm.

exclusively on the 27-kD protein encoded by *PsNlec1* (Perotto et al., 1991).

In Situ Localization of the *PsNlec1* Gene Expression

The pattern of expression of *PsNlec1* transcription was the same in both pea varieties, Wisconsin Perfection and Sprint2. Figure 8A shows that in the mature nodule *PsNlec1* signal was very abundant in all cells of the symbiotic zone (zone III), including uninfected cells. Although traces of signal could be found in zone II (Fig. 8B), the level of expression increased sharply in the same region where the *PsLb* signal first became detectable. Expression of *PsNlec1* remained high throughout zones III to IV, whereas the *PsLb* signal was at a maximum in the youngest part of zone III and showed a gradual decrease in older tissues (Fig. 8, C and D). Unlike the *PsLb* expression, which was confined to the infected cells, the *PsNlec1* signal was also present in the uninfected cells. This was not obvious in normal nodules, where uninfected cells are few, and their peripheral cytoplasm was usually close to the infected cells that had a very strong signal. However, in nodules formed by the nonfixing lipopolysaccharide-deficient mutant B661 (Fig. 8F), expression of the *PsNlec1* gene in uninfected cells could be clearly seen, because the proportion of infected cells was much smaller. In the nonfixing plant mutant Sym13, the *PsNlec1* signal was still quite strong and showed a similar expression pattern (Fig. 8G), confirming that biological nitrogen fixation is not a prerequisite for expression of *PsNlec1* in nodule cells.

Among the nonsymbiotic tissues, the *PsNlec1* signal was found in the vascular strands of the nodule and in the stele of the root to which the nodule is attached. This root expression was confined to the parenchyma cells between the proto-phloem poles and phloem elements. The root expression was also detectable at some distance from the nodule attachment point (Fig. 8H), but we have not studied how far this expression was distributed along the root system. No hybridization was observed with sections of uninoculated roots grown in the absence of supplementary nitrate.

DISCUSSION

We initiated the study of molecular interactions at the plant-*Rhizobium* interface by fractionation of nodule homogenates, by isolating monoclonal antibodies reacting with peribacteroid components, and by subsequently using these antibodies as molecular probes for further biochemical analysis. We report here that one of the glycoprotein antigens identified inside the symbiosome using monoclonal antibodies is encoded by a previously unrecognized lectin-like gene that is expressed strongly in nodules.

Relationships of PsNlec1 and Other Legume Lectins

Legume lectins are a group of plant lectins with highly conserved sequences (Sharon, 1993; Rini, 1995). To evaluate the position of PsNlec1 among other members of the lectin family, we constructed a phylogenetic tree based on the protein sequences of 31 legume lectins (Fig. 4). For a number of legumes, lectins are encoded by multigene families, as in *Pisum sativum*, *Dolichos biflorus*, *Phaseolus vulgaris*, and *Ulex europaeus*. Despite this, the topology of the tree in Figure 4 is in good agreement with the taxonomic positions of most species: for example, members of the tribes and subtribes Diocleinae (ConA, Dlab, Bowrin, and Diocle), Viciinae (Lathsp, Lens, PSL, Lathoch, and Medic), Phaseolinae (PhasvA, Phasv, and Phaslun), and Erythrininae (Erythvar and Erythcor) cluster together. The Bauhiniidae

Figure 8. (On facing page.) In situ hybridization showing expression of *PsNlec1* in the central tissue of a nodule and in the vascular tissue of the root system of a nodulated plant. A, Longitudinal section of the Sprint2 pea nodule hybridized with the antisense *PsNlec1* riboprobe; scale bar = 20 μm. B, 2.8× magnification of the nodule tip in A. C, Section similar to that shown in A hybridized with *PsLb* leghemoglobin probe. D, 2.8× magnification of C. E, As for A but hybridized with the sense control riboprobe. F, Fragment of the late symbiotic zone of the nodule formed by the *Rhizobium* mutant B661; bar = 5 μm. G, Longitudinal section of the nodule formed by the symbiotically defective plant mutant Sym13; bar = 10 μm. H, Cross-section of a nodulated pea root at 1 mm from the nodule, probed with *PsNlec1*; bar = 5 μm. Note that hybridizations with the *PsNlec1* probe (A and B) were developed for longer periods than those for the Lb probe (C and D) to maximize the very weak signal in root tissue (H). Asterisks and stars show infected and uninfected cells, respectively. Arrowheads mark expression in the stele of the root.

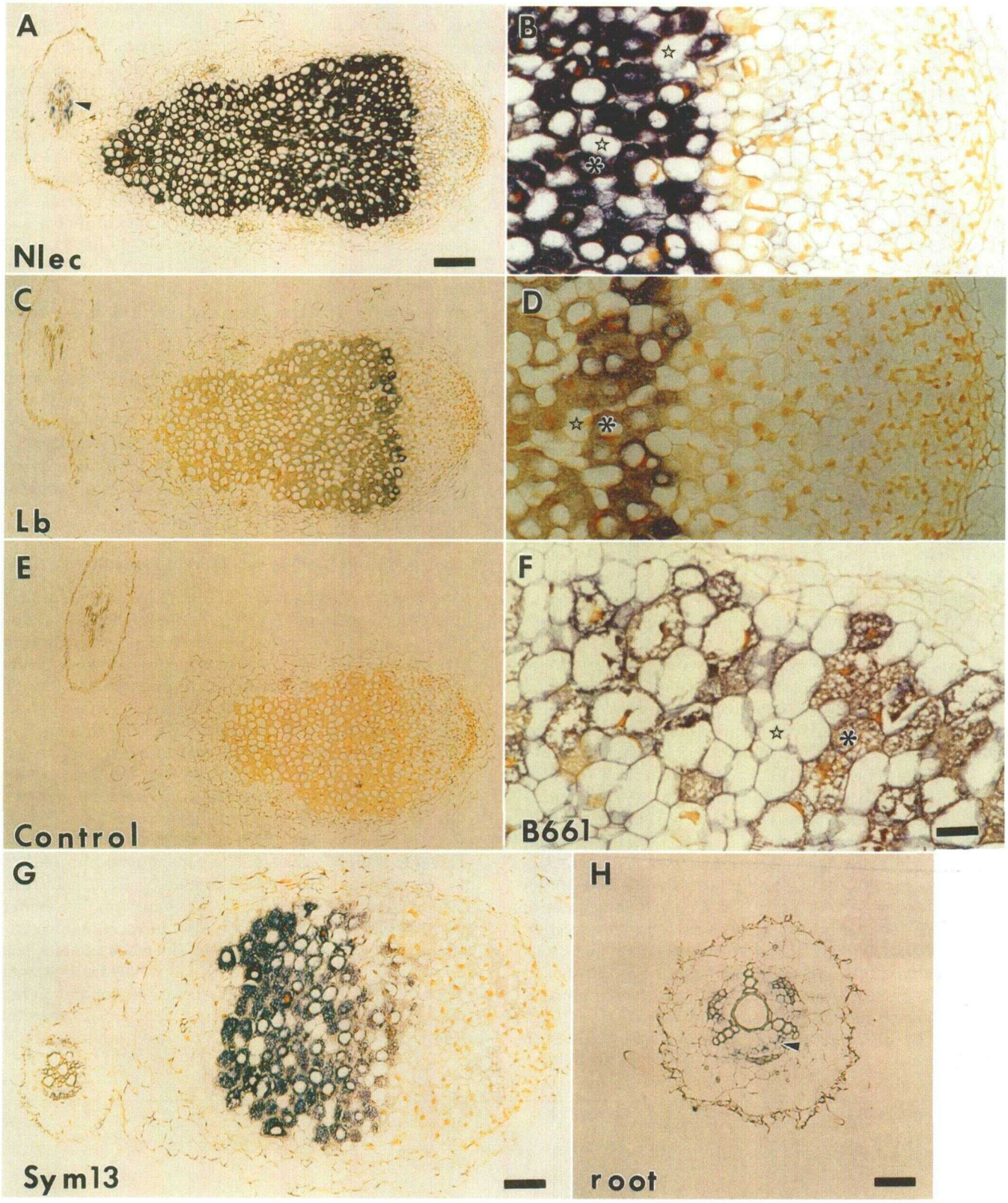

Figure 8. (Legend appears on facing page.)

are assumed to have diverged early in the evolution of the family and display some archaic features (Polhill et al., 1981), and the *Bauhinia purpurea* lectin (Bauhi) branch emerges from the base of the tree in Figure 4 and, consequently, was used to place an outgroup root.

The most notable exceptions to this pattern are PsNlec1 and Blec1. In contrast to PSL, these two vegetative pea lectins (or putative lectins) show very little relatedness to the seed lectins derived from other members of the Viciinae tribe. Thus, genes encoding PsNlec1 and PSL appear to be an example of "paralogous genes," i.e. genes that duplicated before the evolutionary divergence of taxa and in which the two copies may have acquired different functions and evolved independently (Schlegel, 1994). In fact, one would predict that, if lectin-like proteins with similar function to that of PsNlec1 were described in other legume species, their molecular phylogeny would again correlate with the legume taxonomy. In preliminary experiments with diploid *Medicago* species belonging to the *Medicago sativa* complex (*M. sativa* ssp. *quasifalcata* and *M. sativa* ssp. *cocrulea*), we observed significant cross-hybridization between the *PsNlec1* probe and nodule mRNA under conditions in which the *PSL* probe gave no signal. Similarly, when *PsNlec1* probe was used for cross-hybridization with *Eco*RI digests of genomic DNA from *M. sativa* ssp. *quasifalcata* and *M. sativa* ssp. *cocrulea*, a single hybridization signal at the 5.1-kb region was observed (G. Kiss and P. Kaló, personal communication). These results indicate that at least this legume has lectin-related sequences that are closer to *PsNlec1* than to the PSL.

There is a strong correlation between the position of lectins on the tree and their specificities as defined by competitive inhibition of carbohydrate binding. The Glc/Man-specific lectins are split into two compact groups, one with members of the Diocleinae subtribe (which includes concanavalin A) and the other with lectins from the Viciinae tribe, which includes PSL. Gal- and galactosamine-specific lectins form a more heterogeneous group: the *Bauhinia purpurea* lectin is in this group, suggesting that perhaps the "ancestral" lectin was Gal/*N*-acetylgalactosamine specific. Studies of the three-dimensional structures of legume lectins (Bourne et al., 1992; Sharon, 1993; Rini, 1995) revealed a remarkable structural conservation and identified invariant residues that coordinate metal ions and interact with the ligand. In editing our alignment of 31 legume lectin sequences, we took this structural information into account, and also estimated the degree of variability of these ligand-binding-site amino acid residues, as summarized in Table II. Our analysis confirmed the previous observation that, whereas metal-coordinating residues are very conserved, the ones in contact with the ligand are often hypervariable (Young and Oomen, 1992), thus possibly accounting for a wide range of lectin carbohydrate specificities. PsNlec1 and Blec1 protein sequences also conform to this pattern, but the absence of residues G^{99} and G^{216} strongly suggests that PsNlec1 does not have the same sugar-binding site as PSL (where these residues have been shown to be essential). It is still theoretically possible that it could have a different sugar-binding site or that it may correspond to a lectin-like defense protein similar to arcelin or α-amylase inhibitor, which also lack G^{99} (Mirkov et al., 1994).

In summary, it is probable that PsNlec1 has a three-dimensional conformation similar to that of other legume lectins, but, on the basis of sequence comparisons, it is impossible to deduce whether or not the protein has lectin activity and what the carbohydrate specificity might be. The two vegetative pea lectins, PsNlec1 and Blec1, apparently represent a gene family paralogous to PSL. These two groups probably diverged early in the evolution of the family Leguminosae and have presumably acquired a different range of functions.

Localization and Possible Functions of PsNlec1

We have previously reported strong immunogold labeling in Golgi bodies and peribacteroid compartments for the glycoproteins recognized by the MAC266 antibody (Perotto et al., 1991). Several lines of evidence now indicate that the *PsNlec1* gene product is preferentially targeted to the symbiosome compartment: the original MAC266 antigen was isolated from symbiosomes; MAC266 immunogold labeling confirmed a symbiosomal location; and in situ localization of mRNA also showed that the gene is most strongly expressed in infected cells in the nodule. Proteins with hydrophobic N-terminal signal peptides enter the secretory system, and PsNlec1 has such a peptide. The subsequent sorting is believed to depend on the presence of the appropriate "tags" on a protein. *PsNlec1* may thus provide a convenient model for studying symbiosome targeting in infected cells, and one would expect to find some common features between the mechanisms of symbiosomal and vacuolar targeting. However, to investigate the targeting of PsNlec1, it will first be necessary to obtain a monospecific antiserum directed against this protein, because immunolocalization with MAC266 does not distinguish

Table II. *Conservation of the ligand-binding site in legume lectins*

Type	Residue Position/Conservation[a]										
Coordinating divalent cations	D^{81}		E^{119}	D^{121}		N^{125}	D^{129}	H^{136}			
	87%		97%	100%		84%	94%	90%			
Interacting with sugar moiety	D^{81}	G^{99}			F/Y^{123}	N^{125}			G^{216}	A^{217}	E^{218}
	87%	90%			87%	84%			87%	42%	13%
Present in PsNlec/Blec1	+	−	+	+	+	+	+	+	−	−/+	−

[a] Residues known to contribute to ligand binding from crystallographic studies were scored in the alignment of 31 lectin sequences. Numbering is as in the mature PSL sequence.

between PsNlec1 and other glycoproteins recognized by these antibodies (Perotto et al., 1991).

Because expression of *PsNlec1* is strongest in the late symbiotic zone of the nodule, it is unlikely that the protein is involved in the early stages of nodule initiation, as has been hypothesized for other legume lectins (Smit et al., 1992). Furthermore, using antibody MAC266, we could not detect any binding of PsNlec1 to the bacteroid surface in vitro (results not shown). With respect to the presence in the nodule, there is more similarity between PsNlec1 and a nodule form of peanut lectin described recently (Law and Tonder, 1992; VandenBosch et al., 1994). In that study, antibodies raised against peanut seed lectin were used to analyze lectin distribution within the nodule. The signal was most abundant in vacuoles and in extracellular spaces of nodule parenchyma. Remarkably, it was also found in the soluble fraction in the symbiosome lumen, although in lesser quantities than in the parenchyma. From the distribution of labeling in the peanut nodule, the authors speculated that lectins in nodules may serve as a transient storage compound for assimilated nitrogen. Both the size and some aspects of its localization relate this lectin to PsNlec1. It is also noteworthy that in Figure 4 the Gal-specific peanut agglutinin (ArachiG) is one of the nearest neighbors to PsNlec1. It will now be interesting to obtain purified PsNlec1 in sufficient quantities to test for possible lectin activity and to determine its other biochemical characteristics and possible functions in nodule and root development.

Because *PsNlec1* is weakly expressed in roots of nodulated plants and also in uninfected roots grown on nitrate, its expression may reflect the nitrogen regime of the root system on successful nodulation. Thus, PsNlec1 may indeed function as a transient sink for the excess nitrogen produced in the nodules, as proposed for peanut lectin ArachiG (VandenBosch et al., 1994). However, the *PsNlec1* transcript is still present at considerable levels in ineffective or senescing nodules, where no excess of fixed nitrogen is expected. Moreover, transcripts of *PsNlec1* were undetectable in seeds. Thus, regulation of the *PsNlec1* gene is probably quite complex. At the transcriptional level, both the nitrogen status of the root system and developmental signals during nodule formation seem to be regulating gene expression. It will be interesting to know if *PsNlec1* expression is also regulated at the translational or post-translational levels, which might be reflected in the quantities of PsNlec1 protein accumulating in symbiosomes of wild-type and ineffective nodules. Identification of the *PsNlec1* gene paves the way to answering these questions concerning the structure, function, regulation, and targeting of pea nodule lectin.

ACKNOWLEDGMENTS

We thank M. Dobres, R. Casey, T. Bisseling, and F. Watts for supplying clones of Blec1, PSL, PsLb, and pea ubiquitin for use as hybridization probes; G. Kiss for performing the Southern hybridization with *Medicago* DNA; M.-F. LeGal and P. Dahiya for help with protein purification; and D.A. Hopwood for his comments on the manuscript.

Received November 20, 1995; accepted February 16, 1996.
Copyright Clearance Center: 0032–0889/96/111/0049/12.
The accession number for the sequence reported in this article is U31981.

LITERATURE CITED

Borisov AY, Morzhina EV, Kulikova OA, Tchetkova SA, Lebsky VK, Tikhonovich IA (1992). New symbiotic mutants of pea (*Pisum sativum* L.) affecting either nodule initiation or symbiosome development. Symbiosis **14:** 297–313

Borthakur D, Barber CE, Lamb JW, Daniels MJ, Downie JA, Johnston AWB (1986) A mutation that blocks exopolysaccharide synthesis prevents nodulation of peas by *Rhizobium leguminosarum* but not of beans by *Rhizobium phaseoli* and is corrected by cloned DNA from *Rhizobium* or the phytopathogen *Xanthomonas*. Mol Gen Genet **203:** 320–323

Bourne Y, Nesa MP, Rouge P, Mazurier J, Legrand D, Spik G, Montreuil J, Cambillau C (1992) Crystallization and preliminary X-ray diffraction study of *Lathyrus ochrus* isolectin II complexed to the human lactotransferrin N2 fragment. J Mol Biol **227:** 938–941

Bradley DJ, Wood EA, Larkins AP, Galfrè G, Butcher GW, Brewin NJ (1988) Isolation of monoclonal antibodies reacting with peribacteroid membranes and other components of pea root nodules containing *Rhizobium leguminosarum*. Planta **173:** 149–160

Brewin NJ (1991) Development of the legume root nodule. Annu Rev Cell Biol **7:** 191–226

Brewin NJ, Robertson JG, Wood EA, Wells B, Larkins AP, Galfrè G, Butcher GW (1985) Monoclonal antibodies to antigens in the peribacteroid membrane from *Rhizobium*-induced root nodules of pea cross-react with plasma membrane and Golgi bodies. EMBO J **4:** 605–611

Buffard D, Kaminski P, Strosberg D (1988) Lectin gene expression in pea (*Pisum sativum* L) roots. Planta **173:** 367–372

Chawla D, Animashaun T, Hughes R, Harris A, Aitken A (1993) *Bowringia mildbraedii* agglutinin: polypeptide composition, primary structure and homologies with other legume lectins. Biochim Biophys Acta **1202:** 38–46

Chrispeels MJ, Raikhel NV (1991) Lectins, lectin genes, and their role in plant defense. Plant Cell **3:** 1–9

De Block M, Debrouwer D (1993) RNA-RNA *in-situ* hybridization using digoxigenin-labeled probes: the use of high-molecular-weight polyvinyl-alcohol in the alkaline-phosphatase indoxyl-nitroblue tetrazolium reaction. Anal Biochem **215:** 86–89

de Vries SC, Springer J, Wessels JHG (1982) Diversity of abundant mRNA sequences and patterns of protein synthesis in etiolated and greened pea seedlings. Planta **156:** 129–135

Felsenstein J (1993) Phylip (Phylogeny Inference Package). Internet version 3.5. Department of Genetics, University of Washington, Seattle

Franssen HJ, Vijn I, Yang W-C, Bisseling T (1992) Developmental aspects of the *Rhizobium*-legume symbiosis. Plant Mol Biol **19:** 89–107

Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA **85:** 8998–9002

Gowda LR, Savithri HS, Rao DR (1996) The complete primary structure of a unique mannose/glucose-specific lectin from field bean (*Dolichos lab lab*). J Biol Chem **269:** 18789–18793

Grunert S, Jackson RJ (1994) The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. EMBO J **13:** 3618–3630

Hansen JE, Lund O, Engelbrecht J, Bohr H, Nielsen JO, Hansen JES, Brunak S (1995) Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. Biochem J **308:** 801–813

Harlow E, Lane D (1988) Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Kannenberg EL, Rathbun EA, Brewin NJ (1992) Molecular dissection of structure and function in the lipopolysaccharide of *Rhizobium leguminosarum* strain 3841 using monoclonal antibodies and genetic analysis. Mol Microbiol **6:** 2477–2487

Kardailsky IV, Yang W-C, Zalensky AO, Van Kammen A, Bisseling T (1993) The pea late nodulin gene PsNOD6 is homologous to the early nodulin genes PsENOD3/14 and is expressed after the leghaemoglobin genes. Plant Mol Biol **23:** 1029–1037

Kneen BE, LaRue TA, Hirsch AM, Smith CA, Weeden NF (1990) *sym13*—A gene controlling ineffective nodulation in *Pisum sativum*. Plant Physiol **94:** 899–905

Konami Y, Yamamoto K, Osawa T (1991a) The primary structures of two types of the *Ulex europeus* seed lectins. J Biochem **109:** 650–658

Konami Y, Yamamoto K, Osawa T, Irimura T (1992a) Correlation between carbohydrate-binding specificity and amino acid sequence of carbohydrate-binding regions of Cytisus-type anti-H(O) lectins. FEBS Lett **304:** 129–135

Konami Y, Yamamoto K, Osawa T, Irimura T (1992b) The primary structure of the *Cytisus scoparius* seed lectin and a carbohydrate-binding peptide. J Biochem **112:** 366–375

Konami Y, Yamamoto K, Toyoshima S, Osawa T (1991b) The primary structure of the *Laburnum alpinum* seed lectin. FEBS Lett **286:** 33–38

Kouchalakos RN, Bates OJ, Bradshaw RA, Hapner KD (1984) Lectin from sainfoin (*Onobrychis viciifolia* Scop). Complete amino acid sequence. Biochemistry **23:** 1824–1830

Kusui K, Yamamoto K, Konami Y, Osawa T (1991) cDNA cloning and expressionof *Bauhinia purpurea* lectin. J Biochem **109:** 899–903

Law IJ, Tonder HJ (1992) Localization of mannose- and galactose-binding lectins in an effective peanut nodule. Protoplasma **167:** 10–18

Mandaci S, Dobres MS (1993) Sequence of a vegetative homolog of the pea seed lectin gene. Plant Physiol **103:** 663–664

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Mellor RB (1989) Bacteroids in the *Rhizobium*-legume symbiosis inhabit a plant internal lytic compartment: implications for other microbial endosymbioses. J Exp Bot **40:** 831–839

Mirkov TE, Wahlstrom JM, Hagiwara K, Finardi-Filho F, Kjentrup S, Chrispeels MJ (1994) Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-α-amylase inhibitor family of the common bean and its relatives. Plant Mol Biol **26:** 1103–1113

Nakamura K, Matsuoka K (1993) Protein targeting to the vacuole in plant cells. Plant Physiol **101:** 1–5

Nap J-P, Bisseling T (1990) Nodulin function and nodulin gene regulation in root nodule development. *In* PM Gresshoff, ed, Molecular Biology of Symbiotic Nitrogen Fixation. CRC Press, Boca Raton, FL, pp 181–229

Paau AS, Cowles JR (1979) Effect of induced nodule senescence on parameters related to dinitrogen fixation, bacteroid size and nucleic acid content. J Gen Microbiol **111:** 101–107

Pak JH, Hendrickson T, Dobres MS (1992) Predicted sequence and structure of a vegetative lectin in *Pisum sativum*. Plant Mol Biol **18:** 857–863

Perez G, Perez C, Sousa-Cavada B, Moreira R, Richardson M (1991) Comparison of the amino acid sequences of the lectins from seeds of *Dioclea lehmanni* and *Canavalia maritima*. Phytochemistry **30:** 2619–2621

Perotto S, Brewin NJ, Kannenberg EL (1994) Cytological evidence for a host defense response that reduces cell and tissue invasion in pea nodules by lipopolysaccharide-defective mutants of *Rhizobium leguminosarum* strain 3841. Mol Plant-Microbe Interact **7:** 99–112

Perotto S, VandenBosch KA, Butcher GW, Brewin NJ (1991) Molecular composition and development of the plant glycocalyx associated with the peribacteroid membrane of pea root nodules. Development **112:** 763–773

Polhill RM, Raven PH, Stirton CH (1981) Evolution and systematics of the Leguminosae. *In* RM Polhill, PH Raven, eds, Advances in Legume Systematics. Royal Botanic Gardens, Kew, Surrey, UK, pp 1–26

Richardson M, Yarwood A, Rouge P (1987) The amino acid sequence of an atypical single-chain lectin from seeds of *Lathyrus sphaericus* (Retz). FEBS Lett **216:** 145–150

Rini JM (1995) Lectin structure. Annu Rev Biophys Biomol Struct **24:** 551–577

Roth L, Jeon K, Stacey G (1988) Homology in endosymbiotic systems: The term 'symbiosome'. *In* R Palacios, DPS Verma, eds, Molecular Genetics of Plant-Microbe Interactions. APS Press, St. Paul, MN, pp 220–225

Roth L, Stacey G (1989) Bacterium release into host cells of nitrogen-fixing soybean nodules: the symbiosome membrane comes from three sources. Eur J Cell Biol **49:** 13–23

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Scheres B, van de Wiel C, Zalensky AO, Horvath B, Spaink HP, Eck H, Zwartkruis F, Volters A-M, Gloudemans T, Van Kammen A, Bisseling T (1990) The ENOD12 gene product is involved in the infection process during the pea-*Rhizobium* interaction. Cell **60:** 281–294

Schlegel M (1994) Molecular phylogeny of eukaryotes. Trends in Ecology and Evolution **9:** 330–335

Schroeder HE, Gollasch S, Moore A, Tabe LM, Craig S, Hardie DC, Chrispeels MJ, Spencer D, Higgins TJV (1995) Bean α-amylase inhibitor confers resistance to the pea weevil (*Brucus pisorum*) in transgenic peas (*Pisum sativum* L). Plant Physiol **107:** 1233–1239

Sharon N (1990) Legume lectins—a large family of homologous proteins. FASEB J **4:** 3198–3208

Sharon N (1993) Lectin-carbohydrate complexes of plants and animals: an atomic view. Trends Biochem Sci **18:** 221–226

Smit G, Swart S, Lugtenberg BJJ, Kijne JW (1992) Molecular mechanisms of attachment of *Rhizobium* bacteria to plant roots. Mol Microbiol **6:** 2897–2903

VandenBosch KA, Bradley DJ, Knox JP, Perotto S, Butcher GW, Brewin NJ (1989) Common components of the infection thread matrix and the intracellular space identified by immunocytochemical analysis of pea nodules and uninfected roots. EMBO J **8:** 335–342

VandenBosch KA, Rodgers LR, Sherrier DJ, Kishinevsky BD (1994) A peanut nodule lectin in infected cells and in vacuoles and the extracellular matrix of nodule parenchyma. Plant Physiol **104:** 327–337

Vasse J, de Billy F, Camut S, Truchet GL (1990) Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. J Bacteriol **172:** 4295–4306

Vitale A, Chrispeels MJ (1992) Sorting of proteins to the vacuoles of plant cells. BioEssays **14:** 151–159

von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res **14:** 4683–4690

Wilson RC, Long FX, Maruoka EM, Cooper JB (1994) A new proline-rich early nodulin from *Medicago truncatula* is highly expressed in nodule meristematic cells. Plant Cell **6:** 1265–1275

Wood EA, Butcher G, Brewin NJ, Kannenberg EL (1989) Genetic derepression of a developmentally regulated lipopolysaccharide antigen from *Rhizobium leguminosarum* 3841. J Bacteriol **171:** 4549–4555

Yamaguchi O, Kimura M, Araki M, Yamasaki N, Kimura Y, Nakajima S, Takagi S (1993) Chemical structures of two subunits, A-subunit and B-subunit, of galactose-specific isolectins from *Erythrina variegata* seeds. J Biochem **114:** 560–566

Yoshida K, Baba K, Yamamoto N, Tazaki K (1994) Cloning of a lectin cDNA and seasonal changes in levels of the lectin and its messenger-RNA in the inner bark of *Robinia-pseudoacacia*. Plant Mol Biol **25:** 845–853

Young NM, Oomen RP (1992) Analysis of sequence variation among legume lectins–a ring of hypervariable residues forms the perimeter of the carbohydrate-binding site. J Mol Biol **228:** 924–934