Characterization of cDNA for nodulin-75 of soybean: A gene product involved in early stages of root nodule development

(Rhizobium/nodule-like structures/proline-rich proteins/DNA sequence)

Henk J. Franssen, Jan-Peter Nap, Ton Gloudemans, Willem Stiekema*, Hans van Dam, Francine Govers, Jeanine Louwerse*, Albert van Kammen, and Ton Bisseling

Department of Molecular Biology, Agricultural University, De Dreijen 11, 6703 BC Wageningen, The Netherlands

Communicated by Harold J. Evans, February 5, 1987

ABSTRACT Establishment of a nitrogen-fixing root nodule is accompanied by a developmentally regulated expression of nodulin genes, only some of which, the so-called early nodulin genes, are expressed in stages preceding actual nitrogen fixation. We have isolated soybean cDNA clones representing early nodulin genes and have studied clone pENOD2 in detail. The cDNA insert of this clone hybridizes to nodulespecific RNA of 1200 nucleotides in length. The RNA that was hybrid-selected by the cloned ENOD2 DNA was in vitro translated to produce two nodulins with an apparent M_r of 75,000, the N-75 nodulins. These two nodulins differ slightly in charge and one does not contain methionine. The amino acid sequence deduced from the DNA sequence shows that proline accounts for 45% of the 240 residues in these nodulins and the sequence contains at least 20 repeating heptapeptide units. The amino acid composition of none of the (hvdroxy)proline-rich (glyco)proteins described in plants resembles the composition of the N-75 nodulins, especially with respect to the high glutamic acid and the low serine content. This suggests that the N-75 nodulins belong to a hitherto unidentified class of presumably structural proteins. The genes encoding the N-75 nodulins were found to be expressed in nodule-like structures devoid of intracellular bacteria and infection threads, indicating that these nodulins do not function in the infection process but more likely function in nodule morphogenesis.

The formation of nitrogen-fixing nodules on the roots of leguminous plants induced by bacteria of the genera Rhizobium and Bradyrhizobium involves the specific expression of a number of plant genes called nodulin genes (1-3). In a description of nodule development, Vincent (4) distinguishes between three stages in nodule development denoted as "preinfection", "infection and nodule formation", and "nod-ule function". In the preinfection stage, the *Rhizobium* bacteria recognize their host plants and attach to the root hairs, an event that is followed by root hair curling. At the moment, nothing is known about specific plant genes that are involved in this stage. In the next stage, the bacteria enter the roots by infection threads while concomitantly the dedifferentiation of some cortical cells results in the formation of meristems. The infection threads grow toward the meristematic cells; bacteria are released into the cytoplasm of about half of these cells and develop into bacteroids. In the final stage, further differentiation of nodule cells occurs leading up to a nitrogen-fixing nodule. Most studies on the expression of nodulin genes so far have been confined to the final stage of root nodule development. But the steps involved in root nodule formation show that major decisions determining the development of a root nodule are made in the stages preceding the establishment of a nitrogen-fixing nodule. We have

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

shown (5) that nodulin genes are differentially expressed during development and that in pea at least two nodulin genes are transcribed in the second stage of root nodule formation. These genes are referred to as early nodulin genes. Here we report the detailed analysis of a soybean cDNA clone representing an early nodulin gene.

MATERIALS AND METHODS

Growth Conditions for Plants and Bacteria. Soybean plants [*Glycine max* (L) Merr. cv. Williams] were cultured as described for pea plants (6) but at 28°C. At the time of sowing the soybean seeds were inoculated with *Bradyrhizobium japonicum* USDA110 or *Rhizobium fredii* USDA257. Both strains were cultured as described (7).

Isolation of Nodules. Nodules were excised from the roots with a scalpel. For samples prior to 6 days after sowing and inoculation, a 4-cm root segment of the upper part of the main root (where root nodules normally would develop) was harvested. Nodules were frozen in liquid nitrogen and stored at -70° C until use.

Isolation of Nucleic Acids. Total RNA from nodules and roots was isolated as described (2) and $poly(A)^+$ RNA was obtained by oligo(dT)-cellulose chromatography (8). Plasmid DNA was isolated by the alkaline lysis method (8).

Construction of cDNA Library. DNA complementary to $poly(A)^+$ RNA isolated from nodules from 21-day-old plants was synthesized with reverse transcriptase (Anglian Biotechnology, Essex, England) and second strand synthesis was performed under standard conditions (8). The double-stranded cDNA was treated with S1 nuclease and size-fractionated on a sucrose gradient. Double-stranded cDNA with a length of 500 base pairs (bp) or more was tailed with dC and then annealed to *Pst* I-cut oligo(dG)-tailed pBR322 (Boehringer Mannheim) in a 1:1 molar ratio. The hybridized mixture was used to transform *Escherichia coli* RR1 (8). On the average, 5000 transformants were obtained per μg of poly(A)⁺ RNA.

Differential Screening of the cDNA Library. Individual transformants were picked, transferred to 96-well microtiter plates containing LB medium, 15% glycerol, and 12.5 μ g of tetracycline per ml, and grown for 16 hr at 37°C. Two replica filters were made on GeneScreen*Plus* (New England Nuclear). After 16 hr of bacterial growth on LB agar containing 12.5 μ g of tetracycline per ml, the filters were prepared for hybridization according to the GeneScreen*Plus* manufacturer's manual. Probes for differential screening were prepared from poly(A)⁺ RNA isolated from segments of 5-day-old, uninfected roots and from nodules 10 and 21 days after inoculation, as in the construction of the cDNA library except that 10 μ Ci of [³²P]dATP (specific activity = 3200 Ci/mmol; 1 Ci = 37 GBq; New England Nuclear) was used.

Abbreviations: 2D, two-dimensional; ORF, open reading frame. *Present address: Research Institute ITAL, Keyenbergseweg 6, 6704 PJ Wageningen, The Netherlands.

The filters were hybridized for 72 hr at 65°C to either root or nodule ³²P-labeled cDNA in 0.9 M NaCl/90 mM sodium citrate/5× Denhardt's solution (8)/10 mM EDTA/0.5% NaDodSO₄/100 μ g of sonicated, denatured calf thymus DNA per ml/20 μ g of poly(A) per ml. The filters were washed twice in 0.3 M NaCl/30 mM sodium citrate/0.1% NaDodSO₄ for 15 min at room temperature and twice in 75 mM NaCl/7.5 mM sodium citrate/0:1% NaDodSO₄ for 30 min at 65°C.

In Vitro Translation of Total RNA. Total RNA (3 μ g) from roots or nodules was translated *in vitro* in a wheat germ extract (Bethesda Research Laboratories) in a 15- μ l mixture to which 15–30 μ Ci of [³⁵S]methionine or 6 μ Ci of [³H]leucine was added (according to the manufacturer's manual). Translation products were separated by two-dimensional (2D) gel electrophoresis followed by fluorography of the dried gel to preflashed Kodak XAR5 film (2).

Hybrid-Released Translation. For hybrid-released translation, the pENOD2 insert (10–15 μ g of DNA) was denatured and applied to 0.5-cm² discs of diazophenylthioether-paper (Bio-Rad) essentially as described (8). Total soybean RNA from 16-day-old nodules (750 μ g) was then hybridized to the filter-bound DNA in 300 μ l of 50% (vol/vol) deionized formamide/0.1% NaDodSO₄/0.6 M NaCl/4 mM EDTA/80 mM Tris·HCl, pH 7.8. Hybridization was initiated at 40°C and the temperature was slowly decreased to 37°C over a period of 6 hr. After washing, the bound RNA was eluted (8) and dissolved in 3 μ l of H₂O; 1.5 μ l was translated and analyzed as above.

RNA Transfer Blot Analysis. Total soybean RNA was denatured in dimethyl sulfoxide/glyoxal, electrophoresed in 0.8% agarose gels (8), and transferred to GeneScreen (New England Nuclear) as described (2). The blots were prehybridized for 6 hr in 50% (vol/vol) deionized formamide/1 M NaCl/0.05 M Tris·HCl, pH 7.5/5× Denhardt's solution/0.1% NaDodSO₄/100 μ g of denatured salmon sperm DNA per ml and hybridized with nick-translated (8) probes. Hybridization was performed for 16 hr at 42°C. Blots were washed twice for 15 min at 42°C in 0.3 M NaCl/30 mM sodium citrate/0.1% NaDodSO₄ and twice for 30 min at 42°C in 75 mM NaCl/7.5 mM sodium citrate/0.1% NaDodSO₄.

DNA Sequencing. Standard techniques were used for cloning into M13 and pUC vectors (9) and for dideoxy (10, 11) and for Maxam–Gilbert (12) sequencing. The DNA sequence data were stored and analyzed with programs written by R. Staden (13) on a microVAX/VMS computer.

Cytology. Nodules were fixed for 16 hr in 3% glutaraldehyde in 50 mM sodium phosphate buffer (pH 7.2). After fixation the nodules were rinsed with the same buffer, dehydrated in a graded ethanol series, embedded in Technovit 7100 (Kulzer, Wehrheim, F.R.G.), cut into $5-\mu m$ sections, stained with toluidine blue, and examined under a light microscope.

RESULTS

Characterization of pENOD2. Out of a cDNA library prepared against $poly(A)^+$ RNA of soybean root nodules, 10 cDNA clones were isolated that specifically hybridized with the nodule cDNA probe transcribed from 10-day-old plant $poly(A)^+$ RNA. These clones represent early nodulin genes and will be designated pENOD to distinguish them from pNOD clones that represent nodulin genes expressed at later stages of development. Cross-hybridization studies revealed 8 clones with common sequences, of which the clone pENOD2 with an insert length of 1000 bp was chosen for further characterization.

On DNA transfer blots of EcoRI-digested soybean genomic DNA, five restriction fragments were found to hybridize with ³²P-labeled pENOD2, whereas no hybridization was observed with *B. japonicum* DNA (not shown). Hence, the cloned ENOD2 DNA is encoded by the soybean genome and its gene is probably part of a small gene family.

Clone pENOD2 hybridized to a mRNA of 1200 nucleotides; the concentration of the ENOD2 mRNA is highest at day 10 and decreases during further nodule development (Fig. 1A). Thus, the ENOD2 gene is apparently transiently expressed during soybean nodule development, although in some experiments the concentration of ENOD2 mRNA remained nearly constant between 10 and 21 days.

For comparison of the expression of the ENOD2 genes with that of nodulin genes expressed later in development, a leghemoglobin (Lb) cDNA clone (pLb), selected from the cDNA library by hybridization with a soybean Lb cDNA clone (14) made available by K. Marcker (University of Aarhus, Denmark), was used. The difference in time of expression between ENOD and Lb genes is illustrated in Fig. 1B. The RNA transfer blot was first hybridized with pENOD2 and was subsequently probed with pLb. The Lb mRNAs start to appear when the ENOD2 mRNA concentration is already decreasing.

pENOD2 Codes for N-75 Nodulin. To identify the early nodulin encoded by pENOD2, mRNA was hybrid-selected by pENOD2 and translated in vitro in the presence of [³⁵S]methionine. The results showed that the pENOD2encoded polypeptide has an apparent M_r of 75,000 with an isoelectric point around 6.5 (Fig. 2C). In accordance with the nomenclature established for nodulins (15), the identified polypeptide is named N-75. After in vitro translation of the pENOD2-selected RNA in the presence of [³H]leucine, two polypeptides were found, one of which comigrated with the polypeptide detected after translation with [³⁵S]methionine whereas the other, more prominent, polypeptide was slightly more basic (Fig. 2B). The two polypeptides obtained from hybrid-released translation should be closely related and are each encoded by a different member of the ENOD2 gene family. The polypeptides, both referred to as N-75, were also easily detected as nodulins in the 2D pattern of the polypeptides obtained upon in vitro translation of total nodule RNA (Fig. 2A).

The translation products of these mRNAs of 1200 nucleotides in length have an apparent M_r of 75,000, whereas mRNA of that length has a coding capacity for a polypeptide of, at most, M_r 45,000. This notable discrepancy prompted us



FIG. 1. Autoradiographs of two identical RNA transfer blots (A and B) containing 15 μ g of total RNA isolated from 5-day-old uninoculated roots (R) and nodules (N) harvested 10, 14, and 21 days after sowing and inoculation with B. *japonicum* USDA110. (C) Autoradiograph of an RNA transfer blot containing 15 μ g of RNA isolated from nodules harvested 14 days after sowing and inoculation with B. *japonicum* USDA110 (lane 1) and from nodule-like structures collected 4 weeks after inoculation with R. fredii USDA257 (lane 2). The blots were hybridized with ³²P-labeled pENOD2 (A and C) and with pENOD2 and pLb in consecutive order (B). The positions of the rRNAs are indicated by arrowheads.



FIG. 2. Characterization of the early nodulin cDNA clone pENOD2 by hybrid-released translation (A-C) and time-course analysis of the expression of the nodulin-75 genes during nodule development (D-H). Total RNA from 16-day-old soybean nodules (A) and RNA eluted from filter-bound pENOD2 DNA (B and C) were translated in a wheat germ extract in the presence of [³H]leucine (A and B) or $[^{35}S]$ methionine (C). The products obtained were separated by 2D gel electrophoresis and fluorographed. The positions of the N-75 nodulins are indicated by arrowheads. In D-H only the region of the 2D gel within the square indicated in A is shown that represents the [3H]leucine in vitro translation products obtained from RNA of 5-day-old, uninfected soybean root (D), from root segments of infected soybean plants 6 days after sowing and inoculation with B. japonicum USDA110 (E), and from soybean nodules 7 (F), 10 (G), and 13 (H) days after sowing and inoculation with the same strain. In A the molecular weight of the markers is indicated as $M_r \times 10^{-3}$.

to sequence the cDNA insert of pENOD2 to see if the deduced amino acid sequence could explain the peculiar physical properties of the encoded polypeptides. At the same time information on the structure may provide clues on a possible function of the N-75 nodulins. The sequencing data (Fig. 3) reveal that the Pst I insert of the pENOD2 cDNA clone contains 998 nucleotides including a short 3' poly(A) tail but excluding the dCdG nucleotides generated in the cloning procedure. At most 200 nucleotides of the 5' end of the 1200-nucleotide-long mRNA, including the initiation codon and coding sequences for the N terminus of the polypeptide, are thus missing in this clone. Two open reading frames (ORFs) occur in the same strand (Fig. 3). From one ORF, designated ORF-1, 728 nucleotides are found in the cDNA clone (positions 13-741; Fig. 3), and this ORF ends with two successive termination codons. A second ORF, designated ORF-2, not in phase with the first one, comprises 611 nucleotides (positions 14-625; Fig. 3) and ends in a single termination codon. Both ORFs are followed by 3' nontranslated regions of about 250 (ORF-1) and 375 (ORF-2) nucleotides, respectively, in which two potential poly(A) addition signals are present (Fig. 3).

Although both ORFs seem to be able to code for a nodulin of about the same size, several lines of evidence indicate that only ORF-1 corresponds to a nodulin N-75. First, ORF-1 gives rise to a completely different polypeptide (241 amino acids of which none is methionine) than ORF-2 (203 amino acids of which 20 are methionine). Such entirely different polypeptides will differ in physical properties. The two N-75 in vitro translation products will therefore most likely be related polypeptides with a similar amino acid sequence, derived from two different mRNAs, and not from two ORFs of one mRNA. If then one of the N-75 in vitro translation products is shown to contain no methionine (Fig. 2C) the other N-75 nodulin may have a low methionine content. But the ORF-2-derived polypeptide contains 20 methionines, indicating that it is rich in methionine. We therefore deduce that ORF-1 encodes a nodulin-75. This conclusion is supported by the absence of methionine in the amino acid sequence deduced from this ORF.

Moreover, though both ORFs encode a polypeptide containing repeating peptide sequences, the repetitive amino acid sequences occurring in the polypeptide encoded by ORF-1 are better preserved than those present in the ORF-2-derived polypeptide. This indicates that an evolutionary tendency exists for a functional conservation of the polypeptide coded for by ORF-1. We therefore propose that ORF-1 will exclusively be used for the generation of a nodulin N-75.

ORF-1 shows that the N-75 nodulins are peculiar prolinerich proteins. N-75 contains a repetitive sequence of 10 or 11 amino acids that is repeated at least 20 times. Embedded in this repetitive sequence a heptapeptide sequence is found that is conserved in 19 of 20 units (Fig. 3). This repeated heptapeptide sequence is Pro-Pro-Xaa-Glu-Lys-Pro-Pro, in which 17 times Xaa = histidine and 3 times Xaa = tyrosine or leucine. Three or four amino acids that are not as conserved, mainly proline, glutamic acid/glutamine, and tyrosine, flank the heptapeptide repetitive sequence. Two apparent partials of the heptapeptide repeat are found at positions 133 and 451 within the sequence. Neither α -helix nor β -sheet conformations were found using the method of Lim (16). The high proline content of the N-75 nodulins probably explains the discrepancy between the observed apparent M_r of 75,000 and the coding capacity of a 1200nucleotide mRNA. A similar aberrant migration behavior on NaDodSO₄/polyacrylamide gels is found for the proline-rich protein collagen (17).

N-75 Is Involved in Nodule Morphogenesis. To form an idea of the process in which the proline-rich N-75 protein might be involved, we attempted to correlate the beginning of expression of the N-75 genes with a defined stage in root nodule formation. Total RNA was isolated from tap root segments of 6-day-old inoculated plants, where nodules are not yet visible, and from nodules harvested 7, 10, and 13 days after sowing and inoculation. RNA preparations were analyzed by *in vitro* translation using [³H]leucine followed by 2D gel electrophoresis. The area of the 2D gel where N-75 nodulins are found is shown in Fig. 2 D-H. Both N-75 proteins first appear at day 7 (Fig. 2F) and then increase in concentration up to day 13 (Fig. 2 G and H).

By examining the nodule structures formed by R. fredii USDA257, it proved possible to distinguish between the infection process and the differentiation into a nodule structure. On commercial soybean cultivars, this fast-growing *Rhizobium* strain cannot form nitrogen-fixing root nodules but forms nodule-like structures that are not able to fix nitrogen.

Histological examination of these nodule-like structures revealed that they arise from a combination of cell swellings and randomly oriented cortical cell divisions (Fig. 4). In these nodule-like structures, no infected plant cells or infection threads were observed and none of the structures examined had an organization with vascular bundles at the periphery, similar to normal nodules. Apparently the formation of such a nodule-like structure does not require an infection process. By transfer blot analysis, N-75 RNA could be detected in RNA isolated from these nodule-like structures (Fig. 1*C*).

DISCUSSION

A cDNA library from soybean root nodules has been analyzed for copies of mRNA transcripts of early nodulin genes. These genes are expressed in the early stage of root nodule development when the nodule structure is being formed. One of the clones, pENOD2, was characterized in detail. The nucleotide sequence of the *Pst* I insert of pENOD2 has been determined and the derived amino acid sequence shows that 45% of the amino acids of N-75 is proline and that the amino acid sequence is organized in highly repetitive units (Fig. 3).

In an effort to derive a function for N-75 in root nodule development from these sequence data, we have surveyed the occurrence of proline-rich proteins in plants and their



OR	F-1	1	200	00	P	I	i	E	K	- 1	-	P	P	E	Y	, 	L	P	P	F	<u>.</u> ידר	H	E	K	P	F	5	P	E	Y TAC	L	F	P	P	H	E	K	P	P		P	Е	Y	Q	P		P	H	120
OR	F-2	200		007	100	P	M	K	I	K	H	H	L	54	5	I	Ŷ	L	, U	6	L	M	R	1	100	H	H	Q	N	1	[Y	L	L	M	F			R	H	Q	N	T	1	ICC	L	L	M	120
E	K	(P	P	H	· · · · E	:	 N	P	P	•	Р	E	H	Q	_	P	P	H	E		K	P	P	E	H	1	Q -	P	P	H	E	3	K	P	P	P	Е	Y	. 1	E	P	P	H	E	ł	(P	
GA	GAA R	N N	CA H	CCC F		TGA M	GA R	I	CC/ F	ACC I	AC H	CGC R	GAG S	CAC 1	CA	AC N	CAC H	CT L	CA7	rga f	GA. R	AGC S	CCA H	CCA C	GA 2	GCA S	CC. T	AAC N	CCA H	CCI L		TGA M	NGA R	AGO S	CCA H	CCA H		GA(GTA S	TG/ M	AAC N	CAC: H	CCT L		'GA I	GA/ R	AC N	CA H	240
P	P	,	Е	Y	Q	F	,	Р	H	E	1	К	P	Р	P	1	E	Y	Q	P		P	H	E	K	Р		P	Р	E	Y	Q	2	P	P	H	E	K	P	1	P	Р	Ē	Н	Q	F)	P	
CC/	ACC H	Q Q	AA N	TAC I		ACC N	AC H	CT(L	CA1	'GA I	GA. R	AGC S	CA H	CCA F	.CC.	AG. Q	AA1 N	AC T	CAA	ICC I	AC H	CTC L	AT(M	GAG R	iAA.	ACC N	AC H	CAC H	CAI Q	GAA N	TA	CCA T	AC N	CAC H	CT L	CAT M	GAG		SCC.	ACC H	CAC H	CAC Q	SAG S	CAC T		AC <u>(</u> N	AC(H	CT L	360
H	E	~	K	P	P	E	~~	H	Q	P	1	P	H	E	K	1	P	P	P	E		Y	9	P	P	H		E	K	P	P	- P		E	Y	Q	 Р	P	Q	F	3	ĸ	P	P	H	E	1	K	400
CA1	l'GA (R	AG S	H	(aga Q	S	T	S		ACC H	L	M	R		S	H	H	Q		GT/ S	I	N	H			TG/ M	R	N	H		ACC H	Q	AA'I N	T	N	L	I		AGA K	K	AGU	H	H		rga M	K	n. N	480
P CCA H	P CC	GC R	P CAC Q	E GAA N	Y TAC 1	Q CCA	ACO N	P CTC L	P CT L	H CA'	E TGA M	E AAA K	K AGC S	P CCA H	P CC/	- I ACC H	P CAG Q	E AA(N	H CAC T	Q CA	ACC N	CTC	P CCC P	H CAT L	E GA/	K	GCC S	P CAC H	P CAC H	P CCA Q	V GTC	Y GTA	CC T	P CAC H	P CCO P	P CCT L	Y TAT M	E GAG	K AA/	P ACC	AC H	P CAC H	P CA Q	V GTG C	Y TA1	E FGA	ACC N	P CC P	600
P CCT L	Y TA'	l TG/ M	E AG/ R	K AAG S	P CCA	P	- 1 CCC P	P CAG Q	V TA *	V GT(Y Gta	7 ATC	P CAC	P CT	P CC1	H CA	I ATG	E AG/	K	P CC/	F	CA	I TTT	Y AT	E GAC	P GCC/	F	P CGC	P CAT	L TG	E GAG	K GAA	GC	P CAC	Þ CGC	V STC	Y Fac	N AAT	P CCC	Р 200	, I AC	P CTI	Y AT(G 3GC	R CGC	Y TA	F TCC	S CA	720
P CCA	S TC	H Ca/	K Aga	K AAA	N AAC	* TA	AT/	H AAC	CA	CTI	rgc	CT	GCG	TC.	ACA	TG	TT	TTG	GT	ста	аст	'CA.	AAC	TT.	AGA	CCI	rgo	ccc	TTT	GT	CAT	ATA	A A/	AGC	TTI	TT	GTT	гст	GTT	TA	.AG	ATC	TC/	\A Gʻ	ГАС	: AA '	TAT	G	840
тсс	CT	rc1	rgc	AT	GCA	CT	\C 1	TC	TT	C A 4		ŤA.	AAG	GC	гтт	AT	GC	CTÁ	TG	TAI	ΓΑΑ	TA	стс	TA	CTT	TA/	\T T	(CT	сст	TT	CAC	CA	гсо	GAT	AT1	GT	AT	GTC	AAC	:TA	CT/	AGT	GTC	GG	ГТТ	'AT	CTA	T	960
GGC	TAT	ΓΑΑ	٩ΤΑ	AG	ГТТ	TT	TT	TG	TG	гтт	'AA	ÂÂ			444	AA	AA	4 AA		AAC	ccc	cc	ccc	cco	ccc	cco	ccc	TG																					

FIG. 3. Partial restriction map, sequencing strategy, and nucleotide sequence of the *Pst* I fragment from pENOD2. Sequencing was performed by the Maxam-Gilbert method (12) (open circles) and by the dideoxy method of Sanger (10, 11) (closed circles). The arrows depict the direction and extent of sequencing from the indicated site. The dashed lines correspond to sequences originating within the vectors. In the DNA sequence, nucleotides are numbered on the right of the sequence. The predicted amino acid sequence is shown in standard single-letter code for both ORFs and the characteristic heptapeptide repeat is overlined in ORF-1. The two partial repeats are indicated by a dotted line. Termination codons (*) and potential poly(A) addition sites (+++) are also marked. P, *Pst* I; Hf, *Hinf*I; R, *Rsa* I; A, *Acc* I; He, *Hae* III; Hd, *Hind*III; S, *Sph* I; C, *Cla* I.

assumed functions. All proline-rich proteins that have been described in plants so far contain hydroxyproline that is



FIG. 4. (A) Nodule-like structure on soybean roots obtained 4 weeks after inoculation with R. fredii USDA257. (B) Part of a longitudinal section of a nodule-like structure. Cell divisions in the outer and inner cortical cell layers are indicated by arrows. VB, root vascular bundle; LR, lateral root. (Bar = $100 \ \mu m$.)

posttranslationally formed and in most cases subsequently glycosylated. Although it remains to be established whether the soybean N-75 proteins become hydroxylated and glycosylated *in vivo*, soybean nodule tissue has been described as extremely hydroxyproline-rich (18). Also the apparent absence of a class of proline-rich proteins in plants justifies the assumption that the nodulins N-75 belong to one of the classes of hydroxyproline-rich glycoproteins. There are four major classes of hydroxyproline-rich glycoproteins in plants (19): (*i*) the cell wall structural hydroxyproline-rich glycoproteins or extensins (20, 21), (*ii*) the arabinogalactan proteins (22), (*iii*) the solanaceae lectins (23), and (*iv*) hydroxyproline-rich agglutinins (24).

Extensins are associated with the cell walls of most dicotyledonous plants (25). They are assumed to play a role in maintaining the integrity of the primary cell wall (26) and they may be important in controlling growth and development. Extensins have been shown to accumulate in plant cell walls upon wounding (27) and pathogen attack (28, 29) in what is considered a defense response. The amino acid sequences of extensins are characterized by the occurrence of a repeating Ser-Hyp-Hyp-Hyp-Hyp pentapeptide. Neither this pentapeptide nor the typical high serine content is found in the amino acid sequence of N-75 (Fig. 3). The nodulins N-75 are therefore not closely related to extensins.

After comparison of the amino acid composition of representatives of each of the classes mentioned above with N-75, it can be concluded that the amino acid composition of none of the (hydroxy)proline-rich (glyco)proteins analyzed so far resembles the amino acid composition derived for N-75. Nor is N-75 homologous to the hydroxyproline-rich (glyco)proteins P33 (30) and "protein 4" (31), recently identified, using the same criteria as above. We therefore may be dealing with a hitherto unknown class of (hydroxy)proline-rich proteins, characterized by a remarkably low content of serine and a surprisingly high content of glutamic acid. Both the highly repetitive nature of the amino acid sequence and the high proline content suggest that the N-75 nodulins are structural proteins.

To gain a better understanding of the biological function of the N-75 nodulins in root nodule development, we have studied the expression of their genes as a function of time (Fig. 2 *D*-*H*). The expression of the N-75 genes is first detectable at 7 days after sowing and inoculation, when the nodule meristems just emerge through the root epidermis. We were not able to detect N-75 RNA in 7-day-old tap root pieces with visible nodule structures, whereas we could detect N-75 mRNA in excised nodule structures of the same age. Therefore it cannot be excluded that these genes are already expressed earlier than 7 days after sowing and inoculation but in fewer cells and at a similar or lower level than at 7 days.

Irrespective of the possible expression of these genes before 7 days, a strong stimulation of N-75 gene expression occurs from 7 to 13 days. The N-75 proteins should therefore be involved in a developmental event that proceeds during this period. Around day 7 in our growth system the nodules emerged through the root epidermis. This stage of development equals stage VII as described by Calvert et al. (32), and cytological observations have shown that in this stage the meristems start to differentiate into nodule structures (32). Some of the meristem cells have been invaded by infection threads from which the rhizobia are beginning to be released. To examine whether the induction or stimulation of the expression of the N-75 genes is specifically related to the infection process, on the one hand, or the formation of a nodule structure, on the other, we have looked for N-75 gene expression in nodule-like structures formed by R. fredii USDA257. This strain induces the formation of nodule structures devoid of intracellular bacteria and infection threads (Fig. 4), in which, however, early nodulin N-75 RNA is detectable (Fig. 1C). The expression of the N-75 genes in nodules without bacteria or infection threads strongly suggests that N-75 is not involved in the infection process, but more likely in nodule morphogenesis.

Quantitative light microscopic observations on nodule initiation have shown (32) that up to the stage in which the meristems emerge through the epidermis, development can stop. However, when a meristem has reached the "emergence stage" it will continue to develop into a mature nodule. The stimulation of the expression of the N-75 genes coincides therefore with the moment the soybean nodule meristems have reached an apparently critical developmental stage. The expression of the N-75 genes might therefore reflect the definitive "commitment" of the meristems to develop into a nodule. The nature of the involvement of the hydroxyprolinerich nodulins N-75 in this commitment remains to be established.

We acknowledge the stimulating discussions with Dr. Linda Brebnor and thank Wim Dirkse for help in the Maxam-Gilbert sequencing. This work was supported by an Agrigenetics-sponsored Research Contract (H.J.F., T.G., W.S., and J.L.) and by grants from the Netherlands Organization for the Advancement of Pure Research (J.-P.N.) and the Netherlands Organization for Biological Research (F.G.).

- 1. Fuller, F., Künstner, P. W., Nguyen, T. & Verma, D. P. S. (1983) Proc. Natl. Acad. Sci. USA 80, 2594-2598.
- Govers, F., Gloudemans, T., Moerman, M., van Kammen, A. & Bisseling, T. (1985) *EMBO J.* 4, 861–867.
- Lang-Unnasch, N. & Ausubel, F. M. (1985) Plant Physiol. 77, 833–839.
- Vincent, J. M. (1980) in *Nitrogen Fixation*, eds. Newton, W. E. & Orme-Johnson, W. H. (University Park Press, Baltimore), Vol. 2, pp. 103-131.
- Govers, F., Moerman, M., Downie, J. A., Hooykaas, P., Franssen, H. J., Louwerse, J., van Kammen, A. & Bisseling, T. (1986) Nature (London) 323, 564-566.
- Bisseling, T., van den Bos, R. C. & van Kammen, A. (1978) Biochim. Biophys. Acta 539, 1-11.
- Bhuvaneswari, T. V., Turgeon, B. G. & Bauer, W. D. (1980) Plant Physiol. 66, 1027–1031.
- 8. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 9. Messing, J. (1983) Methods Enzymol. 101, 20-78.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- 11. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
- 12. Maxam, A. M. & Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
- 13. Staden, R. (1984) Nucleic Acids Res. 12, 521-538.
- Jensen, E. Ø., Paludan, K., Hyldig-Nielsen, J. J., Jørgensen, P. & Marcker, K. (1981) Nature (London) 291, 677-679.
- 15. Van Kammen, A. (1984) Plant Mol. Biol. Rep. 2, 43-45.
- 16. Lim, V. I. (1974) J. Mol. Biol. 88, 873-894.
- 17. Freytag, J. W., Noelken, M. E. & Hudson, B. G. (1979) Biochemistry 18, 4761-4768.
- 18. Cassab, G. I. (1986) Planta 168, 441-446.
- McNeil, M., Darvill, A. G., Fry, S. C. & Albersheim, P. (1984) Annu. Rev. Biochem. 53, 635-663.
- 20. Van Holst, G. J. & Varner, J. E. (1984) Plant Physiol. 74, 247-251.
- 21. Chen, J. & Varner, J. E. (1985) EMBO J. 9, 2145-2151.
- Van Holst, G. J., Klis, F. M., de Wildt, P. J. M., Hazenberg, C. A. M., Buijs, J. & Stegwee, D. (1981) *Plant Physiol.* 68, 910-913.
- Allen, A. K., Desai, N. N., Neuberger, A. & Creeth, J. M. (1978) Biochem. J. 171, 665–674.
- 24. Leach, J. E., Cantrell, M. A. & Sequeira, L. (1982) Plant Physiol. 70, 1353-1358.
- Lamport, D. T. A. & Catt, J. W. (1981) in *Plant Carbohy-drates II: Encyclopedia of Plant Physiology*, eds. Tanner, W. A. & Loewus, F. A. (Springer, Berlin), Vol. 13B, pp. 133-165.
- Cooper, J. B. (1984) in Structure, Function and Biosynthesis of Plant Cell Walls, eds. Dugger, W. M. & Bartnicki-Garcia, S. (Waverly, Baltimore), pp. 397–399.
- Chrispeels, M. J., Sadara, D. & Lamport, D. T. A. (1974) J. Exp. Bot. 25, 1157–1166.
- Hammerschmidt, R., Lamport, D. T. & Muldoon, E. P. (1984) Physiol. Plant Pathol. 24, 43-47.
- Showalter, A. M., Bell, J. N., Cramer, C. L., Bailey, J. A., Varner, J. E. & Lamb, C. J. (1985) Proc. Natl. Acad. Sci. USA 82, 6551-6555.
- Chen, J. & Varner, J. E. (1985) Proc. Natl. Acad. Sci. USA 82, 4399-4403.
- Hong, J. C., Nagao, R. T. & Key, J. L. (1985) in Abstracts First International Congress of Plant Molecular Biology, ed. Galau, G. A. (Univ. of Georgia Center for Continuing Education, Athens), p. 91 (abstr. PO-1-117).
- 32. Calvert, H. E., Pence, M., Pierce, M., Malik, N. S. A. & Bauer, W. D. (1984) Can. J. Bot. 62, 2375-2384.