JOURNAL OF BACTERIOLOGY, Dec. 1989, p. 6764–6770
0021-9193/89/126764-07$02.00/0

Vol. 171, No. 12

# *nodO*, a New *nod* Gene of the *Rhizobium leguminosarum* Biovar *viciae* Sym Plasmid pRL1JI, Encodes a Secreted Protein

RUUD A. DE MAAGD,* ANDRÉ H. M. WIJFJES, HERMAN P. SPAINK, JOSÉ E. RUIZ-SAINZ, CAREL A. WIJFFELMAN, ROBERT J. H. OKKER, AND BEN J. J. LUGTENBERG

*Department of Plant Molecular Biology, Botanical Laboratory, Leiden University, Nonnensteeg 3, 2311 VJ Leiden, The Netherlands*

Received 23 June 1989/Accepted 20 September 1989

**The region of the *Rhizobium leguminosarum* biovar *viciae* Sym plasmid pRL1JI, responsible for the production and secretion of a previously described 50-kilodalton protein (R. A. de Maagd, C. A. Wijffelman, E. Pees, and B. J. J. Lugtenberg, J. Bacteriol. 170:4424–4427, 1988), was cloned and its nucleotide sequence was determined. A new *nod* gene, *nodO*, preceded by a poorly conserved *nod* box, was identified and its transcriptional start site was determined. Comparison of its predicted protein product with the N-terminal amino acid sequence of the isolated secreted protein showed that *nodO* is the structural gene of this protein, although the nucleotide sequence predicted a protein only 30,002 daltons in size. This comparison also showed that the secreted protein is not the product of N-terminal processing of a larger precursor. A conventional N-terminal signal sequence was not detected in the NodO protein. The NodO protein has significant homology with a part (residues 720 to 920) of the hemolysin protein (HlyA) of *Escherichia coli*. Analysis of the transcriptional regulation of the *nodO* gene revealed that, in contrast with other *nod* promoters in this species, activity of the *nodO* promoter is greatly enhanced in the presence of multiple copies of the *nodD* gene.**

*Rhizobium leguminosarum* is a gram-negative soil bacterium which induces nodules on the roots of plants of the family *Leguminosae* (32). Within these nodules the bacteria, differentiated into bacteroids, fix atmospheric nitrogen.

Bacterial genes, which are essential for nodule formation (*nod* genes) and nitrogen fixation (*fix* and *nif* genes), are located on large Sym (symbiosis) plasmids (5, 11, 14). Expression of *nod* genes is induced by flavonoids, which are excreted by the host plant roots, and requires the *nodD* gene product (10, 19, 21, 23, 24, 29, 35).

In an earlier study we identified a secreted, flavonoid-inducible, Sym plasmid (pRL1JI)-dependent protein of *R. leguminosarum* biovar *viciae* with an apparent molecular size of 50 kilodaltons (kDa) (3). Production of this protein was greatly enhanced in the presence of multiple copies of the *nodD* gene. We have produced mutants lacking this protein and identified a region on the Sym plasmid pRL1JI responsible for its production (2). Depending on the bacterial chromosomal background and the host plant species, mutations in this region either do not affect nodulation or delay nodulation and result in lower nodule numbers per plant. No immunologically cross-reacting proteins were found in strains of other biovars, suggesting that this protein may be unique for *R. leguminosarum* biovar *viciae* strains.

The 50-kDa protein described by us is the first secreted protein reported for *R. leguminosarum*. In this paper we describe the cloning of the pRL1JI region involved in the production of the secreted protein and the determination of the nucleotide sequence of both the structural gene for the protein and the preceding promoter region. The transcriptional regulation of this gene, which appears to be different from that of earlier identified *nod* genes, is also characterized.

* Corresponding author.

## MATERIALS AND METHODS

**Strains and plasmids.** Relevant strains and plasmids used in this study are listed in Table 1.

**Enzymes and chemicals.** Lyophilized large fragment (Klenow) of DNA polymerase I was obtained from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.). A Sequenase version 2.0 kit was obtained from Rijnland Chemische Produkten en Instumentenhandel (Capelle a/d IJssel, The Netherlands). Polynucleotide kinase and reverse transcriptase were obtained from Promega Biotech (Leiden, The Netherlands). All other enzymes and M13 primers were purchased from Boehringer GmbH (Mannheim, Federal Republic of Germany). Other primers for sequencing were obtained from Isogen Bioscience (Amsterdam, The Netherlands). [$\alpha$-$^{35}$S]dATP, [$\alpha$-$^{35}$S]dCTP, and [$\gamma$-$^{32}$P]dATP were purchased from Amersham International plc (Amersham, United Kingdom). All enzymes were used according to the specifications of the manufacturers.

**DNA sequencing.** DNA sequencing was performed on both strands, using the dideoxy chain termination method (26) with the M13 vectors tg130 and tg131 (15) and large fragment (Klenow) of DNA polymerase I. As a control, all sequences were also analyzed by using the Sequenase 2.0 kit with dITP instead of dGTP in the chain termination reactions. Some regions with strong secondary structures were confirmed by running sequence gels supplemented with 50% deionized formamide. Restriction sites used for cloning in M13 were *Hin*dIII, *Bgl*II, *Eco*RI, *Sph*I, *Sal*I, *Pst*I, and *Bam*HI.

**DNA isolation and plasmid constructs.** Recombinant DNA techniques were carried out essentially as described by Maniatis et al. (17). Broad-host-range plasmids were mobilized from *Escherichia coli* to *R. leguminosarum*, using pRK2013 as a helper plasmid (4). Selection of transconjugants was done on YMB medium (12) with the addition of 5 mg of chloramphenicol and 500 mg of streptomycin per liter (with IncQ plasmids) or 2 mg of tetracycline per liter (with IncP plasmids) for plasmid selection and 20 mg of rifampin per liter to select against *E. coli*.

TABLE 1. Strains and plasmids used in this study

| Strain or plasmid | Characteristics | Source or reference |
|---|---|---|
| *E. coli* | | |
| KMBL1164 | Δ(*lac-pro*) *thi* $F^-$ | P. van de Putte |
| JM101 | Δ(*lac-pro*) *supE thi* (F' *traD36 proAB lacI*$^q$ *lacZ*ΔM15) | 36 |
| *R. leguminosarum* | | |
| LPR5045 | bv. *trifolii* RCR5, Sym plasmid cured, Rif$^r$ | 13 |
| RBL5560 | LPR5045 carrying pJB5JI (=pRL1JI *mep*::Tn*5*) | 14, 34 |
| RBL5580 | LPR5045 carrying pRL1JI::Tn*1831* Δ50kb, from within *nodE* to the left | 27 |
| Plasmids | | |
| pIJ1089 | IncP carrying a 30-kb pRL1JI fragment | 5 |
| pIC20R | Intermediary cloning vector | 18 |
| pRK2013 | Helper plasmid for mobilization | 4 |
| M13tg130 | Phage cloning vector for sequencing | 15 |
| M13tg131 | Phage cloning vector for sequencing | 15 |
| pMP220 | IncP vector with promoterless *lacZ* | 27 |
| pMP190 | IncQ vector with promoterless *lacZ* | 27 |
| pMP77 | IncQ vector with promoterless *xylE* | J. Marugg[a] |
| pMP157 | pMP190 containing *nodD* of pRL1JI | 27 |
| pMP240 | pMP220 containing pRL1JI promoter *nodABCIJ* | 3 |
| pMP280 | pMP92 containing *nodD* of pRL1JI | 30 |
| pMP454 | pMP220 carrying *Pst*I-*Bgl*II fragment of pRL1JI containing *nodO* | This study |
| pMP455 | pMP220 carrying *Pst*I-*Bam*HI fragment of pRL1JI containing promoter *nodO* | This study |
| pMP446 | pMP220 carrying *Bam*HI-*Bgl*II fragment of pRL1JI containing *nodO* coding sequence | This study |
| pMP468 | pMP77 containing *Hin*dIII fragment of pMP280 with *nodD* gene of pRL1JI | This study |
| MPM98 | M13tg131 carrying *Bgl*II-*Pst*I fragment of pRL1JI containing promoter *nodO* | This study |
| pMP465 | pMP190 with *Bgl*II fragment of MPM98 containing *nodO* promoter and M13 primer sequence | This study |

[a] Ph.D. thesis, State University of Utrecht, The Netherlands, 1988.

**Determination of transcriptional start site.** Details of the method used for determination of the transcriptional start site are given elsewhere (28). The *Bgl*II-*Bam*HI fragment containing the *nodO* promoter was first cloned in the M13tg131 vector, resulting in plasmid MPM98. Subsequently, a *Bgl*II fragment of MPM98, containing the *nodO* promoter with the M13 primer sequence at the 3' end was cloned in the IncQ vector pMP190, resulting in plasmid pMP465. This plasmid produced fusion mRNA, which could be used for primer extension experiments with the 15-mer M13 sequencing primer. LPR5045 containing pMP465 and pMP280 (an IncP vector containing *nodD* of pRL1JI) was grown for 8 h in the presence of 100 nM naringenin, and mRNA was isolated by methods described previously (31). Primer extension experiments were performed by the method of Maniatis et al. (17), using $^{32}$P-end-labeled DNA primers. The resulting end product was compared on a gel with a sequence ladder of the noncoding strand obtained from MPM98, which was sequenced by the dideoxy chain termination method with $^{32}$P-end-labeled primer.

**Induction assays.** Assays for β-galactosidase activity, using 100 nM naringenin as the *nod* gene inducer, were performed as described previously (27). Each test was performed in duplicate, and the variation of the expression levels was within 20%.

**Immunodetection.** Immunodetection of the secreted NodO protein, using Western blotting (immunoblotting) with rabbit antiserum, was performed as described by de Maagd et al. (2).

**Amino acid sequencing.** Protein was isolated by electroelution from acrylamide gels as described previously (2). Eluted protein was subsequently dialyzed against double-distilled water, precipitated with 9 volumes of acetone, and resolubilized in water for amino acid sequencing. Sequence analysis was performed with a gas phase sequenator (model 470A; Applied Biosystems), using 25% trifluoroacetic acid in water as the conversion reagent. The resulting phenylthiohydantoin amino acids were analyzed on-line by reversed-phase high-pressure liquid chromatography on a phenylthiohydantoin analyzer (model 120A; Applied Biosystems) with a phenylthiohydantoin $C_{18}$ column (2.1 by 220 mm) (Applied Biosystems).

## RESULTS

**Cloning of the pRL1JI region responsible for production and secretion of the 50-kDa protein.** In our earlier study (2) we had demonstrated that pIJ1089, a cosmid clone of pRL1JI, contains a region which is necessary for production of the secreted, naringenin-inducible 50-kDa protein. Using pIJ1089, we subcloned fragments of pRL1JI into the vector pMP220 (27) (Fig. 1). These subclones were subsequently introduced into RBL5580. This strain contains a pRL1JI derivative with a large deletion, starting within the *nodE* gene to the left. This plasmid appeared to be lacking a region necessary for production of the secreted protein (2). Clones which could complement RBL5580 for production and secretion of the protein were selected by immunodetection with a specific antiserum against the secreted protein. This resulted in the isolation of pMP454, containing a 1.6-kilobase (kb) *Pst*I-*Bgl*II fragment sufficient for complementation of production and secretion of the protein in RBL5580. Our earlier obtained Tn*5* insertions in pIJ1089, inhibiting production of the secreted protein, were mapped in the same fragment (2). pMP454, together with the *nodD* clone pMP157, was sufficient to enable the Sym plasmid-cured strain LPR5045 to produce and secrete the protein, showing that besides *nodD* and the 1.6-kb fragment, no other parts of the Sym plasmid pRL1JI are essential for production and secretion of the protein.

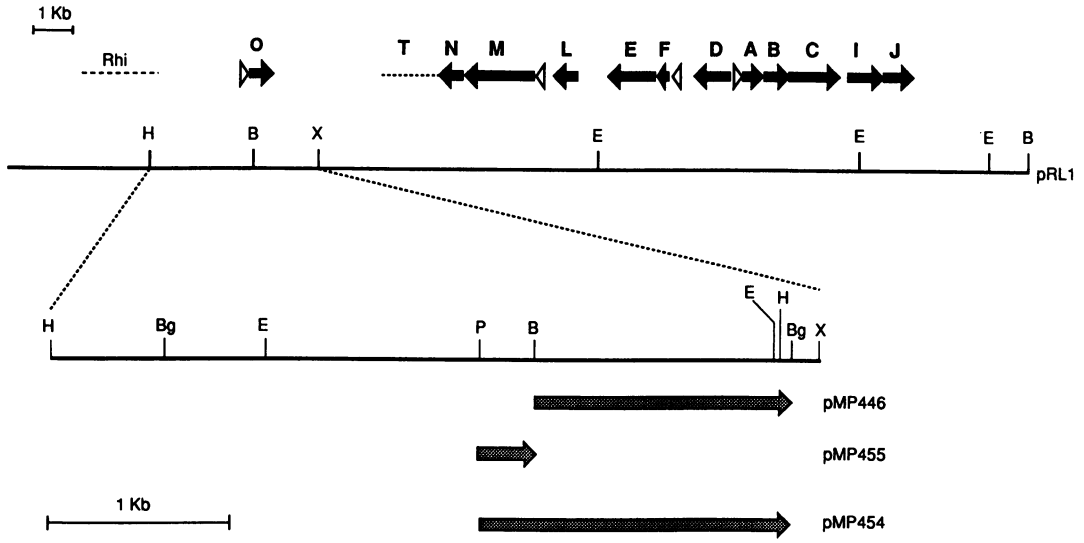


FIG. 1. Restriction fragments of pRL1JI used in this study. Solid arrows show the positions and transcription directions of the known *nod* genes. Open arrowheads represent known *nod* boxes. Dashed lines show the approximate positions of the *nodT* locus (H. C. J. Canter Cremers, H. P. Spaink, A. H. M. Wijfjes, E. Pees, C. A. Wijffelman, R. J. H. Okker, and B. J. J. Lugtenberg, Plant Mol. Biol., in press) and the *Rhi* locus (6). Hatched arrows indicate the subclones of pRL1JI used in this study and their orientation towards the promoterless *lacZ* gene of the vector pMP220 (see text and Table 1). Restriction sites are indicated as follows: B, *Bam*HI; E, *Eco*RI; P, *Pst*I; Bg, *Bgl*II; H, *Hin*dIII.

```
                                                                  PstI
241   CCACGCCTGGAGCTGAGGTTTTCGATCTGCAAAGCACCCTGAGATCAGGTGCTCTGCAGA
301   TTTGTCTTCAGCGTATACGAGGGAAGAAGTTGTGGCCTTCGTCAACGGCCGCCGATCGTC
361   ATAGCCCCCAGTCGTTTTCATATCTGCCGGCCAACTACGAAGGGCGTGCCGTGCGGCCGA
                                  nod-box
421   GATAAACATTTTCGCATCCGTCATTCAAATAGGTCATATCAAAACAATGGATTTCACTAA
                                TS                SD
481   TTCGCTCTTGGAAAAGATAAGGGGCACAGGCGGCGCCCGTTGCCTAATAAGGAGTATATG
                                          BamHI
541   CGATGAATATCAAAGGCAGTGATAACGGCAGTTTTATCAAAGGATCCCCTGAAAACGACA
        M  N  I  K  G  S  D  N  G  S  F  I  K  G  S  P  E  N  D  I  -
601   TAATCGATGGCGGCAAGAAGAACGACTGGATTGATGCCGGCAACGGCGATGATCGGATCA
        I  D  G  G  K  K  N  D  W  I  D  A  G  N  G  D  D  R  I  K  -
661   AAGCTGGTGACGGCCAAGACAGCATCACGGCCGGTCCGGGCCATGACATTGTCTGGGCCG
        A  G  D  G  Q  D  S  I  T  A  G  P  G  H  D  I  V  W  A  G  -
                primer
721   GGAAAGGCTCAGACGTAATCCATGCCGACGGTGGTGACGATCTCTTGTACAGCGACGCCT
        K  G  S  D  V  I  H  A  D  G  G  D  D  L  L  Y  S  D  A  S  -
781   CCTACCCCTTATACGTCACCGACCCTCACCGCGTAATACCGCATAGCGGCGAGGGCGATG
        Y  P  L  Y  V  T  D  P  H  R  V  I  P  H  S  G  E  G  D  D  -
841   ACGTGCTCTACGCCGGCCCTGGCAGCGATATACTTGTGGCTGGTGACGGCGCAGATGTTC
        V  L  Y  A  G  P  G  S  D  I  L  V  A  G  D  G  A  D  V  L  -
901   TGACTGGCGGCGACGACGGCGACGCCTTCGTGTTTCGGTTCCACGACCCTATGGTTGGAA
        T  G  G  D  D  G  D  A  F  V  F  R  F  H  D  P  M  V  G  T  -
961   CAACGCACTGCTATACGAGTGTGATGGATTTCGACACGAAGCAGGACCGCTTTGTCCTGG
        T  H  C  Y  T  S  V  M  D  F  D  T  K  Q  D  R  F  V  L  D  -
1021  ACGCCGCAGATTTCGGTGGTGACCGGAATCTGTTTGATGCAAATTTCATCAATCATTCCA
        A  A  D  F  G  G  D  R  N  L  F  D  A  N  F  I  N  H  S  K  -
1081  AGGGCTTTCCGGGCGAATTTGTGGACACCTTCTACAACGGCGCGGCCGAAGGCGCGCACG
        G  F  P  G  E  F  V  D  T  F  Y  N  G  A  A  E  G  A  H  G  -
                                  primer
1141  GCGAGCACGTCGTGGTAATCACTGATCGAGGCTTTGCGTCTGCCGCTGCCGCCGCGACTG
        E  H  V  V  V  I  T  D  R  G  F  A  S  A  A  A  A  A  T  A  -
1201  CTATTGATCACGAAGCCCGCGGTGACATCATTGTCTTCCATGATCAAAAAACTCTCGGTC
        I  D  H  E  A  R  G  D  I  I  V  F  H  D  Q  K  T  L  G  Q  -
1261  AAGATGGCGAAACTCACGGTGCGACACTAGCCTATGTCGATTCTGCGAACCACGCGCATG
        D  G  E  T  H  G  A  T  L  A  Y  V  D  S  A  N  H  A  H  A  -
      SphI      SalI
1321  CCTTCGCTCATGTCGACAATCTGCACGACATGTCGGATCTTACCTCGCTTACGGCGGAAA
        F  A  H  V  D  N  L  H  D  M  S  D  L  T  S  L  T  A  E  N  -
1381  ATTTCGGCTTCATTTAATTCGATGATCCGAGGAGCGTTCCACCCTTGGGGCGCTTCTCTT
        F  G  F  I  *
1441  TTCCAACATGGCGCAGGGAACTGAAAATAGAAACGACGTGATTTTATTGATCGACTGCAC
1501  CAGTAAAGGTACGCCATTGAAACAAGTTCTCGTCGCCGATGACGACGCCGCCATGCGCCA
1561  CCTGCATCCTGGGGCGGTTGCGCAAGCTGACTTTCTTCTCGCTGGCTGAGGCCAATGCCG
1621  CTATTGGCACTTGATCGCATCAACGATCACCTCATGCGTCGATTGGGTGTTTACCCGGCG
                                                              EcoRI
1681  GCAAGTATTTGAACGTGTCGAACGTGCTGCGCTCGCTAGCCTCCCGGGTGAAACTACGAA
                                                    HindIII
1741  TTCGCXGAATGGCGTCTGCTCCGTGTCTCGACCGATTATCACGTCGAGTTCAAAAGCTTC
1801  TTCTATTCCGTCCCTCATGCCCTCATTCGCCAGCAGGTCGATCTTAGAGCAACGGCGCGC
        BglII
1861  ACCATCGAGATCT
```

**Sequence analysis.** The *Pst*I-*Bgl*II fragment of pMP454 was subsequently sequenced. The resulting sequence, with the features described below, is illustrated in Fig. 2. A screening for sequence homology with the consensus sequence of the *nod* box (a general feature of flavonoid-inducible *nod* genes [25]) revealed a *nod* box-like sequence (Fig. 2) located within a *Pst*I-*Bam*HI fragment. A long open reading frame starting 42 base pairs downstream of this *nod* box is also indicated in Fig. 2. The codon usage of the indicated open reading frame is very similar to that of the *nodA*, *nodB*, and *nodC* genes of fast-growing rhizobia, which suggests that the open reading frame is a structural gene (data not shown). The open reading frame is preceded by a possible ribosome-binding site (Fig. 2). This gene, which we designated *nodO*, codes for a protein of 284 amino acids with a predicted molecular size of 30,002 daltons.

To test whether the gene identified above codes for the secreted protein, we have compared the predicted amino acid sequence with the sequence of the electroeluted protein as determined by gas phase amino acid sequencing. Sequencing successfully identified amino acid residues 4 to 18 of the purified protein, and these matched the predicted amino acids of the open reading frame in the same positions. Residues 1 to 3 could not be identified because of contamination by glycine, probably from the gel electrophoresis used for purifying the protein. These results indicate that *nodO* is the structural gene for the secreted protein. Moreover, these results show that the protein is not produced by N-terminal processing of a larger parental form. Analysis of the amino acid sequence, using the algorithm of Von Heijne

FIG. 2. Nucleotide sequence of the *Pst*I-*Bgl*II fragment of pRL1JI in pMP454 (GenBank accession number M29532). The translated amino acid sequence of the large open reading frame (*nodO*) is given in single-letter code. The primers used for sequencing, the position of the putative *nod* box, the transcription start site (TS), and a putative Shine-Dalgarno sequence (SD) are also indicated.

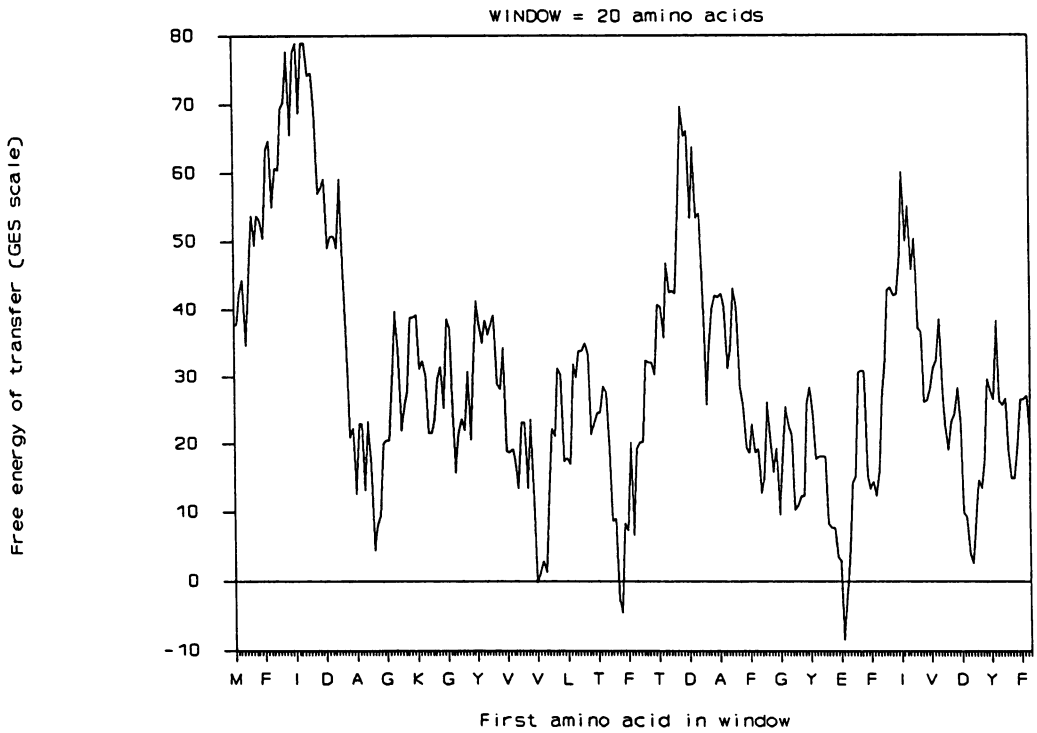


FIG. 3. Hydropathy plot of the NodO protein, produced with the algorithm of Engelman et al. (7), using a window of 20 amino acids. The vertical axis shows the free energy of transfer from water to oil in kilocalories (1 cal = 4.184 J) per mole.

(33), revealed no putative signal sequence involved in the export of protein. A hydropathy profile of the predicted amino acid sequence, made with the algorithm of Engelman et al. (7), is shown in Fig. 3. Almost the entire length of the protein appears to be very hydrophilic, which is consistent with its presence in the growth medium. Furthermore, the protein has a relatively high content of phenylalanine (17 residues) and tyrosine (7 residues) residues.

**NodO is homologous to part of the hemolysin A protein (HlyA) of *E. coli*.** The amino acid sequence of the NodO protein was compared with the protein sequence data base of the National Biomedical Research Foundation. The highest degree of homology was found with the amino acid sequence of the *hlyA* gene product, hemolysin, of *E. coli* (9). This homology had a quality of 120.2 (using the GenDataBase: SWGapPep.Cmp symbol comparison table) and 27% amino acid similarity for the entire length of the NodO protein, as calculated by BESTFIT of the GCG sequence analysis software (University of Wisconsin, Madison). The homology was concentrated in the area of residues 700 to 900 of hemolysin. Figure 4 shows the alignment of the NodO sequence with this part of the hemolysin sequence.

**Transcription analysis of the *nodO* gene.** Promoter activity of pRL1JI fragments containing portions of the *nodO* gene was tested by cloning fragments in front of the promoterless *lacZ* gene in pMP220. The original *nodO*-containing *Pst*I-*Bgl*II fragment in pMP454 (Fig. 1) showed no inducible promoter activity in either direction, suggesting that this fragment contained a complete transcriptional unit. Subsequently, the 0.3-kb *Pst*I-*Bam*HI-fragment of this clone containing the *nod* box sequence described above and the adjacent *Bam*HI-*Bgl*II fragment were subcloned and tested for inducible promoter activity in both directions. Only the former fragment showed naringenin-inducible, *nodD*-dependent promoter activity directed towards the *nodO* reading frame (pMP455 in Fig. 1). The 1.3-kb *Bam*HI-*Bgl*II fragment in pMP446 showed neither production of the protein nor inducible promoter activity. These results indicate the presence of a flavonoid-inducible promoter controlling expression of *nodO* (transcribed from left to right in Fig. 1). Although homology between the consensus sequence of the *nod* box and the promoter region of the *nodO* gene was found, the *nod* box was poorly conserved. Figure 5 shows the *nod* box of the *nodO* gene, aligned with those of the *nodA*, *nodF*, and *nodM* genes of pRL1JI as well as with the consensus sequence defined by Spaink et al. (27). Ten mismatches with the consensus sequence were found, which is more than in any of the other *nod* box sequences determined so far (27).

By using the primer extension method, the transcription start site in the promoter fragment was determined; the results are shown in Fig. 6. Transcription starts 24 base pairs downstream of the *nod* box, a position which is similar to that found for other *nod* promoters of pRL1JI (28), confirming that the identified *nod* box preceding *nodO* is functional.

**Effects of *nodD* gene copy number on transcription of *nodO*.** In our earlier studies we found that, unlike the wild-type situation in which the secreted protein is produced in very low amounts, the introduction of multiple copies of the *nodD* gene leads to increased production (2). To assess whether the number of *nodD* copies affects production of the NodO protein at the transcriptional level, we compared the induction of transcription of the *nodO* promoter with that of the *nodA* promoter of the same Sym plasmid, pRL1JI, by measuring the induction of β-galactosidase activity from both promoters cloned in front of a promoterless *lacZ* gene in the IncP vector pMP220. The construction of the IncP vector containing the *nodO* promoter pMP455 was as described above. Plasmid pMP240, containing the *nodA* pro-

```
nodO   2 NIKGSDNGSFIKGSPENDIIDGGKKNDWIDAGNGDDRIKAGDGQDSITAG 51
hlyA 720 ELIGTTRADKFFGSKFADIFHGADGDDHIEGNDGNDRLYGDKGNDTLSGG 769
nodO  52 PGHDIVWAGKGSDVIHADGGDDLLYSDASYPLYVTDPHRV...IPHSGEG 98
hlyA 770 NGDDQLYGGDGNDKLIGGAGNNYLNGGDGDDELQVQGNSLAKNVLSGGKG 819
nodO  99 DDVLYAGPGSDILVAGDGADVLTGGDDGDAFVF.....RFHDPMVGTTHC 143
hlyA 820 NDKLYGSEGADLLDGGEGNDLLKGGYGNDIYRYLSGYGHHIIDDDGGKDD 869
nodO 144 YTSVMDFDTKQDRFVLDAADFGGDRNLFDANFINHSKGFFGEFVDTFYNG 193
hlyA 870 KLSLADIDFRDVAFRREGNDLIMYKAEGNVLSIGHKNGITFKNWFEKESG 919
```

FIG. 4. Alignment of NodO (top line, residues 2 to 193) and HlyA (bottom line, residues 720 to 919). Identical and similar residues are connected by vertical dashes. The pairs of similarity used here (I and L, V and I, V and L, W and Y, F and Y, D and E, and K and R) each have a score of 0.8 or higher in the PAM250 matrix of Dayhoff (1), in which identical pairs each have a score of 1.5. Gaps in the aligned sequences are indicated ( · ).

moter, was described previously (3). Both constructs were tested in backgrounds with one *nodD* gene copy as well as with multiple *nodD* gene copies. Either the wild-type copy in pRL1JI (RBL5560) or an IncQ-*nodD* clone, pMP468, was used as the source of *nodD*. pMP468 was obtained by cloning the *nodD*-containing *Hin*dIII fragment of pMP280 into the IncQ vector pMP77; results of the induction experiments are shown in Table 2. The induced activity of the *nodA* promoter was raised by only 60% when the number of *nodD* gene copies was increased. In contrast, the induced activity of the *nodO* promoter, which was initially low compared with that of the *nodA* promoter with one *nodD* gene copy, was raised by 650% in the presence of multiple *nodD* gene copies. These results show that the maximum activity of the *nodO* promoter is at least comparable to that found for the *nodA* promoter. Expression of the cloned *nodF* and *nodM* promoters under similar conditions was, as for the *nodA* promoter, raised only slightly by raising the number of *nodD* gene copies (data not shown). These results clearly show that expression of the *nodO* promoter has regulation features which are different from those described for the other known inducible *nod* promoters of pRL1JI.

## DISCUSSION

**NodO is the structural gene for the secreted, naringenin-inducible 50-kDa protein.** In this study we describe the cloning and analysis of the structural gene for a previously described Sym plasmid-dependent, flavonoid-inducible protein of *R. leguminosarum* biovar *viciae* (2). This gene, designated *nodO*, is located in a new transcription unit located at the left of the already identified *nod* genes of pRL1JI. It is under transcriptional control of a so far unidentified *nod* box. The region in which the gene is located is identical to the location of earlier identified Tn*5* insertions in mutants, which could not produce the secreted protein (2). The nodulation locus *nolR* described by Economou et al. (6) was also localized in this region. The exact location of this *nolR* gene, as well as its nucleotide sequence, was also recently determined by this group and appeared to be identical to *nodO* (A. W. B. Johnston, personal communication). It is generally accepted now to name the gene *nodO*.

**Properties of the *nodO* gene and its product.** The *nodO* gene and its product have a number of interesting properties. First, the NodO protein is the first rhizobial protein that has been shown to be secreted into the growth medium. In gram-negative bacteria, in which the outer membrane forms an extra barrier for the export of proteins from the cytoplasm to the exterior, several different mechanisms have evolved to overcome this problem (22). In most known examples of protein transport through the cytoplasmic membrane, the presence of an N-terminal signal sequence is required and export is followed or accompanied by processing by a signal peptidase. Our observation that the secreted protein (NodO protein) shows no evidence of processing, as well as the fact that no apparent signal sequence could be found, suggests that the NodO protein is exported in an unusual manner. Although uncommon, export of proteins lacking N-terminal signal sequences does occur in *E. coli*, as was shown for colicins (22), hemolysin (8), and, very recently, curlin (20).

Second, although the molecular size of the NodO protein was originally estimated at 50 kDa by gel electrophoresis, the translated sequence of the open reading frame of *nodO* allows only for the production of a protein of 30 kDa. The reason for this extremely anomalous behavior of the NodO protein in electrophoresis is yet unclear. Possible explanations are posttranslational modification of the protein or low sodium dodecyl sulfate-binding capacity. Several possibilities are currently being studied in our laboratory.

Third, the regulation of expression of the *nodO* gene at the transcriptional level appears to be different from that of other inducible *nod* genes in this species. We have shown that, although the maximally observed expression level of the *nodO* promoter is the same as that of the *nodA* promoter, this level was only reached when multiple copies of the *nodD* gene were present. With one *nodD* gene copy present, the induced *nodO* promoter showed only 23% of the activity of the induced *nodA* promoter. This result may, at least partly, explain the overproduction of the NodO protein when multiple *nodD* gene copies are present (2). The different behavior of the *nodO* promoter compared with that of other promoters of pRL1JI could be caused by the fact that the *nod* box preceding *nodO* is poorly conserved. As could be expected, a strain with an IncQ clone of *nodD* forms increased amounts of NodD protein (H. Schlamann, personal communication). A low level of NodD protein in the wild-type situation could favor induction of transcription of

```
                                                                    **
nodA     GGGTTGAATATCCATTCCATAGATGATTGCCATCCAAACAATCAATTTTACCAATCTTTCGGATCACTTATAGAAAACCCGG  176 bp to nodA
                                                                    **
nodF     CGAGCCACAATCCATAGTGTGGATGCTTTTGATCCACACAATCAATTTTACCAATGATGCCATATGATCCATAGCAGGGCAG  150 bp to nodF
nodM     GTGGGCGACATCCATATCGTGGATGATAGCTATCCCAACAATCAATTTTACTAATCTGTTTGGATTTATTAGCACGCGCTGG   69 bp to nodM
Consensus        YATCCAY..YUYUGATG....Y.ATC.AAACAATCUATTTTACCAATCY  1-13 bp  AT(T)AG
                                                                   *
NodO     CATTTTCGCATCCGTCATTCAAATAGGTCATATCAAAACAATGGATTTCACTAATTCGCTCTTGGAAAAGATAAGGGGCACA   35 bp to nodO
```

FIG. 5. Comparison of the *nod* boxes of *nodA*, *nodF*, *nodM*, and *nodO* of pRL1JI with the consensus sequence as defined by Spaink et al. (27). Mismatches are underscored. The transcription start sites of the *nodA*, *nodF*, and *nodO* genes at the right end of the sequence are indicated (*). Y, Pyrimidine; U, purine.

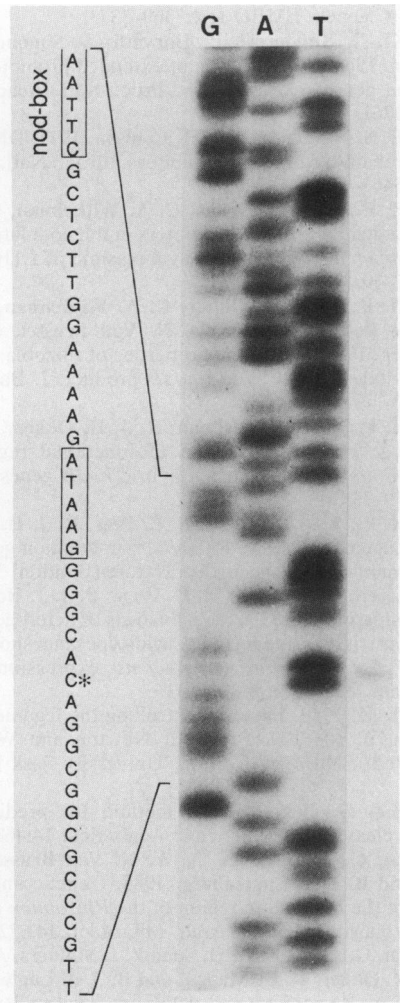


FIG. 6. Determination of the transcriptional start site of *nodO* by primer extension. The right lane shows the primer extension, with sequencing lanes (G, A, and T) of the noncoding strand shown on the left. The surrounding coding strand sequence of the transcription start site (*) is shown. Also shown is the last part of the *nod* box.

the more conserved *nod* promoters, while keeping transcription of *nodO* at a relatively low level. This could provide the cell with a mechanism for fine-tuning *nod* gene expression.

**Function of *nodO* at the molecular level.** At present, the function of the NodO protein at the molecular level, as for most of the other identified *nod* gene products, remains unknown. Although the homology with the hemolysin A protein is significant, it does not provide hard evidence for the function of NodO because hemolysin is much larger than NodO protein (1,023 versus 284 amino acids) and functional regions of hemolysin A have not been identified in detail. However, it is interesting that hemolysin is also a secreted protein without an N-terminal signal sequence. The region of HlyA, which is homologous to NodO, contains a tightly clustered block of repeats of the consensus sequence GGB GBBXLX (16). It was suggested that in HlyA, this block may form an important structural domain separating the N-terminal toxin part of the molecule from the C-terminal secretory domain. Whether NodO has a similar domain structure remains to be determined. NodO protein may have two functions, as it was shown by Economou et al. that mutation of the *nodO* gene affects expression of the *rhiA* product (6), which is located in the cytoplasm (3). This initially seems inconsistent with the extracellular localization of NodO protein and requires further study. The extracellular NodO protein may have some function in the communication between the bacterium and its host plant, possibly through interaction with the host plant cell surface.

TABLE 2. Comparison of induction of expression of the *nodA* and *nodO* promoters in different backgrounds

| Promoter (gene) | β-Galactosidase activity (U, $10^{-3}$) in background: | | | |
|---|---|---|---|---|
| | RBL5560 | | LPR5045(pMP468) | |
| | No inducer | Naringenin | No inducer | Naringenin |
| pMP240 (*nodA*) | 0.2 | 6.8 | 0.2 | 11.6 |
| pMP455 (*nodO*) | 0.8 | 1.6 | 0.7 | 12.0 |

[a] pMP468 is the IncQ vector pMP71 containing the cloned *nodD* gene of pRL1JI. Naringenin at 100 nM was used as the inducer.


## ACKNOWLEDGMENTS

J. Ruiz-Sainz is supported by fellowship BAP-0522-NL of the biotechnology program of the Commission of the European Community.

We thank Pieter D. Wassenaar and J. H. Brouwershaven of the Unilever Research Laboratory, Vlaardingen, The Netherlands, for performing the amino acid sequencing and John M. A. Verbakel for helpful discussions.


## LITERATURE CITED


1. **Dayhoff, M.** 1978. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Md.
2. **De Maagd, R. A., H. P. Spaink, E. Pees, I. H. M. Mulders, A. Wijfjes, C. A. Wijffelman, R. J. H. Okker, and B. J. J. Lugtenberg.** 1989. Localization and symbiotic function of a region on the *Rhizobium leguminosarum* Sym plasmid pRL1JI responsible for a secreted, flavonoid-inducible 50-kilodalton protein. J. Bacteriol. **171**:1151–1157.
3. **De Maagd, R. A., C. A. Wijffelman, E. Pees, and B. J. J. Lugtenberg.** 1988. Detection and subcellular localization of two Sym plasmid-dependent proteins of *Rhizobium leguminosarum* biovar *viciae*. J. Bacteriol. **170**:4424–4427.
4. **Ditta, G., S. Stanfield, D. Corbin, and D. R. Helinski.** 1980. Broad host-range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. USA **77**:7347–7351.
5. **Downie, J. A., Q. S. Ma, C. D. Knight, G. Hombrecher, and A. W. B. Johnston.** 1983. Cloning of the symbiotic region of *Rhizobium leguminosarum*: the nodulation genes are between the nitrogenase and a nifA-like gene. EMBO J. **2**:947–952.
6. **Economou, A., F. K. L. Hawkins, J. A. Downie, and A. W. B. Johnston.** 1989. Transcription of *rhiA*, a gene on a *Rhizobium leguminosarum* bv. *viciae* Sym plasmid, requires *rhiR* and is repressed by flavanoids that induce *nod* genes. Mol. Microbiol. **3**:87–93.
7. **Engelman, D. M., T. A. Steitz, and A. Goldman.** 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. **15**:321–353.
8. **Felmlee, T., S. Pellett, E.-Y. Lee, and R. A. Welch.** 1985. *Escherichia coli* hemolysin is released extracellularly without cleavage of a signal peptide. J. Bacteriol. **163**:88–93.
9. **Felmlee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163**:94–105.
10. **Firmin, J. L., K. E. Wilson, L. Rossen, and A. W. B. Johnston.** 1986. Flavonoid activation of nodulation genes in *Rhizobium* reversed by other compounds present in plants. Nature (London) **324**:90–92.

11. **Hombrecher, G., N. J. Brewin, and A. W. B. Johnston.** 1981. Linkage of genes for nitrogenase and nodulation ability on plasmids in *Rhizobium leguminosarum* and *Rhizobium phaseoli*. Mol. Gen. Genet. **182**:133–136.
12. **Hooykaas, P. J. J., P. M. Klapwijk, M. P. Nuti, R. A. Schilperoort, and A. Rorsch.** 1977. Transfer of the *Agrobacterium* Ti plasmid to avirulent agrobacteria and to rhizobia ex planta. J. Gen. Microbiol. **98**:477–484.
13. **Hooykaas, P. J. J., F. G. M. Schnijdewindt, and R. A. Schilperoort.** 1982. Identification of the Sym plasmid of *Rhizobium leguminosarum* strain 1001 and its transfer to and expression in other Rhizobia and *Agrobacterium tumefaciens*. Plasmid **8**: 73–82.
14. **Johnston, A. W. B., J. L. Beynon, A. V. Buchanon-Wollaston, S. M. Setchell, P. R. Hirsch, and J. E. Beringer.** 1978. High frequency transfer of nodulation ability between strains and species of *Rhizobium*. Nature (London) **276**:634–636.
15. **Kieny, M. P., R. Lathe, and J. P. Lecocq.** 1983. New versatile cloning and sequencing vectors based on bacteriophage M13. Gene **26**:91–99.
16. **Mackman, N., K. Baker, L. Gray, R. Haigh, J.-M. Nicaud, and I. B. Holland.** 1987. Release of a chimeric protein into the medium from *Escherichia coli* using the C-terminal secretion signal of haemolysin. EMBO J. **6**:2835–2841.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
18. **Marsh, J. L., M. Erfle, and E. J. Wykes.** 1984. The pIC plasmid and phage vectors with versatile cloning sites for recombinant selection by insertional inactivation. Gene **32**:481–485.
19. **Mulligan, J. T., and S. R. Long.** 1985. Induction of *Rhizobium meliloti nodC* expression by plant root exudate requires *nodD*. Proc. Natl. Acad. Sci. USA **82**:6609–6613.
20. **Olsen, A., A. Jonsson, and S. Normark.** 1989. Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*. Nature (London) **338**:652–655.
21. **Peters, N. K., J. W. Frost, and S. R. Long.** 1986. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* genes. Science **233**:977–980.
22. **Pugsley, A. P., and M. Schwartz.** 1985. Export and secretion of proteins by bacteria. FEMS Microbiol. Rev. **32**:3–38.
23. **Redmond, J. W., M. Batley, M. A. Djordjevic, R. W. Innes, P. L. Kuempel, and B. G. Rolfe.** 1986. Flavones induce expression of nodulation genes in *Rhizobium*. Nature (London) **323**: 632–635.
24. **Rossen, L., C. A. Shearman, A. W. B. Johnston, and J. A. Downie.** 1985. The *nodD* gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the *nodA, B, C* genes. EMBO J. **4**:3369–3373.
25. **Rostas, K., E. Kondorosi, B. Horvath, A. Simoncsits, and A. Kondorosi.** 1986. Conservation of extended promoter regions of nodulation genes in *Rhizobium*. Proc. Natl. Acad. Sci. USA **83**:1757–1761.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Spaink, H. P., R. J. H. Okker, C. A. Wijffelman, E. Pees, and B. J. J. Lugtenberg.** 1987. Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Plant Mol. Biol. **9**:27–39.
28. **Spaink, H. P., R. J. H. Okker, C. A. Wijffelman, T. Tak, L. Goosen-de Roo, E. Pees, A. A. N. Van Brussel, and B. J. J. Lugtenberg.** 1989. Symbiotic properties of rhizobia containing a flavonoid-independent hybrid *nodD* product. J. Bacteriol. **171**: 4045–4053.
29. **Spaink, H. P., C. A. Wijffelman, R. J. H. Okker, and B. J. J. Lugtenberg.** 1989. Localization of functional regions of the *Rhizobium nodD* product using hybrid *nodD* genes. Plant Mol. Biol. **12**:59–73.
30. **Spaink, H. P., C. A. Wijffelman, E. Pees, R. J. H. Okker, and B. J. J. Lugtenberg.** 1987. *Rhizobium* nodulation gene *nodD* as a determinant of host specificity. Nature (London) **328**:337–340.
31. **Van Slogteren, G. M. S., J. H. C. Hoge, P. J. J. Hooykaas, and R. A. Schilperoort.** 1983. Clonal analysis of heterogenous crown gall tumour tissues induced by wild-type and shooter mutant strains of *Agrobacterium tumefaciens*: expression of T-DNA genes. Plant Mol. Biol. **2**:321–333.
32. **Vincent, J. M.** 1980. Factors controlling the legume-*Rhizobium* symbiosis, p. 103–129. *In* W. E. Newton and W. H. Orme-Johnson (ed.), Nitrogen fixation. University Park Press, Baltimore.
33. **Von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14**:4683–4690.
34. **Wijffelman, C. A., E. Pees, A. A. N. Van Brussel, R. J. H. Okker, and B. J. J. Lugtenberg.** 1985. Genetic and functional analysis of the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. Arch. Microbiol. **143**:225–232.
35. **Wijffelman, C. A., B. Zaat, H. Spaink, I. Mulders, A. A. N. Van Brussel, R. Okker, R. De Maagd, and B. J. J. Lugtenberg.** 1986. Induction of *Rhizobium nod* genes by flavonoids: differential adaptation of promoter, *nodD* gene and inducers for various cross-inoculation groups, p. 123–135. *In* B. Lugtenberg (ed.), Recognition in microbe-plant symbiotic and pathogenic interactions. Springer-Verlag KG, Berlin.
36. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC vectors. Gene **33**:103–119.