JOURNAL OF BACTERIOLOGY, June 1991, p. 3356–3365
0021-9193/91/113356-10$02.00/0

Vol. 173, No. 11

Studies of the *Bradyrhizobium japonicum* $nodD_1$ Promoter: a Repeated Structure for the *nod* Box

SHUI-PING WANG[1] AND GARY STACEY[1,2]*

Center for Legume Research and Department of Microbiology[1] and Graduate Program of Ecology,[2] The University of Tennessee, Knoxville, Tennessee 37996-0845

Received 9 October 1990/Accepted 21 March 1991

Induction of *nod* genes in *Rhizobium* and *Bradyrhizobium* species is dependent on the presence of plant-produced flavonoids, the NodD protein, and the *cis*-acting *nod* box promoter sequence. Although the *nodD* ($nodD_1$) gene in *Rhizobium* species is constitutively expressed, $nodD_1$ expression in *Bradyrhizobium japonium* is inducible by isoflavones in a manner similar to that of the *nodYABC* operon. A consensus *nod* box sequence is found 5′ of the *nodYABC* operon, whereas a presumptive, *nod* box-like sequence is found 5′ of the $nodD_1$ gene. As an initial step toward examining the $nodD_1$ promoter, the transcriptional start sites of the $nodD_1$ and *nodYABC* operons were determined and found to be 44 and 28 bp, respectively, downstream of their respective *nod* box sequences. A series of deletions of the $nodD_1$ promoter were constructed and fused to the *lacZ* gene. Analysis of the activity of these deletions clearly showed that the divergent *nod* box sequence was essential for $nodD_1$ induction by isoflavones or soybean seed extract. The induction of $nodD_1$ expression requires $NodD_1$, as tested in *B. japonicum* and in a heterologous system, *Agrobacterium tumefaciens*. On the basis of these data, we analyzed the published *nod* box sequences and propose a new consensus sequence composed of paired 9-bp repeats. Analysis of the $nodD_1$ *nod* box and synthetic constructs of the *nodYABC nod* box indicate that at least two 9-bp repeats are required for $NodD_1$-mediated induction. Furthermore, insertions between the paired repeats of the *nodYABC nod* box suggest that orientation of the repeats on opposite faces of the DNA helix is essential for maximum *nod* gene expression.

Bradyrhizobium japonicum is a gram-negative bacterium that induces the formation of nodules on soybean roots. The bacteria within these nodules differentiate into bacteroids that fix atmospheric nitrogen. The induction of nitrogen-fixing nodules is a complex, multistep process requiring a number of plant and bacterial genes. The common *nod* genes, *nodDABC*, have been found in all *Rhizobium* and *Bradyrhizobium* species examined (see, e.g., references 8, 17, 23, 24, 29, 32, 33, and 39). These genes are involved in the induction by the bacteria of root hair curling and plant cortical cell division, two of the earliest stages of infection (2, 7, 16, 29). A unique feature of *Bradyrhizobium* species is the presence of an additional gene 5′ of *nodABC* and within the same operon (23, 33). This gene is termed *nodK* in *Bradyrhizobium* (*Parasponia*) sp. (33) and *nodY* in *B. japonicum* (23). Work with *Rhizobium meliloti* showed that *nodABC* are poorly expressed in free-living cells, but can be induced in the presence of a host-produced flavone, luteolin (22, 25). The inducible expression of *nodABC* is dependent on the presence of a functional *nodD* gene which encodes a positive regulatory protein (22). Similar results involving different, but related, inducers have been reported for other *Rhizobium* spp. (reviewed in reference 18). Previous work has shown that the *nodYABC* operon in *B. japonicum* can be induced 200-fold by soybean seed extract or isoflavone compounds such as genistein or daidzein (1, 12). However, although *nodD* ($nodD_1$) genes in *Rhizobium* spp. are constitutively expressed (22, 28, 30), $nodD_1$ expression in *B. japonicum* is inducible by isoflavones or soybean seed extract (1). As is the case for *Rhizobium* spp., *nod* gene induction in *B. japonicum* is dependent on a functional $NodD_1$ (1).

A conserved 47-bp DNA sequence has been found in the promoters of all host-inducible *nod* operons examined (4, 10, 23, 31, 34, 38). Rostas et al. (31) and Spaink et al. (38) showed that this sequence was essential for the induction of *nodABC* in *R. meliloti* and *R. leguminosarum* bv. viciae, respectively. Subsequent studies have shown the *nod* box to be the site at which NodD binds (9, 10, 14). Previously, Nieuwkoop et al. (23) identified a *nod* box sequence 5′ of the *B. japonicum nodYABC* operon that showed high homology to similar sequences from *Rhizobium* spp. Thus, this sequence is probably involved in the inducible expression of this operon. However, a similar, well-conserved *nod* box was not found 5′ of $nodD_1$; this made it difficult to explain the $NodD_1$-dependent, inducible expression of this gene. Upon comparing and aligning the $nodD_1$ promoter sequence with the classic *nod* box sequences, we identified a divergent, presumptive *nod* box-like sequence and proposed that it was involved in $nodD_1$ expression (1). In this study, we tested this hypothesis by constructing a series of deletions of the $nodD_1$ promoter and measuring their activity fused to the *lacZ* gene of *Escherichia coli*. Preliminary to this work, we also defined the start sites of the *nodYABC* and $nodD_1$ transcripts. The results indicate that the presumptive *nod* box is essential for $nodD_1$ expression and this induction is dependent on $NodD_1$. These results led us to question the validity of the previously proposed *nod* box consensus sequence (31, 38). Upon examining the published sequences for *nod* boxes from various *Rhizobium* and *Bradyrhizobium* spp., we identified a novel, repetitive consensus sequence. The validity of this repetitive model for the *nod* box was tested by using synthetic constructs based on the typical *nod*

* Corresponding author.

box of the *nodYABC* operon. A model concerning the interaction of NodD with the *nod* box is described.

MATERIALS AND METHODS

Bacterial strains and plasmids. *B. japonicum* USDA110 and USDA135 were used as wild-type strains (1). *E. coli* S17-1 (3) was used as the recipient for plasmid transformation and as the donor during transconjugation experiments.

Plasmid pZB19 (1), a $nodD_1$-*lacZ* translational fusion, was used for constructing a series of $nodD_1$ promoter deletions. The plasmid vector pRK290 (6) was used because of its broad host range, which includes *B. japonicum* (see Fig. 2). Cointegrates with pRK290 were constructed in vitro as described by Banfalvi et al. (1), transformed into *E. coli* S17-1, and mated into *B. japonicum* USDA135. This strain was chosen over USDA110 since it gives significantly higher expression of $nodD_1$ (1). For experiments involving *Agrobacterium tumefaciens*, plasmids carrying either a $nodD_1$-*lacZ* or *nodY-lacZ* fusion (pZB19 and pZB27, respectively [1]) were ligated to pSUP104, a Q-group plasmid (27, 35), to create plasmids pSW10419-1 and pSW10427-1, respectively. These conjugative plasmids were then transformed into *E. coli* S17-1 and conjugated into Ti plasmid-cured *A. tumefaciens* A136 (provided by E. W. Nester, University of Washington, Seattle). The plasmid pSW375D was constructed by ligating the *Eco*RI $nodD_1$ fragment from pZB48 (1) into the *Eco*RI site of the chloramphenicol gene on pACYC184 (26) and then integrating the latter plasmid into pPP375, a conjugative plasmid (1). In this way, the *B. japonicum* $nodD_1$ gene is constitutively expressed from the chloramphenicol gene promoter. Plasmid pSW375D was introduced into *A. tumefaciens* by conjugation as described above. Other strains and plasmids used in this study are described in the corresponding legends.

Microbiological techniques. *E. coli* cultures were grown in LB medium (19) with the addition of appropriate antibiotics (i.e., 15 μg of tetracycline per ml and/or 100 μg of ampicillin per ml). Medium, growth, and conjugation conditions for *B. japonicum* strains were as described previously (37).

Enzymes, chemicals, and cloning. Restriction endonucleases were obtained from Promega, Madison, Wis., Bethesda Research Laboratories, Gaithersburg, Md., or New England BioLabs, Beverly, Mass. T4 DNA ligase, T4 polynucleotide kinase, nuclease S1, exonuclease III, and avian myeloblastosis virus reverse transcriptase were obtained from Bethesda Research Laboratories. RNase-free DNase (RQ1 DNase) and nuclease *Bal* 31 were obtained from Promega. Reaction conditions for these enzymes were as recommended by the manufacturer. The sources of flavonoid chemicals and *Glycine max* cv. Essex seeds were as described previously (1). Soybean seed extract was produced as described by Banfalvi et al. (1). Standard plasmid isolation and gene cloning techniques were used (19). DNA fragments were routinely recovered from agarose gels by electroelution.

mRNA extraction. *B. japonicum* strains were cultured to early log phase in RDY medium as described previously (37). Genistein was added to a final concentration of 2 μM, and the culture was incubated for an additional 10 h. After 10 min on ice, the cells were collected by centrifugation at 8,800 × *g* for 10 min in a Beckman JA-17 rotor. The cells were resuspended in 20 mM sodium acetate–1 mM EDTA solution (pH 5.5, 1/100 culture volume in a 50-ml centrifuge tube). β-Mercaptoethanol and vanadyl nucleoside were added to the suspension to a final concentration of 10 mM. A sodium dodecyl sulfate (SDS) solution was added to 0.5%, and the suspension was immediately placed into a 65°C water bath. An equal volume of hot phenol (65°C) equilibrated with 20 mM sodium acetate (pH 5.5) was added. After 5 min, the DNA and protein debris were removed by centrifugation at 30,000 × *g* for 20 min at 4°C. The supernatant containing RNA was extracted twice with phenol-chloroform (1:1, vol/vol) and once with chloroform and precipitated by addition of 2.5 volumes of ethanol and 1/10 volume of 3 M sodium acetate. The RNA was collected by centrifugation at 30,000 × *g* for 15 minutes, washed with cold 80% ethanol, dried, and resuspended in distilled water. Any contaminating DNA in the RNA samples was removed by digestion with RNase-free DNase I followed by sequential extraction with phenol, phenol-chloroform, and chloroform. The samples were then precipitated and dried as described above. The RNA was stored at −70°C for later use. All apparatus and solutions (except Tris-containing solutions) used for mRNA experiments were treated with 0.2% diethylpyrocarbonate and autoclaved.

Primer extension. To determine the transcriptional start sites of the $nodD_1$ or *nodYABC* operons, we synthesized two oligonucleotide primers, 5′-CGCCAAGTTCCCTGAACTA GATTTAG-3′ (26-mer) and 5′-GCAGATATCATCCGTT TCCATATTTCAGG-3′ (29-mer), which were complementary to bases +3 to +28 of the $nodD_1$ coding sequence and +3 to +31 of the *nodY* coding sequence, respectively (23) (see Fig. 1B). End labeling was carried out for 1 h at 37°C by using a volume of 25 μl with 0.1 μg of DNA primer and 200 μCi of [α-^{32}P]ATP and using T4 polynucleotide kinase in the buffer suggested by the supplier. Primer extensions were carried out by the method of Kassavetis and Geiduschek (15). One microliter of the above primer was hybridized to 25 μg of RNA overnight at 30°C. The products were separated by electrophoresis on 6% polyacrylamide DNA sequencing gels. Size standards were obtained by using the same primers in dideoxy sequencing reactions on M13 clones of the $nodD_1$-*nodYABC* intergenic region (23).

DNA sequencing. Double-stranded DNA sequencing was used to delineate the extent of the $nodD_1$ promoter deletions. The 26-mer oligonucleotide for $nodD_1$ as described above was used as the primer. All plasmid purifications and reaction conditions were followed as suggested by the manufacturer for double-stranded sequencing with Sequenase Version 2.0 (United State Biochemical Corp.).

Synthetic *nod* box constructs. Two oligonucleotides were synthesized (a 44-mer, 5′-CATCGATCCATCGTGTGGAT GTGTTAGATCTCTATCGAAACAAT-3′, and a 50-mer, 5′-CGATTGTTTCGATAGAGATCTAACACATCCACACGA TGGATCGATGGTAC-3′; Oligos Etc. Inc.) identical in sequence to the *nodYABC nod* box but containing a 6-bp insertion (see Fig. 4). These two oligonucleotides were annealed and cloned between the *Kpn*I and *Cla*I sites in pZB27, a *nodY-lacZ* fusion plasmid (1). In this way, the wild-type *nod* box in the *nodYABC* promoter was replaced by the synthetic mutant *nod* box, creating plasmid pSWI61 (see Fig. 4). The 6-bp insertion in this synthetic *nod* box introduced a unique *Bam*HI site. After digestion with *Bam*HI, treatment with either S1 nuclease or the Klenow fragment of *E. coli* DNA polymerase, and religation, the 2-

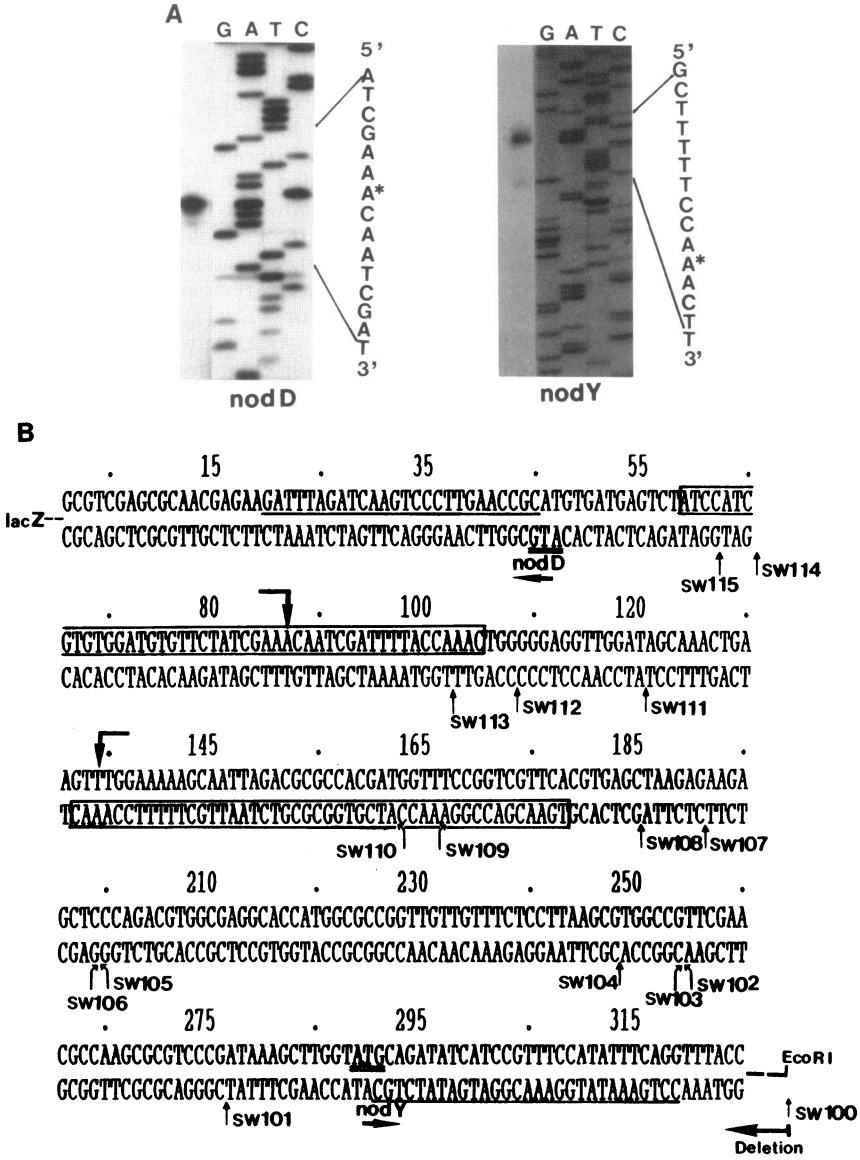


FIG. 1. Determination of the transcriptional initiation sites of the $nodD_1$ and *nodYABC* transcripts of *B. japonicum*. (A) Primer extension of *nodD* (left) and *nodY* (right) mRNA compared with a DNA ladder. Because of the low intensity, the primer extension products were exposed to the X-ray film for a longer time (12 h versus 4 h for the sequencing ladder). The transcriptional start sites are indicated (*). (B) DNA sequence of the intercistronic region of the $nodD_1$ and *nodY* genes. Translational initiation sites for $nodD_1$ and *nodY* are doubly underlined. The $nodD_1$ (lower strand) and *nodY* (upper strand) *nod* box sequences are boxed. Transcriptional start sites are indicated by ⌐ for $nodD_1$ and ⌐ for *nodYABC*. The single underlining indicates the corresponding sequences of the synthetic primers used for primer extension and sequencing. Indicated by an arrow and SW plus a number are the promoter deletions in the $nodD_1$ promoter.

and 10-bp insertional mutants of the *nodYABC* promoter were created (pSWI21 and pSWI101, respectively [see Fig. 4]). Similarly, digestion of pSWI61 with *Cla*I or *Bam*HI and religation removed the upstream sequences, yielding constructs possessing only one or two 9-bp repeats (see Fig. 4). Each of these plasmids was transformed into *E. coli* S17-1 and conjugated into *B. japonicum* USDA110.

Induction and β-galactosidase assay. The induction conditions for *nod* gene expression were described above for mRNA extraction. β-Galactosidase assays were performed with SDS-chloroform-permeabilized cells as described by Miller (21) with minor modifications as described by Wang and Stacey (40).

RESULTS

Transcriptional start sites of the $nodD_1$ and *nodYABC* operons. To determine the transcriptional start sites of the $nodD_1$ and *nodYABC* transcripts, we isolated mRNA from *B. japonicum* USDA110 cells induced for 10 h with 2 μM genistein. Two synthetic oligonucleotides complementary to the coding regions of $nodD_1$ and *nodY*, respectively, were used in primer extension and DNA sequencing reactions with single-stranded M13 clones corresponding to the $nodD_1$-*nodY* intercistronic region (Fig. 1B). The primer-extended products were separated on DNA sequencing gels adjacent to the DNA sequence ladder. As shown in Fig. 1A,

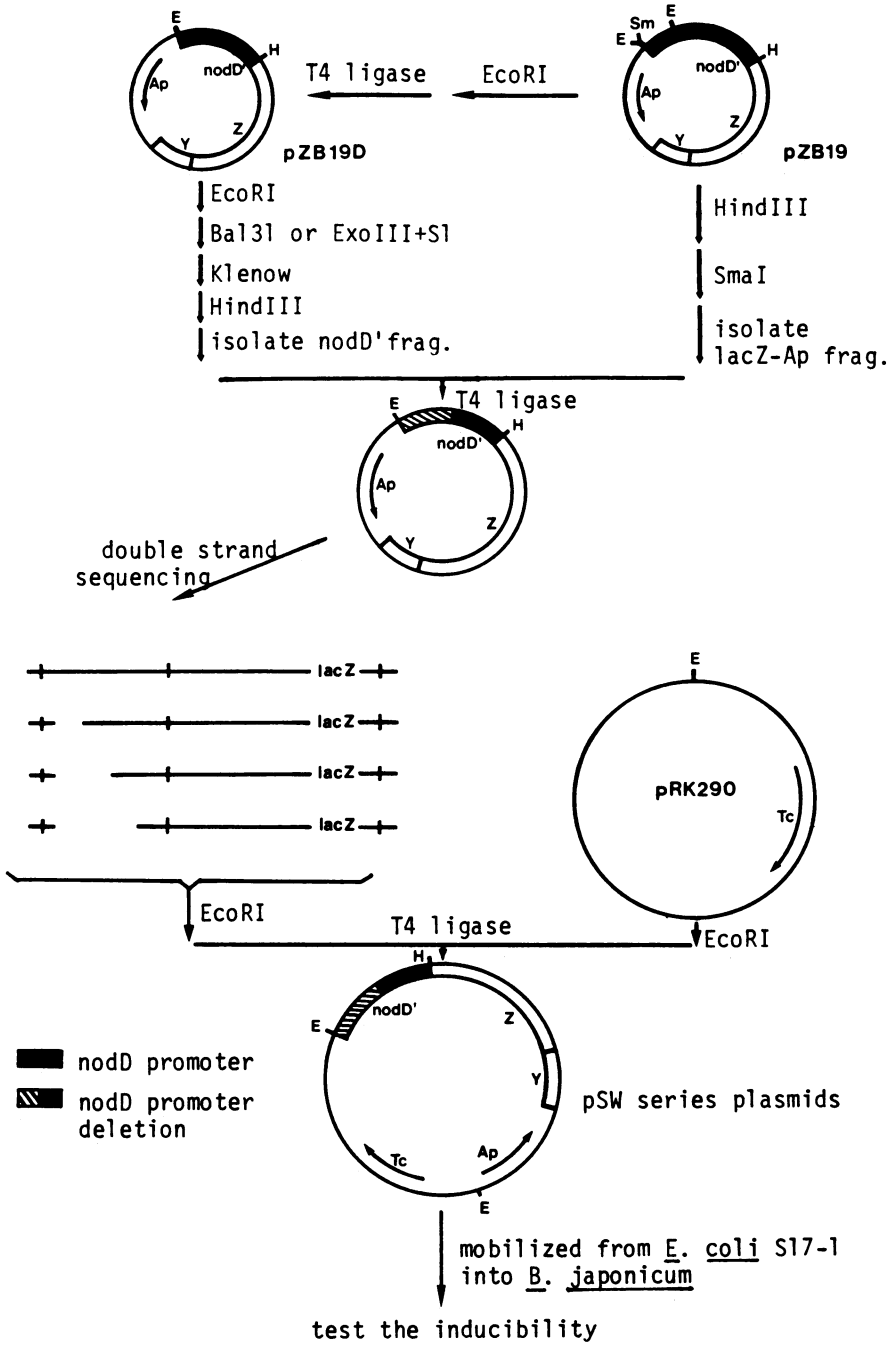


FIG. 2. Strategies for constructing the $nodD_1$ promoter deletion-*lacZ* fusions and testing their inducibility in *B. japonicum*. The plasmids are not drawn to scale. Some of the restriction sites are omitted for clarity. Abbreviations: E, *Eco*RI; H, *Hin*dIII; Sm, *Sma*I; Z, *lacZ*; Y, *lacY*; nodD', $nodD_1$ promoter; Ap, ampicillin resistance; Tc, tetracycline resistance.

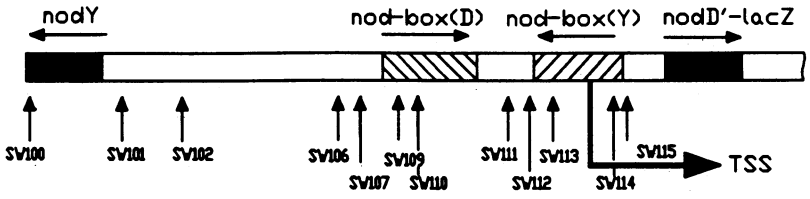
both $nodD_1$ and *nodY* transcripts revealed a single, defined start site. The transcriptional start site for *nodY* is located 28 bp downstream of its previously identified *nod* box. The transcriptional start site for $nodD_1$ is located 44 bp downstream of the presumptive *nod* box (Fig. 1A and B).

The proximity of the $nodD_1$ transcriptional start site to the divergent *nod* box suggests a functional role for this sequence within the promoter (Fig. 1B). It should be noted that the $nodD_1$ transcriptional start lies within the *nodY nod* box, whereas the *nodY* start site lies within the presumptive $nodD_1$ *nod* box. Therefore, there is considerable overlap between the promoters of these two opposing operons.

Deletions of the $nodD_1$ promoter. To test the functionality of the divergent, presumptive *nod* box 5′ of $nodD_1$, we constructed a series of deletions of increasing size initiating well upstream of the $nodD_1$ transcriptional start. As shown in Fig. 2, plasmid pZB19 (1), containing a $nodD_1$-*lacZ* fusion, was used for these constructions. pZB19 was digested with *Eco*RI and ligated, yielding plasmid pZB19D, which possesses a unique *Eco*RI site. pZB19D was then digested with

TABLE 1. Induction of *nodD* promoter deletion *lacZ* fusions in *B. japonicum* USDA135[a]

Plasmid	β-galactosidase (U)		
	inducer: +none	+G	+S.S.E.
pSW100	45	174	182
pSW101	50	124	168
pSW102	37	133	114
pSW106	34	109	164
pSW107	25	131	179
pSW109	34	139	147
pSW110	46	154	146
pSW111	50	32	40
pSW112	51	nt	48
pSW113	47	26	50
pSW114	1	1	0
pSW115	4	1	0



[a] *B. japonicum* USDA135 strains were grown in RDY medium to log phase and then diluted to an optical density at 600 nm of 0.05 to 0.1. The inducer was then added, and the cultures were incubated for approximately 14 h (see Materials and Methods). Miller units are presented (21). Every value is the mean of two assays with a deviation (standard error) less than 10%. The relatively high background levels exhibited by pSW111 to pSW113 are probably due to the influence of vector sequences on expression. The orientation of the cloned insert within the vector was found to influence the level of background expression. The position of the deletion in each of the plasmids is indicated in the schematic figure below by SW plus a number and arrow. Figure 1B shows the position of each deletion with regard to the DNA sequence. Abbreviations: +none, no inducer; +G, 2 μM genistein; +S.S.E., soybean seed extract; nt, not tested.

*Eco*RI, and the resulting linearized DNA was digested either with *Bal* 31 nuclease or with exonuclease III plus S1 nuclease. The Klenow fragment of DNA polymerase was used to repair the DNA ends, and the DNA was subsequently digested with *Hin*dIII. Fragments containing the deleted $nodD_1$ promoter were recovered from an agarose gel by electroelution. The original plasmid, pZB19, was digested with *Sma*I and *Hin*dIII. The vector portion encoding resistance to ampicillin and the *lacZ* gene was also isolated by electroelution from agarose. The vector fragment and the $nodD_1$ promoter fragments were then mixed, ligated, and transformed into *E. coli*. Specific clones were isolated, and the extent of the promoter deletions was determined by double-stranded DNA sequencing. Four classes of deletions were selected for analysis: (i) those ending before the *nod* box; (ii) those ending within the *nod* box; (iii) those ending after the *nod* box but before the transcriptional start site; and (iv) those terminating beyond the transcriptional start site of $nodD_1$ (Fig. 1B; Table 1).

Inducibility of the $nodD_1$ promoter constructs. The expression level of each of the $nodD_1$ promoter deletions was assayed by measuring the β-galactosidase activity. As shown in Table 1, class 1 deletions (i.e., ending before the $nodD_1$ *nod* box [e.g., SW106 and SW107]) were inducible by either soybean seed extract or genistein. In contrast, class 3 deletions (i.e., ending after the *nod* box [e.g., SW111, SW112, and SW113]) were not induced, although they retained a low level of constitutive expression. As expected, the class 4 deletions (i.e., terminating beyond the start of transcription [e.g., SW114 and SW115]) showed a loss of all activity. These results indicate that the presumptive *nod* box is essential for $nodD_1$ expression in *B. japonicum*. Surprisingly, class 2 deletions (e.g., SW109 and SW110) in which only the 5′ portion of the *nod* box sequence was deleted were still inducible by soybean seed extract or genistein.

$NodD_1$ is essential for inducible expression of the *B. japonicum* $nodD_1$ promoter. Previously (1) we provided genetic evidence that $nodD_1$ expression required a functional $NodD_1$. The essentiality of a *nod* box sequence, the presumptive binding site for NodD, further supports the idea

I. the nod-box

Classical consensus:	ATCCATA..	..GGATG..	....	ATCCAAACA	ATCGATTTT ACCAATC
R.m. nodA	ATCCATATC	GCAGATGAT	CGTT	ATCCAAACA	ATCAATTTT ACCAATC
R.m. nodF	ATCCATTTC	ACGGATGGC	CGAC	ATCCAAACA	ATCGATTTT ACCAATC
R.m. nodH	ATTCACAGG	CTGGATCCC	TCTC	ATAAAAACA	ATCGATTTT ACCAATC
R.m. n4	ATCCATAGG	GTGGATGAT	TGCT	ATCCTCATA	ATCGATTTT ACCAATC
R.m. n5	ATCCACCGC	GCGGATAAA	GGTG	-TCCAAACA	ATCGATTTT ACTAATC
R.m. N6	ATCCATAGC	GGAGATGAT	-TGT	TCCCAAACA	ATCGATTTT CACACTC
R.le.nodA	ATCCATTCC	ATAGATGAT	TGCC	ATCCAAACA	ATCAATTTT ACCAATC
R.le.nodF	ATCCATAGT	GTGGATGCT	TTTG	ATCCACACA	ATCAATTTT ACCAATG
R.le.nodM	ATCCATATC	GTGGATGAT	AGCT	ATCCCAACA	ATCAATTTT ACTAATC
R.tr.nodA	ATCCACGCT	GTAGATGAT	TGCG	ATCCAAACA	ATCAATTTT ACCAATC
R.tr.nodF	ATCCATACT	GCGGATGCT	TTCG	ATCCAATCA	ATCAATTTT ACCAATC
B.p. nodA	ATCCATCGT	GTGGATGTA	TTCT	ATCGAAACA	ATCGATTTT ACCAGAT
B.j. nodA	ATCCATCGT	GTGGATGTG	TTCT	ATCGAAACA	ATCGATTTT ACCAAAC
B.j. pRJ4323	ATCCATTCG	GTCGATAGA	TCTG	ATCAAGAAA	ATCAATTTT ACCGATT
B.j. pRJ4198	ATCCGTCAC	CTGGATGCG	TTCC	ATCAAAACA	ATCGATTTT GCGGAGC
B.j. pRJ314	TTCCGTGCC	GTTGATTTC	CATTA	GTAAACATG	ATCGATTTT ACCGCCG
B.j. nodD	TGAACGACC	GGAAACCAT	CGTG	GCGCGTCTA	ATTGCTTTT TCCAAAC
B.j. pRJ310	TGAGCGTGA	TGTCTTCGT	TGAT	CTCGAAGAA	ATCGATTGT TGTGTTG
B.j. pRJ316	CGTCGACCT	CGACAGCGG	CGAC	ATCCAAATCA	ATCGACTTT AGCATGA
B.j. pRJ315	CTCGACTAT	CTCACCACG	CAAT	ATCCAAAC-	ATCGATTCC GGCGCCC

Consensus Repeat: G C A a a

$a_{74}T_{90}C_{88}G_{85}A_{93}T_{89}t_{71}y_{74}t_{74}$

II. Comparision of the consensus repeat sequence

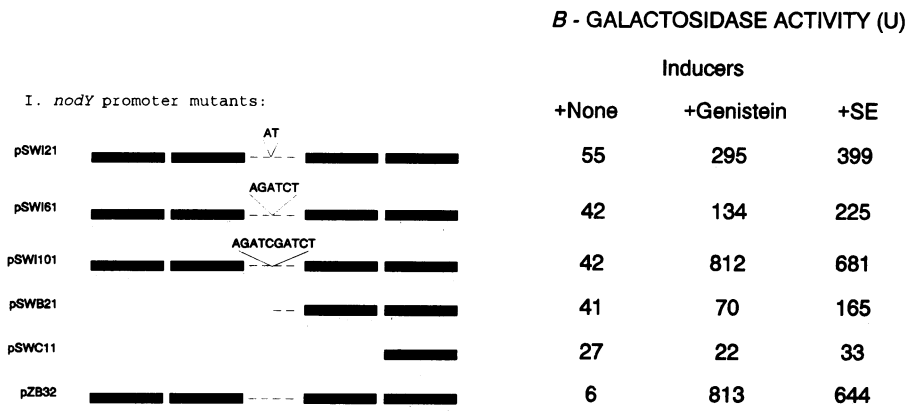
								A	A
MomR/OxyR:	A	T	C	G	A	T	C	G	T
	:	:	:	:	:	:			:
NodD:	A	T	C	G	A	T	T	Y	T
			G	C		A	A		A

FIG. 3. The 47-bp *nod* box and 9-bp consensus repeat sequences. (I) The DNA sequence of 20 *nod* boxes and location of 72 repeat sequences are presented for five different organisms: *R. meliloti* (R.m.) (29), *R. leguminosarum* bv. viciae (R.le.) (32), *R. leguminosarum* bv. trifolii (R.tr.) (30), *Bradyrhizobium* (*Parasponia*) sp. (B.p.) (31), and *B. japonicum* (B.j.) (see reference 12 for the pRj series and reference 23 for $nodD_1$ and *nodA*). The classical consensus sequences were derived from Rostas et al. (31), Nieuwkoop et al. (23), and Fisher and Long (10). The 9-bp conserved repeat sequences are boxed. The subscripts for the consensus repeat sequence are percentages of the occurrences of a base at a specific position found in the 72 repeats examined. The capital letters in the consensus repeat represent the highly conserved base pairs (frequency of occurrence, >80%). (II) Comparison of the consensus repeat sequence of *nod* promoters with that of the MomR/OxyR-regulated promoters. The identical bases between two repeat sequences are indicated (:).

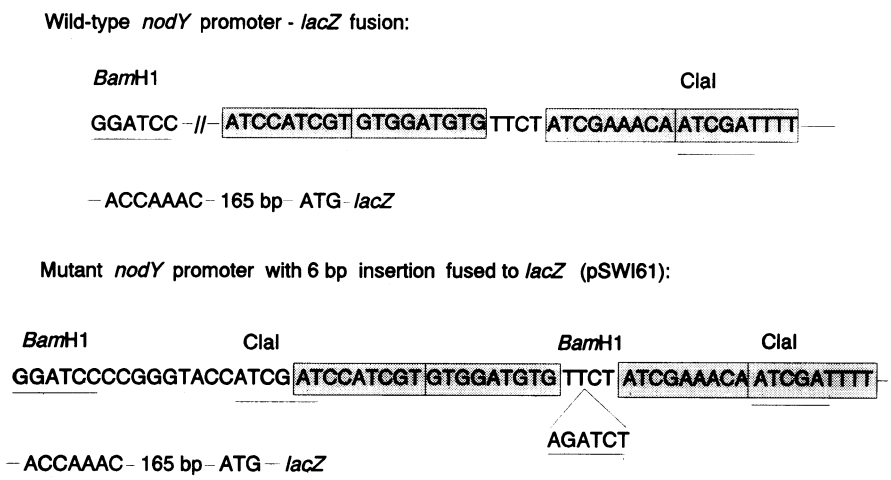
that $nodD_1$ expression is autogenously regulated. To test this further, we used a heterologous system in which the $nodD_1$-*lacZ* or *nodY-lacZ* fusion was introduced in a Ti plasmid-cured *A. tumefaciens* strain. Plasmid pSW375D, encoding the constitutively expressed $nodD_1$ of *B. japonicum*, was also introduced into this strain. The expression of each of the *lacZ* fusion constructs was then tested in the presence or absence of *nod* gene inducers, i.e., 2 μM genistein or soybean seed extract. Inducible expression of β-galactosidase was detected only when the $NodD_1$-expressing plasmid was present. Expression in the *nodY-lacZ* fusion strain was elevated approximately threefold by the addition of genistein (25 to 71 U) and fivefold by the addition of soybean seed extract (25 to 127 U). Expression in the $nodD_1$-*lacZ* fusion strain was elevated approximately 2-fold by the addition of genistein (9 to 17 U) and 2.5-fold by the addition of soybean seed extract (9 to 23 U). The low level of expression is probably due to inefficient transcription from *B. japonicum* promoters in *A. tumefaciens*. However, the induction seen was very reproducible and absolutely dependent on the presence of a functional $NodD_1$.

A novel repetitive consensus sequence for the *nod* box. The fact that a divergent *nod* box is essential for $nodD_1$ induction and that the 5′ portion of this sequence is not necessary for expression (i.e., deletions SW109 and SW110) raises questions about the structure of the previously proposed *nod* box. Therefore, we examined the available *nod* box sequences, searching for a consensus with a better match to the $nodD_1$ *nod* box. This analysis is shown in Fig. 3A and indicates that the *nod* box can be better viewed as a repeat of

A.



B.



C.

FIG. 4. Structure of the synthetic *nod* box constructs and activity of the *nodY* promoter mutant-*lacZ* fusions. (A) β-Galactosidase activity of the various *nodY-lacZ* fusion constructs. The 9-bp repeat sequences are indicated by solid boxes that correspond to the sequences indicated in the boxes in Fig. 4B. The four dashed lines between the two solid boxes represent the 4 bp between the paired repeats in the *nodYABC* promoter. The bases above the corresponding plasmid are the respective insertions in the middle of the 4 bp. The units presented here are mean values of two assays with a deviation of ±10%. The higher background expression levels shown by the pSW plasmids are probably due to the influence of vector sequences on expression (see footnote, Table 1). Abbreviations: +None, no inducer; +Genistein: 2 μM genistein added; +SE: 50 μl of soybean seed extract added. (B) DNA sequence of the wild-type *nodYABC* promoter and the synthetic construct containing the 6-bp insertion. The boxed sequences correspond to the solid boxes in panel A. Some of the restriction sites used in construction of the promoter mutants are underlined and identified above the sequence. (C) Relative activities of the *nodYABC* promoter mutants with different insertions compared with the wild type (panel A, pZB32). The activity of the wild-type promoter was set to 100% (see panel A for values). The inducer was 2 μM genistein.

a 9-bp sequence. Some of the base pairs in the repeat are highly conserved (frequency of 85% or better). Examination of the 9-bp repeat indicates that the conservation is significant within all 72 repeats (average conservation, 82% ± 7.8%). This repeat is found four times within the previously proposed 47-bp conserved *nod* box sequence. However, in the $nodD_1$ *nod* box and three additional putative *nod* box-like sequences previously identified from *B. japonicum*, only two repeats of the 9-bp sequence are present. These two

repeats occur within the $nodD_1$ *nod* box beyond the class 2 deletions (i.e., SW109 and SW110), possibly explaining the inducible expression of these constructs. The 9-bp repeat of the *nod* box is similar (7 of 9 bp identical) to that previously proposed within the promoters regulated by the *E. coli* MomR/OxyR protein (3) (Fig. 3B). This protein is a member of a family of prokaryotic transcriptional regulatory proteins that includes NodD (13).

Validity of the repetitive structure in the *nod* box. As shown in Fig. 3, the *nodYABC nod* box is composed of four 9-bp repeats, arranged as two paired repeats separated by 4 bp. This arrangement indicates that the two paired repeats lie on opposite faces of the DNA helix, since 4 bp is approximately a one-half turn of the helix. If the paired repeats function as unique units, this arrangement on the DNA helix could be essential for optimal NodD-*nod* box interaction. To examine this question, we first synthesized a *nod* box identical in sequence to the *nodYABC nod* box, but containing a 6-bp insertion within the 4 bp separating the paired repeats (pSWI61 in Fig. 4). This insertion was designed to contain a unique *Bam*HI site. As described in Materials and Methods, this *Bam*HI site was then used to construct *nod* boxes containing a 2- or 10-bp insertion between the paired repeats (Fig. 4). These modified *nod* box constructs were then fused to a promoterless *lacZ* gene and introduced into *B. japonicum* USDA110. As shown in Fig. 4, β-galactosidase expression from the *nodYABC* promoter containing a 2-bp insertion was significantly reduced. A further reduction of expression was found with the promoter containing the 6-bp insertion. Both of these insertions would have the effect of rotating the paired repeats in the *nod* box toward the same face of the DNA helix. Complete rotation of the paired repeats obtained by inserting 10 bp resulted in only a slight reduction in β-galactosidase expression. These data suggest that the paired repeats do act as functional units whose orientation on opposite sides of the DNA helix is critical for optimal gene expression.

Our analysis of the $nodD_1$ *nod* box indicated that the presence of only two 9-bp repeats is essential for $NodD_1$-mediated *nod* gene expression. However, this could be a peculiar trait of this *nod* box and may not reflect a general structural requirement for *nod* box function. To test this, we constructed synthetic *nod* boxes of identical sequence to the *nodYABC nod* box, but lacking either the first and second (5′ to 3′) 9-bp repeats or the first, second, and third repeats (Fig. 4, pSWB21 and pSWC11, respectively). These *nod* boxes were also fused to a promoterless *lacZ* gene and introduced into *B. japonicum*. As predicted by the structure of the $nodD_1$ *nod* box, the construct possessing two 9-bp repeats showed inducible expression of β-galactosidase (Fig. 4). The level of expression was significantly reduced from that of the *nodYABC* wild-type promoter; a similar result was seen previously with the $nodD_1$ promoter (1). Also, the level of expression seen with pSWB21 is comparable to that of pSWI61 (Fig. 4), suggesting that in the latter case inducible expression may be due to only the 3′ paired repeats of this promoter. The promoter constructed with only one of the 9-bp repeats exhibited background activity, indicating that a repetitive *nod* box structure is essential for functionality (Fig. 4).

DISCUSSION

Expression of *nod* genes in *Rhizobium* and *Bradyrhizobium* species is induced by the presence of host-produced flavonoids via a process requiring NodD and its corresponding DNA-binding site, the *nod* box. The exceptions to this rule are the $nodD_1$ genes of *Rhizobium* species, which are constitutively expressed (22, 28, 30). The *B. japonicum* $nodD_1$ gene is unusual in that it is induced similarly to other *nod* genes and has a divergent *nod* box-like sequence (1). These facts led us to perform experiments to explain the apparent induction of the *B. japonicum* $nodD_1$ gene in the absence of a well-conserved 5′ *nod* box. As described above, deletions that remove the identified, presumptive *nod* box sequence lead to an inability to induce $nodD_1$ expression, clearly indicating an essential regulatory role for this sequence.

The transcriptional start sites of the few *nod* genes examined in *Rhizobium* species are approximately 25 to 28 bp 3′ of the *nod* box (4, 10). This is also true for the *B. japonicum nodYABC* operon, which is the first example reported for *Bradyrhizobium* spp. However, the transcriptional start site for the $nodD_1$ operon is 44 bp downstream of the divergent *nod* box. This position is indicated by primer extension of $nodD_1$ mRNA and is confirmed by the total loss of expression of deletions that extend just beyond this site (e.g., SW114 and SW115 in Table 1). The reason for the difference in distance between the *nod* box and the beginning of mRNA synthesis for $nodD_1$ and *nodYABC* is not immediately clear. However, it is interesting that the promoters for the $nodD_1$ and *nodYABC* operons clearly overlap with the respective transcriptional start sites lying in the *nod* box of the opposing transcript. We believe that this arrangement is not coincidental, but plays an important, presently unknown role in the coordinate regulation of these two operons. By this reasoning, the placement of the $nodD_1$ transcriptional start is determined by the positioning of the *nodYABC nod* box.

All host-inducible *nod* genes characterized in *Rhizobium* and *Bradyrhizobium* species contain the 47-bp, conserved *nod* box sequence. Previously, a consensus sequence for the *nod* box was proposed, consisting of three highly conserved regions of 7, 5, and 25 bp (5′ to 3′) (10, 23, 31). The functional $nodD_1$ *nod* box does not fit this consensus; nor do other divergent *nod* box sequences reported by Gottfert et al. (12) for *B. japonicum* and by Goethals et al. (11) for *Azorhizobium caulinodans*. Furthermore, deletions in the *B. japonicum* $nodD_1$ *nod* box that remove the 5′ end still allow inducible expression of $nodD_1$, indicating that this region is not essential for the presumptive $nodD_1$ *nod* box function. This is consistent with the fact that this presumptive *nod* box is conserved at its 3′ end and divergent at its 5′ end with respect to the classic *nod* box sequences. These results are different from those of other studies of *Rhizobium* species in which deletions in the 5′ end of the *nod* box resulted in a drastic reduction of inducible expression (38). All of these contradictions led us to examine the published *nod* box sequences for a more universal consensus. This analysis suggests that the *nod* box can best be viewed as a repeat of a 9-bp sequence. In most *nod* box sequences four repeats are apparent, whereas the *B. japonicum* $nodD_1$ *nod* box and related examples possess only two repeats. In those *nod* box sequences with four repeats, two pairs of repeats are found separated by exactly 4 bp, indicating that they are on opposite faces of the DNA helix. Precise insertional mutagenesis of the *nodYABC nod* box indicates that the orientation of the paired repeats on the DNA helix is essential for maximal expression of the *nod* genes. Since the *nodYABC* promoter has high homology to other *nod* box sequences, we believe that this will be generally true of *Rhizobium* as well as *Bradyrhizobium nod* promoters.

NodD is a member of a growing family of prokaryotic

transcriptional regulatory proteins that includes LysR, MomR/OxyR, IlvY, and CysB (13). Members of this LysR family have several common features. They are all positive transcriptional regulatory proteins; they show amino acid conservation, including a helix-turn-helix DNA-binding motif in their N termini; and several are transcribed divergently from operons that they regulate. Similar to the case proposed for the *nod* box, promoters regulated by MomR/OxyR (3) or IlvY (41) also contain four repeats of a 9-bp sequence. The native conformation of one of the members of the LysR family, the CysB protein, has been shown to be a tetramer (20), suggesting that each of the 9-bp repeats may be recognized by one subunit of the active complex. This may also be the case for NodD binding to a *nod* box with four repeats, although it is also possible that a dimer interacts with a paired repeat. Unfortunately, the native structure of the NodD protein has not been determined. Two repeats may be the minimal requirement for the interaction of NodD with the *nod* box sequence, whereas four repeats appear to be required for full activation of *nod* gene expression. The *B. japonicum* $nodD_1$ *nod* box consists of only two 9-bp repeats. This fact may explain the lower level of induction of this operon compared with *nodYABC* (1). This is also consistent with the results for the *nodY* promoter mutant containing two repeats (cf., pSWB21). Moreover, independent induction of $nodD_1$ (i.e., without concomitant induction of *nodYABC*) by substituted isoflavones has been demonstrated (36). The two-9-bp repeat structure of the $nodD_1$ *nod* box may be essential to allow this independent induction and suggests that the $nodD_1$ *nod* box may interact with a different form of NodD (e.g., a dimer). The binding of a multimeric form of $NodD_1$ is supported by the fact that a single 9-bp unit is insufficient for inducible expression (i.e., pSWC11).

In summary, the study of the $nodD_1$ gene in *B. japonicum* led to the elucidation of a novel *nod* box sequence showing significant divergence from the previously accepted consensus sequence. A new consensus sequence is proposed, consisting of four 9-bp repeats, that more closely represents all of the identified *nod* box sequences. The repetitive nature of the *nod* box is supported by mutagenesis experiments that identify functional units within the *nod* box consensus sequence. These observations should be useful for understanding the mechanism of the NodD-*nod* box interaction. They may be useful to other researchers seeking to explain the NodD-flavonoid-dependent induction of genes in the apparent absence of a classic *nod* box sequence (5). Furthermore, the newly proposed *nod* box consensus sequence suggests a number of experiments to describe the native conformation of NodD protein and to more precisely define the structural requirements for DNA binding and transcription initiation. Concomitant with these studies will be efforts to explain the specificity and importance of interaction with the flavonoid inducer. Finally, the specific studies of the *B. japonicum* $nodD_1$-*nodYABC* promoter region indicate that additional regulatory complexity exists, elucidation of which may teach valuable lessons concerning the biology of the *B. japonicum*-soybean symbiosis.

ACKNOWLEDGMENTS

We thank Maria Schell for important technical assistance with this work. We are also grateful to W. Mark Barbour and Gerrit Smit for helpful discussions and critical comments on the manuscript.

This work was supported by Public Health Service grants GM33494-06 and GM40183-03 from the National Institutes of Health and grant 62-600-1636 from the U.S. Department of Agriculture.

REFERENCES

1. **Banfalvi, Z., A. J. Nieuwkoop, M. G. Schell, L. Besl, and G. Stacey.** 1988. Regulation of *nod* gene expression in *Bradyrhizobium japonicum*. Mol. Gen. Genet. **214**:420–424.
2. **Bender, G. L., W. Goydych, B. G. Rolfe, and M. Nayudu.** 1987. The role of *Rhizobium* conserved and host specific nodulation genes in the infection of the non-legume *Parasponia andersonii*. Mol. Gen. Genet. **210**:299–306.
3. **Bolker, M., and R. Kahmann.** 1989. The *Escherichia coli* regulatory protein OxyR discriminates between methylated and unmethylated states of the phage Mu *mom* promoter. EMBO J. **8**:2403–2410.
4. **de Maagd, R. A., A. H. M. Wijfjes, H. P. Spaink, J. E. Ruiz-Sainz, C. A. Wijffelman, R. J. H. Okker, and B. J. J. Lugtenberg.** 1989. *nodO*, a new *nod* gene of the *Rhizobium leguminosarum* biovar *viciae* Sym plasmid pRL1JI, encodes a secreted protein. J. Bacteriol. **171**:6764–6770.
5. **Deshmane, N., and G. Stacey.** 1989. Identification of *Bradyrhizobium nod* genes involved in host-specific nodulation. J. Bacteriol. **171**:3324–3330.
6. **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**:7346–7351.
7. **Dudley, M. E., T. W. Jacobs, and S. R. Long.** 1987. Microscopic studies of cell divisions induced in alfalfa roots by *Rhizobium meliloti*. Planta **171**:289–301.
8. **Egelhoff, T. T., R. F. Fisher, T. W. Jacobs, J. T. Mulligan, and S. R. Long.** 1985. Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC*. DNA **4**:241–248.
9. **Fisher, R. F., T. T. Egelhoff, J. T. Mulligan, and S. R. Long.** 1988. Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. Genes Dev. **2**:282–293.
10. **Fisher, R. F., and S. R. Long.** 1989. DNA footprint analysis of the transcriptional activator proteins NodD1 and NodD3 on inducible *nod* gene promoters. J. Bacteriol. **171**:5492–5502.
11. **Goethals, K., M. Gao, D. Geelen, G. Van den Eede, M. Van Montagu, and M. Holsters.** 1990. Highly divergent *nod* box sequences in *Azorhizobium caulinodans* strain ORS571, p. 536. *In* P. M. Gresshoff, E. Roth, G. Stacey, and W. E. Newton (ed.), Nitrogen fixation: achievements and objectives. Chapman & Hall, New York.
12. **Gottfert, M., J. W. Lamb, R. Gasser, J. Semenza, and H. Hennecke.** 1989. Mutational analysis of the *Bradyrhizobium japonicum* common *nod* genes and further *nod* box-linked genomic DNA regions. Mol. Gen. Genet. **215**:407–415.
13. **Henikoff, S., G. W. Haughn, J. M. Calvo, and J. C. Wallace.** 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA **85**:6602–6606.
14. **Hong, G.-F., J. E. Burn, and A. W. B. Johnston.** 1987. Evidence that DNA involved in the expression of nodulation (*nod*) genes in *Rhizobium* binds to the product of the regulatory gene *nodD*. Nucleic Acids Res. **15**:9677–9689.
15. **Kassavetis, G. A., and E. P. Geiduschek.** 1982. Bacteriophage T4 late promoters: mapping 5′ ends of T4 gene 23 mRNAs. EMBO J. **1**:107–114.
16. **Kondorosi, A., E. Kondorosi, C. E. Pankhurst, W. J. Broughton, and Z. Banfalvi.** 1984. Physical and genetic analysis of a symbiotic region of *Rhizobium meliloti*: identification of nodulation genes. Mol. Gen. Genet. **193**:445–452.
17. **Lamb, J. W., and H. Hennecke.** 1986. In *Bradyrhizobium japonicum* the common nodulation genes, *nodABC*, are linked to *nifA* and *fixA*. Mol. Gen. Genet. **202**:512–517.
18. **Long, S. R.** 1989. *Rhizobium*-legume nodulation: life together in the underground. Cell **56**:203–214.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Miller, B. E., and N. M. Kredich.** 1987. Purification of the *cysB* protein from *Salmonella typhimurium*. J. Biol. Chem. **262**:6006–6009.

21. **Miller, J.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Mulligan, J. T., and S. R. Long.** 1985. Induction of *Rhizobium meliloti nodC* expression by plant exudate requires *nodD*. Proc. Natl. Acad. Sci. USA **82**:6609–6613.
23. **Nieuwkoop, A. J., Z. Banfalvi, N. Deshmane, D. Gerhold, M. G. Schell, K. M. Sirotkin, and G. Stacey.** 1987. A locus encoding host range is linked to the common nodulation genes of *Bradyrhizobium japonicum*. J. Bacteriol. **169**:2631–2638.
24. **Noti, J. D., B. Dudas, and A. A. Szalay.** 1985. Isolation and characterization of nodulation genes from *Bradyrhizobium* sp. (*Vigna*) strain IRc78. Proc. Natl. Acad. Sci. USA **82**:7379–7383.
25. **Peters, N. K., J. W. Frost, and S. R. Long.** 1986. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science **233**:977–980.
26. **Pouwels, P. H., B. E. Engervalk, and W. J. Brammer.** 1990. Cloning vectors. A laboratory manual, p.I-A-iv-9. Elsevier Biomedical Press, Amsterdam.
27. **Priefer, U., R. Simon, and A. Puhler.** 1985. Extension of the host range of *Escherichia coli* vectors by incorporation of RSF1010, replication and mobilization functions. J. Bacteriol. **163**:324–330.
28. **Rolfe, B. G., J. W. Redmond, M. Batley, H. Chen, S. P. Djordjevic, R. W. Ridge, B. J. Bassam, C. L. Sargent, F. B. Dazzo, and M. A. Djordjevic.** 1986. Intercellular communication and recognition in the *Rhizobium*-legume symbiosis, p. 39–47. *In* B. Lugtenberg (ed.), Recognition in microbe-plant symbiotic and pathogenic interactions, vol. H4. Springer-Verlag KG, Berlin.
29. **Rossen, L., A. W. B. Johnston, and J. A. Downie.** 1984. DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nodAB* and *C* required for root hair curling. Nucleic Acids Res. **12**:9497–9508.
30. **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 *nodABC* genes. EMBO J. **4**:3369–3373.
31. **Rostas, K., E. Kondorosi, B. Horvath, A. Simoncsits, and A. Kondorosi.** 1986. Conservation of extended promoter regions of nodulation genes in rhizobia. Proc. Natl. Acad. Sci. USA **83**:1757–1761.
32. **Schofield, P. R., and J. M. Watson.** 1986. DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nodABC* and *nodFE*. Nucleic Acids Res. **14**:2891–2903.
33. **Scott, K. F.** 1986. Conserved nodulation genes from the non-legume symbiont *Bradyrhizobium japonicum* sp. (*Parasponia*). Nucleic Acids Res. **14**:2905–2910.
34. **Shearman, C. A., L. Rossen, A. W. B. Johnston, and J. A. Downie.** 1986. The *Rhizobium leguminosarum* nodulation gene *nodF* encodes a polypeptide similar to acyl-carrier protein and is regulated by *nodD* plus a factor in pea root exudate. EMBO J. **5**:647–652.
35. **Simon, R., U. Priefer, and A. Puhler.** 1983. Vector plasmids for in-vivo and in-vitro manipulations of gram-negative bacteria, p. 98–106. *In* A. Puhler (ed.), Molecular genetics of the bacteria-plant interaction. Springer-Verlag KG, Berlin.
36. **Smit, G., V. Puvanesarajah, R. W. Carlson, and G. Stacey.** 1990. *Bradyrhizobium japonicum nodD* can specifically be induced by soybean seed extract compounds which do not induce the *nodYABC* genes, p. 274. *In* P. Gresshoff, E. Roth, G. Stacey, and W. E. Newton (ed.), Nitrogen fixation: achievements and objectives. Chapman & Hall, New York.
37. **So, J.-S., A. L. M. Hodgson, R. Haugland, M. Leavitt, Z. Banfalvi, A. J. Nieuwkoop, and G. Stacey.** 1987. Transposon-induced symbiotic mutants of *Bradyrhizobium japonicum*: isolation of two gene regions essential for nodulation. Mol. Gen. Genet. **207**:15–23.
38. **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.
39. **Torok, I., E. Kondorosi, T. Stepkowski, J. Postfai, and A. Kondorosi.** 1984. Nucleotide sequence of *Rhizobium meliloti* nodulation genes. Nucleic Acids Res. **12**:9509–9524.
40. **Wang, S.-P., and G. Stacey.** 1990. Ammonia regulation of *nod* genes in *Bradyrhizobium japonicum*. Mol. Gen. Genet. **223**:329–331.
41. **Wek, R. C., and G. W. Hatfield.** 1988. Transcriptional activation at adjacent operators in the divergent-overlapping *ilvY* and *ilvC* promoters of *Escherichia coli*. J. Mol. Biol. **203**:643–663.