MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 2, No. 4, 181-194, 1989

Nucleotide Sequence and Protein Products of Two New Nodulation Genes of *Rhizobium meliloti*, *nodP* and *nodQ*

Julie Schwedock and Sharon R. Long

Department of Biological Sciences, Stanford University, Stanford, CA 94305-5020 U.S.A.
Received 28 October 1988. Accepted 13 March 1989.

Previous studies had suggested the existence of nodulation (*nod*) genes downstream of *nodG* in *Rhizobium meliloti* strain 1021. We have established the DNA sequence and analyzed the translation products of the genes located in this position. Computer analysis of the DNA sequence revealed a number of overlapping putative open-reading frames (ORFs), so we constructed several clones that contained either full-length or truncated ORFs. The protein products of these clones were expressed in both *R. meliloti* and *Escherichia coli in vitro* transcription-translation systems. These assays unambiguously defined the expressed ORFs, which we named *nodP* and *nodQ*. In addition, we found homology to these genes, via Southern hybridizations, elsewhere in the genome of *R. meliloti* strain 1021, and in other species of *Rhizobium*. The *nodP* gene also displayed homology to *E. coli*. A computer search revealed significant homology between NodQ and the GDP binding domain of elongation factor Tu (EF-Tu).

Additional keywords: alfalfa, *in vitro* expression, symbiosis.

Bacteria of the genus *Rhizobium* are able to induce the growth of and invade root nodules on leguminous plants. Inside the host cells of the nodule, the bacteria differentiate into dinitrogen-fixing bacteroids. Neither the bacteria nor the plants are capable of fixing dinitrogen independently (Long 1989).

The genus *Rhizobium* is divided into cross-inoculation groups, based on the legumes they nodulate. *R. meliloti* Dangeard generally nodulates plants of the genera *Medicago*, *Melilotus*, and *Trigonella*, although some strains have a broader host range (more genera), and others have a narrower host range (only some species within a genus).

A variety of genes are necessary for *Rhizobium* species to successfully nodulate their host plants. Tn*5* insertions in the genes *nodDABC* cause a complete inability of *Rhizobium* to nodulate any hosts (Rossen *et al.* 1984; Török *et al.* 1984; Djordjevic *et al.* 1985; Downie *et al.* 1985; Egelhoff *et al.* 1985; Jacobs *et al.* 1985; Debellé *et al.* 1986; Schofield and Watson 1986). Other genes, such as *R. meliloti* genes *nodFE* and *nodH*, are involved in host range and efficiency of nodulation: Tn5 mutations in these genes can result in either inefficient (delayed) nodulation on some hosts or the inability to nodulate other hosts, or both (Djordjevic *et al.* 1983; Downie *et al.* 1983; Hombrecher *et al.* 1984; Schofield *et al.* 1984; Debellé *et al.* 1986; Horvath *et al.* 1986; Swanson *et al.* 1987).

Many of the genes involved in nodulation are inducible by plant exudates (Mulligan and Long 1985; Rossen *et al.* 1985; Shearman *et al.* 1986; Redmond *et al.* 1986; Fisher *et al.* 1987; Zaat *et al.* 1987). The most active factor in alfalfa (*Medicago sativa* L.) exudate has been identified as the flavone luteolin (Peters *et al.* 1986).

The constitutively expressed gene, *nodD1*, is required for the induction of *nodABC* and other inducible nodulation (*nod*) genes (Mulligan and Long 1985, 1989; Fisher *et al.* 1987). In addition, the *nodD* gene product binds the conserved regions known as *nod*-boxes (Rostas *et al.* 1986) that lie upstream of *nodABC*, *nodFE*, and *nodH* (Fisher *et al.* 1988).

Previous studies showed that genes involved in nodulation in *R. meliloti* were located in at least two clusters (Kondorosi *et al.* 1984; Truchet *et al.* 1985). Genetic mapping and DNA sequencing defined these as a cluster containing the common *nod* genes *nodDABC*, and another cluster 5 kilobases (kb) downstream of *nodABC* containing genes involved in host range, *nodFEG* and *nodH* (Debellé and Sharma, 1986; Horvath *et al.* 1986; Swanson *et al.* 1987; Fisher *et al.* 1987). Because these two clusters were present on a DNA fragment that could complement a large Nod^- deletion strain, it was of interest to see whether other *nod* genes were borne on that fragment. Transposon mutagenesis has provided evidence for the presence of genes involved in nodulation lying between the common and host range *nod* clusters (Debellé *et al.* 1986; Swanson *et al.* 1987).

In this study, we evaluate and report the DNA sequence of two new *nod* genes, *nodP* and *nodQ*, found downstream of the host range *nod* cluster. These genes were originally identified by Tn*5* insertions that delay nodulation on alfalfa (Swanson *et al.* 1987). The 5′ end of *nodP* has been sequenced previously and was referred to as ORF216 (Fisher *et al.* 1987). *In vitro* transcription-translation studies and *nodP*::*lacZ* fusion analysis indicate that *nodP* is expressed at a low constitutive level. Its level of expression is affected by the presence or absence of the inducer luteolin only when *nodD* is present in extra copies. Unlike previously studied inducible *nod* genes, there is no apparent *nod*-box consensus sequence between *nodG* and the start of *nodP*.

We have also discovered significant homology to both *nodP* and *nodQ* elsewhere in the genome of Rm1021, which may account for the weak phenotypes of the mutants carrying Tn*5* insertions in these genes.

Nucleotide and/or amino acid sequence data is to be submitted to GenBank as accession number J03676.

MATERIALS AND METHODS

Strains. Strains and plasmids used in this study are listed in Table 1. Those not previously reported are shown in Figure 1 or described below.

Materials. Enzymes were obtained from Bethesda

Table 1. Strains and plasmids

Strain or plasmid	Relevant characteristics or genotype[a]	Source or reference
Escherichia coli		
HB101	F^-; hsdS20(r^-_b, m^-_b); *recA13*	Maniatis *et al.* (1982)
MC1061	*strA*; hsr^-; hsm^+; *lacX74*	Casadaban and Cohen (1980)
JM101	*lacpro*; *supE*; *thi*; F′*tra36*; *proAB*; $lacI^q$	Messing (1983)
W3110	λ^-, IN (*rrnD-rrnE*)1	Hill and Harnish (1981)
Rhizobium meliloti		
RCR2011	Wild type	Meade *et al.* (1982)
Rm1021	Str^r derivative of RCR2011	Meade *et al.* (1982)
A1375	Same as Rm3600, greater than 70-kb deletion pSyma of Rm1021, *nod nif*	Honma and Ausubel (1987)
JT216	*nodP*::Tn*5* derivative of Rm1021	Swanson *et al.* (1987)
JT702	*nodQ*::Tn*5* derivative of Rm1021	Swanson *et al.* (1987)
JSS5	*nodP*::*lacZ* fusion in Rm1021	This study
JSS6	Out-of-frame *nodP*::*lacZ* fusion in Rm1021	This study
JM57	*nodC*::*lacZ* fusion in Rm1021	Mulligan and Long (1985)
Other *Rhizobium* species		
Rhizobium spp. ANU265	Sym^-, heat-cured derivative of NGR234	Morrison *et al.* (1983)
R8401/pRL1JI	*R. leguminosarum* bv. *viciae*	Downie *et al.* (1985)
ANU843	*R. leguminosarum* bv. *trifolii*	Rolfe *et al.* (1980)
Plasmids[b]		
pUC18	Ap^r, *placZ* expression vector	Norrander *et al.* (1984)
pUC118	pUC18 containing M13 origin of replication	Vieira and Messing (1987)
pUC119	Same as pUC118, with reversed polylinker	Vieira and Messing (1987)
pBR322	Ap^r, Tc^r, contains origin of transfer	Bolivar *et al.* (1977); Sutcliffe (1978)
pLAFR1	incP, broad host range vector	Friedman *et al.* (1982)
pRmJT5	incP, 20-kb cosmid-bearing host range genes in pLAFR1	Swanson *et al* (1987)
pRmBE11	30-kb clone-bearing common *nod* and host range genes	Buikema *et al.* (1983)
pRmF30	2.2-kb *Eco*RI fragment containing the 3′ half of *nodP* and the majority of *nodQ* in pUC118	R. Fisher, this laboratory
pRmF31	Same as pRmF30, except in opposite orientation	R. Fisher, this laboratory
pRmF32	3.25-kb fragment containing *nodG* and the 3′ end of *nodP*	Fisher *et al.* (1987)
pZC9	Ap^r, Nm^r derivative of pBR322	D. Biek, personal communication
pRmS63	A 12-kb *Kpn*I fragment of pBE11 containing common *nod* genes and full-length *nodP* and *nodQ*	J. Swanson, this laboratory
pMC931Sp	Contains 9.6-kb *lacZ*, Sp^r cartridge	Mulligan and Long (1985)
pRmJSS30	pRmF30 with one *Eco*RI site filled in	This study
pRmJSS31	pRmF31 with one *Eco*RI site filled in	This study
pRmJSS32	1.8-kb deletion of pRmJSS30	This study
pRmJSS33	0.40-kb deletion of pRmJSS31	This study
pRmJSS34	1.48-kb deletion of pRmJSS30	This study
pRmJSS35	1.07-kb deletion of pRmJSS31	This study
pRmJSS36	330-bp *Sal*I fragment of pRmJSS30 in pUC118	This study
pRmJSS37	Same as pRmJSS36, opposite orientation	This study
pRmJSS38	1.2-kb *Sal*I-*Kpn*I fragment of pRmS63 in pUC118	This study
pRmJSS39	0.4-kb *Sal*I fragment of pRmS63 in pUC118	This study
pRmJSS40	Same as pRmJSS39, opposite orientation	This study
pRmJSS8	Reconstructed fragment in pUC118 containing *nodP* and the majority of *nodQ*	This study
pRmJSS9	Same as pRmJSS8, except *nodP* sequence is 5′ truncated	This study
pRmJSS10	Same as pRmJSS9, except 5′ *nodP* truncation is extended farther	This study
pRmJSS11	*Hin*dIII-*Pst*I fragment of pRmJSS8 in pUC119, containing *nodP* and a 3′ truncation of *nodQ*	This study
pRmJSS12	*Hin*dIII-*Pst*I fragment of pRmJSS9 in pUC119, containing a 5′ truncation of *nodP* and a 3′ truncation of *nodQ*	This study
pRmJSS13	*Hin*dIII-*Pst*I fragment of pRmJSS10 in pUC119, containing a 5′ truncation of *nodP* and a 3′ truncation of *nodQ*	This study
pRmJSS15	Same fragment as pRmJSS38 in pUC119	This study
pRmJSS16	*nodP*::*lacZ* fusion in pUC118	This study
pRmJSS17	27-kDa ORF::*lacZ* fusion in pUC118	This study
pRmJSS16b	Insert of pRmJSS16 in pZC9	This study
pRmJSS17b	Insert of pRmJSS17 in pZC9	This study
pRmJ30	incP, 8.7-kb *Eco*RI fragment containing *nodD1ABC*	Jacobs *et al.* (1985)
pRmSL26	incP, 19.5-kb clone containing *nodD1ABC*	Long *et al.* (1982)
pRmE65	incP, *nodD3* expressed under control of *trp* promoter	Fisher *et al.* (1988)
pRmE43	incP, *nodD1* expressed under control of *trp* promoter	Fisher *et al.* (1988)
pRmJM57	incP, *nodC*::*lacZ* fusion in pRmSL26	Mulligan and Long (1985)
pRmJM61	incP, *nodD1*::*lacZ* fusion in pRmJ30	Mulligan and Long (1985)

[a] The following abbreviations indicate: Str^r, streptomycin resistant; *nod*, nodulation, and *nif*, nitrogen fixing; Ap^r, ampicillin resistant; Tc^r, tetracycline resistant; Nm^r, neomycin resistant; and Sp^r, spectinomycin resistant.
[b] Plasmids based on ColE1 replicon, unless otherwise stated.

Research Laboratories, Gaithersburg, MD, and Promega Biotech, Madison, WI. T4 DNA ligase was obtained from BioRad Laboratories, Richmond, CA. Modified T7 DNA polymerase sequencing kits were obtained from U. S. Biochemical Corp., Cleveland, OH, under the brand name of Sequenase. Radiolabeled nucleotides and the Klenow fragment of *Escherichia coli* DNA polymerase I were obtained from Amersham Corp., Arlington Heights, IL.

Plasmid and strain constructions. pRmS63 (J. A. Swanson, personal communication) is a 12-kb *Kpn*I fragment of pRmBE11 cloned into pUC18, containing the common *nod* region and extending to the *Kpn*I site in pRmJT5 (Table 1; Fig. 1A, B, and E). pZC9 (D. P. Biek, personal communication) is a derivative of pBR322 in which the *Sal*I-*Hin*dIII fragment encoding tetracycline resistance (Tc^r) is replaced by the *Sal*I-*Hin*dIII fragment of Tn*5* encoding neomycin resistance (Nm^r).

The 2.2-kb *Eco*RI fragment of pRmJT5 (Fig. 1A) that appends directly to the left *Eco*RI site of the region for which the DNA sequence was determined previously by Fisher *et al.* (1987: this is the same site as the right-hand *Eco*RI site of Fig. 1B) was subcloned in both orientations in pUC118 to form pRmF30 and pRmF31 (R. F. Fisher, personal communication). These two plasmids were partially digested with *Eco*RI. The 5.4-kb fragments from these digests, corresponding to the cleavage of a single *Eco*RI site, were filled in with T4 DNA polymerase and religated. The plasmids with the filled-in *Eco*RI site at the far end of the clone with respect to the universal sequencing primer hybridization site (i.e., near the *lacZ* promoter) were designated pRmJSS30 and pRmJSS31, respectively (Fig. 1E). These plasmids were used to create nested deletions for single-stranded DNA sequencing.

Because the series of nested deletions derived from pRmJSS30 and pRmJSS31 were not sufficient to complete the DNA sequencing, the following plasmids were constructed. pRmJSS32 was constructed by digesting pRmJSS30 with *Bam*HI and religating. This digest resulted in a 1.8-kb deletion, because there is a *Bam*HI site in the polylinker. Similarly, pRmJSS34 was constructed by digesting pRmJSS30 with *Pst*I and religating, resulting in a 1.48-kb deletion. pRmJSS33 was formed by cutting pRmJSS31 with *Bam*HI and religating, deleting 0.4 kb. pRmJSS35 was constructed by cutting pRmJSS31 with *Bam*HI and *Bgl*II and religating, deleting 1.07 kb, because the cohesive ends of the *Bam*HI site in the polylinker can

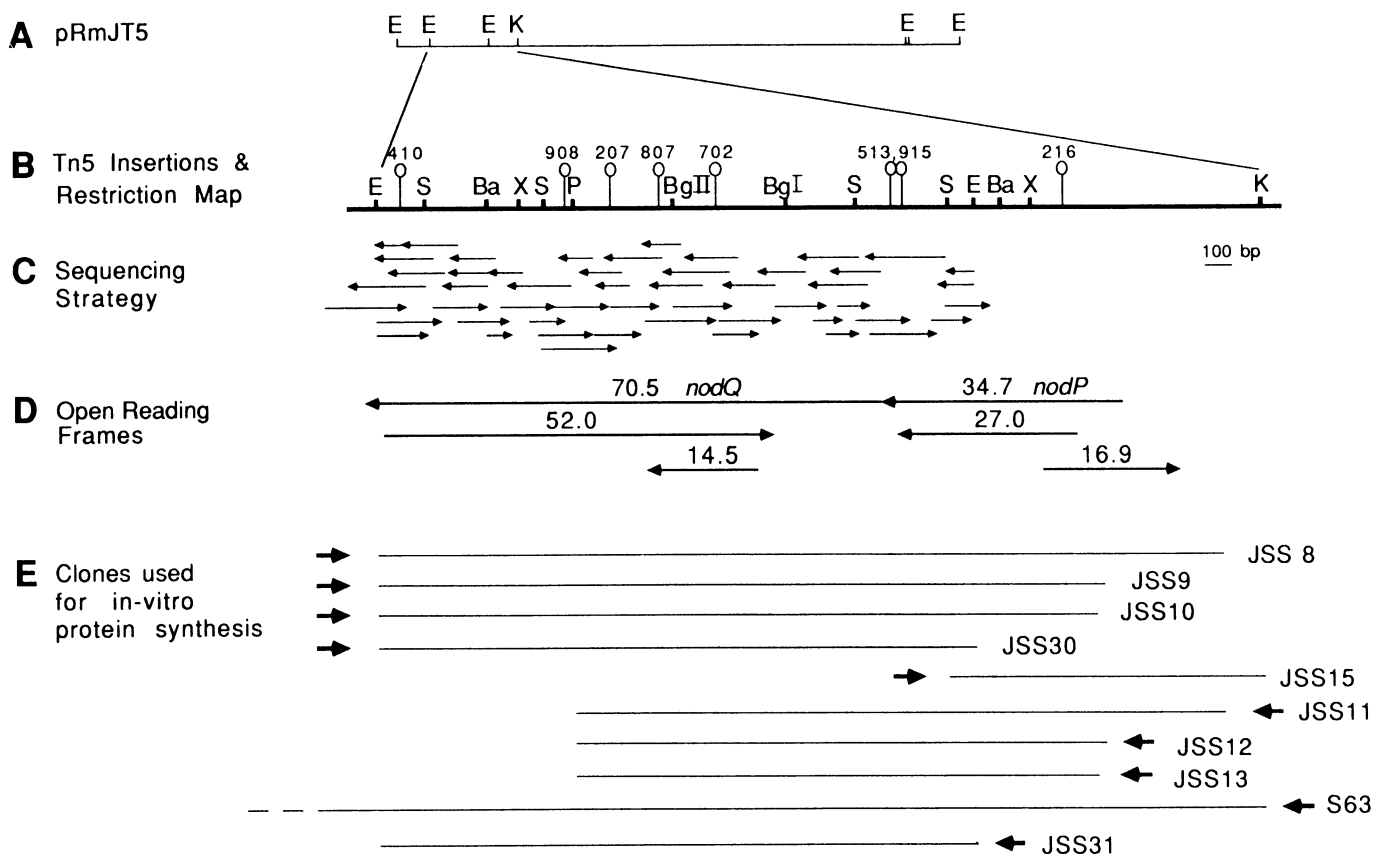

Fig. 1. A, Restriction map of pRmJT5. **B,** Expanded map of the segment used for molecular analysis. (This region represents the rightmost 3.3 kilobase [kb] of pRmS63.) Positions of relevant Tn*5* insertions are shown with accompanying numbers as open circles with lines. Tn*5* 216 has been mapped precisely by sequencing the insertion (Fisher *et al.* 1987; see Fig. 2). The other Tn*5* insertions are mapped to within 200 bp; their relative order is certain, and they are known to be within the 2.2-kb *Eco*RI fragment (Swanson *et al.* 1987). Restriction sites are as follows: E = *Eco*RI, K = *Kpn*I, S = *Sal*I, Ba = *Bam*HI, X = *Xho*I, P = *Pst*I, BgI = *Bgl*I, and BgII = *Bgl*II. **(C)** Arrows indicate the readable sections of individual sequence reactions that were used to define the DNA sequence. **(D)** Arrows indicate putative open-reading frames (ORFs) defined by computer analysis of the nucleotide sequence. Predicted sizes in kilodaltons (kDa) of polypeptides are shown above each ORF. The newly defined genes, *nodP* and *nodQ*, are shown above the corresponding ORFs. **E,** Clones used for protein analysis. (The dashed line at the left of pRmS63 indicates that the insert continues past the *Eco*RI site.) Direction of vector-borne *lac* promoter is shown for each clone by a short, thick arrow. Constructions of clones are described in the text.

anneal to the *Bgl*II cohesive ends. Also, the 330-bp *Sal*I fragment of pRmJSS30 was subcloned into pUC118 in both orientations to create pRmJSS36 and 37.

pRmJSS38 is the 1.2-kb *Sal*I-*Kpn*I fragment of pRmS63 (Fig. 1B and E) subcloned into pUC118, and was used to sequence across the right-hand *Eco*RI site of pRmJSS30, as oriented in Figure 1. A 0.4-kb *Sal*I fragment of pRmS63 was subcloned in both directions in pUC118, creating pRmJSS39 and 40. These were used to sequence across the left-hand *Eco*RI site.

In order to create appropriate clones for *in vitro* transcription-translation analysis, certain DNA sequences had to be reconstructed by ligating three DNA fragments in a directed manner. Nested deletions derived from pRmF32 by Fisher *et al.* (1987) extend only to the *Bam*HI site to the left of the map position of Tn*5* 216. Because the putative open-reading frames (ORFs) span this site and the following *Eco*RI site, some deletion derivatives of pRmF32 and some of the clones described above were ligated to reconstruct the DNA sequence as it occurs in the genome. Three deletion derivatives of pRmF32, numbers 62, 163, and 170 (R. Fisher, personal communication), were digested with *Hin*dIII (the *Hin*dIII site lies in the polylinker) and *Xho*I. The 0.7-, 0.28-, and 0.22-kb fragments of the respective deletions were ligated with the 0.2-kb *Xho*I-*Eco*RI fragment of pRmJSS38 and an *Eco*RI-*Hin*dIII (the *Hin*dIII site is in the polylinker) digest of pRmJSS30, in trimolecular ligations, to create plasmids pRmJSS8-10, respectively (Fig. 1E; Fig. 2).

Plasmids pRmJSS11-13 were created by digesting pRmJSS8-10, respectively, with *Hin*dIII and *Pst*I, and ligating the insert into pUC119, which had been digested with the same enzymes (Fig. 1E).

pRmJSS15 is the *Sal*I-*Kpn*I fragment of pRmJSS38 cloned into pUC119, such that the *lacZ* promoter is in the opposite orientation (Fig. 1E).

Translational *lacZ* fusions were constructed in the two putative ORFs of *nodP* in the following manner. The fusion to the larger ORF (predicted size 34.7 kDa, Fig. 1D) was constructed by digesting pRmJSS38 with *Xho*I and partially filling in the cohesive ends with 2′-deoxyribocytidine triphosphate (dCTP) and 2′-deoxyribothymidine triphosphate (dTTP). This was ligated to the 9.6-kb *Bam*HI fragment of pMC931Sp with its cohesive ends partially filled in with 2′-deoxyriboguanosine triphosphate (dGTP) and 2′-deoxyriboadenosine triphosphate (dATP), creating pRmJSS16. The GAT sequence of the *Bam*HI site is in frame with *lacZ* on the 9.6-kb fragment, which also contains a spectinomycin-resistance (Sp^r) gene. The fusion to the smaller ORF (predicted size 27 kDa, Fig. 1D) was created in a similar manner, except that completely filled-in *Xho*I sites were ligated to completely filled *Bam*HI sites, regenerating *Xho*I sites, and creating pRmJSS17.

Because pUC118 derivatives cannot be conjugationally mobilized, it was necessary for genetic experiments to put the inserts of pRmJSS16 and 17 into a mobilizable plasmid. pZC9 (D. Biek, personal communication), an ampicillin-resistant (Ap^r) and Nm^r derivative of pBR322, was chosen for this purpose. pRmJSS16 and 17 were digested with *Kpn*I and blunted with S1 nuclease, followed by a treatment with the Klenow fragment of DNA polymerase. The plasmids were then digested with *Sal*I. This mixture was ligated with pZC9, which had been digested with *Sma*I and *Sal*I. Plasmids that were Ap^r, Nm^r, and Sp^r were selected, checked by restriction digests, and named pRmJSS16b and pRmJSS17b.

Plasmids pRmJSS16b and pRmJSS17b were mobilized into Rm1021 via triparental mating, as previously described (Ditta *et al.* 1980). Because pBR322 derivatives cannot replicate in *R. meliloti*, integration of the plasmid into the homologous region of the chromosome is the only means of marker rescue. All Sp^r transconjugants were also Nm^r, indicating a single crossover event. The Ap^r gene is not an effective marker in *R. meliloti* (unpublished observations). These strains were verified by Southern blot hybridization analysis of the genome as described below. The strains containing an integrated copy of pRmJSS16b or pRmJSS17b were named JSS5 and JSS6, respectively.

DNA sequencing. Most of the DNA sequencing was carried out by the dideoxy chain termination technique of Sanger *et al.* (1977), in vectors pUC118 and pUC119. Some of the sequencing was done using a modification of this technique (Tabor and Richardson 1987), using a modified T7 DNA polymerase instead of the Klenow fragment of *E. coli* DNA polymerase I. Single-stranded DNA was produced by the method of Vieira and Messing (1987) and isolated as for M13 preparations, which is described in a handbook for M13 cloning and sequencing (Amersham Corp.) A series of nested deletions was created from pRmJSS30-35 by exonuclease III digestion using the procedure of Henikoff (1984). To sequence regions not covered by the exonuclease III deletions, plasmids pRmJSS36 and pRmJSS37 were analyzed. We sequenced across the right-hand *Eco*RI site as shown in Figure 1B using pRmJSS38, a subclone of pRmS63 in pUC118. We sequenced across the left-hand *Eco*RI site in both directions with pRmJSS39 and 40. Overlapping nested deletions were organized, and DNA sequence analysis was conducted using Seqsort, AA (Amino Acid), and RE (Restriction Enzyme) programs as previously described (Egelhoff *et al.* 1985).

Protein products. DNA segments were cloned in the expression plasmids pUC118, pUC119, and pUC18. Plasmids (1 μg) purified by CsCl banding were incubated with coupled transcription-translation extracts from *E. coli* HB101 or *R. meliloti* RCR2011, using modifications of the protocol of Gunsalus *et al.* (1979), as previously described (Fisher *et al.* 1987). Protein products were separated by PAGE (Laemmli 1970) and visualized by autoradiography.

Filter hybridizations. Total bacterial DNA was prepared as previously described by a modification of the method of Meade *et al.* (1982). DNA (3 μg per lane) was digested with restriction enzymes and separated by electrophoresis in a 0.8% agarose gel. DNA was transferred to Genescreen nylon membranes (Du Pont, Wilmington, DE) according to a modification of the technique of Reed and Mann (1985) as described by Rigaud *et al.* (1987). DNA probes were labeled with the random hexamer primer technique (Feinberg and Vogelstein 1983). Filters were hybridized at 65° C in 5× SSC, 50 mM Na phosphate pH 6.8, 5× Denhardt's, 0.25% sodium dodecyl sulfate (SDS), and denatured salmon sperm DNA at 1 mg/ml. Filters were washed four times in 1× SSC, 0.25% SDS at room temperature (22° C) after hybridization.

Computer methods. Amino acid sequences were compared to sequences in the Genetics Computer Group (GCG) GenBank using the Wordsearch and Bestfit programs on a VAX. Putative RNA secondary structures were analyzed with the programs Fold, Stemloop, Squiggles, and Circles. Direct repeats were found with the program Repeat (Devereux *et al.* 1984).

β-Galactosidase assays. The assays were carried out using a modified technique of Miller (1972) as described by Mulligan and Long (1985).

Nodulation tests. Seeds were sterilized, planted on agar slants, and inoculated, as described by Jacobs *et al.* (1985). Plants were observed at 3-day intervals and scored for the number of nodulated plants and number of nodules on each plant.

```
                  15                  30                  45                  60
TGG CGA TCC CGA TCC ACC GCA TGG GCA CCG GTA CCG AAG TCG CGC TCC GCC GTT GCG TAT
                  75                  90                 105                 120
CTC GCT TCC GAT CAC GCC GCC TAT GTC ACC GGA CAG ACC ATT CAC GTG AAC GGC GGT ATG
                 135                 150                 165                 180
GCA ATG ATT TGA AGG CGG TCG GGC CTA CGG ATG AGT GGG CTT GCA TTT GCA TAC GCC AGC
                 195                 210                 225                 240
CTA TCA GCG CAA TGA TGA TAA CGG CAT AAA GGC CAT TGC ACT TTC CGA AAG CTG AGG AAG
        8,11▼    255                 270                 285                 300
CAA GCC ATT ATG GAT AGT GCA CCT GTC AGC AAT ACT GAA CGG TCT CAA CGG AAT AGC CTG
                 315                 330                 345                 360
CGA TTG AGC GCT CCG GTC CCA GCA GCA ATA GCT CGG CCC CAT ATG AAG ACG CTG TCT CGC
                 375                 390                 405                 420
TCG GCG CCG GCG CAT CAG CGC GGA ACG TCA GAT AGC GCA AAC GCT TTA GTG CGG CGT TGC
                 435                 450                 465                 480
TTA GCG CCA TTA CGT CGC GCC ACC GTC TTG CCG CGG TGA TCC CAC GCA TTG GGA TGC CTT
                 495                 510                 525                 540
GAG CGA GCT GAG CTG CCG AGG CGT AAC CCG GAT AGG TTT CCT GAA CAT AGA ACA AGG CCA
                 555                 570                 585                 600
CAA ATG TCT CTT CCC CAT CTT CGG CGG CTT GAA GCC GAA GCG ATC CAT GTC ATT CGA GAA
    MET Ser Leu Pro His Leu Arg Arg Leu Glu Ala Glu Ala Ile His Val Ile Arg Glu
             9,12▼               630                 645                 660
GTT GTT GCG ACA TTC TCC AAT CCG GTC GTG CTT TAC TCG ATC GGC AAA GAC TCC TCG GTA
Val Val Ala Thr Phe Ser Asn Pro Val Val Leu Tyr Ser Ile Gly Lys Asp Ser Ser Val
                 675        10,13▼   690                 705                 720
CTG CTG CAC CTG GCG ATG AAG GCG TTC TAC CCC GCC AAG CCG CCA TTT CCA TTC CTG CAT
Leu Leu His Leu Ala MET Lys Ala Phe Tyr Pro Ala Lys Pro Pro Phe Pro Phe Leu His
                 735                 750                 765                 780
GTA GAT ACC AAA TGG AAG TTC CGG GAG ATG ATC GAG TTT CGC GAC CGG ATG GCG CGA GAG
Val Asp Thr Lys Trp Lys Phe Arg Glu MET Ile Glu Phe Arg Asp Arg MET Ala Arg Glu
                 795                 810 ◆Tn5#216        825                 840
CTC GGC TTC GAT CTC CTC GTC CAC GTC AAT CAG GAC GGG GTC GAG CAG GGC ATC GGG CCA
Leu Gly Phe Asp Leu Leu Val His Val Asn Gln Asp Gly Val Glu Gln Gly Ile Gly Pro
                 855                 870                 885                 900
TTC ACG CAC GGT TCC AAC GTG CAC ACC CAT GTC ATG AAG ACG ATG GGG CTC CGG CAG GCG
Phe Thr His Gly Ser Asn Val His Thr His Val MET Lys Thr MET Gly Leu Arg Gln Ala
                 915                 930                 945                 960
CTC GAG AAA TAC GGT TTC GAC GCG GCG CTC GCA GGC GCG CGG CGC GAC GAG GAG AAG TCG
Leu Glu Lys Tyr Gly Phe Asp Ala Ala Leu Ala Gly Ala Arg Arg Asp Glu Glu Lys Ser
                 975                 990                1005                1020
CGC GCC AAG GAA CGC ATC TTC TCG ATT CGC AGC GCC CAG CAC GGC TGG GAT CCG CAG CGC
Arg Ala Lys Glu Arg Ile Phe Ser Ile Arg Ser Ala Gln His Gly Trp Asp Pro Gln Arg
                1035                1050                1065                1080
CAG CGG CCC GAG ATG TGG AAG ACT TAC AAT ACG CGG GTC GGA CAA GGC GAG ACG ATG CGA
Gln Arg Pro Glu MET Trp Lys Thr Tyr Asn Thr Arg Val Gly Gln Gly Glu Thr MET Arg
                1095                1110                1125                1140
GTC TTC CCG CTT TCC AAC TGG ACC GAA TTC GAC ATC TGG CAG TAC ATC CTG CGC GAG GAA
Val Phe Pro Leu Ser Asn Trp Thr Glu Phe Asp Ile Trp Gln Tyr Ile Leu Arg Glu Glu
                1155                1170                1185                1200
ATT CCG ATC GTG CCG CTT TAT TTC GCG GCC AGG CGC CCG GTC GTC AAG CGA GAG GGT ATG
Ile Pro Ile Val Pro Leu Tyr Phe Ala Ala Arg Arg Pro Val Val Lys Arg Glu Gly MET
                1215                1230                1245                1260
CTG ATC ATG GTC GAC GAC GAC CGC ATG CCC ATC CAA CCC GAA GAG GAG GTT ACC GAA CAG
Leu Ile MET Val Asp Asp Asp Arg MET Pro Ile Gln Pro Glu Glu Glu Val Thr Glu Gln
                1275                1290                1305                1320
CTC GTG CGT TTC CGC ACG CTT GGC TGC TAT CCG CTG ACC GGG GCG GTC GAG TCC GAC GCT
Leu Val Arg Phe Arg Thr Leu Gly Cys Tyr Pro Leu Thr Gly Ala Val Glu Ser Asp Ala
```

Fig. 2. DNA sequence of *nodP* and *nodQ*. The sequence from *Kpn*I to the rightmost *Eco*RI site in Figure 1B was previously reported by Fisher *et al.* (1987). The deduced amino acid sequences for *nodP* and *nodQ* are shown aligned with the nucleotide sequence. The breakpoints of the inserts in pUC118 to form pRmJSS8-10 are indicated with an arrowhead (▼). These are the same as the breakpoints for pRmJSS11-13, respectively, in pUC119. The position of Tn*5* 216 is indicated with a diamond (◆).

RESULTS

DNA sequence and ORFs defining *nodP* and *nodQ*. Tn*5* insertion analysis of the clone pRmJT5 (Fig. 1A) revealed the presence of genes involved in nodulation. Some of the strains containing homogenotized Tn*5* insertions had phenotypes of delayed nodulation on alfalfa (Swanson *et al.* 1987). The restriction map and positions of Tn*5* insertions in the DNA fragment containing the newly defined genes, *nodP* and *nodQ*, are presented in Figure 1B. Two regions have mutant phenotypes: strains containing Tn*5* insertions 216, 702, and 807 exhibit phenotypes of delayed nodulation on alfalfa. Flanking insertions 513 and 915, and 207, 908, and 410 are indistinguishable from wild type (Swanson

```
                1335                1350                1365                1380
GTC ACC GTT CCC GAG ATA TTG CGG GAA ATG CTG ACG GTG CGC ACG TCC GAA CGG CAG AGC
Val Thr Val Pro Glu Ile Leu Arg Glu MET Leu Thr Val Arg Thr Ser Glu Arg Gln Ser
                1395                1410                1425                1440
CGG CTG ATC GAC ACG GAT GAA GTC GGG GCG ATG GAG AAA AAG AAG CGG GAG GGC TAC TTC
Arg Leu Ile Asp Thr Asp Glu Val Gly Ala MET Glu Lys Lys Lys Arg Glu Gly Tyr Phe
                1455                1470                1485                1500
TGA TGT CGT ATG TTC AAT CTA TAC CGC CGC ATG ACA TTG AAG CGC ATC TGG CCG AGC ACG

  MET Ser Tyr Val Gln Ser Ile Pro Pro His Asp Ile Glu Ala His Leu Ala Glu His Asp
                1515                1530                1545                1560
ACA ACA AGT CGA TCC TGA GAT TCA TCA CTT GCG GCT CGG TCG ACG ACG GCA AAT CGA CCC
  Asn Lys Ser Ile Leu Arg Phe Ile Thr Cys Gly Ser Val Asp Asp Gly Lys Ser Thr Leu
                1575                1590                1605                1620
TGA TCG GGC GAC TGC TTT ACG ATG CGA AGC TGG TCT TCG AAG ACC AGC TCG CAA ACC TCG
  Ile Gly Arg Leu Leu Tyr Asp Ala Lys Leu Val Phe Glu Asp Gln Leu Ala Asn Leu Gly
                1635                1650                1665                1680
GGC GTG TCG GCT CTC CCG GCG CCG CCA ACG GCA AGG AGA TCG ATC TCG CCT TGC TTC TCG
  Arg Val Gly Ser Pro Gly Ala Ala Asn Gly Lys Glu Ile Asp Leu Ala Leu Leu Leu Asp
                1695                1710                1725                1740
ACG GGC TTG AGG CCG AGC GCG AGC AGG GCA TCA CCA TCG ACG TCG CCT ATC GCT ATT TCG
  Gly Leu Glu Ala Glu Arg Glu Gln Gly Ile Thr Ile Asp Val Ala Tyr Arg Tyr Phe Ala
                1755                1770                1785                1800
CCA CGT CCA AAC GCA AGT TCA TCG TCG CCG ATA CGC CTG GCC ACG AGG AAT ATA CGC GCA
  Thr Ser Lys Arg Lys Phe Ile Val Ala Asp Thr Pro Gly His Glu Glu Tyr Thr Arg Asn
                1815                1830                1845                1860
ACA TGG TGA CCG GCG CTT CGA CGG CGG ATC TCG CCA TCA TCC TCA TCG ACA GCC GGC AGG
  MET Val Thr Gly Ala Ser Thr Ala Asp Leu Ala Ile Ile Leu Ile Asp Ser Arg Gln Gly
                1875                1890                1905                1920
GCA TTC TCC AGC AGA CCC GGC GCC ACT CCT ATA TAG CCT CAC TCC TCG GCA TCC GCC ATG
  Ile Leu Gln Gln Thr Arg Arg His Ser Tyr Ile Ala Ser Leu Leu Gly Ile Arg His Val
                1935                1950                1965                1980
TCG TGC TGG CCG TCA ACA AGA TCG ATC TCG TCG ATT TTA AAC AAC AGG TGT ACG AGG AAA
  Val Leu Ala Val Asn Lys Ile Asp Leu Val Asp Phe Lys Gln Gln Val Tyr Glu Glu Ile
                1995                2010                2025                2040
TCG TCG CCG ACT ACA TGG CTT TCG CCA AAG AGC TCG GTT TCG CCA GCA TAC GGC CAA TCC
  Val Ala Asp Tyr MET Ala Phe Ala Lys Glu Leu Gly Phe Ala Ser Ile Arg Pro Ile Pro
                2055                2070                2085                2100
CGA TCT CGG CGC GAG ACG GCG ACA ACG TCA TCT CGG CTT CCG CCA ATA CCC CCT GGT ACA
  Ile Ser Ala Arg Asp Gly Asp Asn Val Ile Ser Ala Ser Ala Asn Thr Pro Trp Tyr Arg
                2115                2130                2145                2160
GAG GGG CGG CGC TGC TCG AAT ATC TGG AAA CGG TCG AAC TCG ATC CGA CGG ACC AGG CAA
  Gly Ala Ala Leu Leu Glu Tyr Leu Glu Thr Val Glu Leu Asp Pro Thr Asp Gln Ala Lys
                2175                2190                2205                2220
AGC CTT TCC GCT TCC CGG TTC AGA TGG TCA TGC GGC CGA ACG CGG ATT TTC GCG GCT ATG
  Pro Phe Arg Phe Pro Val Gln MET Val MET Arg Pro Asn Ala Asp Phe Arg Gly Tyr Ala
                2235                2250                2265                2280
CCG GGC AGA TCT CCT GCG GGA GGA TTT CCG TGG GCG ATC CGG TCG TCG TCG CGA AGA CCG
  Gly Gln Ile Ser Cys Gly Arg Ile Ser Val Gly Asp Pro Val Val Val Ala Lys Thr Gly
                2295                2310                2325                2340
GGC AGC GGA CAT CGG TCA AGG CGA TCG TGA CCT ATG ACG GGG AGC TTG CGA CGG CAG GGG
  Gln Arg Thr Ser Val Lys Ala Ile Val Thr Tyr Asp Gly Glu Leu Ala Thr Ala Gly Glu
                2355                2370                2385                2400
AAG GCG AAG CGG TGA CGC TGG TCC TCT CTG ACG AGG TGG ATG CGT CTC GCG GCA ATA TGC
  Gly Glu Ala Val Thr Leu Val Leu Ser Asp Glu Val Asp Ala Ser Arg Gly Asn MET Leu
                2415                2430                2445                2460
TCG TGG CCC CTG GTG CCC GGC CCT TCG TGG CGG ACC AGT TCC AGG CGC ATG TGA TCT GGT
  Val Ala Pro Gly Ala Arg Pro Phe Val Ala Asp Gln Phe Gln Ala His Val Ile Trp Phe
                2475                2490                2505                2520
TCG ATG CGA ACC CGA TGA TGC CGG GAC GAA GCT ATA TCC TGC GCA CGG AGA CCG ACA GCG
  Asp Ala Asn Pro MET MET Pro Gly Arg Ser Tyr Ile Leu Arg Thr Glu Thr Asp Ser Val
```

Fig. 2 *continued.*

et al. 1987).

The overlapping sequencing reactions used to construct the DNA sequence are shown in Figure 1C. The resulting DNA sequence and computer analysis of ORFs defining *nodP* and *nodQ* are shown in Figure 2. The region to the right of the right-hand *Eco*RI site (upstream and 5′ section of *nodP*), previously sequenced by Fisher *et al.* (1987), is included here for clarity. A computer search of the nucleotide sequence for putative RNA secondary structures and direct repeats revealed nothing suggestive of the possible biological roles of *nodP* and *nodQ*.

A number of putative ORFs defined by methionine initiation codons and transcription termination codons are present. All of those larger than 75 amino acids are shown in Figure 1D. The termination codon of the 34.7-kDa ORF overlaps the start codon of the 70.5-kDa ORF (Fig. 2). There are 10 ORFs in this region between 50 and 75 amino acids long (data not shown).

There are three putative ORFs that are interrupted by Tn*5* insertion 216, and another three interrupted by insertions 702 and 807. In order to determine which of the ORFs are expressed, we constructed a number of protein expression clones (Fig. 1E). We performed *in vitro* transcription-translation assays on these clones with both *E. coli* and *R. meliloti* S-30 extracts. The exact positions of the right-hand endpoints of the insert for clones pRmJSS8-13 are indicated in the DNA sequence (Fig. 2).

We used both full-length and truncated clones to show that an expressed gene, *nodP*, corresponds to the 34.7-kDa ORF, not the smaller 27-kDa ORF. Clones containing both ORFs (pRmJSS8 and pRmJSS11) expressed the 34.7-kDa gene product (Fig. 3A, lanes 2 and 5); those containing only the 27-kDa ORF (pRmJSS9, 10, 12, and 13) showed no product (Fig. 3A, lanes 3, 4, 6, and 7).

Analysis of protein products points to the identity of the 70.5-kDa ORF as a second gene, *nodQ*. *In vitro* transcription-translation assays of pRmJSS8-10 produce a protein of approximately 70 kDa in *R. meliloti* extracts

```
              2535                2550                2565                2580
TCA GCG CGA CGG TCA CCA CGC TCA AGC ACC AGG TCA ACA TCA ACA GCT TCA TCC GTG AGG
  Ser Ala Thr Val Thr Thr Leu Lys His Gln Val Asn Ile Asn Ser Phe Ile Arg Glu Ala
              2595                2610                2625                2640
CGG CGA AGT CGC TGC AGA TGA ACG AAG TGG GTG TCT GCA ACA TCT CGA CAC AGG CGC CGA
  Ala Lys Ser Leu Gln MET Asn Glu Val Gly Val Cys Asn Ile Ser Thr Gln Ala Pro Ile
              2655                2670                2685                2700
TTG CCT TCG ACG CCT ACA ATG ACA ACC GGG CGA CGG GCA ATT TCA TCA TCG TCG ACC GGG
  Ala Phe Asp Ala Tyr Asn Asp Asn Arg Ala Thr Gly Asn Phe Ile Ile Val Asp Arg Val
              2715                2730                2745                2760
TGA CGA ATG CCA CGG TTG GTG CGG GGT TGA TCG ATT TTC CGC TCC GGC GCG CAG ACA ACG
  Thr Asn Ala Thr Val Gly Ala Gly Leu Ile Asp Phe Pro Leu Arg Arg Ala Asp Asn Val
              2775                2790                2805                2820
TCC ACT GGC ATG CGC TCG AGG TGA ACA AGA GCG CGC GTA GCG CCA TGA AAA ATC AGC TCC
  His Trp His Ala Leu Glu Val Asn Lys Ser Ala Arg Ser Ala MET Lys Asn Gln Leu Pro
              2835                2850                2865                2880
CTG CCG TTC TCT GGT TCA CCG GGC TTT CCG GCT CCG GAA AAT CGA CCA TCG CGA ACG AGC
  Ala Val Leu Trp Phe Thr Gly Leu Ser Gly Ser Gly Lys Ser Thr Ile Ala Asn Glu Leu
              2895                2910                2925                2940
TCG ACA GGA TCC TCC ACG CTC AGG GCA AGC ATA CTT ACC TGC TCG ACG GCG ACA ATG TGC
  Asp Arg Ile Leu His Ala Gln Gly Lys His Thr Tyr Leu Leu Asp Gly Asp Asn Val Arg
              2955                2970                2985                3000
GTC ACG GCC TCA ACC GGG ACC TCG GCT TTA CCG AGG AGG ACC GGG TAG AGA ACA TCC GCC
  His Gly Leu Asn Arg Asp Leu Gly Phe Thr Glu Glu Asp Arg Val Glu Asn Ile Arg Arg
              3015                3030                3045                3060
GCG TGG CGG AGG TGG CCA AGC TCA TGG CCG ATG CCG GTC TGA TCG TTC TCG TCT CCT TCA
  Val Ala Glu Val Ala Lys Leu MET Ala Asp Ala Gly Leu Ile Val Leu Val Ser Phe Ile
              3075                3090                3105                3120
TCT CGC CGT TCC GCG ACG AGC GGC GGA TGG CGC GGG AAT TGA TGG AGG AGG GCG AGT TCA
  Ser Pro Phe Arg Asp Glu Arg Arg MET Ala Arg Glu Leu MET Glu Glu Gly Glu Phe Ile
              3135                3150                3165                3180
TCG AGA TAT TCG TCG ACA CGC CGC TCG ACG AGT GCG CGC GCC GCG ATC CGA AGG GGC TCT
  Glu Ile Phe Val Asp Thr Pro Leu Asp Glu Cys Ala Arg Arg Asp Pro Lys Gly Leu Tyr
              3195                3210                3225                3240
ACG AGA AGG CGC TCG CCG GCA AGA TCG CGA ACT TCA CCG GCG TAT CCT CGT GCT ATG AGG
  Glu Lys Ala Leu Ala Gly Lys Ile Ala Asn Phe Thr Gly Val Ser Ser Cys Tyr Glu Ala
              3255                3270                3285                3300
CCC CGG AAA ATC CGG AAC TCC ATA TAC GCA CCG TCG GCC ATC AAC CGA ACG ACC TGG CGC
  Pro Glu Asn Pro Glu Leu His Ile Arg Thr Val Gly His Gln Pro Asn Asp Leu Ala Leu
              3315                3330                3345                3360
TCG CGA TCG AGG AAT TCC TTG ACC GCA GGA TTG GAG GAC AAA TGA CGC CGC TTC AAC GCC
  Ala Ile Glu Glu Phe Leu Asp Arg Arg Ile Gly Gly Gln MET Thr Pro Leu Gln Arg Pro
              3375                3390                3405                3420
CAA CAT AGA GAC GGA CTG TAC GCA TCC GAG GAG GGC CGC CTT GCT GAG AAG TGA GGC GAA
  Thr  .
              3435                3450                3465
CTG GAT GTT CTG TCT ATA TGG TCC CGG AAC TAG CCA TTC CGC TTG TTG CGC CGA C
```

Fig. 2 *continued.*

(Fig. 3A, lanes 2–4), consistent with its origin as a truncation-fusion product of the 70.5-kDa predicted ORF, which has the last 17 amino acids replaced by 29 amino acids from the vector DNA sequence. The 45-kDa protein produced from clones pRmJSS11-13 (Fig. 3, lanes 5–7) is probably a truncation-fusion product of the 70.5-kDa ORF that is truncated at the *Pst*I site. It is not consistent with expression of the 52-kDa ORF because the putative translation start site, as well as half of its coding sequence is missing.

The fact that the 52-kDa ORF was not seen to produce a product in *E. coli* extracts even if oriented downstream of the *lacZ* promoter (for example, pRmJSS30; see discussion of promoters below) also reinforces the conclusion that the 70-kDa ORF, not the 52-kDa ORF, represents an expressed gene. pRmJSS30 (Fig. 3, lanes 10) expresses the same truncation product as pRmJSS8-10 in *R. meliloti* S-30 extracts, and pRmJSS31 expresses a truncation product of a similar size. Because the insert is in the opposite orientation with respect to the pUC118 vector, a slightly different product is formed due to different read-through of vector sequences. However, in *E. coli* extracts only pRmJSS31 expresses a truncation product, the faint upper band of a doublet in the overexposed lane 11 of Figure 3B, consistent with the orientation of the *lacZ* promoter and the 70.5-kDa ORF, and inconsistent with the orientation of the 52-kDa ORF.

Comparing expression of clones using *E. coli* and *R. meliloti* S-30s suggested that *nodP* and *nodQ* can be expressed from endogenous promoters recognized only by *Rhizobium* extracts. We have previously observed that the transcription-translation products from *E. coli* extracts do not always exhibit the same protein profiles as the products from *R. meliloti* extracts (data not shown), probably because the two organisms recognize promoters differently (Yelton *et al.* 1987; Fisher *et al.* 1987). The *E. coli* extract recognizes the *E. coli lacZ* promoter in the presence of 0.5 mM cAMP; the *R. meliloti* extract recognizes the *lacZ* promoter somewhat more weakly and is unaffected by cAMP (Fisher *et al.* 1987).

Many of the clones used in this study contain inserts bearing promoters that the *R. meliloti* extract recognizes. pRmJSS8-10 generate no detectable insert-specific protein products in *E. coli* extracts (Fig. 3B, lanes 2–4 and 12). However, when the inserts are in the opposite orientation relative to the *lacZ* promoter, insert-specific proteins are produced by *E. coli* extracts: pRmJSS11 expresses protein products of 40 and 45 kDa (Fig. 3B, lane 5), and pRmJSS12 and pRmJSS13 express a 45-kDa protein product (Fig. 3B, lanes 6 and 7). Thus, the expression of the *R. meliloti* gene products in *E. coli* S-30 extracts depends on the vector promoter. By contrast, all six of these clones express proteins when *R. meliloti* extracts are used (Fig. 3A, lanes 2–7). Even in the absence of 5′ *nodP* sequences, *nodQ* truncation products are weakly expressed in *R. meliloti* extracts irrespective of the orientation of the *lacZ* promoter (Fig. 3A, 70-kDa peptide in lanes 3, 4, and 10; 45-kDa peptide in lanes 6 and 7). In other assays we found no evidence for cryptic vector promoters that could account for such expression (data not shown). This suggests that there is an internal promoter within the *nodP* coding sequence.

Because none of the clones discussed thus far contains the complete *nodQ* gene, it was of interest to express the intact protein to show that its size is consistent with that deduced from the DNA sequence. pRmS63 expresses the full-length products of *nodP* and *nodQ*, of approximately

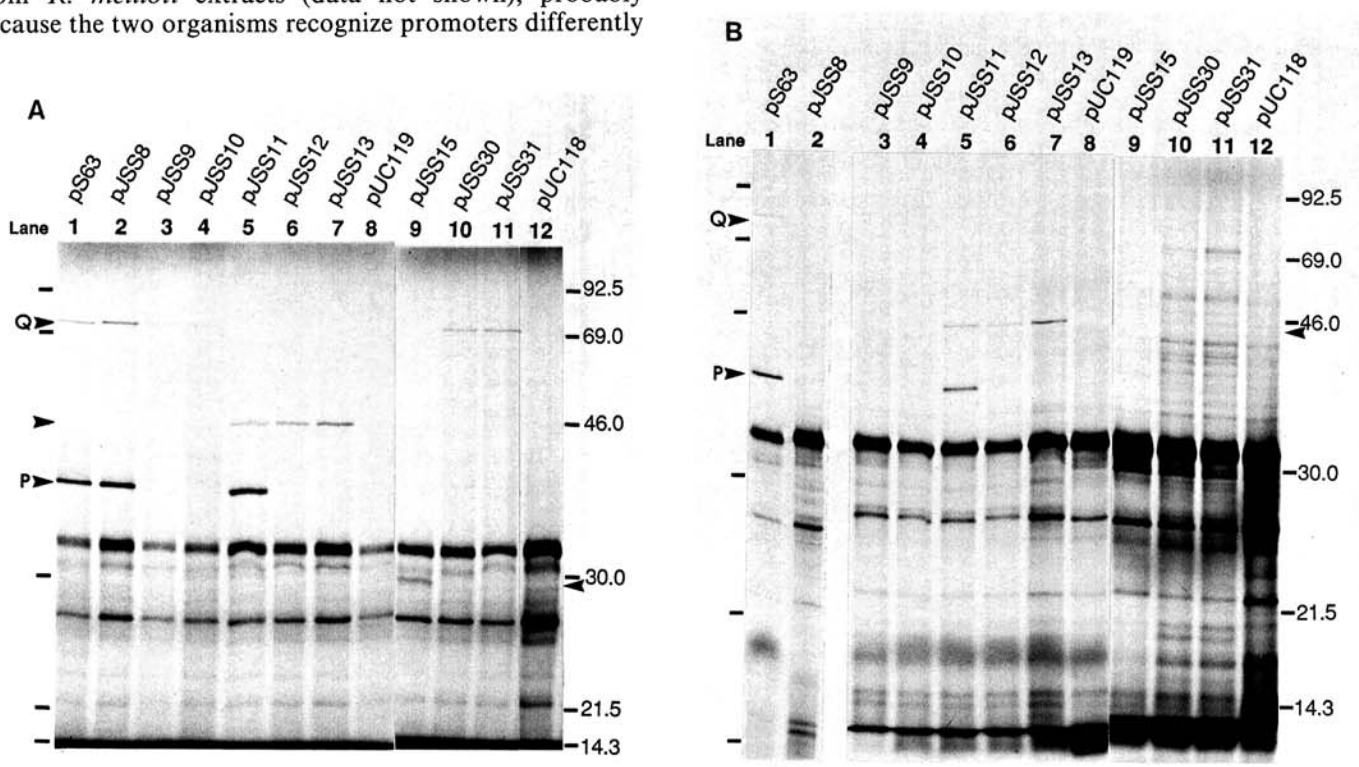

Fig. 3. Coupled transcription-translation assays define polypeptides encoded by the sequenced DNA segment. S-30 extracts used to produce polypeptides were from **(A)** *Rhizobium meliloti* and **(B)** *Escherichia coli*. The plasmid for each reaction is indicated above the individual lanes. Molecular weight standards are indicated on the sides of the gel. NodP and NodQ are indicated with an arrowhead, as are truncation products (see text).

40 and 70 kDa, respectively (Fig. 3, lanes 1). Although this clone contains the *nodDABC* region, these genes are apparently too far from a strong promoter to be detected. The full-length product of *nodQ* is barely distinguishable from the truncation-fusion product of pRmJSS30 because the protein products differ in length by only 12 amino acids.

By several tests, other possible ORFs do not correspond to genes. Expression of the 16.9-kDa ORF was tested; pRmJSS15 (Fig. 3, lanes 9) expresses no insert-specific products in *E. coli* extracts, but produces a faint 28-kDa band in the *R. meliloti* extracts, consistent with a truncation-fusion product of the 34.7-kDa ORF. We have no evidence that the 16.9-kDa ORF or 14.5-kDa ORF is expressed.

To verify the conclusions about which ORF corresponds to *nodP*, we performed *in vivo* assays of gene expression. We fused the *E. coli lacZ* gene in frame to the 34.7-kDa ORF and independently to the 27-kDa ORF and mated each construction into Rm1021. The strain containing the 34.7-kDa ORF fusion, JSS5, has approximately 9 units of β-galactosidase activity; the strain containing the fusion to the 27-kDa ORF, JSS6, has 3 units of activity. Rm1021 without insertions has a basal level of 2 units of activity. The activities were independent of luteolin addition except when *nodD* was overexpressed on a plasmid (see results below and Table 2).

The evidence above is consistent with the expression of the 34.7- and 70.5-kDa ORFs, and these have been named *nodP* and *nodQ*, respectively. It is unexpected that some of the strains carrying Tn*5* insertions in the ORFs of *nodP* and *nodQ* display no aberrant nodulation phenotype on alfalfa. However, it is interesting to note that the Tn*5* insertions in the carboxy terminal halves of each of the genes (in *nodP*, insertions 513 and 915; in *nodQ*, 207, 908, and 410) are those with wild-type nodulation phenotypes. These insertions may produce partially active truncation products. Tn*5* is known to contain promoters recognized by *R. meliloti* that read outward from the insertion endpoints (Corbin *et al.* 1983), so insertions 513 and 915 may not interfere with the functional expression of *nodQ*. The structure of NodQ is interesting in that there is a region of homology to the guanosine diphosphate (GDP) binding domain of elongation factor Tu (EF-Tu) that occurs in the amino terminal half of *nodQ* (see results and discussion below). Nonmutating insertions 207, 908, and 410 are all downstream of this region. Besides the possible effect of mutation position, the existence of homologous copies of the genes elsewhere in the genome may provide functions that restore or partially restore phenotype (see results below).

Table 2. β-Galactosidase assays of *nod* gene-*lacZ* fusions

Strains	Relevant characteristics[a]	Enzyme units[b] +	−
Rm1021	Wild type	2.3	2.0
JM57	*nodC*::*lacZ*	16.3	8.4
JSS5[c]	*nodP*::*lacZ*	9.4	7.6
JSS6[d]	Out-of-frame *nodP*::*lacZ*	3.0	2.8
JSS5/pRmJ30	*nodP*::*lacZ* / *nodD1ABC* on 8.7-kb insert	43.9	7.9
JSS5/pRmSL26	*nodP*::*lacZ* / *nodD1ABC* on 19.5-kb insert	38.8	9.2
JSS5/pRmE65	*nodP*::*lacZ* / *ptrp*::*nodD3*	50.3	44.7
JSS5/pRmE43	*nodP*::*lacZ* / *ptrp*::*nodD1*	79.4	10.7
Rm1021/pRmJM61	Wild type / *nodD1*::*lacZ*	285.5	280.5
JT216/pRmJM61	*nodP*::Tn*5* / *nodD1*::*lacZ*	461.0	430.0
JT702/pRmJM61	*nodQ*::Tn*5* / *nodD1*::*lacZ*	419.5	373.0
Rm1021/pRmJM57	Wild type / *nodD1ABC*::*lacZ*	608.0	31.0
JT216/pRmJM57	*nodP*::Tn*5* / *nodD1ABC*::*lacZ*	723.5	31.0
JT702/pRmJM57	*nodQ*::Tn*5* / *nodD1ABC*::*lacZ*	863.5	33.0

[a] Characteristics before slash indicate those in the genome; those after the slash are on the plasmid.

[b] Cells were assayed according to Miller (1972) during log phase growth in tryptone-yeast extract (TY) medium. The presence or absence of 3 μm luteolin is indicated by + or −, respectively.

[c,d] The units of these strains represent six independent assays, each performed in duplicate. All others represent two independent assays, each performed in duplicate.

Expression of *nodP*::*lacZ* fusion. We tested the expression of β-galactosidase fusions in the presence or absence of different plasmids and luteolin. The results are summarized in Table 2. The strain carrying the *nodP*::*lacZ* fusion, JSS5, shows little response to the addition of luteolin. In contrast, JM57, the strain carrying the *nodC*::*lacZ* fusion, shows a twofold induction of activity. However, the addition of a plasmid carrying *nodD1* causes the activity of the *nodP*::*lacZ* fusion to increase fourfold in the presence of luteolin, and the presence of overexpressed *nodD3* causes a four to fivefold increase in activity that is independent of luteolin addition. Although this is consistent with the expression of other *nod* genes in the presence of the overexpressed *nodD*s, it was unexpected in this case because there is no apparent *nod*-box upstream of *nodP*.

And although JSS5 shows a response to the overexpression of *nodD*, the β-galactosidase activity of JSS6, in which *lacZ* is fused in frame with the 27-kDa ORF, remains at 3 units, even with the addition of *nodD* expression plasmids (data not shown). This is consistent with the fusion of *lacZ* to *nodP* being out of frame in JSS6, confirming the 27-kDa ORF is not the *nodP* gene.

We also tested the effects of mutations in *nodP* and *nodQ* on the expression of plasmid-borne copies of *nodD1*::*lacZ* and *nodC*::*lacZ* fusions (Table 2). There is no large effect on the activities of these constructions in the mutant backgrounds.

Nodulation of *nodP* and *nodQ* mutants on various hosts. Mutant strains JT216 (*nodP*::Tn*5*), and JT702 and J0909 (*nodQ*::Tn*5*) were tested for nodulation activity on the hosts *Melilotus albus* Desr., *Medicago truncatula* Gaertn., *Medicago polymorpha* L., and *Trigonella foenum-graecum* L. In no case were the nodulation activities of these strains significantly different from that of Rm1021.

Homologies to *nodP* and *nodQ*. It has been found previously that the leaky nodulation phenotype of *R. meliloti* strains with a single *nodD* mutation was due to the presence of other *nodD* homologues. Strains carrying a single mutation in any of the three *nodD* genes are not completely Nod$^-$. However, a strain with all three genes interrupted displays a Nod$^-$ phenotype (Honma and Ausubel 1987). Therefore, it was thought that *nodP* and *nodQ*, mutations of which display subtle nodulation phenotypes, might have homologues in the *R. meliloti* genome.

Southern blots of total DNA from Rm1021 and its deletion derivative A1375 were probed with the 430-bp *Sal*I-*Sst*I fragment internal to *nodP*, and homologies to other sequences within Rm1021 were found (Fig. 4A, lanes 1–4).

The washing conditions of the hybridization were such that sequences with approximately 70% or more homology with the probe would be detected. An *Xho*I digest of total Rm1021 DNA revealed an extra 0.9-kb band in addition to the expected 1.9- and 4.8-kb bands (Fig. 4A, lane 1). A surprising result is that the expected 1.9-kb band of Rm1021 is still present in the deletion strain, although the entire region should be missing (Fig. 4A, lane 2). An *Eco*RI digest had an extra band of 9.4 kb in addition to the expected 2.2- and 15-kb bands (Fig. 4A, lane 3). And as above, the 2.2-kb band present in wild type is still present in the deletion strain (Fig. 4A, lane 4). In addition, we found homology to *Rhizobium* spp. strain ANU265 (Fig. 4A, lane 7, two bands), *R. leguminosarum* bv. *viciae* Jordan strain R8401/pRL1JI (Fig. 4A, lane 6), and *R. l.* bv. *trifolii* strain ANU843 (Fig. 4A, lane 8), as well as to *E. coli* DNA (Fig. 4A, lane 5).

A probe containing sequence internal to *nodQ* (1.45-kb *Eco*RI-*Bgl*I fragment) was used to demonstrate that more than one such sequence is present in Rm1021. An *Xho*I digest contained an extra band of approximately 600 bp as well as the expected bands of 1.2 and 1.9 kb (Fig. 4B, lane 1); an *Eco*RI digest had the expected 2.2-kb fragment (Fig. 4B, lane 3). And as before, the 1.9-kb *Xho*I band and the 2.2-kb *Eco*RI band are still unexpectedly present in the deletion strain (Fig. 4B, lanes 2 and 4). Homology between the *nodQ* probe and ANU265 was also found (Fig. 4, lane 7).

In order to demonstrate that the DNA isolated from A1375 was indeed that of a deletion strain, the same blot used to compose Figure 4 was probed with an internal *nodC* probe and an internal *syrM* probe. These probes cover regions to the left and right, respectively, of the *nodPQ* region as oriented in Figure 1. The probes confirmed that the A1375 DNA did contain a deletion: both of these probes failed to hybridize with this DNA, though they revealed the appropriate bands in the lanes containing Rm1021 DNA (data not shown). In addition, DNA from two other deletion derivatives of *R. meliloti*, A1376 and GMI255, was subjected to the same treatment. These showed the same patterns of hybridization as did A1375 (data not shown): although flanking DNA sequences confirmed deletion, *nodPQ* bands were still present. Similar results were obtained with *Bam*HI digested DNA, in that some of the bands of expected sizes were not missing in the deletion strains. Thus, it appears that *nodP* and *nodQ* are reiterated in highly conserved form elsewhere in the *R. meliloti* genome.

As shown in Figure 4, *nodP* is homologous to a DNA sequence in *E. coli*. In addition, a *nodP* homologue exists in *Azospirillum brasilense* Tarrand *et al.*, a soil bacterium that fixes dinitrogen and forms a loose association with

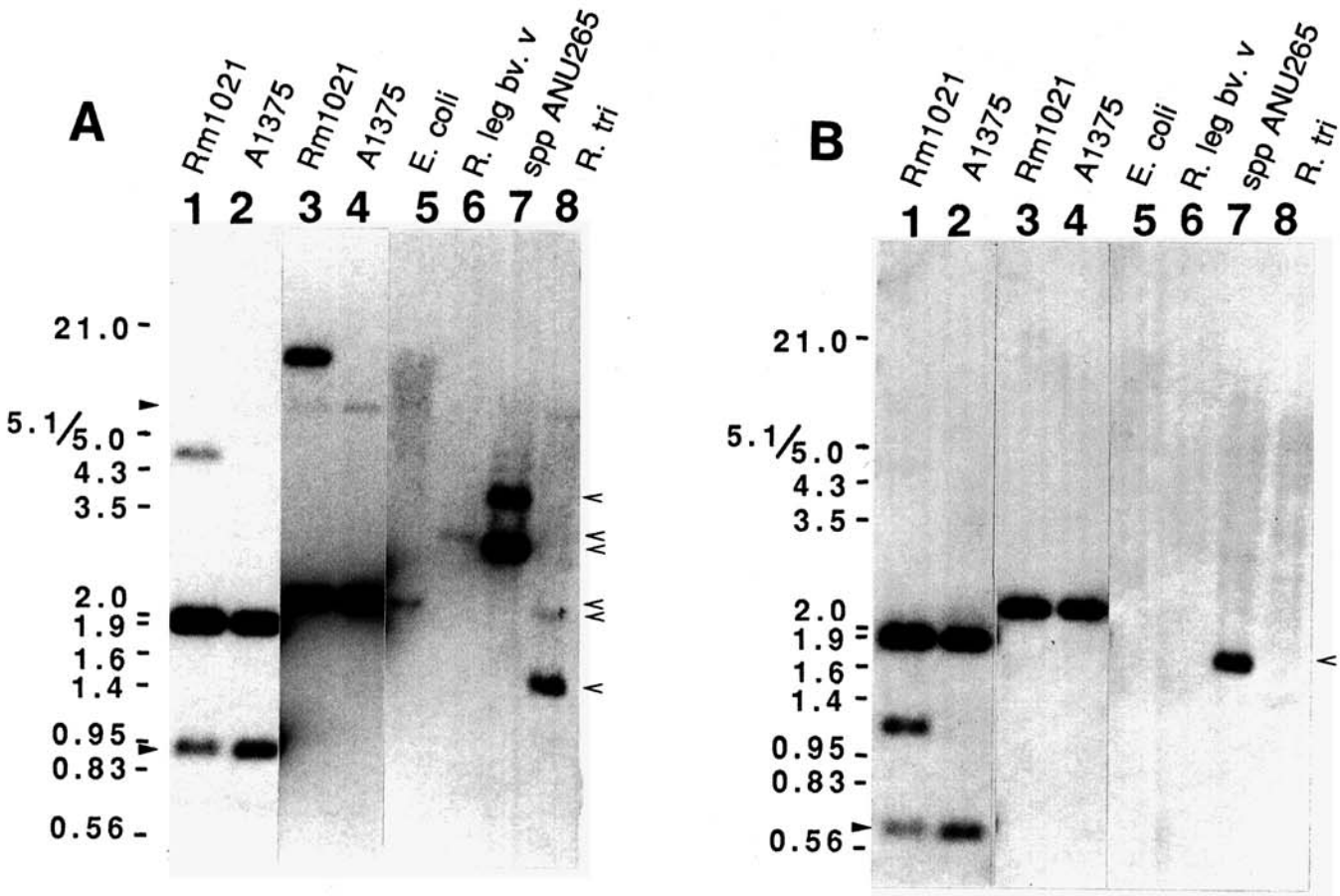

Fig. 4. Southern blot hybridization of various *Rhizobium* species to probes with sequences internal to **(A)** *nodP* and **(B)** *nodQ*. The lanes contain the following genomic DNA: lanes 1, *Xho*I digested Rm1021; lanes 2, *Xho*I digested deletion derivative A1375; lanes 3, *Eco*RI digested Rm1021; lanes 4, *Eco*RI digested A1375; lanes 5, *Eco*RI digested *Escherichia coli* W3110; lanes 6, *Eco*RI digested R8401/pRL1JI; lanes 7, *Eco*RI digested ANU265; and lanes 8, *Eco*RI digested ANU843. Filled arrows (►) indicate extra bands of hybridization in Rm1021; open arrows (◄) indicate hybridization to species other than *R meliloti*.

plant roots, based on a comparison of deduced amino acid sequences (Vieille *et al.* 1988).

A computer search revealed significant homology between the amino acid sequence of the *nodQ* protein product and the procaryotic type EF-Tu of *Saccharomyces cerevisiae* mitochondria, *E. coli*, and *Euglena gracilis* chloroplast. (Note that at the level of Southern blot hybridization no homology was found to *E. coli* DNA.) In addition, homology to the nuclear encoded eucaryotic type elongation factor 1-alpha chain of *S. cerevisiae* and *Artemia salina* (brine shrimp) was found. Figure 5 shows the regions of homology between the amino acid sequence of *nodQ* and *E. coli* EF-Tu. These regions include those identified as forming the binding pocket for GDP and GTP (guanosine triphosphate), as discussed below.

DISCUSSION

R. meliloti nod genes have been defined by transposon mutagenesis, mapping, and sequencing of DNA segments that complemented Nod$^-$ strains. This study extends these procedures to a region downstream of *nodFE* and *nodG*, where previous studies (Swanson *et al.* 1987) indicated the likelihood of additional *nod* genes. We have determined the nucleotide sequence and the protein products for two new genes, *nodP* and *nodQ*. The phenotypes of mutants in these genes are not severely Nod$^-$ on alfalfa; in this respect, they are more similar to *nodFE* and *nodG* than to *nodABC* and *nodH*. Unlike these other *nod* genes, however, *nodP* and *nodQ* are not preceded by a *nod*-box. They appear to be expressed at a low basal level. However when NodD1 (the activator required for inducible *nod* expression) is overproduced, the addition of luteolin leads to a fourfold induction of expression.

There are a number of possible explanations for this phenomena. *nodP* and *nodQ* could be in an operon with *nodFE* and *nodG*. This would easily explain the induction seen in strains which overproduce NodD1, because there is a *nod*-box upstream of *nodF*. However, there are more than 400 bp between the end of *nodG* and the start of *nodP*, and 500 bp between *nodE* and *nodG* (Fisher *et al.* 1987). The region between *nodG* and *nodP* does contain a small ORF of approximately 8 kDa, but we have no evidence for its expression at this time.

Another explanation is that the induction of *nodP::lacZ* seen in the presence of overproduced NodD1 is an artifact of overexpression of the *nodFE* operon coupled with inefficient transcription termination. In a wild-type strain, there may be a transcription termination signal between *nodE* and *nodP*, which is not 100% efficient. When the

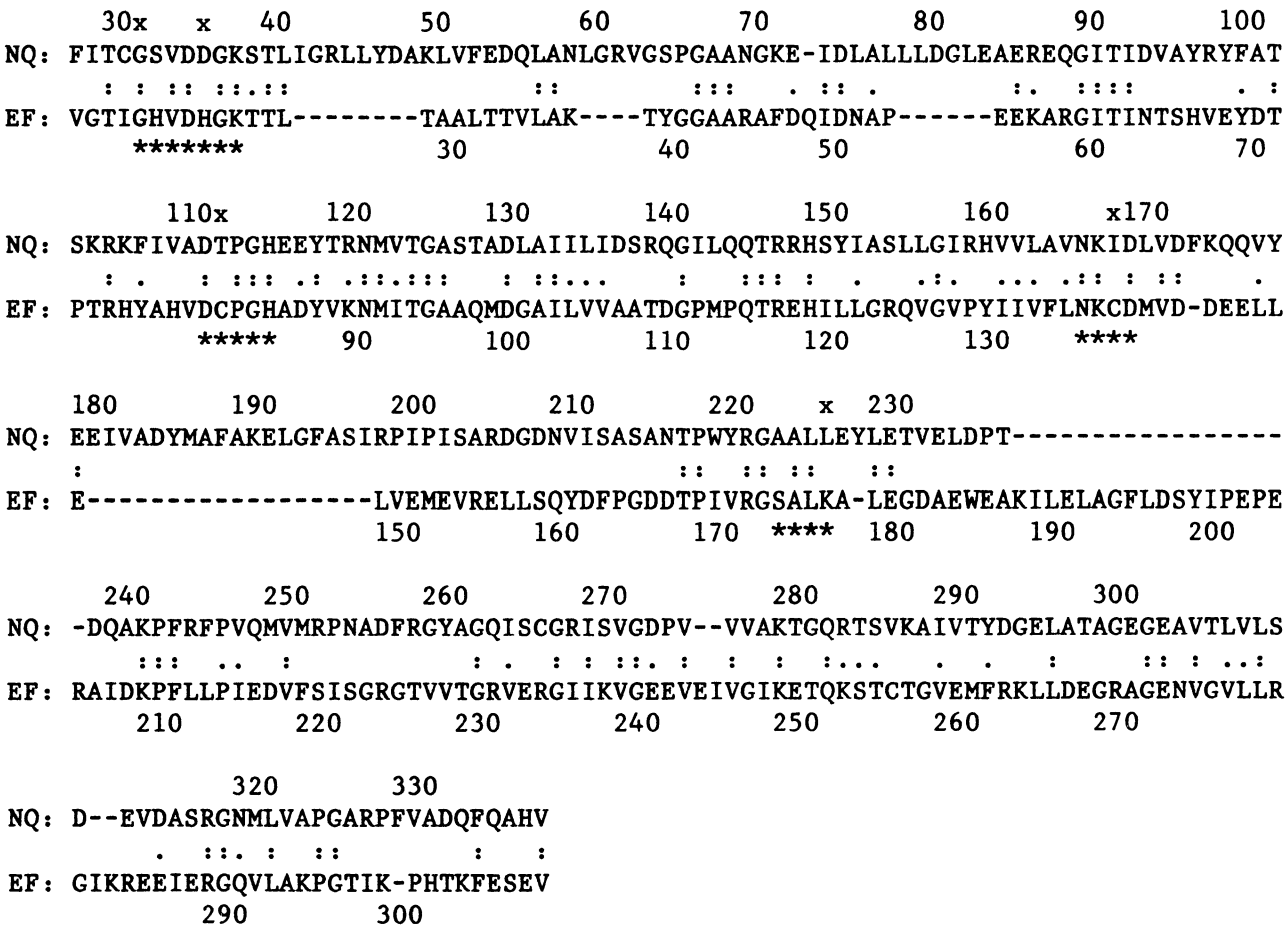

Fig. 5. Amino acid homology between NodQ (NQ, top line) and *Escherichia coli* elongation factor Tu (EF-Tu) (EF, bottom line). Identical amino acids are indicated by a colon (:), and conservative amino acid changes are indicated by a period (.). The four sets of asterisks below the EF-Tu sequence indicate the bases which establish the GDP (guanosine 5′-diphosphate) binding pocket. X's above the NodQ sequence indicate amino acids in the binding pocket that are the least conserved between elongation factors Tu from various procaryotic, mitochondrial, and chloroplast sources.

nodFE operon is artificially overexpressed, read-through of the termination signal may be detected as increased expression of *nodP*.

Yet another possibility is that NodD binding has less sequence specificity than was previously thought. A weak NodD binding site that bears little resemblance to the current consensus *nod*-box may be present upstream of *nodP*. Preliminary gel shift assay results suggest that NodD3 binds the region upstream of *nodP* with 14% of the strength that it binds the *nodH nod*-box (data not shown). NodD3 will shift a nonpromoter DNA fragment at 2% of the amount that it will shift the *nodH nod*-box.

A number of experiments should help sort through these possibilities. These include S1 nuclease mapping of transcription start sites, analysis of *nodP*::*lacZ* fusions in strains deleted or altered in the putative promoter regions, and more extensive analysis of the region upstream of *nodP* via gel retardation assays.

The leaky phenotype of mutations in *nodP* and *nodQ* might be explained by extra copies of these genes in the genome, because we observed homology to internal gene probes. These extra copies are highly conserved at the DNA level, because sites internal to the coding region for three different restriction enzymes are conserved. The role of additional gene copies can be tested by identifying and mutating those extra homologues. In the case of *nodD*, for which there are three functional copies, constructing a triple *nodD* mutant results in a strain with a severe Nod$^-$ phenotype (Honma and Ausubel 1987). However, the presence of additional copies does not account for the different phenotypes of Tn*5* insertions at various positions within the genes. One possibility is that the proteins have more than one domain; only an understanding of the biochemical function of the *nod* gene products will show whether this is the case.

Another interesting observation is the conservation of homology to *nodP* in several other *Rhizobium* species and *E. coli*, and at the level of DNA sequence and deduced amino acid sequence to *A. brasilense* (Vieille *et al.* 1988). This implies that the function of *nodP* might be common among microbes. We plan to map the homology found in the *E. coli* genome, which we hope will lead us to a possible function for *nodP*.

The homology between NodQ and the GDP binding domain of EF-Tu suggests possibilities for the function of the *nodQ* protein product. Although the best homology to NodQ is with the *S. cerevisiae* mitochondrial EF-Tu, we have shown a comparison with EF-Tu of *E coli*, for which the structural details, determined by X-ray crystallography (La Cour *et al.* 1985; Jurnak 1985), are the best known (Fig. 5). Four stretches of amino acids form the GDP binding pocket, and the homology between EF-Tu and NodQ in these regions is striking (Fig. 5). The first stretch involved in the pocket (EF-Tu Gly18 through Lys24, and correspondingly NodQ Gly31 through Lys37) fits the common motif of Gly-X-X-X-X-Gly-Lys that is found in many purine binding or processing proteins, such as ATPases (adenosine triphosphatases), Fe-nitrogenases, and myosin (La Cour *et al.* 1985). In addition, one of the nonconserved amino acids in this part of the pocket (EF-Tu His22 and NodQ Asp35) is also not well-conserved between the above-mentioned elongation factors.

The second and third stretches of amino acids that form the GDP binding pocket of EF-Tu (Asp80 through His84 and Asn135 through Asp138) are also homologous to the corresponding regions of NodQ, with the exception of those amino acids that are not well-conserved between elongation factors. The third stretch is the loop that interacts with the guanine base.

The fourth part of the pocket of the *E. coli* EF-Tu (Ser173 through Lys176) is not well-conserved between the procaryotic type EF- Tu and the eucaryotic type EF1-alpha chain. It is interesting to note that NodQ is more homologous to the eucaryotic type EF1-alpha chain in the region after the third part of the pocket than it is to the procaryotic type EF-Tu.

To determine whether the homology between NodQ and EF-Tu has a functional basis, we plan to determine whether NodQ displays any GDP or GTP binding properties. Further experiments in the characterization of these genes include cloning and mutating the second copy of these genes, mapping the transcription start sites, testing the nodulation behavior of double mutants of *nodPQ* on alfalfa and other *R. meliloti* host plants, and examining the regulatory effects these genes may have on other known nodulation genes. Until the second copies of these genes have been analyzed, the degree of importance of these genes in the process of nodulation cannot be fully understood.

REFERENCE ADDED IN PROOF

A concurrent study of *nodP* and *nodQ* has been conducted by Cervantes *et al.* (in press). They have introduced these genes into *R. trifolii* and *R. leguminosarum* and have reported the effects that these genes have on root hair curling and nodulation of homologous and heterologous host plants.

ACKNOWLEDGMENTS

We are grateful to the members of our research group for invaluable advice and discussion. We thank D. Biek and J. Swanson for use of unpublished plasmids; R. Fisher, M. Yelton, and T. Egelhoff for materials and advice on coupled transcription-translation experiments; and C. Elmerich for communication of unpublished data. We also wish to acknowledge J. Dénarié and C. Rosenberg for discussions of their independent and concurrent study of *R. meliloti nodP* and *nodQ*. We are grateful to J. Swanson and J. Ogawa for their advice on DNA isolation and Southern blotting techniques; and we are indebted to R. Fisher, M. Yelton, B. Rushing, and B. White for their critical review of the manuscript and to A. Bloom for her assistance with its preparation.

This research was supported by contract AS03-82-ER12084 from the Department of Energy to S. R. L. J. S. was supported by a Molecular Biology Training Grant in Cell and Molecular Biology from the National Institutes of Health to Stanford University, Stanford, CA.

LITERATURE CITED

Bolivar, F., Rodriguez, R. L., Greene, P. J., Bethlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95-113.

Buikema, W. J., Long, S. R., Brown, S. E., Van Den Bos, R. C., Earl, C., and Ausubel, F. M. 1983. Physical and genetic characterization of *Rhizobium meliloti* symbiotic mutants. J. Mol. Appl. Genet. 2:249-260.

Casadaban, M. J., Chou, J., and Cohen, S. N. 1980. *In vitro* gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signal. J. Bacteriol. 143:971-980.

Casadaban, M. J., and Cohen, S. N. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. 138:179-207.

Cervantes, E., Sharma, S. B., Maillet, F., Vasse, J., Truchet, G., and Rosenberg, C. The *Rhizobium meliloti* host-range *nodQ* gene encodes a protein that shares homology with translation elongation and initiation factors. Mol. Microbiol. In press.

Corbin, D., Barran, L., and Ditta, G. 1983. Organization and expression of *Rhizobium meliloti* nitrogen fixation genes. Proc. Natl. Acad. Sci. USA 80:3005-3009.

Debellé, F., Rosenberg, C., Vasse, J., Maillet, F., Martinez, E., Dénarié, J., and Truchet, G. 1986. Assignment of symbiotic developmental phenotypes to common and specific nodulation (*nod*) genetic loci of *Rhizobium meliloti*. J. Bacteriol. 168:1075-1086.

Debellé, F., and Sharma, S. B. 1986. Nucleotide sequence of *Rhizobium meliloti* RCR2O11 genes involved in host specificity of nodulation. Nucleic Acids Res. 14:7453-7472.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. 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.

Djordjevic, M. A., Schofield, P. R., and Rolfe, B. G. 1985. Tn*5* mutagenesis of *Rhizobium trifolii* host-specific nodulation genes result in mutants with altered host-range ability. Mol. Gen. Genet. 200:463-471.

Djordjevic, M. A., Zurkowski, W., Shine, J., and Rolfe, B. G. 1983. Sym plasmid transfer to various symbiotic mutants of *Rhizobium trifolii*, *R. leguminosarum*, and *R. meliloti*. J. Bacteriol. 156:1035-1045.

Downie, J. A., Hombrecher, G., Ma, Q.-S., Knight, C. D., Wells, B., and Johnston, A. W. B. 1983. Cloned nodulation genes of *Rhizobium leguminosarum* determine host-range specificity. Mol. Gen. Genet. 190:359-365.

Downie, J. A., Knight, C. D., Johnston, A. W. B., and Rossen, L. 1985. Identification of genes and gene products involved in the nodulation of peas by *Rhizobium leguminosarum*. Mol. Gen. Genet. 198:255-262.

Egelhoff, T. T., Fisher, R. F., Jacobs, T. W., Mulligan, J. T., and Long, S. R. 1985. Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC*. DNA 4:241-248.

Feinberg, A. P., and Vogelstein, B. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6-13.

Fisher, R. F., Egelhoff, T. T., Mulligan, J. T., and Long, S. R. 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.

Fisher, R. F., Swanson, J., Mulligan, J. T., and Long, S. R. 1987. Extended region of nodulation genes in *Rhizobium meliloti* 1021. II. Nucleotide sequence, transcription start sites, and protein products. Genetics 117:191-201.

Friedman, A. M., Long, S. R., Brown, S. E., Buikema, W. J., and Ausubel, F. M. 1982. Use of a *cos* derivative of pRK290 in constructing a clone bank of *Rhizobium meliloti* DNA. Gene 18:289-296.

Gunsalus, R. P., Zurawski, G., and Yanofsky, C. 1979. Structural and functional analysis of cloned deoxyribonucleic acid containing the *trpR-thr* region of the *Escherichia coli* chromosome. J. Bacteriol. 140:106-113.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-359.

Hill, C. W., and Harnish, B. W. 1981. Inversion between ribosomal RNA genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 78:7069-7072.

Hombrecher, G., Götz, R., Dibb, N. J., Downie, J. A., Johnston, A. W. B., and Brewin, N. J. 1984. Cloning and mutagenesis of nodulation genes from *Rhizobium leguminosarum* TOM, a strain with extended host range. Mol. Gen. Genet. 194:293-298.

Honma, M. A., and Ausubel, F. M. 1987. *Rhizobium meliloti* has three functional copies of the *nodD* symbiotic regulatory gene. Proc. Natl. Acad. Sci. USA 84:8558-8562.

Horvath, B., Kondorosi, E., John, M., Schmidt, J., Torok, I., Gyorgypal, L. Z., Barabas, I., Wieneke, U., Schell, J., and Kondorosi, A. 1986. Organization structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. Cell 46:335-344.

Jacobs, T. W., Egelhoff, T. T., and Long, S. R. 1985. Physical and genetic map of a *Rhizobium meliloti* nodulation gene region and nucleotide sequence of *nodC*. J. Bacteriol. 162:469-476.

Jurnak, F. 1985. Structure of the GDP domain of EF-Tu and location of the amino acids homologous to *ras* oncogene proteins. Science 230:32-36.

Kondorosi, E., Banfalvi, Z., and Kondorosi, A. 1984. Physical and genetic analysis of a symbiotic region of *Rhizobium meliloti*: Identification of nodulation genes. Mol. Gen. Genet. 193:445-452.

La Cour, T. F. M., Nyborg, J., Thirup, S., and Clark, B. F. C. 1985. Structural details of the binding of guanosine diphosphate to elongation factor Tu from *E. coli* as studied by X-ray crystallography. EMBO J. 4:2385-2388.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.

Long, S. R., Buikema, W. E., and Ausubel, F. M. 1982. Cloning of *Rhizobium meliloti* nodulation genes by direct complementation of Nod$^-$ mutants. Nature (London) 298:485-488.

Long, S. R. 1989. *Rhizobium*-legume nodulation: Life together in the underground. Cell 56:203-214.

Maniatis, T., Fritsch, E. F., and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 545 pp.

Meade, H. M., Long, S. R., Ruvkun, G. B., Brown, S. E., and Ausubel, F. M. 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn*5* mutagenesis. J. Bacteriol. 149:114-122.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. 101:20-78.

Miller, J. H. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 466 pp.

Morrison, N. A., Hau, C. Y., Trinick, M. J., Shine, J., and Rolfe, B. G. 1983. Heat curing of a Sym plasmid in a fast-growing *Rhizobium* sp. that is able to nodulate legumes and the nonlegume *Parasponia* sp. J. Bacteriol. 153:527-531.

Mulligan, J. T., and Long, S. R. 1989. Extended nodulation gene region in *Rhizobium meliloti* 1021. III. Interaction of three copies of *nodD* and regulatory locus in the control of *Rhizobium meliloti* nodulation and exopolysaccharide genes. Genetics 85:6602-6606.

Mulligan, J. T., and Long, S. R. 1985. Induction of *Rhizobium meliloti nodC* expression by plant exudate requires *nodD*. Proc. Natl. Acad. Sci. USA 82:6609-6613.

Norrander, J., Kempe, T., and Messing, J. 1984. Construction of improved M13 vectors using oligodeoxyribonucleotide-directed mutagenesis. Gene 26:101-106.

Peters, N. K., Frost, J. W., and Long, S. R. 1986. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science 233:917-1008.

Redmond, J. W., Batley, M., Djordjevic, M. A., Innes, R. W., Kuempel, P. L., and Rolfe, B. G. 1986. Flavones induce expression of nodulation genes in *Rhizobium*. Nature 323:632-634.

Reed, K. C., and Mann, D. A. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. 13:7207-7221.

Rigaud, G., Grange, T., and Pictet, R. 1987. The use of NaOH as transfer solution of DNA onto nylon membrane decreases the hybridization efficiency. Nucleic Acids Res. 15:857.

Rolfe, B. G., Gresshoff, P. M., and Shine, J. 1980. Rapid screening for symbiotic mutants of *Rhizobium* and white clover. Plant Sci. Lett. 19:277-284.

Rossen, L., Johnston, A. W. B., and Downie, J. A. 1984. DNA sequence of the *Rhizobium leguminosarum* nodulation genes *nodAB* and *C* required for root hair curling. Nucleic Acids Res. 12:9497-9508.

Rossen, L., Shearman, C. A., Johnston, A. W. B., and Downie, J. A. 1985. The *nodD* gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the *nod A, B, C* genes. EMBO J. 4:3369-3373.

Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A., and Kondorosi, A. 1986. Conservation of extended promoter regions of nodulation genes in *Rhizobium*. Proc. Natl. Acad. Sci. USA 83:1757-1761.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Schofield, P. R., Ridge, R. W., Rolfe, B. G., Shine, J., and Watson, J. M. 1984. Host-specific nodulation is encoded on a 14 kb DNA fragment in *Rhizobium trifolii*. Plant Mol. Biol. 3:3-11.

Schofield, P. R., and Watson, J. M. 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.

Shearman, C. A., Rossen, L., Johnston, A. W. B., and Downie, J. A. 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.

Sutcliffe, J. G. 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. 43:77-90.

Swanson, J., Tu, J. K., Ogawa, J. M., Sanga, R., Fisher, R., and Long, S. R. 1987. Extended region of nodulation genes in *Rhizobium meliloti* 1021. I. Phenotypes of Tn*5* insertion mutants. Genetics 117:181-189.

Tabor, S., and Richardson, C. C. 1987. Sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767-4771.

Török, I., Kondorosi, E., Stepkowski, T., Posfai, J., and Kondorosi, A. 1984. Nucleotide sequence of *Rhizobium meliloti* nodulation genes. Nucleic Acids Res. 12:9509-9524.

Truchet, G., Debellé, F., Vasse, J., Terzaghi, B., Garnerone, A.-M., Rosenberg, C., Batut, J., Maillet, F., and Dénarié, J. 1985. Identification of a *Rhizobium meliloti* P-sym-2011 region controlling the host specificity of root hair curling and nodulation. J. Bacteriol. 164:1200-1210.

Vieille, C., Onyeocha, I., Baca, B.-E., Palacios, R., Flores, M., and Elmerich, C. 1988. Genetic analysis of a 90 Md plasmid of *Azospirillum brasilense* Sp7. Pages 209-210 in: Molecular Genetics of Plant-Microbe Interations. R. Palacios and D. P. S. Verma, eds. American Phytopathological Society, St. Paul, MN. 401 pp.

Vieira, J., and Messing, J. 1987. Production of single stranded plasmid DNA. Methods Enzymol. 153:3-11.

Yelton, M. M., Mulligan, J. T., and Long, S. R. 1987. Expression of *Rhizobium meliloti nod* genes in *Rhizobium* and *Agrobacterium* backgrounds. J. Bacteriol. 169:3094-3098.

Zaat, S. A. J., Wijffelman, C. A., Spaink, H. P., Van Brussel, A. A. N., Okker, R. J. H., and Lugtenberg, G. J. J. 1987. Induction of the *nodA* promoter of *Rhizobium leguminosarum* Sym plasmid PRL1JI by plant flavanones and flavones. J. Bacteriol. 169:198-204.