MOLECULAR PLANT-MICROBE INTERACTIONS
Vol. 1, No. 3, pp. 145-149, 1988

Identification and Characterization of the *nod*D Gene in *Rhizobium leguminosarum* strain 1001

Andrea Squartini,[1] Ron J. M. van Veen,[2] Tonny Regensburg-Tuink,[2] Paul J. J. Hooykaas,[2] and Marco P. Nuti[1]

[1] Dipartimento di Biotecnologie Agrarie, Università di Padova, via Gradenigo 6, 35131 Padova, Italy, and [2]Department of Plant Molecular Biology, University of Leiden, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands.
Received 30 November 1987. Accepted 21 March 1988.

A gene library of the symbiotic 240-kb plasmid of *Rhizobium leguminosarum* strain 1001 was constructed in pUC18. The clones showing homology with a 6.6-kb fragment containing *nod*EFDABC from the *Sym* plasmid pRL1JI were detected by colony hybridization. Additional probes from the symbiotic region of pRL1JI were used to localize the corresponding genes on the map of pRle1001*a*. The relative positions of *nod* and *nif* gene clusters are different than those of pRL1JI. A comparison of the amino acid sequence for NodD from pRle1001*a* with NodD proteins from other *Rhizobium* species showed a high degree of sequence conservation at the amino terminus of the protein.

Additional keywords: Sym plasmids, nodulation genes, NodD protein, *Rhizobium leguminosarum*.

Most of the functions needed for the interaction of fast-growing *Rhizobium* with leguminous plants are encoded by large plasmids (Nuti *et al.* 1977) called *Sym* (symbiotic) plasmids (Hooykaas *et al.* 1981). Such plasmids share various degrees of homology among different species of *Rhizobium*. A set of genes, the *nod* (nodulation) genes, is responsible for the formation of nodules on the roots of the host plant. This process occurs by a highly specific series of recognition events, some of which are controlled by *nod* genes acting as host-specificity determinants.

Some of the *nod* genes, namely *nod*DABC, appear to be present in virtually all *Rhizobium* and were therefore referred to as common *nod* genes. Nevertheless, *nod*D seems to play a role in the host specificity as well (Spaink *et al.* 1987), indicating that, despite the overall conservation of this genetic unit, minor differences may have a significant effect in terms of inducibility by different plant compounds. Flavones and flavanones can in fact turn on some of the *nod* operons, a process that requires the *nod*D product, probably as a positive regulator (Rossen *et al.* 1985).

Probes for the common *nod* and *nif* genes can be used to screen gene libraries from different *Rhizobium* to find the corresponding *nod* and *nif* regions and compare the arrangement of symbiotic regions in different species and strains. Our effort was devoted to *Rhizobium leguminosarum* strain 1001. This strain, originally isolated from *Pisum sativum* nodules, carries a 240-kilobase pair (kbp) symbiotic plasmid, pRle1001*a*, also called p*Sym*1 (Hooykaas *et al.* 1982). A circular restriction map of the whole plasmid was defined by Prakash *et al.* (1982), including the homology with other *Sym* plasmids, with the *Agrobacterium tumefaciens* Ti plasmids, and with the *nif* structural genes (Prakash *et al.* 1981).

Address correspondence to Marco P. Nuti.

Nucleotide and/or amino acid sequence data is to be submitted to GenBank as accession number J03671.

© 1988 The American Phytopathological Society

The aims of this work were: to identify and clone the genetic determinants for nodulation of *Vicia* and *Pisum* in *R. leguminosarum* 1001; to compare the organization of genes in the symbiotic region with that found in pRL1JI, the source of the *nod* and *nif* probes; and to further characterize the *nod*D regulatory gene.

MATERIALS AND METHODS

Bacteria and culture conditions. The bacteria used throughout this study were *R. leguminosarum* strains RCR 1001 (Rothamsted Collection, U.K.) and LPR 1105 (1001 rif^r; P. Hooykaas, University of Leyden, The Netherlands), *Escherichia coli* strains 490 (PC 2480 Phabagen Collection, University of Utrecht, The Netherlands), HB 101 (Boyer and Roulland-Dussoix 1969), KA 817 *dam*$^-$ *dcm*$^-$ (P. van de Putte, University of Leyden), and JM 101 (Messing *et al.* 1981). *Rhizobium* was grown on yeast marmitol broth (Hooykaas *et al.* 1977) or tryptone-yeast extract medium (Beringer 1974), at 28° C. *E. coli* was grown on Luria-Bertani (LB) medium (10 g/L of Bacto-tryptone, 5 g/L of yeast extract, 5 g/L of NaCl, pH 7) at 37° C. Solidified media contained 18 g/L of agar.

Plasmids. The following plasmids were used: pIJ 1047 (5.5-kbp *Hind*III fragment, *fix* region of pRL1JI [Hombrecher *et al.* 1984], cloned in pBR322), pIJ 1238 (5-kbp *Hind*III fragment, 24-K region of pRL1JI [Hombrecher *et al.* 1984], cloned in pSUP 202), pIJ 1246 (6.6-kbp *Eco*RI fragment, *nod* region of pRL1JI [Hombrecher *et al.* 1984], cloned in pSUP 202), and pSA30 (7-kbp *Eco*RI fragment, *nif*HDK of *Klebsiella pneumoniae* [Cannon *et al.* 1979], cloned in pACYC 184). The large symbiotic plasmid of *Rhizobium* was isolated as described by Prakash *et al.* (1981). Plasmids from *E. coli* were isolated as described by Birnboim and Doly (1979).

DNA manipulation techniques. Restriction endonuclease digestions, DNA cloning, agarose gel electrophoresis, Southern blotting, and nick translation were carried out as described by Maniatis *et al.* (1982). DNA transformation of *E. coli* was performed by the procedure of Dagert and

Ehrlich (1979). The 240-kbp symbiotic plasmid was digested with *Bam*HI, and the fragments were cloned into a linearized pUC18 (Norrander *et al.* 1983). *E. coli* 490 was transformed as described.

Southern hybridization. Filters were prehybridized in 5× SSPE (20× SSPE stock: 3.6 M NaCl; 0.2 M NaH_2P0_4; 0.02 M Na_2EDTA, pH 6.5), 5× Denhardt solution (0.02% polyvinylpyrrolidone; 0.02% Ficoll; 0.02% bovine serum albumin), 0.5 mg of denatured herring sperm DNA per milliliter, for 5 hr at 72° C; filters were subsequently hybridized, under stringent conditions, in the same solution containing in addition 0.3% sodium dodecyl sulfate (SDS), and the alkali-treated probe.

Filters were washed twice with 5 × SSPE, 0.1% SDS at 72° C for 30 min; twice with 0.5× SSPE, 0.1% SDS at 72° C for 30 min; and once with 0.5× SSPE at room temperature. Air-dried filters were wrapped in plastic, and an X-ray film was exposed at −70° C with an intensifying screen.

Colony hybridization. The *E. coli* recombinant clones were plated at a density of 1,000 colony-forming units per plate onto nitrocellulose filters (Schleicher & Schuell), previously laid on well-dried large LB plates supplemented with carbenicillin at 50 μg/ml. Plates were incubated overnight at 37° C. Two replica filters were obtained by touching the master filter, and colonies were allowed to grow for 8 hr on new plates. Colonies were lysed by leaving the filter, colony side up, on Whatman paper 3MM saturated with 0.5 M NaOH, 0.05% Triton X-100 for 10 min, and then on paper with 0.5 M NaOH, 1.5 M NaCl for 10 min. This treatment was followed by a neutralizing step on paper with 1 M Tris, pH 8.3, for 5 min and finally on paper with 0.5 M Tris, pH 8.5, 1.5 M NaCl for 15 min. Filters were air-dried and baked for 24 hr at 80° C in a vacuum oven. The hybridization conditions were the same as for Southern filters.

DNA sequencing. Fragments of interest were cloned in

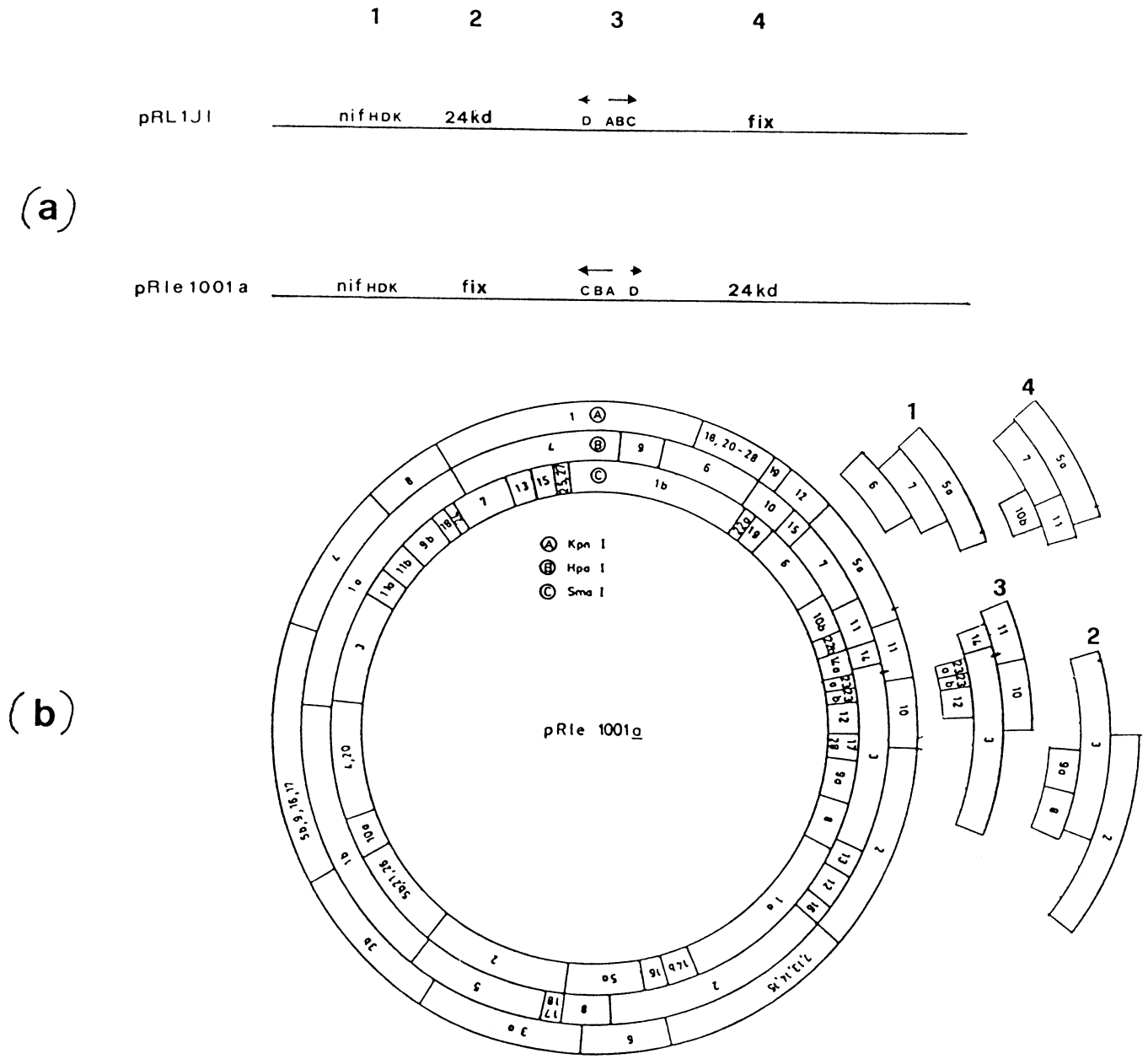

Fig. 1. (a), Location of the homology regions between pRL1J1 and pRle1001*a*. Relative order of the fragments used as probes in pRL1J1 (24 kDa indicates a gene for a 24-kDa protein), compared to that found on pRle1001*a*. **(b),** Restriction map of pRle1001*a*. The fragments hybridizing with the different probes are shown as excised clusters; numbers refer to the probes used.

M13 mp 18/19 (Norrander *et al.* 1983). The transfection of *E. coli* JM 101 and the preparation of template DNA were performed as described by Messing *et al.* (1981). Both strands were sequenced by the method of Sanger *et al.* (1977). Reagents were purchased from Amersham. ^{35}S-ATP was used as label at 8 μCi per group of four reactions; Klenow DNA-polymerase was from Boehringer. Samples were boiled for 3 min and run on a 0.4-mm thick polyacrylamide gel, in Tris-borate buffer at 35 mA/1,300 V. The gel was dried with a Bio-Rad gel dryer and Kodak X-0 mat film was exposed overnight at room temperature. Data were analyzed by means of the Beckman Microgenie program on an IBM personal computer.

RESULTS

Regions of pRle1001*a* are colinear with the *nod* segment from pRLlJI. Two different clones were detected by colony hybridization with pIJ 1246 (6.6-kb *nod* fragment from pRLlJI) as a probe in the *Bam*HI plasmid bank of pRle1001*a* (p*Sym*1). These contained adjacent fragments, which partially overlapped the region containing the *Kpn*I fragments 10 and 11 shown on the map of pRle1001*a* (Fig. 1). A detailed restriction map of the cloned region aligned with that of pRLlJI is given in Figure 2. The position of the *nod* genes in pRLlJI is also shown. Many restriction sites in the *nod* regions of pRLlJI and p*Sym*1 turn out to be conserved, which underscores the close evolutionary relatedness of these regions.

The *nod* and *nif* regions of pRle1001*a* and pRLlJI contain homologous sequences but are arranged in a different order. Four different ^{32}P-labeled probes from the plasmid pRLlJI were hybridized to Southern blots of pRle1001*a* digested with the enzymes *Bam*HI, *Bgl*II, *Eco*RI, *Hind*III, *Hpa*I, *Kpn*I, *Pst*I, *Sma*I, *Sst*I, and *Xba*I. The following probes were used: pSA30 (*nif*HDK of *K. pneumoniae*); pIJ 1238 containing the gene for a 24-kDa protein close to the *nif* structural genes in pRLlJI; pIJ 1246 with *nod*FEDABC of pRLlJI; and pIJ 1047 with the pRLlJI *fix* region. The relative positions of these gene clusters on pRLlJI (Hombrecher *et al.* 1984) are shown in the upper part of Figure 1. The regions of pRle1001*a* hybridizing with the probes are also shown in Figure 1. Using the *nif*HDK as a reference point, the locations of the other clusters in pRle1001*a* are different than those of the corresponding clusters in pRLlJI. The *fix* genes map next to the *nif* structural genes on pRle1001*a*, whereas the 24-kDa protein gene is distal to the *nod* area. The analyses of the cloned *nod* region showed that the order of the *nod* genes is the same, but their orientations are in opposite directions toward the structural *nif* genes.

Nucleotide sequencing of *nod*D from pRle1001*a*. The sequence of a 1,439-bp stretch containing a 969-bp open reading frame (ORF) is shown in Figure 3. The sequencing strategy is shown in Figure 2. The ORF starts at position 406 and stops at position 1,374. The consensus for a *nod* box is boxed. The nucleotide sequence of the ORF shows a high degree of homology with that of previously published *nod*D genes.

Inverted repeat sequences of 10 bp map immediately downstream from the end of ORF and are separated by 10 bp. A consensus amino acid sequence for DNA binding proteins, as described by Shearman *et al.* (1986), maps at the N terminus of the gene and is shown in brackets in Figure 3. By comparing the nucleotide sequence with a data bank, a homology of 94.7% was found between a 19-bp stretch spanning position 55–73 and the reverse strand of a region flanking the T-DNA of the *A. tumefaciens* octopine Ti plasmid pTi15955 (Barker *et al.* 1983; position 23,932–23,950).

Computer-assisted NodD protein sequence analysis. The amino acid sequence deduced from DNA sequence was compared to that of four other *nod*D genes. The protein alignment is shown in Figure 4. The homology at protein level is 90.7% with NodD from *R. leguminosarum* (pRLlJI) (Shearman *et al.* 1986), 74.7% with *R. meliloti* 1021

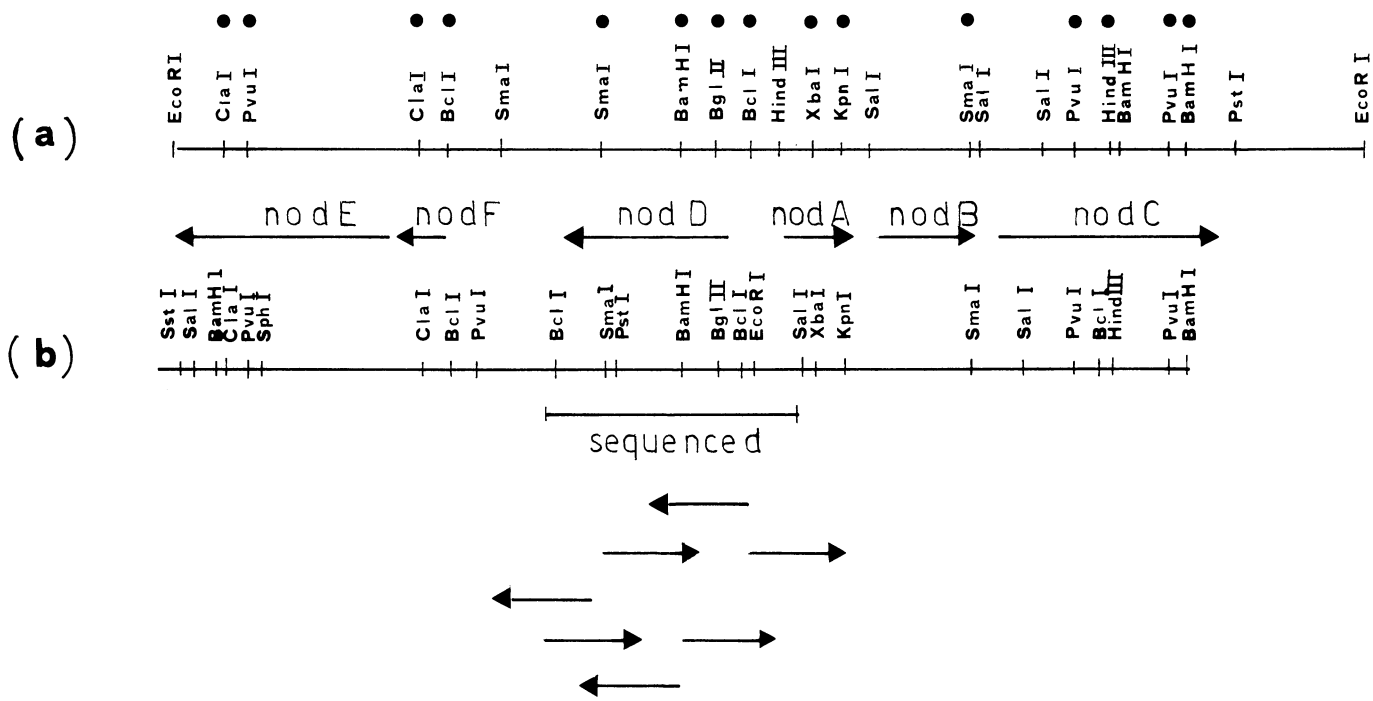

Fig. 2. (a), 6.6-kilobase pair *Eco*RI fragment from the *nod* region of pRLlJI used as a probe. The restriction sites conserved in pRle1001*a* are dotted. **(b),** Cloned region of pRle1001*a*. Arrows indicate the sequencing strategy.

(Egelhoff *et al.* 1985), 72.3% with NodD from *R. trifolii* ANU843 (Schofield and Watson 1986), and 64.8% with NodD from *Bradyrhizobium parasponiae* (Scott 1986). A computer prediction of the secondary structure by the algorithm of Garnier *et al.* (1978) for our NodD is given in Figure 4.

DISCUSSION

Homology between pRle1001*a* and pRLlJI. As expected, these two *R. leguminosarum* plasmids share sufficient homology to allow the identification and cloning of corresponding regions. However, there are some noteworthy differences. Although the nucleotide sequence of the *nod*D genes is greater than 90% conserved, there are significant differences nevertheless in the restriction map of the regions examined (see Fig. 2). Furthermore, the order of the gene clusters is different, as shown in Figure 1. This suggests that a large fragment encoding the *fix*, the *nod* genes, and the gene for the 24-kDa protein has undergone an inversion event.

Nucleotide sequence. The putative start site of a second ORF (strongly homologous to *nod*A) on the complementary strand is shown in Figure 3. This start site for *nod*A was chosen by comparing the *nod*A sequence from two *R. leguminosarum* strains with that from an *R. meliloti* strain. The inverted complementary sequence found downstream of *nod*D is a potential transcriptional terminator. A 19-bp sequence at the N terminus of *nod*A is also found on the pTi15955 from *A. tumefaciens*. Its location is 49 bp to the right of the right border of the T_R-DNA. The significance of this homology, if any, is unknown.

NodD protein sequence analysis. The higher degree of amino acid sequence conservation is at the amino terminus of the gene. A sequence comparison of five genes indicates a strong consensus at the first 40 amino acid residues (Fig. 4).

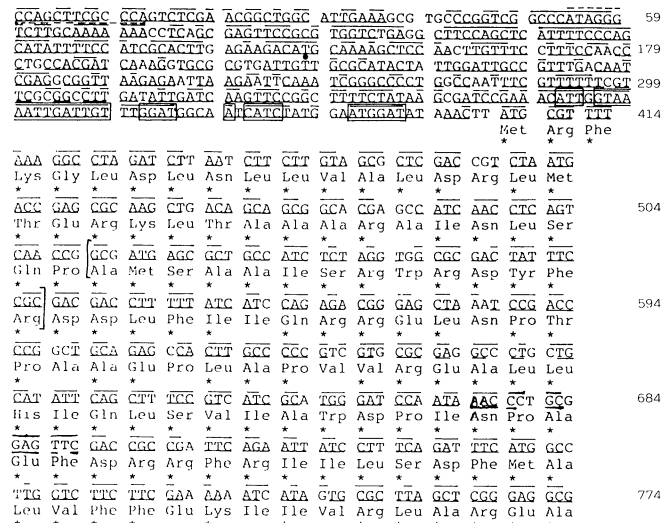

Fig. 3. Nucleotide sequence of the *nod*D gene of pRle1001*a*. The conserved nucleotides between this plasmid and pRLlJI are underlined; those conserved in the *Rhizobium meliloti* 1021 megaplasmid are overlined. Asterisks represent the conserved amino acids between pRle1001*a* and pRLlJI. The boxes indicate the consensus for the *nod* box. The dot (.) indicates the presumed start of *nod*A on the complementary strand. The broken line from position 55 to 73 overlines the region conserved in pTi15955. A consensus amino acid sequence for DNA-binding proteins is shown in brackets.

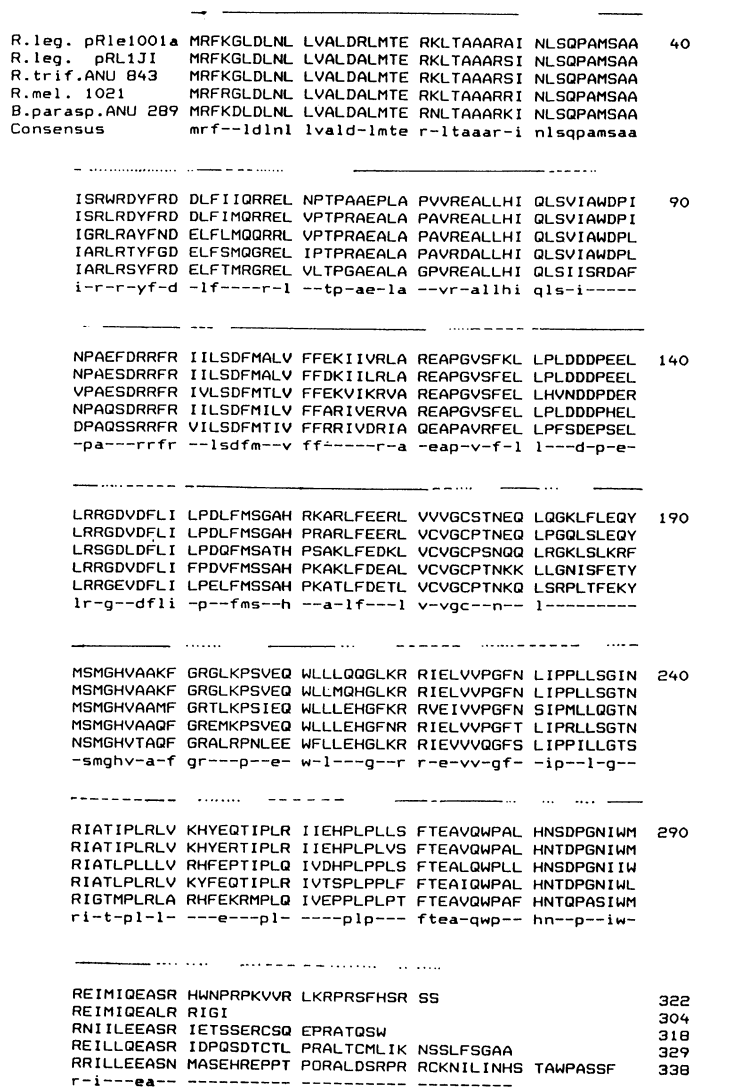

Fig. 4. Protein alignment of different NodDs. The computer prediction for the secondary structure is indicated above the sequence. Solid line (————) represents α-helix structures, broken line (- - - -) represents β-sheet, and dotted line (. . . .) indicates the reverse turn, as described by Garnier *et al.* (1978).

According to the computer prediction, the secondary structure of the 1–40 amino acid region is an α-helix.

The predicted secondary structures of five NodD proteins are shown in Figure 4. These show that the unconserved regions are in most cases characterized as reverse turn.

The border regions between an α-helix and a β-sheet domain (e.g., from residue 75 to 89, 120 to 128, and 170 to 175) are generally better conserved than other parts of the sequence. The highest degree of NodD amino acid conservation is between *R. leguminosarum* and *R. trifolii* proteins. The *B. parasponiae* sequence is the most divergent and is more closely related to the *R. meliloti* protein than that of the two *R. leguminosarum* strains. The region from residue 157 to 170 is predicted as an α-helix and is quite variable in all the species.

The amino acid sequence between residues 271 and 290 is relatively well conserved and precedes the only part of the protein that does not show any consensus. From residue 290 up to the carboxy terminus end, the various NodD are completely unrelated and vary in length. Even the two Nod proteins from *R. leguminosarum* differ in length by 18 residues. The predominant secondary structure predicted for this end is reverse turn.

It has now been shown (Spaink *et al.* 1987) that the *nod*D gene plays an important role in host specificity. The predicted secondary structures of five *nod*D proteins that we have described could provide a useful guide for structural and functional analyses of these proteins.

ACKNOWLEDGMENTS

We thank John Allan Downie for providing the probes and Frank Cannon for critically reading the manuscript. Mark Van Haaren, Leo Melchers, Amke den Dulk-Ras, and Norbert Sedee are gratefully acknowledged for discussion and cooperation. We thank Lidia Zaltron for typing the manuscript.

This research was supported by CNR (Italy), Special Grant I.P.R.A., paper 1,612.

An account of this work was presented at the third International Symposium on Plant-Microbe Interactions, Montreal, Quebec, July-August 1986.

LITERATURE CITATIONS

Barker, R. F., Idler, K. B., Thompson, D. V., and Kemp, J. D. 1983. Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti plasmid pTi 15955. Plant Mol. Biol. 2:335-350.

Beringer, J. E. 1974. R-factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. 84:188-198.

Birnboim, H. C., and Doly, J. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acid Res. 7:1513-1523.

Boyer, H. W., and Roulland-Dussoix, D. 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. 41:459-472.

Cannon, F. C., Riedel, G. E., and Ausubel, F. M. 1979. Overlapping sequences of *Klebsiella pneumoniae nif* DNA cloned and characterized. Mol. Gen. Genet. 174:59-66.

Dagert, M., and Ehrlich, S. D. 1979. Prolonged incubation in calcium chloride improves the competence of *E. coli* cells. Gene 6:23-28.

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

Garnier, J., Osguthorpe, D. J., and Robson, B. 1978. Analysis of the accuracy and implications of single methods for predicting the secondary structure of globular proteins. J. Mol. Biol. 120:97-120.

Hombrecher, G., Gotz, 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 an extended host range. Mol. Gen. Genet. 194:293-298.

Hooykaas, P. J. J., Klapwijk, P. M., Nuti, M. P., Schilperoort, R. A., and Rorsch, A. 1977. Transfer of the *Agrobacterium tumefaciens* Ti plasmid to avirulent agrobacteria and to *Rhizobium ex planta*. J. Gen. Microbiol. 98:477-484.

Hooykaas, P. J. J., Van Brussel, A. A. N., den Dulk-Ras, A., Van Slogteren, G. M. S., and Schilperoort, R. A. 1981. *Sym* plasmids of *Rhizobium trifolii* expressed in different rhizobial species and in *Agrobacterium tumefaciens*. Nature 291:351-352.

Hooykaas, P. J. J., Snijdewint, F. G. M., and Schilperoort, R. A. 1982. Identification of the *Sym* plasmid of *Rhizobium leguminosarum* strain 1001 and its transfer to and expression in other rhizobia and *Agrobacterium tumefaciens*. Plasmid 8:73-82.

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

Messing, J., Crea, R., and Seeburg, P. M. 1981. A system for shotgun DNA sequencing. Nucleic Acid Res. 9:309-321.

Norrander, J., Kempe, T., and Messing, J. 1983. Improved M13 vectors using oligonucleotide-directed mutagenesis. Gene 26:101-106.

Nuti, M. P., Ledeboer, A. M., Lepidi, A. A., and Schilperoort, R. A. 1977. Large plasmid in different *Rhizobium species*. J. Gen. Microbiol. 100:241-249.

Prakash, R. K., Schilperoort, R. A., and Nuti, M. P. 1981. The large plasmids of fast-growing rhizobia: Homology studies and location of structural nitrogen fixation (*nif*) genes. J. Bacteriol. 145:1129-1136.

Prakash, R. K., Van Veen, R. J. M., and Schilperoort, R. A. 1982. Restriction endonuclease mapping of a *Rhizobium leguminosarum sym* plasmid. Plasmid 7:271-280.

Rossen, L., Shearman, C. A., Johnston, A. W. B., and Downie, J. A. 1985. The *nod*D 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.

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

Schofield, P. R., and Watson, J. M. 1986. DNA sequence of *Rhizobium trifolii* nodulation genes reveals a reiterated and potentially regulatory sequence preceding *nod*ABC and *nod*FE. Nucleic Acid Res. 14:2891-2903.

Scott, K. 1986. Conserved nodulation genes from the non-legume symbiont *Bradyrhizobium* sp. (*Parasponia*). Nucleic Acid Res. 14:2905-2919.

Shearman, C. C., Rossen, L., Johnston, A. W. B., and Downie, J. A. 1986. The *Rhizobium leguminosarum* nodulation gene *nod*F encodes a polypeptide similar to acyl-carrier protein and is regulated by *nod*D plus a factor in pea root exudate. EMBO J. 5:647-652.

Spaink, M. P., Wijffelman, C. A., Pees, E., Okker, R. J. M., and Lugtenberg, B J. J. 1987. *Rhizobium* nodulation gene *nod*D as a determinant of host specificity. Nature 328:337-339.