The EMBO Journal vol.8 no.10 pp.2811 – 2818, 1989

Genetic analysis and cellular localization of the *Rhizobium* host specificity-determining NodE protein

Herman P.Spaink, Jeremy Weinman[1], Michael A.Djordjevic[1], Carel A.Wijffelman, Robert J.H.Okker and Ben J.J.Lugtenberg

Department of Plant Molecular Biology, Leiden University, Nonnensteeg 3, 2311 VJ Leiden, The Netherlands and [1]Plant Microbe Interaction Group, Research School Biological Sciences, Australian National University, PO Box 475, Canberra City, Australia

Communicated by A.van Kammen

The nucleotide sequence of the *nodE* gene of *Rhizobium trifolii* strain ANU843 was determined. Like the *nodE* gene of *R.leguminosarum* strain 248 it encodes a protein with a predicted mol. wt of 42.0 kd. The predicted NodE proteins of *R.trifolii* and *R.leguminosarum* have a homology of 78%. Using antibodies raised against the NodE protein of *R.trifolii* it was shown that the NodE products of *R.leguminosarum* and *R.trifolii* are localized in the cytoplasmic membrane. Furthermore, these NodE proteins are predicted to contain at least two non-polar transbilayer α-helices. The *nodE* genes of *R.trifolii* and *R.leguminosarum* were separated from the *nodF* genes that precede them in the respective *nodFE* operons. Using the resulting clones, in which NodE was produced under control of the flavonoid-inducible *nodABCIJ* promoter of *R.leguminosarum*, it was shown that the NodE product is the main factor that distinguishes the host range of nodulation of *R.trifolii* and *R.leguminosarum*. Hybrid *nodE* genes, which consist of a 5′ part of the *R.leguminosarum nodE* gene and a 3′ part of the *R.trifolii* gene, were constructed. From the properties of these hybrid genes it was concluded that a central region of 185 amino acids at the most, containing only 44 non-identical amino acids, determines the difference in the host-specific characteristics of the two NodE proteins.

Key words: host-specific nodulation/hybrid genes/intracellular localization/NodE/*Rhizobium*

Introduction

The symbiosis between bacteria of the genus *Rhizobium* and their leguminous host plants results in the formation of root nodules in a species-specific way, in that a particular bacterial species can nodulate only a limited number of plant species. In *R.leguminosarum*, with plants of the genera *Vicia, Lathyrus, Pisum* and *Lens* as hosts, *R.trifolii*, with plants of the genus *Trifolium* as hosts, and *R.meliloti* with plants of the genera *Medicago, Melilotus* and *Trigonella* as hosts, genes involved in the nodulation and host range determination are localized on large Sym (for symbiosis) plasmids. Several of these so called *nod* genes, which are highly homologous and which show a similar operon organization, designated *nodABC(IJ), nodD* and *nodFE*, have been identified (Schofield and Watson, 1986; Shearman *et al.*, 1986; Spaink *et al.* 1986; Fisher *et al.*, 1987). Besides these homologous *nod* genes the *R.leguminosarum* and *R.trifolii* Sym plasmids also contain the genes *nodL* and *nodMN* of which no counterparts are known in *R.meliloti* (Surin and Downie, 1988). In *R.meliloti* the *nod* genes *nodG* and *nodH*, which have no counterparts in *R.leguminosarum* and *R.trifolii*, are present on the Sym plasmid (Debellé and Sharma, 1986; Horvath *et al.*, 1986). The *nod* genes identified so far are regulated at the transcriptional level as one regulon in which the constitutively expressed *nodD* product acts as a positive regulator (Innes *et al.*, 1985; Mulligan and Long, 1985; Shearman *et al.*, 1986; Spaink *et al.*, 1986). This *nod* regulon is only transcribed in the presence of an inducer produced by the host plants. The inducing compounds have been identified as flavones, flavanones or closely related compounds (Firmin *et al.*, 1986; Peters *et al.*, 1986; Redmond *et al.*, 1986; Wijffelman *et al.*, 1986). A strongly conserved regulatory domain, designated the *nod* box, is present upstream of all inducible *nod* operons and appears to be essential for the activity of the inducible promoters (Rostas *et al.*, 1986; Schofield and Watson, 1986; Spaink *et al.*, 1987a).

The genes *nodABCIJ* are not involved in the host specificity of nodulation on various host plants since they can be replaced by the corresponding genes of other *Rhizobium* species without affecting the host range. Therefore these *nod* genes are referred to as common *nod* genes (for a review see Djordjevic *et al.*, 1987). Although *nodD* has also been considered as a common *nod* it has recently been shown that *nodD* can be a determinant of host specificity, presumably as a result of the interaction of its product with structurally distinct sets of flavonoids in a species-specific way (Horvath *et al.*, 1987; Spaink *et al.*, 1987b). It is clear that other *nod* genes are also involved in the determination of the host range. For *R.meliloti* it has been shown that the *nodH* gene is of crucial importance for the specific nodulation on *Medicago* host plants (Debellé and Sharma, 1986; Horvath *et al.*, 1986). This *nod* gene is supposed to be involved in the modification of an extracellular *nodABC* dependent signal to a *Medicago* specific form (Faucher *et al.*, 1988). For *R.leguminosarum* and *R.trifolii* the operon containing the *nodFE* genes appears to be very important for the host specificity (Downie *et al.*, 1983; Wijffelman *et al.*, 1985; Djordjevic *et al.*, 1986). Djordjevic *et al.* (1985) have shown that a *R.trifolii* strain ANU843, which contains a Tn*5* insertion in the *nodE* gene, acquired the ability to nodulate peas. Presently this is the only example of a host range limiting influence of the *nodE* gene. In this paper the role of the *nodE* genes in the host specificity of *R.leguminosarum and R.trifolii* is studied in more detail. Furthermore, the products encoded by the *nodE* genes were cellularly localized using antibodies. We show that, in isogenic *R.trifolii* strains, the source of the *nodE* gene alone determines whether *R.leguminosarum* or *R.trifolii* host plants are efficiently nodulated. By constructing a series of

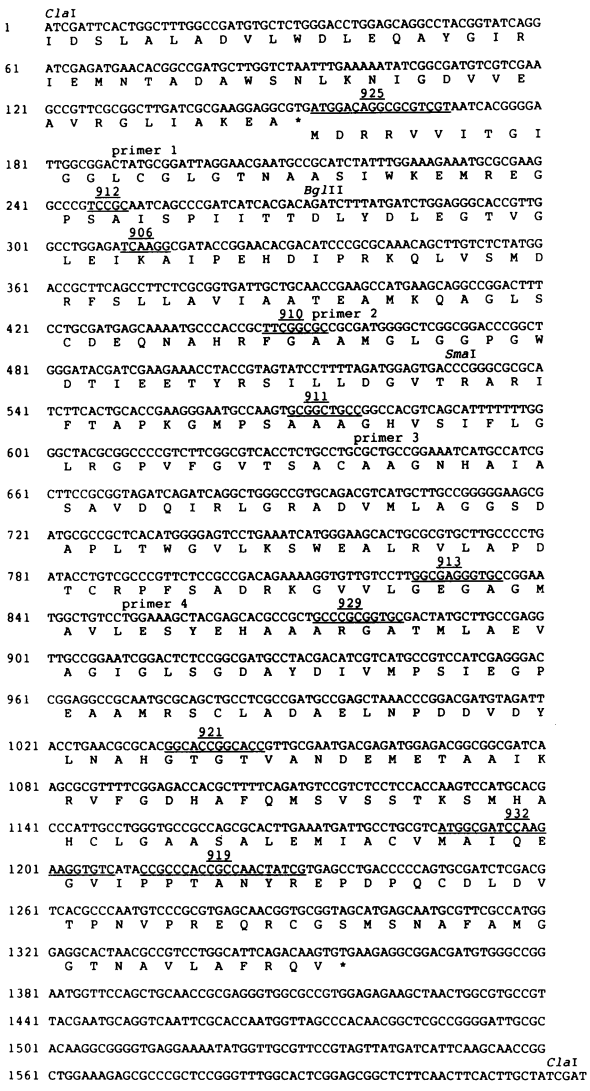

Fig. 1. Nucleotide sequence of the *R.trifolii nodE* gene. Shown is the sequence of the *Cla*I fragment illustrated in Figure 3A (fragment III). The predicted protein sequence of the NodE protein and part of the NodF protein is indicated. Also indicated are the sites of recombination in the *nodE* hybrid genes, the position of the synthetic oligonucleotide primers used for nucleotide sequencing of constructed hybrid *nodE* genes and the relevant restriction sites.

hybrid *nodE* genes a region was localized which determines the difference in host-specific characteristics of the parental NodE proteins.

Results

Genetic analysis and immunological detection of NodE protein

In order to compare the *nodE* genes of *R.leguminosarum* and *R.trifolii*, the nucleotide sequence of *R.trifolii* strain ANU843 was determined (Figure 1). Like the predicted NodE protein of *R.leguminosarum* strain 248 (Shearman *et al.*, 1986; Figure 2) this *nodE* gene encodes a 42.0kd protein. These predicted proteins have a homology of 78% (Figure 2). To obtain antibodies against NodE protein we constructed plasmid pMP1150 (Figure 3B) in which the indicated 3′ part of the *nodE* gene of *R.trifolii* is fused in frame to the 5′ terminal part of the *Escherichia coli lacZ* gene present in pIC20H (Marsh *et al.*, 1984). *Escherichia*

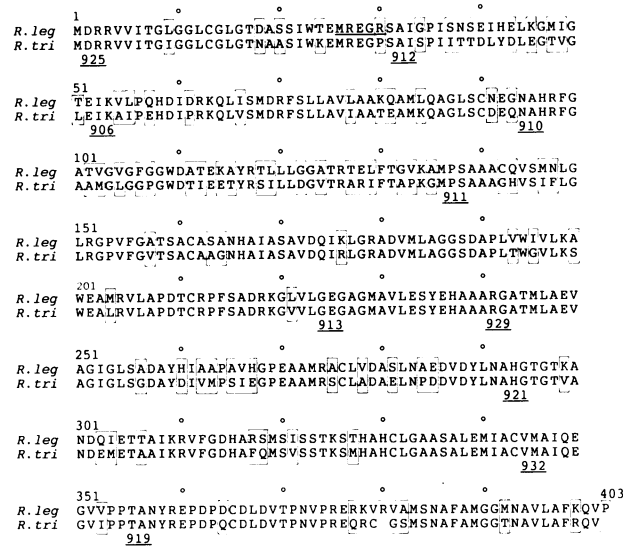

Fig. 2. Comparison of the NodE proteins of *R.leguminosarum* and *R.trifolii* and localization of the fusion site in NodE hybrids. The non-conserved amino acids in both proteins are indicated by boxes. The indicated *nodE* hybrid genes (underlined numbers) produce products with a N-terminal moiety of the *R.leguminosarum* NodE protein. The underlined amino acids at positions 27−31 are corrections on the data of Shearman *et al.* (1986).

coli strain JM101 containing plasmid pMP1150 produced an isopropyl-β-D-galactopyranoside (IPTG)-inducible protein which migrates as a 47 kd protein in SDS−PAGE. Since this was the only detected IPTG-inducible protein produced by this strain, we isolated this protein for immunization of a rabbit although the predicted mol. wt of the expected fusion protein was only 38 kd. Using the procedure described in Materials and methods, serum was obtained that reacted specifically with the produced fusion protein in *E.coli* on immunoblots (Figure 4, lane 1). This serum also reacted with flavonoid-inducible proteins produced by the *R.trifolii* wild-type strain ANU843 and *R.leguminosarum* wild-type strain 248. These proteins migrate as 51 and 45 kd proteins respectively in SDS−PAGE (Figure 4, lanes 5 and 7). The serum also reacted very weakly with a few non-inducible proteins in both strains.

To test whether the detected flavonoid-inducible proteins indeed are the predicted NodE proteins of *R.trifolii* and *R.leguminosarum* we tested several *nodE*::Tn*5* mutants derived from *R.trifolii* strain ANU843 and *R.leguminosarum* strain 248. In the *R.trifolii* Tn*5* mutant strains K11, H7, C7 and ANU297 and *R.leguminosarum* Tn*5* mutant strains K11, H7, C7 and ANU297 and *R.leguminosarum* Tn*5* mutant strains RBL601 and RBL605 no flavonoid-inducible proteins were detected by immunoblotting (results not shown) showing that the major reaction of the serum is with the NodE proteins. Although the predicted NodE proteins are very homologous, the mobilities of the proteins in SDS−PAGE are considerably different and do not agree with the predicted mol. wts for both proteins of 42.0 kd. When the NodE proteins of *R.trifolii* and *R.leguminosarum* were produced in *E.coli* (Figure 4, lanes 2 and 3) using the transcriptional fusion plasmids pMP1154 and pMP1210 (Figure 3) the same differences in electrophoretic mobilities were observed as in *Rhizobium*.

The serum against NodE was used to analyze the presence of homologous flavonoid-inducible proteins in other fast-

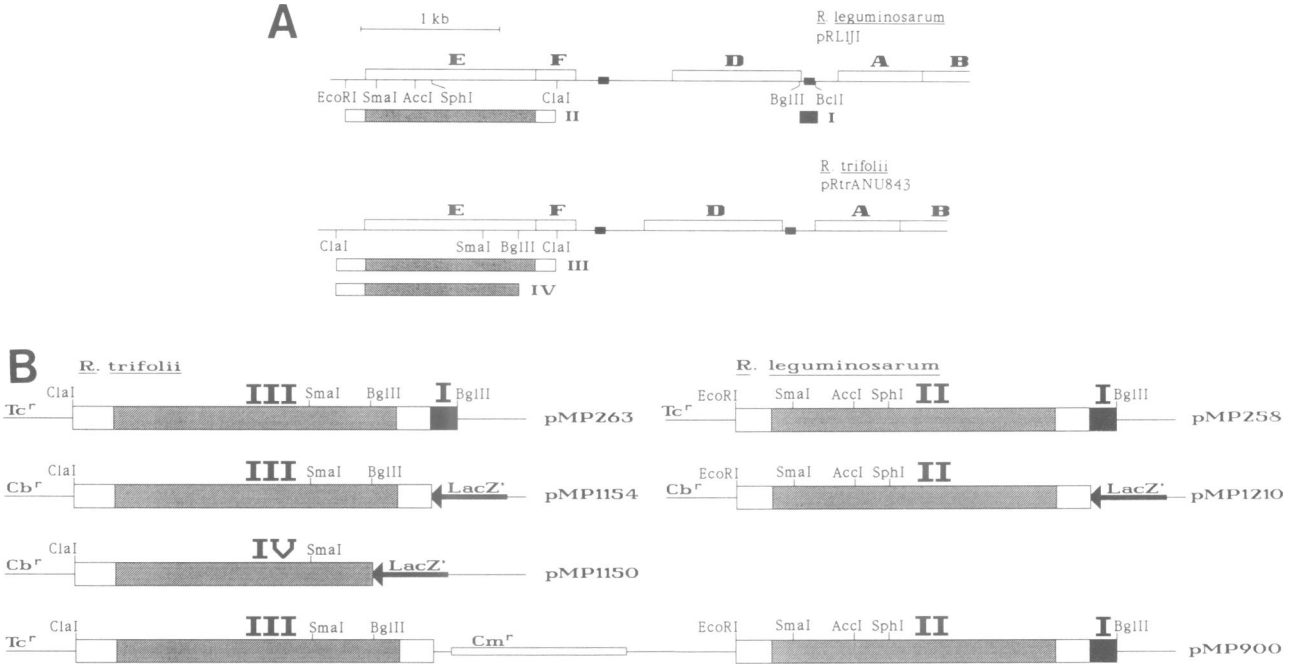

Fig. 3. Construction of plasmids. (**A**) Restriction fragments derived of the *R.leguminosarum* and *R.trifolii nod* region used in this study. The used restriction fragments are designated I, II, III and IV. The position of the restriction sites and *nod* genes is according to published data (Schofield and Watson, 1986; Spaink *et al.*, 1989) and the data of Figure 1. *nod* boxes are indicated by solid boxes upstream of the inducible *nod* operons. (**B**) Constructed plasmids. The restriction fragments indicated in (A) were used. The restriction fragments of plasmids pMP1154 and pMP1210 were cloned in the vector pIC20H (Marsh *et al.*, 1984) resulting in an in-frame fusion of the *lacZ* 5' part and the *nodF* gene as confirmed by nucleotide sequencing. Other fragments were cloned in the vector pMP92. The fragment containing the gene encoding chloramphenicol acetyl transferase (Cm^r) was derived from the vector pMP190 (Spaink *et al.*, 1987a).

growing *Rhizobium* strains. The additional four *R.trifolii* strains and eight *R.leguminosarum* strains tested (Table II) also produced a naringenin-inducible protein with apparent mol. wts of 51 kd and 44−45 kd for the *R.trifolii* and *R.leguminosarum* proteins respectively (results not shown). A protein that was inducible by luteolin, the reported inducer of the *nod* genes of *R.meliloti* (Peters *et al.*, 1986) and which migrates as a 43 kd protein in SDS−PAGE, was detected in *R.meliloti* strain 102F28 (results not shown). Using the same test procedure, neither this protein nor any other inducible proteins were detected in the two other *R.meliloti* strains tested, 2011 and LPR2.

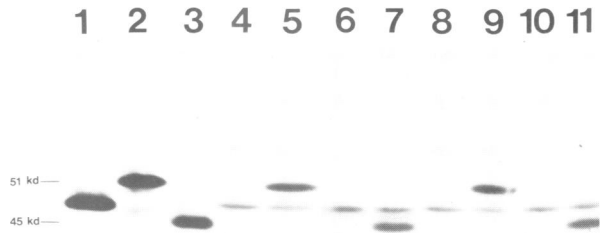

Fig. 4. Immunological detection of the NodE proteins of *R.trifolii* strain ANU843 and *R.leguminosarum* strain 248. **Lanes: 1**, *E.coli* containing pMP1150 plus IPTG; **2**, *E.coli* containing pMP1154 plus IPTG; **3**, *E.coli* containing pMP1210 plus IPTG; **4**, strain ANU843 without inducer; **5**, strain ANU843 plus naringenin; **6**, strain 248 without inducer; **7**, strain 248 plus naringenin; **8**, K11.263 without inducer; **9**, K11. pMP263 plus naringenin; **10**, K11.pMP258 without inducer; **11**, K11.pMP258 plus naringenin. Induction of *nod* gene transcription was performed as described previously (Spaink *et al.*, 1987a). For *Rhizobium*, protein samples produced by equal cell numbers (2×10^8) were used. The calculated apparent mol. wts of the detected inducible proteins are indicated at the edge of the figure.

Intracellular localization of the NodE protein

Cells of *R.leguminosarum* strain 248 and of strain RBL1246, the latter containing the Sym plasmid of *R.trifolii* strain ANU843, were grown in the presence of the inducer naringenin and were fractionated into outer membrane, cytoplasmic membrane, periplasmic and cytoplasmic components. Quantities of these fractions and the components of the spent growth medium, each derived of equal cell numbers, were analyzed on an immunoblot using anti-NodE antibodies. The results (Figure 5) showed that the NodE proteins of *R.leguminosarum* and *R.trifolii* were only detectable in the cytoplasmic membrane. When five times larger quantities of these components were analyzed, also a relatively small amount (<5%) of the NodE protein was detected in the outer membrane fraction (results not shown). NodE protein could not be removed from the cytoplasmic membrane fraction by incubation of this fraction with 1 M solutions of KCl or NaCl and subsequent purification on a sucrose gradient, indicating that the presence of NodE protein in this cellular fraction is not the result of an aspecific binding by ionic interactions to this membrane. In conclusion, the NodE proteins of *R.leguminosarum* and *R.trifolii* are localized in the cytoplasmic membrane.

According to the algorithm of von Heijne (1986) the predicted NodE proteins do not contain a signal sequence. The occurrence of non-polar transbilayer helices in NodE was analyzed using the algorithms of Engelman *et al.* (1986) and Chou and Fasman (1978) (Figure 6). According to the polarity scale of Engelman *et al.*, (1986), an α-helical stretch

of 20 amino acids with a free energy of transfer from water to oil of < -20 kcal/mol is a probable transmembrane protein structure. On this basis the NodE proteins of *R.trifolii* and *R.leguminosarum* are predicted to have at least two common transmembrane helices. These start at position 154 and 329 in both proteins. The NodE protein of *R.leguminosarum* is predicted to have an additional transmembrane helix starting at position 72. The other apolar regions in the NodE proteins with an energy of transfer of < -20 kcal/mol are probably not α-helical structures and therefore no predictions can be made about a possible transmembrane character of these regions.

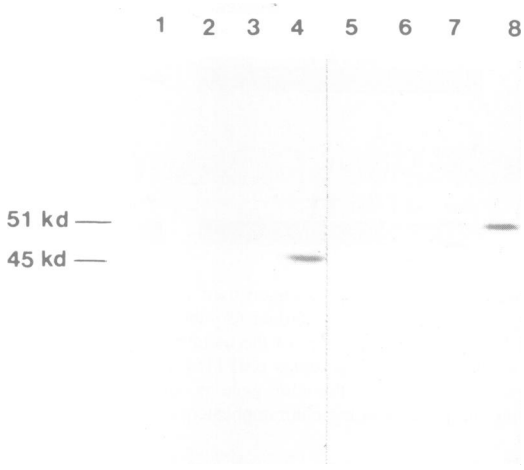

Fig. 5. Intracellular localization of the NodE protein of *R.leguminosarum* and *R.trifolii*. **Lanes 1–4** are fractions from *R.leguminosarum* strain 248. **Lanes 5–8** are fractions derived from strain RBL1246 containing the *R.trifolii* Sym plasmid of strain ANU843. Protein fractions derived of equal cell numbers were analyzed. **Lanes: 1,5** medium components; **2,6** cytoplasmic plus periplasmic fractions; **3,7** outer membrane fractions; **4,8** cytoplasmic membrane fractions. The calculated apparent mol. wts of the detected proteins are indicated at the edge of the figure.

NodE of R.leguminosarum and R.trifolii as a determinant of host specificity

To separate the *nodE* genes from the *nodF* genes, the latter ones preceding the *nodE* genes in the respective *nodFE* operons, the IncP plasmids pMP263 and pMP258 (Figure 3) were constructed. In these plasmids the NodE proteins of *R.trifolii* and *R.leguminosarum* respectively are expressed under control of the flavonoid-inducible promoter of the *nodABCIJ* operon of *R.leguminosarum*. Since identical regulatory regions are present in pMP263 and pMP258, these plasmids are well suited to compare the influence of respective *R.trifolii* and *R.leguminosarum nodE* genes on the nodulation process. Moreover, the promoter of the *nodABCIJ* operon, which controls the expression of a common nodulation operon, was used to rule out a possible host-specific influence of the *nodFE* upstream region. The plasmids pMP263 and pMP258 were introduced into the *R.trifolii nodE* mutant strain K11. The resulting strains produce the NodE proteins of *R.trifolii* and *R.leguminosarum* respectively, after induction with naringenin (Figure 4, lanes 9 and 11) in amounts that are similar to those of the induced wild-type strains (Figure 4, lanes 5 and 7).

The derivatives of strain K11, containing the plasmids pMP263, pMP258 and the control pMP92 in which no *nodE* gene was cloned, were tested for nodulation on several host plants of either *R.trifolii* or *R.leguminosarum*. The results (Table I) show that the *nodE* mutant strain K11 itself is severely affected in its nodulation ability on its original host plants. Its properties vary from an almost complete inability to nodulate the host, e.g. in the case of *Trifolium pratense*, to a reduction in nodulation frequency by 75% after 3 weeks of inoculation in the case of *Trifolium repens*. With the exception of nodulation of *Pisum sativum*, mutant strain K11 has not acquired the ability to nodulate hosts of *R.leguminosarum* to a considerable extent. Identical results were obtained using the *R.trifolii nodE* mutant strains ANU258, H7, C7 and ANU297 (results not shown). As expected, plasmid pMP263 restored the mutant phenotype

Table I. Host-specific properties of *R.trifolii nodE* mutant strain K11 containing a cloned *nodE* gene of *R.trifolii* or *R.leguminosarum*

Tested plant species	Nodulation characteristics[a] and average number of nodules per nodulated plant with each strain			
	ANU843 (*R.trifolii* wild-type)	K11.pMP92 (*nodE*::Tn*5*)	K11.pMP263 (*nodE R.trifolii*)	K11.pMP258 (*nodE R.leguminosarum*)
Trifolium				
T.arvense	+ (5)	–	+ (5)	–
T.campestre	+ (2)	–	+ (3)	–
T.lappaceum	+ (3)	–	+ (3)	–
T.pratense	+ (5)	–	+ (6)	–
T.repens	+ (8)	± (3)	+ (11)	± (4)
Vicia				
V.hirsuta	–	–	–	+ (10)
V.sativa	–	–	–	+ (4)
Lathyrus				
L.montanum	–	–	–	+ (2)
L.pratensis	–	–	–	+ (3)
Pisum				
P.sativum	–	+ (15)	–	+ (20)

[a] –, 0–15% of plants nodulated; ±, 20–40% of plants nodulated; +, >80% of plants nodulated.

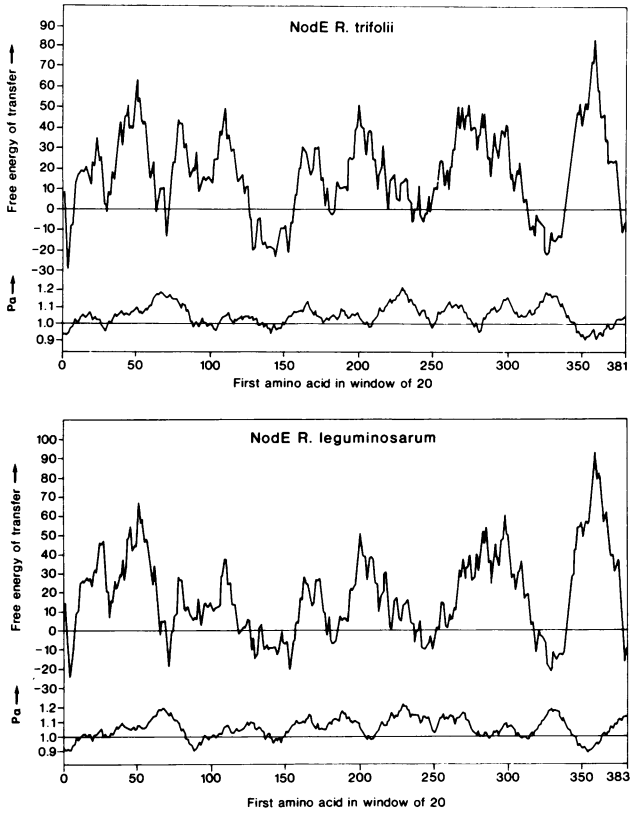

Fig. 6. Prediction of nonpolar transbilayer helices in NodE. The NodE proteins of *R.leguminosarum and R.trifolii* were scanned with the algorithm of Engelman *et al.* (1986) and Chou and Fasman (1978) using a window of 20 amino acids. The polarity values according to Engelman *et al.*, are indicated with the free energy of transfer from water to oil (in kcal/mol) on the vertical axis of the upper panels. The Chou and Fasman conformational parameters are indicated with Pα on the vertical axis of the lower panels. A value of Pα of >1 favors an α-helical conformation of the analyzed stretch of amino acids.

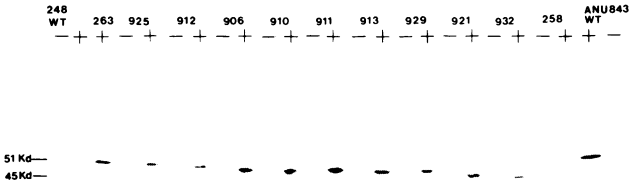

Fig. 7. Immunological detection of hybrid NodE proteins. The hybrid *nodE* genes present in strain K11 were tested in the absence (−) and presence (+) of the inducer naringenin. Protein samples were derived from 5×10^7 cells except in the case of the strains containing plasmids pMP921, pMP932 and pMP258 in which case 1×10^8 cells were used. The calculated apparent mol. wts of the parental NodE proteins are indicated at the edge of the figure.

of strain K11 to that of the wild-type strain ANU843 (Table I). In strong contrast, the NodE product of *R.leguminosarum* produced by pMP258 confers on strain K11 the ability to nodulate efficiently on the hosts of *R.leguminosarum* tested, whereas the nodulation phenotype of strain K11 on the *R.trifolii* host plants is not altered by the presence of this plasmid. We therefore conclude that in isogenic *R.trifolii* strains the source of the *nodE* gene determines whether *R.leguminosarum* or *R.trifolii* host plants are efficiently nodulated.

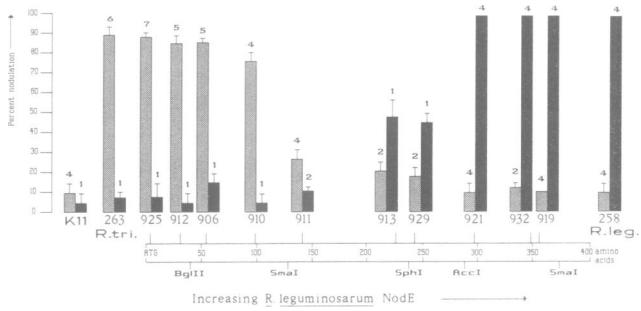

Fig. 8. Host-specific properties of *nodE* hybrid genes present in *R.trifolii nodE* mutant strain K11. The sites of recombination of the hybrid *nodE* genes are indicated on the bottom of the figure together with the restriction sites used for the initial classification of the hybrid genes. Hatched bars indicate nodulation frequencies on *T.pratense*, black bars indicate nodulation frequencies on *V.sativa*. The average numbers of nodules per nodulated plant are indicated above the bars. Standard deviations from two independent experiments are indicated. Also indicated are the results with strain K11 containing the parental *R.trifolii* and *R.leguminosarum nodE* genes or in the absence of the *nodE* gene.

Localization of the region of the NodE protein that determines the difference in host specificity

To localize one or more regions of the NodE proteins that are involved in the determination of the difference in host specificity we constructed hybrid *nodE* genes by *in vivo* homologous recombination. To obtain hybrid *nodE* genes we have constructed pMP900 (Figure 3B). This plasmid contains the *nodE* genes of *R.leguminosarum* and *R.trifolii* located in direct repeat, with a unique *Eco*RI restriction site and the gene coding for chloramphenicol-acetyl-transferase (CAT) located in between these genes. Homologous recombination between these *nodE* genes located in trandem will result in plasmids that contain a hybrid *nodE* gene and no longer possess the *Eco*RI restriction site and the CAT gene. Consequently, *Eco*RI endonuclease treatment of plasmid preparations of pMP900, followed by transformation, enriches for plasmids that contain hybrid *nodE* genes. Using plasmid pMP900 a set of 40 *nodE* hybrid genes was obtained that contain a 5′ moiety of the *R.leguminosarum nodE* gene and a 3′ moiety of the *R.trifolii nodE* gene. These hybrid *nodE* genes could be divided into five classes using the restriction sites indicated in Figure 3B. The recombination site of three representatives of each class was localized by nucleotide sequencing. Several hybrid genes appeared to have indentical recombination sites and of such hybrids only one representative was investigated further (Figure 1). In Figure 2 the positions of these recombination sites are indicated together with the deduced protein sequences of both the *R.leguminosarum* and *R.trifolii nodE* products. No deletions, duplications or multiple cross-overs were detected in any of the sequenced hybrid *nodE* genes.

The representative *nodE* hybrid genes were introduced in the *R.trifolii nodE* mutant strain K11 and the resulting strains, grown in the absence and presence of the inducer naringenin, were analyzed on immunoblots using antiserum against NodE protein. The results (Figure 7) indicate that all tested *nodE* hybrid-containing strains produce an inducible NodE protein in similar amounts. The electrophoretic mobilities of these products indicate that several regions of the NodE protein are involved in determining the difference of

Table II. Bacterial strains

Designation	Characteristics	Source of reference
E.coli		
KMBL1164	*del(lac-pro), thi*,F$^-$	P.van der Putte, Leiden
JM101	*del(lac-pro), supE, thi,* (F'*traD36, ProAB, LacI^{q}ZdelM15*)	Messing and Vieira (1982)
R.trifolii		
ANU843	wild-type	Djordjevic *et al.* (1985)
RCR5	wild-type	Hooykaas *et al.* (1982)
RBL52, RBL53	wild-type	A.A.N.van Brussel, Leiden
0403	wild-type	Gardiol and Dazzo (1986)
ANU258, ANU297	ANU843 *nodE*::Tn*5*	Djordjevic *et al.* (1985)
K11, H7, C7	ANU843 *nodE*::Tn*5*	Djordjevic *et al.* (1985)
R.leguminosarum		
248	wild-type	Josey *et al.* (1979)
128C53	wild-type	Brewin *et al.* (1980)
K1a, k1b, K2d	wild-type	de Maagd *et al.* (1989b)
RCC1012, RCC1044	wild-type	de Maagd *et al.* (1988)
LPR1	wild-type	Hooykaas *et al.* (1982)
PRE	wild-type	Winarno and Lie (1979)
RBL601, RBL605	pRL1JI, *nodE*::Tn*5*	Wijffelman *et al.* (1985)
RBL1387	248 cured of sym plasmid	Priem and Wijffelman (1984)
RBL1246	RBL1387 pRtrANU843::Tn*5*	E.Pees, Leiden
R.meliloti		
102F28	wild-type	S.Long, Stanford
2011	SU47 Sm^r	Debellé and Sharma (1986)
LPR2	wild-type	Hooykaas *et al.* (1982)

electrophoretic mobility of the *R.leguminosarum* and *R.trifolii* NodE proteins.

The derivatives of *R.trifolii* strain K11, containing the representative *nodE* hybrid genes or the parental *nodE* genes, were tested for their nodulation ability on *T.pratense* and *Vicia sativa*. The results (Figure 8) show that hybrids which contain a recombination site in the 5' part of the *nodE* gene, bordered by the recombination site in *nodE910*, have a nodulation phenotype that is not significantly different from that of the *R.trifolii* parental *nodE* gene. Hybrids which contain a recombination site in the 3' part of the *nodE* gene, bordered by the recombination site in *nodE921*, have a phenotype that is almost identical to that of the *R.leguminosarum* parental *nodE* gene. These results indicate that the 47 non-conserved amino acids located in the 94 amino acid long N terminal region and 117 amino acid long C terminal region do not contribute to the difference in the host specificity of the NodE proteins of *R.trifolii* and *R.leguminosarum*. The hybrids *nodE911*, *nodE913* and *nodE929*, which resulted from a recombination in the central region of nodE confer a phenotype that is different from that caused by both the parental *nodE* genes. Application of the non-parametric test of Mann and Whitney (1947) with $\alpha = 0.05$ showed that the observed nodulation frequencies were also significantly different from that of strain K11 lacking a *nodE* gene-containing plasmid. This indicates that the NodE proteins produced by these hybrid *nodE* genes have still some beneficial effect in a host-specific way on the nodulation process. In conclusion, a central region of the NodE protein of 185 amino acids at the most, containing 44 non-conserved amino acids, is involved in the determination of the difference in host specificity conferred by NodE proteins.

Discussion

We have shown that the NodE proteins of *R.trifolii* and of *R.leguminosarum* are a crucial factor in the determination of the host range of nodulation. These proteins appear to be very homologous both in their predicted primary (Figure 2) and secondary (Figure 6) structures. By studying the properties of a series of hybrid *nodE* genes we showed that only a central region of the NodE protein, containing 44 non-conserved amino acids, is involved in the determination of the differences in host-specific characteristics between the NodE proteins of *R.trifolii* and *R.leguminosarum* (Figure 8).

In addition to their different host-specific characteristics, the NodE proteins of *R.trifolii* strain ANU843 and *R.leguminosarum* strain 248 are also different in their mobility during SDS–PAGE. It is very unlikely that this difference in mobility is the result of a difference in post-translational modification of these NodE proteins since (i) the NodE proteins produced in *E.coli* have the same difference in electrophoretic mobility and (ii) this difference appears to be caused by several regions of the NodE protein as concluded from the electrophoretic mobilities of the hybrid NodE proteins. A more likely explanation is that the 91 amino acids that are different between the NodE proteins have a strong influence on the migration rate during SDS–PAGE. This explanation is consistent with the observation that the substitution of only one amino acid already can alter the electrophoretic mobility of a protein substantially (Noel *et al.*, 1979). Surprisingly, the NodE

homologues of four other *R.trifolii* strains and eight other tested *R.leguminosarum* strains appeared to be very similar in their electrophoretic mobilities to the NodE proteins of *R.trifolii* strain ANU843 and *R.leguminosarum* strain 248 respectively. These results suggest that the difference in apparent mol. wts of the NodE proteins of *R.trifolii* and *R.leguminosarum* is a general feature for these species and it therefore may reflect a strong structural conservation of these NodE proteins. In contrast to the results with *R.meliloti* strain 102F28, a NodE protein was not detected in *R.meliloti* strain 2011 although this strain contains a functional *nodE* gene that is very homologous to the *nodE* genes of *R.leguminosarum* and *R.trifolii* (Fisher *et al.*, 1987). This result may be caused by the low level of expression of the inducible *nod* genes in this strain (Mulligan and Long, 1985).

The biochemical function of NodE protein is unknown. We have not found any significant homology of the NodE protein with other protein sequences present in the NBRF or EMBL data libraries. To obtain a clue to the biochemical function we have determined the localization of the NodE protein in the *Rhizobium* cell. From experiments using polyclonal antibodies against the NodE protein we conclude that the NodE protein is localized in the cytoplasmic membrane of the bacterial cell. Furthermore, application of the algorithm of Engelman *et al.* (1986) suggested that the NodE protein is a transmembrane protein with regions present in the cytoplasmic and periplasmic compartments. The host-specific reaction towards the plant therefore is most probably transduced by means of a product that is specifically produced, modified or transported by the NodE protein. The strong homology of the NodE proteins suggest that they are involved in a very similar biochemical process. The localization of the *nodF* and *nod*E genes in one operon suggests that the proteins coded by these two genes are involved in this same process. This suggestion is supported by the observation that the efficiency of translation of the *nodE* gene is dependent on the translation of the *nodF* gene both in *R.trifolii* and *R.leguminosarum* (H.P.Spaink, unpublished results). This translational coupling was already suggested by the positions of the stop codon of *nodF* and the start codon of *nodE* (Gold, 1988). The predicted product of the *nodF* gene, which precedes the *nodE* gene in the *nodFE* operon, has been shown to be homologous with the *E.coli* acyl carrier protein (ACP) (Shearman *et al.*, 1986), especially with the phosphopantetheine binding region. Furthermore, NodF protein is very hydrophilic (data not shown) and therefore probably, like the ACP, localized in the cytoplasm. The ACP protein has been shown also to function as a transglucosylase factor involved in the synthesis of β1,2-linked glucans (Therisod and Kennedy, 1987; Therisod *et al.*, 1986). These β1,2-linked glucans constitute the backbone of the membrane-derived oligosaccharides of *E.coli* which very much resemble the cyclic β1,2-linked glucans found in several rhizobia and *Agrobacterium* (e.g. see Amemura *et al.*, 1983; Dell *et al.*, 1983; Batley *et al.*, 1987; Miller *et al.*, 1987). The production of the known cyclic β1,2-linked glucans in *R.meliloti* does not require the induction of the *nod* genes but is dependent on the identified chromosomal *ndvA* and *ndvB* loci (Stanfield *et al.*, 1988). However, NodF and NodE protein could be involved in the production, modification or transport of a host-specific, not yet identified, β1,2-linked glucan. The important role of the host plant lectins, which are specific sugar-binding proteins, in the determination of host specificity (Diaz *et al.*, 1989) also indicates that the NodE protein is involved in the production of a host-specific signal of carbohydrate nature.

Materials and methods

Bacterial strains and plasmids

Bacterial strains are listed in Table II. Strain RBL1246 was constructed by mobilization of the Sym plasmid of strain ANU251 (Djordjevic *et al.*, 1986) to strain RBL1387, a cured derivative of *R.leguminosarum* strain 248, using pRL5522 as a helper plasmid. The construction of plasmids is described in Figure 3. pMP92 is a IncP vector (Spaink *et al.*, 1987a) derived from pTJS75 (Schmidhauser and Helinksi, 1985). Plasmids were mobilized from *E.coli* strain KMBL1164 to *Rhizobium* using pRK2013 as a helper plasmid (Ditta *et al.*, 1980) as described previously (Spaink *et al.*, 1987a). Hybrid *nodE* containing plasmids were constructed from pMP900 (Figure 3B) according to the methods described previously (Spaink *et al.*, 1989) using the restriction endonuclease *Eco*RI to enrich for hybrid *nodE* containing plasmids. The Cm^r gene present in pMP900 was derived from pMP190 (Spaink *et al.*, 1987a).

DNA manipulation and sequencing

Recombinant DNA techniques were carried out as described by Maniatis *et al.* (1982). DNA sequencing was performed using the dideoxy chain termination method (Sanger *et al.*, 1977) using the M13 vectors system (Messing and Vieira, 1982). Sequence gels were supplemented with 40% formamide to prevent band compressions. The site of recombination of *nodE* hybrid genes was determined using four oligonucleotides of 20 nucleotides long, indicated in Figure 1, which were a gift from Mogen International NV (Leiden, The Netherlands). Enzymes, M13 DNA primer and unlabeled nucleotides were obtained from Boehringer (Mannheim, FRG), freeze-dried large fragment (Klenow) of DNA polymerase I was obtained from BRL (Gaithersburg, MD, USA) and [α-^{35}S] dCTP was purchased from Amersham International plc (Amersham, UK).

Plant tests

Plant tests were performed on solid agar medium as described previously, except for *P.sativum* plants which were grown on gravel (van Brussel *et al.*, 1982). Nodulation tests were scored 18 days after inoculation. Plant seeds were gifts of the Botanical Garden of Leiden University (The Netherlands) and the Zentral Institut für Genetik und Kulturpflanzenforschung (Gaterleben, DDR). Seeds from *V.sativa* subspecies *nigra* were harvested in this laboratory.

Preparation of antibodies

Cells of strain JM101 harboring plasmid pMP1150 (Figure 3B), produce a protein with an apparent mol. wt of 47 kd under control of the *E.coli lacZ* promoter. To obtain large quantities of the presumed β-galactosidase−NodE fusion protein, these cells were grown in the presence of 20 mg/l isopropyl-β-D-galactopyranoside (IPTG). Subsequently, the cells were fractionated into membrane components (insoluble fraction) and cytoplasmic/periplasmic components (soluble fraction) as described previously (de Maagd and Lugtenberg, 1986). The insoluble fraction was resuspended and incubated overnight in a solution of 5 mM EDTA and 0.1% Nonidet. The resulting insoluble fraction, ~50% of which consisted of the fusion protein, was electrophoresed through an 11% SDS−polyacrylamide gel. After a short staining period with Coomassie Brilliant Blue, the part of the gel that contained the fusion protein was excised and the fusion protein was removed from the gel slice by electro-elution. One hundred micrograms of the fusion protein obtained by this method was injected into a rabbit at intervals using methods described previously (de Maagd *et al.*, 1989a). The serum of the rabbit contains antibodies with a titer of 10^5 against the apparent 47 kd fusion protein as determined by testing serial dilutions on immunoblots.

Electrophoresis and immunoblotting

SDS−PAGE was performed in 11% polyacrylamide gels (Lugtenberg *et al.*, 1975). The mol. wt markers ovalbumin (45 kd) and glutamate dehydrogenase (55 kd) were obtained from Sigma (St Louis, MO, USA). Proteins were electrophoretically transferred to nitrocellulose sheets and the blots were incubated with 4000-fold diluted antibodies and washed as described previously (de Maagd *et al.*, 1989a). Bound antibodies were visualized using alkaline phosphatase-conjugated goat anti-rabbit immunoglobulin G, obtained from Sigma according to the method described by Ey and Ashman (1986).

Fractionation of Rhizobium cells

Cells of *R.leguminosarum* strain 248 or derivatives of this strain were fractionated into outer membrane, cytoplasmic membrane, cytoplasmic and periplasmic fractions according to the methods described previously (de Maagd and Lugtenberg, 1986). *nod* genes were induced by the addition of 400 nM naringenin to the growth medium. Efficiency of membrane separation was tested by assaying NADH oxidase activity and 2-keto-3-deoxyoctonate (KDO) content, markers for cytoplasmic membrane and outer membrane respectively.

Acknowledgements

We thank Henk Roest and Jan Verdoes for their help in part of the experiments and Ruud de Maagd for critically reading the manuscript and valuable discussions. We also thank Mogen International NV (Leiden, The Netherlands) for the kind gift of oligonucleotides.

References

Amemura,A., Hisamatsu,M. and Mitani,H. (1983) *Carbohydr. Res.*, **114**, 277–285.

Batley,M., Redmond,J.W., Djordjevic,S.P. and Rolfe,B.G. (1987) *Biochim. Biophys. Acta*, **901**, 119–126.

Brewin,N.J., De Jong,T.M., Phillips,D.A. and Johnston,A.W.B. (1980) *Nature*, **288**, 77–79.

Chou,P.Y. and Fasman,G.D. (1978) *Adv. Enzymol.*, **47**, 46–148.

de Maagd,R.A. and Lugtenberg,B.J.J. (1986) *J. Bacteriol.*, **167**, 1083–1085.

de Maagd,R.A., Van Rossum,C. and Lugtenberg,B.J.J. (1988) *J. Bacteriol.*, **170**, 3782–3785.

de Maagd,R.A., de Rijk,R., Mulders,I.H.M. and Lugtenberg,B.J.J. (1989a) *J. Bacteriol.*, **171**, 1136–1142.

de Maagd,R.A., Spaink,H.P., Pees,E., Mulders,I.H.M., Wijfjes,A., Wijffelman,C.A., Okker,R.J.H. and Lugtenberg,B.J.J. (1989b) *J. Bacteriol.*, **171**, 1151–1157.

Debellé,F. and Sharma,S.B. (1986) *Nucleic Acids Res.*, **14**, 7453–7472.

Dell,A., York,W.S., McNeil,M., Darvill,A.G. and Albersheim,P. (1983) *Carbohydr. Res.*, **117**, 185–200.

Diaz,C.L., Melchers,L.S., Hooykaas,P., Lugtenberg,B.J.J. and Kijne,J.W. (1989) *Nature*, **338**, 579–581.

Ditta,G., Stanfield,S., Corbin,D. and Helinski,D.R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 7347–7351.

Djordjevic,M.A., Schofield,P.R. and Rolfe,B.G. (1985) *Mol. Gen. Genet.*, **200**, 463–471.

Djordjevic,M.A., Innes,R.W., Wijffelman,C.A., Schofield,P.R. and Rolfe,B.G. (1986) *Plant Mol. Biol.*, **6**, 389–401.

Djordjevic,M.A., Gabriel,D.W. and Rolfe,B.G. (1987) *Annu. Rev. Phytopathol.*, **25**, 145–168.

Downie,J.A., Ma,Q.S., Knight,C.D., Hombrecher,G. and Johnston,A.W.B. (1983) *EMBO J.*, **2**, 947–952.

Engelman,D.M., Steitz,T.A. and Goldman,A. (1986) *Annu. Rev. Biophys. Chem.*, **15**, 321–353.

Ey,P.L. and Ashman,L.K. (1986) *Methods Enzymol.*, **121**, 497–509.

Faucher,C., Maillet,F., Vasse,J., Rosenberg,C., van Brussel,A.A.N., Truchet,G. and Dénarié,J. (1988) *J. Bacteriol.*, **170**, 5489–5499.

Firmin,J.L., Wilson,K.E., Rossen,L. and Johnston,A.W.B. (1986) *Nature*, **324**, 90–92.

Fisher,R.F., Swanson,J.A., Mulligan,J.T. and Long,S.R. (1987) *Genetics*, **117**, 191–201.

Gardiol,A.E. and Dazzo,F.B. (1986) *J. Bacteriol.*, **168**, 1459–1462.

Gold,L. (1988) *Annu. Rev. Biochem.*, **57**, 199–233.

Hooykaas,P.J.J., Schnijdewindt,F.G.M. and Schilperoort,R.A. (1982) *Plasmid*, **8**, 73–82.

Horvath,B., Kondorosi,E., John,M., Schmidt,J., Török,I., Györgypal,Z., Barabas,I., Wieneke,U., Schell,J. and Kondorosi,A. (1986) *Cell*, **46**, 335–343.

Horvath,B., Bachem,C.W., Schell,J. and Kondorosi,A. (1987) *EMBO J.*, **6**, 841–848.

Innes,R.W., Kuempel,P.L., Plazinksi,J., Canter Cremers,H.C.J., Rolfe,B.G. and Djordjevic,M.A. (1985) *Mol. Gen. Genet.*, **201**, 426–432.

Josey,D.P., Beynon,J.L., Johnston,A.W.B. and Beringer,J.E. (1979) *J. Appl. Bacteriol.*, **46**, 343–350.

Leigh,J.A. and Lee,C.C. (1988) *J. Bacteriol.*, **170**, 3327–3332.

Lugtenberg,B.J.J., Meyers,J., Peters,R., van der Hoek,P. and van Alphen,L. (1975) *FEBS Lett.*, **58**, 254–258.

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

Mann,H.B. and Whitney,D.R. (1947) *Math. Stat.*, **18**, 50–60.

Marsh,J.L., Erfle,M. and Wykes,E.J. (1984) *Gene*, **32**, 481–485.

Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.

Miller,K.J., Reinhold,V.N., Weissborn,A.C. and Kennedy,E.P. (1987) *Biochim. Biophys. Acta*, **901**, 112–118.

Mulligan,J.T. and Long,S.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6609–6613.

Noel,D., Nikaido,K. and Ames,G.F. (1979) *Biochemistry*, **18**, 4159–4165.

Peters,N.K., Frost,J.W. and Long,S.R. (1986) *Science*, **233**, 977–980.

Priem,W.J.E. and Wijffelman,C.A. (1984) *FEMS Microbiol. Lett.*, **25**, 247–251.

Redmond,J.W., Batley,M., Djordjevic,M.A., Innes,R.W., Kuempel,P.L. and Rolfe,B.G. (1986) *Nature*, **323**, 632–635.

Rostas,K., Kondorosi,E., Horvath,B., Simoncsits,A. and Kondorosi,A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1757–1761.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schmidhauser,T.J. and Helinski,D.R. (1985) *J. Bacteriol.*, **164**, 446–455.

Schofield,P.R. and Watson,J.M. (1986) *Nucleic Acids Res.*, **14**, 2891–2903.

Shearman,C.A., Rossen,L., Johnston,A.W.B. and Downie,J.A. (1986) *EMBO J.*, **5**, 647–652.

Spaink,H.P., Okker,R.J.H., Wijffelman,C.A., Pees,E. and Lugtenberg,B. (1986) In Lugtenberg,B. (ed.), *Recognition in Microbe–Plant Symbiotic and Pathogenic Interactions*. ASI Series, Vol. H4. Springer-Verlag, Berlin, pp. 55–68.

Spaink,H.P., Okker,R.J.H., Wijffelman,C.A., Pees,E. and Lugtenberg,B.J.J. (1987a) *Plant Mol. Biol.*, **9**, 27–39.

Spaink,H.P., Wijffelman,C.A., Pees,E., Okker,R.J.H. and Lugtenberg,B.J.J. (1987b) *Nature*, **328**, 337–340.

Spaink,H.P., Wijffelman,C.A., Okker,R.J.H. and Lugtenberg,B.J.J. (1989) *Plant Mol. Biol.*, **12**, 59–73.

Stanfield,S.W., Ielpi,L., O'Brochta,D., Helinski,D.R. and Ditta,G.S. (1988) *J. Bacteriol.*, **170**, 3523–3530.

Surin,B.P. and Downie,J.A. (1988) *Mol. Microbiol*, **2**, 173–184.

Therisod,H. and Kennedy,E.P. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8235–8238.

Therisod,H., Weissborn,A.C. and Kennedy,E.P. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7236–7240.

van Brussel,A.A.N., Tak,T., Wetselaar,A., Pees,E. and Wijffelman,C.A. (1982) *Plant Sci. Lett.*, **27**, 317–325.

von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.

Wijffelman,C.A., Pees,E., van Brussel,A.A.N., Okker,R.J.H. and Lugtenberg,B.J.J. (1985) *Arch. Microbiol.*, **143**, 225–232.

Wijffelman,C., Zaat,B., Spaink,H.P., Mulders,I., van Brussel,T., Okker,R., Pees,E., de Maagd,R. and Lugtenberg,B. (1986) In Lugtenberg,B. (ed.), *Recognition in Microbe–Plant Symbiotic and Pathogenic Interactions*. ASI Series, Vol. H4. Springer-Verlag, Berlin, pp. 123–136.

Winarno,R. and Lie,T.A. (1979) *Plant Soil.*, **51**, 135–142.

Received on March 23, 1989; revised on June 14, 1989.