JOURNAL OF BACTERIOLOGY, June 1992, p. 4026–4035
0021-9193/92/124026-10$02.00/0
Copyright © 1992, American Society for Microbiology
Vol. 174, No. 12

Molecular Characterization and Regulation of the Rhizosphere-Expressed Genes *rhiABCR* That Can Influence Nodulation by *Rhizobium leguminosarum* Biovar viciae

M. T. CUBO,[1]† A. ECONOMOU,[1] G. MURPHY,[2] A. W. B. JOHNSTON,[1]‡ AND J. A. DOWNIE[1]*

John Innes Institute[1] and Cambridge Laboratory,[2] John Innes Centre, Colney Lane, Norwich NR4 7UH, United Kingdom

Received 27 November 1991/Accepted 15 April 1992

A group of four *rhi* (rhizosphere-expressed) genes from the symbiotic plasmid of *Rhizobium leguminosarum* biovar viciae has been characterized. Although mutation of the *rhi* genes does not normally affect nodulation, in the absence of the closely linked nodulation genes *nodFEL*, mutations in the *rhi* genes can influence the nodulation of the vetch *Vicia hirsuta*. The DNA sequence of the *rhi* gene region reveals four large open reading frames, three of them constituting an operon (*rhiABC*) transcribed convergently toward the fourth gene, *rhiR*. *rhiABC* are under the positive control of RhiR, the expression of which is repressed by flavonoids that normally induce *nod* gene expression. This repression, which requires the *nodD* gene product (the transcriptional activator of *nod* gene expression), may be due to a *cis* effect caused by a high level of NodD-dependent expression from the adjacent *nodO* promoter, which is transcribed divergently from *rhiR*. RhiR shows significant similarities to a subfamily of transcriptional regulators that includes the LuxR and UvrC-28K proteins. RhiA shows limited homology to a short domain of the lactose permease, LacY, close to a region thought to be involved in substrate binding. No strong homologies were found for the other *rhi* gene products. It appears that RhiA and RhiB are cytoplasmic, whereas RhiC is a periplasmic protein, since it has a typical N-terminal transit sequence and a *rhiC-phoA* protein fusion expresses alkaline phosphatase activity. The biochemical role of the *rhi* genes has not been established, but it appears that they may play a role in the plant-microbe interaction, possibly by allowing the bacteria to metabolize a plant-made metabolite.

There are many bacterial genes involved in the interaction between rhizobia and their legume hosts. In *Rhizobium* spp. many of these genes are present on large symbiotic plasmids. Initially, these genes were identified by isolating mutants unable to fix nitrogen or form nodules, but several nodulation (*nod*) genes in which mutations have little or no effect on nodulation have now been identified. Currently, over 30 different *nod* genes have been identified among a wide variety of rhizobia (32), and in general they are under the control of positively acting transcriptional regulators encoded by *nodD* genes. Several of the *nod* gene products are involved in the biosynthesis of low-molecular-weight signalling molecules that are specifically recognized by legumes. It is now clear that different rhizobia make different but related signalling molecules which are substituted glycolipids consisting of acylated oligoglucosamine signal molecules (43, 51). The biosynthesis of these nodulation factors involves most of the *nod* gene products; different substitutions of the glycolipid, such as the presence of sulfate or the type of acyl group, are mediated by *nod* gene products and determine host specificity in the interaction between the bacterium and legume (43, 51).

In addition to the various *nod* genes and genes involved in nitrogen fixation, several other symbiotic-plasmid genes have been identified. For example, genes involved in the synthesis of melanin (3, 9) and bacteriocins (28) are found on symbiotic plasmids. In *Rhizobium meliloti*, Boivin et al. (4) identified genes involved in the catabolism of trigonelline, a secondary metabolite secreted by some legumes. Similarly, Murphy et al. (36) have identified *R. meliloti* genes that synthesize opinelike compounds in nodules; these compounds are thought to be secreted and stimulate rhizosphere growth of strains of *R. meliloti* that have the capacity to utilize the opinelike metabolite. van Egeraat (53) demonstrated that homoserine (a metabolite accumulated and secreted by pea roots) could be used as a carbon and nitrogen source by many pea-nodulating biovars of *Rhizobium leguminosarum* but not by biovars trifolii and phaseoli or by *R. meliloti*; Johnston et al. (29) showed that this characteristic is encoded on the symbiotic plasmid.

Another characteristic found in most strains of *R. leguminosarum* bv. viciae but absent from other biovars of *R. leguminosarum* is the ability to make large amounts of a protein with an M_r of 24,000 (17). This protein is expressed by bacteria in the rhizosphere of peas but not by nodule bacteroids. The gene encoding this protein was called *rhiA* and was found to be located within a 10-kb region of DNA between the *nod* genes and the nitrogenase genes of the symbiotic plasmid pRL1JI of *R. leguminosarum* bv. viciae (Fig. 1). The *rhiA* gene has been detected in all strains of *R. leguminosarum* bv. viciae tested and was absent from other biovars of *R. leguminosarum* or other *Rhizobium* strains tested (16, 17). Using a *rhiA-lacZ* fusion, Economou et al. (22) noted that the expression of the *rhiA* gene is decreased in the presence of flavonoids and that this effect depends on NodD, a transcriptional activator that induces *nod* gene expression in the presence of flavonoids.

Although the *rhiA* gene is specific for strains of *R. leguminosarum* bv. viciae, its expression is affected by flavonoids, and it is located adjacent to *nod* and *nif* genes, no

* Corresponding author.
† Present address: Departamento Microbiologia, Facultad de Biologia, Universidad de Sevilla, Seville 41080, Spain.
‡ Present address: School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, United Kingdom.

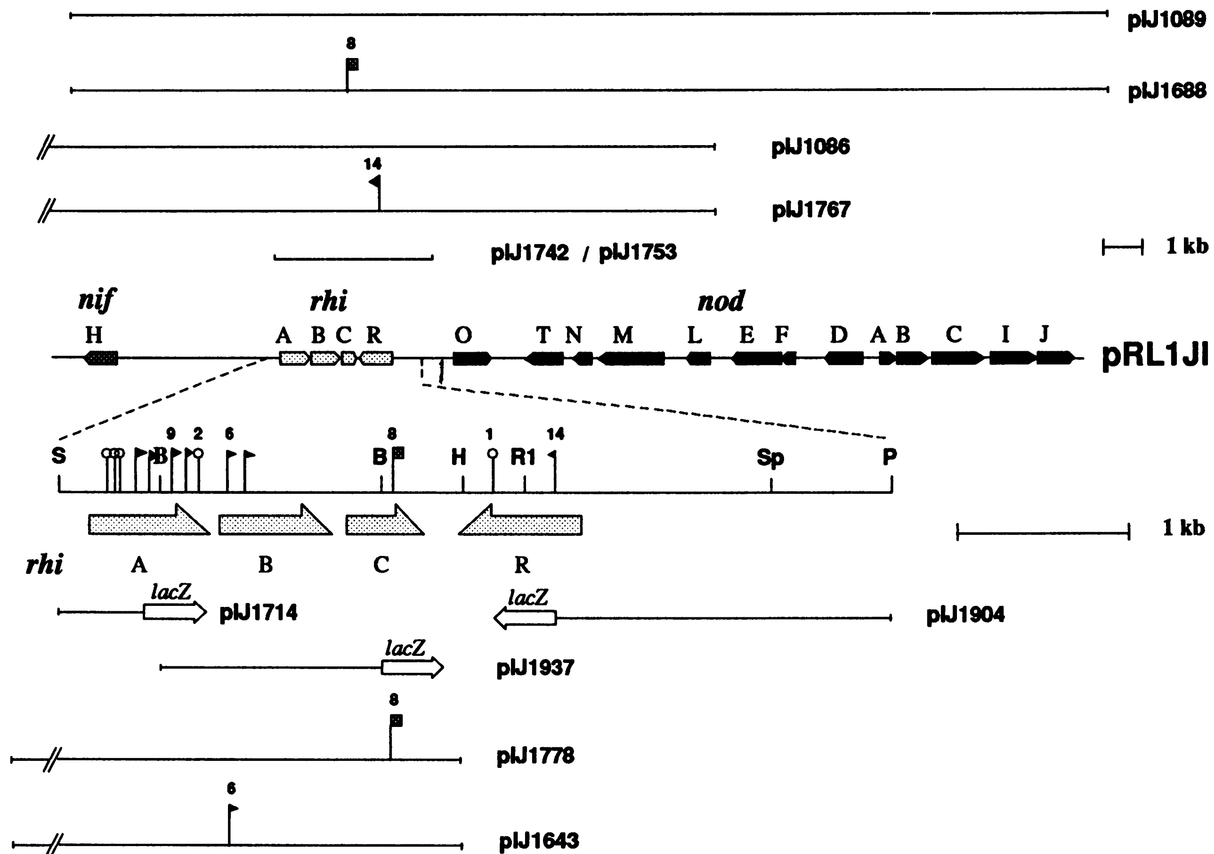

FIG. 1. Physical and genetic map of the *R. leguminosarum* bv. viciae *rhi* and *nod* genes on pRL1JI. The gene sizes and orientations are indicated as arrows, and an enlargement of the *rhi* gene region is shown. The sizes of the various plasmids are indicated; //, gaps introduced for purposes of presentation. The plasmid sizes indicated in the lower part of the figure correspond with the enlarged *rhi* gene region shown. The locations of various mutations are indicated: ○, Tn*5* mutations mapped previously (17); ▶, transposon insertions that make *lacZ* fusions; ▩, the Tn*phoA* mutation in *rhiC*. The numbers above representative mutations correspond with allele numbers (e.g., 14 = *rhiR14*::Tn*5lacZ*) described in the text and Table 1. Plasmids carrying *lacZ* fusions made in vitro are indicated, with the positions of *lacZ* fusions shown. Restriction sites: B, *Bam*HI; H, *Hin*dIII; P, *Pst*I; R1, *Eco*RI; S, *Sma*I; Sp, *Sph*I.

effect on symbiotic phenotype was observed in strains mutated in *rhiA* (16, 17, 22). In an attempt to uncover a function for *rhiA* and for the *rhiR* gene which is required for its expression (22), we have undertaken a molecular characterization of the *rhiA-rhiR* gene region. We report here the identification and characterization of four genes, *rhiABCR*, whose products appear to influence the interaction between *R. leguminosarum* bv. viciae and the vetch *Vicia hirsuta*, one of its host legumes.

MATERIALS AND METHODS

Microbiological techniques. Media and general growth conditions for *Rhizobium* strains were as described by Beringer (2), with the appropriate antibiotics (49). The flavone hesperetin was made up as a 1 mM solution in methanol and added to give a final concentration of 1 μM. *Escherichia coli* strains were grown in L broth at 37°C (34). Plasmids were mobilized in triparental conjugational matings from *E. coli* into *Rhizobium* strains by using a helper plasmid.

β-Galactosidase activities were determined as described by Rossen et al. (39). Alkaline phosphatase assays were performed as described by Brickman and Beckwith (6) with slight modifications: cultures shaken for 40 h at 28°C were grown in 5 ml of Y medium. Cells from 1.5 ml of the culture were pelleted by centrifugation and resuspended in 1.5 ml of 1 M Tris-HCl, pH 8.0. An aliquot (0.3 ml) was used to measure the turbidity at 600 nm with a Titertek Multiscan Plus spectrophotometer. To cells in the remaining 1.2 ml, 10 μl of Triton X-100 (25%, vol/vol) was added, and the cells were incubated for 20 min at 28°C. Aliquots (1 ml) were then assayed at 28°C for alkaline phosphatase activity with *p*-nitrophenylphosphate (Sigma). The reaction was stopped by addition of 200 μl of 1 M K_2HPO_4. Before the optical density at 420 nm was measured, the samples were cleared by centrifugation (1 min). For those samples which had an optical density at 420 nm of >1, the cells were diluted in Tris-HCl buffer before assaying. Units of activity were calculated as described previously (12).

Plant tests were as described by Knight et al. (31); about 10^6 bacteria were inoculated onto *V. hirsuta*. Bacteria were isolated from surface-sterilized nodules as described previously (31).

Microbiological strains and plasmids. Bacterial strains and plasmids are shown in Table 1. Plasmid pIJ1714 was constructed by subcloning a 0.5-kb *Sma*I-*Pst*I fragment carrying the *rhiA* promoter into the *Sma*I-*Pst*I sites of pMP220. pIJ1742 and pIJ1753 were constructed by subcloning a 4.2-kb *Sma*I-*Sph*I fragment from pIJ1089 into the *Pvu*II site

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Relevant properties	Source
Rhizobium strains		
8400	*R. leguminosarum* bv. phaseoli cured of its Sym plasmid	33
8401	Str^r derivative of 8400	33
A34	8401 carrying pRL1JI	18
A69	8401/pRL1JI*nod*Δ*99*(*nodF-nifH*)::Km	20
A160	8401/pRL1JI*rhiR1*::Tn*5*	17
Plasmids		
pIJ1086	*nifH rhiABCR nodMNTO* in pLAFR1	19
pIJ1089	*nifHDK rhiABCR nodABCDEFIJLMNTO* in pLAFR1	19
pIJ1242	pIJ1089*rhiR1*::Tn*5*	17
pIJ1643	5-kb *Hin*dIII fragment carrying *rhiA rhiB6*::Tn*3HoHo1 rhiC* cloned into pKT230	22
pIJ1688	pIJ1089*rhiC8*::Tn*phoA*	This work
pIJ1714	*rhiA* promoter in pMP220	This work
pIJ1742	*rhiABCR* in pKT230	This work
pIJ1750	pIJ1086*rhiA9*::Tn*5lacZ*	This work
pIJ1751	pIJ1086::Tn*5lacZ* insertion upstream of *rhiA*	This work
pIJ1753	*rhiABCR* in pUC119	This work
pIJ1767	pIJ1086*rhiR14*::Tn*5lacZ*	This work
pIJ1778	*rhiA rhiB rhiC8*::Tn*phoA* in pLAFR3	This work
pIJ1904	*rhiR14-lacZ* in pLAFR3	This work
pIJ1937	*rhiBC-lacZ* in pMP220	This work
pIM229	*rhiABC* in pLAFR3	This work

of pKT230 and the *Sma*I site of pUC119. pIJ1904 was made by subcloning the 5-kb *Pst*I fragment carrying the *rhiR-lacZ* fusion from pIJ1767 into the *Pst*I site of pLAFR3. pIJ1937 was made by subcloning the 1,310-bp *Bam*HI fragment from pIJ1753 into pMP220.

Genetic manipulations. Transformation, cloning, ligations, and restriction mapping were done essentially as described by Maniatis et al. (34). DNA sequence determination of cloned fragments was done by the dideoxy chain termination method (40), and both strands were sequenced at least twice with overlapping clones.

Tn*5lacZ* mutagenesis of pIJ1086 was carried out as described previously (50) with *E. coli* A118. Derivatives of pIJ1086 containing Tn*5lacZ* were transferred to strain 8401(pRL1JI), and 600 transconjugants were screened for repression of the β-galactosidase activity in the presence of hesperetin. Transconjugants with this phenotype were screened for the absence of the RhiA protein by using anti-RhiA antiserum (17). The locations of transposon insertions were mapped with *Eco*RI, *Bam*HI, and *Hin*dIII. One of the plasmids carrying the mutation *rhiA9*::Tn*5lacZ* was called pIJ1750, and the plasmid carrying the mutation *rhiR14*::Tn*5lacZ* was called pIJ1767. The control plasmid, pIJ1751, is a derivative of pIJ1086 carrying Tn*5lacZ* ≈5 kb upstream of *rhiA*. Tn*phoA* mutagenesis of pIJ1089 was carried out as described by Davis and Johnston (12). Derivatives of pIJ1089 containing Tn*phoA* were transferred to strain 8401. From 1,500 transconjugants, 3 were identified as being repressed for alkaline phosphatase activity in the presence of hesperetin. The locations of the transposon insertions were mapped with *Eco*RI, *Bam*HI, and *Hin*dIII, and one of the plasmids carrying the mutation *rhiC8*::Tn*phoA* was called pIJ1688.

Strain A124 was constructed by transferring the *rhiC8*::Tn*phoA* allele from pIJ1688 to pRL1JI by homologous recombination as described previously (18). pIM229 was made by cloning a 5-kb *Hin*dIII fragment (from pIJ1089) carrying *rhiABC* into the *Hin*dIII site of pLAFR3, and the derivative of pIM229 (pIJ1778) carrying *rhiC8*::Tn*phoA* was made by recombining the transposon onto pIM229 from its homologous location on pRL1JI essentially as described previously (18). Restriction enzyme analysis of pIJ1778 confirmed the correct location of the transposon in the *rhiC* gene.

Computing methods. Sequence data were assimilated by using the DB system of Staden (44). Predictions of coding probability were made by the positional base preference method (45). Data base searches were made with the program PROSRCH (8), which implements an exhaustive inexact string matching algorithm by using the PAM100 similarity table and with the program TFASTA (38).

Identification of RhiA protein on sodium dodecyl sulfate-polyacrylamide gels. Proteins were isolated from *Rhizobium* strains as described by Dibb et al. (17), electrophoretically separated with sodium dodecyl sulfate-polyacrylamide gels, and stained with Coomassie blue or transferred to nitrocellulose and stained with anti-RhiA antiserum (17) as described by Bradley et al. (5).

RESULTS

***rhi* genes can influence nodulation.** Previous work (17, 22) failed to identify any phenotype caused by mutations in *rhiA*. A similar observation was made with the *nodO* gene (14, 21, 22), but it was later found that in the absence of other *nod* genes (*nodFEL*) mutation of *nodO* had a clear effect on nodulation (20). A similar test system was used to determine whether the *rhi* gene region influences nodulation. pIJ1086 (Fig. 1), which carries the *rhi* gene region along with *nodO* and the *nodLMNT* genes, complements the deletion mutant A69 for nodulation of vetch (Fig. 2), although the level of nodulation observed is about half of that seen with a wild-type strain. This reduced level of nodulation is consistent with the absence of the *nodFE* genes from strain A69(pIJ1086) (20).

The derivatives of pIJ1086 (Fig. 1) carrying *rhiA9*::Tn*5lacZ* (pIJ1750) or *rhiR14*::Tn*5lacZ* (pIJ1767) were tested for their ability to complement strain A69 for nodulation of

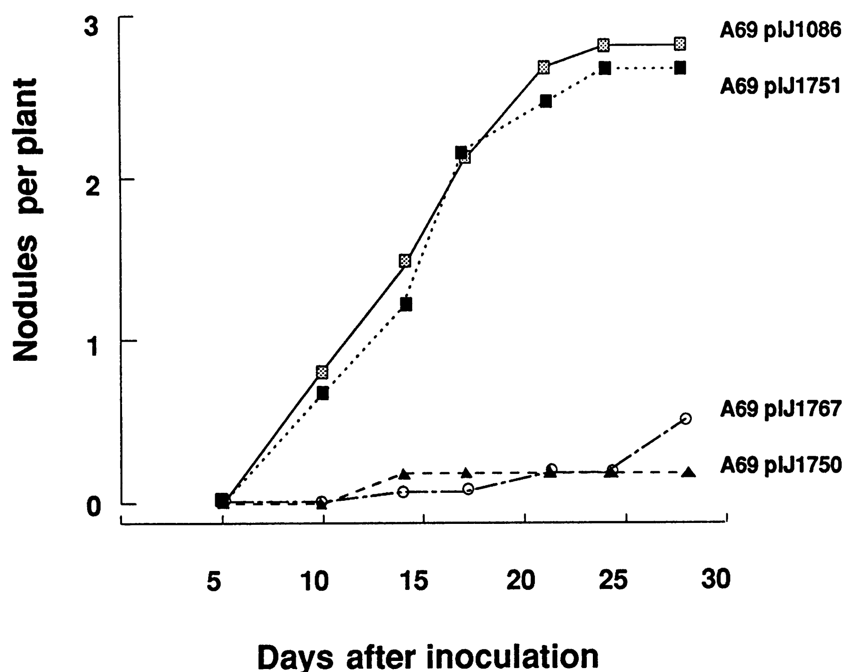

FIG. 2. Nodulation of *V. hirsuta* by strain A69 carrying various plasmids. Freshly germinated *V. hirsuta* was inoculated with derivatives of strain A69 carrying the following plasmids: pIJ1086 (*nodL nodMNT nodO rhiABC rhiR*), pIJ1750 (pIJ1086*rhiA9*::Tn*5lacZ*), pIJ1751 (pIJ1086::Tn*5lacZ* upstream of *rhiA*), and pIJ1767 (pIJ1086*rhiR14*::Tn*5lacZ*). The data presented are from one experiment and show the average number of nodules per plant, obtained by using 16 plants for each strain.

vetch. As shown (Fig. 2), the level of nodulation was significantly less than that found with A69 carrying the control plasmid pIJ1086. To test whether this was a nonspecific effect caused by the insertion of Tn*5lacZ* in pIJ1086, a nodulation test was carried out with strain A69 containing pIJ1751, which is a derivative of pIJ1086 carrying Tn*5lacZ* inserted in the region to the left of the *rhi* genes. In this case (Fig. 2) the level of nodulation was similar to that observed with A69(pIJ1086), confirming that the reduction of nodulation observed with *rhiA* and *rhiR* mutations is not due to a nonspecific effect (e.g., plasmid stability) caused by insertion of the transposon. Therefore, the *rhi* genes do appear to contribute in some way to the efficient nodulation of vetch, but in normal laboratory nodulation tests (in which all of the *nod* genes are present), it is not possible to measure this effect with *rhi* mutants. Since the *rhi* genes are not under the same regulatory control as the *nod* genes, we decided not to call them *nod* genes but to retain the gene designation *rhi* (for rhizosphere) previously used to define *rhiA* and *rhiR* (17, 22).

DNA sequence of the *rhi* genes. The *rhi* region was first identified on the basis of *rhiA* mutations that blocked (or *rhiR* mutations that severely reduced) the formation of the *rhiA* gene product. The mutations are located in a region of about 3 kb of DNA close to *nodO*, and all of them mapped between a *Sma*I site and an *Sph*I site (Fig. 1). This 4.2-kb *Sma*I-*Sph*I fragment was subcloned into the *Pvu*II site of pKT230 to form pIJ1742. A strain of *R. leguminosarum* lacking a symbiotic plasmid but carrying pIJ1742 made the RhiA protein, indicating that both the *rhiA* and *rhiR* genes are on pIJ1742.

The DNA sequence of this region was determined (Fig. 3). Within the *rhi* gene region, four large open reading frames (ORFs) were identified; computer-assisted analysis of the DNA sequence indicated that they had a high probability of encoding proteins. Three of these ORFs are transcribed from one strand and one is transcribed from the other strand (Fig. 1 and 3).

The first ORF (Fig. 3), nucleotides (nt) 186 to 888, is preceded by a good potential ribosome binding site and encodes a protein with a predicted molecular weight of 24,885. This corresponds with the *rhiA* gene product since (i) the previously described *rhiA* mutations map within this region, (ii) the estimated M_r of the RhiA protein is 24,000, and (iii) the amino terminus of the purified RhiA protein was found by protein sequencing to be MSLXVSYVDKEMXD (15), corresponding with the predicted ORF.

Downstream of *rhiA* is a second ORF (positions 931 to 1587; Fig. 3) which encodes a protein with a predicted molecular weight of 23,142. We propose to call this gene *rhiB* since it is in the same operon as *rhiA* (see below). The *rhi-6*::Tn*3HoHo1* allele (22) maps within *rhiB* (Fig. 1) and was found to be under regulatory control similar to that over *rhiA*.

The third ORF (which we will call *rhiC*) extends from nt 1679 to 2125 and encodes a protein with a predicted molecular weight of 15,299. Significantly, the N terminus of the predicted *rhiC* gene product is similar to N-terminal transit sequences found in proteins exported across the bacterial inner membrane (54). It consists of a hydrophobic stretch of 15 amino acids preceded by a positively charged residue (Fig. 3) and followed by the residues Ala-Gln-Gly that could constitute a signal peptidase recognition site (54). To test whether the predicted RhiC protein is exported across the membrane, a RhiC-PhoA protein fusion was made with Tn*phoA*, which contains the *phoA* gene lacking a transit sequence. When *phoA* inserts in frame into a gene that encodes a protein with an N-terminal transit sequence, the alkaline phosphatase domain is translocated across the cy-

```
   1 CCCTATGGGTTAATCGTGGTCTGTCAGGGAGTAAATCTACTCATTTACGTCACCCGGAAATGATTATAGCTGAATTGGTGTCTCCAATAACACTCGCCTC

                                                                                        M  S  L  H  V
 101 AGTTGGAGCTGGGCCAAGAAGCGGTCCAAGCGATTGGATCAGTCCTGGCGTGAAGACACAAAATCAATGGAAAAGGGAGAGTTAAATGTCCTTGCATGTC

     S  Y  V  D  K  E  M  T  D  H  A  R  A  S  Q  P  G  S  A  A  L  A  Q  G  T  Q  Y  S  L  L  L  K  N  Q
 201 AGCTACGTAGACAAGGAAATGACGGATCATGCCCGTGCATCACAGCCGGGAAGCGCAGCGCTTGCCCAAGGAACGCAATATTCGTTATTGCTGAAGAATC

      S  A  Q  P  W  T  F  Y  V  Y  Q  K  M  P  Q  P  V  A  N  V  F  S  L  A  W  F  C  S  P  Y  Q  I  R
 301 AATCGGCGCAACCTTGGACCTTCTATGTCTATCAAAAGATGCCTCAACCTGTTGCCAATGTCTTCTCCTTGGCATGGTTCTGCTCTCCGTATCAAATTCG

      V  G  N  Q  I  K  F  T  W  E  L  A  Y  N  F  V  W  S  D  T  G  Q  L  I  P  G  V  D  F  F  A  S  G
 401 GGTTGGCAATCAAATCAAGTTCACTTGGGAGCTCGCCTATAATTTCGTCTGGAGCGACACGGGACAACTGATTCCTGGCGTGGACTTTTTTGCCTCCGGG

     V  E  D  C  S  P  S  G  R  N  T  T  T  F  S  L  S  D  G  P  G  L  T  A  P  I  K  G  D  P  A  G  S  L
 501 GTGGAGGACTGCAGCCCCAGCGGGAGAAACACCACTACTTTTTCATTAAGCGATGGACCTGGCCTGACCGCGCCGATCAAGGGGGATCCTGCAGGATCAC

      V  I  N  D  A  G  N  V  P  N  N  R  F  S  V  G  I  G  M  S  G  T  G  T  Y  V  A  Q  A  G  T  N  L
 601 TGGTCATCAACGATGCTGGCAATGTGCCAAACAACCGGTTCTCCGTGGGCATCGGCATGTCCGGAACAGGGACTTATGTTGCACAAGCGGGCACCAATCT

      L  H  T  F  T  P  T  P  S  Y  W  I  A  A  G  T  N  V  T  I  G  S  V  L  S  I  D  T  I  T  Q  T  R
 701 GCTCCACACGTTCACGCCAACTCCGAGCTATTGGATTGCGGCGGGAACGAATGTCACTATCGGGTCGGTGCTCAGCATCGACACGATCACCCAGACCAGG

     E  A  K  F  P  S  A  V  F  N  L  V  G  V  L  Q  E  D  N  T  W  D  I  N  P  A  *
 801 GAAGCCAAGTTTCCCTCTGCCGTCTTCAATCTCGTAGGCGTGCTCCAAGAGGACAATACCTGGGATATCAACCCGGCTTGATGGCAGGGTGAATTGGCTT

                                  M  G  V  P  V  V  A  D  S  W  H  S  F  P  H  T  S  I  L  P  E  E  P  T
 901 CAGCTCCGCCCGTCAGGGCGGAACTGATGCATGGGAGTGCCTGTTGTTGCTGATAGTTGGCACTCCTTTCCTCACACATCAATATTGCCCGAGGAACCAA

      M  F  A  G  P  P  I  G  A  V  C  P  F  A  G  Q  V  A  P  I  S  S  S  V  N  T  I  W  S  N  T  P  C
1001 CGATGTTTGCAGGGCCTCCAATCGGTGCGGTGTGTCCCTTTGCGGGGCAAGTTGCCCCTATCTCGAGCTCGGTCAATACAATCTGGAGCAACACGCCCTG

      A  S  S  G  E  A  A  G  T  N  A  E  A  P  I  S  Y  V  E  A  Q  G  W  M  L  C  D  G  R  Y  L  R  A
1101 CGCAAGCTCTGGCGAGGCTGCGGGTACAAATGCGGAAGCTCCTATCAGCTACGTCGAAGCCCAGGGATGGATGCTTTGCGATGGACGCTATCTCAGGGCA

     A  V  Y  P  E  L  Y  A  V  L  G  G  L  Y  G  E  R  N  S  T  A  D  L  E  F  R  I  P  D  Y  R  G  L  F
1201 GCCGTCTATCCAGAACTCTACGCCGTTCTCGGCGGACTCTATGGAGAGCGGAATTCTACGGCCGATCTGGAATTTCGGATTCCAGACTACCGTGGGCTGT

      L  R  G  F  D  A  G  G  G  M  D  P  D  A  K  R  R  L  D  P  T  G  N  N  V  A  N  V  V  G  S  L  Q
1301 TCCTGCGCGGTTTCGATGCCGGTGGCGGTATGGACCCTGACGCGAAACGACGGCTAGATCCAACCGGCAATAATGTTGCGAATGTTGTTGGATCACTCCA

      C  D  A  L  Q  V  H  A  H  P  Y  E  I  T  T  P  A  G  I  S  Q  Q  G  N  A  A  G  T  S  I  S  S  K
1401 ATGCGATGCTCTGCAAGTTCACGCGCATCCCTATGAGATAACGACGCCGGCGGGAATTTCGCAGCAGGGCAATGCCGCCGGAACTTCCATCTCAAGCAAA

     S  T  G  S  P  E  N  P  A  R  T  A  L  E  T  R  P  K  N  V  A  V  N  Y  L  I  K  F  R  *
1501 TCGACAGGTTCACCCGAAAATCCGGCACGCACAGCTCTTGAGACGCGTCCCAAGAACGTGGCCGTGAACTACCTGATCAAGTTTCGGTAACTTGAAAGGG

                                                                                       M  T  A  T  L  R  A  F
1601 GTGTAGGTCCTTTCATTCGATTTAGAGACATCGGTCCGGAAGCGACATCTTGCCGGCCGGCAAAAGAGGAGAACCTTAATGACTGCTACTTTGAGAGCTT
                                                               ↓
      G  W  L  A  A  F  A  L  T  V  T  F  A  Q  G  A  A  A  E  E  Q  Q  K  G  K  V  G  A  K  P  V  E  T
1701 TCGGATGGCTAGCTGCTTTCGCTCTGACCGTCACGTTCGCGCAGGGTGCCGCGGCTGAGGAGCAGCAAAAAGGCAAGGTGGGGGCGAAGCCCGTAGAAAC

      G  V  V  I  R  G  V  T  L  A  G  P  V  G  N  P  G  T  S  T  G  K  T  C  D  F  S  G  E  P  V  D  P
1801 TGGCGTCGTGATCCGCGGCGTCACGCTTGCCGGTCCGGTTGGAAATCCGGGAACGTCAACCGGGAAAACATGCGACTTCAGCGGTGAGCCCGTGGATCCA

     S  G  R  L  E  G  A  S  V  N  C  R  P  N  G  N  Q  A  N  T  T  P  G  L  P  A  R  F  N  A  Y  C  M  I
1901 TCGGGGCGTTTGGAGGGGGCAAGCGTCAACTGCAGGCCAAATGGCAACCAGGCGAATACAACGCCAGGGCTACCCGCACGCTTTAACGCGTATTGCATGA

      N  A  P  V  K  S  A  R  L  I  Q  A  A  R  P  E  N  A  N  H  C  D  L  S  G  I  T  P  K  D  A  T  G
2001 TAAATGCGCCGGTTAAAAGTGCGCGGCTGATCCAAGCTGCCCGACCAGAAAATGCCAACCACTGCGATCTGTCGGGGATCACCCCAAAAGACGCGACTGG

      Q  F  G  G  A  V  W  R  *
2101 TCAATTCGGGGGGGCCGTCTGGCGGTAGGCGGTCATTGGACGACTTCTGGGGAAACAAATTCAAAGCACGTCCTCGCAGCCCCCCAATCGATGCTGCCTG

2201 CTCGACCTACCGAAAACTGTGTTGCGCCTGTCTATCAATTCAGACGCAGATTTAGCTCTGCATTTCAACGGGGCTACTATCTGATCATGATGGCTGTCGT

2301 GATCCAACAGGCCCGATCTTTGAAGTGGCCGAAAACTCCATATGCTTCAGCGAATGATTCTGAGACGAAAGCTTTCGGCAATCATCTGAGGTGTATTAAC
                                                  AGT
2401 CACATTGAGCTTTCGCTGGATGTTCAAAATGACATTTTGAATGGTTCGATGGGATCGTCCGAGAATCGTTGCGATCTCTATTGCTGTTTTGCCTGCGGCA

2501 CACCAATGGATGATTTCGCGCTCTCGCGTTGTGATCATGGGTAAGGCCTGAATTTTTCGACTGGCATCATTCCCAATTTGGGCGCGGAGCAAGCTGTGCG

2601 CGTAAGCCGCCATCAGATAGAGCGCGCTTCGGTCGCAGGTTGAGAGCTCTACCTTTTCCGCTCCAAAACTGACAATCGATTGAAACCCAGCGGCAGTGTG

2701 TAAAGGAACGCTGAAGCCGTCAATCAACTTGAATTCGCGTGCCTCATCCATGACACGACGGCTCTGCCGATCGAGTTTTTGATCTCGCAGAGCTTCCGAC

2801 CAAACAAATGCGTGATCGCATGTCTTGCTAAGATGCACTATGGGGTCGGCGTGGACATAATTGTTTTCACGATAACGATCAAACCACCCGACTGACCAAT

2901 TGCCCAAAACGAAATAGGAATCAATTCGCTCAATAGGCGAGGGATGCCGGATATCGCGAAATAACTAAACCCGAAATACTGCGAAATTTTTCCAAATAA

3001 GAGTAACACATCGTCTTTGCTCTTTGCGCTGGCTGACTCTGACAGGAAATCGAACACCAGATTCGAGACCGCTGAGGATTCTTCCTTCACGACAACCTCC
                                                 GTG                                 GTG
3101 CGGTTCGGTGCAAGATTAACTTCCTCACATCTAAGCAGTGAGTGTCCAGTTCCGTGGGCAGCTAAACTCTGAAAACCATATTATATACCCTTGATGCATA

3201 CAACCAAAAAGCGCATCTTCAGGCGGGATCGTCCTCGGCAAACCGTTTGGCGCCGCTCGTCTGAAAAACAATTCGAGCCGTCAATTTTAACCGAGTGCGT
```

FIG. 3. Nucleotide sequence of the *rhi* region. The predicted amino acid sequences of the *rhiABC* genes are presented. The *rhiR* gene (nt 3090 to 2347) reads from the opposite strand of the depicted sequence. The two potential RhiR start codons are indicated (GTG), and the stop codon is also shown (TGA). The translation of RhiR using the first of the two GTG codons is shown in Fig. 4. Possible ribosome binding sites are underlined. A potential leader peptidase cleavage site in the RhiC protein is indicated by a vertical arrow.

toplasmic membrane to the periplasm, where its activity can be measured (35). Tn*phoA* mutant derivatives of pIJ1089 were screened for alkaline phosphatase activity, and the mutations were mapped with restriction enzymes. One of the transposon insertions mapped within *rhiC* at about nt 1990 (Fig. 3) and was called *rhiC8*::Tn*phoA*. Since the plasmid (pIJ1688) carrying this allele conferred alkaline phosphatase activity to strains grown on plates (using the chromogenic

```
RHIR        1  VKEESSAVSNLVFDFLSESASAKSKDDVLLLFGKISQYFGFSYFAISGIPSPIERIDSYFVLGN
                          .... . .*..*    .....     .*. .. * ..    .. .*.*
LUXR           MKDINADDTYR-IINKIKACRSNNDINQCLSDMTKMVHCEYYLLAIIYPHSMVKSDISILDN
                      .            *         * ... *....    . .      .*
28KD-UVRC      MQDKDFFSWRRTMLLRFQRMETAEEVYHEIELQAQQLEYDYYSLCVRHPVPFTRPKVAFYTN

RHIR       65  WSVGWFDRYRENNYVHADPIVHLSKTCDHAFVWSEALRDQKLDRQSRRVMDEAREFKLIDGF
               ..  * . * ..* .. ****. *.. . .. * .   .. ....*..*..**..  **.**
LUXR           YPKKWRQYYDDANLIKYDPIVDYSNSNHSPINWNIFENNA-VNKKSPNVIKEAKTSGLITGF
               ..  * . *...*..  **     . ..  ..*...*                       *
28KD-UVRC      YPEAWVSYYQAKNFLAIDPVLNPENFSQGHLMWNDDLFSE-AQPLWEAARAHGLRRGVHSVF

RHIR      127  SVPLHTAA-GFQSIVSFGAEKVELSTCDRSALYLMAAYAHSLLRAQIGNDASRKIQALPMIT
               * * ***. **  *     .** .  ..      .. .    * ..  .  .  . ..    *
LUXR           SFPIHTANNGFGMLSFAHSEKDNYIDSLFLHACMNIPLIVPSLVDNYRKINIANNKSNNDLT
                .   * * **    *  .      .*  .* .      * *                 ..
28KD-UVRC      NAA-QTGALGF-LSFSRCSRREIPILSDELQLKMQL-LVRESLMALMRLNDEIVMTPEMNFS
                                  ____________________
RHIR      188  TREREIIHWCAAGKTAIEIATILGRSHRTIQNVILNIQRKLNVVNTPQMIAESFRLRIIR
               .**.*    *   ** .  * .*** *.**     . * * ***..*  * *....    *
LUXR           KREKECLAWACEGKSSWDISKILGCSERTVTFHLTNAQMKLNTTNRCQSISKAILTGAIDCPYFKS
               .**.** .* * ***.  ** **  *. * .     * * * *  *  *    .     *
28KD-UVRC      KREKEILRWTAEGKTSAEIAMILSISENTVNFHQKNMQKKINAPNKTQVACYAAATGLI
```

FIG. 4. Amino acid sequence alignment of the RhiR protein with the *V. fisheri* LuxR (15) and the *E. coli* UvrC-28K (42) proteins. Gaps (-) were introduced to allow an optimal alignment. Shown are identical residues (*) and conservative substitutions (⊙) versus RhiR. The proposed helix-turn-helix motif is indicated by a line.

substrate 5-bromo-4 chloro-3-indolylphosphate) and in liquid medium (see below), it is concluded that *rhiC* encodes a protein that crosses the inner membrane. Since the RhiC protein does not contain hydrophobic domains other than the potential transit sequence, we conclude that it is probably a periplasmically located protein.

The fourth ORF is transcribed convergently toward *rhiC*. The ORF predicted from computer analysis extends from about nt 3100 to about nt 2400; four separate transposon mutations have been identified in this region, two of which are shown (Fig. 1). Previously, one of these mutations (*rhiR1*::Tn*5*) was used to identify the regulatory gene, *rhiR*, required for *rhiA* expression (22). With antiserum to the RhiA protein, all four transposon mutations in this region were found to reduce strongly (by about 98%) the formation of RhiA protein, indicating that they are all within *rhiR*. However, there is no ATG codon in the proposed *rhiR* ORF until position 2750. Since one of these mutations is upstream of this ATG and computer predictions indicated that the region from about position 3100 is coding, we looked for alternative translational starts. Two potential GTG translational starts (positions 3090 and 3057) are indicated on the DNA sequence in Fig. 3. The start of the predicted RhiR protein sequence (Fig. 3) that was chosen (nt 3090) was based on the preceding potential ribosome binding site and on homology with other proteins; its translation is shown in Fig. 4. A search for characteristic protein motifs identified that amino acid residues 203 to 222 correspond with a helix-turn-helix motif (37), characteristic of DNA-binding proteins (Fig. 4).

Protein homologies. The predicted coding regions of the *rhiA*, *rhiB*, *rhiC*, and *rhiR* genes were used to search against protein sequences from the EMBL, GenBank, and Swissprot data bases. No significant similarities were detected with RhiB or RhiC. Neither of these proteins shows a hydrophobic domain typical of those found in transmembrane proteins, although as mentioned above, RhiC does have a hydrophobic domain near the N terminus and is probably periplasmic.

A short region of similarity was found between RhiA and the *Klebsiella pneumoniae* and *E. coli* lactose carriers encoded by the *lacY* gene. This homology was restricted to the region adjacent to the LacY residue cysteine 148, which has been implicated in substrate recognition since it is protected by substrates against attack by maleimide (52). Unlike LacY (a transmembrane protein), RhiA has no hydrophobic domains and is a soluble cytoplasmic protein (13, 17). Therefore, the limited homology between LacY and RhiA may indicate similarities between substrate recognition sites.

The *rhiR* gene product showed extensive and strong similarity with two other bacterial proteins, LuxR and UvrC-28K. An alignment of their protein sequences (Fig. 4) shows extensive homology throughout the length of the proteins. The LuxR protein, in the presence of the small diffusible molecule *N*-(3-oxo-hexdroyl)-homoserine lactone, activates transcription of the genes required for bioluminescence (41) in the marine bacterium *Vibrio fischeri* (the *lux* regulon). It has been suggested that the UvrC-28K protein (encoded by an ORF upstream of *uvrC*) may be required to enhance transcription of the UvrC gene (42) in induced conditions (DNA damage caused by UV light). In addition to this extensive homology with LuxR and UvrC-28K, there was a high degree of localized homology between RhiR and other DNA-binding and regulatory proteins (Fig. 5). Part of this homologous region corresponds with the domain that was noted earlier to contain a probable helix-turn-helix DNA-binding motif. However, it is significant that the most highly conserved region among RhiR, LuxR, and UvrC-28K and the other DNA-binding proteins is the region immediately upstream of the helix-turn-helix domains (Fig. 5). This points to homology with a particular family of regulators, the two-component (sensor-kinase) regulatory group of proteins (25). The conserved region (boxed) consists of (i) a highly conserved Leu at position 186, (ii) Ser or Thr at position 187, (iii) a hydrophilic residue at position 188, (iv) a highly conserved Arg-Gln pair at positions 189 to 190, (v) a less highly conserved block of residues at positions 191 to 199, and (vi) a Gly at position 200 preceding the helix-turn-helix

```
RhiR        ITTREREIIHWCAAGKTAIEIATILGRSHRTIQNVILNIQRK
LuxR        LTKREKECLAWACEGKSSWDISKILGCSERTVTFHLTNAQMK
UvrC-28KD   FSKREKEILRWTAEGKTSAEIAMILSISENTVNFHQKNMQKK
GerE        LTKREREVFELLVQDKTTKEIASELFISEKTVRNHISNAMQK
UhpA        LTKRERQVAEKLAQGMAVKEIAAELGLSPKTVHVHRANLMEK
ComA        LTPRECLILQEVEKGFTNQEIADALHLSKRSIEYSLTSIFNK
DegU        LTRRECEVLQMLADGKSNRGIGESLFISEKTVKNHVSNILQK
RcsB        LSPKRSRVLRLFAEGFLVTEIAKKLNGRIKTISSQKKSAMMK
FixJ        LSERERQVLSAVVAGLPNKSIAYDLDISPRTVEVHRANVMAK
NarL        LTPRERDILKLIAQGLPNKMIARRLDITESTVKVHVKHMLKK
MalT        LTQREWQVLGLIYSGYSNEQIAGELEVAATTIKTHIRNLYQK
RcsA        LSRTESSMLRMWMAGQGTIQISDQMNIKAKTVSSHKGNIKRK
UvrC-23KD   LSERELQIMLMITKGQKVNEISEQLNLSPKTVNSYRYRMFSK
BvgA        LSNRELTVLQLLAQGMSNKDIADSMFLSNKTVSTYKTRLLQK
NodW        LSPREQAVMKLVATGLMNKQVAAELGLAEITVKIYRGHVMKK
```

FIG. 5. Alignment of RhiR and other regulatory proteins. The potential helix-turn-helix domain is overlined and underlined, and extended homology beyond it is illustrated by boxed identical and highly conserved residues. The sequences are taken from LuxR (15), UvrC-28K (42), GerE (10), UhpA (23), ComA (55), DegU (27), RcsB (47), FixJ (11), NarL (46), MalT (7), RcsA (48), UvrC-23K (42), BvgA (1), and NodW (24). The sequence of RhiR shown extends from amino acid residues 186 to 227.

at positions 203 to 222, which is followed by a conserved Lys at position 227 (numbered with respect to RhiR). The conservation of the residues around the Arg-Gln pair is significantly stronger than the conservation of amino acid residues within the helix-turn-helix domains of the group of protein listed in Fig. 5. While this may indicate that a second *rhi* regulator (encoding a sensor) may be present, the LuxR, UvrC-28K, and RhiR proteins are a distinct subgroup of regulators and no additional regulators (sensors) have been found to act in conjunction with LuxR and UvrC-28K.

Regulation of the *rhi* genes. It was established previously (22) that the *rhiR* gene is required for *rhiA* expression. To confirm that the *rhiA* promoter is contained within the 186-nt region between the *Sma*I site and the *rhiA* translational start (Fig. 3), a *Sma*I-*Pst*I fragment was subcloned into pMP220, a *lacZ* fusion vector. As shown in Table 2, the plasmid formed (pIJ1714) expresses β-galactosidase activity, the expression of which is *rhiR* dependent. As was observed previously, the expression of *rhiA-lacZ* could be repressed by flavonoids (Table 2) and this flavonoid effect is *nodD* dependent. On the basis of these observations it is concluded that the intact *rhiA* promoter is present between the *Sma*I site (position 1) and the translation start of *rhiA*.

Like *rhiA* expression, *rhiB* expression is repressed by flavonoids and is *rhiR* dependent. This was measured previously with the *rhiB-lacZ* fusion created by the *rhiB6*::Tn*3HoHo1* allele. A *Hin*dIII fragment carrying *rhiA*, *rhiB6*::Tn*3HoHo1*, and *rhiC* was subcloned to form pIJ1643. Since this fragment lacks *rhiR*, it was possible to show that *rhiB* expression in pIJ1643 requires *rhiR* (Table 2).

TABLE 2. Measurements of *rhi* gene expression by using *rhi-lacZ* fusion

Strain	*rhi* gene(s) present	Fusion	β-Galactosidase activity (U)[a]	
			−Hesperitin	+Hesperetin
8401(pIJ1714)	None	*rhiA-lacZ*	124	130
A160(pIJ1714)	*rhiABC*	*rhiA-lacZ*	130	125
A34(pIJ1714)	*rhiABCR*	*rhiA-lacZ*	1,125	758
8401(pIJ1643)	*rhiA*	*rhiAB-lacZ*	27	29
A160(pIJ1643)	*rhiABC*	*rhiAB-lacZ*	26	34
A34(pIJ1643)	*rhiABCR*	*rhiAB-lacZ*	1,450	776
8401(pIJ1937)	*rhiB*	*rhiBC-lacZ*	100	105
A34(pIJ1937)	*rhiABCR*	*rhiBC-lacZ*	120	115

[a] β-Galactosidase activity was measured as described previously (39). Hesperetin, when added, was present at 1 μM. The *lacZ* fusion vector used to construct pIJ1714 and pIJ1937 usually gives a background level of 100 to 120 U of activity, whereas the transposon-*lacZ* reporter used to make pIJ1643 usually gives a background of about 25 to 30 U of activity.

Similarly, pIJ1778, which contains the *rhiC8*::Tn*phoA* allele on the subcloned *Hin*dIII fragment carrying *rhiA*, *rhiB*, and *rhiC8*::Tn*phoA*, was used to measure *rhiC* expression. Strain A160(pIJ1778) (which lacks *rhiR*) expressed only 100 U of alkaline phosphatase, compared with 940 U for strain A31(pIJ1778), in which *rhiR* is present. Therefore, *rhiC* expression requires *rhiR* and the *rhiABC* genes appear to constitute an operon under *rhiR* control. The absence of promoters between *rhiA* and *rhiB* or *rhiB* and *rhiC* was confirmed, because the 1,310-bp *Bam*HI fragment (in pIJ1937) carrying the *rhiA-rhiC* region but lacking the DNA region upstream of *rhiA* had no promoter activity (Table 2).

NodD could exert its inhibitory effect on *rhiABC* expression via a direct effect on the *rhiABC* promoter or indirectly by affecting the expression of RhiR, the transcriptional activator of the *rhiABC* genes. The *rhiR14*::Tn*5lacZ* allele on pIJ1767 was used to monitor the effects of flavonoids and NodD on *rhiR* gene expression. As shown in Fig. 6, *rhiR* is expressed normally during free-living culture and the level of its expression is decreased by the flavone hesperetin, but only if the *nodD* gene is present. This indicates that the effects of flavonoids on *rhiABC* expression could be caused by an inhibition of *rhiR* expression. Two models can be proposed to account for this *nodD*-dependent effect: (i) it could be due to a direct effect of NodD on the *rhiR* promoter, or (ii) it could be due to a *cis* effect caused by the very strongly expressed (*nodD*-dependent) *nodO* promoter which is transcribed divergently from *rhiR* (Fig. 1). To distinguish these two possibilities, the *rhiR14*::Tn*5lacZ* allele was subcloned away from the *nodO* promoter with *Pst*I, which cuts upstream of the *nodO* promoter (21). The resulting plasmid, pIJ1904, retained a high level of *rhiR* expression (1,200 U), but this activity was not inhibited by hesperetin when *nodD* was present. We therefore conclude that the *nodD*-dependent inhibition of *rhi* gene expression may be due to a *cis* effect, possibly caused by the high level of *nodD*-dependent transcription at the *nodO* promoter. However, the results presented do not eliminate the possibility that NodD also has a direct effect on the *rhiABC* promoter. The *nodD*-dependent inhibition of *rhiABC* expression by flavonoids may be relevant to the possible role of the *rhi* genes in nodulation, but the biological significance (if any) is not clear.

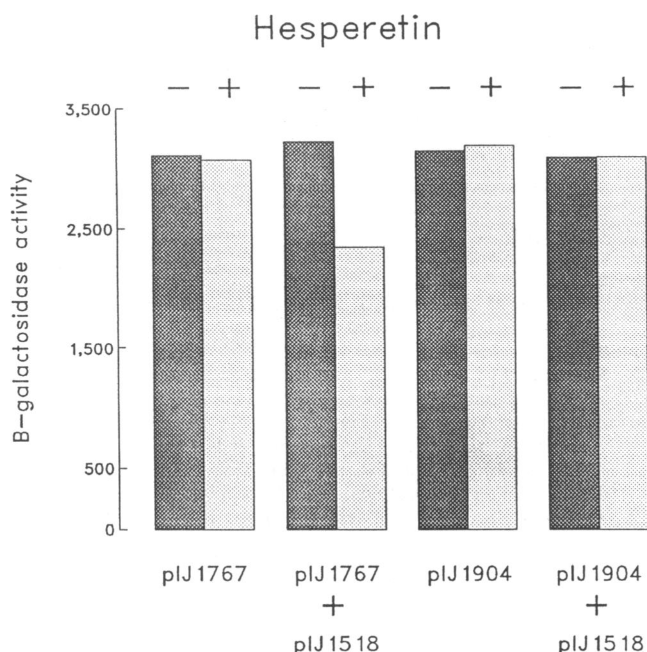

FIG. 6. NodD- and hesperetin-dependent repression of *rhiR* expression. β-Galactosidase activities were measured by using derivatives of strain 8401 containing *rhiR14*::Tn*5lacZ* on pIJ1767 plus pIJ1518 (*nodD*) or *rhiR14*::Tn*5lacZ* subcloned away from the *nodO* promoter on plasmid pIJ1904 plus pIJ1518.

DISCUSSION

In strains of *R. leguminosarum* bv. viciae grown to stationary phase in liquid or solid media, the *rhiA* gene product is the most prominent single protein in the cytoplasmic fraction. RhiA is also a prominent protein in bacteria isolated from the rhizosphere of peas but is absent from bacteroids isolated from pea nodules (17). Since RhiA is found in *R. leguminosarum* bv. viciae and is absent from the closely related biovars trifolii and phaseoli, it was anticipated that it may play a role in the symbiotic interaction between *R. leguminosarum* bv. viciae and its host legumes. This idea was further strengthened by the observation that *rhiA* transcription is repressed by flavonoids that normally induce *nod* gene expression and that this repression is *nodD* dependent (22). However, previous studies failed to identify any role for *rhiA* either in symbiosis or in normal free-living growth conditions (16, 17), even though *rhiA* was established to be present as a single-copy gene.

It is now evident that the *rhi* genes do play a role in the early stages of the symbiotic interaction, but this role in nodulation was uncovered only in a strain that lacked the *nodFE* genes. A similar observation was made with the *nodO* gene (20). One explanation could be that the *rhi* genes influence NodO expression or secretion or modify NodO in some way. However, mutants lacking the *rhi* genes secrete normal amounts of NodO (21). Furthermore, it is clear from the work presented here that the *rhi* genes are under regulatory control different from that of the *nod* genes. Four *rhi* genes have been identified. *rhiABC* are in one operon, whose expression is controlled by the *rhiR* gene product, a protein with no similarity to either NodD or SyrM, both of which are regulators of *nod* gene expression (32). Interestingly the flavonoid- and *nodD*-dependent inhibition of *rhiA* expression may be mediated via *rhiR*. In all growth conditions tested, *rhiR* was expressed constitutively, except in the presence of flavonoids, in which case we found a *nodD*-dependent repression of *rhiR*.

The RhiR protein is homologous to regulator proteins from a family of signal-transducing proteins that allow bacteria to respond to their environment. Generally, these consist of two components (25), a sensor which senses environmental stimuli (such as osmotic stress, nutrient deprivation, or specific metabolites) and a regulator which is usually a DNA-binding protein that activates transcription after being phosphorylated by the sensor protein. It is significant that the sequence of strongest homology between RhiR and these regulators is the region immediately upstream of the proposed helix-turn-helix domains thought to be involved in DNA binding. The high degree of conservation of this region (Fig. 5) indicates that it may be involved in an extension of the DNA-binding domain, a suggestion also proposed by Kahn and Ditta (30). Whereas the homology between RhiR and many of the DNA-binding proteins is limited, there is end-to-end homology with LuxR and UvrC-28K. This indicates that these three proteins belong to a distinct subgroup of regulators, in accord with the proposal of Henikoff et al. (26), who studied homologies among this group of regulatory proteins.

The role of the *rhi* genes in the symbiotic interaction remains unclear. Genes involved in host-specific nodulation but independent of *nodD* control have been identified in *Bradyrhizobium japonicum*. These *nodVW* genes are required for nodulation of some legumes and appear to make up a two-component regulatory system that allows extension of the host range (24). However, the genes regulated by *nodVW* were not identified, and their functional role in host specificity was not established.

The cellular location of the *rhiABC* gene products may give some hints as to possible types of functions. While RhiA is cytoplasmic and RhiB is also likely to be cytoplasmic, the RhiC protein is most probably periplasmic. Many periplasmic proteins (in conjunction with membrane transport proteins) are involved in nutrient acquisition. Although no inner membrane transport proteins appear to be encoded by the *rhiABC* genes, other as yet unidentified *rhi* genes under the control of RhiR might encode such proteins. Therefore, one possible type of function is the utilization of some specific substrate(s) secreted by (some of) the legumes nodulated by *R. leguminosarum* bv. viciae. The symbiotic plasmid pRL1JI does contain genes that allow *R. leguminosarum* bv. viciae to utilize homoserine (29), a metabolite exuded in quantity by pea roots (53). However, homoserine does not influence *rhi* gene expression, and mutations affecting the *rhi* genes do not influence the ability of *R. leguminosarum* bv. viciae to utilize homoserine (data not shown). It is possible that some other rhizosphere metabolite is metabolized by the *rhi* gene products, and the limited homology between LacY and RhiA may point toward some sugar-containing metabolite.

ACKNOWLEDGMENTS

We thank Nick Brewin, Elmar Kannenberg, Fiona Hawkins, and Rob Okker for helpful discussions. Andrea Davies gave invaluable help with bacterial strains and plasmids, and we thank Alan Radford for help with computing. We also thank David Hopwood and Nick Brewin for critically reading the manuscript.

The work was supported by a grant in aid from the AFRC, European Community Sectoral Training Grants (to T.C. and A.E.), and a postdoctoral grant (to T.C.) from the Spanish Ministry of Education and Science.

REFERENCES

1. **Arico, B., J. F. Miller, C. Roy, S. Stibitz, D. Monack, S. Falkow, R. Gross, and R. Rappuoli.** 1989. Sequences required for expression of *Bordetella pertussis* virulence factors share homology with prokaryotic signal transduction proteins. Proc. Natl. Acad. Sci. USA **86:**6671–6675.
2. **Beringer, J. E.** 1974. R factor transfer in *Rhizobium leguminosarum*. J. Gen. Microbiol. **84:**188–198.
3. **Beynon, J. L., J. E. Beringer, and A. W. B. Johnston.** 1980. Plasmids and host range in *Rhizobium leguminosarum* and *Rhizobium phaseoli*. J. Gen. Microbiol. **120:**421–429.
4. **Boivin, C., L. R. Barran, C. A. Malpica, and C. Rosenberg.** 1991. Genetic analysis of a region of the *Rhizobium meliloti* pSym plasmid specifying catabolism of trigonelline, a secondary metabolite present in legumes. J. Bacteriol. **173:**2809–2817.
5. **Bradley, D. J., E. A. Wood, A. P. Larkins, G. Galfre, G. W. Butcher, and N. J. Brewin.** 1988. Isolation of monoclonal antibodies reacting with peribacteroid membranes and other components of pea root nodules containing *Rhizobium leguminosarum*. Planta **173:**149–160.
6. **Brickman, E., and J. Beckwith.** 1975. Analysis of the regulation of Escherichia coli alkaline phosphatase synthesis using deletions and ϕ80 transducing phages. J. Mol. Biol. **96:**307–316.
7. **Cole, S. T., and O. Raibaud.** 1986. The nucleotide sequence of the *malT* gene encoding the positive regulator of the *Escherichia coli* maltose region. Gene **42:**201–208.
8. **Collins, J. F., A. F. W. Coulson, and A. Lyall.** 1988. The significance of protein sequence similarities. Comp. Applic. Biosci. **4:**67–71.
9. **Cubo, M. T., A. M. Buendia, J. M. Beringer, and J. E. Ruiz-Sainz.** 1988. Melanin production by *Rhizobium* strains. Appl. Environ. Microbiol. **54:**1812–1817.
10. **Cuttings, S., and J. Mandelstam.** 1986. The nucleotide sequence and transcription during sporulation of the *gerE* gene of *Bacillus subtilis*. J. Gen. Microbiol. **132:**3013–3024.
11. **David, M., M. L. Daveran, J. Batut, A. Dedieu, O. Domergue, J. Ghai, C. Hertig, P. Boistard, and D. Kahn.** 1988. Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. Cell **54:**671–683.
12. **Davis, E. O., and A. W. B. Johnston.** 1990. Analysis of three *nodD* genes in *Rhizobium leguminosarum* biovar *phaseoli*; *nodD1* is preceded by *nolE*, a gene whose product is secreted from the cytoplasm. Mol. Microbiol. **4:**921–932.
13. **de Maagd, R. A., C. A. Wijffelman, E. Pees, and B. J. J. Lugtenberg.** 1988. Detection and subcellular localization of two Sym plasmid-dependent proteins of *Rhizobium leguminosarum* biovar *viciae*. J. Bacteriol. **170:**4424–4427.
14. **de Maagd, R. A., A. H. M. Wijfjes, H. P. Spaink, J. E. Ruiz-Sainz, C. A. Wijffelman, R. J. H. Okker, and B. J. J. Lugtenberg.** 1989. *nodO*, a new *nod* gene of the *Rhizobium leguminosarum* biovar *viciae* Sym plasmid pRL1JI, encodes a secreted protein. J. Bacteriol. **171:**6764–6770.
15. **Devine, J. H., C. Countryman, and T. O. Baldwin.** 1988. Nucleotide sequence of the *luxR* and *luxI* genes and the structure of the primary regulatory region of the *lux* regulon of *Vibrio fischeri* ATCC7744. Biochemistry **27:**827–842.
16. **Dibb, N. J.** 1983. Plasmid determined proteins in *Rhizobium leguminosarum*. Ph.D. thesis. University of East Anglia, Norwich, United Kingdom.
17. **Dibb, N. J., J. A. Downie, and N. J. Brewin.** 1984. Identification of a rhizosphere protein encoded by the symbiotic plasmid of *Rhizobium leguminosarum*. J. Bacteriol. **158:**621–627.
18. **Downie, J. A., G. Hombrecher, Q.-S. Ma, C. D. Knight, B. Wells, and A. W. B. Johnston.** 1983. Cloned nodulation genes of *Rhizobium leguminosarum* determine host-range specificity. Mol. Gen. Genet. **190:**359–365.
19. **Downie, J. A., Q.-S. Ma, C. D. Knight, G. Hombrecher, and A. W. B. Johnston.** 1983. Cloning of the symbiotic region of *Rhizobium leguminosarum*: the nodulation genes are between the nitrogenase genes and the *nifA*-like gene. EMBO J. **2:**947–952.
20. **Downie, J. A., and B. P. Surin.** 1990. Either of two *nod* gene loci can complement the nodulation defect of a *nod* deletion mutant of *Rhizobium leguminosarum* bv. *viciae*. Mol. Gen. Genet. **222:**81–86.
21. **Economou, A., W. D. O. Hamilton, A. W. B. Johnston, and J. A. Downie.** 1990. The *Rhizobium* nodulation gene *nodO* encodes a Ca^{2+}-binding protein that is exported without N-terminal cleavage and is homologous to haemolysin and related proteins. EMBO J. **9:**349–354.
22. **Economou, A., F. K. L. Hawkins, J. A. Downie, and A. W. B. Johnston.** 1989. Transcription of *rhiA*, a gene on a *Rhizobium leguminosarum* bv. *viciae* Sym plasmid, requires *rhiR* and is repressed by flavonoids that induce *nod* genes. Mol. Microbiol. **3:**87–93.
23. **Friedrich, M. J., and R. J. Kadner.** 1987. Nucleotide sequence of the *uhp* region of *Escherichia coli*. J. Bacteriol. **169:**3556–3563.
24. **Gottfert, M., P. Gross, and H. Hennecke.** 1990. Proposed regulatory pathway encoded by the *nodV* and *nodW* genes determinants of host specificity in *Bradyrhizobium japonicum*. Proc. Natl. Acad. Sci. USA **87:**2680–2684.
25. **Gross, R., B. Arico, and R. Pappuoli.** 1989. Families of bacterial signal-transducing proteins. Mol. Microbiol. **3:**1661–1667.
26. **Henikoff, S., J. C. Wallace, and J. P. Brown.** 1989. Finding protein similarities with nucleotide sequence databases. Methods Enzymol. **183:**111–132.
27. **Henner, D. J., M. Yang, and E. Ferrari.** 1988. Localization of *Bacillus subtilis sacU*(Hy) mutations to two linked genes with similarities to the conserved procaryotic family of two-component signalling systems. J. Bacteriol. **170:**5102–5109.
28. **Hirsch, P. R.** 1979. Plasmid-determined bacteriocin production by *Rhizobium leguminosarum*. J. Gen. Microbiol. **113:**219–228.
29. **Johnston, A. W. B., J. E. Burn, A. Economou, E. O. Davis, F. K. L. Hawkins, and M. J. Bibb.** 1988. Genetic factors affecting host range in *Rhizobium leguminosarum*, p. 378–384. *In* R. Palacios and D. P. S. Verma (ed.), Molecular genetics of plant-microbe interactions. APS Press, St. Paul, Minn.
30. **Kahn, D., and G. Ditta.** 1991. Modular structure of FixJ: homology of the transcriptional activator domain with the −35 binding domain of sigma factors. Mol. Microbiol. **5:**987–997.
31. **Knight, C. D., L. Rossen, J. G. Robertson, B. Wells, and J. A. Downie.** 1986. Nodulation inhibition by *Rhizobium leguminosarum* multicopy *nodABC* genes and analysis of early stages of plant infection. J. Bacteriol. **166:**552–558.
32. **Kondorosi, A., E. Kondorosi, M. John, J. Schmidt, and J. Schell.** 1991. The role of nodulation genes in bacterium-plant communication, p. 115–136. *In* J. K. Setlow (ed.), Genetic engineering, vol. 13. Plenum, New York.
33. **Lamb, J. W., G. Hombrecher, and A. W. B. Johnston.** 1982. Plasmid-determined nodulation and nitrogen-fixation abilities in *Rhizobium phaseoli*. Mol. Gen. Genet. **186:**449–452.
34. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
35. **Manoil, C., and J. Beckwith.** 1985. Tn*phoA*: a transposon probe for protein export signals. Proc. Natl. Acad. Sci. USA **82:**8129–8133.
36. **Murphy, P. J., N. Heycke, S. P. Trenz, P. Ratet, and F. J. de Bruijn.** 1988. Synthesis of an opine-like compound, a rhizopine, in alfalfa nodules is symbiotically regulated. Proc. Natl. Acad. Sci. USA **85:**9133–9137.
37. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53:**293–321.
38. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
39. **Rossen, L., C. A. Shearman, A. W. B. Johnston, and J. A. Downie.** 1985. The *nodD* gene of *Rhizobium leguminosarum* is autoregulatory and in the presence of plant exudate induces the *nodABC* genes. EMBO J. **4:**3369–3373.
40. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143:**161–178.
41. **Shadel, G. S., and T. O. Baldwin.** 1991. The *Vibrio fischeri* LuxR protein is capable of bidirectional stimulation of transcrip-

tion and both positive and negative regulation of the *luxR* gene. J. Bacteriol. **173**:568–574.

42. **Sharma, S., T. F. Stark, W. G. Beattie, and R. E. Moses.** 1986. Multiple control elements for the *uvrC* gene unit of *Escherichia coli*. Nucleic Acids Res. **14**:2301–2318.
43. **Spaink, H. P., D. M. Sheeley, A. A. N. van Brussel, J. Glushka, W. S. York, O. Geiger, E. P. Kennedy, V. N. Reinhold, and B. J. J. Lugtenberg.** 1991. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium leguminosarum*. Nature (London) **354**:125–130.
44. **Staden, R.** 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res. **10**:2951–2961.
45. **Staden, R.** 1984. Measurements of the effects that coding for a protein has on DNA sequence and their use for finding genes. Nucleic Acids Res. **12**:9509–9524.
46. **Stewart, V., J. Parales, Jr., and A. M. Merkel.** 1989. Structure of genes *narL* and *narX* of the *nar* (nitrate reductase) locus in *Escherichia coli* K-12. J. Bacteriol. **171**:2229–2234.
47. **Stout, V., and S. Gottesman.** 1990. RcsB and RcsC: a two-component regulator of capsule synthesis in *Escherichia coli*. J. Bacteriol. **172**:659–669.
48. **Stout, V., A. Torres-Cabassa, M. R. Maurizi, D. Gutnick, and S. Gottesman.** 1991. RcsA, an unstable positive regulator of capsular polysaccharide synthesis. J. Bacteriol. **173**:1738–1747.
49. **Surin, B. P., and J. A. Downie.** 1988. Characterization of the *Rhizobium leguminosarum* genes *nodLMN* involved in efficient host specific nodulation. Mol. Microbiol. **2**:173–183.
50. **Surin, B. P., J. M. Watson, W. D. O. Hamilton, A. Economou, and J. A. Downie.** 1990. Molecular characterisation of the nodulation gene *nodT* from two biovars of *Rhizobium leguminosarum*. Mol. Microbiol. **4**:245–252.
51. **Truchet, G., P. Roche, P. Lerouge, J. Vasse, S. Camut, F. de Billy, J.-C. Prome, and J. Denarie.** 1991. Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root organogenesis in alfalfa. Nature (London) **351**:670–673.
52. **Trumble, W. R., P. V. Viitanen, H. K. Sarkar, M. S. Poonian, and H. R. Kaback.** 1984. Site-directed mutagenesis of *cys sub*(148) in the Lac carrier protein of *Escherichia coli*. Biochem. Biophys. Res. Commun. **119**:860–867.
53. **van Egeraat, A. W. S. M.** 1975. The possible role of homoserine in the development of *Rhizobium leguminosarum* in the rhizosphere of pea deedlings. Plant Soil **42**:381–386.
54. **Von Heijne, G.** 1986. A new method for predicting signal cleavage sites. Nucleic Acids Res. **14**:4683–4690.
55. **Weinrauch, Y., N. Guillen, and D. Dubnau.** 1989. Sequence and transcription mapping of *Bacillus subtilis* competence genes *comB* and *comA*, one of which is related to a family of bacterial regulatory determinants. J. Bacteriol. **171**:5362–5375.