Nucleotide Sequence and Functional Analysis of the Two nifH Copies of Rhizobium ORS571

By FRANÇOISE NOREL AND CLAUDINE ELMERICH*
Unité de Physiologie Cellulaire, Département des Biotechnologies, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France

(Received 2.5 September 1986; revised 24 November 1986)

In Rhizobium ORS571, which is able to form both root and stem nodules on the tropical plant Sesbania rostrata and to grow in the free-living state at the expense of N₂, two copies of the nifH gene, which codes for the nitrogenase Fe-protein, were characterized. One copy, nifH₁, was localized in the nifHDK operon; the other, nifH₂, was localized elsewhere. nifH₁ and nifH₂ differed in only six nucleotides and both encoded a polypeptide of 296 amino acids with a single change at position 282, a serine in nifH₁ and a threonine in nifH₂. At position 102, arginine in the sequence Gly-Arg-Gly-Val-Ile-Thr might, as in Rhodospirillum rubrum, be the site of 'switch-off' inactivation. Compared with other nifH genes, the highest S_AB was found with Bradyrhizobium japonicum and Rhizobium sp. (Parasponia). A sequence identical (nifH₁) or very similar (nifH₂) to the nif consensus sequence was found upstream from the initiation codon. Homology with sequences upstream from Klebsiella pneumoniae nif promoters, which affect nifA-mediated activation, was also found in both cases. From mutants carrying either a single nifH₁ or nifH₂ deletion, or a double nifH₁–nifH₂ deletion, it appeared that both genes were functional ex planta and in planta and that no other nifH copy seemed to be present in ORS571. The optimal expression of nifH₁ and of nifH₂ might depend on different physiological conditions.

INTRODUCTION

The fast-growing Rhizobium strain ORS571, isolated from stem nodules of the tropical legume Sesbania rostrata, forms nodules on both stems and roots of the host plant (Dreyfus & Dommergues, 1981) and can grow in the free-living state at the expense of N₂ as the sole nitrogen source (Elmerich et al., 1982; Dreyfus et al., 1983). The nitrogenase complex that catalyses the reduction of N₂ to NH₃ has been isolated from a variety of diazotrophs (Eady & Smith, 1979) including Rhizobium ORS571 (Kush et al., 1985). The nitrogenase of Rhizobium ORS571 was subject to 'switch-off' when ammonia was added to a N₂-fixing culture (Kush et al., 1985). In all cases, nitrogenase was shown to be composed of two protein components, a MoFe-protein (component I) and an Fe-protein (component II). The MoFe-protein is a tetramer composed of two non-identical subunits (a and ß) and the Fe-protein is a dimer composed of two identical subunits (Eady & Smith, 1979). In a few cases, the complete or partial amino acid sequence of both proteins has been determined (Tanaka et al., 1977; Hausinger & Howard, 1982; Hase et al., 1984). Moreover, nitrogenase isolated from various organisms shows a remarkable similarity in component composition and the formation of active heterologous complexes has been reported (Emerich & Burris, 1978). The genes coding for the nitrogenase polypeptides (nifHDK) were initially characterized in Klebsiella pneumoniae, where they are organized in a single transcription unit (Elmerich et al., 1978; Merrick et al., 1978). The nifH gene encodes the two identical subunits of component II whereas nifD and nifK encode the ß and ß subunits of component I, respectively. As shown by hybridization with K. pneumoniae nif

Abbreviation: ORF, open reading frame.

0001-3736 © 1987 SGM
cloning of the structural genes for nitrogenase from a variety of N\(_2\) fixers including bacteria of the genus Rhizobium (Hennecke, 1981; Elmerich et al., 1982; Quinto et al., 1982; Scott et al., 1982; Downie et al., 1983; Scott et al., 1983a, b; Fuhrmann & Hennecke, 1984). In Rhizobium ORS571, identification and cloning of DNA fragments homologous to \(K.\) pneumoniae \(nifH, D, K\) and \(E\) were previously reported (Elmerich et al., 1982; Norel et al., 1985a; Donald et al., 1986). By site-directed mutagenesis, it was shown that these genes are organized in two independent transcription units, \(nifHDK\) and \(nifE\) (Donald et al., 1986; Denède et al., 1987). The nucleotide sequence of \(nifH\) from ten eubacteria and from a methanogenic archaeabacterium is now established: \(Anabaena\) 7120 (Mevarech et al., 1980), \(K.\) pneumoniae (Scott et al., 1981; Sundaresan & Ausubel, 1981), \(R.\) meliloti (Torök & Kondorosi, 1981), \(Rhizobium\) sp. (\(Parasponia\)) (Scott et al., 1983a), \(R.\) trifolii (Scott et al., 1983b), \(Bradyrhizobium\) japonicum (Fuhrmann & Hennecke, 1984), \(Azotobacter\) vinelandii (Brigle et al., 1985), \(Rhizobium\) phaseoli (Quinto et al., 1985), \(Clostridium\) pasteurianum (Chen et al., 1986), \(Azotobacter\) chroococcum (Robson et al., 1986) and \(Methanococcus\) voltae (Souillard & Sibold, 1986). From these results, \(nifH\) appears as the most conserved translated gene in bacteria. Another interesting feature of \(nifH\) is the presence of multiple copies of the gene in several diazotrophs including \(R.\) phaseoli (Quinto et al., 1982), \(Anabaena\) 7120 (Rice et al., 1982), \(B.\) japonicum (Prakash & Atherly, 1984), \(Calothrix\) sp. (Kallas et al., 1983), \(Rhodopseudomonas\) capsulata (Scolnik & Haselkorn, 1984), \(Azotobacter\) chroococcum (Robson et al., 1986), \(C.\) pasteurianum (Chen et al., 1986), and \(Rhizobium\) ORS571 (Norel et al., 1985a, b; Donald et al., 1986). A functional analysis of the reiterated copies was done in \(Rps.\) capsulata (Scolnik & Haselkorn, 1984), \(R.\) phaseoli (Quinto et al., 1985), and \(Azotobacter\) chroococcum (Robson et al., 1986).

Because of the unusual properties of strain ORS571, forming both root and stem nodules, fixing N\(_2\) under both free-living and symbiotic conditions and having a nitrogenase subject to ‘switch-off’ by inactivation of component II, we decided to determine the nucleotide sequence of the two \(nifH\) copies and to study their expression. Results reported here show that the two copies are almost identical, closely related to \(nifH\) of \(B.\) japonicum (Fuhrmann & Hennecke, 1984) and of \(Rhizobium\) sp. (\(Parasponia\)) (Scott et al., 1983a), and that they are functional both \(ex\) \(planta\) and \(in\) \(planta\).

**METHODS**

**Bacterial strains, bacteriophages, plasmids and growth conditions.** Bacterial strains and bacteriophages are listed in Table 1. \(Rhizobium\) strains were grown as described by Elmerich et al. (1982). \(Escherichia\) coli strains were grown in Luria Broth medium (Maniatis et al., 1982). The final concentrations of antibiotics for \(Rhizobium\) are indicated in Table 5; those for \(E.\) coli were (\(\mu\)g ml\(^{-1}\)): carbenicillin, 100; chloramphenicol, 25; tetracycline, 5; kanamycin, 40; gentamicin, 20; spectinomycin, 25. Carbenicillin was used to select the Amp\(^8\) phenotype. Plasmids are listed in Table 1 or described in Fig. 1. Plasmids containing \(K.\) pneumoniae \(nif\) genes were previously described (Norel et al., 1985a) and are not listed in Table 1. Plasmid pRK2013 was used to mobilize IncP Tra\(^{-}\) plasmids. Marker exchange in ORS571 was done using pPH11J and R751 as described by Denède et al. (1987). The cloning vector pHE3 and the methods for direct selection of recombinant DNA in \(E.\) coli RR28 were previously described (Hennecke et al., 1982).

**Molecular biology techniques.** DNA isolation, endonuclease digestion, ligation, alkaline phosphatase and Bal31 treatment, filling-in of protruding ends, DNA transformation, nick translation, nitrocellulose filter hybridization, and colony hybridization were done essentially as described by Maniatis et al. (1982). DNA sequencing was done by the method of Sanger et al. (1977) with phage M13 derivatives (Messing & Vieira, 1982) that were propagated in \(E.\) coli TG1. Sequence data were compiled by computer analysis by using the program devised by Staden (1983). To prepare a Sp\(^8\) cartridge with SalI–SmaI extremities, a 3.4 kb BamHI fragment from pGV764 containing the Sp\(^8\) gene was cloned into pUC9 to yield pPC942. From this plasmid, the SalI–SmaI fragment containing the Sp\(^8\) gene was recovered.

**Plant test and nitrogenase assay.** Root inoculation of \(Sesbania\) rostrata plantlets and nitrogenase assays were done as described by Elmerich et al. (1982).
RESULTS AND DISCUSSION

Cloning of the second copy of nifH

Hybridization, with a nifH intragenic probe prepared from K. pneumoniae or R. phaseoli, of Rhizobium ORS571 total DNA digested with several restriction endonucleases revealed two fragments in most cases, except with PstI where a single 9.5 kb fragment was detected (Norel et al., 1985a). The same result was observed when the R1 fragment (Fig. 1), was used as a probe (Norel et al., 1985a). This suggested that in ORS571 the two nifH copies were carried by two PstI fragments of about the same size. The nifH copy adjacent to nifDK will be referred to as nifH1 and the second nifH copy as nifH2.

To clone the PstI fragments containing nifH1 and nifH2, size-fractionated fragments (between 7 and 12 kb) of ORS571 total DNA were ligated at the unique PstI site of pHE3. Transformants were screened by colony hybridization using the R1 fragment as probe and recombinant plasmids containing a 9.5 kb PstI DNA insert were analysed. The physical map of pRS68, which contains nifH1 is shown in Fig. 1. Another plasmid, termed pRS54 (Fig. 1), appeared to carry nifH2. The localization of nifH2 in the insert was established by using the R1 fragment as a probe. Homology to R1 was limited to the 1.65 kb PstI-SmaI fragment and a strong homology was found within the 0.6 kb XhoI fragment (R2 in Fig. 1). The R2 fragment was used as a probe against pRS hydrolysed by XhoI and BamHI. Only the R1 fragment hybridized with R2. Moreover, R1 and R2, when used as probes against ORS571 total DNA restricted by several endonucleases, revealed hybridizing fragments of similar size (e.g. 13 and 16 kb BamHI fragments, 7 and 20 kb BglII fragments, 2-8 and 20 kb SmaI fragments, a 0.6 kb XhoI fragment, and a 9.5 kb PstI fragment) (data not shown). The restriction map of the insert carried by pRS54 (Fig. 1) was presented in a preliminary report (Norel et al., 1985b). The entire 9.5 kb PstI fragment from pRS54 was used as a probe against plasmids carrying several nif genes of K. pneumoniae. No homology was detected except with nifH.
Fig. 1. Restriction map of *Rhizobium* ORS571 DNA fragments carrying the first copy of *nifH* (*nijH*) in the *nif* *HDKE* cluster and the second copy of *nifH* (*nijH2*). (a) Physical map of the DNA region carrying the *nif* *HDKE* cluster. The localization of the *nif* *HDKE* genes and the direction of transcription (Norel *et al.*, 1985a) are indicated by boxes and arrows, respectively. The localization of the adjacent *nif* region is indicated by --- (Denèche *et al.*, 1987). Plasmids pRS2 and pRS82 were previously described (Elmerich *et al.*, 1982; Denèche *et al.*, 1987). Plasmid pRS68 contained the 9.5 kb *PstI* fragment cloned into pHE3. Plasmid pRS72, used as a source of DNA for sequencing, was constructed as follows. The 1 kb *PstI*--*SalI* fragment carrying *nijH1* was cloned into pUC8 to yield pRS67 (not shown). The adjacent 0.5 kb *SalI*--*XhoI* fragment was subsequently cloned at the *SalI* site of pRS67 in the correct orientation to yield pRS72. The R1 fragment was purified from pRS72. (b) Physical map of the DNA region containing *nijH2* whose localization is indicated by 5-pointed stars. The 9.5 kb *PstI* fragment was cloned into pHE3 and pLA29-17 to yield pRS54 and pRS55, respectively. The 1-65 kb *PstI*--*SmaI* fragment was cloned into pUC9 and pUC8 to yield pRS56 and pRS57 (not shown), respectively, which were used for sequencing. The R2 fragment was purified from pRS54. Plasmid pRS71 was constructed as follows. The 1.1 kb *SalI*--*SmaI* fragment of pRS54, which contained *nijH2*, was substituted by the 3.4 kb *SalI*--*SmaI* fragment of pPC942, which contained a *Spf* gene, to yield pRS66 (not shown). The deletion was subsequently transferred into pRS55, by homologous recombination between pRS55 and pRS66 in *E. coli* MC1061. Restriction sites: *B*, *BamHI*; *Bg*, *BglII*; *H*, *HindIII*; *P*, *PstI*; *Pv*, *PvuI*; *R*, *EcoRI*; *S*, *SalI*; *Sph*, *SphI*; *Sm*, *SmaI*; *X*, *XhoI*.

therefore assumed that no DNA sequence homologous to any *nif* gene of *K. pneumoniae* was located in the vicinity of *nijH2*. Plasmids pRS68 and pRS54 were both used as probes against ORS571 total DNA. This allowed us to establish or to corroborate the physical map of DNA fragments surrounding the probes (Fig. 1). Our data are in agreement with restriction maps of the two *nifH* regions of ORS571 also reported by Donald *et al.* (1986).

**DNA sequence analysis of the two *nifH* coding regions**

The complete nucleotide sequence of ORS571 *nifH1* and *nijH2* was established. Fig. 2 gives an outline of the sequencing strategy employed. About 1500 nucleotides were sequenced in each region.

**Nucleotide sequence of *nifH1***. The orientation of *nifH1* was deduced from hybridization experiments (Norel *et al.*, 1985a). Computer analysis of the three possible reading frames in both strands revealed a single frame large enough in the correct orientation to encode a *nifH*
DNA sequence of *Rhizobium ORS571* nifH copies

Fig. 2. Strategy used for sequencing *Rhizobium ORS571* nifH genes. Arrows indicate direction and length of sequenced fragments. A, Sequences obtained after cloning nifH1 restriction fragments from pRS2 and pRS72 and nifH2 restriction fragments from pRS54 and pRS56 into M13 derivatives. B, Fragments obtained after *Bal31* treatment of pRS72 (nifH1) and pRS56(nifH2) and subcloning into M13 derivatives. Restriction sites as in Fig. 1. H, *Hae*III. Stars indicate the extent and direction of transcription of the nifH genes.

Polypeptide. The nucleotide sequence of nifH1 and the predicted amino acid sequence are shown in Fig. 3. From the presumptive ATG start codon at position +1 to the TGA stop codon at position 889, an open reading frame (ORF) accounts for a polypeptide of 296 amino acids. Purified nitrogenase component II had an estimated *M*ₐ of 36000 and a pI of 5.3 (Kush et al., 1985). The *M*ₐ of the polypeptide deduced from the sequence is 31951, thus slightly different from that of purified component II. The discrepancy between the two estimations may be due to the acidic nature of the polypeptide, which in SDS-PAGE migrates as a polypeptide with a higher apparent *M*ₐ. As no other in-frame ATG was found upstream from the presumptive initiation codon and as the *M*ₐ matches with other nifH products (Eady & Smith, 1979), we assumed that the identified ORF was the correct one. The sequence upstream from the start codon was determined for about 450 bp. A 5'-AGGAG-3' sequence, which is a potential ribosome-binding site (Shine & Dalgarno, 1974), is located seven nucleotides upstream from the suggested initiation codon.

A second ORF, located downstream from nifH1, was found. This ORF starts with an ATG located in position 959 and is preceded by a 5'-AGGA-3' sequence eight nucleotides upstream from the potential initiation codon. The predicted amino acid sequence of this ORF is presented.
in Fig. 3. On the basis of amino acid sequence homology with the nifD gene from B. japonicum (Kaluza & Hennecke, 1984), this ORF has been assigned to the ORS571 nifD gene. It is well established that in K. pneumoniae nifH is the first gene in a polycistronic operon, followed by nifD, nifK and nifY. A similar arrangement (nifHDK) was found in two fast-growing rhizobia, R. meliloti (Ruvkun et al., 1982) and R. leguminosarum (Schetgens et al., 1984). In two members of the slow-growing rhizobia, B. japonicum and Rhizobium sp. (Parasponia), nifD and nifK form one operon and nifH is located elsewhere (Kaluza et al., 1983; Scott et al., 1982). A similar organization (i.e. nifH and nifDK transcribed separately) was found in strain IRC78 which is representative of the tropical cowpea Rhizobium miscellanly (Yun & Szalay, 1984). In the region flanking the 3’ end of ORS571 nifH1 coding region, an inverted repeat was found 11 nucleotides downstream from the nifH1 termination codon (Fig. 3). The mRNA transcribed from this region could potentially form a stem characteristic of a loop terminator structure. The potential stem is formed by 10 consecutive GC pairs. The structure, with a free energy of \(\Delta G = -38.2 \text{ kcal mol}^{-1} (\text{ -159.8 kJ mol}^{-1})\) at 25 °C as estimated by the method of Tinoco et al. (1973), should be stable. Such a GC-rich inverted repeat structure was already described in B. japonicum in the 3’ end flanking region of nifH and nifK, and it was suggested that these sequences could function as transcription terminators (Fuhrmann & Hennecke, 1984; Thöny et al., 1985). By using Tn5 mutagenesis in the ORS571 nifHDK homologous sequences, we showed that the ORS571 nitrogenase structural genes were organized into a single transcription unit (Denéfle et al., 1987). However, the potential inverted repeat structure preceding nifD might have a regulatory effect on ORS571 nifDK transcription. Similar inverted repeat structures have also been described downstream from each of the nitrogenase structural genes in A. vinelandii and in A. chroococcum (Brigle et al., 1985; Kennedy et al., 1985). In A. vinelandii, the nifHDK genes are part of the same operon (Brigle et al., 1985) and in A. chroococcum, four transcripts were found to hybridize to a nifH1 probe during nitrogenase derepression (Jones et al., 1984; Kennedy et al., 1985). The size of the transcripts was compatible with a termination near the stop codons of nifH, nifD and nifK within the nifHDK intergenic regions (Kennedy et al., 1985). Thus a transcriptional regulatory function of the inverted repeat structures could not be ruled out.

**Nucleotide sequence of nifH2.** Computer analysis of the sequenced 1.65 kb PstI-SmaI fragment revealed a single ORF large enough to encode a nifH product. Moreover the ORF, which would code for a product of 296 amino acids, shows a high degree of homology, both at the nucleotide and at the amino acid sequence level, with the nifH1 coding region. The nucleotide and amino acid sequences of the nifH2 coding region are shown in Fig. 3. No in-frame ATG was found upstream from the potential start codon at position +1. Only six nucleotides were changed between nifH1 and nifH2. Five of these changes affect the third nucleotide of the corresponding codon and have no effect at the amino acid level. The only changed amino acid between the two genes is residue 282, a serine in nifH1 and a threonine in nifH2, two amino acids which are chemically related.

Reiterated nifH copies have been found in other N2-fixing organisms. In C. pasteurianum (Chen et al., 1986), one of the three nifH sequences that were identified (namely nifH1) encodes a 273 amino acid sequence identical to that determined with the isolated Fe-protein. A second

---

**Fig. 3.** Complete nucleotide sequence of *Rhizobium* ORS571 nifH1 and nifH2 coding and flanking regions. The nucleotide sequence of the nifH1 non-coding strand is shown in lines with 1 in the margin. In the ORFs, the derived amino acids are printed above the corresponding codons. Numbering of both nucleotides and amino acids starts at the beginning of the nifH1 coding sequence. The nucleotide sequence of nifH2 non-coding strand is shown in lines with 2 in the margin. Nucleotides in the nifH2 coding region identical to nifH1 are indicated by dots. Changes in nucleotides or amino acids are mentioned in the text. Downstream from nifH1, the first 42 amino acid residues of nifD are presented. Amino acids identical to the B. japonicum nifD gene (Kaluza & Hennecke, 1984) are indicated by +. Structural features in the nucleotide sequence of nifH1 and nifH2 flanking regions are marked as follows: boxed region, consensus nif promoter sequence (Ausubel, 1984); underlined sequences, ribosome-binding site and consensus for upstream activator sequence of nif promoters (Buck et al., 1986); arrows, inverted repeat. *, Nucleotides also found in B. japonicum nifH 5’ non-coding region (Fuhrmann & Hennecke, 1984).
copy was sequenced and, in the 272 amino acids that it probably encodes, 23 residues differed from the \textit{nifH1}-encoded Fe-protein. The third \textit{nifH}-related sequence was found to be less homologous to \textit{nifH1} and \textit{nifH2}. On the other hand, the coding regions of the three \textit{nifH} copies found in \textit{R. phaseoli} are identical (Quinto et al., 1985).

The flanking regions of the \textit{nifH2} coding sequence were determined for about 500 nucleotides upstream from the initiation codon and about 250 nucleotides downstream from the stop codon. A 5'-AGGA-3' potential ribosome binding site was found nine nucleotides upstream from the \textit{nifH2} potential start codon. From data reported in Fig. 3, it appeared that nucleotide sequences upstream from \textit{nifH1} and \textit{nifH2} are different. This situation differs from that found in \textit{R. phaseoli} (Quinto et al., 1985) where the sequences are identical up to nucleotide 96 upstream from the presumptive initiation codon. Beyond, one of the \textit{nifH} regions diverges while the two others remain identical. The \textit{R. phaseoli} \textit{nifH} potential transcriptional regulatory signals are located within the region which is conserved upstream from the three \textit{nifH} genes. In the region downstream from the \textit{nifH2} coding sequence (Fig. 3) an inverted repeat structure is located seven nucleotides after the stop codon. A potential stem is formed by nine consecutive basepairs, eight of which are GC pairs. It would be slightly less stable than that found in the case of \textit{nifH1}. No ORF was detected downstream from \textit{nifH2}.

**Potential promoter regions in the two \textit{nifH} sequences.** The two \textit{nifH} promoter regions were compared to other known \textit{nif} promoters and to the \textit{nif} promoter consensus sequence which was shown to be located 11 bp upstream from the transcription start of \textit{nif} genes or of other genes subject to \textit{ntrC}/\textit{nifA} control (Ausubel, 1984). Striking similarities with the \textit{nif} consensus sequence were found, in particular \textit{nifH1} (Table 2). No sequence reminiscent of a typical \textit{E. coli} RNA polymerase binding site (5'-TTGACTAC-G-3') (Rosenberg & Court, 1979) was found in either of the two \textit{nifH} promoter regions. This was in agreement with the fact that \textit{nifH}, \textit{D} and \textit{K} products were not expressed when plasmid pRS2 was introduced into \textit{E. coli} minicells. The sequences upstream from \textit{K. pneumoniae} \textit{nif} promoters, which affect \textit{nifA}-mediated activation, contain a 5'-A-G-N_7-TGT-N_4-T-N_5-ACA-3' consensus sequence (Buck et al., 1986). Two regions similar to this consensus were found in the \textit{nifH2} promoter region between nucleotides −253 to −238 and between nucleotides −282 to −267 (Fig. 3). In the \textit{nifH1} promoter region, only the sequence between nucleotides −280 and −265 was identical to the consensus (Fig. 3). However, since no specific deletions were constructed in these sequences, we do not know if they are required for a fully efficient transcription of the \textit{nifH} gene as was the case in \textit{B. japonicum} (Alvarez-Morales et al., 1986). It also remains to be determined whether the slight difference between the two ORS571 \textit{nifH} genes consensus sequences might account for a different regulation of their expression.

The two \textit{nifH} non-coding regions were compared to the \textit{nifH} promoters from various rhizobia. Little homology was found to the \textit{nifH1} non-coding region. However, the \textit{nifH2} sequence located downstream from the potential promoter region, between nucleotides −152 and −76, exhibited

<table>
<thead>
<tr>
<th>Origin</th>
<th>Position</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{nif} consensus†</td>
<td>−</td>
<td>5'-CTGGCACNNNNNTTGCA-3'</td>
</tr>
<tr>
<td>\textit{K. pneumoniae nifH}‡</td>
<td>−57 to −41</td>
<td>5'-CTGGTATGTTCCCTGCA-3'</td>
</tr>
<tr>
<td>\textit{B. japonicum nifH}§</td>
<td>−178 to −162</td>
<td>5'-TTGGCACGGCTGTGCT-3'</td>
</tr>
<tr>
<td>ORS571 \textit{nifH1}</td>
<td>−196 to −180</td>
<td>5'-CTGGCACACCCCGTGCA-3'</td>
</tr>
<tr>
<td>ORS571 \textit{nifH2}</td>
<td>−169 to −153</td>
<td>5'-ATGGCACGGTCTGTGCT-3'</td>
</tr>
</tbody>
</table>

DNA sequence of Rhizobium ORS571 nifH copies

Table 3. Binary matching coefficients (S_{AB}) calculated for Rhizobium ORS571 nifH1 and nifH from different N_2-fixing organisms at the DNA and deduced amino acid sequence levels

\[ S_{AB} = 2 \times \frac{\text{(no. of identical nucleotides or amino acids between A and B)}}{\text{(total no. of nucleotides or amino acids in A) + (total no. of nucleotides or amino acids in B)}} \]

<table>
<thead>
<tr>
<th>Organism</th>
<th>Nucleotide sequence</th>
<th>Amino acid sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anabaena 7120</td>
<td>0.63</td>
<td>0.72</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td>Azotobacter vinelandii</td>
<td>0.72</td>
<td>0.73</td>
</tr>
<tr>
<td>A. chroococcum</td>
<td>0.72</td>
<td>0.71</td>
</tr>
<tr>
<td>Rhizobium meliloti</td>
<td>0.72</td>
<td>0.79</td>
</tr>
<tr>
<td>R. phaseoli</td>
<td>0.74</td>
<td>0.80</td>
</tr>
<tr>
<td>R. trifolii</td>
<td>0.71</td>
<td>0.77</td>
</tr>
<tr>
<td>Bradyrhizobium japonicum</td>
<td>0.78</td>
<td>0.92</td>
</tr>
<tr>
<td>Rhizobium sp. (Parasponia)</td>
<td>0.80</td>
<td>0.93</td>
</tr>
<tr>
<td>Clostridium pasteurianum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nifH1</td>
<td>0.51</td>
<td>0.61</td>
</tr>
<tr>
<td>nifH2</td>
<td>0.50</td>
<td>0.62</td>
</tr>
<tr>
<td>Methanococcus voltae</td>
<td>0.44</td>
<td>0.49</td>
</tr>
</tbody>
</table>

a strong homology to sequences located downstream from the nifH promoter of B. japonicum (Fig. 3) and Rhizobium sp. (Parasponia) (data not shown).

The significance of this conserved region is not yet known. In B. japonicum this presumably untranslated leader region which could potentially form secondary structures is not required for nifA-mediated expression of nifH promoter in E. coli. (Kaluza et al., 1985). However, it cannot be ruled out that this region is required for optimal nifH expression in the homologous host (Kaluza et al., 1985).

Comparison of amino acid and DNA sequences of nifH1 and nifH from other organisms. The sequence of the Fe-protein in ORS571 shows similarities to the Fe-protein sequences from other organisms. In particular, the protein contains the five cysteine residues which are invariant in all species and which are supposed to be involved in the binding of the iron acid-labile sulphur cluster to the apoprotein. The sequence conservation is particularly clustered around the cysteine residues and in the N-terminal region which, according to Robson (1984), might contribute to an adenine-nucleotide-binding domain. Based on the nucleotide and amino acid sequences, similarity coefficients (S_{AB}) (Fox et al., 1977) were calculated between the ORS571 nifH1 gene and other nifH genes (Table 3). The highest degree of homology was found with two members of the slow-growing rhizobia, Rhizobium sp. (Parasponia) and B. japonicum. It was previously suggested that nifH genes and 16S rRNA within a given bacterium may have evolved in parallel (Hennecke et al., 1985). Moreover, from analysis of both 16S rRNA and nifH homology between rhizobia it was found that fast- and slow-growing rhizobia fall into two groups (Hennecke et al., 1985). Based on nifH S_{AB} comparison, it appears that ORS571 is much more related to rhizobia belonging to the slow growers, in agreement with taxonomic studies.

In ORS571, nitrogenase is subject to 'switch-off' which, as in photosynthetic bacteria, results from a specific inactivation of the Fe-protein (Kush et al., 1985). This phenomenon, which was discovered with free-living bacteria, was also observed in bacteroids (this laboratory, unpublished). In Rhodospirillum rubrum, inactivation was shown to be due to a covalent modification of the Fe-protein. The modifying group is adenosine diphosphoribose linked through the terminal ribose to a guanidino nitrogen of arginine. The modified site in the protein is Gly-Arg(ADP-ribose)-Gly-Val-Ile-Thr (Pope et al., 1985). This amino acid sequence, which is located around residues 100 and 106, is highly conserved in all nifH products whose sequence has been established or deduced from the nucleotide sequence, with the exception of Anabaena 7120 (Mevarech et al., 1980) and C. pasteurianum (Chen et al., 1986) where Val is replaced by Ile. In
ORS571 nifH1 and nifH2 products, the Gly-Arg-Gly-Val-Ile-Thr sequence is found between residues 101 and 107 (Fig. 3). Although some features of the ‘switch off/on’ phenomenon in ORS571 seemed to be different from that observed in photosynthetic bacteria (Kush et al., 1985), it is possible that Arg in position 102 is the site of the ‘switch-off’ modification.

G + C content and codon usage in ORS571 nifH genes. Codon usage in ORS571 nifH genes is very biased (Table 4). Four amino acids are coded by a single coding triplet (glutamine, lysine, aspartate, and cysteine). An asymmetrical codon usage occurs for isoleucine, glutamate, arginine, and glycine. In total, 21 of the 60 coding triplets are not used. The absence of tryptophan, which is another feature of most nijH genes, should be noted. The G + C content in the nifH coding region is 64% and a very high G + C content was found in the third position of codons, 92% as compared to 61% in the first position and to 39% in the second. This probably reflects the high G + C content of the ORS571 genome, which is 66% (Dreyfus, 1982).

Functional analysis of the two nifH copies

We previously reported that nifH1 deletion mutant 57182 had a Nif+ Nod+ Fix+ phenotype (Denifle et al., 1987). The plasmid used to construct this mutant was pRS82 (Fig. 1) in which the two XhoI fragments carrying nifHI were replaced by a KmR cartridge from Tn5. It is likely that in pRS82 the nif/DK transcription is initiated from the KmR gene promoter. The Nif+ Fix+ phenotype of mutant 57182 led us to conclude that a second copy of nifH was functional (Denifle et al., 1987). However, we had no proof that this second nifH gene was the nifH2 copy detected by hybridization and studied here. Thus, it was of interest to obtain both a nifH2 mutant and a double nifHI-nifH2 mutant.

Table 5 indicates the procedure followed to obtain strain 57171 that carries a deletion of nifH2 and strain 57172 that carries a double nifH1-nifH2 deletion. The method, similar to that described by Ruvkun & Ausubel (1981), was based on the incompatibility between pRK290 and pPH1JI and between pLA29-17 and R751. It was checked by hybridization that deletions in strains 57171 and 57172 had occurred at the correct location (data not shown). The mutants were used to inoculate Sesbania rostrata plantlets. Nitrogenase activity both ex planta and in planta was determined. The double deletion mutant 57172 had a Nif− Nod+ Fix− phenotype, indicating that no functional nifH gene other than nifH1 and nifH2 was present in ORS571. Moreover,
DNA sequence of Rhizobium ORS571 nifH copies

Selection was performed on minimal LSN solid medium (Elmerich et al., 1982), containing appropriate antibiotics: gentamicin (Gm), 10 μg ml⁻¹; kanamycin (Km), 100 μg ml⁻¹; spectinomycin (Sp), 50 μg ml⁻¹; tetracycline (Tc), 10 μg ml⁻¹; trimethoprim (Tp), 400 μg ml⁻¹.

Table 5. Site-directed mutagenesis of the two nifH copies

<table>
<thead>
<tr>
<th>Donor strain</th>
<th>Recipient strain</th>
<th>Selected marker</th>
<th>Frequency</th>
<th>Designation of selected strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>HB101(pRK2013)</td>
<td>ORS571 (pRS71)</td>
<td>Tc–Sp</td>
<td>10⁻²</td>
<td>ORS571(pRS71)</td>
</tr>
<tr>
<td>MC1061(pRS71)</td>
<td>ORS571(pRS71)</td>
<td>Tp</td>
<td>10⁻³</td>
<td>57171*</td>
</tr>
<tr>
<td>HB101(R751)</td>
<td>57171</td>
<td>Tp–Sp (100% Tc⁵)</td>
<td>10⁻⁷</td>
<td>57171(pRS82)</td>
</tr>
<tr>
<td>HB101(pRK2013)</td>
<td>57171</td>
<td>Tp–Tc</td>
<td>10⁻⁹</td>
<td>57172†</td>
</tr>
<tr>
<td>MC1061(pRS82)</td>
<td>57171 (pRS82)</td>
<td>Gm–Km</td>
<td>10⁻⁷</td>
<td>57171(pRS82)</td>
</tr>
</tbody>
</table>

* Strain 57171 contains plasmid R751.
† Strain 57172 contains plasmid pPH1JI.

Strains 57171 and 57182 had a nitrogenase specific activity ex planta which was 70% and 30%, respectively, of that observed with the wild-type strain. This could suggest that in the wild-type strain nifH1 accounts for about 70% of N₂ fixation ex planta and nifH2 for 30%, in agreement with the value of 30% of ORS571 total nitrogenase activity assigned to the nifH2 gene that was previously calculated from data obtained with crude extracts of mutant 57182 (Denèfle et al., 1987). An unexpected result was the divergence obtained in planta, since the nitrogenase specific activity of strain 57182 was about 80%, whereas with strain 57171 the nitrogenase activity was only 15% and the inoculated plantlets looked in poor condition. As the amino acid sequences of the two nifH genes deduced from the nucleotide sequences are almost identical, it is likely that the two Fe-proteins have similar specific activities. Two hypotheses could account for this observation: (i) the deletion in strain 57171 might have a polar effect on the transcription of a fix gene located downstream; (ii) the expression of the two nifH genes might be slightly different ex planta and in planta. The first hypothesis seems unlikely since the entire 1.65 kb PstI–SmaI fragment, deleted in strain 57171, was sequenced and no other ORF was detected downstream from nifH2. In R. phaseoli, it was shown that at least two of the three nifH genes can be functionally expressed (Quinto et al., 1985). In Rps. capsulata only one nifH gene is functional in the wild-type strain but a second copy can be activated when the first one is mutated (Scolnik & Haselkorn, 1984). In A. chroococcum, the second nifH copy was transcribed under molybdenum deficiency conditions (Robson et al., 1986). The significance of these reiterated nifH sequences whether functional or not under defined physiological conditions is not yet known but their existence raises a puzzling and interesting question which might be solved by further investigations based, in particular, on the study of nifH–lacZ fusions.

The authors wish to thank Dr J.-P. Aubert for discussions, Dr D. Dean for reading the manuscript, Dr R. Deblaere for the gift of pGV764, Ms N. Desnoues and A. Paquelin for skilful technical assistance, and the personnel of the Centre d’Informatique Scientifique at the Pasteur Institute for computer facilities. This work was supported by a grant (no. GBI-5-028F) from the Commission of the European Communities and by research funds from the University Paris 7.

REFERENCES


KALLAS, T., REBIERE, M. C., RIPPKA, R. & TANDEAU DE MARSAC, N. (1983). The structural nif genes of the cyanobacteria *Gloeoeche* sp. and *Calothrix* sp. share homology with those of *Anabaena* sp. but the *Gloeoeche* genes have a different arrangement. *Journal of Bacteriology* 155, 427–431.


