Sequence and structural organization of a *nifA*-like gene and part of a *nifB*-like gene of *Herbaspirillum seropedicae* strain Z78

E. M. Souza,1 S. Funayama,1 L. U. Rigo,1 M. G. Yates2 and F. O. Pedrosa1*

1Universidade Federal do Paraná, Department of Biochemistry, PO Box 19046, 81531 Curitiba, PR, Brazil
2AFRC Institute of Plant Science Research, Nitrogen Fixation Laboratory, University of Sussex, Brighton, UK

(Received 12 December 1990; revised 13 February 1991; accepted 18 March 1991)

The deduced amino acid sequence derived from the sequence of a fragment of DNA from the free-living diazotroph *Herbaspirillum seropedicae* was aligned to the homologous protein sequences encoded by the *nifA* genes from *Azorhizobium caulinodans*, *Rhizobium leguminosarum*, *Rhizobium meliloti* and *Klebsiella pneumoniae*. High similarity was found in the central domain and in the C-terminal region. The *H. seropedicae* putative *NifA* sequence was also found to contain an interdomain linker similar to that conserved among rhizobial *NifA* proteins, but not *K. pneumoniae* or *Azotobacter vinelandii*. Analysis of the regulatory sequences found 5' from *nifA* indicated that the expression of this gene in *H. seropedicae* is likely to be controlled by NifA, NtrC and RpoN, as judged by the presence of specific NifA- and NtrC-binding sites and characteristic -24/-12 promoters. Possible additional regulatory features included an 'anaerobox' and a site for integration host factor. The N-terminus of another open reading frame was found 3' from *nifA* and tentatively identified as *nifB* by amino acid sequence comparison. The putative *nifB* promoter sequence suggests that expression of *H. seropedicae* *nifB* may be activated by NifA and dependent on RpoN.

Introduction

*Herbaspirillum seropedicae* is a free-living diazotroph found in soils and in the rhizosphere of several grasses, and has been classified as a member of the Spirillaceae (Baldani et al., 1986). It shows a broader pH range for N2-dependent growth and higher nitrogenase tolerance to oxygen than *Azospirillum* spp. It also shows low (<20%) rRNA homology to *Azospirillum* spp. (Falk et al., 1986).

In *Azospirillum brasilense*, regulation of *nif* expression may be analogous to that of *Klebsiella pneumoniae* (Pedrosa & Yates, 1984). Potential *nifA* and *ntrC* mutants of *Azospirillum brasilense* were complemented for nitrogen fixation by homologous genes from *K. pneumoniae*. However, the organization of these genes and the control of nitrogen assimilation in *Azospirillum brasilense* may be different from that in *K. pneumoniae* (Bozouklian & Elmerich, 1986; Bozouklian et al., 1986).

The dependence of *nif* genes on RpoN and NifA for expression has also been demonstrated in *Rhizobium meliloti* (Ronson et al., 1987) and *Azotobacter vinelandii* (Toukodian & Kennedy, 1986; Santero et al., 1988). Activation by NifA also occurs in *Rhizobium* and *Bradyrhizobium* spp. (Gussin et al., 1986) and in *Azorhizobium caulinodans* (de Bruijn et al., 1988), and the presence of a -24/-12 promoter sequence suggests a dependence on RpoN in the majority of *nif* promoters sequenced to date.

The *nifA* genes from several organisms have been cloned and sequenced. The NifA amino acid sequences inferred from the determined nucleic acid sequences share a high degree of homology in domains D and E, as defined by Drummond et al. (1986). Domain D, the central domain, appears to contain sequences homologous to ATP-binding proteins (Walker et al., 1982). Domain E, located in the C-terminal region, contains a helix-turn-helix motif characteristic of DNA-binding proteins; NifA also shares high homology with NtrC in domains D and E (Drummond et al., 1986).

A recombinant plasmid capable of complementing a *nifA* mutant of *A. brasilense* was identified from a gene library of *H. seropedicae* strain Z78 (Souza et al., 1991). In this paper we present the nucleotide sequence of a *nifA*-
like gene from H. seropedicae. This sequence shows a high level of homology to nifA genes from other organisms, especially to that of A. caulinodans. The presence of −24/−12 promoter sequences and binding-sites for NtrC and NifA in the promoter of H. seropedicae nifA suggests a requirement for these proteins and RpoN for the expression of this gene. Additional regulatory factors may include an ‘anaerobox’ sequence and a site for integration host factor proteins.

**Methods**

*Bacterial strains and plasmids.* Escherichia coli strains were usually grown in LB (Maniatis et al., 1982) or 2 × YT broth (Bankier et al., 1987) supplemented with antibiotics as required. E. coli 71-18 was kept in minimal medium supplemented with thiamin (5 mg l−1) (Maniatis et al., 1982). The phagemids pTZ18R and pTZ19R were used as sequencing vectors and the helper phage M13K07 was grown in liquid culture and titrated on a lawn of E. coli 71-18 (Mead et al., 1986).

**DNA labelling and hybridization.** DNA probes were labelled with [α-32P]dCTP by the hexadeoxynucleotide-primed method (Feinberg & Vogelstein, 1984). Restriction fragments from pEMS101 were separated by electrophoresis and transferred to a nylon membrane as described by Maniatis et al. (1982). Hybridizations were done under low stringency conditions (35%, v/v, formamide, 42 °C) and were washed twice with 2 × SSC (1 × SSC is 0.15 M-NaCl, 0.015 M-trisodium citrate, pH 7.0) at room temperature and twice with 2 × SSC at 65 °C (Maniatis et al., 1982).

**Cloning and molecular biology methods.** All enzymes were obtained from commercial sources and used according to the manufacturers’ instructions. Transformation, DNA purification from agarose gels, electrophoresis and single-strand DNA purification for enzymic sequencing were done as described by Maniatis et al. (1982) and Bankier et al. (1987).

**Results**

*A nifA-like gene is present in H. seropedicae*  
Plasmid pEMS1 contains a 30 kb DNA fragment, and was isolated from a genomic library of *H. seropedicae* strain Z78 constructed in the cosmid pVK102 (Souza et al., 1991). This recombinant plasmid was capable of complementing the nifA-like mutant (FP10) of Azospirillum brasilense. Plasmid pEMS101 was constructed by deleting the 21 kb BamHI fragment from pEMS1 and is also capable of complementing FP10 (Souza et al., 1991).

To locate and confirm the presence of the nifA-like gene on plasmid pEMS101, hybridization experiments were performed using the following two probes derived from pMC71A (Buchanan-Wollaston et al., 1981): (a) a 3 kb SalI fragment containing the complete nifA and
part of nifB and nifL from K. pneumoniae (probe A); and (b) a 1.4 kb NruI fragment containing 69 bp of nifL and most of the nifA from K. pneumoniae (probe B). The restriction pattern of pEMS101 cleaved with SalI, PstI, BglII and PstI/BglII is shown in Fig. 2(a). Hybridization of these restriction fragments with probe A and probe B are shown in Fig. 2(b) and 2(c), respectively. The strong hybridization signals observed with both probes, although under low stringency conditions, indicated the presence of a nifA-like gene in H. seropedicae. The nifA-specific probe (probe B) did not hybridize to the 1.7 kb PstI, 2.3 kb BglII or 1.6 kb PstI/BglII fragments. This negative result suggested the additional presence of nifB or nifL on pEMS101, adjacent to the putative nifA gene in H. seropedicae. Hybridization of probes A and B to a single 2.0 kb SalI fragment (Fig. 2b, c) positioned the majority of the nifA-like gene of H. seropedicae in this fragment. The absence of hybridization of the 1.7 kb PstI fragment to probe B (Fig. 2c) suggested that the nifA-like gene is restricted to the 1.1 kb SalI/PstI fragment (compare Fig. 1 with Fig. 2).

Nucleotide sequence of the nifA-like gene from H. seropedicae

The DNA region most likely to contain the entire nifA-like gene from H. seropedicae, as determined by hybridization comprises a 2.0 kb SalI fragment, part of a 1.7 kb EcoRI fragment and an interconnecting region (Fig. 1). These fragments were sequenced in both directions and the final sequence was found to be 4037 bp long, 3287 bp of which are shown in Fig. 3. Analysis of this sequence by a base positional method (Staden, 1984) indicated the presence of two potential coding regions: a complete open reading frame (ORF) and the N-terminus of a 3' ORF. Hybridization and genetic complementation studies (Souza et al., 1991) suggest that the complete ORF corresponds to the H.
Fig. 3 (continued on the following two pages). Nucleotide sequence of the \( nfA \)-like region of \( H. \) seropedicae. The putative \( nf \) regulatory sequences and the potential ribosome binding sites are underlined. The sequence encoding NifA starts at position 776 and stops at position 2404 while the sequence encoding NifB starts at 2797. The stop codon of the \( nfA \) gene is indicated by an asterisk.
seropedicae nifA-like gene. This ORF comprises 1629 nucleotides, contains only one in-frame methionine initiation codon at position 776, terminates at position 2404 and codes for a protein of 542 amino acid residues (Mr, 60 880). A putative Shine-Dalgarno sequence was found 4 bp from the translation initiation codon (AUG) (Fig. 3).

The H. seropedicae NifA-like protein deduced from the nucleotide sequence was compared to the inferred amino acid sequences of NifA proteins from K. pneumoniae (Buikema et al., 1985; Drummond et al., 1986), R. meliloti (Buikema et al., 1985), R. leguminosarum (Gronger et al., 1987) and Azorhizobium caulinodans (Nees et al., 1988) (Fig. 4). The homology in the N-terminal region of all proteins, domains A and C as defined by Drummond et al. (1986), was very poor: there were only seven identical amino acid residues and several gaps had to be introduced to reveal similarity. However, in the central domain (domain D) and in the C-terminal domain (domain E) extensive similarity was found. Overall, the NifA protein of H. seropedicae showed the highest degree of homology with that of Azorhizobium caulinodans (47%) and the lowest with that of K. pneumoniae (38%).

Finally, the H. seropedicae amino acid sequence also contains a conserved interdomain, between domain D and domain E (Fig. 4), similar to that present in NifA proteins of Rhizobium species (Buikema et al., 1985; Gronger et al., 1987), Bradyrhizobium japonicum (Fischer et al., 1988), Azorhizobium caulinodans (Nees et al., 1988) and Rhodobacter capsulatus (Masepohl et al., 1988), but which is absent from K. pneumoniae or Azotobacter vinelandii (Bennett et al., 1988). Although these linking regions are not highly homologous, two cysteine residues, separated by four residues, are conserved and there are also several conservative substitutions.

Regulatory sequences

The promoter region of the nifA-like gene of H. seropedicae is complex. Three potential -24/-12 promoters occur at about 110 bp 5' from the translation initiation site: GGGCATGAAATTTGCT, at position 615; CGGCCGCTTTTTTCTGCC, at position 662; and CGGTAAATACGGCAG, at position 651. Of these three sequences, the first is the best match to the consensus sequence and the third is the poorest match and least likely to constitute a functional -24/-12 promoter (Fig. 3 and Fig. 5b). Furthermore, a putative NifA-binding site (TGTCGTAATCCGACA, at
Fig. 4 (continued on the following page). Alignment of the deduced amino acid sequence from the *H. seropedicae* NifA-like protein relative to other homologous NifA proteins. HsNifA, *H. seropedicae* NifA; AcNifA, *Azorhizobium caulinodans* NifA; RINifA, *R. leguminosarum* NifA; RmNifA, *R. meliloti* NifA; and KpNifA, *K. pneumoniae* NifA. The consensus sequence shown required the simultaneous matching of amino acids in all five proteins. Conservative substitutions (Higgins & Sharp, 1988) are indicated by asterisks, and mismatches are indicated by dots. Domains A, C, D and E are boxed as originally proposed by Drummond et al. (1986), although the homology in the N-terminus shared among the five proteins is not sufficiently high to define domains A and C. The interdomain linker (I) is denoted by a horizontal bar and the conserved cysteine residues by the symbol +.
position 504) about 110 bp 5' from the -24/-12 motif at 615, suggests this region as a likely promoter (Fig. 3 and Fig. 5b). Another -24/-12 promoter sequence (GWCTACCGGCTTST, at position 356) and a NtrC-binding site (GCACGCACGTGGTGCAC, at 257) are located further 5' to the initiation site (Fig. 3 and Fig. 5a).

The 'anaerobox' and the integration host factor (IHF) site

There are two additional features which may have a role in the control of the expression of *H. seropedicae* *nif* *A*. First, a potential 'anaerobox' sequence was identified at position 1538 (TTCATCAAGGTCAA), 3' from the translation initiation site. This structure has two mismatches (underlined) with respect to the 'anaerobox' consensus sequence (TTGATNNNNATCAA) (Nees et al., 1988). A putative repressive function has been ascribed to an 'anaerobox' in a coding region (Nees et al., 1988).

Secondly, starting 49 bp 5' from the putative NifA-binding site in the *H. seropedicae* *nif* *A*-like promoter (position 455) is the sequence AATCAAATGGGATA which differs from the IHF consensus sequence (WATCAANNNNTTR, where W = A, T, and R = A or G) in only one conserved thymine residue (underlined) (Yang & Nash, 1989; Goodrich et al., 1990). This factor has the ability to introduce bends in DNA (Yang & Nash, 1989), presumably to assist contact between upstream activators and the sigma factor and is normally located between the 5' activator binding site and the promoter.

The *nifB*-like sequence

The N-terminal of another ORF was found 3' from the *nifA* gene. The protein deduced from this sequenced region spans the first 162 amino acid residues, starting at position 2797 (Fig. 3). Hybridization studies had previously indicated the presence of sequences homologous to either *nifB* or *nifL* in this region of pEMS101 (Fig. 2). Comparison of the partial amino acid sequence, deduced from this ORF, with the NifB proteins from *Azotobacter vinelandii* (Joerger & Bishop,
(a) NtrC-activated promoters

<table>
<thead>
<tr>
<th>Promoter</th>
<th>UAS</th>
<th>-24</th>
<th>-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>KpflnA</td>
<td>TGCACTATATTGTCAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpnifL</td>
<td>TGCACTATATTGTCAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HsNifA</td>
<td>AGCAGGATCTGTCAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td>GCAGYN5TGGGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Putative NifA upstream activating sequence (UAS) in nifA promoters

<table>
<thead>
<tr>
<th>Promoter</th>
<th>UAS</th>
<th>-24</th>
<th>-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcnifA</td>
<td>GGTTCATTAAAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KpnflA</td>
<td>GGGGGCAAGGTTTACAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HsNifA</td>
<td>TGTCGTATATCACGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td>TGT-N10ACA-NTGGYAYR-N4-TTGGTACAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) Putative NifA upstream activating sequence (UAS) in nifB promoters

<table>
<thead>
<tr>
<th>Promoter</th>
<th>UAS</th>
<th>-24</th>
<th>-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>KpnfB</td>
<td>TGTCGATCCGCAAGGTTTACAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RmnfB</td>
<td>TGTCGATCCGCAAGGTTTACAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AvNifB</td>
<td>TGTCGATCCGCAAGGTTTACAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HsNifB</td>
<td>TGTCGATCCGCAAGGTTTACAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td>TGT-N10ACA-NTGGYAYR-N4-TTGGTACAGACAGGATTTCGCT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5. Comparison of H. seropedicae nifA and nifB promoters with NtrC- and NifA-activating promoters of K. pneumoniae (Kp), R. meliloti (Rm), Azotobacter vinelandii (Av) and Azorhizobium caulinodans (Ac).

Consensus sequences are according to Dixon (1988).
+ The NifA binding site in the nifL promoter of K. pneumoniae is located downstream from the transcription initiation site (Drummond et al., 1983).

Fig. 6. Alignment of the deduced N-terminus of the NifB protein according to amino acid homology. HsNifB, H. seropedicae NifB; AvNifB, Azotobacter vinelandii NifB; and BjNifB, B. japonicum NifB. The consensus sequence shown required the simultaneous matching of amino acids in all three proteins. Conservative substitutions (Higgins & Sharp, 1988) are indicated by asterisks, and mismatches are indicated by dots. The conserved cysteine cluster is boxed.

1988) and Bradyrhizobium japonicum (Noti et al., 1986) showed extensive homology (Fig. 6). A cluster of cysteine residues (cys-X2-cys-X2-cys-X2-cys), characteristic of NifB proteins (Buikema et al., 1987) was found between residues 80 and 94, and most (approximately 90%) of the 20 amino acid residues spanning each side of this cluster were found to be conserved in the H. seropedicae NifB-like protein (Fig. 6). A possible -24/-12 pro-
motorm and two potential NifA-binding sites (TGTCGTGTCATGACA, at position 2620 and TGTCGCTATCCTCACA, at position 2661) were identified 5' from the initiation codon of the H. seropedicae nifB-like gene (Fig. 3 and Fig. 5c). A possible IHF-binding motif occurs at position 2708–2720 (Fig. 3).

Discussion

Our results show that H. seropedicae contains nifA- and nifB-like genes which code for proteins similar in structure to NifA and NifB proteins from other species. As expected, the NifA homology was confined to the C-terminal half of the proteins. The degree of homology of domains D (central domain) and E (DNA-binding site) of H. seropedicae was higher with the respective domains of Azorhizobium caulinodans and rhizobia, than with those of H. japonicum and K. pneumoniae. Our results also indicate that the nifA-like gene of H. seropedicae may be regulated by NtrC, NifA and RpoN, as is nifA in K. pneumoniae (Merrick, 1983; Drummond et al., 1983).

A conserved interdomain is present in the NifA proteins from R. meliloti, B. japonicum, R. leguminosarum, Azorhizobium caulinodans, Rhod. capsulatus and H. seropedicae, but is absent from those of K. pneumoniae and Azotobacter vinelandii. Two highly conserved cysteine residues are found in this region which, together with two other cysteine residues located at the end of the central domain (Fig. 4), have been proposed to be part of a metal-binding site which could act as a sensor for oxygen. Under low oxygen levels the bound metal would be reduced, thus stabilizing NifA in an active form, whereas at high oxygen concentrations the bound metal would be oxidized, destabilizing NifA and preventing activation of nif transcription (Fischer et al., 1988). In B. japonicum the four cysteine residues are essential for NifA function and metal chelators prevent B. japonicum NifA from activating nifD transcription (Fisher et al., 1988). Metal chelators have no effect on transcription activation by NifA of K. pneumoniae. Recently, it has been demonstrated that the NifA from B. japonicum is indeed sensitive to oxygen whereas K. pneumoniae NifA, which lacks this interdomain, is not (Fischer & Hennecke, 1987). The presence of this interdomain in the NifA-like protein of H. seropedicae suggests that its activity may be controlled by the prevailing oxygen levels.

The arrangement of a -24/-12 promoter element and a NifA-binding site about 100 bp 5' in the nifA-like gene of H. seropedicae is typical of nif promoters which are activated by NifA protein and require RpoN as σ-factor (Gussin et al., 1986; Dixon, 1988). Most promoters of nifA genes do not contain a prominent NifA upstream activating sequence. However, a potential NifA-binding site was found 5' from the nifA of Azorhizobium caulinodans (Nees et al., 1988), where it may act as a negative autoregulator under non-symbiotic N2-fixing conditions (de Bruijn et al., 1988). Furthermore, a putative NifA-binding site was also found in the intercistronic region of the fixRnifA operon from B. japonicum, which is autoactivated by the NifA protein (Thöny et al., 1989). The presence of a -24/-12 promoter element and a correct spacing in respect to the upstream activating sequence, favour the idea that NifA may autogenously activate transcription of the nifA gene in H. seropedicae.

An NtrC-binding site and a -24/-12 promoter sequence are located 505 bp and 407 bp 5' from the putative initiation codon of the H. seropedicae nifA gene, suggesting a possible role for NtrC in nifA transcription.

The IHF-binding site has been reported in other nif promoters: it occurs in the nifH promoter region between the NifA-binding site and the promoter (Santer et al., 1989) and the IHF protein enhances NifA-mediated expression of nifH in vitro (Hoover et al., 1990). The nifU promoter, on the other hand, contains multiple NifA-binding sites, one of which abuts the IHF-binding site at which NifA and the IHF compete for binding, thus inhibiting nifU transcription (Cannon et al., 1990). The proximity of the NifA-binding site to the IHF-binding site could lead to competition between NifA and IHF, affecting nifA expression in H. seropedicae.

The partial N-terminal amino acid sequence of the ORF 3' from nifA in H. seropedicae showed extensive homology with those of the NifB proteins from Azotobacter vinelandii and B. japonicum. The homology of the N-terminal of the putative NifB of H. seropedicae is closer to that of B. japonicum (68%) or Azotobacter vinelandii (66%) than to that of K. pneumoniae (53%) (not shown). The cysteine cluster, postulated to bind metal during the processing of FeMo-co (Buikema et al., 1987), is conserved, as well as most of the amino acid residues surrounding it. A -24/-12 promoter sequence and two NifA-binding sites were identified in the nifB promoter region (Fig. 5c), suggesting that transcription of nif promoters in H. seropedicae could be regulated by the NifA protein and dependent on the σ54 factor (RpoN). The presence of a doublet of NifA-binding sites in the promoter region of nifB-like gene has been described before in R. meliloti (Buikema et al., 1987) and Azotobacter vinelandii (Joerg & Bishop, 1988) and may be involved in transcriptional control.

This is the first nifA-like gene of a member of the Spirillaceae to be sequenced, as well as the first evidence supporting the existence of the nifA in the Spirillaceae
since the work of Pedrosa & Yates (1984). Our results indicate that expression of the nifA gene in H. seropedicae may be RpoN- and NtrC-dependent, and thus analogous to that of K. pneumoniae, but that it may also be regulated by the NiFA protein. Whether this regulation, if it occurs, is negative or positive remains to be determined. We have found no evidence for a nifL type gene in H. seropedicae and it is possible that oxygen (and nitrogen) regulation of nif expression in this organism is mediated by the interdomain region of the NiFA protein.

We are especially thankful to Christina Kennedy, Ray Dixon, Martin Drummond, Martin Buck and Paul Woodley for helpful suggestions and to Richard Pau and Mike Merrick for help with the computer analysis. We also thank Mike Merrick and Barry Smith for reading the manuscript and Mrs Rosemary Foote for typing. The DNA sequencing reported here was undertaken and completed by the senior author (E. M. S.) while on a visiting fellowship at the IPSR Nitrogen Fixation Laboratory.


