

GENE 09018

Sequences of *nifX*, *nifW*, *nifZ*, *nifB* and two ORF in the *Frankia* nitrogen fixation gene cluster

(Actinomycete; actinorhizal; root nodule; nitrogenase; *nif*)

Olivia T. Harriott *, Thomas J. Hosted * and David R. Benson

Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT06268-3044, USA

Received by C.R. Hutchinson: 24 January 1995; Revised/Accepted: 16 March/28 March 1995; Received at publishers: 25 April 1995

SUMMARY

The actinomycete *Frankia alni* fixes N₂ in root nodules of several non-leguminous plants. It is one of the few known N₂-fixing members of the high-GC Gram⁺ lineage of prokaryotes. Thus, we have undertaken a study of its nitrogen fixation gene (*nif*) organization to compare with that of the more extensively characterized proteobacteria. A cosmid (pFN1) containing the *nif* region of *Fa* CpI1 was isolated from a cosmid library using the *nifHDK* genes of *Fa* CpI1 as a probe. A 4.5-kb *Bam*HI fragment that mapped downstream from the previously characterized *nifHDK* genes was cloned and sequenced. Based on nt and aa sequence similarities to *nif* from other N₂-fixing bacteria, eight ORF were identified and designated *nifX*, *orf3*, *orf1*, *nifW*, *nifZ*, *nifB*, *orf2* and *nifU*. A region that hybridized to *Rhizobium meliloti* and *Klebsiella pneumoniae nifA* did not appear to contain a *nifA*-like gene. We have revised the map of the *Fa nif* region to reflect current information.

INTRODUCTION

Strains of *Frankia* are strictly aerobic euactinomycetes that form mutualistic N₂-fixing root nodules on some nonleguminous plants (actinorhizal plants) (Benson and Silvester, 1993). At least five *nif* genes (*HDKAB*) have been detected by heterologous hybridization in *Frankia* (Simonet et al., 1988; Mullin and An, 1990), and *nifHD* and *K*, that encode the three polypeptides of the nitro-

genase complex, have been sequenced (Normand and Bousquet, 1989; Normand et al., 1992; Hirsch et al., 1995). *NifB* from *Kp* synthesizes a low-molecular-mass Fe-S cluster precursor of the FeMo cofactor of nitrogenase (Shah et al., 1994). *NifA* activates σ^{54} -mediated transcription of *nif* and some *fix* operons in response to low pO₂ and/or nitrogen starvation in diazotrophs from the α and γ groups of the purple bacteria (Dean and Jacobson, 1992).

This report describes the characterization of a 4.5-kb DNA fragment containing eight *Frankia nif* genes. Although heterologous *nifA* probes hybridized to some cloned fragments of this *nif* region, no nt or aa similarity to *nifA* is present. This region is linked to the structural genes for the nitrogenase Fe protein and MoFe protein.

Correspondence to: Dr. D.R. Benson, Department of Molecular and Cell Biology, Box U-44, University of Connecticut, Storrs, CT 06279-3044, USA. Tel. (1-203) 486-4258; Fax (1-203) 486-1784; e-mail: dbenson@uconnvm.uconn.edu

* Present addresses: (O.T.H.) Center for Agricultural Molecular Biology, Rutgers University, Cook College, New Brunswick, NJ 08903, USA. Tel. (1-908) 932-1586; (T.J.H.) Eli Lilly Co., Lilly Corporate Center, Indianapolis, IN 46285, USA. Tel. (1-317) 276-5318.

Abbreviations: aa, amino acid(s); *Ac*, *Azorhizobium caulinodans*; *Av*, *Azotobacter vinelandii*; bp, base pair(s); *Fa*, *Frankia alni*; kb, kilobase(s) or 1000 bp; *Kp*, *Klebsiella pneumoniae*; *nif*, genes encoding proteins necessary for nitrogen fixation; nt, nucleotide(s); ORF (Orf), open reading frame(s); *Rc*, *Rhodobacter capsulatus*; *Rm*, *Rhizobium meliloti*.

EXPERIMENTAL AND DISCUSSION

(a) Isolation of the *nif* region of *Fa* CpI1

Plasmids and strains used in this study are listed in Table I. The plasmid pFN1 was isolated from a pLAFR3

TABLE I

Bacterial strains and plasmids

Strain or plasmid ^a	Genotype or properties	Reference
Strains:		
<i>Fa</i> CpI1	<i>Comptonia peregrina</i> isolate, wild type	(Callaham et al., 1978)
<i>E. coli</i> XL1-Blue	<i>supE44 hsdR17 recA1 endA1 gyrA 46 thi relA1 lac</i> [F' <i>proAB</i> ⁺ <i>lacI</i> <i>lacZ</i> ΔM15 Tn10(Tc ^R)]	Stratagene, La Jolla, CA, USA
Plasmids:		
pBluescriptII	Am ^R Tc ^R , pUC18-derived cloning vector	Stratagene
pFR140	Am ^R (pUC19), 7.1-kb <i>Bam</i> HI insert with <i>nifHDK</i>	(Rocheffort and Benson, 1990)
pFN1	Tc ^R [pLAFR3], 28-kb <i>Bam</i> HI insert	(Hosted, 1992)
pFN2.1	pBluescriptSK ⁻ , 2.1-kb <i>Sst</i> I insert	This study
pFNB4.5	pBluescriptSK ⁻ , 4.5-kb <i>Bam</i> HI insert	This study
pJE294	<i>Kp nifA</i>	(Santero et al., 1989)
pCE300	<i>Rm nifA</i>	C. Earl
pFQ149	<i>Fa</i> ArI3 4.5-kb <i>nif</i> clone containing <i>nifB</i>	(Simonet et al., 1988)

^a *Escherichia coli* strains were grown at 37°C in Terrific Broth or 2 × YT medium. When required ampicillin (100 µg/ml) or tetracycline (10 µg/ml) was added. *Fa* CpI1 was grown in defined succinate medium (Norridge and Benson, 1986).

cosmid library of *Fa* CpI1 genomic DNA by probing with CpI1 *nifHDK* genes in pFR140. A 4.5 kb *Bam*HI fragment was subcloned from pFN1 and designated pFNB4.5 (Fig. 1). In its restriction map, and hybridization pattern to heterologous *nifA* probes, the insert in pFNB4.5 is identical to that in pFQ149, a clone containing the homologous region from the closely related *Fa* ArI3 (Simonet et al., 1988). pFNB4.5 hybridized strongly to appropriately sized bands of restriction digested *Fa* CpI1 genomic DNA confirming that pFNB4.5 originated from *Fa* CpI1 (not shown).

(b) Sequence analysis of pFNB4.5

The strategy used to sequence pFNB4.5 and the resulting sequence are recorded in Figs. 1 and 2. High-%G + C Gram⁺ bacteria like *Frankia* generally have a G or C in the third codon position. Thus, ORF were initially circumscribed on the basis of third position bias. ORF were identified by the similarity of their inferred aa sequences to aa sequences of *nif* genes from other N₂-fixing bacteria. Altogether, eight *nif* genes were identified including *nifX*,

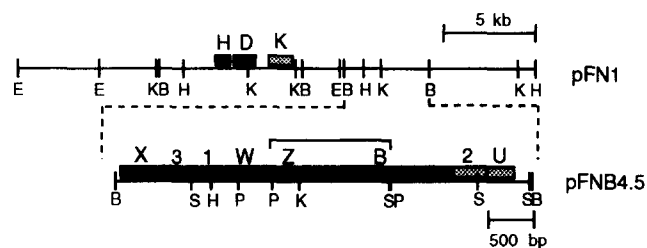


Fig. 1. Physical and genetic map of pFNB4.5 and its position relative to *nifHDK*. B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sst*I. The area under the bracket spanning *nifZ* and *nifB* represents the region that hybridized to *Rm* and *Kp nifA*. Stippled boxes represent known genes whose sequences have not been reported.

orf1, *orf3*, *nifW*, *nifZ*, *nifB*, *orf2* and *nifU*. All are transcribed in the same direction and, with the exception of *nifB*, are preceded by GGAG or GAGG potential Shine-Dalgarno regions. The most probable Shine-Dalgarno region for *nifB* is GGACG beginning 12 bp before the putative GTG start codon. The genes reported here overlap in six instances. Sufficient sequence information was obtained to identify *orf2* and *nifU* but is not reported because the complete sequence of both strands was not obtained. The characteristics of each gene product are listed in Table I.

(c) *nifX*, *orf3*, *orf1*, *nifW* and *nifZ*

The *Fa* CpI1 *nifX* gene is upstream from *orf3* as in *Av* and *Rc* (Jacobson et al., 1989; Moreno-Vivian et al., 1989). Overlapping the *nifX* GTG start codon is a TGA stop codon in frame with upstream codons having >90% GC bias in the third position, probably indicating an additional ORF. *Frankia* NifX has 30% predicted aa sequence identity to *Rc*, 27% to *Av* and 28% to *Kp nifX* genes. Accounting for conservative substitutions, *Fa* NifX has 54% similarity to NifX from all three microorganisms. The role of NifX is unclear but has been suggested to help negatively regulate nitrogenase synthesis (Gosink et al., 1990).

A 444-bp ORF that overlaps *nifX* by four nt shows 40% predicted aa sequence identity, and 62% aa similarity, to Orf3 of *Av* and Orf4 of *Rc* (Jacobson et al., 1989; Moreno-Vivian et al., 1989). Overlapping the 3' end of *orf3* by 11 nt, is a 249-bp ORF with 33% derived aa sequence identity, and 55% aa sequence similarity, to *Ac* Orf1 (Arigoni et al., 1991). The functions of Orf3 and Orf1 are not known.

Several direct repeats found at the 3' end of *nifZ* are reflected in the aa sequence as an Ala + Pro-rich region. The role of NifZ is unclear, although studies in *Av* suggest that NifZ might be required for the maturation of the MoFe protein (Jacobson et al., 1989).

(d) *nifB*

NifB is essential for nitrogen fixation and is highly conserved (Dean and Jacobson, 1992). It synthesizes an Fe-S cluster that forms part of the FeMo-cofactor that is present at the active site of the molybdenum-iron protein of nitrogenase (Shah et al., 1994). *Fa nifB* has a predicted molecular mass of 55 078 Da (see Table II), a value that is close to that of other NifB proteins. It is most similar to *Anabaena* NifB with 58% aa identity and 75% similarity, followed by *Av* NifB (53% and 69%), *Rm* NifB (50% and 66%) and *Kp* NifB (49% and 69%). The nine conserved cysteine residues found in other NifB proteins are also present in *Fa* NifB.

The start of *nifB* overlaps the 3' end of *nifZ* by 2 nt, suggesting that *nifB* is cotranscribed with *nifX*, *orf3*, *orf1*, *nifW* and *nifZ*. However, several repeated sequences, ranging in length from 11 to 33 nt, and within the 3' region of *nifZ*, may suggest the presence of a regulatory region such as occur in promoter regions of genes in *Streptomyces* sp. (Delic et al., 1992).

In the only other NifB sequence from a Gram⁺ organism, *Clostridium pasteurianum*, *nifB* is fused to the 3' end of *nifN* and is transcribed as part of the *nifEN-B* operon (Dean and Jacobson, 1992; J.-S. Chen, personal communication). In the Gram⁻ diazotrophs *Av*, *Kp*, *Anabaena* sp. strain 7120 and *Rc*, *nifB* is transcribed from its own promoter (Buikema et al., 1987; Joerger and Bishop, 1988; Klipp et al., 1988; Masepohl et al., 1988; Mulligan and Haselkorn, 1989).

Sequence of the region directly downstream from *nifB* indicates the presence of two other ORF. We have design-

nated one as *orf2*; it has 42% sequence identity and 64% similarity to *orf6* from *Av*. The other ORF was designated *nifU* due to its predicted aa similarity to the carboxyl domain of NifU from other diazotrophs. Truncated forms of *nifU* have been reported from *Rc* and *Bradyrhizobium japonicum* (Dean and Jacobson, 1992). The reading frame bias and aa alignments, including placement of conserved cysteine residues, are consistent with these designations.

(e) Hybridization to *Rm* and *Kp nifA*

Plasmid pFNB4.5 has a restriction map identical to that of pFQ149 (Simonet et al., 1988). pFQ149 contains a 4.5-kb *Bam*HI insert from *Fa* strain ArI3 that hybridized with *Rm nifA* and *nifB*. We probed CpI1 chromosomal DNA, pFNB4.5 and pFN1 with a 945 bp *Pst*I-*Eco*RV fragment of pJE294 containing an internal fragment of *Kp nifA*. *Kp nifA* did not hybridize to digested *Fa* CpI1 chromosomal DNA but did hybridize to the 4.5-kb *Bam*HI fragment of pFN1 and pFNB4.5, the 3.5-kb *Hind*III-*Bam*HI and 1.2-kb *Pst*I fragments of pFNB4.5 (Fig. 1). Identical results were obtained when *Rm nifA* from pCE300 was used as a probe (not shown). These observations confirm previously reported hybridization results (Simonet et al., 1988).

The absence of *nifA* in pFNB4.5 suggests that the hybridization obtained with cloned fragments of pFNB4.5 was artefactual. Short regions (25–50 nt) present upstream from *nifB* with 65–70% similarity to *nifA* probes might be sufficient to allow some binding to high-copy-number *Frankia nif* DNA sequences on Southern blots. Therefore, we conclude that a conserved *nifA* is not present in *Fa* CpI1, although a significantly diverged homologue cannot be ruled out.

(f) Conclusions

(1) The *nifX-orf3-orf1-nifW-nifZ-nifB-orf2-nifU* region of *Fa* CpI1 is about 3 kb away from the *nifH-nifD-nifK* region.

(2) *nifA*-hybridizable DNA was detected in the *nifZ* region but an ORF corresponding to *nifA* was not found. Therefore *Frankia* strains have not yet been shown to possess a *nifA* homologue.

TABLE II

Properties of predicted polypeptides from pFNB4.5 genes

ORF	Positions ^a (nt)	Number of aa ^b	Molecular mass ^b (Da)	<i>pI</i>
<i>nifX</i>	55–486	143	15 407	6.78
<i>orf3</i>	483–926	147	16 166	4.73
<i>orf1</i>	916–1164	82	9 081	4.74
<i>nifW</i>	1161–1559	132	14 822	5.08
<i>nifZ</i>	556–2134	192	19 002	5.14
<i>nifB</i>	2131–3663	510	55 078	5.06

^a Numbers refer to positions in the sequence shown in Fig. 2. Includes the first position in the initiation codon and the third position in the stop codon.

^b Includes the N-terminal Met.

ACKNOWLEDGEMENTS

We thank P.W. Betts for his expert technical assistance, P. Normand for pFQ149 and T. Hoover for pJES294.

REFERENCES

Arigoni, F., Kaminski, P.A., Hennecke, H. and Elmerich, C.: Nucleotide sequence of the *fixABC* region of *Azorhizobium caulinodans* ORS571:

- similarity of the *fixB* product with eukaryotic flavoproteins, characterization of *fixX*, and identification of *nifW*. *Mol. Gen. Genet.* 225 (1991) 514–520.
- Benson, D.R. and Silvester, W.B.: Biology of *Frankia* strains, actinomycete symbionts of actinorhizal plants. *Microbiol. Rev.* 57 (1993) 293–319.
- Buikema, W.J., Klingensmith, J.A., Gibbons, S.L. and Ausubel, F.M.: Conservation of structure and location of *Rhizobium meliloti* and *Klebsiella pneumoniae nifB* genes. *J. Bacteriol.* 169 (1987) 1120–1126.
- Callahan, D., DelTredici, P. and Torrey, J.G.: Isolation and cultivation in vitro of the actinomycete causing root nodulation in *Comptonia*. *Science* 199 (1978) 899–902.
- Dean, D.R. and Jacobson, M.R.: Biochemical genetics of nitrogenase. In: Stacey, G., Burris, R.H. and Evans, H.J. (Eds.), *Biological Nitrogen Fixation*. Chapman and Hall, New York, NY, 1992, pp. 763–834.
- Delic, I., Robbins, P. and Westpheling, J.: Direct repeat sequences are implicated in the regulation of two *Streptomyces* chitinase promoters that are subject to carbon catabolite control. *Proc. Natl. Acad. Sci. USA* 89 (1992) 1885–1889.
- Gosink, M.M., Franklin, N.M. and Roberts, G.P.: The product of the *Klebsiella pneumoniae nifX* gene is a negative regulator of the nitrogen fixation (*nif*) regulon. *J. Bacteriol.* 172 (1990) 1441–1447.
- Hirsch, A.M., McKhann, H.I., Reddy, A., Liao, J., Fang, Y. and Marshall, C.R.: Assessing horizontal transfer of *nifHDK* genes in eubacteria: nucleotide sequence of *nifK* from *Frankia* strain HFPCcI3. *Mol. Biol. Evol.* 12 (1995) 16–27.
- Hoover, T.R., Imperial, J., Ludden, P.W. and Shah, V.K.: Homocitrate is a component of the iron-molybdenum cofactor of nitrogenase. *Biochemistry* 28 (1989) 2768–2771.
- Hosted, T.J.: Studies on *Frankia* Symbiotic Genes. Ph. D. Thesis, University of Connecticut, Storrs, CT, 1992.
- Jacobson, M.R., Cash, V.L., Weiss, M.C., Laird, N.F., Newton, W.E. and Dean, D.R.: Biochemical and genetic analysis of the *nifUSVWZM* cluster from *Azotobacter vinelandii*. *Mol. Gen. Genet.* 219 (1989) 49–57.
- Joerger, R.D. and Bishop, P.E.: Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. *J. Bacteriol.* 170 (1988) 1475–1487.
- Klipp, W., Masepohl, B. and Puhler, A.: Identification and mapping of nitrogen fixation genes of *Rhodobacter capsulatus*: duplication of a *nifA-nifB* region. *J. Bacteriol.* 170 (1988) 693–699.
- Lee, S. and Rasheed, S.: A simple procedure for maximum yield of high-quality plasmid DNA. *Biotechniques* 9 (1990) 676–679.
- Masepohl, B., Angermuller, S., Hennecke, S., Hubner, P., Moreno-Vivian, C. and Klipp, W.: Nucleotide sequence and genetic analysis of the *Rhodobacter capsulatus orf6-nifU1SVW* gene region: possible role of NifW in homocitrate processing. *Mol. Gen. Genet.* 238 (1993) 369–382.
- Masepohl, B., Klipp, W. and Puhler, A.: Genetic characterization and sequence analysis of the duplicated *nifA/nifB* gene region of *Rhodobacter capsulatus*. *Mol. Gen. Genet.* 212 (1988) 27–37.
- Meijer, W.G. and Tabita, R.: Isolation and characterization of the *nifUSVW-rpoN* gene cluster from *Rhodobacter sphaeroides*. *J. Bacteriol.* 174 (1992) 3855–3866.
- Moreno-Vivian, C., Schmehl, M., Masepohl, B., Arnold, W. and Klipp, W.: DNA sequence and genetic analysis of the *Rhodobacter capsulatus nifENX* gene region: homology between NifX and NifB suggests involvement of NifX in processing of the iron-molybdenum cofactor. *Mol. Gen. Genet.* 216 (1989) 353–363.
- Mulligan, M.E. and Haselkorn, R.: Nitrogen fixation (*nif*) genes of the cyanobacterium *Anabaena* species strain PCC 7120. *J. Biol. Chem.* 264 (1989) 19200–19207.
- Mullin, B.C. and An, C.S.: The molecular genetics of *Frankia*. In: Schwintzer, C.R. and Tjepkema, J.D. (Eds.), *The Biology of Frankia and Actinorhizal Plants*. Academic Press, San Diego, CA, 1990, pp. 195–214.
- Noridge, N.A. and Benson, D.R.: Isolation and nitrogen-fixing activity of *Frankia* sp. strain CpI1 vesicles. *J. Bacteriol.* 166 (1986) 301–305.
- Normand, P. and Bousquet, J.: Phylogeny of nitrogenase sequences in *Frankia* and other nitrogen-fixing microorganisms. *J. Mol. Evol.* 29 (1989) 436–447.
- Normand, P., Gouy, M., Cournoyer, B. and Simonet, P.: Nucleotide sequence of *nifD* from *Frankia alni* strain ArI3: phylogenetic inferences. *Mol. Biol. Evol.* 9 (1992) 495–506.
- Paul, W. and Merrick, M.: The nucleotide sequence of the *nifM* gene of *Klebsiella pneumoniae* and identification of a new *nif* gene: *nifZ*. *Eur. J. Biochem.* 170 (1987) 259–265.
- Rocheffort, D.A. and Benson, D.R.: Molecular cloning, sequencing, and expression of the glutamine synthetase II (*glnII*) gene from the actinomycete root nodule symbiont *Frankia* sp. strain CpI1. *J. Bacteriol.* 172 (1990) 5335–5342.
- Sambrook, J., Fritsch, E.F. and Maniatis, T.: *Molecular Cloning. A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- Santero, E., Hoover, T., Keener, J. and Kustu, S.: In vitro activity of the nitrogen fixation regulatory protein NifA. *Proc. Natl. Acad. Sci. USA* 86 (1989) 7346–7350.
- Shah, V.K., Allen, J.R., Spangler, N.J. and Ludden, P.W.: In vitro synthesis of the iron-molybdenum cofactor of nitrogenase. *J. Biol. Chem.* 269 (1994) 1154–1158.
- Simonet, P., Normand, P. and Bardin, R.: Heterologous hybridization of *Frankia* DNA to *Rhizobium meliloti* and *Klebsiella pneumoniae nif* genes. *FEMS Microbiol. Lett.* 55 (1988) 141–146.
- Wang, S., Dean, D.R., Chen, J. and Johnson, J.L.: The N-terminal and C-terminal portions of NifV are encoded by two different genes in *Clostridium pasteurianum*. *J. Bacteriol.* 173 (1991) 3041–3046.