Mol Gen Genet (1989) 216:353–363

MGG

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

Conrado Moreno-Vivian*, Manfred Schmehl, Bernd Masepohl, Walter Arnold, and Werner Klipp

Lehrstuhl für Genetik, Fakultät für Biologie, Universität Bielefeld, Postfach 8640, D-4800 Bielefeld 1, Federal Republic of Germany

Summary. *Rhodobacter capsulatus* genes homologous to *Klebsiella pneumoniae nifE*, *nifN* and *nifX* were identified by DNA sequence analysis of a 4282 bp fragment of *nif* region A. Four open reading frames coding for a 51188 (NifE), a 49459 (NifN), a 17459 (NifX) and a 17472 (ORF4) dalton protein were detected. A typical NifA activated consensus promoter and two imperfect putative NifA binding sites were located in the 377 bp sequence in front of the *nifE* coding region. Comparison of the deduced amino acid sequences of *R. capsulatus* NifE and NifN revealed homologies not only to analogous gene products of other organisms but also to the α and β subunits of the nitrogenase iron-molybdenum protein. In addition, the *R. capsulatus nifE* and *nifN* proteins shared considerable homology with each other. The map position of *nifX* downstream of *nifEN* corresponded in *R. capsulatus* and *K. pneumoniae* and the deduced molecular weights of both proteins were nearly identical. Nevertheless, *R. capsulatus* NifX was more related to the C-terminal end of NifY from *K. pneumoniae* than to NifX. A small domain of approximately 33 amino acid residues showing the highest degree of homology between NifY and NifX was also present in all *nifB* proteins analyzed so far. This homology indicated an evolutionary relationship of *nifX*, *nifY* and *nifB* and also suggested that NifX and NifY might play a role in maturation and/or stability of the iron-molybdenum cofactor. The open reading frame (ORF4) downstream of *nifX* in *R. capsulatus* is also present in *Azotobacter vinelandii* but not in *K. pneumoniae*. Interposon-induced insertion and deletion mutants proved that *nifE* and *nifN* were necessary for nitrogen fixation in *R. capsulatus*. In contrast, no essential role could be demonstrated for *nifX* and ORF4 whereas at least one gene downstream of ORF4 appeared to be important for nitrogen fixation.

Key words: *Rhodobacter capsulatus* – Nitrogen fixation – DNA sequence analysis – *nifE*, *nifN*, *nifX* genes – Protein comparisons

Introduction

The purple non-sulfur photosynthetic bacterium *Rhodobacter capsulatus*, formerly called *Rhodopseudomonas capsulata* (Imhoff et al. 1984), is able to fix atmospheric dinitrogen. Several genes involved in the nitrogen fixation process were previously identified and cloned (Avtges et al. 1983, 1985; Kranz and Haselkorn 1985; Ahombo et al. 1986; Klipp et al. 1988; for a review see Haselkorn 1986). In contrast to *Klebsiella pneumoniae*, the *R. capsulatus nif* genes are not clustered (Willison et al. 1985). Physical mapping of Tn*5* induced Nif$^-$ mutants revealed three not closely linked *nif* gene clusters in *R. capsulatus* and these were designated *nif* regions A, B and C (Klipp et al. 1988). Within *nif* region A, DNA fragments hybridizing to *K. pneumoniae nifE*, *nifS*, *nifB* and *nifA* were localized (Klipp et al. 1988). DNA sequencing also established the existence of *nifA* and *nifB* analogous genes in this region (Masepohl et al. 1988). A reiteration of *nifA* and *nifB* (*nifA*/*nifB* copy II; Masepohl et al. 1988) is located in *nif* region B downstream of the *nifHDK* operon, which encodes the subunits of the nitrogenase (Avtges et al. 1983; Schumann et al. 1986), and the regulatory gene *nifR4* (Kranz and Haselkorn 1985; Ahombo et al. 1986). Three additional genes (*nifR1–R3*), which are involved in transcriptional control of other *nif* genes, were identified by Kranz and Haselkorn (1985) and mapped in *nif* region C. The existence of typical NifA activated promoters in front of both copies of *nifB* (Masepohl et al. 1988) and in front of *nifH* (Pollock et al. 1988) suggests that expression of non-regulatory *R. capsulatus nif* genes is controlled by the same mechanisms as in other nitrogen fixing organisms.

Biological nitrogen fixation is catalyzed by the nitrogenase enzyme complex, which is composed of the Fe-protein and the MoFe-protein. The Fe-protein is a dimer of identical subunits (NifH), which contains a single [4Fe-4S] cluster. The MoFe-protein is a tetramer of two NifD and two NifK subunits, which includes 2 iron-molybdenum cofactor (FeMoco) centers and about 16 additional iron and acid-labile sulfur ions (for review see Orme-Johnson 1985). In *K. pneumoniae* the *nifH*, *nifD* and *nifK* genes are part of an operon which also includes *nifT* and *nifY*, two genes of unknown function (Arnold et al. 1988).

Six *K. pneumoniae* genes are known to be involved in FeMoco biosynthesis: *nifH*, *nifQ*, *nifB*, *nifV*, *nifE* and *nifN*. A clear function has been assigned only to *nifV*, which encoded a homocitrate synthase (Hoover et al. 1987). Imperial et al. (1984) proposed that *nifQ* is involved in the processing of molybdenum. A central role for *nifB* in FeMoco biosynthesis is supported by the existence of analogous genes in a variety of different organisms including *Rhizo-*

* *Present address:* Departamento de Bioquimica, Facultad de Ciencias, Universidad de Cordoba, E-14071 Cordoba, Spain

Offprint requests to: W. Klipp

Table 1. Bacterial strains and plasmids

Designation	Characteristics[a]	Source or reference
Bacterial strains		
Rhodobacter capsulatus B10S	Spontaneous Sm^r mutant of *R. capsulatus* B10	Klipp et al. (1988)
Escherichia coli JM83	Host for pUC8 and pSVB plasmids	Vieira and Messing (1982)
E. coli S17-1	RP4-2 (Tc::Mu) (Km::Tn*7*) integrated in the chromosome	Simon et al. (1983)
E. coli S605	*E. coli* C600 Met::Tn*5*	Klipp and Pühler (1984)
Plasmids		
pUC8	Ap^r, Lac$^+$	Vieira and Messing (1982)
pSVB20	Ap^r, Lac$^+$	Arnold and Pühler (1988)
pSVB23	Ap^r	Arnold and Pühler (1988)
pSVB25	Ap^r	Arnold and Pühler (1988)
pSUP202	Ap^r, Tc^r, Cm^r, *mob*	Simon et al. (1983)
pSUP301	Ap^r, Km^r, *mob*	Simon et al. (1983)
pSUP2021	pSUP202::Tn*5*	Simon et al. (1983)
pSUP10141	pSUP101::Tn*5*-Tc	Simon et al. (1983)
pWKR31.2	6.8 kb *Hin*dIII fragment of *R. capsulatus* cloned in pSUP202	Klipp et al. (1988)
pWKR189	3.6 kb *Bam*HI fragment carrying Gm^r cloned from pPH1JI[b] into pUC8	This work
pWKR300	3 kb *Eco*RI fragment carrying *nifEN* cloned into pSUP202	This work
pWKR300.1I[c]	pWKR300 *nifE*::[Km][d]	This work
pWKR300.2I/II[c]	pWKR300 *nifE*::[Km][d]	This work
pWKR300.3I[c]	pWKR300 *nifN*::[Km][d]	This work
pWKR300.4I/II[c]	pWKR300 *nifN*::[Km][d]	This work
pWKR42	0.9 kb *Eco*RI fragment carrying *nifX* cloned into pSUP202	This work
pWKR298I/II[c]	pWKR42 *nifX*::[Km][e]	This work
pCMV171	2.3 kb *Hin*dIII-*Xho*I fragment carrying ORF4 cloned into pSUP301	This work
pCMV180I/II[c]	pCMV171 ORF4::[Gm][f]	This work
pCMV200I/II[c]	pCMV171 ORF4::[Gm][f], (Tc^r)[g]	This work

[a] Only characteristics relevant for this study are listed. Sm, streptomycin; Ap, ampicillin; Tc, tetracycline; Cm, chloramphenicol; Km, kanamycin; Gm, gentamicin
[b] Hirsch and Beringer (1984)
[c] Roman numbers refer to the orientation of the cloned interposon
[d] Tn*5* deletion mutant unable to transpose
[e] Km^r interposon derived from pSUP2021
[f] Gm^r interposon derived from pWKR189
[g] A *Pst*I fragment from pSUP10141 carrying Tc^r was cloned into pCMV180I and pCMV180II

bium meliloti (Buikema et al. 1987), *R. leguminosarum* (Rossen et al. 1984; Buikema et al. 1987), *Bradyrhizobium japonicum* (Fuhrmann et al. 1985; Noti et al. 1986). *Azotobacter vinelandii* (Joerger and Bishop 1988) and *R. capsulatus* (Masepohl et al. 1988). DNA sequence analysis revealed a high degree of homology between the *nifEN* encoded gene products and the two subunits of the MoFe-protein from *K. pneumoniae* (Arnold et al. 1988) and *A. vinelandii* (Brigle et al. 1987). These homologies in amino acid sequence indicate that the *nifEN* proteins can form a complex structurally analogous to the MoFe-protein (Brigle et al. 1987). In *K. pneumoniae* and *A. vinelandii*, the *nifE* and *nifN* genes are organized in one transcription unit together with *nifX*, a gene of unknown function (Klipp and Pühler 1986; Brigle et al. 1987), whereas *nifE* and *nifN* are separated from each other in *R. meliloti* (Aguilar et al. 1987).

In this work we present the complete nucleotide sequence and describe the genetic characterization of a DNA fragment located in *R. capsulatus nif* region A, which includes *nifE*, *nifN* and *nifX*.

Materials and methods

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are listed in Table 1.

Media and growth conditions. R. capsulatus was cultivated at 30° C in RCV medium described by Weaver et al. (1975) or on PY plates as described previously (Klipp et al. 1988). Growth on N_2 (Nif$^+$ phenotype) was tested in ammonium-free RCV medium in microtiter plates incubated in an anaerobic jar (GasPak, BBL Microbiology Systems). *Escherichia coli* strains were grown at 37° C in LB medium (Miller 1972). The antibiotic concentrations used were as described elsewhere (Masepohl et al. 1988).

DNA biochemistry. DNA isolation, restriction enzyme analysis, agarose gel electrophoresis and cloning procedures were performed using standard methods (Maniatis et al. 1982). Restriction endonucleases, T4 DNA ligase and Klenow polymerase were obtained from Bethesda Research Laboratories and were used as recommended by the supplier.

DNA sequencing. Sequencing was performed for both DNA strands according to the chemical degradation method (Maxam and Gilbert 1980) using α-[^{32}P]dCTP or γ-[^{32}P]ATP for 3′ or 5′ labeling, as previously described (Arnold et al. 1988). Various defined restriction fragments from the recombinant plasmid pWKR31.2 were cloned into the appropriate restriction sites of the sequencing vectors

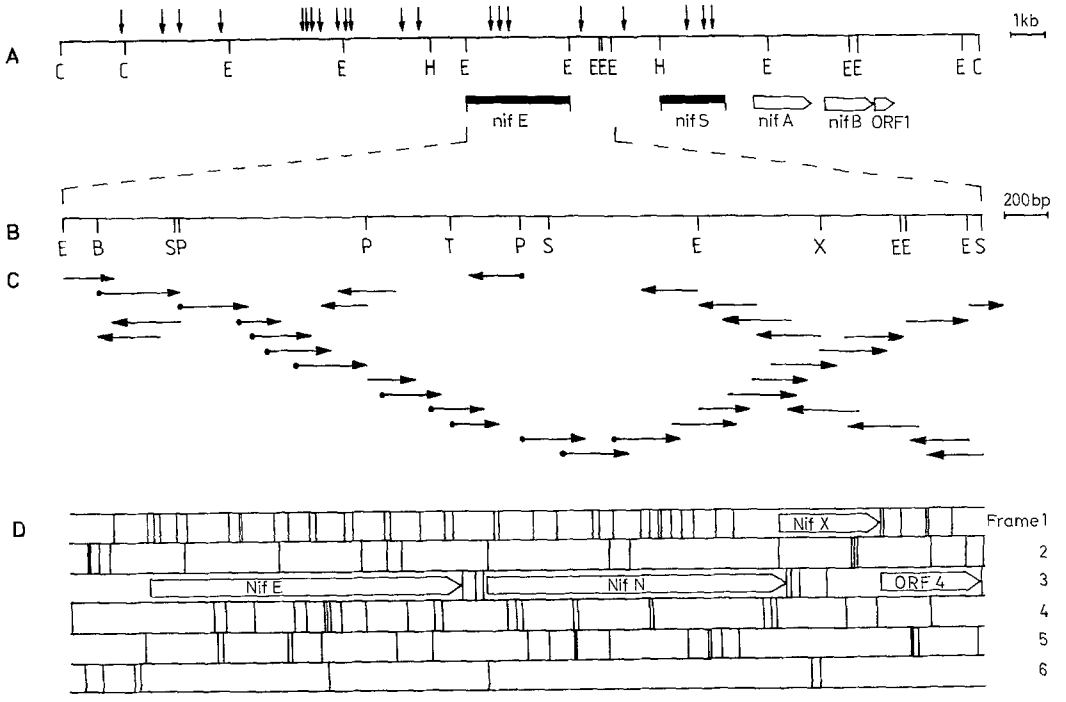

Fig. 1 A–D. Sequence analysis of a *Rhodobacter capsulatus* DNA region with homology to *Klebsiella pneumoniae nifE* and identification of open reading frames. In **A**, *vertical arrows* above the physical map of *R. capsulatus nif* region A indicate the location of Nif:: Tn*5* insertions (Klipp et al. 1988). *Open arrows* below the physical map show location, size and orientation of *nifA*, *nifB* and ORF1 (Masepohl et al. 1988). *Filled bars* indicate regions of homology to *K. pneumoniae nifE* and *nifS*. A detailed restriction map of the sequenced region is given in **B** at larger scale and the sequencing strategy is presented in **C**. *Arrows* indicate the direction and extent of each sequenced clone and *arrows marked by a dot* show DNA regions from which the sequence of both strands was obtained from the same clone. The distribution of translational termination codons in all six reading frames and the location of four open reading frames is indicated in **D**. The designation NifE, NifN and NifX of these coding regions is based on amino acid sequence homology to analogous genes from other organisms (Figs. 3A, B, 4A). Abbreviations: *Bam*HI (B), *Bst*EII (T), *Cla*I (C), *Eco*RI (E), *Hin*dIII (H), *Pst*I (P), *Sma*I (S), *Xho*I (X)

pSVB20, pSVB23 and pSVB25. Fragments for overlapping sequencing were generated by partial digestion with *Sau*3A in the presence of ethidium bromide (Parker et al. 1977) followed by digestion with *Bam*HI, resulting in plasmids with different deletions. Nested deletions were also induced by digestion with *Eco*RI followed by a fill-in reaction by Klenow polymerase and partial digestion with *Hae*III. All restriction sites, with the exception of two *Pst*I sites, were confirmed by overlapping sequencing. Sequences were compared using the computer programs previously described (Masepohl et al. 1988).

Construction of plasmids for interposon mutagenesis. The 2986 bp *Eco*RI fragment (Fig. 5) carrying *nifE* and part of *nifN* was cloned into the mobilizable vector plasmid pSUP202. The resulting hybrid plasmid pWKR300 was mutagenized by transposon Tn*5* in *E. coli* S605 selecting for the enhanced kanamycin (Km) resistance of strains carrying Tn*5* on multicopy plasmids (Klipp and Pühler 1984). Plasmids with the desired *nifE*:: Tn*5* and *nifN*:: Tn*5* insertions were subsequently digested with *Xho*I. Religation and selection for Km resistance resulted in plasmids containing a Tn*5* deletion derivative unable to transpose. The *nifX* mutation was constructed by cloning the *Xho*I fragment of Tn*5* encoding Km resistance from plasmid pSUP2021 into the pSUP202 derivative pWKR42, which contains the *R. capsulatus* 918 bp *Eco*RI fragment. To construct an ORF4 mutation, the 2.3 kb *Hin*dIII-*Xho*I fragment was first cloned into the mobilizable vector plasmid pSUP301, resulting in pCMV180I/II. Two internal *Eco*RI fragments were substituted by an *Eco*RI fragment encoding gentamicin (Gm) resistance. This fragment was isolated from plasmid pWKR189, which was generated by cloning a *Bam*HI fragment of pPH1JI (Hirsch and Beringer 1984) into the multiple cloning site of pUC8. Since the ampicillin resistance of pCMV180I/II cannot be used in *R. capsulatus*, one of the two *Pst*I restriction sites of this plasmid was used to insert a *Pst*I fragment encoding the tetracycline (Tc) resistance from plasmid pSUP10141.

Homogenotization of interposons. Plasmids containing interposons were mobilized from *E. coli* S17-1 into *R. capsulatus* B10S by filter matings (Masepohl et al. 1988). Since these plasmids are not able to replicate in *R. capsulatus*, selection for the interposon mediated Km or Gm resistance and subsequent test for loss of the vector encoded Tc resistance resulted in the desired interposon mutants. Correct homogenotization of the interposon was verified by Southern hybridization of different restriction enzyme digests of total DNA isolated from these *R. capsulatus* mutants.

Results

DNA sequence analysis

Figure 1A shows the physical map of the *R. capsulatus nif* region A as well as Tn*5* insertions resulting in a Nif$^-$ pheno-

```
GAATTCACCCGCCGCATGGGGATGCTGGTGCGCAAGGACAACCTCGGCTTCGGCCTGCGGTCCTGGCGCTATGCCGCCATCGTGACGAATGGCGTGATCG  100

AGGCCTGGTTCGAAGAGCCCGGCCTGATGGACAACTGCCCCGAGGATCCCTATGGCGTCTCCAGCCCGGAAAACGTTCTGGCCTGGCTGAAAACCGCGAA  200

AGTCGCCTGAGCGATCCGACAGAGGTTCCGAAAGCCCCGCCCGTCCGGCGTGGCTTCTGTTGCAGGCCGTCCGACAATTTGTCGGATATGCTCCAAACCC  300

                                                                    M  S  E  A  L  K  S  K      NifE
GGCGCCGTGGCTGCAAAGACCGGGTCGTTTCACCTGGCACGTTGCTTGCTTCATCCCTCCTGACAGGGGGATGCATCATGTCCGAAGCCTTGAAGTCGAA  400
                                                    •••
  I  A  D  V  L  N  E  P  G  C  A  T  N  S  T  K  T  D  V  L  R  K  R  G  C  A  E  R  L  T  P  G  A
AATCGCCGATGTCCTGAACGAGCCGGGCTGTGCCACCAACTCGACCAAGACCGATGTTCTGCGCAAGAGGGGCTGCGCCGAGCGGCTGACCCCGGGGGCC  500

 A  A  G  G  C  A  F  D  G  A  M  T  A  L  Q  P  I  V  D  V  A  H  L  V  H  A  P  A  A  C  W  S  N
GCGGCCGGTGGCTGCGCCTTTGACGGCGCGATGACCGCGCTGCAGCCGATCGTGGATGTCGCCCATCTGGTCCATGCCCCCGCCGCCTGCTGGTCGAACG  600

G  W  D  N  R  S  S  A  S  S  G  S  E  L  Y  R  K  G  F  T  T  D  L  S  E  L  D  I  V  M  G  H  G  E
GCTGGGACAACCGCTCTTCGGCCTCGTCGGGCTCGGAGCTTTACCGCAAGGGCTTCACCACCGATCTGTCGGAACTCGACATCGTCATGGGCCACGGCGA  700

  K  K  L  Y  R  A  L  R  P  V  I  E  A  E  S  P  A  A  V  F  V  Y  A  T  C  V  T  A  L  I  G  D  D
GAAGAAACTTTACCGCGCGCTCCGCCCGGTGATCGAGGCGGAAAGCCCCGCCGCCGTCTTCGTCTATGCGACCTGCGTGACGGCGCTGATCGGCGACGAT  800

 L  G  A  V  C  G  A  A  T  A  K  W  G  A  P  C  V  P  V  G  V  P  G  F  A  G  S  K  N  L  G  N  K
CTGGGTGCGGTCTGCGGGGCGGCGACGGCGAAATGGGGCGCGCCCTGCGTTCCGGTCGGTGTGCCGGGGTTCGCGGGGTCGAAGAACCTGGGCAACAAGC  900

L  G  G  E  A  L  L  D  R  V  V  G  A  L  E  P  E  T  V  T  P  C  D  V  N  I  I  G  D  Y  G  L  S  G
TGGGCGGAGAGGCGCTTCTGGATCGCGTCGTCGGCGCGCTGGAACCCGAGACGGTGACGCCCTGTGACGTCAACATCATCGGCGATTACGGCCTGTCGGG 1000

  E  L  W  Q  V  K  P  L  L  D  K  L  G  I  R  I  L  G  S  I  A  G  D  A  R  Y  K  Q  V  A  M  A  H
CGAGCTGTGGCAGGTGAAGCCGCTGCTCGACAAGCTCGGCATCCGCATTCTGGGCTCGATCGCCGGGGATGCGCGCTACAAGCAGGTGGCGATGGCGCAT 1100

 R  A  K  V  T  M  L  V  C  S  Q  A  L  I  N  V  A  R  K  M  Q  E  R  Y  G  I  P  Y  F  E  G  S  F
CGGGCCAAGGTGACGATGCTGGTCTGTTCGCAGGCGCTGATCAATGTCGCGCGCAAGATGCAGGAGCGCTACGGCATCCCCTATTTCGAGGGGTCATTCT 1200

Y  G  I  S  D  T  S  Q  S  L  R  R  I  C  E  L  L  V  D  Q  G  A  P  K  D  L  L  N  R  C  E  V  L  V
ACGGCATTTCCGACACTTCGCAATCGCTGCGCAGGATATGCGAATTGCTGGTGGATCAGGGCGCGCCGAAGGACCTTCTGAACCGCTGCGAGGTGCTGGT 1300

  A  R  E  E  A  K  A  W  A  A  L  K  P  F  R  P  R  V  A  G  R  R  V  L  L  Y  T  G  G  H  K  S  W
CGCCAGGGAAGAGGCAAAGGCCTGGGCGGCGCTGAAACCATTTCGCCCGCGCGTCGCGGGGCGTAGGGTGCTGCTTTACACCGGCGGGCACAAATCCTGG 1400

 S  V  A  S  A  L  Q  E  L  G  M  E  V  V  G  T  S  M  R  K  V  T  A  N  D  R  D  R  V  I  E  I  M
TCCGTCGCCTCGGCGCTGCAGGAACTGGGGATGGAGGTCGTCGGCACCTCGATGCGCAAGGTGACGGCGAATGACCGCGACCGGGTGATCGAGATCATGG 1500

G  D  D  K  H  M  C  E  N  M  A  P  R  E  M  Y  Q  E  C  C  A  R  R  A  D  V  L  L  S  G  G  R  S  Q
GCGACGACAAGCACATGTGTGAAAACATGGCCCCGCGCGAGATGTATCAAGAATGCTGCGCACGCCGGGCGGATGTGCTTTTGTCGGGCGGGCGGTCGCA 1600

  F  V  A  L  K  A  L  V  P  S  V  D  V  N  Q  E  K  H  E  P  Y  A  G  Y  M  G  M  V  D  L  V  R  A
ATTCGTCGCCCTGAAGGCGCTGGTGCCCTCGGTCGATGTCAATCAGGAAAAGCACGAACCCTATGCCGGGTATATGGGCATGGTCGATCTGGTTCGCGCC 1700

 I  D  R  S  V  N  N  P  M  W  A  D  L  R  A  P  A  P  W  D  A  S  L  T  G  S  V  V  S  V  P  S  G
ATCGACCGATCGGTGAACAACCCGATGTGGGCCGATCTGCGCGCCCCGGCGCCCTGGGATGCGTCCTTGACCGGATCGGTCGTGTCCGTCCCTTCGGGCC 1800

P  A  R  *
CTGCCCGGTGACCGGGATTTCCGCGCCGATGGGGCGAGTCCCGGTGCTGGTGTCGCAACAGGTCGTCGCGGCGCTGCTGATCCCGGTCGCGCCCAGTCTG 1900

                                                  M  A  V  L  T  H  S  R  R  A  L  S  T  N  P  L  K  T  S      NifN
GTCGCGATGACCAATTCCGTCACCGATTTCGAGGATTGCTGACATGGCCGTGCTGACCCATTCCCGTCGCGCCCTTTCGACCAATCCGCTGAAGACCTCG 2000
                                    ••••
 A  P  L  G  A  A  M  A  Y  L  G  I  E  G  A  V  P  L  F  H  G  A  Q  G  C  T  A  F  G  V  V  H  L
GCTCCCCTTGGCGCGGCCATGGCCTATCTGGGCATCGAGGGCGCGGTGCCGCTGTTTCACGGTGCGCAGGGCTGCACCGCCTTCGGCGTGGTTCACCTTG 2100

V  R  H  F  K  E  A  V  P  L  Q  T  T  A  M  N  E  V  S  T  I  L  G  G  G  E  Q  I  E  E  A  I  D  N
TCCGCCATTTCAAGGAAGCCGTGCCGCTGCAGACCACCGCGATGAACGAGGTCTCGACGATCCTTGGCGGGGGCGAGCAGATCGAGGAAGCCATCGACAA 2200

  I  R  K  R  A  N  P  K  F  I  G  I  A  S  T  A  L  T  E  T  R  G  E  D  I  A  G  E  L  R  A  M  Q
CATCCGCAAGCGGGCGAACCCGAAATTCATCGGCATCGCCTCCACCGCGCTGACCGAAACCCGGGGCGAGGATATCGCGGGCGAATTGCGCGCGATGCAG 2300

 V  R  R  K  D  W  V  G  T  A  V  V  H  V  I  T  P  D  F  E  G  G  Q  Q  D  G  W  A  K  A  V  E  A
GTCCGGCGCAAGGATTGGGTCGGGACCGCGGTCGTTCATGTCATCACCCCGGATTTCGAGGGCGGTCAGCAGGACGGCTGGGCGAAGGCTGTCGAGGCGA 2400

I  V  A  A  L  V  P  V  T  A  E  R  D  P  D  L  R  Q  V  T  L  L  V  P  S  C  F  T  T  A  E  I  D  E
TCGTCGCCGCGCTGGTCCCGGTGACGGCCGAACGCGACCCCGATCTGCGGCAGGTGACGCTGCTGGTGCCGAGCTGCTTCACCACCGCCGAGATTGACGA 2500

  A  V  R  M  I  R  A  F  G  L  S  P  I  V  L  P  D  L  S  T  S  L  D  G  H  L  S  D  D  W  S  G  H
GGCGGTGCGGATGATCCGCGCCTTTGGTCTGTCGCCCATCGTGCTGCCCGATCTCTCCACTTCGCTCGACGGGCATCTGTCGGATGACTGGTCGGGCCAT 2600

 S  L  G  G  T  R  L  D  D  I  A  R  I  P  R  S  A  V  T  L  A  I  G  E  Q  M  R  A  A  A  P  M  I
TCGCTCGGCGGCACCCGTCTGGACGACATCGCCCGCATTCCCCGTTCGGCCGTGACGCTGGCGATCGGCGAGCAGATGCGCGCCGCCGCGCCGATGATCG 2700

E  D  R  A  L  V  P  Y  R  V  F  Q  S  L  T  G  L  K  V  V  D  A  F  V  R  V  L  M  E  L  S  G  M  Q
AGGACCGCGCGCTGGTGCCCTATCGCGTCTTTCAAAGCCTGACCGGGCTGAAGGTCGTCGATGCCTTTGTCCGGGTGCTGATGGAGCTTTCGGGGATGCA 2800

  D  P  P  P  S  T  K  R  D  R  A  R  M  M  D  A  A  L  D  A  H  F  F  T  G  G  L  R  V  A  I  G  A
GGACCCGCCCCCCTCGACCAAGCGCGACCGGGCGCGGATGATGGATGCGGCGCTTGATGCGCATTTCTTCACCGGGGGCTTGCGCGTCGCCATCGGCGCC 2900

 D  P  D  L  M  F  A  L  S  T  A  L  V  S  M  G  A  E  I  V  T  A  V  T  T  T  Q  N  S  A  L  I  E
GATCCCGATCTGATGTTCGCGCTCTCCACCGCGCTTGTGTCGATGGGGGCGGAGATCGTCACCGCCGTCACCACCACGCAGAATTCCGCGCTGATCGAGA 3000

K  M  P  C  A  E  V  I  L  G  D  L  G  D  V  E  R  G  A  G  Q  A  E  A  Q  I  L  I  T  H  S  H  G  R
AAATGCCCTGCGCCGAGGTCATTCTGGGCGATCTGGGCGATGTCGAACGCGGCGCAGGGCAGGCCGAGGCGCAGATCCTGATCACCCACAGCCACGGCCG 3100
```

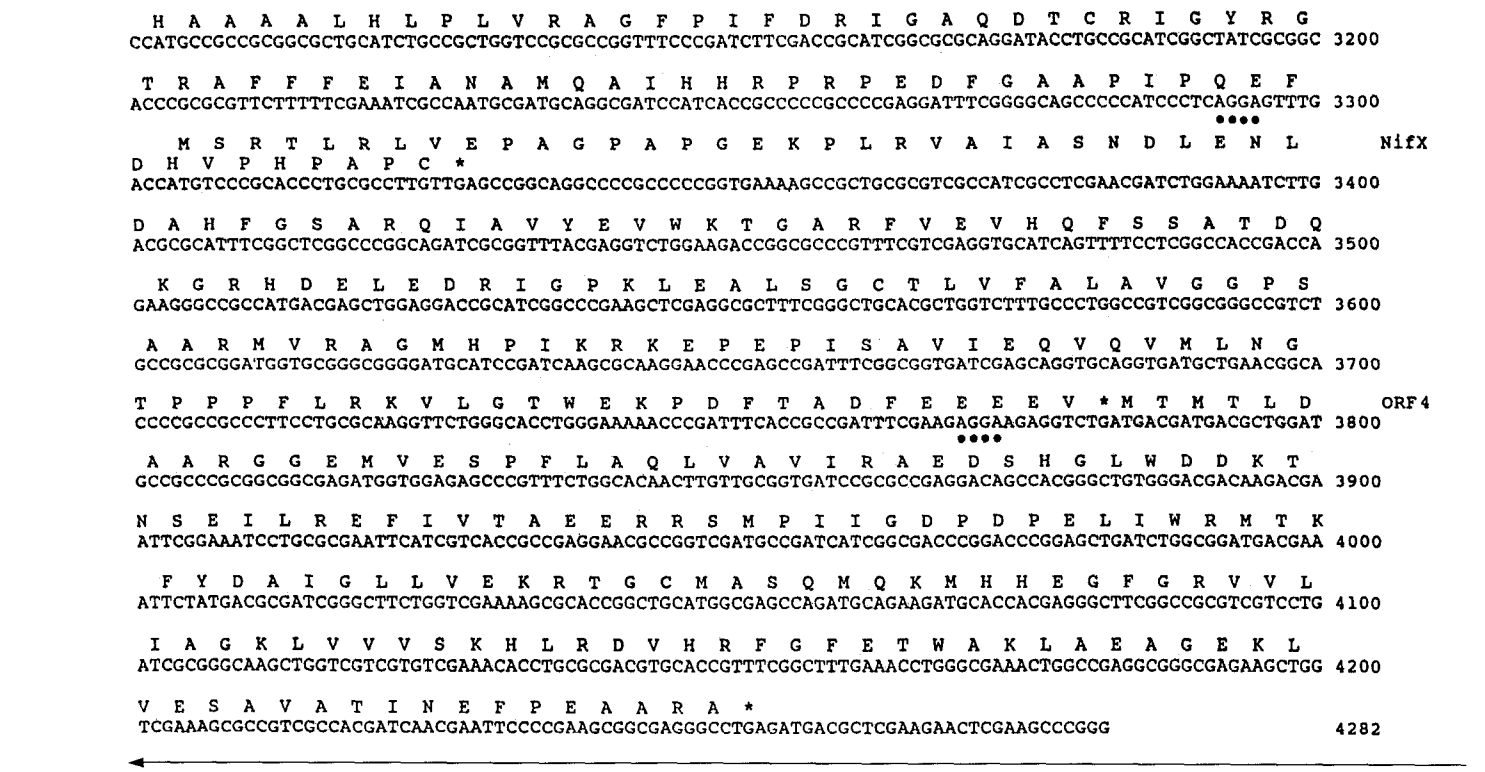

Fig. 2. Nucleotide sequence of a *Rhodobacter capsulatus* DNA fragment containing the genes *nifE*, *nifN*, *nifX* and the open reading frame ORF4. The nucleotide sequence is given in the 5′→3′ direction and the predicted amino acid sequences are indicated by single letter code. Translational stop codons are marked by *asterisks*. Possible ribosome binding sites in front of each coding region are indicated by *dots*. Two DNA motifs at positions 261 and 280 which resemble the putative NifA binding site (TGT-N_{10}-ACA) are *underlined*. Nucleotides which conform to the consensus sequence are marked by *filled bars*, mismatches are indicated by *open bars*. A *nif* consensus promoter sequence at position 334 is *underlined* and nucleotides which are identical to the promoter found in front of the *R. capsulatus nifB* gene (Masepohl et al. 1988) are marked by *filled bars*

type (Klipp et al. 1988). The 4.2 kb *Eco*RI-*Sma*I fragment, including the 3 kb *Eco*RI fragment homologous to *K. pneumoniae nifE*, was now subjected to DNA sequence analysis. A detailed restriction map and the sequencing strategy is shown in Fig. 1B, C. The complete nucleotide sequence of this 4282 bp *Eco*RI-*Sma*I fragment is presented in Fig. 2. On the basis of the availability of AUG start codons and the codon usage frequency (Staden and McLachlan 1982) characteristic of *R. capsulatus nif* genes, only four possible coding regions could be defined. These four open reading frames (ORFs) were orientated in the same direction as the *nifA* and *nifB* genes located approximately 4 kb further downstream. The first ORF encoded a protein of 477 amino acids with a predicted molecular weight of 51188 dalton. By aligning the deduced amino acid sequence with NifE sequences of *K. pneumoniae* (Arnold et al. 1988; Setterquist et al. 1988) and *A. vinelandii* (Dean and Brigle 1985; Brigle and Dean 1988) we confirmed that this ORF represents the *R. capsulatus nifE* gene (Fig. 3A).

The second ORF started 133 bp downstream of this *nifE* analogous gene and encoded a protein of 461 amino acids with a deduced molecular weight of 49459 dalton (Fig. 3B). The identification of this coding region as the *R. capsulatus nifN* gene was based on the high degree of homology between the predicted amino acid sequence and NifN protein sequences of *R. meliloti* (Aguilar et al. 1987), *K. pneumoniae* (Arnold et al. 1988; Setterquist et al. 1988) and *A. vinelandii* (Brigle et al. 1987).

A third ORF with a coding capacity of 159 amino acids was located immediately downstream of *nifN* (Fig. 2). Location, molecular weight (17459 dalton) and homology to the amino acid sequence of the *K. pneumoniae nifX* protein (Arnold et al. 1988) indicate that this open reading frame is the *R. capsulatus nifX* gene (Fig. 4A).

The 5′ end of *nifX* overlaps the 3′ end of *nifN* by 26 nucleotides. In addition, the UGA stop codon of *nifX* overlaps the AUG start codon of another potential open reading frame (called ORF4) by one base (ORF1–3 were already described by Masepohl et al. 1988). ORF4 codes for a protein of 156 amino acids with a calculated molecular weight of 17472 dalton. An overall homology of 37% between ORF4 and an ORF located at a corresponding position in *A. vinelandii* (Brigle et al. 1988; Jacobson et al. 1988) was found (Fig. 4B).

The start codons of *nifN*, *nifX* and ORF4 are preceded by putative Shine-Dalgarno sequences (AGGA). A possible ribosome binding site (GGA) according to Stormo et al. (1982) is also located in front of *nifE* (Fig. 2). The *R. capsulatus nifE* gene is preceded by a DNA sequence resembling the consensus *nif* promoter CTGG-N_8-TTGC (Beynon et al. 1983). However, no sequence which agreed perfectly with the putative NifA binding site TGT-N_{10}-ACA (Buck et al. 1986) could be identified within the 333 bp sequence upstream of the consensus *nif* promoter. Instead, two sequences (TGC-N_{10}-ACA and TGT-N_{10}-CCA), which both exhibit one mismatch to the consensus putative NifA binding site, are located in this region. As shown in Fig. 2, these two elements are separated by only 3 bp.

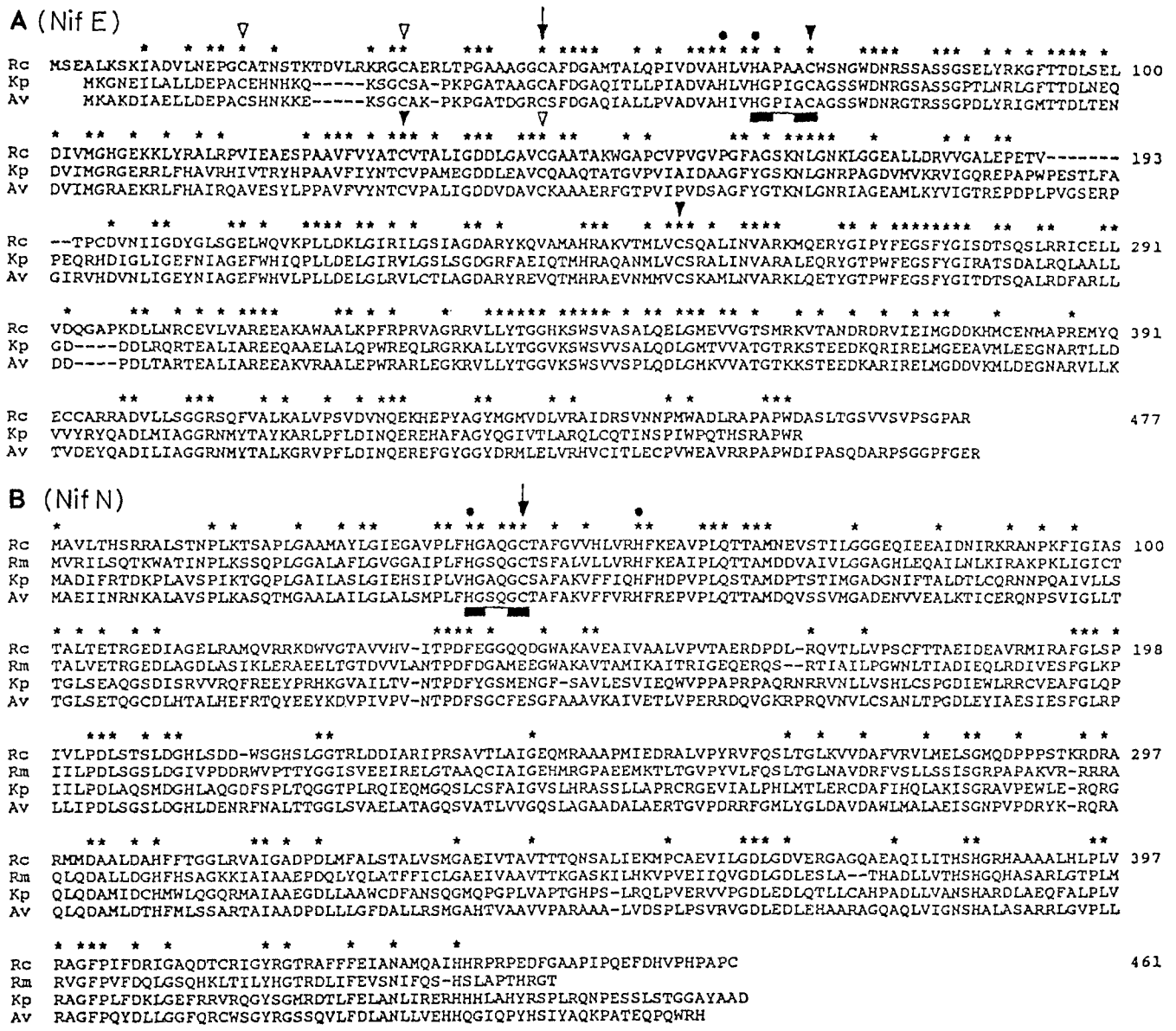

Fig. 3A and B. Comparison of the predicted amino acid sequences of the *Rhodobacter capsulatus* (Rc) *nifE* and *nifN* proteins with analogous gene products from *Klebsiella pneumoniae* (Kp), *Azotobacter vinelandii* (Av) and *Rhizobium meliloti* (Rm). In **A**, the amino acid sequences of *nifE* proteins are aligned for maximum matching and amino acids identical in all three proteins are marked by *asterisks*. Conserved cysteine residues are indicated by *open triangles*. *Filled triangles* represent cysteine residues also conserved in the *nifD* proteins from these three organisms. The cysteine residue conserved also in NifE/NifN (Fig. 4) and NifD/NifK comparisons is marked by an *arrow*. The H-G/A-X_2-G/A-C motif, conserved in all *nifE nifN*, *nifD* and *nifK* proteins, is indicated by *filled bars*. In addition, histidine residues conserved in all *nifE/nifD* proteins **A** and in all *nifN/nifK* gene products **B** are indicated by *dots*. In **B**, the sequences of the *nifN* proteins are compared and the same symbols are used. Only the *R. capsulatus* NifE and NifN amino acid sequences are numbered on the right border

Characteristics of proteins encoded by the R. capsulatus *genes* nifE, nifN, nifX, *and the open reading frame ORF4*

An alignment of the amino acid sequence of the *R. capsulatus nifE* protein with the *K. pneumoniae* (Arnold et al. 1988; Setterquist et al. 1988) and *A. vinelandii* (Dean and Brigle 1985; Brigle and Dean 1988) *nifE* gene products is shown in Fig. 3A. Homology of 47% to *K. pneumoniae* and 50% to *A. vinelandii* was found. Conserved amino acids occur along the entire length of the proteins and seven conserved cysteine residues (marked by triangles in Fig. 3A) are present in all three *nifE* proteins. The amino acid sequence of the *nifN* gene product from *R. capsulatus* was compared with *nifN* proteins from *R. meliloti* (Aguilar et al. 1987), *K. pneumoniae* (Arnold et al. 1988; Setterquist et al. 1988) and *A. vinelandii* (Brigle et al. 1987). The alignment revealed an overall homology of 44% to *R. meliloti*, 35% to *K. pneumoniae* and 39% to *A. vinelandii* (Fig. 3B). The *nifN* gene products show a higher degree of conservation in their N-terminal regions. Only one of the five cysteine residues present in the *R. capsulatus nifN* protein is conserved in the *nifN* proteins of the other three organisms (marked by an arrow in Fig. 3B).

A high degree of homology between NifE and NifN

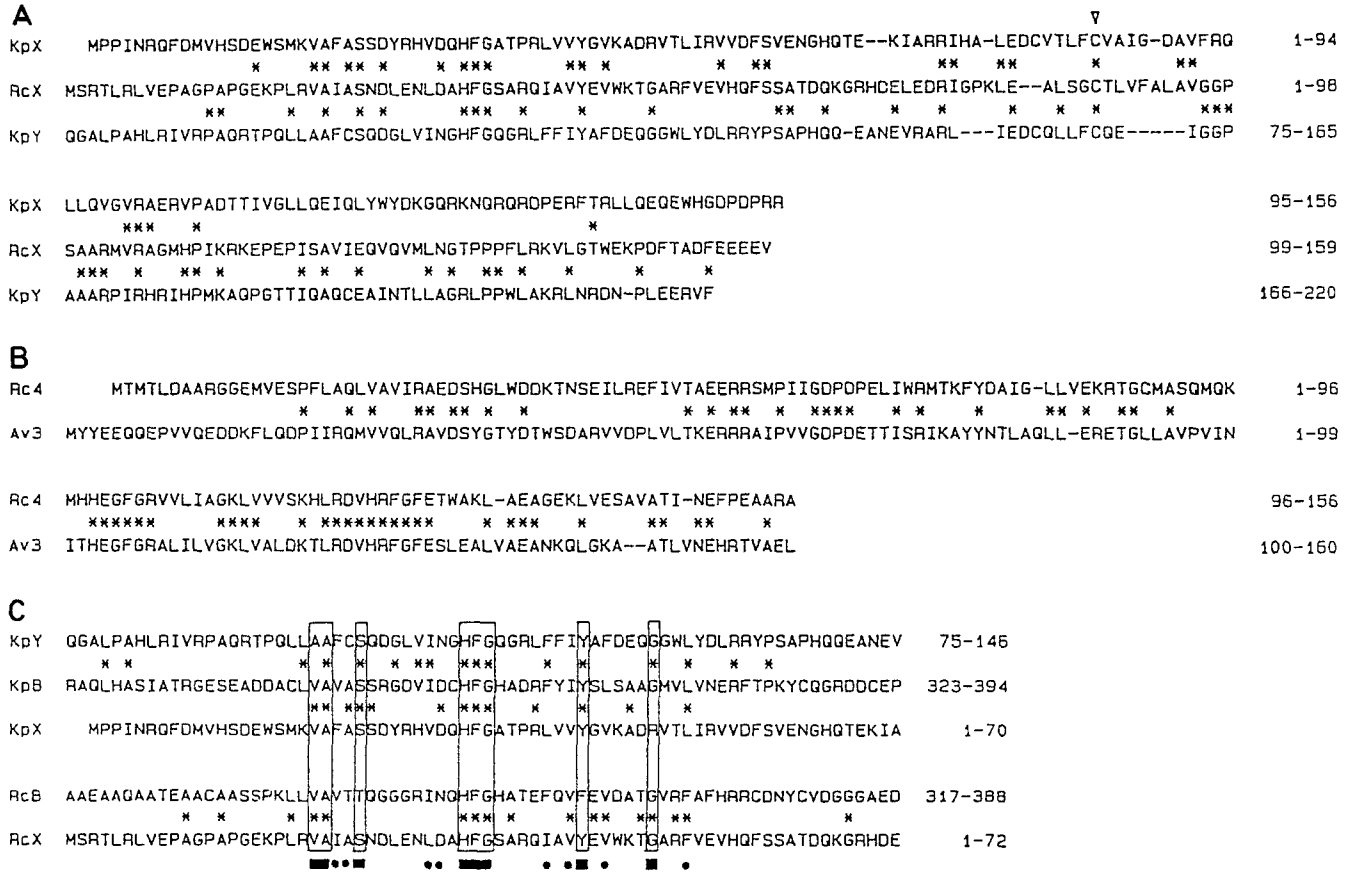

Fig. 4A–C. Comparison of the amino acid sequences of *Rhodobacter capsulatus* NifX and ORF4 with analogous gene products from *Klebsiella pneumoniae* and *Azotobacter vinelandii* and identification of a conserved domain between NifX, NifY and NifB. In **A**, the complete amino acid sequence of *R. capsulatus* NifX (RcX) is aligned for maximum matching with the complete sequence of *K. pneumoniae* NifX (KpX) and with the C-terminal part of *K. pneumoniae* NifY (KpY). Identical amino acids of adjacent sequences are indicated by *asterisks*. A conserved cysteine residue is marked by a *triangle*. A comparison of *R. capsulatus* ORF4 (Rc4) with *A. vinelandii* ORF3 (Av3) is given in **B**. Identical amino acid residues are marked by *asterisks*. In **C**, parts of NifY (KpY), NifB (KpB) and NifX (KpX) from *K. pneumoniae* (upper sequences) and from *R. capsulatus* NifB (RcB) and NifX (RcX) (lower sequences) are compared. *Asterisks* indicate identical amino acids only for adjacent sequences. Amino acids identical in at least four of the five sequences are *boxed* and the consensus is marked by *filled bars*. *Dots* in the consensus indicate amino acids which are similar in size and polarity (Doolittle 1985) in all five sequences. The exact location of the compared amino acid residues within the different proteins is indicated on the right border

with the two structural proteins NifD and NifK of the MoFe-protein was found in *A. vinelandii* (Dean and Brigle 1985; Brigle et al. 1987), *K. pneumoniae* (Arnold et al. 1988) and between NifN and NifK in *R. meliloti* (Aguilar et al. 1987). This conservation between gene products involved in FeMoco biosynthesis and the subunits of nitrogenase MoFe-protein was also confirmed for *R. capsulatus* (data not shown). Comparison of the amino acid sequence of the *R. capsulatus nifE* gene product with the *R. capsulatus nifD* protein (Schumann et al. 1986) revealed an overall homology of 23%, including four cysteine and two histidine residues which are conserved (marked by filled triangles and dots, respectively, in Fig. 3A). Since the *R. capsulatus nifK* sequence is not available, we compared *R. capsulatus* NifN to the amino acid sequence of NifK from *K. pneumoniae* (Holland et al. 1987; Arnold et al. 1988). Even in this interspecies comparison, an overall similarity of 21%, one conserved cysteine and three conserved histidine residues were found (marked by an arrow and dots, respectively, in Fig. 3B).

Structural homology between NifD and NifK was also reported (Thöny et al. 1985; Holland et al. 1987) whereas NifE and NifN of *A. vinelandii* showed only very poor similarities when compared with each other (Brigle et al. 1987). In contrast, about 21% of the amino acids of the *R. capsulatus nifE* and *nifN* gene products are homologous. The conserved amino acids are distributed along the entire length of the proteins and only one cysteine residue is conserved between NifE and NifN. This cysteine residue (marked by arrows in Fig. 3A, B) is not only conserved between NifE/N but also between NifE/D and NifN/K.

In *K. pneumoniae*, *nifDK* and *nifEN*, which code for proteins with remarkable amino acid sequence homology, are followed by additional genes in their respective transcription units. The genes *nifT*, encoding a low molecular weight polypeptide, and *nifY* are located directly downstream of *nifDK*; *nifX* is located downstream of *nifEN*. Therefore, we compared the amino acid sequence of the *R. capsulatus nifX* gene product not only with NifX but also with NifY from *K. pneumoniae* (Fig. 4A). A significant degree of sequence conservation was found between *R. capsulatus* NifX and the 146 C-terminal amino acids of the *K. pneumoniae nifY* gene product (27%) whereas both *nifX* proteins shared only 18% homology. Relationship between

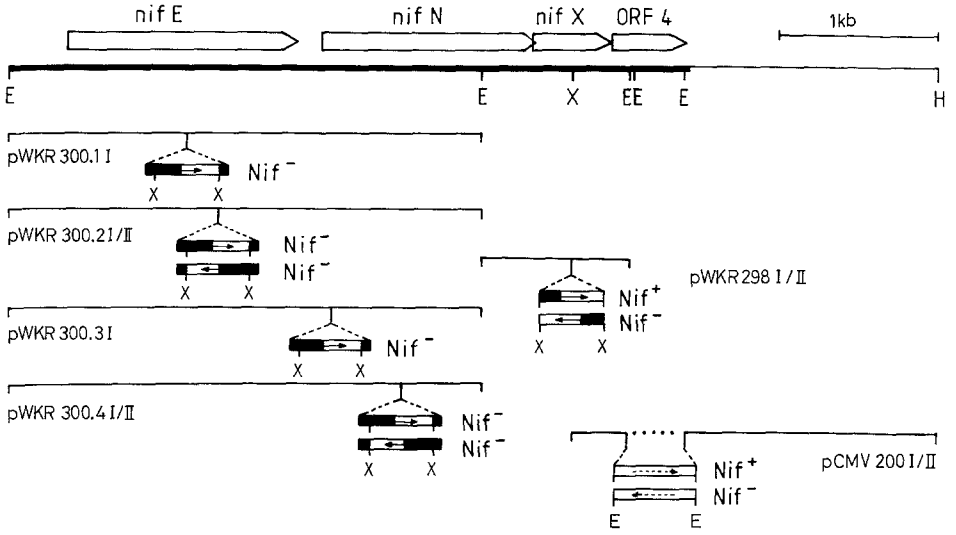

Fig. 5. Interposon mutagenesis of the *Rhodobacter capsulatus* region carrying *nifENX* and ORF4. The restriction map of the *R. capsulatus nifE-N-X*-ORF4 gene region is given for the enzymes *Eco*RI (E), *Hin*dIII (H) and *Xho*I (X). The sequenced DNA fragment is marked by a *heavy line* and the location of coding regions as deduced from the nucleotide sequence is symbolized by *open arrows*. Plasmids used for the construction of *R. capsulatus* interposon mutants are indicated below the restriction map. The cloned fragment and the point of interposon insertion is drawn for each plasmid. Interposons are not drawn to scale. *Arrows* give the direction of transcription of the kanamycin resistance gene and *black bars* symbolize parts of the inverted repeats of Tn*5*. For the construction of ORF4 mutations, two *Eco*RI fragments (*dotted line*) were substituted by a gentamicin resistance (Gm^r) interposon. The direction of transcription of the Gm^r gene is indicated by *dotted arrows*. The Nif phenotype of each *R. capsulatus* mutant, isolated by homogenotization of the corresponding interposons into the genome, is given for each construct

NifX and NifY has recently been reported also for *A. vinelandii* and *K. pneumoniae* (Brigle et al. 1988; Jacobson et al. 1988). Comparison of *A. vinelandii* and *R. capsulatus* NifX revealed an overall homology of 31%, which is distributed over the entire length of the proteins (data not shown).

The location of *nifX* downstream of the *nifEN* genes, which are known to be essential for FeMoco synthesis, suggested a possible function of NifX in the biosynthesis of FeMoco. Therefore, we compared NifX and NifY from *K. pneumoniae* and NifX from *R. capsulatus* with the corresponding NifB proteins. The *nifB* gene product is also necessary for the biosynthesis of FeMoco. As shown in Fig. 4C, the alignment of the C-terminal parts of *K. pneumoniae* NifB (Buikema et al. 1987; Arnold et al. 1988) and NifY (Arnold et al. 1988) with the N-terminal part of *K. pneumoniae* NifX (Arnold et al. 1988) and a similar comparison for NifB (Masepohl et al. 1988) and NifX from *R. capsulatus* revealed significant homologies. A protein domain encompassing approximately 33 amino acids including the motif His-Phe-Gly is highly conserved. If not only identical amino acids are considered but also amino acids which resemble each other in polarity and size (Doolittle 1985), 16 out of the 33 residues corresponded in all five proteins (Fig. 4C).

Genetic analysis of the nifE-N-X-*ORF4 gene region*

In order to confirm the Nif^- phenotype of *nifE* and *nifN* mutants and to analyze the role of *nifX* and ORF4 in nitrogen fixation, plasmids with defined interposon insertions were constructed. Interposons within each of the four coding regions were subsequently homogenotized into the *R. capsulatus* genome (for details see Materials and methods) and the Nif phenotype was determined (Fig. 5).

A. *nifE* and *nifN* mutants. To construct insertions in these two genes, the 2986 bp *Eco*RI fragment carrying *nifE* and most of *nifN* was first mutagenized by Tn*5* in *E. coli*. The resulting random Tn*5* insertions were mapped by restriction analysis and two different mutations in *nifE* and *nifN*, respectively, were chosen. To avoid further transposition of Tn*5*, one of the two internal *Xho*I fragments of the transposon, which carries most of IS*50R* including the transposase gene, was deleted. To exclude effects caused by a possible read-through of the transcript initiated at the promoter of the kanamycin resistance gene, *nifE* and *nifN* mutations carrying the remaining *Xho*I fragment in both orientations were isolated. As shown in Fig. 5, homogenotization of these six interposon insertions in *nifE* and *nifN* (pWKR300.1–4) resulted in a Nif^- phenotype.

B. *nifX* mutants. Mutations in *nifX* were constructed by cloning a kanamycin resistance interposon into the *Xho*I site of the 918 bp *Eco*RI fragment (Fig. 5). The resulting plasmids pWKR298I and pWKR298II contain the *Xho*I fragment from Tn*5*, which encodes the kanamycin resistance gene, in both orientations. Homogenotization of these two *nifX*: : [Km^r] insertions revealed a Nif^+ phenotype only if the direction of transcription of the kanamycin resistance gene was in agreement with the direction of *nifE-N-X*-ORF4, but a Nif^- phenotype if orientated in opposite direction.

C. ORF4 mutants. The *Hin*dIII-*Xho*I fragment containing the C-terminal part of *nifX*, ORF4 and the adjacent DNA region was cloned and the internal two *Eco*RI fragments encompassing most of the ORF4 coding region were replaced by a Gm^r interposon resulting in an interposon/deletion mutation (Fig. 5). The two resulting ORF4 deletions carrying the Gm^r interposon in both orientations were

introduced into the *R. capsulatus* chromosome by homologous recombination. A Nif$^+$ phenotype was found if the direction of transcription of the gentamicin resistance gene was identical to the direction of *nifE-N-X*-ORF4 and a Nif$^-$ phenotype was observed if orientated in the opposite direction (Fig. 5).

Discussion

DNA sequence analysis of an *R. capsulatus* DNA region including a fragment homologous to *K. pneumoniae nifE* (Klipp et al. 1988) proved the existence of *R. capsulatus nifE*, *nifN* and *nifX* genes. In addition, an open reading frame (ORF4) downstream of *R. capsulatus nifX* was homologous to a possible coding region found at the same position in *A. vinelandii*, but showed no homology to any of the known *nif* genes in *K. pneumoniae* (Arnold et al. 1988). The arrangement of the *R. capsulatus nifENX* genes is identical to those of *K. pneumoniae* and *A. vinelandii*. In contrast to these two organisms, the *nifENX* gene region of *R. capsulatus* is not located adjacent to the nitrogenase structural genes *nifHDK* (Klipp et al. 1988).

Comparison of the deduced amino acid sequences of the *R. capsulatus nifE* and *nifN* gene products revealed homologies not only with analogous proteins from other organisms (Fig. 3A, B) but also with NifD from *R. capsulatus* and NifK from *K. pneumoniae*. Structural homologies between NifE and NifN and the corresponding subunits of the nitrogenase MoFe-protein NifD and NifK were previously reported also for *A. vinelandii* (Dean and Brigle 1985; Brigle et al. 1987) and *K. pneumoniae* (Arnold et al. 1988). Our data confirmed the strict conservation of four cysteine and two histidine residues in all *nifE*/*nifD* proteins (Fig. 3A) and one cysteine and three histidine residues in all *nifN*/*nifK* proteins (Fig. 3B). Arnold et al. (1988) proposed a H-G-X_2-G-C motif, which is conserved in NifE, NifN, NifK and NifD of *K. pneumoniae*, as an intersubunit FeMoco binding site. This motif with the consensus sequence H-G/A-X_2-G/A-C is also present in the *nifE* and *nifN* proteins of *R. capsulatus* and the other organisms compared in this study (Fig. 3A, B).

In contrast to *A. vinelandii* (Brigle et al. 1987), a high degree of homology was found between *R. capsulatus* NifE and NifN. This conservation in amino acid sequence and the homology previously found between NifD and NifK (Thöny et al. 1985; Holland et al. 1987) support the hypothesis that all four genes have evolved from one common ancestral gene.

Downstream of the *nifN* gene of *R. capsulatus* we identified an open reading frame analogous to *K. pneumoniae* and *A. vinelandii nifX* (Arnold et al. 1988; Brigle et al. 1988; Jacobson et al. 1988). As shown in Fig. 4A, the amino acid sequence of *R. capsulatus* NifX shares a higher degree of homology to the C-terminal part of *K. pneumoniae* NifY (Arnold et al. 1988) than to *K. pneumoniae* NifX. Nevertheless, we propose to call this *R. capsulatus* gene *nifX* since: (1) the deduced molecular weight of 17459 dalton corresponds better to NifX (18229 dalton) than to NifY (24691 dalton) from *K. pneumoniae*; (2) the map position downstream of *nifEN* is in agreement with the localization of *nifX* in *K. pneumoniae*; and (3) an overlapping of the 3′ end of *nifN* and the 5′ end of *nifX* is found in both organisms.

A small protein domain of about 33 amino acids, which shows the highest degree of homology between the *R. capsulatus nifX* gene product and NifX and NifY from *K. pneumoniae*, is also present in the C-terminal part of NifB from both organisms. It is remarkable that the same 16 amino acids, which are identical or similar in these five proteins (marked in Fig. 4C), are also conserved in the *nifB* gene products of *R. meliloti* (Buikema et al. 1987), *R. leguminosarum* (Rossen et al. 1984; Buikema et al. 1987), *B. japonicum* (Noti et al. 1986) and only 1 amino acid is different in NifB of *A. vinelandii* (Joerger and Bishop 1988). The homology between *nifB*, *nifX* and *nifY* indicated an evolutionary relationship among these genes. The *nifX* and *nifY* genes may have had as their origins a *nifB*-like gene and have now evolved to something else. Considering in addition the sequence relationship between *nifE*, *nifN*, *nifD* and *nifK*, indicating that these four genes may also originate from a single gene, one can speculate that only a few genes were sufficient for ancient nitrogen fixation. Since it is known that NifB is essential for FeMoco biosynthesis the homology between NifB, NifX and NifY suggests an involvement of NifX and NifY in maturation and/or stability of FeMoco. In addition, the conserved domain may be a common site of contact with the FeMoco allowing, in the case of NifX, optimal function of the NifEN complex; NifY may play a similar role for the NifDK complex.

A *nifY* analogous gene could not yet be identified in *R. capsulatus* but it is worth noting that a comparison of NifX and ORF4 revealed identical molecular weights, almost identical amino acid composition and an overall homology of 19% (data not shown). However, the conserved domain found among NifX, NifY and NifB (Fig. 4C) was not found in ORF4 and, therefore, a role for ORF4 in FeMoco synthesis remains speculative.

To analyze the role of *nifE*, *nifN*, *nifX* and ORF4 in nitrogen fixation, we constructed defined interposon insertion and interposon/deletion mutants. All interposon insertion mutants within *nifE* and *nifN*, independent of the orientation of the cartridge carrying the kanamycin resistance gene, revealed a Nif$^-$ phenotype. This result demonstrates that *nifE* and *nifN* are essential for nitrogen fixation in *R. capsulatus* as already shown for *K. pneumoniae* (Roberts et al. 1978). *B. japonicum* (Ebeling et al. 1987), *A. vinelandii* (Brigle et al. 1987) and for *nifN* in *R. meliloti* (Aguilar et al. 1987). The Nif phenotype of *nifX* and ORF4 mutants is dependent on the orientation of the interposon. A Nif$^-$ phenotype is only observed if the direction of transcription of the interposon encoded antibiotic resistance gene is opposite to the direction of *nifE-N-X*-ORF4 transcription. This result suggests that *nifX* and ORF4 are not essential for nitrogen fixation in *R. capsulatus* whereas at least one gene required for nitrogen fixation could be located downstream of ORF4. However, no possible open reading frame could be localized within the 200 bp sequenced so far downstream of ORF4 (data not shown). Until now we could not exclude that the transcription starting from the interposons interferes with the expression of *nifEN* resulting in the observed Nif$^-$ phenotype. The homology in amino acid sequence and the conserved map position of *nifX* and ORF4 downstream of *nifEN* in different organisms suggest a role of *nifX* and ORF4 in nitrogen fixation. Since there is no duplication of the *nifE-N-X*-ORF4 gene region (data not shown) as previously found for *R. capsulatus nifA* and *nifB* (Klipp et al. 1988) one possible explanation of the Nif$^+$

phenotype of *nifX* and ORF4 mutants is the existence of genes with analogous functions. This hypothesis is supported by the homology found between NifX and NifY in *K. pneumoniae* and the existence of conserved domains between NifX, NifY and NifB, which implies a functional relationship of these gene products. Another possibility for the observed Nif^+ phenotype of these mutants is that *nifX* and ORF4 are necessary for efficient nitrogen fixation under growth conditions that have not been used.

The organization of *nifE*, *nifN*, *nifX* and ORF4 into a single transcription unit is supported by the existence of a typical *nif* consensus promoter only in front of *nifE*. In addition, overlapping of *nifN*/*nifX* and *nifX*/ORF4 indicates translational coupling of these genes (Oppenheim and Yanofsky 1980). A similar overlapping between *nifN* and *nifX* was also found in *K. pneumoniae* (Arnold et al. 1988). However, it is remarkable that the *R. capsulatus nifE* and *nifN* genes are separated by 133 bp whereas these two genes are closely linked in *K. pneumoniae* (Arnold et al. 1988; Setterquist et al. 1988) and *A. vinelandii* (Brigle et al. 1987). The promoter structure in front of *R. capsulatus nifE* agreed with the consensus sequence of NifA activated promoters found in other organisms (Beynon et al. 1983) and a comparison of the *R. capsulatus nifE*, *nifB* (Masepohl et al. 1988) and *nifH* (Pollock et al. 1988) promoters revealed an even higher degree of homology: 15 out of 17 bp are identical to the *nifB* promoter (Fig. 2) and 13 out of 17 bp to the *nifH* promoter. In contrast to the *nifB* and *nifH* promoters, which are preceded by perfect putative NifA binding sites TGT-N_{10}-ACA (Buck et al. 1986), only two imperfect motifs could be localized in front of *R. capsulatus nifE*. One or both of these structures, which are separated only by 3 bp (Fig. 2), could be involved in NifA dependent activation of the *nifE* promoter in *R. capsulatus*. Two imperfect NifA binding sites (TGT-N_{10}-GCA) were also found in front of the *R. leguminosarum dctA* gene, which is known to be activated by NifA in this organism (Ronson 1988).

Acknowledgements. We are grateful to A. Pühler for helpful advice and encouragement. We thank D.R. Dean for kindly providing unpublished sequence data, U.B. Priefer for critically reading the manuscript and helpful discussions, and S. Hennecke for excellent technical assistance. One of us (C.M.V.) was a postdoctoral fellow of the Alexander von Humboldt-Stiftung (FRG). This work was supported by financial grants from Deutsche Forschungsgemeinschaft (Kl 556/1-1 and Pu28/14-1B).

References

Aguilar OM, Reiländer H, Arnold W, Pühler A (1987) *Rhizobium meliloti nifN* (*fixF*) gene is part of an operon regulated by a *nifA*-dependent promoter and codes for a polypeptide homologous to the *nifK* gene product. J Bacteriol 169:5393–5400

Ahombo G, Willison JC, Vignais PM (1986) The *nifHDK* genes are contiguous with a *nifA*-like regulatory gene in *Rhodobacter capsulatus*. Mol Gen Genet 205:442–445

Arnold W, Pühler A (1988) A family of high-copy-number plasmid vectors with single end-label sites for rapid nucleotide sequencing. Gene 80:171–179

Arnold W, Rump A, Klipp W, Priefer UB, Pühler A (1988) Nucleotide sequence of a 24206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. J Mol Biol 230:715–738

Avtges P, Scolnik PA, Haselkorn R (1983) Genetic and physical map of the structural genes (*nifH, D, K*) coding for the nitrogenase complex of *Rhodopseudomonas capsulata*. J Bacteriol 156:251–256

Avtges P, Kranz RG, Haselkorn R (1985) Isolation and organization of genes for nitrogen fixation in *Rhodopseudomonas capsulata*. Mol Gen Genet 201:363–369

Beynon J, Cannon M, Buchanan-Wollaston V, Cannon F (1983) The *nif* promoters of *Klebsiella pneumoniae* have a characteristic primary structure. Cell 34:665–671

Brigle KE, Dean DR (1988) Revised nucleotide sequence of the *Azotobacter vinelandii nifE* gene. Nucleic Acids Res 16:5214

Brigle KE, Weiss MC, Newton WE, Dean DR (1987) Products of the iron-molybdenum cofactor-specific biosynthetic genes, *nifE* and *nifN*, are structurally homologous to the products of the nitrogenase molybdenum-iron protein genes, *nifD* and *nifK*. J Bacteriol 169:1547–1553

Brigle K, Beynon J, Newton W, May H, Dean D (1988) The isolation, sequence analysis, and mutagenesis of the *nifX* region from *Azotobacter vinelandii* and *Klebsiella pneumoniae*. In: Bothe H, de Bruijn FJ, Newton WE (eds) Nitrogen fixation: Hundred years after. Fischer, Stuttgart New York, p 320

Buck M, Miller S, Drummond M, Dixon R (1986) Upstream activator sequences are present in the promoters of nitrogen fixation genes. Nature 320:374–378

Buikema WJ, Klingensmith JA, Gibbons SL, Ausubel FM (1987) Conservation of structure and location of *Rhizobium meliloti* and *Klebsiella pneumoniae nifB* genes. J Bacteriol 169:1120–1126

Dean DR, Brigle KE (1985) *Azotobacter vinelandii nifD*- and *nifE*-encoded polypeptides share structural homology. Proc Natl Acad Sci USA 82:5720–5723

Doolittle RF (1985) Proteins. Sci Am 253:74–83

Ebeling S, Hahn M, Fischer HM, Hennecke H (1987) Identification of *nifE*-, *nifN*- and *nifS*-like genes in *Bradyrhizobium japonicum*. Mol Gen Genet 207:503–508

Fuhrmann M, Fischer HM, Hennecke H (1985) Mapping of *Rhizobium japonicum nifB*-, *fixBC*- and *fixA*-like genes and identification of the *fixA* promoter. Mol Gen Genet 199:315–322

Haselkorn R (1986) Organization of the genes for nitrogen fixation in photosynthetic bacteria and cyanobacteria. Annu Rev Microbiol 40:525–547

Hirsch PR, Beringer JE (1984) A physical map of pPH1JI and pJB4JI. Plasmid 12:139–141

Holland D, Zilberstein A, Zamir A, Sussman JL (1987) A quantitative approach to sequence comparisons of nitrogenase MoFe protein α- and β-subunits including the newly sequenced *nifK* gene from *Klebsiella pneumoniae*. Biochem J 247:277–285

Hoover TR, Robertson AD, Cerny RL, Hayes RN, Imperial J, Shah VK, Ludden PW (1987) Identification of the V factor needed for synthesis of the iron-molybdenum cofactor of nitrogenase as homocitrate. Nature 329:855–857

Imhoff JF, Trüper HG, Pfennig N (1984) Rearrangement of the species and genera of the phototrophic "purple nonsulfur bacteria". Int J Syst Bacteriol 34:340–343

Imperial J, Ugalde RA, Shah VK, Brill WJ (1984) Role of the *nifQ* gene product in the incorporation of molybdenum into nitrogenase in *Klebsiella pneumoniae*. J Bacteriol 158:187–194

Jacobson MR, Brigle KE, Bennett LT, Setterqist RA, Wilson MS, Cash VL, Beynon J, Newton WE, Dean DR (1988) Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. J Bacteriol, in press

Joerger RD, Bishop PE (1988) Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. J Bacteriol 170:1475–1487

Klipp W, Pühler A (1984) Determination of coding regions on multicopy plasmids: analysis of the chloramphenicol acetyltransferase gene of plasmid pACYC184. In: Pühler A, Timmis K (eds) Advanced molecular genetics. Springer, Berlin Heidelberg New York, pp 224–235

Klipp W, Pühler A (1986) Transcriptional analysis of the nitrogen fixation region (*nif*-region) of *Klebsiella pneumoniae* in *Escherichia coli*. In: Broughton WJ, Pühler A (eds) Nitrogen

fixation, vol 4 Molecular biology. Clarendon Press, Oxford, pp 95–105
Klipp W, Masepohl B, Pühler A (1988) Identification and mapping of nitrogen fixation genes of *Rhodobacter capsulatus*: Duplication of a *nifA-nifB* region. J Bacteriol 170:693–699
Kranz RG, Haselkorn R (1985) Characterization of *nif* regulatory genes in *Rhodopseudomonas capsulata* using *lac* gene fusions. Gene 40:203–215
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Masepohl B, Klipp W, Pühler A (1988) Genetic characterization and sequence analysis of the duplicated *nifA/nifB* gene region of *Rhodobacter capsulatus*. Mol Gen Genet 212:27–37
Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–560
Miller JH (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
Noti JD, Folkerts O, Turken AN, Szalay AA (1986) Organization and characterization of genes essential for symbiotic nitrogen fixation from *Bradyrhizobium japonicum* I110. J Bacteriol 167:774–783
Oppenheim DS, Yanofsky C (1980) Translational coupling during expression of the tryptophan operon of *Escherichia coli*. Genetics 95:785–795
Orme-Johnson WH (1985) Molecular basis of biological nitrogen fixation. Annu Rev Biophys Biophys Chem 14:419–459
Parker RC, Watson RM, Vinograd J (1977) Mapping of closed circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. Proc Natl Acad Sci USA 74:851–855
Pollock D, Bauer CE, Scolnik PA (1988) Transcription of the *Rhodobacter capsulatus nifHDK* operon is modulated by the nitrogen source. Construction of plasmid expression vectors based on the *nifHDK* promoter. Gene 65:269–275
Roberts GP, MacNeil T, MacNeil D, Brill WJ (1978) Regulation and characterization of protein products coded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*. J Bacteriol 136:267–279
Ronson CW (1988) Genetic regulation of C4-dicarboxylate transport in rhizobia. In: Bothe H, de Bruijn FJ, Newton WE (eds) Nitrogen fixation: Hundred years after. Fischer, Stuttgart New York, pp 547–551
Rossen L, Ma QS, Mudd EA, Johnston AWB, Downie JA (1984) Identification and DNA sequence of *fixZ*, a *nifB*-like gene from *Rhizobium leguminosarum*. Nucleic Acids Res 12:7123–7134
Schumann JP, Waitches GM, Scolnik PA (1986) A DNA fragment hybridizing to a *nif* probe in *Rhodobacter capsulatus* is homologous to a 16S rRNA gene. Gene 48:81–92
Setterquist R, Brigle KE, Beynon J, Cannon M, Ally A, Cannon F, Dean DR (1988) Nucleotide sequence of the *nifE* and *nifN* genes from *Klebsiella pneumoniae*. Nucleic Acids Res 16:5215
Simon R, Priefer U, Pühler A (1983) A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in gram negative bacteria. Biotechnology 1:784–791
Staden R, McLachlan AD (1982) Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res 10:141–156
Stormo GD, Schneider TD, Gold LM (1982) Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res 10:2971–2996
Thöny B, Kaluza K, Hennecke H (1985) Structural and functional homology between the α and β subunits of the nitrogenase MoFe protein as revealed by sequencing the *Rhizobium japonicum nifK* gene. Mol Gen Genet 198:441–448
Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268
Weaver PF, Wall JD, Gest H (1975) Characterization of *Rhodopseudomonas capsulata*. Arch Microbiol 105:207–216
Willison JC, Ahombo G, Chabert J, Magnin JP, Vignais PM (1985) Genetic mapping of the *Rhodopseudomonas capsulata* chromosome shows non-clustering of genes involved in nitrogen fixation. J Gen Microbiol 131:3001–3015

Communicated by H. Hennecke

Received September 22, 1988