Periplasmic nitrate-reducing system of the phototrophic bacterium *Rhodobacter sphaeroides* DSM 158: transcriptional and mutational analysis of the *napKEFDABC* gene cluster

Francisca REYES[1,2], Mónica GAVIRA[2], Francisco CASTILLO and Conrado MORENO-VIVIÁN[3]
Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad de Córdoba, 14071-Córdoba, Spain

The phototrophic bacterium *Rhodobacter sphaeroides* DSM 158 is able to reduce nitrate to nitrite by means of a periplasmic nitrate reductase which is induced by nitrate and is not repressed by ammonium or oxygen. Recently, a 6.8 kb *Pst*I DNA fragment carrying the *napABC* genes coding for this periplasmic nitrate-reducing system was cloned [Reyes, Roldán, Klipp, Castillo and Moreno-Vivián (1996) Mol. Microbiol. **19**, 1307–1318]. Further sequence and genetic analyses of the DNA region upstream from the *napABC* genes reveal the presence of four additional *nap* genes. All these *R. sphaeroides* genes seem to be organized into a *napKEFDABC* transcriptional unit. In addition, a partial open reading frame similar to the *Azorhizobium caulinodans yntC* gene and the *Escherichia coli yjcC* and *yhjK* genes is present upstream from this *nap* gene cluster. The *R. sphaeroides napK* gene codes for a putative 6.3 kDa transmembrane protein which is not similar to known proteins and the *napE* gene codes for a 6.7 kDa transmembrane protein similar to the *Thiosphaera pantotropha* NapE. The *R. sphaeroides napF* gene product is a 16.4 kDa protein with four cysteine clusters that probably bind four [4Fe-4S] centres. This iron–sulphur protein shows similarity to the NapF and NapG proteins of *E. coli* and *Haemophilus influenzae*. Finally, the *napD* gene product is a 9.4 kDa soluble protein which is also found in *E. coli* and *T. pantotropha*. The 5′ end of the *nap* transcript has been determined by primer extension, and a σ^{70}-like promoter has been identified upstream from the *napK* gene. The same transcriptional start site is found for cells growing aerobically or anaerobically with nitrate. Different mutant strains carrying defined polar and non-polar insertions in each *nap* gene were constructed. Characterization of these mutant strains demonstrates the participation of the *nap* gene products in the periplasmic nitrate reduction in *R. sphaeroides*.

INTRODUCTION

Three different types of nitrate-reducing systems have been described in bacteria [1,2]. The first type is a cytoplasmic assimilatory nitrate reductase, which enables the utilization of nitrate as the nitrogen source for biosynthesis. This enzyme is repressed by ammonium, but is not affected by oxygen [3]. The second type is a membrane-bound respiratory nitrate reductase, which catalyses nitrate respiration and the first step of denitrification to allow ATP synthesis by using nitrate as an alternative electron acceptor under anaerobic conditions. This enzyme is repressed by oxygen, but is insensitive to ammonium [1,4]. The third nitrate-reducing system is a periplasmic nitrate reductase found in some Gram-negative bacteria. This enzyme is repressed by neither ammonium nor oxygen and probably participates in redox balance and/or aerobic nitrate respiration [1,2,5,6]. However, expression of the *aeg-46.5 locus* of *Escherichia coli*, which includes the *nap* genes of this bacterium, is induced anaerobically by the transcriptional regulatory protein Fnr and in response to nitrate by the NarP protein [7].

Periplasmic nitrate reductases have been studied at the biochemical and/or genetic level in several Gram-negative bacteria, such as *Alcaligenes eutrophus* [8], *Thiosphaera pantotropha* [5,9], *E. coli* [10], *Rhodobacter capsulatus* [11] and *Rhodobacter sphaeroides* [2,6,12–14]. These enzymes are heterodimers composed of a catalytic 90 kDa subunit, containing a molybdopterin guanine dinucleotide cofactor and one [4Fe-4S] centre, and a 13–19 kDa bihaem *c*-type cytochrome. Both subunits are encoded by the *napAB* genes respectively. The electron transfer to the periplasmic nitrate reductase requires a membrane-bound tetrahaem cytochrome *c*, encoded by the *napC* gene [2,5]. Two additional genes, *napE*, encoding a monotopic integral membrane protein, and *napD*, coding for a soluble cytoplasmic protein, are also found in the complete *napEDABC* cluster of *T. pantotropha* [5]. Finally, three genes (*napFGH*) coding for putative iron–sulphur proteins have been described in the *aeg-46.5 locus* of *E. coli*, although this gene cluster lacks a *napE* homologous gene [10].

The phototrophic bacterium *R. sphaeroides* DSM 158 is neither a nitrate-assimilating nor denitrifying bacterium because it lacks both soluble assimilatory and membrane-bound respiratory nitrate reductases and it is also devoid of nitrite reductase. However, this strain contains a periplasmic nitrate reductase which uses nitrate as an ancillary oxidant to dissipate the excess of photosynthetic reducing power, with subsequent release of nitrite into the medium [2,6,13,14]. We have recently identified and cloned a 6.8 kb *Pst*I fragment carrying the *nap* gene region coding for the periplasmic nitrate reductase from *R. sphaeroides*. The sequencing of a 3882 bp *Sma*I–*Xho*I segment of this region revealed the presence of *napABC* genes [2]. In this work, we report the sequence and genetic analyses of the entire *nap* region of *R. sphaeroides* DSM 158 and the identification of four additional *nap* genes. The transcriptional start site of *napK*

Abbreviations used: Gm^R, gentamicin resistance; Km^R, kanamycin resistance; ORF, open reading frame; Tc^R, tetracycline resistance.

[1] Present address: Nitrogen Fixation Laboratory, John Innes Centre, Norwich NR4 7UH, U.K.
[2] Contributed equally to this work.
[3] To whom correspondence should be addressed (e-mail bb1movic@uco.es).

The nucleotide sequence data reported in this paper have been deposited in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession number Z46806.

was determined, and the location and possible functions of the proteins encoded by the *napKEFDABC* gene cluster are also discussed.

EXPERIMENTAL

Bacterial strains and growth conditions

Bacterial strains and plasmids used in this work are listed in Table 1. *E. coli* strains were grown aerobically in Luria–Bertani medium [16] and *R. sphaeroides* strains were grown at 30 °C on peptone–yeast extract plates or RCV medium [21] with KNO_3 and L-glutamate (1 g/l each) under phototrophic (light-anaerobiosis) or heterotrophic (dark-aerobiosis) conditions, as previously described [22,23].

Enzyme assay, analytical methods and DNA manipulations

Nitrate reductase was assayed in whole cells or cell extracts by measuring the nitrite produced in the reaction with reduced Methyl Viologen as electron donor [22]. Nitrite was measured colorimetrically [24] and protein was determined by the Lowry procedure [25] using BSA standard. DNA isolation, restriction enzyme analysis, agarose-gel electrophoresis and DNA-cloning procedures were carried out by standard methods [16].

DNA sequencing

DNA sequencing of the appropriate restriction fragments subcloned into pSVB vectors [17] was performed automatically (ABI 310; Perkin–Elmer) by the chain-termination method [26]. DNA was isolated by using the Qiagen-tip column kit (Diagen) and amplified by PCR (DNA Thermal Cycler 480; Perkin–Elmer) using commercial AmpliTaq DNA polymerase and a mixture of deoxynucleotides and fluorescent-labelled dideoxynucleotides. Sequences were determined on both strands of DNA through the different restriction sites. Sequence analysis was carried out with the Genetics Computer Group software package programs and database searches were performed using the BLAST program (NCBI). The nucleotide sequence data reported in this paper have been deposited in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number Z46806.

RNA extraction and primer extension analysis

Total RNA was isolated from *R. sphaeroides* cells grown aerobically or anaerobically with nitrate by the hot phenol method [27]. RNA concentration and integrity were determined [16]. Primer extension was carried out using the synthetic oligonucleotide 5′-CAGCAAAGCCGAGCCCAGAACG-3′, which is complementary to the *napK* coding region (positions 111 to 90 relative to the putative translational start site). After end labelling with T4 polynucleotide kinase (Boehringer) and [γ-^{32}P]dATP, extension reaction was performed with the avian myeloblastosis virus reverse transcriptase (Promega) under conditions described elsewhere [28]. To determine the size of the extension products, a DNA sequencing of the appropriate plasmid was carried out with the same oligonucleotide as a primer.

Construction of *R. sphaeroides nap*::Gm or *nap*::Km insertion mutants

The construction of insertion mutants in the *napABC* genes has been previously described [2]. To obtain defined polar and non-polar insertion mutants in the *napKEFD* genes, different *R. sphaeroides* wild-type DNA fragments were cloned into plasmids with a *mob*-Tc^R fragment from Tn5-B13 [19], to allow further mobilization and selection. The restriction sites shown in Figure 1 were used to insert the appropriate interposons. Plasmids containing the desired mutations were mobilized from *E. coli* S17-1 into *R. sphaeroides* DSM 158S by filter matings [23]. As the

Table 1 Bacterial strains and plasmids used in this work

Ap^R, ampicillin-resistance marker; Cm^R, chloramphenicol-resistance marker; Sm^R, streptomycin-resistance marker; Gm^R, gentamicin-resistance marker; Tc^R, tetracycline-resistance marker; Km^R, kanamycin-resistance marker.

Strain or plasmid	Relevant characteristics	Source or reference
Bacterial strains		
R. sphaeroides DSM 158	Wild-type, Nap$^+$	[13]
R. sphaeroides DSM 158S	Spontaneous Sm^R mutant of DSM 158, Nap$^+$	[13]
E. coli S17-1	Tra$^+$, host for mobilizable plasmids	[15]
E. coli DH5α	Host for pSVB and pBluescript plasmids	[16]
Plasmids		
pSUP202	Ap^R, Tc^R, Cm^R, Mob$^+$	[15]
pSUP2021	pSUP202::Tn5	[15]
pSVB plasmids	Ap^R, Lac$^+$	[17]
pUC plasmids	Ap^R, Lac$^+$	[18]
pBluescript SK$^+$/KS$^+$	Ap^R, *lacZ′*, f1*ori*, T7 promoter	Stratagene
pTn5-B13	pBR325::Tn5-B13 (Tc^R, Mob$^+$)	[19]
pWKR189	Ap^R, Gm^R	[20]
pFR10	6.8-kb *Pst*I fragment (*nap*) in pSVB20	[2]
pFR22Gm1/2	*napA*::Gm^R in both orientations, Tc^R, Mob$^+$	[2]
pFRHKm1/2	*napB*::Km^R in both orientations, Tc^R, Mob$^+$	[2]
pFRWKm1/2	*napC*::Km^R in both orientations, Tc^R, Mob$^+$	[2]
pFR31	Ap^R, Km^R, Tc^R, Mob$^+$	This work
pFR89	0.9 kb *Kpn*I–*Xho*I (*napKEF*) in pBluescript	This work
pFR24Gm1/2	*napK*::Gm^R in both orientations, Tc^R, Mob$^+$	This work
pFR80Km1/2	*napD*::Km^R in both orientations, Tc^R, Mob$^+$	This work
pFR86Km1/2	*napF*::Km^R in both orientations, Tc^R, Mob$^+$	This work
pFR90Km1/2	*napE*::Km^R in both orientations, Tc^R, Mob$^+$	This work

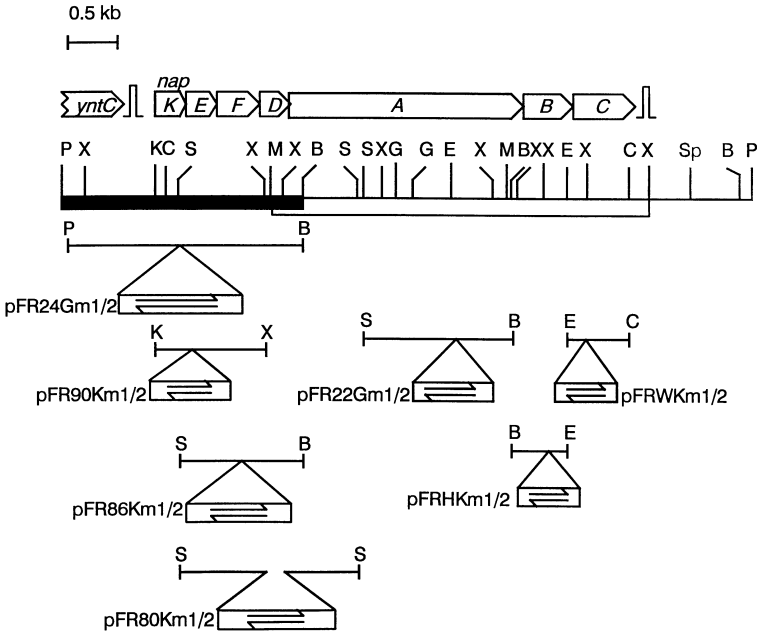


Figure 1 Physical restriction map and mutational analysis of the 6.8 kb *Pst*I fragment carrying the *R. sphaeroides nap* region

The physical map is given for the enzymes *Bam*HI (B), *Bgl*II (G), *Cla*I (C), *Eco*RI (E), *Kpn*I (K), *Pst*I (P), *Sal*I (S), *Sma*I (M), *Sph*I (Sp) and *Xho*I (X). The 2178 bp *Pst*I–*Bam*HI fragment sequenced in this work is marked by a black bar and the 3882 bp *Sma*I–*Xho*I region previously sequenced [2] is indicated by an open bar. The location, sizes and orientations of the *napKEFDABC* genes are symbolized by open arrows. DNA structures probably involved in transcriptional termination are also shown downstream from the *yntC* and *napC* genes respectively. Plasmids containing the Gm^R or the Km^R cassettes inserted in both orientations at different restriction sites are also indicated. The direction of the antibiotic-resistance genes located on the interposons, which are not drawn to scale, is symbolized by arrows.

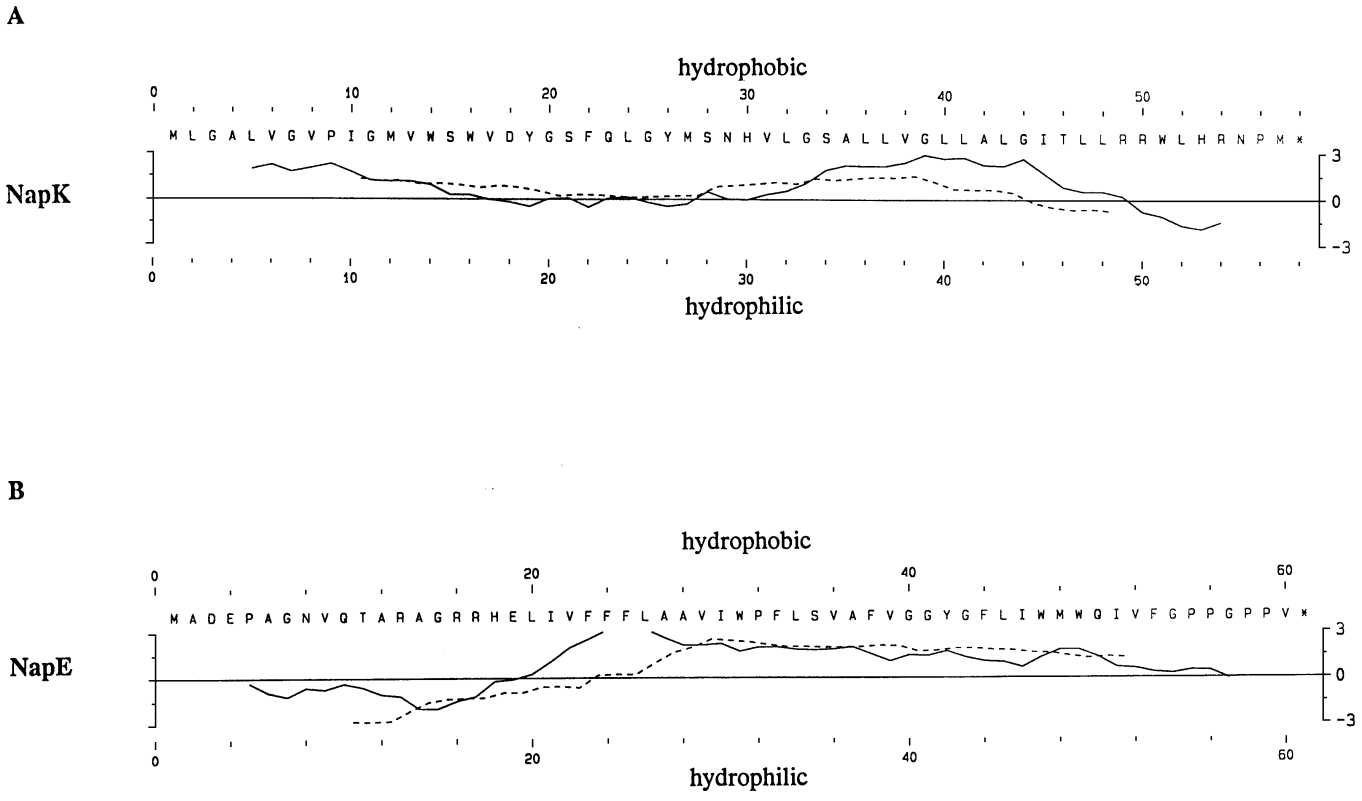


Figure 2 Hydropathy plots of *R. sphaeroides* NapK and NapE proteins

The hydropathy profiles of *R. sphaeroides* NapK (**A**) and NapE (**B**) proteins are shown according to Kyte and Doolittle [35]. The amino acid sequences of both proteins are also given.

plasmids used are not able to replicate in *R. sphaeroides*, the insertion mutation was crossed onto the genome by double homologous recombination, selected by the interposon-encoded kanamycin resistance (Km^R) or gentamicin resistance (Gm^R) and verified by the absence of the vector-encoded tetracycline resistance (Tc^R) [2].

To construct insertion mutants in the *napD* gene, a *Sal*I fragment was cloned into the *Xho*I site of pFR31. The internal *Xho*I fragment of the *napD* gene was replaced by the *Xho*I fragment with the Km^R gene from pSUP2021, which was cloned in both orientations to yield plasmids pFR80Km1 and pFR80Km2. To obtain insertion mutations in the *napF* gene, a *Sal*I–BamHI fragment was cloned into pSVB20 and the *mob*-Tc^R cassette from Tn*5*-B13 [19] was cloned into the *Bam*HI site. An *Apa*I Km^R cassette was obtained from pSUP2021 after successive subcloning steps into pBluescript KS^+ and pUC18 and inserted in both orientations into the *Apa*I site of *napF* to yield plasmids pFR86Km1 and pFR86Km2. To construct *napE* insertion mutants, a *Kpn*I–*Xho*I fragment was cloned into pBluescript KS^+ to obtain pFR89 and a *Cla*I Km^R cassette derived from pSUP2021 was inserted in both orientations into the *Nar*I site of *napE* after partial digestion. Plasmids pFR90Km1 and pFR90Km2 were finally obtained after cloning of the *mob*-Tc^R cassette into the *Bam*HI site. To obtain insertion mutations in the *napK* gene, the *Pst*I–*Bam*HI fragment was cloned into a pUC8Δ*Sal*I vector, and a Gm^R interposon derived from plasmid pWRK189 [20] was cloned in both orientations into the *Sal*I site present in the *napK* gene. The definitive constructions pFR24Gm1 and pFR24Gm2 were obtained after cloning of the *mob*-Tc^R cassette from Tn*5*-B13 [19].

RESULTS AND DISCUSSION

DNA nucleotide sequence of the *nap* gene region of *R. sphaeroides* DSM 158

We have recently cloned the *R. sphaeroides nap* gene region encoding the periplasmic nitrate reductase. This region seems to contain the putative *nap* promoter because it is expressed in *R. capsulatus* B10, a strain lacking nitrate reductase activity, with the same regulatory properties as in *R. sphaeroides* [2]. The previous nucleotide sequence of a 3882 bp *Sma*I–*Xho*I segment revealed the presence of the structural *napABC* genes. However, additional *nap* genes are clustered with the *napABC* genes in other bacteria [5,10,29]. To complete the nucleotide sequence upstream from the *R. sphaeroides napABC* genes, a 2178 bp *Pst*I–*Bam*HI fragment was sequenced. This nucleotide sequence overlaps with that reported earlier by 297 bp (Figure 1) [2]. Sequence analysis revealed the presence of four open reading frames (ORFs) with the appropriate *R. sphaeroides* codon usage and oriented in the same direction as *napABC* genes. The stop codon of each ORF overlaps the start codon or the ribosome-binding site of the next, suggesting that the four ORFs are co-transcribed with the *napABC* genes. Taking into account the sequence comparisons discussed below, we have designated these ORFs *napKEFD*, thus resulting in a *napKEFDABC* gene cluster (Figure 1). A partial ORF was also present between the first nucleotide of the sequence and a TGA stop codon at position 538–540. Sequence comparisons (not shown) revealed that the predicted 179 C-terminal amino acid residues of this partial ORF are very similar (35 % identity) to the corresponding residues of the product of a gene (*yntC*) located between the *ntrBC* and *ntrYX* genes of the nitrogen regulatory system of *Azorhizobium caulinodans* [30]. High similarity was also found to URF-2, which is the product of a gene within the mercury-resistance region of plasmid R100 and transposon Tn*501* (37 % identity) [31], to the

A

```
          ++   *      +  +  +* * ++ *  ** + *** ++*+ * ++***+ *+
RsNapE    MADEPAGNVG TA.RAGRRHE LIVFFFLAAV IWPFLSVAFV GGYGFLIWMW  49
TpNapE     MIDSAKETD RP.KHRKRDE VIAFLILAVV IWPILSVAIV GGYGFLVWMS  48
PsNapE    MAEPIATLEQ TSFRQKRRRE LLTFVVLAFG IWPVVAVGTV ASYGFAVWAY  50

          **++*****
RsNapE    QIVFGPPGPP V                                           60
TpNapE    QIIFGPPGPM H                                           59
PsNapE    QIVYGPPGPH DITPARPN                                    68
```

B

```
                +**   *  +                    *****  +      ++**
RsNapF      MSVSRRD LLAVRLTDRP .......... APIRPPWTRE AD..MARCTG   35
EcNapF    MKIDASRRG ILTGRWRKAS .......... NGIRPPWSGD ESHFLTHCTR   39
HiNapF    MTVENLPRRQ FLRGKFSTLS CLENNQKQNF VGIRPPWSVE NSIFVARCTR   50

          * +*  +*    * ++  +*  *   *   ** +**+*+ ** + + *
RsNapF    CAACAEACPA GI.VRMEAGL PQIAFAGTEC SFCGACAEAC PAPVFDIARP   84
EcNapF    CDACINACEN NILQRGAGGY PSVNFKNNEC SFCYACAQAC PESLFSPRFT   89
HiNapF    CGDCLSVCET NILVKGDAGF PEVRFDNGEC TFCGKCVDAC KQPIFYPRDQ  100

               +       +++   +  +  + + +     +  ++    +   + +
RsNapF    AFAHLAA.VT EGCLAQDGVA CMACADICPE AGISIRPRIG GPALPELSPS  133
EcNapF    RAWDLQFTIG DACLAYQSVE CRRCQDSCEP MAIIFRPTLS GIYQPQLNSQ  139
HiNapF    LPWSHKIDIS VVV                                         113

          ++ +++++    ++  + +       ++
RsNapF    LCTGCGACLS VCPAEALTIH LREPAHA                          160
EcNapF    LCNGCGACAA SCPVSAITAE Y.LHAH                           164
```

C

```
                                        +++*++   + +   +   +
RsNapD                          MREPEA HISSAILRVR PGEEAAVVRR  26
TpNapD    MREPDRTPIQ RRDILTGKLK QDSGGESRFL HISSAIVTAR PDRAADLARH  50
EcNapD                            MHTNW QVCSLVVQAK SERISDISTQ  25

            + **+**+   ++  + ++ +  *    +   +            *** ++  *
RsNapD    VMAVPGCEVA ..AAGEGRLV VLIETRETGA PRAALTELTL LEGVHSACMV  74
TpNapD    FATLPGTEVH ..AVQGAKIV LVLEGASVGE IRQPHGAISV MEGVFSANLV  98
EcNapD    LNAFPGCEVA VSDAPSGQLI VVVEAEDSET LIQTIESVRN VEGVLAVSLV  75

          ++* +
RsNapD    YEQVEALKTL GEKA                                        88
TpNapD    FEQILPADER EALS                                       112
EcNapD    YHQQEEQGEE TP                                          87
```

Figure 3 Amino acid sequence comparisons of the NapE (A), NapF (B) and NapD (C) proteins

(**A**) The predicted amino acid sequence of *R. sphaeroides* NapE (RsNapE) is aligned with the sequences of *T. pantotropha* NapE (TpNapE) [5] and the putative product of a partial ORF of *Pseudomonas* sp. (PsNapE) [36]. (**B**) *R. sphaeroides* NapF (RsNapF) is aligned with the corresponding NapF proteins of *E. coli* (EcNapF) [10] and *H. influenzae* (HiNapF) [29]. (**C**) The amino acid sequence of *R. sphaeroides* NapD (RsNapD) is aligned with the sequences of *T. pantotropha* (TpNapD) [5] and *E. coli* NapD proteins (EcNapD) [10]. Sequences are aligned for maximum matching, and identical residues at the same position are given in bold. Plus signs indicate that the amino acid residues of the *R. sphaeroides* sequence are also present in one of the other two sequences. Amino acid residues conserved in the three sequences are marked with asterisks. The putative transmembrane domain of NapE is underlined. In NapF, the Cys clusters for binding of the [4Fe–4S] iron–sulphur centres are indicated with double lines and the putative double-arginine motif for protein export is underlined.

E. coli yhjK [32] and *yjcC* [33] gene products (37 % and 29 % identity, respectively) and to *R. capsulatus* ORF25 (31 % identity), which is located next to the *mopAB* genes coding for molybdenum-pterin-binding proteins [34]. However, the physiological roles of these proteins are still unknown.

Two start codons, separated by only 6 bp, are possible for the putative *R. sphaeroides napK* gene. Although no typical ribosome-binding site is found upstream from *napK*, the purine-rich GAUGAG sequence preceding the second AUG could facilitate the ribosome binding to initiate translation at this second AUG (see Figure 4B). Therefore NapK could be a small 57-amino acid polypeptide with a molecular mass of 6282 Da and a calculated pI of 10.9. Hydropathy analysis reveals that NapK could be an integral membrane polypeptide with two transmembrane helices (residues 1–17 and 31–48; Figure 2A), but BLASTP searches show no significant similarity of NapK to database protein sequences.

Table 2 Biochemical properties and molecular composition of the Nap proteins of *R. sphaeroides* DSM 158

Data given for NapA and NapB correspond to the mature periplasmic apoproteins after cleavage of the signal peptides at the predicted sites [2]. M, Plasma membrane; P, periplasm; C, cytoplasm; MGD, molybdopterin-guanine dinucleotide; [Fe–S], iron–sulphur centres; MH, transmembrane helix.

Characteristic	NapK	NapE	NapF	NapD	NapA	NapB	NapC
Amino acid residues	57	60	160	88	802	130	227
Molecular mass (Da)	6282	6649	16360	9366	90000	14574	25531
Isolectric point	10.9	7.5	5.3	5.3	6.2	5.6	6.5
Charge (at pH 7.0)	+2	0	−4	−3	−21	−4	−5
Possible location	M	M	P	C	P	P	M
Features and cofactors	MH	MH	4 × [Fe–S]	None	MGD [Fe–S]	2 × haem *c*	MH 4 × haem *c*
Molecular composition (%)							
A+G	19	23	27	24	19	13	15
S+T	9	3	9	9	10	12	15
D+E	2	5	10	15	13	12	12
H+K+R	9	6	9	14	14	12	15
I+L+M+V	42	28	20	29	20	21	19
F+W+Y	12	18	3	1	11	5	10

The *napE* gene is preceded by a typical Shine–Dalgarno sequence and codes for a small protein of 60 residues with a molecular mass of 6649 Da and a pI of 7.5. Hydropathy analysis (Figure 2B) and sequence comparisons (Figure 3A) indicate that *R. sphaeroides* NapE is a single transmembrane polypeptide which shows high similarity to *T. pantotropha* NapE (55% identity) [5] and to the product of an ORF that is divergently transcribed from the *Pseudomonas* sp. *nirU* gene encoding the copper nitrite reductase (49% identity) [36]. NapE and NapK do not show significant similarity to each other at the amino acid sequence level but they share some properties (Table 2). They are small integral membrane proteins with a high aromatic amino acid content (12% NapK, 18% NapE). The positive charges at the C-terminus of NapK and the N-terminus of NapE (Figure 2 and 3A) are probably involved in the transmembrane insertion and are probably located in the cytoplasmic side of the membrane, as deduced by the positive-inside rule [37], whereas some of the aromatic and hydrophobic residues could be involved in protein binding. NapE protein also contains a proline-rich C-terminus (GPPGPP, Figure 3A), which could participate in protein–protein interactions, like other repetitive and non-repetitive proline-rich regions [38]. We propose that NapK and NapE could be associated with NapC in the membrane to facilitate the reduction of this membrane-bound cytochrome *c* from the quinol pool. It is known that small integral membrane proteins, such as the 40-amino acid SoxD subunit of the *Sulfolobus acidocaldarius* quinol oxidase [39], are associated with membrane redox complexes. Possible interactions between *T. pantotropha* NapE and a quinol oxidase have also been suggested [5]. Alternatively, a putative NapKE complex could anchor the nitrate reductase to the periplasmic side of the membrane by the proline-rich C-terminal motif of NapE. In *E. coli*, the integral membrane protein DmsC is required for anchoring of DMSO reductase to the membrane and for quinol oxidation [40].

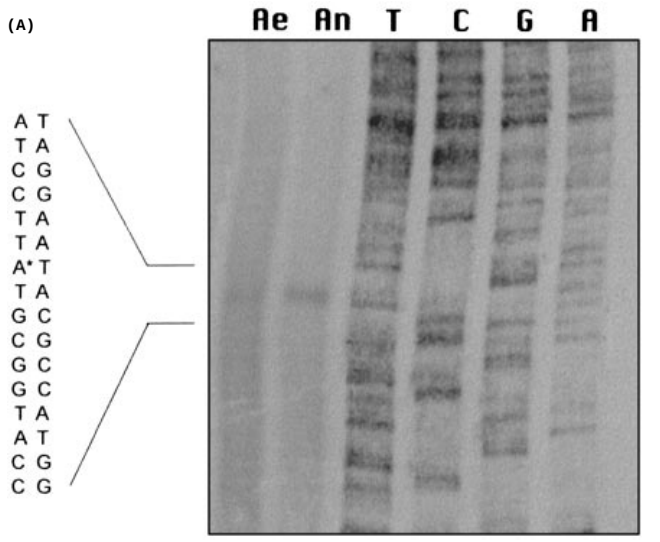
The *R. sphaeroides napF* gene is predicted to start with a GUG codon located five nucleotides downstream from the stop codon of *napE* but lacks a typical Shine–Dalgarno sequence. NapF is a putative soluble 160-amino acid protein with a molecular mass of 16360 Da and a pI of 5.3 (Table 2) containing four Cys clusters which probably bind four [4Fe–4S] centres (Figure 3B). *R. sphaeroides* NapF does not contain a typical signal peptide for protein export but includes in the N-terminus a motif matching with the double-arginine signal sequence for export of redox proteins [41,42], although this motif is not followed by a hydrophobic region. An interesting difference between the *nap* gene clusters previously reported is the presence of genes coding for iron–sulphur proteins. Thus, whereas three genes (*napFGH*) are present in *E. coli* [10], none of them is found in *T. pantotropha* [5]. Membrane-associated redox enzymes, such as DMSO reductase, fumarate reductase, nitrate reductase, formate dehydrogenase and succinate dehydrogenase, contain iron–sulphur subunits involved in the electron transfer from quinones [40]. *R. sphaeroides* NapF shows a low level of overall similarity (33–38% identity) to the NapF (Figure 3B) and NapG (not shown) proteins of *E. coli* [10] and *H. influenzae* [29]. Regions with similarities are centred around the Cys clusters required to bind the [4Fe–4S] centres and around the double-arginine motif, which is followed by an IRPPW sequence conserved in all NapF and NapG proteins (Figure 3B).

The *R. sphaeroides napD* gene product is predicted to be a soluble cytoplasmic protein of 88 residues with a molecular mass of 9366 Da and pI of 5.3 (Table 2). As shown in Figure 3C, little similarity was found between *R. sphaeroides* NapD and the corresponding NapD proteins of *T. pantotropha* (36% identity) [5] and *E. coli* (26% identity) [10]. BLASTP analysis indicates that NapD shows a slight similarity to certain regions of the HypE protein of *R. capsulatus*, which is probably involved in the processing of the NiFe-containing hydrogenase [43]. Although without experimental evidence, a role for NapD in formation of mature periplasmic nitrate reductase in *T. pantotropha* has been suggested [5].

In summary, the genetic organization of the *R. sphaeroides nap* gene cluster differs from those previously reported [5,10] in that both *napE* and *napF* genes are included, and a gene (*napK*) coding for an entirely novel protein is also present.

Primer extension analysis and identification of the *R. sphaeroides nap* promoter

Between *R. sphaeroides yntC* and *napK* genes there was a 395 bp non-coding region including an inverted repeat sequence located immediately downstream from *yntC*. This possible hairpin structure with a deduced formation energy of −79.5 kJ/mol may be involved in the termination of transcription of the *yntC* gene. To establish the real location of the *nap* promoter and to



(B)

```
GCCGCCCCACTCTCCATCGCCGCCGCCCTCGACCTGATCGCGGCATCACGAGAGGAATGA
A  A  P  L  S  I  A  A  A  L  D  L  I  A  A  S  R  E  E  *   YntC

CGCGGACCGGCGTGGCCTGCCCCATGGGGTGGTCCGGTTTCAATGGTCGTGCATGGCAGG

TCTGGGCGCACGGTTTCAGCAACGCGTCTCAGAGCCTGGAGTTGAAGCTCTCGCAACATC

CGTTCTCCCTGGGACGACCAAGTTCGATGAAGGGCGGTACGGCCGCCGACCGCGGCGATC

CAGGTGCGCACCGCAGTCGCGATGAACTCAGGGCCATAGCATGAGCGCGGGAACCCGAAT

                          AANTGTGANNTNNNTCANAT  (Crp)
                           AATAANNNNTTGATA  (IHF)
GCGGAGACGTGAATACGGGGGCCTGAAAATGTCAGCTTGATAAAAACGGCGCTGACTGGA

          -35(σ70)                  -10(σ70)      +1   KpnI
AAGCAGGCTTTGCCAAAGGGGAAGGTTACTGTACCATTATCCTTATGCGGTACCAAGGAT

          SacII              ClaI                    ApaI
CATCCAAGGAAACCCGCGGTTGGCTTTCCGCATCGATGAGCTTCATGCTCGGGGCCCTTG
                              (M  S  F) M  L  G  A  L  V  NapK

                          SalI
TCGGGGTGCCGATTGGCATGGTCTGGAGCTGGGTCGACTACGGCAGCTTCCAGCTTGGGT
  G  V  P  I  G  M  V  W  S  W  V  D  Y  G  S  F  Q  L  G  Y

ACATGAGCAATCACGTTCTGGGCTCGGCTTTGCTGGTGGGGCTTCTCGCGCTCGGCATCA
  M  S  N  H  V  L  G  S  A  L  L  V  G  L  L  A  L  G  I  T

CGCTGCTGCGTCGCTGGCTTCACCGCAATCCGATGTGAGGAGGGGCGCATGGCCGACGAG
  L  L  R  R  W  L  H  R  N  P  M  *           M  A  D  E   NapE
```

Figure 4 Primer extension analysis of the *R. sphaeroides nap* transcript (A) and nucleotide sequence of the DNA region between *yntC* and *napE* genes containing the *napK* gene and the *nap* promoter (B)

(**A**) Total RNA was isolated from cells grown aerobically (Ae) or anaerobically (An) in the presence of nitrate. A 22 nt synthetic primer complementary to the *napK* coding region (underlined in **B**) was used for the extension reaction, as described in the Experimental section. A dideoxy sequencing ladder carried out with the same primer is also shown. The transcriptional start point is indicated with an asterisk. (**B**) The nucleotide sequence of the *R. sphaeroides nap* promoter region is given in the 5′ to 3′ direction and the predicted amino acid sequences for the C-terminus of YntC, NapK and the N-terminus of NapE proteins are indicated by single-letter code. A possible hairpin structure downstream from the *yntC* gene is symbolized by thick arrows and the location of the *Kpn*I, *Sac*II, *Cla*I, *Apa*I and *Sal*I sites is shown. The transcriptional start point is in boldface and marked with +1. The −35 and −10 boxes of a σ^{70}-like promoter are indicated by bars (black for nt matching the *E. coli* consensus sequence and white for mismatching). A putative regulatory AT-rich sequence resembling the binding sites for integration host factor (WAYAA-N_4-TTGATW with W = A/T and Y = C/T) [45]) and Crp protein (AANTGTGA-N_6-TCANAT) [44] is also indicated in bold letters. Positions in the consensus IHF- and Crp-binding sites matching with the *R. sphaeroides* sequence are indicated in boldface. The TTGAT sequence of this AT-rich region also fits to the first box of the Fnr-binding site (TTGAT-N_4-ATCAA) [46]. Palindromic sequences around the AT-rich region are shown by thin arrows. Possible start codons for *napK* and *napE* are shown in italics and putative ribosome-binding sites are marked by bullets. In the NapK amino acid sequence, the first methionine and the two following residues are given in parentheses, since it is assumed that the second start codon is used.

Table 3 Methyl Viologen nitrate reductase activity in the *R. sphaeroides* wild-type and the non-polar insertion mutants in the *nap* genes

R. sphaeroides wild-type strain and non-polar insertion mutants in the different *nap* genes were grown phototrophically (light-anaerobiosis) in the presence of glutamate and nitrate to an A_{680} of 0.8. Nitrate reductase activity was assayed in whole cells using reduced Methyl Viologen as an artificial electron donor. Data correspond to four separate experiments and are expressed as mean ± S.E.M. [Gm] and [Km] symbolize non-polar insertions of respectively the gentamicin and kanamycin cassettes cloned in the indicated *nap* genes.

Strain	Nitrate reductase activity (nmol of nitrite produced/min per mg)
DSM 158 (wild-type, *napKEFDABC*)	43 ± 6
pFR24Gm2 (*napK*::[Gm]*EFDABC*)	38 ± 5
pFR90Km2 (*napKE*::[Km]*FDABC*)	34 ± 7
pFR86Km2 (*napKEF*::[Km]*DABC*)	3 ± 1
pFR80Km1 (*napKEFD*::[Km]*ABC*)	7 ± 2
pFR22Gm1 (*napKEFDA*::[Gm]*BC*)	0 ± 0
pFRHKm1 (*napKEFDAB*::[Km]*C*)	3 ± 1
pFRWKm1 (*napKEFDABC*::[Km])	45 ± 6

demonstrate that *napK* is transcribed, the 5′ end of the *nap* transcript was determined by primer extension using a 22 bp primer that hybridizes to the *napK* coding region (Figure 4). Cells grown aerobically or anaerobically with nitrate were used to isolate total RNA. Only one transcript, starting at the same position under both aerobic and anaerobic conditions, was detected. The transcript starts 51 nt upstream from the first possible AUG for *napK* and 60 nt upstream from the second AUG. A larger amount of the extension reaction product was observed with RNA isolated from anaerobically grown cells (Figure 4A).

A DNA sequence resembling a σ^{70} promoter was found upstream from the *nap* transcriptional start point, with the TTGCCA-N_{16}-TACCAT sequence as −35 and −10 putative boxes (Figure 4B). An AT-rich sequence of 21 bp with only a 24% G+C, which contrasts with the high G+C content (more than 66%) of *R. sphaeroides*, was found between 59 and 79 bp upstream from the transcriptional start site. This AT-rich region resembles slightly the binding sites for the Crp regulatory protein [44] and the integration host factor [45]. Also, a TTGAT sequence fits perfectly to one half of the Fnr-binding site [46]. The Fnr site of the *E. coli ndh* gene matches only the consensus sequence in this TTGAT half-site but seems to be necessary for efficient anaerobic repression of *ndh* [47]. However, whereas expression of *nap* genes is induced anaerobically by Fnr in *E. coli*, the nitrate reductase activity is not repressed by oxygen in *R. sphaeroides* [2]. AT-rich regions upstream from σ^{70}-like promoters, containing putative integration-host-factor-binding sites and palindromic sequences, are required for optimal expression of photosynthetic genes in *Rhodobacter* [48]. These results demonstrate that *R. sphaeroides napK* is transcribed from a σ^{70}-dependent promoter instead of a nitrogen-regulated promoter. This agrees with the normal expression of the *R. sphaeroides nap* genes in *R. capsulatus* mutants defective in *ntrA* or *ntrC* genes [2] and with our previous results, indicating that the periplasmic nitrate reductase of *R. sphaeroides* DSM 158, unlike assimilatory nitrate reductase, is not affected by ammonium or the C/N balance [2,12].

Mutational analysis of the *R. sphaeroides nap* genes

As discussed above, *napEFD* homologous genes have been identified in other bacteria but their physiological roles are unknown because a mutational analysis of these genes has not

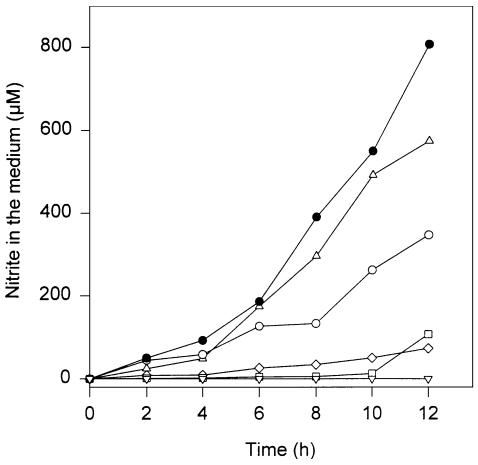


Figure 5 *In vivo* nitrite production by the *R. sphaeroides* DSM 158 wild-type strain and the non-polar insertion mutants in the *nap* genes

Cells were grown phototrophically with L-glutamate and nitrate. When the cultures showed an A_{680} of approx. 0.8, cells were collected, washed twice and resuspended in fresh medium with nitrate. Nitrite released into the medium was determined at the times indicated. Data correspond to one of four separate experiments. ●, DSM 158 (wild-type); ○, pFR242Gm2 (*napK*$^-$); △, pFR90Km2 (*napE*$^-$); □, pFR86Km2 (*napF*$^-$); ◇, pFR80Km1 (*napD*$^-$) and ▽, pFR22Gm1 (*napA*$^-$), pFRHKm1 (*napB*$^-$) or pFRWKm1 (*napC*$^-$).

been reported. To test if the *R. sphaeroides napKEFD* genes are required for a functional periplasmic nitrate reduction pathway and to analyse the possible functions of the different Nap proteins, plasmids carrying defined insertions in all the *nap* genes were constructed and crossed on to the *R. sphaeroides* genome (Figure 1). Transconjugants were grown with nitrate to determine the *in vivo* production of nitrite and the nitrate reductase activity with Methyl Viologen as artificial electron donor. Insertions of Gm^R or Km^R cassettes upstream from *napA* showed polar effects depending on the cassette orientation, suggesting that *napKEFDABC* genes are organized into an operon (not shown). As previously reported, *napABC* genes are essential for *in vivo* nitrate reduction because nitrite is not formed by mutant strains defective in these genes [2]. However, whereas the *napC* insertion mutant had normal Methyl Viologen nitrate reductase activity, the mutant with a non-polar insertion in *napB* showed a very low activity (Table 3), indicating that direct transfer of electrons from Methyl Viologen to NapA, without mediation of the cytochrome *c* subunit, takes place with a very low efficiency. In fact, the loss of the NapB subunit during purification of the *R. capsulatus* enzyme leads to a marked decrease in activity when Methyl Viologen is used as electron donor [11]. However, a direct electron transfer from Methyl Viologen to napA could be possible in *E. coli* [49].

Mutants with non-polar insertions in the *napKE* genes also showed a normal Methyl Viologen nitrate reductase activity (Table 3), although *in vivo* nitrite production was slightly lower than in the wild-type strain (Figure 5), indicating that neither gene is essential but they could be necessary for optimal nitrate reduction. These phenotypes agree with a possible role for NapK and NapE allowing optimal protein interactions between NapC and a quinol oxidase or a membrane redox complex, as discussed above. With regard to the non-polar insertion in the *napD* gene, both nitrite production and nitrate reductase activity in whole cells were much lower than in the wild-type strain (Figure 5, Table 3), which is in accordance with a role for NapD in the formation of a mature or stable complex, as previously suggested [5].

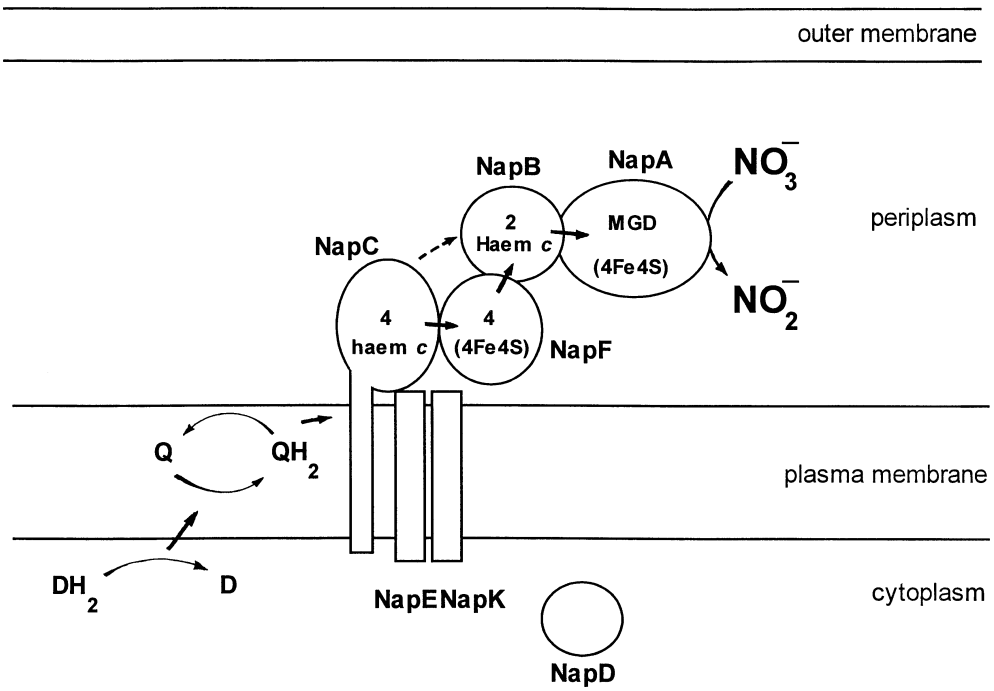


Figure 6 Model for the topological organization of the Nap proteins and the electron flow in periplasmic nitrate reduction in *R. sphaeroides* DSM 158

DH_2, cytoplasmic electron donor; QH_2, quinol pool; MGD, molybdopterin guanine dinucleotide cofactor. Arrows indicate the possible physiological electron flow from the cytoplasm to the catalytic subunit of periplasmic nitrate reductase. The dashed arrow indicates that a more inefficient electron flow between NapC and NapB is also possible.

NapF protein seems to be necessary for optimal nitrate reduction both *in vivo* and *in vitro* because nitrite production and Methyl Viologen nitrate reductase activity were very low in the *napF* non-polar insertion mutant (Figure 5, Table 3). If the *R. sphaeroides* NapF protein were located in the cytoplasm acting as the electron donor for the NapC membrane-bound cytochrome *c*, it would be expected that Methyl Viologen nitrate reductase activity of the *napF* mutant would be unaffected, as found for the *napC* mutant (Table 3). Therefore the low activity observed in the *napF* insertion mutant suggests that NapF protein is probably located in the periplasm, facilitating an optimal electron flow from both Methyl Viologen and NapC to the periplasmic NapAB complex. Thus the absence of NapF decreases both *in vitro* Methyl Viologen nitrate reductase activity and *in vivo* nitrite production (Table 3, Figure 5). An alternative explanation for the low nitrate reductase activity found in this mutant could be a decreased expression of the downstream *napABC* genes. However, this possibility seems unlikely because similar levels of nitrate reductase are found in the wild-type strain and the non-polar *napK* and *napE* insertion mutants (Table 3), suggesting that expression of the *nap* genes is not affected as a consequence of decreased readthrough from the Gm^R or Km^R genes.

NapF lacks a typical signal peptide for protein export but includes a sequence at its N-terminus (Figure 3B) matching with the motif proposed for transport and processing of redox proteins [41,42]. The possible periplasmic location of NapF suggested by the phenotype of the *napF* insertion mutant would support the notion that the twin-arginine motif is necessary for the translocation to the periplasm of this iron–sulphur protein, as recently demonstrated for the *Pseudomonas stutzeri* nitrous oxide reductase [42]. Figure 6 shows a model for the proposed location of the *R. sphaeroides* Nap proteins and the organization of the electron-transport pathway to nitrate. However, a cytoplasmic location of NapF playing a role in the assembly or processing of the iron–sulphur centre of NapA to produce a fully active and functional nitrate reductase, instead of a role in electron transfer, cannot be excluded.

We thank Dr. F. J. Florencio (University of Seville, Spain) for the facilities given for the primer extension analysis and Dr. W. Kipp (University of Bochum, Germany) for providing strains and plasmids. We also thank Dr. D. Richardson (University of East Anglia, U.K.) for critical reading of the manuscript and Dr. A. R. Franco for helpful advice on computer analysis. C.M.-V. thanks Dr. S. J. Ferguson for his hospitality during a stay at the Department of Biochemistry, University of Oxford, U.K. The financial support of DGICYT (grant PB95 0554 CO2 02) and Junta de Andalucía (CVI 0117), Spain, and Alexander von Humbolt Foundation, Germany, are gratefully acknowledged. F.R. and M.G. are recipients of fellowships from the Ministerio de Educación y Ciencia and University of Cordoba (Spain) respectively.

REFERENCES

1 Berks, B. C., Ferguson, S. J., Moir, J. W. B. and Richardson, D. J. (1995) Biochim. Biophys. Acta **1232**, 97–173
2 Reyes, F., Roldán, M. D., Klipp, W., Castillo, F. and Moreno-Vivián, C. (1996) Mol. Microbiol. **19**, 1307–1318
3 Goldman, B. S., Lin, J. T. and Stewart, V. (1994) J. Bacteriol. **176**, 5077–5085
4 Zumft, W. G. (1997) Microbiol. Mol. Biol. Rev. **61**, 533–616
5 Berks, B. C., Richardson, D. J., Reilly, A., Willis, A. C. and Ferguson, S. J. (1995) Biochem. J. **309**, 983–992
6 Roldán, M. D., Reyes, F., Moreno-Vivián, C. and Castillo, F. (1994) Curr. Microbiol. **29**, 241–245
7 Darwin, A. J. and Stewart, V. (1995) J. Mol. Biol. **251**, 15–29
8 Siddiqui, R. A., Warnecke-Eberz, U., Hengsberger, A., Schneider, B., Kostka, S. and Friedrich, B. (1993) J. Bacteriol. **175**, 5867–5876
9 Berks, B. C., Richardson, D. J., Robinson, C., Reilly, A., Aplin, R. T. and Ferguson, S. J. (1994) Eur. J. Biochem. **220**, 117–124
10 Grove, J., Tanapongpipat, S., Thomas, G., Griffiths, L., Crooke, H. and Cole, J. (1996) Mol. Microbiol. **19**, 467–481
11 Richardson, D. J., McEwan, A. G., Page, M. D., Jackson, J. B. and Ferguson, S. J. (1990) Eur. J. Biochem. **194**, 263–270
12 Dobao, M. M., Martinez-Luque, M., Moreno-Vivián, C. and Castillo, F. (1994) Can. J. Microbiol. **40**, 645–650
13 Moreno-Vivián, C., Roldán, M. D., Reyes, F. and Castillo, F. (1994) FEMS Microbiol. Lett. **115**, 279–284
14 Castillo, F., Dobao, M. M., Reyes, F., Blasco, R., Roldán, M. D., Gavira, M., Caballero, F. J., Moreno-Vivián, C. and Martínez-Luque, M. (1996) Curr. Microbiol. **33**, 341–346
15 Simon, R., Priefer, U. and Pühler, A. (1983) Biotechnology **1**, 784–791
16 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) in Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
17 Arnold, W. and Pühler, A. (1988) Gene **70**, 171–179
18 Vieira, J. and Messing, J. (1982) Gene **19**, 259–268
19 Simon, R., Quandt, J. and Klipp, K. (1989) Gene **80**, 161–169
20 Moreno-Vivián, C., Hennecke, S., Pühler, A. and Klipp, W. (1989) J. Bacteriol. **171**, 2591–2598
21 Weaver, P. F., Wall, J. D. and Gest, H. (1975) Arch. Microbiol. **105**, 207–216
22 Moreno-Vivián, C., Cárdenas, J. and Castillo, F. (1986) FEMS Microbiol. Lett. **34**, 105–109
23 Moreno-Vivián, C., Schmehl, M., Masepohl, B., Arnold, W. and Klipp, W. (1989) Mol. Gen. Genet. **216**, 353–363
24 Snell, F. D. and Snell, C. T. (1949) in Colorimetric Methods of Analysis, 3rd edn., vol. 2, pp. 804–805, Van Nostrand Reinhold, Princeton, NJ
25 Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) J. Biol. Chem. **193**, 265–275
26 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
27 Reyes, J. C. and Florencio, F. J. (1995) Plant. Mol. Biol. **27**, 789–799
28 Muro-Pastor, M. A., Reyes, J. C. and Florencio, F. J. (1996) J. Bacteriol. **178**, 4070–4076
29 Fleischmann, R. D., Adams, M. D. and 38 other authors (1995) Science **269**, 496–512
30 Pawlowski, K., Klosse, U. and de Bruijn, F. J. (1991) Mol. Gen. Genet. **231**, 124–138
31 Brown, N. L., Misra, T. K., Winnie, J. N., Schmidt, A., Seiff, M. and Silver, S. (1986) Mol. Gen. Genet. **202**, 143–151
32 Sofia, H. J., Burland, V., Daniels, D. L., Plunkett, G. and Blattner, F. R. (1994) Nucleic Acids Res. **22**, 2576–2586
33 Blattner, F. R., Burland, V. D., Plunkett, G., Sofia, H. J. and Daniels, D. L. (1993) Nucleic Acids Res. **21**, 5408–5417
34 Masepohl, B. and Klipp, W. (1996) Arch. Microbiol. **165**, 80–90
35 Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132
36 Ye, R. W., Fries, M. R., Bezborodnikov, S. G., Averill, B. A. and Tiedje, J. M. (1993) Appl. Environ. Microbiol. **59**, 250–254
37 von Heijne, G. (1992) J. Mol. Biol. **225**, 487–494
38 Williamson, M. P. (1994) Biochem. J. **297**, 249–260
39 Lübben, M., Warne, A., Albracht, S. P. J. and Saraste, M. (1994) Mol. Microbiol. **13**, 327–335
40 Weiner, J. H., Rothery, R. A., Sambasivarao, D. and Trieber, C. A. (1992) Biochim. Biophys. Acta **1102**, 1–18
41 Berks, B. C. (1996) Mol. Microbiol. **22**, 393–404
42 Dreusch, A., Bürgisser, D. M., Heizmann, C. W. and Zumft, W. G. (1997) Biochim. Biophys. Acta **1319**, 311–318
43 Colbeau, A., Richaud, P., Toussaint, B., Caballero, F. J., Elster, C., Delphin, C., Smith, R. L., Chabert, J. and Vignais, P. M. (1993) Mol. Microbiol. **8**, 15–29
44 Eiglmeier, K., Honoré, N., Iuchi, S., Lin, E. C. C. and Cole, S. T. (1989) Mol. Microbiol. **3**, 869–878
45 Goodrich, J. A., Schwartz, M. L. and McClure, W. R. (1990) Nucleic Acids Res. **18**, 4993–5000
46 Spiro, S. and Guest, J. R. (1990) FEMS Microbiol. Rev. **75**, 399–428
47 Green, J., Anjum, M. F. and Guest, J. R. (1997) Microbiology **143**, 2865–2875
48 Bauer, C. E. (1995) in Anoxygenic Photosynthetic Bacteria (Blankensip, R. E., Madigan, M. T. and Bauer, C. E., eds.), pp. 1221–1234, Kluwer Academic Publishers, Dordrecht
49 Metheringham, R., Tyson, K. L., Crooke, H., Missiakas, D., Raina, S. and Cole, J. A. (1996) Mol. Gen. Genet. **253**, 95–102

Received 24 October 1997/26 January 1998; accepted 9 February 1998