Mol Gen Genet (1993) 239:235–240

MGG

Identification of a potential transcriptional regulator of hydrogenase activity in free-living *Bradyrhizobium japonicum* strains

C. Van Soom[1], C. Verreth[1], M.J. Sampaio[2], J. Vanderleyden[1]

[1] F.A. Janssens Laboratory of Genetics, Catholic University of Leuven, Willem de Croylaan 42, B-3001 Heverlee, Belgium
[2] EMBRAPA-CENARGEN, Brasilia, Brazil

Received: 19 October 1992 / Accepted: 8 December 1992

Abstract. In *Bradyrhizobium japonicum*, Tn*5* insertions in a particular chromosomal DNA fragment result in a Hup$^-$ phenotype in free-living conditions without affecting hydrogenase (Hup) activity in the symbiotic state. By determination of the nucleotide sequence of this region, we were able to identify the nature of the inactivated genes. The fragment is located 9 kb downstream of the hydrogenase structural genes and contains one incomplete and three complete open reading frames. They are designated *hypD'*, *hypE*, *hoxX* and *hoxA* respectively, since the deduced amino acid sequences display very strong homology with genes involved in the regulation of hydrogenase activity in *Escherichia coli*, *Rhodobacter capsulatus*, *Azotobacter vinelandii* (*hypD'* and *hypE*) and *Alcaligenes eutrophus* (*hoxX* and *hoxA*). This is the first report on transcriptional activators of the *hup* genes in *B. japonicum*. Implications of these findings with respect to regulation of hydrogenase synthesis by hydrogen, oxygen and nickel in free-living *B. japonicum* are discussed.

Key words: Hydrogenase activity – Transcriptional regulation – *Bradyrhizobium japonicum* – *Alcaligenes eutrophus*

Hup$^+$ strains of *Bradyrhizobium japonicum* are capable of expressing hydrogenase activity in symbiosis as well as in free-living conditions. Free-living expression of hydrogenase activity requires the presence of micro-aerophilic conditions, hydrogen (Maier et al. 1978) and nickel (Stults et al. 1986). For *B. japonicum* the genetic determinants necessary to confer Hup activity during symbiosis have been cloned in cosmid pHU1, isolated from *B. japonicum* strain USDA122 (Haugland et al. 1984; Lambert et al. 1987). To obtain hydrogenase activity in free-living conditions, additional genetic information is necessary, that extends for an unknown distance into a 5.5 kb *Eco*RI fragment present in cosmid pHU52 (Lambert et al. 1987). In this paper, we report the cloning and nucleotide sequence of the corresponding region in *B. japonicum* CB1809, a Hup$^+$ strain, which is currently recommended as a commercial inoculant for soybean in the Cerrados region of Brazil.

Communicated by J. Schell

Correspondence to: J. Vanderleyden

Cloning of hup*-specific DNA from* B. japonicum *CB1809*

A mixture of two 20mers, synthesized according to the published nucleotide sequence of the *B. japonicum* hydrogenase structural genes (Sayavedra-Soto et al. 1988) was used to screen a genomic library in pLAFR1 (Friedman et al. 1982) constructed from a partial *Eco*RI digest of DNA from strain CB1809, according to Meinkoth and Wahl (1984). The cosmid DNA of two positive clones, pFAJ1001 and pFAJ1002, was isolated and further characterized. Both clones contain four *Eco*RI fragments in common, 12.9, 3.0, 2.5 and 0.6 kb in size. The physical organization was determined using standard mapping techniques and appears to be identical to the *Eco*RI map of the *hup* region in *B. japonicum* USDA122 (Fig. 1; Haugland et al. 1984). In order to isolate cosmid clones homologous to the pHU52 cosmid clone, the "right" 4.9 kb *Eco*RI fragment of pFAJ1002 was used to rescreen the pLAFR1 library, and also a pHC79 (Hohn and Collins 1980) library. The latter library was constructed from a partial *Eco*RI digest of *B. japonicum* CB1809 DNA and ligated in pHC79 since this cosmid accepts larger inserts. Four additional clones were isolated bringing the DNA region covered to 55 kb. The physical organization of the entire region was determined using standard mapping techniques (Ausubel et al. 1987). To confirm the presence of *hup*-specific DNA, the cosmid DNAs were hybridized with a 2.7 kb *Xho*I fragment, isolated from plasmid pRWH3 and containing part of the hydrogenase structural genes of *Rhizobium leguminosarum* (kindly provided by Dr. T. Ruiz-Argüeso). A strongly hybridizing 12.9 kb *Eco*RI fragment, present in cosmid clones pFAJ1001, pFAJ1002, pFAJ1010, and in

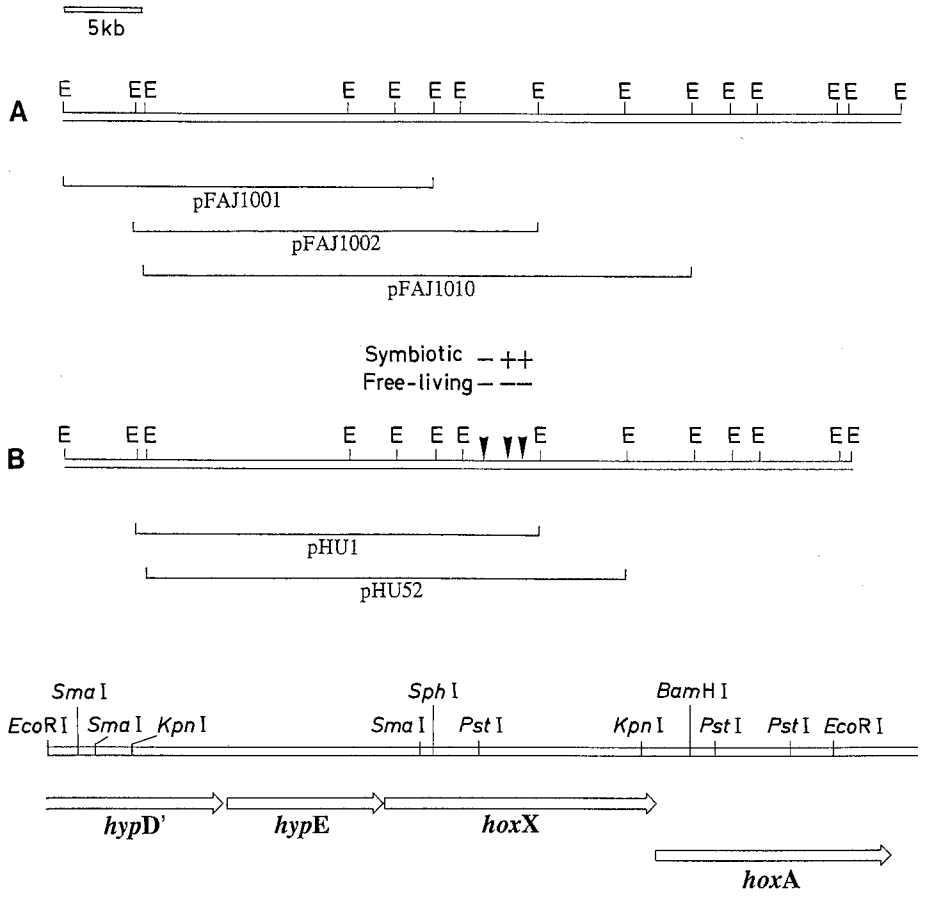

Fig. 1A, B. *Eco*RI restriction maps of the *hup* regions of *B. japonicum* CB1809 (**A**) and USDA122DES (**B**, Haugland et al. 1984). In **B**, the positions of the Tn*5* insertion mutations described by Lambert et al. (1987) are indicated by *arrowheads*, and the corresponding Hup phenotypes are denoted in symbiotic and free-living conditions

Fig. 2. Genetic organization of the 5408 bp sequenced region, containing the *hypD, hypE, hoxX* and *hoxA* homologues

the CB1809 genomic DNA, was detected. This indicates that the hydrogenase structural genes are located on this fragment, as confirmed by partial DNA sequence analysis (results not shown). Figure 1 shows the physical organization of the *hup* region of strain CB1809, and a comparison with the *hup* region of USDA122. The fact that the physical maps of the *hup* region in the two different strains are identical points to a high degree of nucleotide sequence conservation, and consequently, also, of gene organization. Hence the gene(s) previously identified as being required for free-living hydrogenase activity must be present in pFAJ1010, with one border located between 3.5 kb and 2 kb from the "right" *Eco*RI site in the 4.9 kb *Eco*RI fragment, and extending into the 5.5 kb *Eco*RI fragment (Fig. 1).

Nucleotide sequence

A series of overlapping subclones, covering the entire 4.9 kb *Eco*RI fragment and part of the 5.5 kb *Eco*RI fragment, were constructed in pUC19. Double-stranded DNA sequencing was carried out using the AutoRead Sequencing kit (Pharmacia-LKB) with an A.L.F. sequencer. Two gaps were filled using synthetic primers (synthesized by Pharmacia), so that both strands were sequenced completely. In the 5408 bp sequence (Figs. 2 and 3) four open reading frames were identified, the first being incomplete. All are transcribed from the same DNA strand as the hydrogenase structural genes, and are preceded by putative Shine-Dalgarno sequences. No typical promoter consensus sequence can be identified upstream of ORF2 and ORF4. Upstream of ORF3, a potential $-24/-12$ type promoter is present (nucleotides 1974–1987). The incomplete ORF of 1044 bp encodes 347 amino acids. ORF2, ORF3 and ORF4 could code for polypeptides of 321 amino acids (deduced molecular mass 33.6 kDa), 566 amino acids (deduced molecular mass 61.9 kDa) and 484 amino acids (deduced molecular mass 53.3 kDa), respectively.

The GC% of the entire DNA sequence is 66.2%, but reaches as much as 86.8% when only the GC bias at the third codon position is considered. Also the codon usage of the potentially coding regions corresponds well with the codon usage in *B. japonicum* group III genes (Ramseier and Göttfert, 1991).

Fig. 3. DNA sequence and deduced amino acid sequence of the *B. japonicum hypD', hypE, hoxX* and *hoxA* genes. ORF1/*hypD*, positions 1–1044; ORF2/*hypE*, 1116–2081; ORF3/*hoxX*, 2091–3791; ORF4/*hoxA*, 3788–5242. The putative Shine-Dalgarno sequences upstream of *hypE, hoxX* and *hoxA* are *underlined*, and a potential $-24/-12$ promoter consensus sequence upstream of *hoxX* is *doubly underlined*. A potential membrane-spanning region and the histidine kinase domain in HoxX are *boxed*. The sequence has been deposited in the EMBL database under the accession number Z17373

```
GAATTCTGCGGCGGCCATACCCACGCGATCTCGCGTTACGGTCTGGAGGACATGCTGCCTGCGAACGTACGG   72
 E F C G G H T H A I S R Y G L E D M L P A N V R
ATGATTCACGGTCCCGGCTGTCCGGTCTGCGTGCTGCCGGCCGGACGCATCGACATGGCGATCCGGCTCGCG  144
 M I H G P G C P V C V L P A G R I D M A I R L A
ATGCGGCCCGACATCATCCTGTGCGTCTATGGCGACCTGATGCGCGTCCCGGGCTCGCAGGGCGCATCGCTC  216
 M R P D I I L C V Y G D L M R V P G S Q G A S L
CTGAAGGCGAAGGCCCGTGGTGCCGACATCCGCATGGTCTATTCCACGATCGACGCGATCCGGATCGCCGAG  288
 L K A K A R G A D I R M V Y S T I D A I R I A E
GACAATCCCGGGCGCGAGGTGGTGTTCTTCGCCATCGGCTTCGAGACGACGACGCCGCCGACCGCGGTCATG  360
 D N P G R E V V F F A I G F E T T T P P T A V M
ATCCGATTGGCCGGGAAGAAGCAACTCGAAAACTTCAGCGTGTTCTGCAACCACGTGCTGACGCCGCCGGCG  432
 I R L A G K K Q L E N F S V F C N H V L T P P A
ATGCAGAACATTCTGGAGAGCCCCGACATCCGCAACATCGGCCGCGTCGAGATCGACGGCTTTGTCGGACCC  504
 M Q N I L E S P D I R N I G R V E I D G F V G P
GCCCACGTCTCGACCATCATCGGTACCGCGCCTTACGAGTTCTTTGCCGAGGAGTTCGGCAAGCCTGTGGTG  576
 A H V S T I I G T A P Y E F F A E E F G K P V V
ATCGCGGGCTTCGAGCCGCTCGACATGATGCAGGCGATCCTGATGCTGGTGCGGCAGGTCAACGAGCACAGG  648
 I A G F E P L D M M Q A I L M L V R Q V N E H R
CATGAGGTCGAGAACCAGTACAGCCGCGCCGTGACCCGCGACGGCAATCTGCGCGCCAAGGAAGAGGTGTCC  720
 H E V E N Q Y S R A V T R D G N L R A K E E V S
GACATCTTCGAGCTGCGCGATCAGTTCGAGTGGCGCGGGCTCGGGCAGGTGCCCTATAGCGGGTTGAAGCTG  792
 D I F E L R D Q F E W R G L G Q V P Y S G L K L
AAGCGCGCTTACGCGAAATACGACGCTGAGGTCCGCTTCGACATGAACGAGCTGCGCGTCGACGACAATCCG  864
 K R A Y A K Y D A E V R F D M N E L R V D D N P
GCCTGCGAATGCGGCGCGATCCTGCGCGGTGTGAAGAAGCCGGTGGATTGCAAGCTGTTCGGCACCGTCTGC  936
 A C E C G A I L R G V K K P V D C K L F G T V C
ACACCGGAGACGCCGATGGGCTCCTGCATGGTCTCCTCCGAAGGCGCCTGCGCCGCGCACTGGACTTACGGC 1008
 T P E T P M G S C M V S S E G A C A A H W T Y G
CGCTTCCGCGATCATCAGCAGAGGCGGGCGTCATGAAGGCCTATCAGCGCAAGCTCGACATCAGGAACGGCT 1080
 R F R D H Q Q R R A S -
GCGTCGACCTGTCCCACGGCTCGGGCGGCCGCGCCATGGCGCAGTTGATCTCCGGCCTGTTCCACGAAGCCT 1152
                                 M A Q L I S G L F H E A
TCGGCAATGAATGGCTGGCGCGCGGCAACGACCAGTCGGCGTTCGACGTCCGCGCCGGGCGCATGGTGATGA 1224
 F G N E W L A R G N D Q S A F D V R A G R M V M
CGACCGACGGCTATGTGGTCTCGCCGCTGTTCTTTCCCGGCGGCAATATCGGCTCGCTCGCGGTGCACGGCA 1296
 T T D G Y V V S P L F F P G G N I G S L A V H G
CCGTCAACGACATCGCCATGGCCGGCGCGAAGCCGCTCTATCTCTCGGCGAGCTTCATCATCGAGGAGGGCT 1368
 T V N D I A M A G A K P L Y L S A S F I I E E G
TTCGCTTCGCCGACCTCAAGCTGATCGCGGAGTCGATGGGGGCGGCGGCGCGCGAGGCCGACGTCCACATCA 1440
 F R F A D L K L I A E S M G A A A R E A G D V H I
TCACCGGTGACACCAAGGTGGTCGAGCGCGGCAAGGCCGACGGCCTGTTCATTTCGACCGCGGGCGTCGGCG 1512
 I T G D T K V V E R G K A D G L F I S T A G V G
TCGTGCCTGACGGGCTCGATCTGTCGGCCGAGAAGGCCCGCGTCGGCGATCGCGTCCTGATCTCCGGCACGC 1584
 V V P D G L D L S A E K A R V G D R V L I S G T
TCGGCGACCACGGCGTCGCCATCATGTCCAAGCGGCAGAATCTCGCTTTTGAGACCGAGATCGTCTCGGACT 1656
 L G D H G V A I M S K R Q N L A F E T E I V S D
CGGCGTCGCTGCACGATCTCGTCGCCAGGATGGTGCAGGCCGGCGGTCGTGGCATCCGCCTGATGCGCGATC 1728
 S A S L H D L V A R M V Q A G G R G I R L M R D
CCACGCGCGGCCGGCCTTGCCGCGACGCTGAACGAGATCGCCCAGCAGTCCAACCTCGGCTTCCATCTCCTCG 1800
 P T R G G L A A T L N E I A Q Q S N L G F H L L
AGGAGGCGATCCCGGTGAAGCCGGGCGTGGCTGCGGCCTGCGAGCTGCTTGGACTTGATCCGCTCCACGTCG 1872
 E E A I P V K P G V A A A C E L L G L D P L H V
CCAATGAGGGCAAGCTGGTCGCGATCGTCGCGCCGGAGGCCGCCGGAGCGGTGCTGGCCGCGATGCGGGCGC 1944
 A N E G K L V A I V A P E A A G A V L A A M R A
ATCCGCTCGGCCGCGACGCCGCCGACATTGGCGAGGCCGTTGCCGACGATCATCATTTCGTGCAGATGGCGA 2016
 H P L G R D A A D I G E A V A D D H H F V Q M A
CGAGCTTCGGCGGCGGTCGAATCGTCGATTGGCTGTCGGGCGAGCAACTGCCGAGAATCTGTTGAGGCTTCG 2088
 T S F G G G R I V D W L S G E Q L P R I C -
AGATGCGCATCCTGCTGTTATCGCATTCCTTCAACAGCCTGACGCAGCGGCTCCATGTCGAGTTGCGCGAGC 2160
    M R I L L L S H S F N S L T Q R L H V E L R E
GCGGCCACGAGGTTTCGGTCGAGCTGGACATCCACGCCGACGTCACGCGCGAGAGCGTGGCGCTGCATCGGC 2232
 R G H E V S V E L D I H A D V T R E S V A L H R
CGGACCTGGTGATCGCGCCGTTCCTGAAGCGCGCGATCCCCGACGACGTCTGGCGGAGTGTGCGCTGCCTGG 2304
 P D L V I A P F L K R A I P D D V W R S V R C L
TCGTCCATCCCGGGCCGCCCGGCGACCGCGGTCCGGCCGCGCTCGACTGGGCGGTTCTGGAAGGCGTCGCCG 2376
 V V H P G P P G D R G P A A L D W A V L E G V A
AGTGGGGCGTCACGGTGCTGCAGGCGGATGGCGAGTTCGATGCCGGTCCGGTCTGGGCGTTCCGCAGCTTCC 2448
 E W G V T V L Q A D G E F D A G P V W A F R S F
CGATGCGGCGCGCGGCCAAGTCCAGCATCTACCGCAACGAGCTAACCTCCTGCGCGGTCGCGGCCGTGCTCG 2520
 P M R R A A K S S I Y R N E L T S C A V A A V L
AAGCGGTCGCTGCGATCGAAGCCGGACAGGCAGCCCGCAGCCGATGCAGGTGGGCGACCCGCGCATTCGCG 2592
 E A V A A I E A G Q A A P Q P M Q V G D P R I R
TGCGCGGACCTTGCCGTCAGGCGGACCGGACGATCGACTGGCAGCACGACAGCACCGCGACCGTGCTGCGCA 2664
 V R G P C R Q A D R T I D W Q H D S T A T V L R
AGATCAACAGCGCCGATGGCATGCCGGGCCTCGTCGATAGCCTGTTCTTCCAGGAGGTCAGGCTGTTCGATG 2736
 K I N S A D G M P G L V D S L F F Q E V R L F D
```

```
CGCACGAAGCGCACGATATCTCCGGCGTGCCCGGCACCGTCATCGCGCAATGCGAGGGCGCGCTGGCGCGTG 2808
 A H E A H D I S G V P G T V I A Q C E G A L A R
CCACGGTCGATGGCGCGGTCTGGATCGGCCATGTCAGGCGACTCGCGCCGAAGAGCCTGAAGCTGCCGGCTG 2880
 A T V D G A V W I G H V R R L A P K S L K L P A
CGAAAGTCTTCGCCGCGGAGGCGGCGTATTTGCCGCACCGGCCTGGCTGCGGCTACGCCCCGATCCGGTATC 2952
 A K V F A A E A A Y L P H R P G C G Y A P I R Y
GCGAGCATGGCGAAGTCGGCGAGCTCGCGTTCTCCTTCTACAACGGCGCCATGGCGACCGGCGACTGCGAAG 3024
 R E H G E V G E L A F S F Y N G A M A T G D C E
CCTTGCTCGGCGCCTATCGCGCGGCGCTCGACCAGGGTGTTGCTGCTGACCGGCGGCCCCGACT 3096
 A L L G A Y R A A L E R P T R V L L L T G G P D
ATTGGTCGAACGGCATCCATCTTGCCGGGATCGAAGCGGCCGAAAGCGCCGCGGACGAGTCCTGGCGCAACA 3168
 Y W S N G I H L A G I E A A E S A A D E S W R N
TCAATGCGATCGACGATCTCGCCCGCGCGATCATCGAGACCACCGATCGCCTGGTCGTCTCGGTCATTCGCG 3240
 I N A I D D L A R A I I E T T D R L V V S V I R
GCAATGCGGGTGCGGGCGGCGTGTTCCTGAGTCTTCGCCGCCGACGAGGTCTGGGCGAGCGACCAGGTCATCC 3312
 G N A G A G G V F L S L A A D E V W A S D Q V I
TCAATCCGCATTACAAGGACATGGGCAATCTCTACGGCTCGGAATACTGGACCTATCTGCTGCCGCGCCGTG 3384
 L N P H Y K D M G N L Y G S E Y W T Y L L P R R
CGGGGGCCGCGAACGCAACCCGGATCACGCAATGCCGGCTCCCGATGGGCGTGGCGGAAGCCCGGCGCCTCG 3456
 A G A A N A T R I T Q C R L P M G V A E A R R L
CCATCGTCGACCGCGTGCTGTCGGGCGAGGCGCTCGCCGACGCAAGCCTCGTGCGGAGCGGCGCGGCCATGG 3528
 A I V D R V L S G E A L A D A S L V R S G A A M
CGTCGGATGCCGGCTTTGCCGCGCGCCTTGCCGCCAAGCAGCAGCGGCGAGCCGCCGATGAGGCCGAAAAGC 3600
 A S D A G F A A R L A A K Q Q R R A A D E A E K
CTCTGCAATCCTACCGCGAGGAGCTGCGGCGGATGAAGCTGAATTTCTACGGCTTCGATCCGAGCTATC 3672
 P L Q S Y R D E E L R R M K L N F Y G F D P S Y
ATGTCGCGCGCTACAATTTCATCCATAAGGTACCGAAGTCGCGCACGCCGCTGACCATCGCAGGTCACCGTA 3744
 H V A R Y N F I H K V P K S R T P L T I A G H R
TCAGGCGTGCGCCAGGTCGGCCTGTCGGCATGGCGGTGTCATCGTGAGCATTCAGGGAACTATTCTCGTCGT 3816
 I R R A P G R P V G M A V S S -
                                 V S I Q G T I L V V
CGACGACGAGGTCCGCTCGCAGGAAGCGTTGCGGCGCGTGCTCAGGGAGGATTTCGAGGTGCTCTGCGTCGG 3888
  D D E V R S Q E A L R R V L R E D F E V L C V G
CAACGCGACCGATGCGGAGAAGCTGCTGGAGGGCGAGATCGTCCACGCGATCCTCTGCGACCAGCGGATGCC 3960
  N A T D A E K L L E G E I V H A I L C D Q R M P
GCACGAATCCGGCGTGAGCTTCCTGAAGCGCGTCCGCGAGCTCTGGCCGGATCCGGTGCGGATGATCATTTC 4032
  H E S G V S F L K R V R E L W P D P V R M I I S
CGGCTATTCGGAGTCCGAGGACATCATCGCGGGACTGAACGAGGCCGGCATCTATCAATACATCACCAAACC 4104
  G Y S E S E D I I A G L N E A G I Y Q Y I T K P
CTGGCAGCCGGACCAGCTGGTCGAAACCGTCAAGGAGGCGGTTCAGCTCTATCGGCTGCAGAAGGAGACCGA 4176
  W Q P D Q L V E T V K E A V Q L Y R L Q K E T E
GACCGCCGGCGTCGACGTCAAGGCGACCTCCGGGCACATCAAGAAGGTCGTCTCGGTCAAGCGCGGGGTGGC 4248
  T A G V D V K A T S G H I K K V V S V K R G V A
CAAGCAGCTCTACGATTTCGACCGCATCGTGCATTCGACGGAAAGCCCGATGCACGCCGTGATCGAGCTCGG 4320
  K Q L Y D F D R I V H S T E S P M H A V I E L G
CAGGCGCGCCGCCGACTACGATATCTCGGTTCTGATCACCGGCGATCCGGCACCGGCAAGGAGCTTCTGGC 4392
  R R A A D Y D I S V L I T G E S G T G K E L L A
TCGCGCGATCCATTACGGCTCGGCGCGGGCCAACAGGGCGTTCGTGGTCGAGAACTGCGGCGCGCTGCCGGA 4464
  R A I H Y G S A R A N R A F V V E N C G A L P D
CGAGCTCCTGGAAAGCGAGCTGTTCGGTTGCAAGAAGGGCGCCTTCACCGGCGCCTACCAGGATCGCATCGG 4536
  E L L E S E L F G C K K G A F T G A Y Q D R I G
CCTGTTCGAGGTCGCCGATGGCGGCACGATCTTCCTCGACGAGATCGGCGAGACCTCGCCGGCGTTCCAGGT 4608
  L F E V A D G G T I F L D E I G E T S P A F Q V
CAAGCTGCTGCGTGTGCTGCAGGAGAGCGAGATCCGGCCGCTCGGCGCCGCACGTTGCCGCAAGGTCGACGT 4680
  K L L R V L Q E S E I R P L G A A R C R K V D V
TCGCGTGGTGGCGGCGACCAATCGCGACCTGGAGGCGGAGGTCGAGGCCGGCCGCTTCCGGCGACCTGTACT 4752
  R V V A A T N R D L E A E V E A G R F R R P V L
ACCGCCTCGCCGCATTCCCGGTGCACATGCCGGCGCTGCGCGAGCGCCCGATGGACATCCCGCTGATTGCGG 4824
  P P R R I P G A H A G A A R A P D G H P A D C G
AGGGCGTGCTGTCGGCGGTCAAGAGTTCCTTCAACCGGCCGAACTCCTCTTCGCGCGTTCGGCGCTGGAGGA 4896
  G R A V G G Q P A E L L F A R S A L E E
ATTCGGCAAATATCACTGGCCGGGCAACGTCCGCGAGCTCCAGAACGAGATCCAGCGCATGGCCGTGCTCGC 4968
  F G K Y H W P G N V R E L Q N E I Q R M A V L A
CGACCGGGACGAGCTGGCCGCACCGCCGCTGCTCGGCCGCCGCAACGGGAAGCGATCCGCGCCGCTGCCGGC 5040
  D R D E L A A P P L L G R R N G K R S A P L P A
CCACGGCAGGCTGAACGGATCGGCCTCGCTGAAAGACAAGGTGGAGGATCTCGAGAAGTCGGTTATCATGAA 5112
  H G R L N G S A S L K D K V E D L E K S V I M N
CTGCCTGGAGAGGAACGACGGAAACATCAGTCGTGTCGCCAGCGAGCTGGGTCTCTCACGGGTAGGGCTCAG 5184
  C L E R N D G N I S R V A S E L G L S R V G L R
GAACAAGCTGTCCAGGTATGATCTGAGAAAAAATGCCAAAGGCGACGCATTCTCCTGAAGCGAGCGAGGCGG 5256
  N K L S R Y D L R K N A K G D A F S -
ATTCGCTGCTCAAGCTGGTCACCAGCGGCGGCGCGATCGAGTTCGACCGCGAGCGCGAAGGCGCCTGGATCG 5328
AAGTCATCCGCAAGATGGACGAGGTCTATTCGGACCTGTTGCGCTACGAGGTCGACCTCGAGCACAAGAACG 5400
CGGAGCTC 5408
```

Nucleotide sequence comparison with other genes

A database search revealed a strong homology between the nucleotide sequences of *B. japonicum* ORF1, ORF2, ORF4 and part of ORF3 and genes involved in hydrogen metabolism found in *E. coli, Rhodobacter capsulatus, Azotobacter vinelandii,* and *Alcaligenes eutrophus*. Comparisons of the corresponding deduced amino acid sequences indicated functional conservation between different species.

The deduced amino acid sequence of the part of ORF1 that was sequenced displayed 77.2% similarity with the deduced amino acid sequence of the *A. vinelandii* ORF8 (Chen and Mortenson 1992), 69.9% similarity with the deduced amino acid sequence of the *R. capsulatus hypD* (J. Caballero, C. Delphin, P. Richaud, B. Toussaint, A. Colbeau, P. Vignais, unpublished results) and 65.5% similarity with the deduced amino acid sequence of the *E. coli hypD* (Lutz et al. 1991). Based on this homology between ORF1 and the *hypD* gene of various species, we propose to call the incomplete ORF1 *hypD'*. No function has yet been assigned to the HypD gene product. The C–X–X–C motif that is found at the amino-terminal end of the gene product in five different species can play a role in the binding of Fe-S clusters, and is assumed to have a function in metal binding. Metal-binding proteins might be involved in metal transport, in metal cluster synthesis, and in assembly and insertion of metal clusters into the hydrogenase holoprotein.

The predicted product of ORF2 is highly homologous to the translation product of the *R. capsulatus* and *E. coli hypE* genes (65.5% and 62% similarity, respectively) (J. Caballero, C. Delphin, P. Richaud, B. Toussaint, A. Colbeau, P. Vignais, unpublished results; Lutz et al. 1991). A Tn*5* insertion in this region, previously shown to abolish hydrogenase activity both in free-living and symbiotic conditions (Lambert et al. 1987), must be located in this gene, which we call *hypE*.

The 84 carboxyterminal amino acids of ORF3 display almost 60% similarity with the carboxyterminal region of the *A. eutrophus hoxX* sequence. Of the latter, only the sequence of the 3′ terminal 261 nucleotides has been published, and is located immediately upstream of the *hoxA* transcriptional activator sequence (Eberz and Friedrich 1991). It is suggested by B. Friedrich (personal communication) that the *hoxX* gene product might be a sensor protein, forming a two-component sensor/effector couple with the *hoxA* gene product, since the HoxX amino acid sequence contains a potential membrane-spanning region and an amino acid motif characteristic of a histidine kinase. A membrane location is very plausible for a sensor protein that perceives environmental stimuli outside the cell and transduces these stimuli inside the cell by phosphorylation of the effector, which is then activated to regulate the transcription. The deduced amino acid sequence of the *B. japonicum* ORF3 also contains a potential membrane-spanning region and an amino acid motif characteristic for a histidine kinase (Fig. 3) (Ninfa and Bennett 1991). Two Tn*5* insertions in the 4.9 kb *Eco*RI fragment on the *B. japonicum* chromosome which give a Hup^- phenotype only in free-living bacteria (Lambert et al. 1987), can both be localized in ORF3. This points to a functional role of the ORF3 gene product in the regulation of hydrogenase activity in free-living *B. japonicum,* and we propose to call this gene *hoxX*. Since the two aforementioned mutants could only be complemented by cosmid pHU52, which contains an additional 5.5 kb *Eco*RI fragment (Lambert et al. 1987), it can be assumed that *hoxX* forms an operon with the distally located ORF4, which extends for more than 300 nucleotides into the 5.5 kb *Eco*RI fragment. This assumption is supported by the close proximity of the two ORFs, and by the absence of promoter consensus sequences upstream of ORF4.

The deduced amino acid sequence of ORF4 displays extensive homology to that of the *A. eutrophus hoxA* gene product (61.9% similarity) (Eberz and Friedrich 1991), and to other transcriptional regulators of the NtrC family, such as the deduced amino acid sequence of *R. capsulatus* $HupR_1$ (53.3% similarity) (Richaud et al. 1991), and *E. coli* HydG (48.7% similarity) (Stoker et al. 1989).

All these gene products belong to the family of two-component sensor/effector systems, and share some functional characteristics (Fig. 4). The homology between the different polypeptides is most striking in the central region, where a conserved G–X–X–G–X–G–K–E sequence has been proposed to be an ATP-binding site. The amino-terminal region displays more variation. In other NtrC-like regulators, this part of the protein has been shown to be involved in the regulation of the regulatory activity through NtrB-mediated phosphorylation of conserved aspartic acid residues (positions 12 and 54) or lysine residues (position 105). The carboxyterminal region is also less conserved. A helix-turn-helix motif, characteristic of DNA-binding proteins, was identified in the region extending from amino acids 451 to 470 by computer analysis (method of Dodd an Egan 1991). All these data provide strong evidence for the identification of a NtrC-like transcriptional activator. Although the postulated *hoxX* gene product shows no homology to NtrB, its structural characteristics make it a likely candidate for the sensor component of a two-component sensor/effector couple with the *hoxA* gene product. This regulatory system is essential for hydrogenase activity in free-living *B. japonicum,* but is not required in symbiotic conditions. In *B. japonicum,* free-living *hup* gene expression is regulated at the transcriptional level by nickel, hydrogen and oxygen (Kim and Maier 1990; Kim et al. 1991). A region upstream of the hydrogenase structural genes, with its left border located between bases −168 and −118 relative to the start site of transcription, is necessary for regulation by all three components (Kim et al. 1991). A sensor/effector system would be a suitable candidate for this kind of regulation: upon activation by a sensor protein, an NtrC-like effector protein could bind to an upstream activating sequence, and thus activate transcription from a −24/−12 promoter. Indeed such a promoter consensus sequence can be identified at −47/−31 from the transcription initiation site (Sayavedra-Soto et al. 1988). Several facts, however, suggest a more complex regulatory pathway. Firstly, in *B. japonicum,* ribulose-biphosphate carboxylase is coordinately ex-

```
KpNtrC       MQRGIAWIVDDDSSIRWVLERALTGAGLSCTTFESGNEVLDALTTKTPDVLLS    53
EcHydG     MTHDNIDILVVDDDISHCTILQALLRGWGYNVALANSGRQALEQVREQVFDLVLC    55
AeHoxA     MSDKQATVLVVDDETRSQDALRRTLDE-EFRVLTVSSADEARALLLRQPVSVILC    54
BjHoxA      VSIQGTILVVDDEVRSQEALRRVLRE-DFEVLCVGNATDAEKLLEGEIVHAILC    53
RcHupR1    MAASAPAILLVDDEPHSLAAMKLALED-DFDVLTAQGAEAAIAILEEEWVQVIIC    54
                    .***.     .   *                  .       ..

KpNtrC     DIRMPGMDGLALLKQIKQRHPMLPVIIMTAHSDLDAAVSAYQQ-GAFDYLPKPFD   107
EcHydG     DVRMAEMDGIATLKEIKALNPAIPVLIMTAYSSVETAVEALKT-GALDYLIKPLD   109
AeHoxA     DQRMPGLTGVEFLKEVRERWPEIVRIVISGYTDSEDIIAGVNEAGIYQYILKPWV   109
BjHoxA     DQRMPHESGVSFLKRVRELWPDPVRMIISGYSESEDIIAGLNEAGIYQYITKPWQ   108
RcHupR1    DQRMPGRTGVDFLTEVRERWPETVRIIITGYTDSASMMAAINDAGIHQFLTKPWH   109
           * **    *.  *  ..   *    ....  .     .      *   .. **

KpNtrC     IDE-AVALVDRAISHYQEQQQPR--------------------NAPINSPTADII   141
EcHydG     FDNLQATWKKRSHTHSIDAETP----------------------AVTASQFGMV   141
AeHoxA     PDHLIDTVRQAVEAQGLQGDMHRLDLELRTSTPVLRQRSSQKLASAQSAFNFERI   164
BjHoxA     PDQLVETVKEAVQLYRLQKETETAGVDVKATSGHIKKVVSVKRGVAKQLYDFDRI   163
RcHupR1    PEQLLSSARNAARMFTLARENERLSLEMRLLNSTSESRVEKRRRALREGMGFETI   164
             .   .                                              .

KpNtrC     GERP--AMQDVFRIIGRLSRSSISVLINGESGTGKELVAHALHRHSPRAKAPFIA   194
EcHydG     GKSP--AMQHLLSEIALVAPSEATVLIHGDSAR-KELVARGLHASSARSEKPLVT   193
AeHoxA     VRAPGSPLDAVCEVAARVARYDLPVMVLGESGTGKELLARAIHYASPRAARAFVS   219
BjHoxA     VHSTESPMHAVIELGRRAADYDISVLITGESGTGKELLARAIHYGSARANRAFVV   218
RcHupR1    LRTPNSAMTGAIALARQFASFDVPVLLRGEPGSGRAQLARAMHYVSLRSDKPFYE   219
                 .         .     *.. *.    .  .*  .*  * *.

KpNtrC     LNMAAIPKDLIESELFGHEKGAFTGANTVRQGRFEQADGGTLFLDEIGDMPLDVQ   249
EcHydG     LNCAALNESLLESELFGHEKGAF-GADKRREGPFVEADGGT-CLGEIGDISPMMQ   246
AeHoxA     ENCAAVPDNLLESELFGHKRGAFTGAYEDHAGLFQRANGGTIFLDEIGDTSPAFQ   274
BjHoxA     ENCGALPDELLESELFGCKKGAFTGAYQDRIGLFEVADGGTIFLDEIGETSPAFQ   273
RcHupR1    INLAGLPEDLAMIELFGARRGVLPGGVA-KIGLAQKADRGTLFVAGVEAASPALQ   273
           *   .   *   **** .*   *      *    *  **  .  .       *

KpNtrC     TRLLRVLADGQFYRVGGYAPVKVDVRIIAATHQNLELRVQEGKFREDLF--HRLN   302
EcHydG     VRLLRAIQEREVQRVGSNQIISVDVRLIAATHRDLAAEVNAGRFFQDLY--YRLN   299
AeHoxA     VKLLRVLQEGEVRPVGSPRWIPVDVRVIAATHCNLESDVHAGRFREDLY--YRIA   327
BjHoxA     VKLLRVLQESEIRPLGAARCRKVDVRVVAATNRDLEAEVEAGRFRRPVLPPRRIP   328
RcHupR1    LALLRMLADGAITPLGGQETASTNLRLITGAAADLRAMVAEGRFRADLY--YALS   326
             *** . .     .*        .*... .   *   *  *.*   .     .

KpNtrC     VIRVHLPPLRERREDIPRLARHFLQIAARELGVEAKQLHP-ETEMALTRLAWPGN   356
EcHydG     VVAIEVPSLRQRREDIPLLAGHFLQRFA-ERNRRGKRFYA-PGLDLLIHYDWPGN   352
AeHoxA     GVTISMPPLRERSGDLQPIAAKLLEQVAQELARPG-LYFGGDALAAMMAYPWPGN   381
BjHoxA     GAHAGAA----RAPDGHPADCGGRAVGGQEFLQPAELLFARSALEEFGKYHWPGN   379
RcHupR1    AGEIALPPLRARRGDVALLAQSMLAEAAVRHGKQA-LGFDAAALEFLENYDWPGN   380
                     *  *                                    ****

KpNtrC     VRQLENTCRWLTVMAAGQEVLTQDLPSELFETAI-PDNPTQMLPDSWATLLGQWAD   411
EcHydG     IRELENAVERAVVLLTGEYISERELPLGIASTPI-PLGQSQDI-------------   394
AeHoxA     IRELRNEIYRAVALSSGEEIRAQ-LFSRKVLHGQ-PGTVKRGPHVQTFPQSGT---   432
BjHoxA     VRELQNEIQRMAVLADRDELAAPPLLGRRNGKRSAPLPAHGRLNGSAS--------   427
RcHupR1    LRELHNEVTRMLIFAQ-DNVLGAELISRHILQA-APSESGADRSAEEVMTADGT--   432
           .* * *             .    *          *

KpNtrC     RALRSGHQNLLSEAQPEMERTLLTTALRHTQGHKQEAARLLGWGRNTLTRKLKEL   466
EcHydG     ------------QPLVEVEKEVILAALEKTGGNKTEAARQLGITRKTLLAKLSR    436
AeHoxA     ----------LQERLDAIEAVVLKEALLRHRWNKTHAAKELGLSRGGLRQKLLRF   477
BjHoxA     ----------LKDKVEDLEKSVIMNCLERNDGNISRVASELGLSRVGLRNKLSRY   472
RcHupR1    ----------LKDRIELIEMRILRETLTRNRWNKSRAAAELGLSRVGLRAKLDRY   477
                          .*  ..   *          *  **  *  *  **

KpNtrC     GME                469
AeHoxA     GLEEK              482
BjHoxA     DLRKNAKGDAFS       484
RcHupR1    GIEHPAGRVQEEEED    492
```

Fig. 4. Multiple alignment of the *B. japonicum* (Bj) HoxA protein sequence with other transcriptional regulators of the NtrC family using the CLUSTAL program. *Klebsiella pneumoniae* (Kp) NtrC (Drummond et al. 1986); *E. coli* (Ec) HydG (Stoker et al. 1989); *A. eutrophus* (Ae) HoxA (Eberz and Friedrich 1991) and *R. capsulatus* (Rc) $HupR_1$ (Richaud et al. 1991). Perfectly conserved positions are marked with asterisks and a dot represents a conservative substitution (L–V–I–M; K–R; F–Y–W; D–E; Q–N; T–S–A). Gaps introduced for optimal alignment are marked by dashes. The putative ATP-binding site is boxed

pressed with hydrogenase in free-living conditions, but not in nodules (Simpson et al. 1979; Purohit et al. 1982). Secondly, in *A. eutrophus,* it has been shown that expression of the *hoxA* gene itself is transcriptionally regulated by the same environmental conditions as the hydrogenase structural genes (Schwartz et al. 1991). Thirdly, if the HoxX protein were the sensor protein responsible for regulation of free-living hydrogenase activity in *B. japonicum,* a nickel-binding site should be present that can sense changes in nickel concentrations. No such amino acid motif could be detected in the deduced amino acid sequence. The presence of other genes that might be necessary to obtain Hup activity, located downstream of *hoxA,* cannot be ruled out. The function and regulation of *hoxX* and *hoxA* genes are currently under investigation and should provide further insight into the regulation of hydrogenase activity in free-living *B. japonicum.*

Acknowledgements. The authors are grateful to Dr. R. De Mot for help with computer analysis. C.V.S. is a recipient of a fellowship from the Belgian "Nationaal Fonds voor Wetenschappelijk Onderzoek". This work was partly financed by the EEC–STDII program (project N° TS2–0199–C(GDF)).

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1987) Current protocols in molecular biology. John Wiley New York

Chen JC, Mortensen LE (1992) Identification of six open reading frames from a region of the *Azotobacter vinelandii* genome likely involved in dihydrogen metabolism. Biochim Biophys Acta 1131:199–202

Dodd IB, Egan JB (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res 18:5019–5026

Drummond M, Whitty P, Wooton JC (1986) Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae*. EMBO J 5:441–447

Eberz G, Friedrich B (1991) Three *trans*-acting regulatory functions control hydrogenase synthesis in *Alcaligenes eutrophus*. J Bacteriol 173:1845–1854

Friedman AM, Long SR, Brown SE, Buikema WJ, Ausubel FM (1982) Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. Gene 18:289–296

Haugland RA, Cantrell MA, Beaty JS, Hanus FJ, Russell SA, Evans HJ (1984) Characterization of *Rhizobium japonicum* hydrogen uptake genes. J Bacteriol 159:1006–1012

Hohn B, Collins J (1980) A small cosmid for efficient cloning of large DNA fagments. Gene 11:291–298

Kim H, Maier RJ (1990) Transcriptional regulation of hydrogenase synthesis by nickel in *Bradyrhizobium japonicum*. J Biol Chem 265:18729–18732

Kim H, Choonbal Y, Maier RJ (1991) Common *cis*-acting region responsible for transcriptional regulation of *Bradyrhizobium japonicum* hydrogenase by nickel, oxygen and hydrogen. J Bacteriol 173:3993–3999

Lambert GR, Harker AR, Cantrell MA, Hanus FJ, Russell SA, Haugland RA, Evans HJ (1987) Symbiotic expression of cosmid-borne *Bradyrhizobium japonicum* hydrogenase genes. Appl Environ Microbiol 53:422–428

Lutz S, Jacobi A, Schlensog V, Böhm R, Sawers G, Böck A (1991) Molecular characterization of an operon (*hyp*) necessary for the activity of the three hydrogenase isoenzymes in *Escherichia coli*. Mol Microbiol 5:123–135

Maier RJ, Campbell NER, Hanus FJ, Simpson FB, Russell SA, Evans HJ (1978) Expression of hydrogenase activity in free-living *Rhizobium japonicum*. Proc Natl Acad Sci USA 75:3258–3262

Meinkoth J, Wahl G (1984) Hybridization of nucleic acids immobilized on solid supports. Anal Biochem 138:267–284

Ninfa AJ, Bennett RL (1991) Identification of the site of phosphorylation of the bacterial protein kinase/phosphatase NR_{II}. J Biol Chem 266:6888–6893

Purohit K, Becker RR, Evans HJ (1982) D-Ribulose-1,5-bisphosphate carboxylase/oxygenase from chemolithotrophically grown *Rhizobium japonicum* and inhibition by D-4-phosphoerythronate. Biochim Biophys Acta 715:230–239

Ramseier TM, Göttfert M (1991) Codon usage and G+C content in *Bradyrhizobium japonicum* genes are not uniform. Arch Microbiol 156:270–276

Richaud P, Colbeau A, Toussaint B, Vignais PM (1991) Identification and sequence analysis of the $hupR_1$ gene, which encodes a response regulator of the NtrC family required for hydrogenase expression in *Rhodobacter capsulatus*. J Bacteriol 173:5928–5932

Sayavedra-Soto LA, Powell GK, Evans HJ, Morris RO (1988) Nucleotide sequence of the genetic loci encoding subunits of *Bradyrhizobium japonicum* uptake hydrogenase. Proc Natl Acad Sci USA 85:8395–8399

Schwartz E, Tran-Betcke A, Eberz G, Gewinner P, Friedrich B (1991) Regulation of the hydrogenase genes of *Alcaligenes eutrophus*. Proceedings of Third International Conference on Molecular Biology of Hydrogenases, Portugal

Simpson FB, Maier RJ, Evans HJ (1979) Hydrogen-stimulated CO_2 fixation and coordinate induction of hydrogenase and ribulosebiphosphate carboxylase in a H_2-uptake positive strain of *Rhizobium japonicum*. Arch Microbiol 123:1–8

Stoker K, Reijnders WNM, Oltman LF, Stouthamer A.H. (1989) Initial cloning and sequencing of *hydHG*, an operon homologous to *ntrBC* and regulating the labile hydrogenase activity in *Escherichia coli* K-12. J Bacteriol 171:4448–4456

Stults LW, Sray WA, Maier RJ (1986) Regulation of hydrogenase biosynthesis by nickel in *Bradyrhizobium japonicum*. Arch Microbiol 146:280–283