Mol Gen Genet (1988) 214:97–107

MGG

© Springer-Verlag 1988

Cloning and sequencing of the genes encoding the large and the small subunits of the H_2 uptake hydrogenase (*hup*) of *Rhodobacter capsulatus*

Michèle Leclerc, Annette Colbeau, Béatrice Cauvin, and Paulette M. Vignais

Biochimie Microbienne, Département de Recherche Fondamentale, Centre d'Etudes Nucléaires, 85 X, F-38041 Grenoble Cedex, France

Summary. The structural genes (*hup*) of the H_2 uptake hydrogenase of *Rhodobacter capsulatus* were isolated from a cosmid gene library of *R. capsulatus* DNA by hybridization with the structural genes of the H_2 uptake hydrogenase of *Bradyrhizobium japonicum*. The *R. capsulatus* genes were localized on a 3.5 kb *Hin*dIII fragment. The fragment, cloned onto plasmid pAC76, restored hydrogenase activity and autotrophic growth of the *R. capsulatus* mutant JP91, deficient in hydrogenase activity (Hup$^-$). The nucleotide sequence, determined by the dideoxy chain termination method, revealed the presence of two open reading frames. The gene encoding the large subunit of hydrogenase (*hupL*) was identified from the size of its protein product (68108 dalton) and by alignment with the NH_2 amino acid protein sequence determined by Edman degradation. Upstream and separated from the large subunit by only three nucleotides was a gene encoding a 34256 dalton polypeptide. Its amino acid sequence showed 80% identity with the small subunit of the hydrogenase of *B. japonicum*. The gene was identified as the structural gene of the small subunit of *R. capsulatus* hydrogenase (*hupS*). The *R. capsulatus* hydrogenase also showed homology, but to a lesser extent, with the hydrogenase of *Desulfovibrio baculatus* and *D. gigas*. In the *R. capsulatus* hydrogenase the Cys residues (13 in the small subunit and 12 in the large subunit) were not arranged in the typical configuration found in [4Fe–4S] ferredoxins.

Key words: Hydrogenase structural genes – *Rhodobacter capsulatus* – *hupS, hupL* – *hup* sequence – (NiFe)hydrogenase

Introduction

The hydrogenase enzymes catalyse the "activation" or reversible oxidation of molecular hydrogen according to the reaction $H_2 \rightleftarrows 2H^+ + 2e^-$. They play a central role in the chemolithotrophic growth of a variety of taxonomically diverse bacteria, and function as H_2 uptake hydrogenases to provide cells with reducing power needed for CO_2 fixation, or to supply H_2 electrons in H_2 dependent oxidative phosphorylation. H_2 uptake hydrogenases have been identified and characterized in lithoautotrophic bacteria such as *Alcaligenes*, in nitrogen fixing rhizobia and *Azotobacter* cells, in anaerobic sulfate reducers, in photosynthetic bacteria, and in facultative aerobic *Escherichia coli* cells. H_2 uptake hydrogenases are mostly membrane bound NiFe sulfur proteins. Surveys of the literature concerning those hydrogenases have been published recently (Vignais et al. 1985; Friedrich et al. 1986; Gogotov 1986).

The hydrogenase of the purple nonsulfur photosynthetic bacterium *Rhodobacter capsulatus* (formerly *Rhodopseudomonas capsulata*, Imhoff et al. 1984) is the best studied of the hydrogenases from photosynthetic bacteria. Colbeau and colleagues isolated a hydrogenase protein from membranes of *R. capsulatus* strain B10. The 65 ± 2 kDa protein contained nickel and could transfer electrons from H_2 to artificial electron acceptors such as benzyl-viologen and methylene blue (Colbeau and Vignais 1981, 1983; Colbeau et al. 1983). Recently, Seefeldt et al. (1987) showed that the *R. capsulatus* hydrogenase from strain B100 is a heterodimer with subunits of 67 kDa and 31 kDa. Both the large and small subunits of the *R. capsulatus* hydrogenase acted in a manner similar to the membrane-bound dimeric hydrogenases purified from *Azotobacter vinelandii, Alcaligenes latus* and *A. eutrophus* H16, and cross-reacted with antibodies raised against the *Bradyrhizobium japonicum* hydrogenase. As a result, it is possible that these immunologically related membrane-bound hydrogenases also share common physical properties (Seefeldt et al. 1987).

These findings encouraged us to use a DNA fragment from *B. japonicum*, which carried the structural genes of the *B. japonicum* hydrogenase, to probe our cosmid library of *R. capsulatus* genomic DNA and isolate homologous DNA sequences. The *B. japonicum* probe used was a *hup* gene contained on the 5.9 kb *Hin*dIII fragment of cosmid pHUI (Cantrell et al. 1983) and pHU52 (which is pHUI with an extra 5.5 kb *Eco*RI fragment; Lambert et al. 1985). The *B. japonicum hup* gene product was expressed in *E. coli* and cross-reacted with antibodies against the large subunit (60 kDa) of the hydrogenase from *B. japonicum* (Zuber et al. 1986). The same probe (pHUI) was used successfully by Tibelius et al. (1987) to isolate *hup* genes from *A. chroococcum*.

We are now presenting data on the cloning and sequencing of the structural genes of the (NiFe)hydrogenase from *R. capsulatus*. The sequence of the cloned DNA from *R. capsulatus* revealed the presence of two open reading frames. One was aligned with the NH_2-terminal amino acids of the large subunit of hydrogenase, and the other one encoded a 34 kDa protein which exhibited homology with the small subunit of three other sequenced hydrogenases, the (NiFeSe)hydrogenase from *Desulfovibrio baculatus* (Menon et al. 1987), the (NiFe)hydrogenase from *Desulfovibrio gigas* (Li et al. 1987) and the (NiFe)hydrogenase from *B. japonicum* (Sayavedra et al. 1988).

Offprint requests to: P.M. Vignais

Materials and methods

Bacterial strains and cultures. Bacterial strains and plasmids are listed in Table 1. *E. coli* strains were grown aerobically at 37° C in liquid Luria Bertani (LB) medium (Maniatis et al. 1982). Solid LB medium contained 1.5% agar, unless specified otherwise. *R. capsulatus* strains were grown routinely in the light in a minimal salts (RCV) medium (Weaver et al. 1975; Hillmer and Gest 1977) supplemented with DL-malate (30 mM) and L-glutamate (7 mM) as C and N sources, respectively, or in the dark in YPS medium which contained 0.3% Bacto-peptone, 0.3% yeast extract, 2 mM $CaCl_2$, and 2 mM $MgSO_4$ (Weaver et al. 1975). Solid medium was prepared with 1.2% agar. Petri plates with solid medium were incubated at 30° C in white light (about 2500 lux) anaerobically in GasPak jars (BBL Microbiology Systems, Becton Dickinson and Co., Cockeysville, MD 21030, USA) under an atmosphere of H_2, N_2 and CO_2. For autotrophic cell growth, no C compound was added to RCV medium, and the N source was 7 mM ammonium sulfate. Antibiotics, added to growth media as indicated, were kanamycin (Km), 20 µg/ml; rifampicin (Rif) 40 µg/ml; ampicillin (Ap) 100 µg/ml; and tetracycline (Tc). The Tc concentration in growth media for *E. coli* was 10 µg/ml and for *R. capsulatus* the concentration was 0.5 µg/ml.

Bacterial matings. Bacterial matings were made with the triparental cross system of Ditta et al. (1980) using pRK2013 in *E. coli* HB101 as the mobilizing plasmid. Both *E. coli* strains and *R. capsulatus* were grown aerobically overnight at 30° C in the dark in liquid YPS medium. *E. coli* cultures were diluted 10-fold and *R. capsulatus* was diluted by one half with YPS medium. Incubation of the three cultures was continued for 2 additional hours to allow cells to reach exponential growth. Finally 25 µl of *E. coli* (pRK2013) culture, 25 µl of the culture of *E. coli* carrying the recombinant plasmid, and 50 µl of *R. capsulatus* mutant strain were mixed and spread onto solid YPS medium. After 24 h of aerobic growth at 30° C in the dark, the bacterial layer was scraped off the solid medium, resuspended in 1 ml of salts solution (RCV without phosphate, C or N compounds) and used to inoculate liquid or solid medium. After suitable dilution, the cell suspension was grown either on solid YPS medium containing Rif (40 µg/ml) and Tc (0.5 µg/ml) to estimate the number of transconjugants, or on RCV-agar plates without organic carbon to select for Hup^+ transconjugants. Autotrophic growth of Hup^+ transconjugants in liquid cultures was carried out in 3 ml RCV medium devoid of organic carbon. The inoculated medium in cotton-plugged tubes was incubated under an atmosphere of H_2, N_2, CO_2, in the light, at 30° C, in GasPak jars. Absorbance of the cultures was measured at 660 nm in samples withdrawn at time intervals over a period of one week.

Isolation of DNA. Total DNA was prepared by the method of Marmur (1961). Plasmid DNA was obtained by the alkaline method of Birnboim and Doly (1979). It was purified on cesium chloride gradient for large scale preparations (Ish-Horowicz and Burke 1981), or by successive extractions with a phenol-chloroform solution (1:1, v/v) and chloroform for small scale preparations using only 3 ml of culture fluid. Endonuclease DNA digests were analyzed by electrophoresis in agarose gels (0.7%). Electrophoresis was performed overnight in Tris-borate-EDTA buffer, pH 8.3, with a voltage input of 1 volt/cm.

Table 1. Bacterial strains and plasmids

Strain or plasmid	Genotype or phenotype	Comments and references
Escherichia coli		
HB101	F^- *recA proA ara xyl mtl lacY galK rpsL supE hsdS*	Boyer and Roulland-Dussoix (1969)
TB1	Δ(*lac-pro*), *strA*, *ara*, *thi*, ϕ80, *lacZ* ΔM15, *hsdR*	Gibco BRL (host for pUC plasmids)
JM83(pUC18 or pUC19)	*ara*, Δ(*lac-proAB*), *thi*, *strA* ϕ80, *lacZ* ΔM15, Ap^r	Gibco BRL (source of pUC18 or pUC19)
Rhodobacter capsulatus		
B10	Wild type, Hup^+	Marrs (1974)
JP91	Hup^-, Km^s, obtained from JP57 (Hup^-, Km^r)	Willison et al. (1987)
JP91A	Hup^-, Rif^r, obtained from JP91	
Plasmids		
pLAFR1	Tc^r, cosmid vector	Friedman et al. (1982)
pRK2013	Km^r, helper plasmid	Ditta et al. (1980)
pRK290	Tc^r, plasmid vector	Ditta et al. (1980)
pRK292	Tc^r, plasmid vector	Ditta et al. (1980)
pHC79	Ap^r, Tc^r, cosmid vector	Hohn and Collins (1980)
pAC70	Tc^r (pLAFR1 20 kb *Eco*RI insert)	This work
pAC76	Tc^r (pRK292 3.5 kb *Hin*dIII insert)	This work
pBC1	Ap^r (pHC79, 40 kb *Bam*HI insert)	This work
pUC18	pBR322 derivative	Yanisch-Perron et al. (1985)
pUC19	pBR322 derivative	Yanisch-Perron et al. (1985)

DNA cloning procedure. DNA was hydrolysed by restriction endonucleases as specified by the manufacturer (Boehringer Mannheim, W. Germany). Routinely, the ligation of a DNA fragment with a linearized and dephosphorylated plasmid vector was made under the following conditions: DNA fragments to DNA vector in a ratio of about 4:1 with a total DNA concentration of about 0.2 mg/ml, T4 DNA ligase (0.5 unit per µg DNA), total volume 20 µl. Incubation time was 16 h at 15° C. After ligation, the hybrid DNA was introduced into *E. coli* by transformation. Competent *E. coli* cells were prepared and stored at −80° C as described by Morrison (1979). After transformation with the ligation mixture (200–400 ng DNA per 0.15 ml competent cells), transformed *E. coli* cells were selected on LB plates in the presence of tetracycline (Tc; 10 µg/ml).

Construction of gene banks of R. capsulatus. Two genomic DNA libraries of *R. capsulatus* strain B10 were constructed. The first one using the cosmid vector pLAFR1 (Friedman

et al. 1982) and 20 kb *Eco*RI inserts was described by Colbeau et al. (1986). The second gene bank was constructed with 40 kb *Bam*HI fragments. The 40 kb fragments were obtained by partial digestion of total DNA with the restriction endonuclease *Bam*HI (0.1 units/μg DNA, 1 h at 37° C) and purified on a 10%–40% sucrose gradient (Maniatis et al. 1982). They were cloned into the 6.4 kb cosmid vector pHC79 (Hohn and Collins 1980). Ligation and in vitro packaging were performed as described (Colbeau et al. 1986).

After selection on LB plates containing 100 μg ampicillin (Ap)/ml, 5000 clones per μg DNA were obtained from which 80% were Tc sensitive. Fifteen hundred Ap resistant, Tc sensitive cells were stored individually at −80° C.

DNA-DNA hybridization. For hybridization, DNA fragments were labeled with α-[^{32}P]dCTP (ca. 110 TBq/mmol, Amersham, UK) by nick-translation using a kit according to the instructions by the manufacturer (Boehringer, Mannheim).

Colony hybridization. The gene bank was grown overnight on LB plates with 10 μg Tc/ml. Each plate contained about 2–3 × 10^3 colonies. A replica was made on Hybond nylon membrane (Amersham, UK) which was placed on LB/Tc plates and incubated at 37° C overnight. Colonies were lysed by placing the nylon membrane onto Whatman 3MM paper wetted with 0.5 N NaOH, and incubated for 3 min. The nylon membrane with alkaline-solubilized cells was neutralized by blotting onto Whatman filter paper wetted with 1.0 M Tris-HCl buffer, pH 8, and then washed with chloroform. DNA was fixed for 5 min under UV light.

Filter hybridization. Unlabeled DNA was separated in an agarose gel (0.7%). After DNA denaturation with a solution of 0.5 M NaOH and 1.0 M NaCl, the agarose gel was washed several times in 25 mM phosphate buffer, pH 6.5. DNA was transferred to a Hybond nylon filter by electroblotting in an IBI apparatus (Genofit, Geneva, Switzerland), operating at 30 V with 25 mM phosphate buffer, pH 6.5. After 3 h, the transferred DNA was fixed onto the membrane by exposure to UV light for 5 min. The DNA-loaded membrane was prehybridized for 1 h with salmon sperm DNA, followed by hybridization for 16 h at 42° C as described by Maniatis et al. (1982) under high or low stringency conditions with 50% or 35% (v/v) formamide, respectively. The ^{32}P-labeled DNA probe was denatured for 15 min at 95° C, cooled for 2 min in an ice ethanol bath, and then added to the hybridization mixture (2 × 10^6 cpm/ml/7 cm^2 of Hybond membrane). After hybridization, the Hybond membranes were washed 3 times with a solution containing 0.30 M NaCl, 20 mM NaH_2PO_4, and 2 mM EDTA (2 SSPE) and 0.1% sodium dodecyl sulfate (SDS), at 45° C, before autoradiography (Maniatis et al. 1982). Autoradiograms were prepared by overnight exposure at −80° C to Fuji Nif type X-Ray film.

DNA sequencing. For nonrandom sequence analysis, DNA fragments were prepared as follows. The 3.5 kb *Hin*dIII fragment of pAC70 was cloned into pBR322 and cleaved either at the unique *Cla*I site of pBR322 or at the *Stu*I site of the insert. It was shortened by progressive enzymatic digestion with the exonuclease *Bal*31 (Ponzc et al. 1982) to generate a series of subfragments sequentially shortened by about 150–200 bp. After treatment with Klenow enzyme to generate a blunt cloning end, fragments were subcloned at the *Hin*dIII and *Sma*I sites of pUC18 and pUC19 (Yanisch-Perron et al. 1985). Transformed TB1 colonies were grown on LB agar plates containing Ap (100 μg/ml) and 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) (2%). The presence of *hup* DNA was monitored by colony hybridization with the 3.5 kb *Hin*dIII fragment of pAC76, ^{32}P-labeled by nick translation. Plasmids of labeled colonies were denatured and used for sequencing by the dideoxy chain termination procedure (Sanger et al. 1977; Hattori et al. 1986) after purification by ultracentrifugation on cesium chloride.

Specific oligonucleotide primers were synthesized using phosphoramidite chemistry with an Applied Biosystems synthesizer (model 380A). Ten nanograms of specific primer was annealed to 2 μg of template DNA (Strauss et al. 1986). Sequencing reactions were performed with α-[^{35}S]dATP (37 Tbq/mmol) using the Amersham sequencing kit. The procedure of B. Roe (described in Focus vol. 9(3) 1987) was used for sequencing DNA fragments extending over 200 bases. Deoxy-7-deazaguanosine (Boehringer Mannheim) was used in place of dGTP to reduce compression of bands on sequencing gels (Mizusawa et al. 1986). Samples were prepared on 5% acrylamide (Bio-Rad), 8.3 M urea (Fluka, Buchs, Switzerland) sequencing gels (0.4 mm thickness) with Tris-borate-EDTA buffer (Maniatis et al. 1982) using a Sequi-Gen^{TM} nucleic acid sequencing cell (38 × 50 cm) from Bio-Rad (Richmond, Calif, USA). After electrophoresis at 1850 V, gels were rinsed in 10% acetic acid 10% methanol solution to remove urea and then dried on a slab-gel dryer, model SE 1160 (Hoefer Scientific Instruments, San Francisco, Calif, USA). Autoradiograms were prepared by exposure for 1 or 2 days to Kodak "no screen" DEF-5 X-Ray film.

Sequence data analysis (alignment of the DNA sequences, determination of the open reading frames, translation in amino acids, codon usage and comparison with other Rhodospirillaceae, nucleotide sequence comparisons, prediction of secondary structures and hydrophobicity patterns) was performed on a Prime computer connected to B.I.S.A.N.C.E. (Base Informatique sur les Séquences d'Acides Nucléiques pour les Chercheurs Européens; centre Inter Universitaire de Traitement Informatique, Paris, France) where sequence data banks and data processing programs are accessible.

Hydrogenase assays. Hydrogenase was assayed spectrophotometrically with methylene blue (whole cells) or benzylviologen (membranes and purified protein) as electron acceptor (Colbeau and Vignais 1981). Protein concentrations were determined by the method of Bradford (1976) with bovine serum albumin as standard.

Purification of hydrogenase and NH_2 terminal sequence determination. Hydrogenase was prepared from *R. capsulatus* cells grown anaerobically in the light. Whole cells from 5 l culture fluid were disrupted with a French pressure cell. Hydrogenase in the cell-free extract was precipitated with polyethylene glycol 6000 and purified by DEAE cellulose liquid column chromatography in the presence of 0.2% Triton X-100 (Colbeau et al. 1983). The large hydrogenase subunit was isolated on preparative SDS acrylamide gels (Laemmli 1970), then visualised and eluted by the method

of Hager and Burgess (1980). Protein samples were concentrated with an immersible CX-10 filtration unit (Millipore Bedford, Mass, USA) and salts and SDS were removed by molecular sieving on a 1 ml Biogel P10 column. Biogel P10 was equilibrated in 70% (v/v) formic acid. The hydrogenase protein in 70% formic acid was placed in siliconized Eppendorf tubes and dried in a Speed-Vac evaporator. Amino terminal sequence analysis was performed by Edman degradation with 1 nmol of protein using an Applied Biosystems (Calif. USA) protein sequenator (model 470 A) at the Laboratoire de Chimie Bactérienne (CNRS, Marseille, France).

Results

A DNA sequence homologous to the hup *structural gene from* B. japonicum *is present in* R. capsulatus *genomic DNA*

The 5.9 kb *Hin*dIII fragment from pHUI (Lambert et al. 1985; Zuber et al. 1986) was used to probe homologous DNA sequences in *R. capsulatus* genomic DNA. Indeed, one *Hin*dIII fragment of 3.5 kb and three *Eco*RI fragments of 1.4, 1.0 and 0.6 kb hybridized with the *B. japonicum hup* structural gene under high stringency conditions (Fig. 1A).

Cloning the structural hup *gene from* R. capsulatus *by colony hybridization*

The 2200 clones of the *R. capsulatus* B10 gene bank (Colbeau et al. 1986) were grown on Hybond nylon film. After denaturation with NaOH, the colonies were screened by hybridization with the 5.9 kb *Hin*dIII fragment from pHUI, ^{32}P-labeled by nick translation. Nine colonies had a radioactivity level higher than the noise level. All the labeled colonies were picked, purified, and their cosmids analysed by digestion with *Eco*RI and *Hin*dIII restriction endonucleases. The restriction pattern was the same for all of the clones tested. Clone AC70 was chosen for further study. The restriction endonuclease map of the insert DNA of AC70 is depicted in Fig. 2.

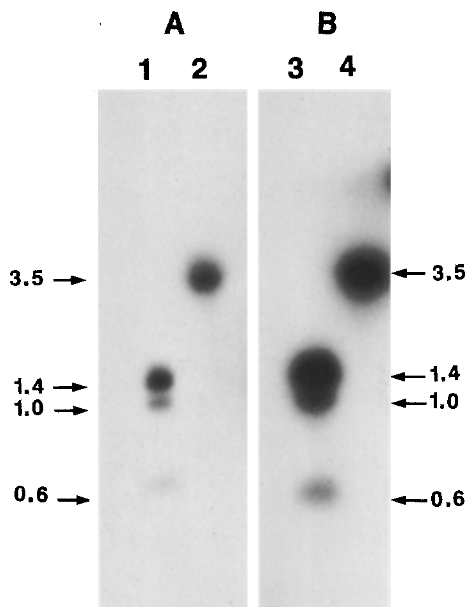

Fig. 1 A, B. Hybridization of *Rhodobacter capsulatus* DNA with the hydrogenase structural gene from *Bradyrhizobium japonicum*. Autoradiographs obtained with a ^{32}P-labeled 5.9 kb *Hin*dIII DNA fragment from pHUI used as a probe (sizes indicated in kb). **A** *Eco*RI and *Hin*dIII digested genomic DNA. *Lane 1, Eco*RI digest; lane 2, *Hin*dIII digest. **B** *Eco*RI and *Hin*dIII digested pAC70. Lane 3, *Eco*RI digest; lane 4, *Hin*dIII digest

In separate experiments, it was determined by DNA hybridization which of the restriction fragments from the

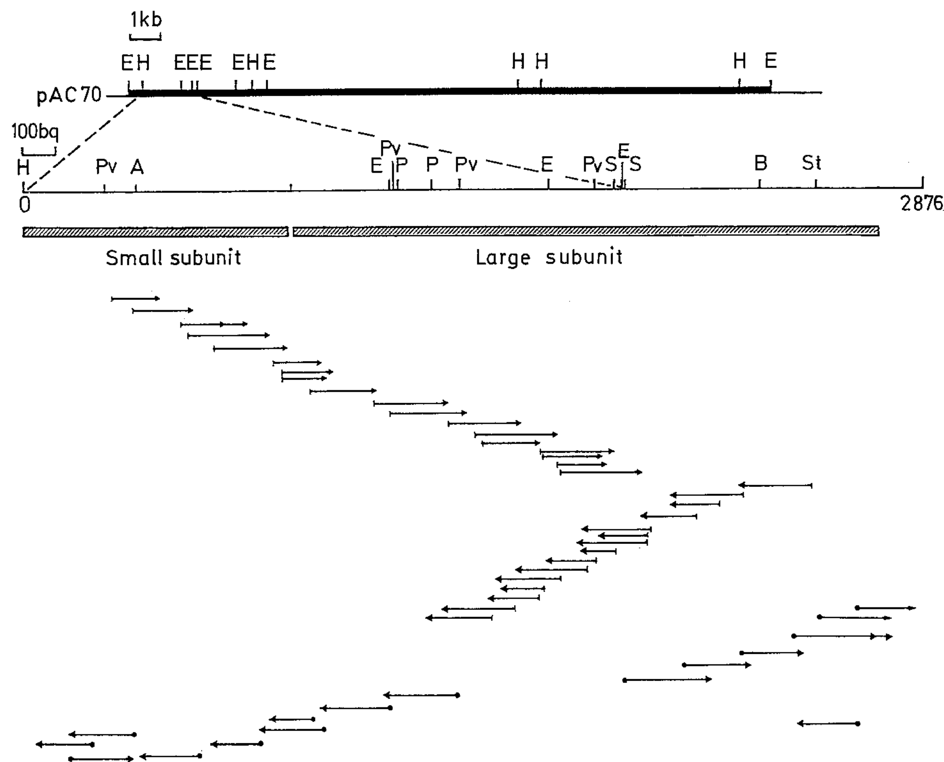

Fig. 2. Restriction map of pAC70 and nucleotide sequencing strategy. The restriction sites (A, *Ava*I; B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I; St, *Stu*I) of the *hup* DNA region are shown. *Arrows* indicate the direction and extent of sequencing: ↦ clones generated by *Bal*31; ↔ regions obtained by specific primer directed sequencing. The region encoding the small and large subunit is shown by *shadowed bars* in the upper part of the figure; the direction of transcription is from left to right

insert DNA of pAC70 was homologous with the *hup* gene from *B. japonicum*. Plasmid was hydrolysed by *Hin*dIII and by *Eco*RI. The digested fragments were separated in an agarose gel, and then transferred to a Hybond nylon membrane to test for hybridization with the ^{32}P-labeled 5.9 kb *Hin*dIII fragment from pHUI (Fig. 1B). A 3.5 kb *Hin*dIII fragment and three *Eco*RI fragments of 1.4, 1.0 and 0.6 kb showed strong homology to the *B. japonicum* probe. It is important to note that fragments of the same size had been detected by hybridizing the same probe with *R. capsulatus* DNA isolated from the genome or from the recombinant plasmid pAC70 and digested by *Hin*dIII and *Eco*RI (Fig. 1). This suggested that the DNA sequence of the insert in pAC70 was similar to native DNA, and that no major rearrangement of the DNA sequence had occurred during the construction of the gene bank. The 3.5 kb *Hin*dIII fragment from pAC70, which strongly hybridized to the *B. japonicum* probe, was subcloned into the *Hin*dIII unique site of plasmid pRK292 to produce plasmid pAC76.

Identification of hup *DNA by complementation of the Hup$^-$ mutant*

Among the Hup$^-$ derivatives of *R. capsulatus* B10 tested in complementation experiments, one mutant, JP91, had its hydrogenase activity restored after conjugation with *E. coli* carrying either plasmid pAC70 or pAC76 (Fig. 3). Mutant JP91 is a Km^s revertant of strain JP57 (Km^r), reported earlier to be unable to grow autotrophically and to contain undetectable levels of hydrogenase activity (Willison et al. 1987). Restoration of hydrogenase activity in Hup$^+$ transconjugants of JP91 was also accompanied by restored synthesis of hydrogenase protein. Hydrogenase synthesis was monitored by rocket immunoelectrophoresis (data not shown) using antibodies raised against *R. capsulatus* hydrogenase (Colbeau et al. 1983).

These data confirmed that the 3.5 kb *Hin*dIII fragment contains a *hup* gene. However, the *hup* gene was probably not present in its entirety in the 3.5 kb fragment since less than 1% of the transconjugants were Hup$^+$ after mating of JP91 with strains AC70 or AC76. Apparently, normal *hup* gene function in JP91 was recovered by marker rescue.

Isolation of a second clone containing the 3.5 kb HindIII *fragment with its upstream region*

To obtain the complete *hup* gene a second gene bank with 40 kb *Bam*HI inserts was prepared and screened by colony hybridization with the ^{32}P-labeled 3.5 kb *Hin*dIII fragment from *R. capsulatus*. From the 16 colonies which hybridized to the probe, 1 colony, BC1, was selected. The insert of pBC1 digested with *Bam*HI gave several restricted fragments, 2 of which, 9.6 kb and 8.0 kb in size, hybridized with the radioactive probe. The *hup* gene together with its upstream DNA region of interest was localized in the 9.6 kb *Bam*HI fragment as determined by hybridization with a ^{32}P-labeled 0.38 kb *Hin*dIII-*Ava*I subfragment (shown on the left hand side in Fig. 2).

Analysis of the NH_2 terminal sequence of the large subunit of R. capsulatus *hydrogenase*

To identify unambiguously that the gene isolated was the *hup* structural gene encoding the large subunit of *R. capsulatus* hydrogenase, the NH_2 terminal amino acid sequence of the purified hydrogenase protein was determined (see Materials and methods), and found to be Thr – Thr – Gln – Thr – Pro – Asn – Gly – Phe – Thr – Leu – Asp – Asn – Ala – Gly – Lys – Arg – Ile – Val – Val.

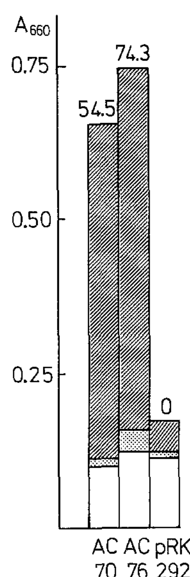

Fig. 3. Restoration of autotrophic growth and hydrogenase activity in the Hup$^-$ mutant JP91 by recombinant cosmids pAC70 and pAC76. *Rhodobacter capsulatus* strain JP91 was first mated on agar gel with *Escherichia coli* strains AC70, AC76 or HB101(pRK292). Transconjugants were transferred to liquid culture and grown anaerobically under light for 3 days □, 5 days ▩ and 8 days ▨. The hydrogenase activity expressed in μmol methylene blue/h per mg protein was determined after 8 days of growth (on top of the *bars*)

Sequencing strategy

The 3.5 kb *Hin*dIII fragment of pAC76 was digested with restriction endonucleases. Hybridization of the resulting restriction fragments with the *B. japonicum* probe indicated that the homologous sequence was located in the 2.55 kb *Hin*dIII-*Stu*I fragment of pAC76 (data not shown). This 2.55 kb fragment was then sequenced by non-random sequence analysis performed on 2 sets of deletion clones as described in Materials and methods. The sequencing strategy and restriction map of the sequenced fragment on plasmid pAC76 are shown in Fig. 2. The region upstream of the *Hin*dIII site, carried on a 9.6 kb *Bam*HI fragment from pBC1 subcloned into pUC18, was sequenced twice using a synthetic 17-mer oligonucleotide which hybridized to a sequence located 40 nucleotides downstream from the *Hin*dIII site.

Sequence analysis

The nucleotide and derived amino acid sequences for the *hup* DNA from *R. capsulatus* are presented in Fig. 4. Two open reading frames (ORF1 and ORF2) could be identified. The NH_2 terminal protein sequence derived from the ORF2, after the translational initiator (ATG) at position 1081, was identical to the NH_2 terminal amino acid sequence of the large subunit of *R. capsulatus* hydrogenase determined by Edman degradation. ORF2 was therefore

```
TTG TCG GAC ATC GAA ACT TTC TAT GAC GTG ATG CGG CGT CAG GGG ATC ACC CGG CGC AGC TTC ATG AAG TCT GTT CGC TCA    81
Leu Ser Asp Ile Glu Thr Phe Tyr Asp Val Met Arg Arg Gln Gly Ile Thr Arg Arg Ser Phe Met Lys Ser Val Arg Ser
                                                      1
CCG CAG CAC GTT TTG GGG CTT GGG CCC TCC TTC GTG CCG AAA ATC GGC GAA GCC ATG GAA ACG AAG CCG CGC ACC CCG GTC   162
Pro Gln His Val Leu Gly Leu Gly Pro Ser Phe Val Pro Lys Ile Gly Glu Ala Met Glu Thr Lys Pro Arg Thr Pro Val     9

GTC TGG GTG CAT GGG CTT GAA TGC ACC TGC TGT TCG GAA AGC TTC ATC CGC TCG GCG CAT CCG CTG GCC AAG GAT GTC GTG   243
Val Trp Val His Gly Leu Glu Cys Thr Cys Cys Ser Glu Ser Phe Ile Arg Ser Ala His Pro Leu Ala Lys Asp Val Val    36

CTC TCG ATG ATC TCG CTC GAT TAC GAC GAC ACG CTG ATG GCC GCC GCC GGT CAC GCC GCC GAA GCG GCC TTC GAG GAA ACC   324
Leu Ser Met Ile Ser Leu Asp Tyr Asp Asp Thr Leu Met Ala Ala Ala Gly His Ala Ala Glu Ala Ala Phe Glu Glu Thr    63

ATC GCC AAA TAC AAG GGC AAC TAC ATC CTG GCC GTC GAG GGC AAC CCG CCG CTC AAC GAA GAC GGG ATG TTC TGC ATC ACC   405
Ile Ala Lys Tyr Lys Gly Asn Tyr Ile Leu Ala Val Glu Gly Asn Pro Pro Leu Asn Glu Asp Gly Met Phe Cys Ile Thr    90

GGC GGC AAG CCC TTT GTC GAG AAG CTG CGC CAC GCC GCC GAG GGC GCC AAG GCG ATC ATC AGC TGG GGG GCC TGT GCG TCT   486
Gly Gly Lys Pro Phe Val Glu Lys Leu Arg His Ala Ala Glu Gly Ala Lys Ala Ile Ile Ser Trp Gly Ala Cys Ala Ser   117

TAT GGC TGC GTG CAG GCC GCG GCG CCG AAC CCC ACC CAG GCG ACG CCG GTG CAC AAG GTG ATC ACC GAC AAG CCG ATC ATC   567
Tyr Gly Cys Val Gln Ala Ala Ala Pro Asn Pro Thr Gln Ala Thr Pro Val His Lys Val Ile Thr Asp Lys Pro Ile Ile   144

AAG GTC CCG GGC TGC CCG CCG ATC GCC GAG GTG ATG ACC GGC GTC ATC ACC TAC ATG CTG ACC TTC GAC CGG ATG CCG GAA   648
Lys Val Pro Gly Cys Pro Pro Ile Ala Glu Val Met Thr Gly Val Ile Thr Tyr Met Leu Thr Phe Asp Arg Met Pro Glu   171

CTG GAC CGT CAG GGC CGC CCG GCG ATG TTC TAC AGC CAG CGC ATC CAC GAC AAA TGC TAC CGC CGC CCG CAT TTC GAC GCC   729
Leu Asp Arg Gln Gly Arg Pro Ala Met Phe Tyr Ser Gln Arg Ile His Asp Lys Cys Tyr Arg Arg Pro His Phe Asp Ala   198

GGG CAA TTC GTC GAA CAC TGG GAC GAC GAA AAC GCG CGC AAG GGC TAT TGC CTC TAC AAG ATG GGC TGC AAG GGC CCG ACC   810
Gly Gln Phe Val Glu His Trp Asp Asp Glu Asn Ala Arg Lys Gly Tyr Cys Leu Tyr Lys Met Gly Cys Lys Gly Pro Thr   225

ACC TAC AAC GCC TGT TCG ACC GTG CCG CTG GAA CGG CGG CGT CAC TTC CCG ATC CAG TCC GGT CAC GGC TGC ATC GGC TGT   891
Thr Tyr Asn Ala Cys Ser Thr Val Pro Leu Glu Arg Arg Arg His Phe Pro Ile Gln Ser Gly His Gly Cys Ile Gly Cys   252

TCC GAG GAC GGG TTC TGG GAT CAG GGC AGC TTC TAT GAC CGG CTG ACC ACG ATC AAG CAA TTC GGC ATC GAG GCC ACG GCC   972
Ser Glu Asp Gly Phe Trp Asp Gln Gly Ser Phe Tyr Asp Arg Leu Thr Thr Ile Lys Gln Phe Gly Ile Glu Ala Thr Ala   279

GAC CAG ATC GGC TGG ACC GCC ACC GGC CTT GTG GGC GCC GCC GTC GCC GCC CAT GCC GCG GTC TCC GTG CTC AAA CGC GCG  1053
Asp Gln Ile Gly Trp Thr Ala Thr Gly Leu Val Gly Ala Ala Val Ala Ala His Ala Ala Val Ser Val Leu Lys Arg Ala   306
                                        1
CAG AAA AAG AAC GAG GAG GCG taa gcc ATG ACG ACC CAA ACG CCG AAC GGC TTC ACC CTC GAC AAC GCC GGC AAG CGC ATC  1134
Gln Lys Lys Asn Glu Glu Ala ***     Met Thr Thr Gln Thr Pro Asn Gly Phe Thr Leu Asp Asn Ala Gly Lys Arg Ile    17

GTC GTC GAT CCC GTC ACC CGG ATC GAA GGC CAC ATG CGC TGC GAA GTG AAC GTC AAC GAT CAG GGC ATC ATC ACC AAC GCC  1215
Val Val Asp Pro Val Thr Arg Ile Glu Gly His Met Arg Cys Glu Val Asn Val Asn Asp Gln Gly Ile Ile Thr Asn Ala    44

GTC TCG ACC GGG ACG ATG TGG CGC GGC CTT GAA GTG ATC CTG AAG GGC CGC GAT CCG CGC GAC GCC TGG GCC TTC ACC GAA  1296
Val Ser Thr Gly Thr Met Trp Arg Gly Leu Glu Val Ile Leu Lys Gly Arg Asp Pro Arg Asp Ala Trp Ala Phe Thr Glu    71

CGG ATC TGC GGT GTC TGC ACC GGC ACC CAT GCG CTG ACC TCG GTC CGC GCG GTC GAA AGC GCG CTG GGG ATC ACC ATC CCC  1377
Arg Ile Cys Gly Val Cys Thr Gly Thr His Ala Leu Thr Ser Val Arg Ala Val Glu Ser Ala Leu Gly Ile Thr Ile Pro    98

GAC AAT GCG AAT TCG ATC CGC AAC ATG ATG CAG CTG AAC CTG CAG ATC CAC GAC CAT ATC GTG CAT TTC TAC CAC CTG CAT  1458
Asp Asn Ala Asn Ser Ile Arg Asn Met Met Gln Leu Asn Leu Gln Ile His Asp His Ile Val His Phe Tyr His Leu His   125

GCG CTG GAT TGG GTG AAC CCG GTC AAT GCG CTG CGC GCC GAT CCG AAG GCG ACC TCG GAA CTG CAG CAG ATG GTT TCG CCC  1539
Ala Leu Asp Trp Val Asn Pro Val Asn Ala Leu Arg Ala Asp Pro Lys Ala Thr Ser Glu Leu Gln Gln Met Val Ser Pro   152

AGC CAT CCG CTG TCG TCC CCC GGC TAT TTC CGC GAC GTG CAG AAC CGG CTG AAG AAA TTC GTC GAA TCC GGG CAG CTG GGG  1620
Ser His Pro Leu Ser Ser Pro Gly Tyr Phe Arg Asp Val Gln Asn Arg Leu Lys Lys Phe Val Glu Ser Gly Gln Leu Gly   179

CTG TTC AAG AAC GGC TAC TGG GAC AAT CCG GCC TAC AAG CTG CCG CCC GAA GCC GAT CTG ATG GCG ACG ACG CAT TAT CTG  1701
Leu Phe Lys Asn Gly Tyr Trp Asp Asn Pro Ala Tyr Lys Leu Pro Pro Glu Ala Asp Leu Met Ala Thr Thr His Tyr Leu   206

GAA GCG CTG GAT CTG CAA AAG GAA GTG GTC AAG GTC CAC ACG ATC TTC GGC GGC AAG AAC CCG CAT CCG AAC TGG CTT GTG  1782
Glu Ala Leu Asp Leu Gln Lys Glu Val Val Lys Val His Thr Ile Phe Gly Gly Lys Asn Pro His Pro Asn Trp Leu Val   233

GGC GGT GTG CCC TGT CCG ATC AAC GTC GAT GGC GTG GGC GCG GTC GGT GCG ATC AAC ATG GAG CGG CTG AAC CTC GTC TCC  1863
Gly Gly Val Pro Cys Pro Ile Asn Val Asp Gly Val Gly Ala Val Gly Ala Ile Asn Met Glu Arg Leu Asn Leu Val Ser   260

TCG ATC ATC GAC CGC TGC ACC GAA TTC ACC CGC AAC GTC TAT CTG CCC GAC CTC AAG GCC ATC GGC GGC TTC TAC AAG GAA  1944
Ser Ile Ile Asp Arg Cys Thr Glu Phe Thr Arg Asn Val Tyr Leu Pro Asp Leu Lys Ala Ile Gly Gly Phe Tyr Lys Glu   287

TGG CTC TAT GGC GGC GGG CTG TCG GGG CAA TCG GTG CTG TCC TAT GGC GAC ATC CCG GAA AAC CCG AAT GAT TTC AGC GCC  2025
Trp Leu Tyr Gly Gly Gly Leu Ser Gly Gln Ser Val Leu Ser Tyr Gly Asp Ile Pro Glu Asn Pro Asn Asp Phe Ser Ala   314

GGT CAG CTG CAT CTG CCG CGC GGG GCG ATC ATC AAC GGC AAC CTG AAC GAG GTG CAT GAC GTC GAC ACG ACC GAC CCC GAA  2106
Gly Gln Leu His Leu Pro Arg Gly Ala Ile Ile Asn Gly Asn Leu Asn Glu Val His Asp Val Asp Thr Thr Asp Pro Glu   341
```

Fig. 4.

```
CAG GTG CAG GAA TTC GTC GAC CAT TCC TGG TAT GAT TAC GGC GAG CCG GGC ATG GGG CTG CAT CCC TGG GAC GGC CGG ACC  2187
Gln Val Gln Glu Phe Val Asp His Ser Trp Tyr Asp Tyr Gly Glu Pro Gly Met Gly Leu His Pro Trp Asp Gly Arg Thr   368

GAG CCG AAA TTC GAG CTG GGC CCG AAC CTG AAA GGC ACC CGC ACC AAC ATC GAG AAC ATC GAC GAA GGC GCG AAA TAT TCC  2268
Glu Pro Lys Phe Glu Leu Gly Pro Asn Leu Lys Gly Thr Arg Thr Asn Ile Glu Asn Ile Asp Glu Gly Ala Lys Tyr Ser   395

TGG ATC AAG GCG CCG CGC TGG CGC GGC AAT GCG ATG GAG GTG GGG CCG CTG GCC GCT ACA TCG TCG GTT ACG CGC AAG GGC  2349
Trp Ile Lys Ala Pro Arg Trp Arg Gly Asn Ala Met Glu Val Gly Pro Leu Ala Ala Thr Ser Ser Val Thr Arg Lys Gly   422

CAC GAG GAC ATC AAG AAC CAG GTC GAG GGC CTG CTG CGC GAC ATG AAC CTG CCG GTT TCG GCG CTG TTC TCG ACG CTG GGC  2430
His Glu Asp Ile Lys Asn Gln Val Glu Gly Leu Leu Arg Asp Met Asn Leu Pro Val Ser Ala Leu Phe Ser Thr Leu Gly   449

CGG ACT GCG GCG CGG GCT CTG GAG GCG GAA TAC TGC TGC CGT CTG CAA AAG CAC TTC TTC GAC AAG CTG GTC ACC AAC ATC  2511
Arg Thr Ala Ala Arg Ala Leu Glu Ala Glu Tyr[Cys][Cys]Arg Leu Gln Lys His Phe Phe Asp Lys Leu Val Thr Asn Ile   476

AAG AAC GGC GAC AGC TCG ACC GCG AAT GTC GAG AAA TGG GAT CCC TCG ACC TGG CCG AAG GAG GCC AAG GGC GTG GGC ATG  2592
Lys Asn Gly Asp Ser Ser Thr Ala Asn Val Glu Lys Trp Asp Pro Ser Thr Trp Pro Lys Glu Ala Lys Gly Val Gly Met   503

ACC GAG GCG CCG CGC GGC GCG CTG GGC CAT TGG GTC AAG ATC AAG GAC GGC CGC ATC GAG AAC TAT CAA TGC GTC GTG CCG  2673
Thr Glu Ala Pro Arg Gly Ala Leu Gly His Trp Val Lys Ile Lys Asp Gly Arg Ile Glu Asn Tyr Gln[Cys]Val Val Pro   530

ACG ACC TGG AAC GGC TCG CCC CGC GAC AGC AAG GGC AAT ATC GGC GCC TTC GAG GCC TCG CTG CTG AAC ACG AAG ATG GAA  2754
Thr Thr Trp Asn Gly Ser Pro Arg Asp Ser Lys Gly Asn Ile Gly Ala Phe Glu Ala Ser Leu Leu Asn Thr Lys Met Glu   557

CGC CCC GAA GAG CCG GTC GAG ATC CTG CGC ACG CTG CAC AGC TTC GAT CCG TGC CTG GCC TGT TCG ACG CAT GTG ATG TCG  2835
Arg Pro Glu Glu Pro Val Glu Ile Leu Arg Thr Leu His Ser Phe Asp Pro[Cys]Leu Ala[Cys]Ser Thr His Val Met Ser   584

GCC GAA GGC CCC CCT GAC CAC CGT CAA GGT CCG GTA GGG GGA TGC CAT GAA GGG AGT TTC CGA CGA AAG GAT CAA TGC CCC  2916
Ala Glu Gly Pro Pro Asp His Arg Gln Gly Pro Val Gly Gly[Cys]His Glu Gly Ser Phe Arg Arg Lys Asp Gln[Cys]Pro   611

CGT CCG TGG CCC GGA tgaaatcttcgaggcctcgcgactgaccgtcgacgccacccgcgaggacctagaaagcatccggcggcgtacctccgtctatgtcta  3018
Arg Pro Trp Pro Gly ***                                                                                     616
```

Fig. 4. Nucleotide and derived amino acid sequences for the small and the large subunits of *Rhodobacter capsulatus* hydrogenase. Coding regions are written in *capital letters* (and in *italics* for the putative signal peptide). Inverted repeats are represented by *facing arrows*. The position of the putative Shine-Dalgarno sequences is *underlined*. Cysteinyl residues are included in *boxes*. Amino acid residues of the small and large subunits are numbered from the N terminal amino acid of the mature protein in each subunit

identified as the gene encoding the large subunit (*hupL*). The gene contained 1851 nucleotides including the initiation ATG codon and terminated with a TGA stop codon at position 2932. A hairpin structure with a 19 bp loop and a 7 bp stem containing 6 GC pairs, beginning at position 2946, followed by another one at position 2992 (not shown) may be involved in the termination of transcription. An "AGGAGG" sequence, complementary to the 3′ terminus of *R. capsulatus* 16 S rRNA (Youvan et al. 1984) and forming a perfect ribosome binding site, was located 14 bases upstream from the start codon. However, no promoter could be identified in that region. The gene was translated into 616 amino acids corresponding to a molecular mass of 68108 Da.

Immediately upstream and separated by only three nucleotides was the other open reading frame (ORF1) ending with a TAA codon at position 1075. The NH_2 terminus of the predicted amino acid sequence of ORF1 could be aligned with that of the small subunit of the hydrogenase from *B. japonicum* (Sayavedra et al. 1988). As shown in Fig. 5 most of the residues are identical in the two proteins.

Of particular interest was a region which exhibited a sequence similar to the NH_2 amino acid terminus of the small subunit of *B. japonicum* hydrogenase, arrived at by protein sequencing (L.C. Seefeldt and D.J. Arp, personal communication). This homology permitted us to assign position number one to the methionine residue immediately preceding the NH_2 terminus of the mature protein. The

```
A : Q L S D I E T F Y D V M R R Q G I T R R S F M K S V R S P Q
B : M G A A T E T F Y S V I R R Q G I T R R S F H K F C S L T A

A : H V L G L G P S F V P K I G E A M E T K P R T P V V W V H G
B : T S L G L G P L A A S R I A N A L E T K P R V P V I W M H G

A : L E C T C C S E S F I R S A H P L A K D V V L S M I S L D Y
B : L E C T C C S E S F I R S A H P L V K D A V L S M I S L D Y

A : D D T L M A A A G H A A E A A F E E T I A K Y K G N Y I L A
B : D D T I M A A A G H Q A E A I L E E T R A K H K G Q Y I L A

A : V E G N P P L N E D G M F C I T G G K P F V E K L R H A A E
B : V E G N P P L N E G G M F C I D G G K P F V E K L K M M A E

A : G A K A I I S W G A C A S Y G C V Q A A A P N P T Q A T P V
B : D A M A I I A W G A C A S W G C V Q A A K P N P T Q A T P I

A : H K V I T D K P I I K V P G C P P I A E V M T G V I T Y M L
B : D K V I T N K P I I K V P G C P P I A E V M T G V V T F I T

A : T F D R M P E L D R Q G R P A M F Y S Q R I H D K C Y R R P
B : T F G K L P E L A R Q G R P K M F Y S Q R I H D K C Y R R P

A : H F D A G Q F V E H W D D E N A R K G Y C L Y K M G C K G P
B : H F D A G Q F V E E W D D E A A R K G Y C L Y K M G C K G P

A : T T Y N A C S T V P L E R R R H F P I Q S G H G C I G C S E
B : T T Y N A C S T V R W N G G V S F P I Q S G H G C I G C S E

A : D G F W D Q G S F Y D R L T T I K Q F G I E A T A D Q I G W
B : D G F W D K G S F Y D R L T N I K Q F G I E K N A D Q I G M

A : T A T G L V G A A V A A H A A V S V L K R A Q K K N E E A -
B : V A A G A V G A A V A A H A A V T A V K R L A T K R E D A D

A : - - -
B : H N S
```

Fig. 5. Comparison of the deduced polypeptide sequences of the leader peptides (in *italics*) and small subunits of the hydrogenases from *Rhodobacter capsulatus* (*line A*) and *Bradyrhizobium japonicum* (*line B*). Common residues are *boxed*

Table 2. Amino acid composition of the large subunit of *Rhodobacter capsulatus* hydrogenase and codon usage

Amino acid	Number of residues	Codon	% use	Amino acid	Number of residues	Codon	% use
Ala	39	GCT	5.1	Lys	29	AAA	17.2
		GCC	38.5			AAG	82.8
		GCA	0.0				
		GCG	56.4	Met	14	ATG	100.0
Arg	34	CGT	8.8	Phe	18	TTT	0.0
		CGC	64.7			TTC	100.0
		CGA	5.9				
		CGG	20.6	Pro	42	CCT	2.4
		AGA	0.0			CCC	35.7
		AGG	0.0			CCA	0.0
						CCG	61.9
Asn	38	AAT	21.1				
		AAC	78.9	Ser	32	TCT	0.0
						TCC	18.8
Asp	35	GAT	37.1			TCA	0.0
		GAC	62.9			TCG	59.4
						AGT	3.1
Cys	12	TGT	16.7			AGC	18.8
		TGC	83.3				
				Thr	37	ACT	2.7
Gln	18	CAA	38.9			ACC	59.5
		CAG	61.1			ACA	2.7
						ACG	35.1
Glu	38	GAA	55.3				
		GAG	44.7	Trp	15	TGG	100.0
Gly	58	GGT	8.6	Tyr	14	TAT	57.1
		GGC	69.0			TAC	42.9
		GGA	3.4				
		GGG	19.0	Val	44	GTT	6.8
						GTC	54.5
His	22	CAT	63.6			GTA	2.3
		CAC	36.4			GTG	36.4
Ile	29	ATT	0.0	*	1	TGA	100.0
		ATC	100.0			TAA	0.0
		ATA	0.0			TAG	0.0
Leu	49	TTA	0.0				
		TTG	0.0				
		CTT	4.1				
		CTC	8.2				
		CTA	0.0				
		CTG	87.8				

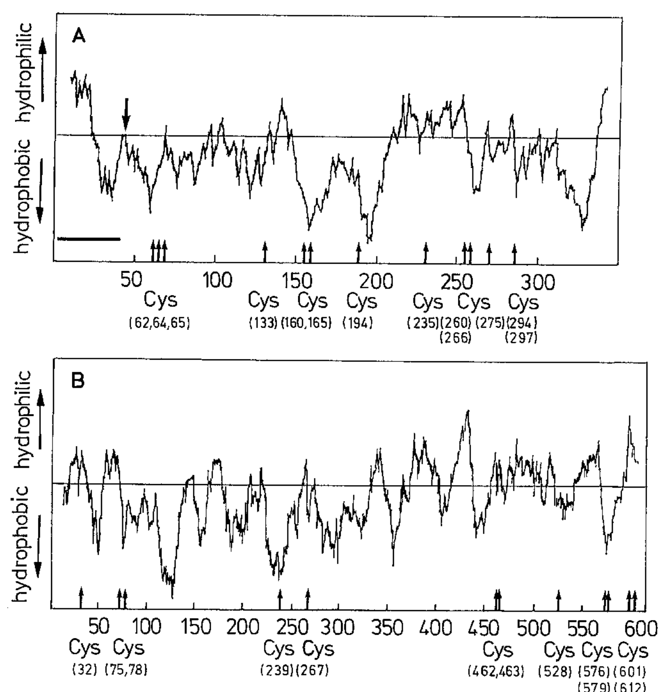

Fig. 6A, B. Hydrophilic plot for **A** the small subunit and **B** the large subunit of *Rhodobacter capsulatus* hydrogenase. A moving window of 19 amino acid residues was used. Positions of Cys are indicated with an *arrow* and *number* refer to positions of Cys in the deduced amino acid sequence of each protein. The region of the putative signal peptide upstream of the small subunit protein is indicated by a *bar*

two predicted amino acid sequences could even be aligned further upstream of the NH_2 protein terminus (Fig. 5). The upstream amino acid sequence would represent a signal sequence whose starting site for translation for *R. capsulatus* was not determined; the processing site would be Ala/Leu for the *B. japonicum* protein and Ala/Met for *R. capsulatus* protein (*hupS*).

The strong similarity between the predicted amino acid sequences of the small subunit of *B. japonicum* hydrogenase and of the product of ORF1 lead us to conclude that ORF1 encodes the small subunit of *R. capsulatus* hydrogenase.

The amino acid composition with codon usage of the large subunit is given in Table 2. There is a high selectivity of codon usage. For example, out of the six leucine codons, only three are used by preference to a CTG codon, or out of three isoleucine codons, only the ATC codon is used. The third position is a G or a C in more than 86% of total codons. Overall, the G+C content is 64.1% and 62.8% for the small and large subunits, respectively. This compares well with the base composition of *R. capsulatus* chromosomal DNA (Yu et al. 1982). This codon usage, compiled according to Staden and McLachlan (1982), is similar to that found in sequenced genes from *R. capsulatus* and other Rhodospirillaceae (cf. Leclerc 1988).

The small and large subunits contain 13 and 12 Cys residues, respectively. In the large subunit, 7 Cys are localised near the COOH terminal region after residue 450 and 3 are near the NH_2 terminus before residue 80. In the small subunit, cysteines appear to be more evenly distributed in the protein sequence.

The polarity index, calculated from the amino acid composition, indicates that 54% and 59% of the residues from the large and the small subunit, respectively, are hydrophobic. The hydropathy plot of the large subunit (Levitt 1976) shows that the region near the NH_2 terminus is very hydrophobic while that near the COOH terminus is more hydrophilic (Fig. 6B). The small protein is more hydrophobic than the large one. Three large hydrophobic domains where most of the Cys residues of the small subunit are located are shown on the hydrophilic profile of the protein (Fig. 6A). The putative signal peptide is characterized by the presence of charged amino acids in the NH_2 terminus region followed by a very hydrophobic stretch preceding the processing locus.

Sequence homology analysis

The NH_2 terminal amino acid sequence of the large subunit of the *R. capsulatus* hydrogenase shows strong homology

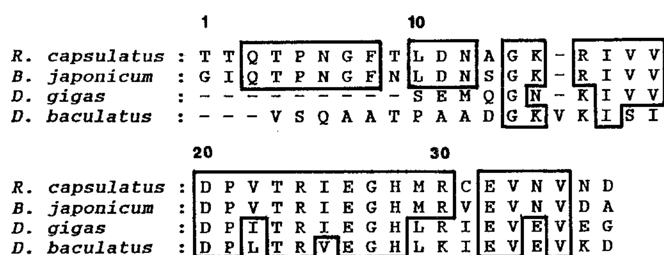

Fig. 7. Alignment of the NH_2 terminal sequences of the large subunit of the hydrogenases from *Rhodobacter capsulatus*, *Bradyrhizobium japonicum*, *Desulfovibrio gigas*, and *D. baculatus*. *Numbers* refer to positions of the residues in the *R. capsulatus* sequence. Conserved residues between at least two sequences are *boxed*

with that of *B. japonicum* hydrogenase (Sayavedra et al. 1988), as well as similarities with both the *D. gigas* (Prickril et al. 1987) and *D. baculatus* (Menon et al. 1987) hydrogenases (Fig. 7). Protein alignment was done using a modification of the algorithm of Needleman and Wunsch (1970). The start points of the sequences of *R. capsulatus* and *B. japonicum* coincided with one another, while the first amino acid of the large subunit of *D. baculatus* and *D. gigas* hydrogenases are aligned with the fourth and the tenth amino acid residues of *R. capsulatus*, respectively.

The predicted amino acid sequence of *R. capsulatus* hydrogenase was compared with each of the 4 other hydrogenases so far sequenced (Table 3). The *R. capsulatus* sequence could not be aligned with the (Fe)hydrogenase of *D. vulgaris*. However, around 30% sequence homology was found between the *R. capsulatus* hydrogenase and the (NiFe)hydrogenase of *D. gigas* and the (NiFeSe) hydrogenase of *D. baculatus*. More than 69% of the amino acid residues of the (NiFe)hydrogenase of *B. japonicum* were identical to the (NiFe)hydrogenase of *R. capsulatus* and in particular all of the 13 Cys of the small subunit and the 6 Cys of the large subunit of the *B. japonicum* enzyme were conserved in the *R. capsulatus* enzyme.

Discussion

The cosmid clone AC70, described in this report, is the third clone from the gene bank of *R. capsulatus* DNA found to complement Hup$^-$ mutants isolated from strain B10. Two other clones (AC57 and AG202) have been shown previously to complement two different series of Hup$^-$ mutants (Colbeau et al. 1986). In accord with the fact that the recombinant plasmids complement different Hup$^-$ mutants, the restriction patterns of DNA inserts in pAC57, pAG202 and pAC70 are totally unlike one another (A. Colbeau, unpublished results). The isolation of three unrelated recombinant plasmids complementing distinct Hup$^-$ mutants indicate that several *hup* genes may be involved in the synthesis of hydrogenase.

In *E. coli*, two unlinked hydrogenase genes (*hydA* and *hydB*) that code for hydrogenase synthesis and activity were isolated by Sankar et al. (1985). They map between 58 and 59 min in the *E. coli* chromosome. Two additional loci (*hydC* and *hydD*) located at 77 min on the *E. coli* chromosome map were identified by Wu and Mandrand-Berthelot (1986). However, in neither case was the structural gene(s) for the *E. coli* hydrogenase identified. In *Alcaligenes eutrophus* H16, structural and regulatory genes encoding the hydrogen oxidizing enzyme system (*hox*) reside on the 450 kb plasmid, pGH1. In this system, both structural and regulatory genes are clustered in a DNA sequence of ca. 80 kb representing several transcription units (Kortlücke et al. 1987).

The present report brings to five the number of hydrogenases whose structural genes have been sequenced. These are the periplasmic (Fe)hydrogenase from *D. vulgaris* (Voordouw and Brenner 1985), the periplasmic (NiFeSe) hydrogenase from *D. baculatus* (Menon et al. 1987), the periplasmic (NiFe)hydrogenase from *D. gigas* (Li et al. 1987) and the membrane-bound (NiFe)hydrogenase from *B. japonicum* (Sayavedra et al. 1988).

The structural genes of *R. capsulatus* were cloned by homology with the gene encoding the large subunit of *B. japonicum* hydrogenase. The nucleotide sequence of the cloned fragment (Fig. 4) revealed the presence of two open reading frames (ORF1 and ORF2). The protein sequences were deduced using a table of codon frequencies (Staden and McLachlan 1982) established from sequenced genes in other Rhodospirillaceae. ORF1 and ORF2 were identified as the genes encoding the small and the large subunits, respectively, on the following basis. ORF1 and ORF2 en-

Table 3. Comparison between the derived amino acid sequence of *Rhodobacter capsulatus* hydrogenase and the hydrogenases from *Desulfovibrio vulgaris*, *D. gigas*, *D. baculatus* and *Bradyrhizobium japonicum*

Microorganism	Type of hydrogenase	Percent of conserved residues	Number of conserved cysteines	Reference
R. capsulatus	(NiFe)	small s.u.: 100%	13 (13)[a]	This work
	membrane bound	large s.u.: 100%	12 (12)	
D. vulgaris	(Fe)	small s.u.: –[b]	– (0)	Voordouw and Brenner (1985)
	periplasmic	large s.u.: –[b]	– (18)	
D. gigas	(NiFe)	small s.u.: 29%	9 (12)	Li et al. (1987)
	periplasmic	large s.u.: 34%	6 (18)	
D. baculatus	(NiFeSe)	small s.u.: 32%	11 (13)	Menon et al. (1987)
	periplasmic	large s.u.: 25%	8 (12)	
B. japonicum	(NiFe)	small s.u.: 80%	13 (13)	Sayavedra et al. (1988)
	membrane-bound	large s.u.: 69%	6 (6)	

s.u., subunit

[a] The total number of Cys residues is indicated in parenthesis

[b] The *R. capsulatus* and *D. vulgaris* sequences were not alignable

coded polypeptides of 34256 Da and 68108 Da, respectively. Their molecular masses were similar to those of the small and large hydrogenase subunits determined by SDS gel electrophoresis (Colbeau et al. 1983; Seefeldt et al. 1987). The product of ORF2 was identified as the large subunit by alignment with the NH_2 terminal sequence of the protein obtained by Edman degradation. The protein products of ORF1 and ORF2 exhibited significant amino acid sequence similarities with the small and large subunits, respectively, of the hydrogenases from *B. japonicum* and from two *Desulfovibrio* species (Table 3). All these facts furnish strong circumstantial evidence which suggests that the two genes sequenced in this study are the structural genes of the *R. capsulatus* hydrogenase. The identification of the transcript(s) of the hydrogenase structural genes will provide further evidence that the genes belong to the same operon. In preliminary experiments it was observed that each individual gene hybridized to the same mRNA species (data not shown; unpublished results).

The gene encoding the small protein subunit is located upstream of the gene encoding the large one (Fig. 4). The two genes are transcribed in the same direction and, due to their close proximity and to the absence of recognisable promoters and termination signals for transcription, they probably belong to the same operon. A similar arrangement was found in *D. baculatus, D. gigas* and *B. japonicum* for the genes of the small and large subunits of hydrogenase. In *D. vulgaris*, the two structural genes belong to the same transcriptional unit, but the order is changed and the large subunit precedes the smaller one (Voordouw and Brenner 1985).

A perfect sequence complementarity to the 16 S rRNA 3′end (..AGGAGG..), the Shine-Dalgarno sequence, is located at the 3′ end of the ORF1 gene. It is separated by only 2 nucleotides from the stop codon TAA and by 8 nucleotides from the initiation codon ATG of the second (ORF2) gene. With this structure, ribosomes should remain simultaneously in contact with the termination codon of the first gene and the initiation codon of the second, since they physically span about 35 bases of mRNA. The presence of the Shine-Dalgarno sequence should prevent dissociation from mRNA of the 30 S subunit of the terminating ribosome, and thereby enable uninterrupted translation of ORF2.

The NH_2 terminus of the predicted protein sequence for the large subunit corresponds to the first amino acid immediately after a methionine start codon indicating that this polypeptide does not contain a leader sequence. On the other hand, the small subunit is preceded by a signal peptide, identified by its homology with the signal peptide of the small subunit of *B. japonicum* hydrogenase (Fig. 5). These observations suggest that a single signal peptide may be used to insert both hydrogenase subunits into the membrane.

The distribution of the cysteinyl residues does not present the typical [4Fe−4S] ferredoxin pattern (Adman et al. 1973; Bruschi and Guerlesquin 1988). It is at variance with that found for *D. vulgaris* hydrogenase, where the large subunit was shown to contain at its NH_2 terminus a duplication of the typical sequence of 4 Cys residues for the coordination of [4Fe−4S] clusters (Voordouw and Brenner 1985). No such characteristic arrangement was found in *R. capsulatus*. On the other hand, among the Cys residues in the 4 hydrogenases presenting sequence homologies (Table 3), identification of the Cys residues common to the 4 hydrogenases may provide a means of predicting the ligation sites of the iron and nickel atoms. Seven Cys residues are conserved in the hydrogenase small subunits of the 4 species, 5 of which are located in the stretch: ClykmGCKGPtTYnaC ...... gCIGCsE. In this stretch beginning with Cys at position 215 and ending at position 254 there are 14 residues (written in bold face) conserved in all 4 proteins. Only 4 Cys were found conserved in all 4 large subunits. In the *R. capsulatus* sequence, they appear at positions 74 and 77 near the NH_2 terminus in the stretch grdprdawafteriCgvCtg and at positions 575 and 578 near the carboxyl terminus (Fig. 4). These conserved Cys could be involved in the active site of the enzyme.

From the comparison of the predicted protein sequences, the five hydrogenases sequenced up to now can be divided into two groups. The first group consists of the (Fe)hydrogenase of *D. vulgaris* which presents the typical Cys arrangement of [4Fe−4S] ferredoxins and the second group includes both the (NiFe)hydrogenases and the (NiFeSe)hydrogenases. The similarities found between the hydrogenases of the second group had not been anticipated, and the presence of Se in the socalled (NiFe)hydrogenases is worth exploring.

Acknowledgements. The authors gratefully thank Prof. H.J. Evans (Corvallis, USA) for a generous gift of plasmid pHUI, for making available before publication the complete sequence of the hydrogenase from *B. japonicum* and for permission to make use of the *B. japonicum* hydrogenase sequence before its publication. They also thank Dr. Arp for communicating the N-terminal protein sequences of the hydrogenase of *B. japonicum* prior to publication, Dr. M. Bruschi and J. Bonicel from the Laboratoire de Chimie Bactérienne du CNRS (Marseille) for help in the determination of the NH_2 terminal protein sequence of *R. capsulatus* hydrogenase, Drs. R.L. Uffen, P. Richaud and J.C. Willison for stimulating discussions, and the authors are grateful to Jacqueline Boyer for typing the manuscript. This work was supported by grants from the Centre National de la Recherche Scientifique (UA 1130, AIP-AFME 3231).

References

Adman ET, Sieker LC, Jensen LH (1973) The structure of a bacterial ferredoxin. J Biol Chem 248:3987–3996

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7:1513–1523

Boyer HW, Roulland-Dussoix D (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J Mol Biol 41:459–472

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Bruschi M, Guerlesquin F (1988) Structure, function and evolution of bacterial ferredoxins. FEMS Microbiol Rev 54:155–176

Cantrell MA, Haugland RA, Evans HJ (1983) Construction of a *Rhizobium japonicum* gene bank and use in isolation of a hydrogen uptake gene. Proc Natl Acad Sci USA 80:181–185

Colbeau A, Vignais PM (1981) The membrane-bound hydrogenase of *Rhodopseudomonas capsulata*. Stability and catalytic properties. Biochim Biophys Acta 662:271–284

Colbeau A, Vignais PM (1983) The membrane-bound hydrogenase of *Rhodopseudomonas capsulata* is inducible and contains nickel. Biochim Biophys Acta 748:128–138

Colbeau A, Chabert J, Vignais PM (1983) Purification, molecular properties and localization in the membrane of the hydrogenase

of *Rhodopseudomonas capsulata*. Biochim Biophys Acta 748:116–127

Colbeau A, Godfroy A, Vignais PM (1986) Cloning of DNA fragments carrying hydrogenase genes of *Rhodopseudomonas capsulata*. Biochimie 68:147–155

Ditta G, Stanfield S, Corbin D, Helinski DR (1980) Broad host range DNA cloning system for Gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. Proc Natl Acad Sci USA 77:7347–7351

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

Friedrich B, Kortlüke C, Hogrefe C, Eberz G, Silber B, Warrelmann J (1986) Genetics of hydrogenase from aerobic lithoautotrophic bacteria. Biochimie 68:133–145

Gogotov IN (1986) Hydrogenases of phototrophic microorganisms. Biochimie 68:181–187

Hager DA, Burgess RR (1980) Elution of proteins from SDS polyacrylamide gels, removal of SDS and renaturation of enzymatic activity: results with subunit of *E. coli* RNA polymerase, wheat germ DNA topoisomerase and other enzymes. Anal Biochem 109:76–86

Hattori M, Sakaki Y (1986) Dideoxy sequencing method using denatured plasmid templates. Anal Biochem 152:232–238

Hillmer P, Gest H (1977) H_2 metabolism in the photosynthetic bacterium *Rhodopseudomonas capsulata:* H_2 production by growing cultures. J Bacteriol 129:720–731

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

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

Ish-Horowicz D, Burke JF (1981) Rapid and efficient cosmid cloning. Nucleic Acids Res 9:2989–2998

Kortlücke C, Hogrefe C, Eberz G, Pühler A, Friedrich B (1987) Genes of lithoautotrophic metabolism are clustered on the megaplasmid pHG1 in *Alcaligenes eutrophus*. Mol Gen Genet 210:122–128

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lambert GR, Cantrell MA, Hanus FJ, Russell SA, Haddad KR, Evans HJ (1985) Intra and interspecies transfer and expression of *Rhizobium japonicum* hydrogen uptake genes and autotrophic growth capability. Proc Natl Acad Sci USA 82:3232–3236

Leclerc M (1988) Génétique de l'hydrogénase chez *Rhodobacter capsulatus:* séquençage de gènes et identification des ARNm. Thesis, University of Grenoble, France

Levitt M (1976) A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol 104:59–107

Li C, Peck HD Jr, LeGall J, Przybyla AE (1987) Cloning, characterization, and sequencing of the genes encoding the large and small subunits of the periplasmic (NiFe)hydrogenase of *Desulfovibrio gigas*. DNA 6:539–551

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Marmur J (1961) A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J Mol Biol 3:208–218

Marrs B (1974) Genetic recombination in *Rhodopseudomonas capsulata*. Proc Natl Acad Sci USA 71:971–973

Menon NK, Peck Jr HD, Le Gall J, Przybyla AE (1987) Cloning and sequencing of the genes encoding the large and small subunits of the periplasmic (NiFeSe) hydrogenase of *Desulfovibrio baculatus*. J Bacteriol 169:5401–5407

Mizusawa S, Nishumura S, Seela F (1986) Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res 14:1319–1324

Morrison DA (1979) Transformation and preservation of competent bacterial cells by freezing. Methods Enzymol 68:326–331

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Poncz M, Solowiejczyk D, Ballantine M, Schwartz E, Surrey S (1982) "Nonrandom" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc Natl Acad Sci USA 79:4298–4302

Prickril BC, He SH, Li C, Menon N, Choi ES, Przybyla AE, Dervartanian DV, Peck Jr HD, Fauque G, Le Gall J, Teixeira M, Moura I, Moura JJG, Patil D, Huynh BH (1987) Identification of three classes of hydrogenase in the genus *Desulfovibrio*. Biochem Biophys Res Commun 149:369–377

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sankar P, Lee JH, Shanmugam KT (1985) Cloning of hydrogenase genes and fine structure analysis of an operon essential for H_2 metabolism in *Escherichia coli*. J Bacteriol 162:353–360

Sayavedra L, Powell G, Morris R, Evans HJ (1988) Nucleotide sequence of the genetic loci encoding subunits of *Bradyrhizobium japonicum* uptake hydrogenase. 7th International Congress on N_2 Fixation, Cologne, FRG, March (Abst. 12–44)

Seefeldt LC, MacCollum LC, Doyle CM, Arp DJ (1987) Immunological and molecular evidence for a membrane-bound dimeric hydrogenase in *Rhodopseudomonas capsulata*. Biochim Biophys Acta 914:299–303

Staden R, McLachlan AD (1982) Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res 10:141–174

Strauss EC, Kobori JA, Siu F, Hood LE (1986) Specific primer directed DNA sequencing. Anal Biochem 154:353–360

Tibelius KH, Robson RL, Yates MF (1987) Cloning and characterization of hydrogenase genes from *Azotobacter chroococcum*. Mol Gen Genet 206:285–290

Vignais PM, Colbeau A, Willison JC, Jouanneau Y (1985) Hydrogenase, nitrogenase and hydrogen metabolism in the photosynthetic bacteria. Adv Microbial Physiol 26:155–234

Voordouw G, Brenner S (1985) Nucleotide sequence of the gene encoding the hydrogenase from *Desulfovibrio vulgaris* (Hildenborough). Eur J Biochem 148:515–520

Weaver PF, Wall JD, Gest H (1975) Characterization of *Rhodopseudomonas capsulata*. Arch Microbiol 105:207–216

Willison JC, Magnin JP, Vignais PM (1987) Isolation and characterization of *Rhodobacter capsulatus* strains lacking endogenous plasmids. Arch Microbiol 147:134–143

Wu LF, Mandrand-Berthelot MA (1986) Genetic and physiological characterization of new *Escherichia coli* mutants impaired in hydrogenase activity. Biochimie 68:167–179

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13*mp18* and *pUC19* vectors. Gene 33:103–119

Youvan DC, Alberti M, Begusch H, Bylina EJ, Hearst JE (1984) Reaction center and light-harvesting I genes from *Rhodopseudomonas capsulata*. Proc Natl Acad Sci USA 81:189–192

Yu PL, Hohn B, Falk H, Drews G (1982) Molecular cloning of the ribosomal RNA genes of the photosynthetic bacterium *Rhodopseudomonas capsulata*. Mol Gen Genet 188:392–398

Zuber M, Harker AR, Sultana MA, Evans HJ (1986) Cloning and expression of *Bradyrhizobium japonicum* uptake hydrogenase structural genes in *Escherichia coli*. Proc Natl Acad Sci USA 83:7668–7672

Communicated by H. Hennecke

Received February 2, 1988