Cloning, DNA Sequencing, and Characterization of a nifD-Homologous Gene from the Archaeon Methanosarcina barkeri 227 Which Resembles nifD1 from the Eubacterium Clostridium pasteurianum

YUEH-TYNG CHIEN and STEPHEN H. ZINDER*
Section of Microbiology, Cornell University, Ithaca, New York 14853

Received 5 May 1994/Accepted 18 August 1994

L. Sibold, M. Henriquet, O. Possot, and J.-P. Aubert (Res. Microbiol. 142:5–12, 1991) cloned and sequenced two nifH-homologous open reading frames (ORFs) from Methanosarcina barkeri 227. Phylogenetic analysis of the deduced amino acid sequences of the nifH ORFs from M. barkeri showed that nifH1 clusters with nifH genes from alternative nitrogenases, while nifH2 clusters with nifH1 from the gram-positive eubacterium Clostridium pasteurianum. The N-terminal sequence of the purified nitrogenase component 2 (the nifH gene product) from M. barkeri was identical with that predicted for nifH2, and dot blot analysis of RNA transcripts indicated that nifH2 (and nifDK2) was expressed in M. barkeri when grown diazotrophically in Mo-containing medium. To obtain nifD2 from M. barkeri, a 4.7-kbp BamHI fragment of M. barkeri DNA was cloned which contained at least five ORFs, including nifH2, ORF105, and ORF125 (previously described by Sibold et al.), as well as nifD2 and part of nifK2. ORF nifD2 is 1,596 bp long and encodes 532 amino acid residues, while the nifK2 fragment is 135 bp long. The deduced amino acid sequences for nifD2 and the nifK2 fragment from M. barkeri cluster most closely with the corresponding nifDK1 gene products from C. pasteurianum. The predicted M. barkeri nifD2 product contains a 50-amino acid insert near the C terminus which has previously been found only in the clostridial nifD1 product. Previous biochemical and sequencing evidence indicates that the C. pasteurianum nitrogenase is the most divergent of known eubacterial Mo-nitrogenases, most likely representing a distinct nif gene family, which now also contains M. barkeri as a member. The similarity between the methanogen and clostridial nif sequences is especially intriguing in light of the recent findings of sequence similarities between gene products from archaea and from low-G+C gram-positive eubacteria for glutamate dehydrogenase, glutamine synthetase I, and heat shock protein 70. It is not clear whether this similarity is due to horizontal gene transfer or to the resemblance of the M. barkeri and C. pasteurianum nitrogenase sequences to an ancestral nitrogenase.

Nitrogenase, the two-component enzyme complex responsible for reduction of N₂ to NH₃, shows high levels of conservation of structure, function, and amino acid sequence across wide phylogenetic ranges (11). In typical Mo-nitrogenases, component 1 (also called the MoFe-protein or dinitrogenase) is an α₂β₂ tetramer encoded by nifD and nifK. It contains two unusual metal clusters, an 8Fe-8S cluster called the P cluster and the FeMo-cofactor, which has the proposed composition 7Fe-9S-Mo (22) and is considered to be the site of N₂ reduction. Component 2 (also called the Fe-protein or dinitrogenase reductase) is a homodimer with a single 4Fe-4S cluster linking the subunits (18) and is encoded by nifH. There is an especially high degree of sequence conservation in nifH1 gene products (11, 18). Two discoveries in the 1980s greatly increased the understanding of nitrogenase diversity; the discovery of alternative nitrogenases and the discovery of nitrogen fixation in methanogenic archaea (archaeobacteria).

Bishop et al. (5) originally proposed the existence of alternative nitrogenases, generally expressed in the absence of Mo, which contain either V (vanadium) instead of Mo in their component 1 (V-nitrogenases) or contain only Fe (Fe-nitrogenases). V- and Fe-nitrogenase genes have been called vnf and anf, respectively, or are given numbers such as nifD and nif3, respectively, in the case of Azotobacter vinelandii. Evidence for the presence of alternative nitrogenases or their genes has now been obtained for a wide variety of eubacteria (14, 15, 21, 24, 26, 51, 57). Alternative nitrogenases show biochemical properties and amino acid sequences which are distinct from those of Mo-nitrogenases (6).

In 1984 and 1985, nitrogen fixation was reported for the methanogenic archaea Methanosarcina barkeri (7, 34) and Methanococcus thermolithotrophicus (3), the first description of this process outside the eubacterial (bacterial) domain. Nitrogen fixation has since been detected in a wide range of methanogens (2, 29). Sibold and colleagues detected sequences in methanogen DNA which hybridized to eubacterial nifH probes (37, 45) and went on to clone and sequence several nifH-homologous open reading frames (ORFs) from methanogens (44, 46–48) and a nifD ORF from M. thermolithotrophicus (48).

The phylogeny of deduced nifH gene products, with an emphasis on alternative and methanogen nitrogenases, is depicted in Fig. 1. There are essentially four major clusters. Cluster I is defined by the standard eubacterial Mo-nitrogenases (such as A. vinelandii nifH1). Cluster II consists of alternative nitrogenases (nifH3 from A. vinelandii and nifH3 from Clostridium pasteurianum) and three methanogen nitrogenase ORFs. The appearance of A. vinelandii nifH2 (vnfH) in the eubacterial Mo-nitrogenase cluster is anomalous, since the

---

* Corresponding author. Mailing address: Section of Microbiology, Wing Hall, Cornell University, Ithaca, NY 14853. Phone: (607) 255-2415. Fax: (607) 255-3094. Electronic mail address: sbt1@cornell.edu.
corresponding nifDK genes cluster with alternative nitrogenases (41). Cluster III consists of nifH1 from C. pasteurianum (and five other closely related nifH ORFs in that organism which are apparently products of gene duplication [53]) and nifH2 from M. Barkeri (44). This is the first instance of clustering of a methanogen nitrogenase with a eubacterial Mo-nitrogenase. Cluster IV is highly divergent and consists of predicted gene products solely from methanogens, such as Methanococcus voltae. No evidence for nitrogenase function has been obtained for these genes.

So far, there is limited information on the structure and function of methanogen nitrogenases. In previous studies on M. Barkeri 227, we showed that diazotrophic growth was stimulated by Mo (27) and that the partially purified nitrogenase complex from M. Barkeri was binary, with low specific activity (28). We also showed that the component 2 from M. Barkeri reacted in Western blots (immunoblots) with antibody to Rhodospirillum rubrum component 2 and that this protein was not detected in blots from ammonia-grown cells, suggesting repression. In this study, we show that nifH2 is expressed in M. Barkeri 227 when growing diazotrophically in standard Mo-containing medium. We also describe the cloning and sequencing of the nifD2 gene and part of nifK2 from M. Barkeri and show that as with nifH2, their predicted amino acid sequences show greatest similarity with those from C. pasteurianum.

**Materials and Methods**

**Bacterial strains and plasmids.** M. Barkeri 227 (ATCC 43241, DSM 1538, and OCM 35) was obtained from our own culture collection. Escherichia coli DH5α was used for transformation and DNA sequencing experiments and was obtained from S. Winans, Cornell University. Plasmid pUC19 was used for subcloning and DNA sequencing.

**Medium and growth conditions.** The growth medium for M. Barkeri was that of Lobo and Zinder (27) as recently modified so that the culture headspace gas was N₂ and the medium was buffered by 12 mM phosphate (pH 7.0) rather than CO₂−HCO₃⁻ (10). Cultures for enzyme purification were grown in 10-liter amounts in a 14-litre fermenter vessel at 35°C. Once significant growth and methanogenesis occurred, the cultures were slowly sparged with N₂ gas (~10 ml/min), 5 ml of 20% (wt/vol) Na₂S was added every other day, and 1 ml of 1 M FeSO₄-7H₂O was added twice during the course of growth. Methanol was added as required to concentrations as great as 100 mM. Cells were harvested within 10 to 14 days when they reached an optical density at 600 nm of ca. 1.0 to 1.5 (1-cm cuvette), measured with a Beckman DU-50 spectrophotometer (Beckman Instruments, Fullerton, Calif.).

**Nitrogenase purification and N-terminal sequence analysis.** Both nitrogenase components were purified anaerobically as previously described (28), with some minor modifications (10). Briefly, the first step was changed from protamine precipitation (28), which provided variable results, to a 0 to 0.5 M NaCl step gradient using an Econo-Pac Q column (Bio-Rad, Richmond, Calif.), followed by desalting on an FPLC (fast protein liquid chromatography) Fast desalting column. Desalted extract from this step was loaded on a Poros Q/M perfusion anion-exchange column (PerSeptive Biosystems), and samples were eluted with an NaCl gradient, using an FPLC apparatus (Pharmacia LKB Biotechnologies, Piscataway, N.J.). Fractions containing nitrogenase components were further purified with a Superose 12 column. Once purified, the nitrogenase components were labile as previously described (28). The purified component 2 was subjected to sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) using a Hoefer Minigel apparatus (Hoefer Scientific Instruments, San Francisco, Calif.) and transferred to a Bio-Rad polyvinylidene difluoride membrane, and the N-terminal sequence of the purified component 2 was determined by the Edman degradation performed at the Cornell Biotechnology Center Analysis Facility.

**Western blot analysis.** Western immunoblot analysis was performed as described by Maniatis et al. (31). A rabbit polyclonal serum raised against component 2 of R. rubrum was kindly provided by Paul Ludden, University of Wisconsin, and was diluted 1:5,000 for use. After incubation with the antisera for 30 min, the Zeta-Probe membrane (Bio-Rad Laboratories) was washed and placed in a sealable bag containing a 1:5,000 dilution of alkaline phosphatase-coupled goat antirabbit antibody (Promega, Middletown, Wis.) and gently shaken for 30 min at room temperature. After the membrane was washed, the immunoreactive bands were visualized by soaking the membrane in 10 ml of alkaline phosphatase substrate (Promega) containing nitroblue tetrazolium and 5-bromo-4-chloro-3-indoly phosphate. The color development was stopped by rinsing the membrane with water. A set of prestained SDS-PAGE standard proteins (Bio-Rad Laboratories) was used for molecular weight estimation.

**Cloning and nucleotide sequencing.** A fragment containing a portion of M. Barkeri nifH2 was amplified by PCR using oligonucleotides corresponding to positions 150 to 174 and 902
to 925 of the published \textit{nifH2} sequence from \textit{M. barkeri} (44) (see Fig. 5). With these oligonucleotides, a 785-bp-long nucleotide fragment was obtained after PCR using a Hybaid thermal cycler (Labnet, Woodbridge, N.J.). The 5' oligonucleotide had an extra nine bases at its 5' end, giving a \textit{HindIII} site and a three-base cap, and there was an \textit{AccI} site at position 774; thus, the fragment was cleaved with these enzymes and was cloned into pUC19. To ensure that it was the desired fragment, the 3' and 5' ends of this cloned fragment were sequenced by the dideoxy-chain termination method (42) (see below). The PCR fragment was labeled with digoxigenin (DIG) by using the Genius random priming labeling system (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) and was used in Southern blot, RNA dot blot, and colony hybridization experiments. The hybridized DNA in these experiments was detected by the Genius chemiluminescence detection kit (Boehringer Mannheim) as described in the accompanying instructions.

\textit{M. barkeri} DNA was digested with \textit{BamHI}, and the resulting fragments were ligated into \textit{BamHI}-phosphatase-treated pUC19. DNA was transformed into \textit{E. coli} DH5\textalpha, and a \textit{BamHI} library of 1,200 clones was constructed. A positive clone, termed pYTC001, was detected by colony hybridization performed in the presence of 50% formamide at 42°C, using the DIG-labeled probe and chemiluminescence detection.

DNA sequencing was performed by the dideoxy-chain termination method (42), using a Sequenase DNA sequencing kit (United States Biochemical Corp., Cleveland, Ohio). DNA fragments were subcloned into pUC19, and double-stranded DNA sequencing was performed with universal oligonucleotide primers directed against the plasmid vector and oligonucleotides directed at previously sequenced parts of the cloned DNA.

**RNA isolation.** \textit{M. barkeri} cells (20 ml), grown with either \textit{N}_2 or \textit{NH}_4\textsuperscript{+}, were harvested from exponential-phase cultures and mixed rapidly with 20 g of crushed ice and 15 mM sodium azide. After the ice melted, cells were sedimented at 20,000 rpm for 1 min. The pellet was frozen with liquid \textit{N}_2, and the resulting brittle pellet was ground with a sterile mortar and pestle. The ground cells were resuspended in 6 ml of TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA) containing 250 mM sucrose and were treated with 600 ml of 10% SDS, 660 ml of 5 M \textit{NaCl}, and 800 ml of hexadecyltrimethylammonium bromide-0.7 M \textit{NaCl} (20). The lysate was extracted twice with hot phenol-chloroform (1:1) and once with chloroform-isooamyl alcohol (24:1) and precipitated with ethanol. The precipitated nucleic acids were redissolved in 20 ml Tris-HCl (pH 8.0)-10 mM MgCl\textsubscript{2}-2 mM CaCl\textsubscript{2}-100 mg of RNase-free DNase per ml.

After incubation for 60 min at 37°C, SDS, EDTA, and NaCl were added to final concentrations of 1%, 50 mM, and 0.2 M, respectively, and the remaining nucleic acids (RNA) were purified by phenol-chloroform extraction and ethanol precipitation. Pellets were dried in a SpeedVac SC 110 vacuum concentration system (Savant Instruments, Inc., Farmingdale, N.Y.) for 10 min, and the pellet was carefully resuspended in 100 \textmu l of diethylypyrocarbonate (Sigma Chemical Co., St. Louis, Mo.),-treated H\textsubscript{2}O. RNA samples were immediately frozen at \textdegree70°C.

**Dot blot DNA-RNA hybridization.** RNA samples (0.5 to 10 \textmu l) isolated from \textit{N}_2- or \textit{NH}_4\textsuperscript{+}-grown cells were denatured by adding 0.5 ml of ice-cold 10 mM NaOH-1 mM EDTA. The denatured RNA was then applied onto a sheet of Zeta-Probe membrane (Bio-Rad), using the dot blot apparatus. The membrane was rinsed briefly in 2 \times SSC (0.3 M NaCl, 0.03 M sodium citrate, [pH 7.0])-0.1% SDS and was dried at 80°C for 30 min in a vacuum oven. The mRNA was then detected by hybridization with DIG-labeled \textit{nifH2} or \textit{nifDK2} probe (5 ng/ml) overnight at 42°C. The \textit{nifH2} probe was the cloned product of the PCR, while the \textit{nifDK2} probe was a subclone of the 3' end of the cloned fragment from the \textit{PsrI} site to the \textit{BamHI} site (see Fig. 5). After hybridization, the membrane
was washed twice for 5 min in 2× SSC-0.1% SDS solution at room temperature and then twice for 15 min with 0.5× SSC-0.1% SDS at 65°C. The hybridized mRNA was then detected by the Genius chemiluminescence detection kit as described in the Genius (Boehringer Mannheim) protocol booklet.

**Phylogenetic analysis.** Amino acid sequences were used instead of DNA sequences to eliminate biases due to different G+C ratios (36) and were obtained from the combined GenBank/Swissprot/PIR database except for *M. barkeri* nifH1 and nifH2 and *Methanobacterium ivanovii* nifH2, which were obtained from Sibold et al. (44), and the nifD sequence for *Frankia alni* (36). The amino acid sequence of the nifH2 product from *M. barkeri* (44) was corrected as described in Results.

Amino acid sequences were aligned by using PILEUP, part of the Genetics Computer Group software package (17), followed by a few minor changes made manually. Gaps present in more than five sequences were removed before phylogenetic analysis (49). The PHYLIP 3.5c phylogeny inference software package (16), in the form of compiled executable programs for Macintosh computers, was used for comparison of the protein sequences. The primary sequence analysis was done by using the PROTDIST program with a Dayhoff amino acid comparison matrix. This program gave a distances expressed in expected changes per amino acid position, including multiple forward and back mutations.

The resulting distance matrices were converted to phylogenetic trees by using the program FITCH, which uses the Fitch-Margoliash least-squares distance method matrix and does not assume a constant evolutionary clock. TREEDRAW was used to draw the unrooted phylogenetic trees presented. Unrooted trees produced by KITCH, NEIGHBOR, UPGMA, and PROPARS methods showed topologies essentially identical with those presented here. Dendrograms produced by DRAWGRAM resembled those previously published for a larger set of eubacterial and a smaller set of methanogenous nif genes (36), but since there is no true outgroup sequence, unrooted trees were used in this analysis.

**Nucleotide accession number.** The nucleotide sequences for nifD2 from *M. barkeri* and the first 135 bp of nifK2 have been submitted to GenBank and have received accession number U11291.

## RESULTS

**Expression of the nifH2 gene in *M. barkeri* grown diazotrophically in Mo-containing medium.** Since the predicted products of the nifH1 and nifH2 genes in *M. barkeri* have different N-terminal amino acid sequences (MTRKIAFYGK and MRQIAIYGK, respectively), we determined the N-terminal sequence of purified nitrogenase component 2 from *M. barkeri* grown in standard Mo-containing medium. Figure 2 shows that a single band was obtained in SDS-polyacrylamide gels from purified component 2 and that this band reacted with antibody to *R. rubrum* component 2 as shown previously (28). The first 10 amino acids of this purified polypeptide were identical with the sequence predicted for *M. barkeri* nifH2, indicating that nifH2 was expressed in Mo-containing medium.

We also examined whether mRNA for nifH2 could be detected in cells growing diazotrophically in Mo-supplemented medium. Figure 3 shows an RNA-DNA dot blot analysis demonstrating that RNA hybridizing to a nifH2 probe or a probe containing fragments of both the nifD and nifK genes (see Fig. 5) could be detected in cells grown with N₂, while cells grown with NH₄⁺ only showed background reactions with the probe which did not increase with increasing RNA concentration.

**Cloning of nifDK2 genes from *M. barkeri*.** We generated and cloned a 786-bp fragment of the nifH2 gene by using PCR (see Materials and Methods and Fig. 5) and used this fragment as a probe to clone downstream regions, since nifD genes are often downstream of nifH (11). To verify that the PCR product was the correct sequence, we sequenced its 5' and 3' ends. The sequence clearly was from the nifH2 gene, but there were a few discrepancies between the sequence that we obtained and the published sequence (44), including a major one unlikely to be a PCR artifact since it involves the addition of three bases rather than a simple substitution. Starting with nucleotide 236 in the published sequence, the following sequence was given: 5'-GGGTGGGTAATGAGAAAAATCC-3', which encodes GWELEKS, amino acids 10 to 16 of the nifH2 gene product. For the same region, we found a sequence lacking the three underlined nucleotide bases, encoding GGIGKS. The latter amino acid sequence is conserved in all other nifH sequences described to date and is believed to play a role in nucleotide binding (11, 18, 40).

The DIG-labeled nifH2 PCR product was used as a probe to detect hybridizing fragments in Southern blots of restriction enzyme-digested *M. barkeri* DNA. As shown in Fig. 4, a 1.85-kb fragment from HindIII-digested DNA hybridized with the probe, as expected from previous results (44). A 4.7-kb fragment was detected in BamH1 digests. We cloned this 4.7-kb BamHI fragment of *M. barkeri* DNA into pUC19, calling the new construct pYTC001. A map of the cloned fragment based on restriction analysis and subsequent sequencing analysis is shown in Fig. 5. The clone contains at least five ORFs which

---

**FIG. 5.** Restriction map of a BamHI fragment from *M. barkeri* containing nifHDK2 genes cloned into pUC19(pYTC001). The original clone of Sibold et al. (44) is defined by the internal HindIII fragment. The DNA fragment (PCR Oligos) used as a probe in the hybridizations presented in Fig. 3 and 4 is indicated by arrows. B, BamHI; H, HindIII; E, EcoRI; P, PstI; A, AccI; Hp, HpaI; Kp, KpnI.
The predicted nifD2 ORF is 1,596 bp long, begins using the final A of the UAA stop codon of ORF125, and has 5′-AGAG-3′, which is within ORF125. The G+C content of nifD2 was 43.7%, close to the value of 44.8% for the nifH2 gene and 42% for the genomic DNA in M. barkeri (1). The ORF encodes a polypeptide 532 amino acids long with a predicted molecular mass of 60.3 kDa, close to the larger of the subunit molecular masses (57 and 62 kDa) detected by PAGE for the purified component 1 from M. barkeri (28). There is an 11-bp intergenic region after the nifD2 ORF followed by the first 135 bp of the nifK2 ORF. A GAGG sequence 8 bp upstream of the initial AUG codon of nifK2 also provides a potential ribosome-binding site.

Figure 7 is an alignment of the amino acid sequences for the predicted nifD2 gene product in M. barkeri and selected other nifD gene products. The predicted amino acid ligands for the FeMo-cofactor (11, 22, 23) are present in the predicted M. barkeri nifD2 product, as are cysteine ligands which can coordinate the P cluster. Most noticeably, both the M. barkeri nifD2 and C. pasteurianum nifD1 deduced amino acid sequences have an insert of approximately 50 amino acids not present in other nifD products. The sequence identity in these inserts was near 33%, while that for the entire polypeptides is near 51%, and areas of identity and similarity could be detected throughout these products.

Figure 8 shows an alignment of the first 45 amino acids of the predicted product of the M. barkeri nifK2 ORF with sequences of other nifK products. Both the M. barkeri and C. pasteurianum sequences lacked a long N-terminal region present in the A. vinelandii and Frankia sp. nifK products (as well as other eubacterial nifK products (11). There was a high degree of sequence identity in the N terminus of the M. barkeri and C. pasteurianum gene products, as well as the other nifK products.

Phylogenetic analysis of the nifD2 product from M. barkeri.

Figure 9 shows that the equivalents of clusters I, II, and III in Fig. 1 can also be defined for nifD gene products. No corresponding nifD products have been described for nif cluster IV. As was the case for nifH1 products, the nifD2 product from M. barkeri clustered with the nifD1 product of C. pasteurianum. As has been noted before (6, 48), the nifD1 product from M. thermodithiotrophicus showed much less association with the alternative nitrogenases than did the nifH1 product. All methods of phylogenetic analyses used (see Materials and Methods) gave similar tree topologies, and all showed association between nifD2 from M. barkeri and nifD1 from C. pasteurianum.

Analysis of the 45-amino acid N-terminal fragment of the M. barkeri nifK2 gene product also showed a close association with nifK1 from C. pasteurianum (data not presented).

Discussion

N-terminal sequencing analysis of purified nitrogenase component 2 and RNA-dot blot analysis both indicate that the nifHDK2 genes were expressed in M. barkeri 227 cells grown in Mo-supplemented medium. While our results do not eliminate the possibility of expression of nifHDK1 genes under these conditions, the N-terminal sequence data indicate that the nifH2 gene product was responsible for most of the detectable (albeit low) nitrogenase activity detected in extracts. This is consistent with the similarity between nifH2 in M. barkeri and nifH1 in C. pasteurianum, which encodes the Fe-protein for its Mo-nitrogenase. An attractive hypothesis is that nifH1 in M. barkeri encodes the Fe-protein of a V- or Fe-nitrogenase. V stimulation of diazotrophic growth of M. barkeri Fusaro and 227 has been reported (4).

The arrangement of the nifHDK2 genes in M. barkeri (Fig. 5) shows similarity to the arrangement of nifHDK1 genes in M.
FIG. 7. Alignment of selected nifD amino acid sequences and sequence similarity between M. barkeri nifD2 and C. pasteurianum nifD1. Organism names are as in Fig. 1. A, residues which are identical in all nifD products; *, residues which are considered to be conserved in nifD sequences from eubacterial Mo-nitrogenases, including C. pasteurianum (11); M, residues which are FeMo-cofactor ligands; P, residues which are P-cluster ligands; I, identical amino acid residues; *, conservative substitutions (I, L, V, M; D, E; K, R, H; S, T; G, A; F, Y; and N, Q).
barkeri, M. thermolithotrophicus, and M. ivanovii, suggesting a methanogen-specific gene arrangement. All of these clusters have two small ORFs located between nifH and nifD encoding products 105 and 122 to 128 amino acids long (44, 48). The deduced products of both of these ORFs show sequence similarity to the glnB gene product in gram-negative bacteria (44). glnB encodes the Pn protein, which plays a role in the regulation of transcription of genes and activity of enzymes involved in nitrogen metabolism (39), suggesting a similar regulatory role for the products of these ORFs in methanogens. While the orders of these genes are similar, the spacings are not identical. For example, in the case of M. thermolitho-
throphicus (48), there was an 8-bp overlap between nifD1 and nifK1 and almost 200 bases between ORF 128 and nifD2, while in M. barkeri, there was no overlap between nifD2 and nifK2 but there was a 1-bp overlap between ORF 125 and nifD2.

The presence of nifH mRNA only in N2-grown cells is consistent with repression of nifH by ammonia, as has been found in other free-living diazotrophs (33). In eubacteria, there is a specific regulatory sequence, often called the -24, -12, or σ54 consensus sequence, typically found upstream of nitrogen-related genes (33). No evidence for this element could be found upstream of M. barkeri nif genes, as is the case for other methanogen nif genes (44, 46, 47). In terms of potential archaeal promoters, in the 209 nucleotides preceding the nifH2 gene (Fig. 3), the closest matches to a consensus methanogenic archaeal promoter of 5'-TTA(T/A)ATA-3' (38) are 5'-TTA TAAA-3' at 72 bp upstream and 5'-TTATTTA-3' at 137 bp upstream of the start of the nifH2 gene. There appears to be no promoter sequence upstream of the nifD2 gene, similar to the M. thermolithotrophicus nifD1 gene, which lacks an upstream promoter sequence even though it was apparently not cotranscribed with nifH1 (48).

It has long been recognized that the C. pasteurianum Mo
itrogenase is divergent from other eubacterial Mo-nitro-
genases. C. pasteurianum nitrogenase has a higher K_i for inhibition of N_2 reduction by H_2 (30) and shows different specificities for nucleotides and nucleotide analogs (54). C. pasteurianum nitrogenase components fail to complement components from other eubacterial Mo-nitrogenases and can form inactive complexes with them (13). Previous analyses of sequences of nif structural genes in C. pasteurianum nitrogenase show them to be highly divergent from other eubacterial Mo-nitrogenases (11, 35, 56), including those from Frankia sp., a eubacterium in the high-G+C branch of the gram-positive line (55). It was recently proposed that the C. pasteurianum nitrogenase represents a separate gene family (36).

Our analysis has shown that the nifH2 and nifD2 gene products and the N-terminal 45 amino acids of the nifK2 products from M. barkeri all show closest similarity to the corresponding nifHDK1 gene products in C. pasteurianum. While it is not surprising that nifD2 from M. barkeri should show the same phylogenetic affiliations as nifH2, this is not always the case, since nifD2 in A. vinelandii, nifD in Frankia sp., and nifD1 in M. thermolithotrophicus all show different branch positions from their corresponding nifH gene products (Fig. 1 and 9) (6, 36, 41, 48). The most striking similarity between the M. barkeri nifD2 and C. pasteurianum nifD1 sequences is the presence of an insert of approximately 50 amino acids near position 403, an insert heretofore found only in the C. pasteurianum nifD1 gene product (9). Kim et al. (23) showed that the insert forms a loop spanning a site above the FeMo-cofactor in component 1 considered to be a part of a region of contact between the two components. This loop causes the cofactor to be buried more deeply in the C. pasteurianum nitrogenase than in others, and the poor reactivity of the C. pasteurianum nitrogenase components toward other nitrogenases has been attributed to the presence of this loop (23). Of the amino acids in the insert which are solvent exposed and are proposed (23) to interact with the Fe-protein (Lys-412, Asp-414, Asp-416, and Asn-419 in Fig. 7), only Lys-412 (Lys-385 in the C. pasteurianum sequence) is conserved in M. barkeri, making it a likely candidate for interaction with component 2.

We have examined the phylogenetic proximity of the M.
barkeri nifHD gene products to nif clusters I and II. Using the distances derived from the PROTDIST program, the M. barkeri nifH2 gene product shows the closest similarity to the nifH1 product from C. pasteurianum (0.36 amino acid change per position) and is roughly equidistant from the eubacterial Mo-nitrogenases (cluster I) (0.487 ± 0.034 [standard deviation] amino acid change per position) and the alternative nitrogenase cluster (cluster II) (0.455 ± 0.027 change per position). For the nifD2 gene product from M. barkeri, the distance from C. pasteurianum nifD1 is 0.75 change per position, while that from cluster I is 1.02 ± 0.07 change per position. The distances from cluster II were 1.46 ± 0.10 when the nifD1 product from M. thermolithotrophicus was included in the cluster and 1.50 ± 0.05 if it was omitted. Similar results were obtained for nifD1 genes from C. pasteurianum. Thus, the
niFD gene products from the M. Barkeri-C. pasteuriunam cluster show more affinity with eubacterial Mo-nitrogenases than with alternative ones, consistent with previous analysis of C. pasteuriunam (11, 35, 36, 56).

One can advance various evolutionary scenarios to explain this similarity between the sequences of proteins between two organisms, M. Barkeri and C. pasteuriunam, which, according to phylogenetic analysis based on 16S rRNA sequences (55), have not had a common ancestor for over 3 billion years. One possible explanation is horizontal transfer of genes, which is given some credence by the fact that clostridia and methanogens are likely to be found in close physical proximity in anaerobic sediments. Arguing against a relatively recent transfer of genes between the two genera is the difference in the G+C values of the genes (ca. 35% for C. pasteuriunam and ca. 44% for M. Barkeri) and the presence of ORF105 and ORF125 located between the niFD and niF2 genes in M. Barkeri but absent in C. pasteuriunam (9). It should be mentioned that the five nearly identical copies of niFH1 in C. pasteuriunam (53) are indicative of significant genetic rearrangement in this organism.

An alternative hypothesis, similar to one put forward by Young (56), is that the ancestor of the archaea and the eubacteria had two sets of niF genes, one encoding a Mo-nitrogenase resembling the present-day clostridial nitrogenase and the other resembling an alternative nitrogenase. Cluster I would represent a more recent radiation within the eubacteria, most likely from the ancestor of the clostridial nitrogenase. It is not clear where the extremely divergent cluster IV niFH genes in methanogens fit in the evolution of niF genes. That they may encode products which are not nitrogenases is suggested by their high degree of divergence relative to other niFH genes, the lack of niFD- or glbB-like ORFs found downstream from them, and the inability to detect nitrogen fixation by M. voltae (29, 47) or expression of niFH2 in M. thermolithotrophicus (48). Recently, it was shown that the iron proteins of protocorophyllide and chlorine reductases involved in bacterochlorophyll synthesis show significant sequence similarity to niFH gene products (8).

There are other instances of similarity between gene sequences from archaea and members of the low-G+C gram-positive branch of the eubacteria, including the sequences of two enzymes which play a central role in nitrogen metabolism: glutamine synthetase I (25, 52) and glutamate dehydrogenase (4, 12). More than nitrogen-related genes may be involved in this phenomenon, since similarity between heat shock protein-70 sequences between archaea and gram-positive bacteria has also been found (19). This similarity is not universal, since other archaeal genes, such as those involved in tryptophan synthesis in methanogens (32), do not show any affinity with those in gram-positive organisms. As above, one can invoke gene transfer between the groups (or perhaps cell fusion) or a common heritage of duplicated copies of the genes in question. The impact of these findings, including those reported here, on our understanding of bacterial phylogeny is unclear at this point but should become clearer as more gene sequences from diverse organisms become available.

ACKNOWLEDGMENTS

This research was supported by USDA competitive grant No. 90-37120-2200 and by USDA Hatch funds.

REFERENCES


