

Characterization of the Upstream Region of the Formate Dehydrogenase Operon of *Methanobacterium formicicum*

PRAMATHESH S. PATEL AND JAMES G. FERRY*

Department of Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Received 22 February 1988/Accepted 12 May 1988

The *fdhA* and *fdhB* genes of *Methanobacterium formicicum*, which code for the α and β subunits of formate dehydrogenase, were cotranscribed as part of a large transcript. By using Northern (RNA) gel blot analysis, the transcription start site was located within a 1.6-kilobase *BglII-NcoI* fragment 4.3 kilobases upstream from the *fdhA* gene. The precise transcription start site within the fragment was determined with the aid of primer extension analysis and S1 nuclease protection studies. A putative promoter sequence for structural genes of methanogenic archaeobacteria is proposed based on a comparison of DNA sequences of the upstream region of methanogen operons for which transcription initiation sites are known. Comparison of the DNA sequence of the upstream region of the *fdh* operon of *M. formicicum* with the sequence upstream of the *fdhF* gene of *Escherichia coli* revealed regions of considerable identity.

Methanogens are taxonomically classified as archaeobacteria which are phylogenetically and physiologically distinct from eubacteria and eucaryotes (2, 10, 11). Initial studies indicate that methanogen genes are arranged in multigene transcriptional units similar to eubacterial operons (6, 8, 9, 20, 21, 25). However, the subunit composition of the archaeobacterial DNA-dependent RNA polymerases are different from the eubacterial enzymes (12, 13, 27), which suggests that the structures of the methanogen and eubacterial promoters may also be different. A consensus promoter sequence has been proposed for stable RNA genes of *Methanococcus* spp. that is distinct from the eubacterial consensus (14, 29, 32-34). The RNA polymerase binding site and the transcription initiation site for the polypeptide-encoding *hisA* gene and the *mcr* operon of *Methanococcus vannielii* have been determined, and the results indicate that both promoters of stable RNA and structural genes contain a common sequence of TATA (7, 28).

Methanobacterium formicicum synthesizes formate dehydrogenase, an iron-sulfur enzyme containing molybdopterin and flavin adenine dinucleotide (17, 23, 24). The genes encoding the two subunits (*fdhAB*) are cloned and sequenced (25). The activity of this enzyme is regulated in response to the amount of molybdate in the growth medium (17). The levels of *fdh*-specific message are inversely proportional to the intracellular concentration of molybdenum and the amount of formate dehydrogenase protein (17). Here we report that the *fdhA* and *fdhB* genes are cotranscribed as part of an approximately 12-kilobase (kb) operon. The transcription start site and the sequence upstream of it were determined.

MATERIALS AND METHODS

Bacterial strains and plasmids. *M. formicicum* JF-1 was grown in medium with or without sodium molybdate supplementation as described previously (17). Plasmid pUCFD18, a 10.6-kb fragment of *M. formicicum* DNA containing the *fdh* genes cloned into pUC18 (25), was a gift from P. Schendel, Genetics Institute, Boston, Mass. Plasmid pAR21 was generated by cloning the 2.0-kb *SalI-EcoRI* fragment of

pUCFD18 into the multiple cloning site of pUC19 (see Fig. 1 and 4).

DNA manipulations. DNA manipulations were performed by using established procedures (3, 16), except ³²P radiolabeling of probes was performed with commercially available kits for random priming (Pharmacia, Inc., Piscataway, N.J.) or end labeling (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) as described by the manufacturers. Radiolabeled nucleotides were purchased from New England Nuclear Laboratories, Boston, Mass. Double-stranded DNA sequencing, using the dideoxy chain termination procedure, was as described previously (19). The avian myeloblastosis virus reverse transcriptase and Sequenase enzymes used for sequencing were purchased from Boehringer Mannheim Biochemicals and United States Biochemical Corp., Cleveland, Ohio, respectively. Erase-A-Base kit (Promega Corp., Madison, Wis.) and pUC19 vector (30) were used to generate clones in DNA sequence analysis.

RNA isolation and Northern (RNA) gel blot analysis. RNA was isolated from cells as described previously (17). Northern blot analysis was performed as described previously (18). In brief, 10 μ g of glyoxylated total cellular RNA was electrophoresed in a 1.1% agarose gel. The RNA was electroblotted from the gel onto a GeneScreen hybridization membrane (New England Nuclear). Prehybridization and hybridization were performed as recommended by the manufacturers, except that hybridization was performed at 45°C for 18 to 24 h. After hybridization, the filters were washed three times at 20°C for 5 min in 2 \times SSC (2 \times SSC = 0.3 M sodium chloride plus 0.03 M sodium citrate) containing 0.1% sodium dodecyl sulfate followed by two 30-min washes at 65°C in the same solution and finally rinsed in 0.2 \times SSC. The filters were exposed for 18 to 30 h at -80°C.

Primer extension analysis and S1 protection. Primer extension analyses were performed as described previously (1) except [α -³⁵S]dATP (2.0 μ M) was used instead of [α -³²P]dATP. A 21-nucleotide DNA primer (5'-CGTCCTTGAAT TCAAAGAGG-3') complementary to the antisense strand of the *fdh* operon and contained within the coding region of the 1.6-kb *BglII-NcoI* fragment was used (Fig. 1). For each reaction, a total of 10 ng of the primer and 10 μ g of total *M. formicicum* RNA was used. A sequencing reaction performed with the above primer and pAR21 (see Fig. 4) was

* Corresponding author.

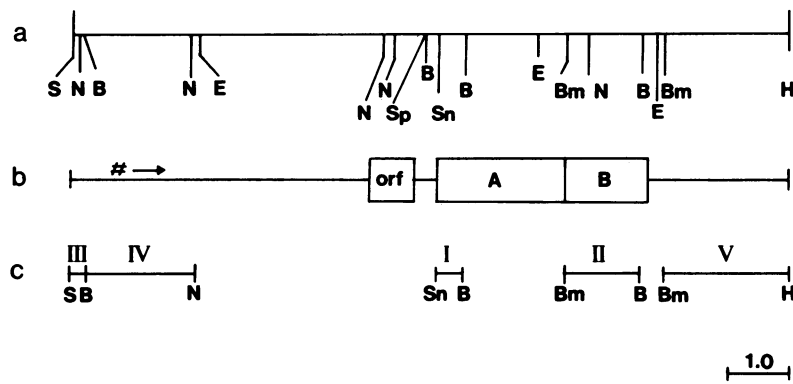


FIG. 1. *M. formicicum* DNA insert in plasmid pUCFD18. (a) Restriction map of the 10.6-kb fragment of *M. formicicum* DNA containing the *fdh* genes. (b) Locations of the *fdhA* and *fdhB* genes, the open reading frame (ORF) based on partial sequencing of the cloned fragment (25; unpublished results), the transcription initiation site (#), and the direction of transcription (arrow) of the *fdh* operon. (c) Fragments used as probes in Northern gel blot analyses. Bar, 1.0 kb. Abbreviations used for the restriction enzyme cleavage sites are as follows: S, *SalI*; B, *BglII*; N, *NcoI*; E, *EcoRI*; Sp, *SphI*; Sn, *SnaBI*; Bm, *BamHI*; H, *HindIII*.

electrophoresed in adjoining lanes. S1 nuclease protection studies were performed as described previously (31). A DNA probe was synthesized with [α - 35 S]dATP, the primer described above, and pAR21 (see Fig. 4). The probe and total cellular RNA (100 μ g) were hybridized at 42°C for 12 h. S1 nuclease digestion (2,260 U; Miles Scientific, Div. Miles Laboratories, Inc., Naperville, Ill.) was performed at 23°C for 30 min. The protected fragment and a sequencing reaction performed with the probe described above and pAR21 (see Fig. 4) were electrophoresed in adjoining lanes.

RESULTS

Cotranscription of the *fdhA* and *fdhB* genes. The *fdhA* and *fdhB* genes, encoding the α and β subunits of formate dehydrogenase from *M. formicicum*, overlap by 1 base pair (bp) (25); thus, Northern gel blot analyses were performed to determine whether the genes are cotranscribed. The DNA probes were restriction endonuclease fragments from either the *fdhA* or the *fdhB* gene (probes I and II, Fig. 1). Both *fdh*-specific probes hybridized to an approximately 12-kb transcript (Fig. 2). The relative amounts of transcript between cells grown in media with or without molybdate supplementation were in general agreement with that previously reported (17). The results (Fig. 2) indicated that genes *fdhA* and *fdhB* were cotranscribed as part of a large message. The large size of the transcript is unprecedented; thus, to ensure that the mRNA was not complexed with proteins or glycolipids (which may have retarded migration during electrophoresis), the RNA samples were sequentially extracted with phenol, chloroform, ether, isobutanol, and isoamyl alcohol. Northern gel blot analyses of these RNA samples showed no smaller transcripts; however, smearing of the signal increased with each successive extraction, indicating progressive degradation of the large transcript (data not shown).

Localization of the transcription start site. To locate the approximate transcription start site, DNA fragments upstream and downstream from the *fdhA* and *fdhB* genes (Fig. 1) were used as probes in Northern gel blot analyses with total cellular RNA. The *BglIII-NcoI* fragment upstream from the *fdhA* gene (probe IV, Fig. 1) and the *BamHI-HindIII* fragment downstream from the *fdhB* gene (probe V, Fig. 1) hybridized to a message of the same size as that to which the *fdhA* and *fdhB* gene probes hybridized (Fig. 3). Nine dif-

ferent overlapping DNA fragments contained between fragments I and IV (Fig. 1), when used as probes, gave the same results as shown in Fig. 2 (data not shown). However, the *SalI-BglII* fragment (probe III, Fig. 1) did not hybridize to the large message, but instead hybridized to two smaller messages of approximately 1.3 and 3.0 kb. The relative

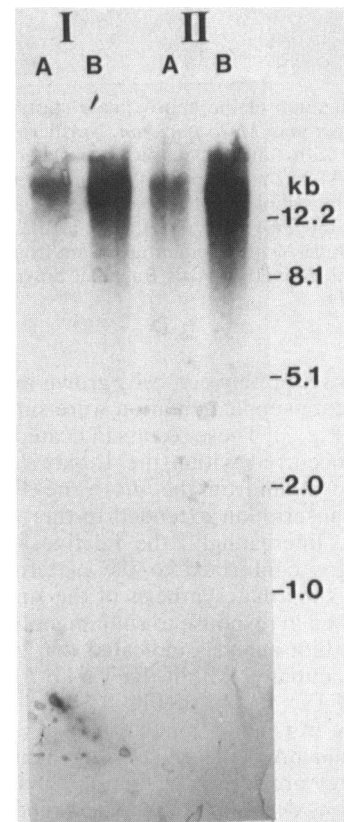


FIG. 2. Estimation of the size of the *fdh* transcript and evidence for cotranscription of the *fdhA* and *fdhB* genes. Northern gel blot analyses were performed with total cellular RNA isolated from cells (17) containing (lane A) 2.22 and (lane B) 0.03 nmol of molybdenum per mg of protein. The radiolabeled probes were (I) the 0.5-kb *SnaBI-BglII* fragment of the *fdhA* gene and (II) the 1.3-kb *BamHI-BglII* fragment of the *fdhB* gene (Fig. 1).

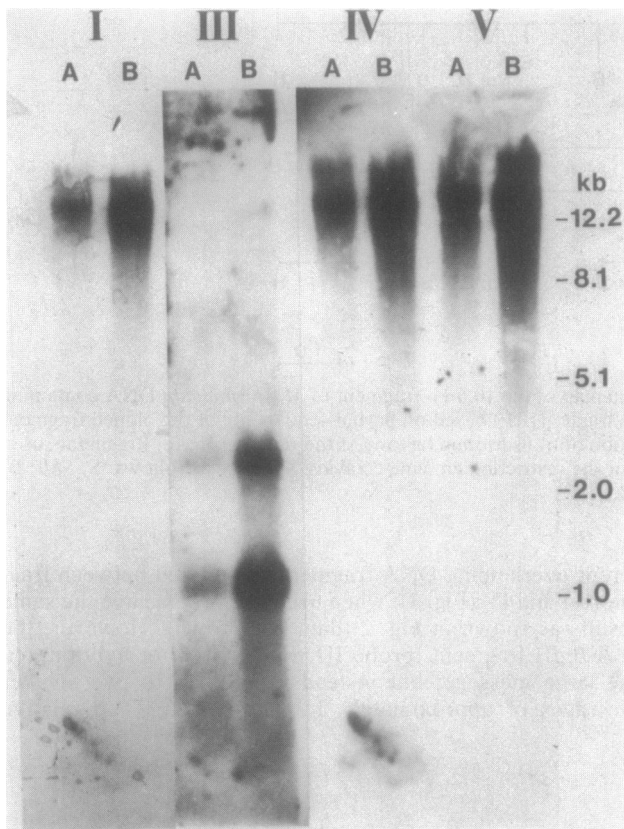


FIG. 3. Localization of the approximate start site of transcription for the *fdh* operon of *M. formicicum*. Northern gel blot analyses were performed with total cellular RNA isolated from cells (17) containing (lane A) 2.22 and (lane B) 0.03 nmol of molybdenum per mg of protein. The radiolabeled probes were (I) the 0.5-kb *Sna*BI-*Bgl*II fragment of the *fdhA* gene, (III) the 0.2-kb *Sall*-*Bgl*II fragment, (IV) the 1.6-kb *Bgl*II-*Nco*I fragment upstream from the *fdhA* gene, and (V) the 2.2-kb *Bam*HI-*Hind*III fragment downstream from the *fdhB* gene (Fig. 1).

amounts of transcript between cells grown in media with or without molybdate supplementation were similar (17) to the large message (Fig. 3). These results indicated that transcription initiation occurred within the 1.6-kb *Bgl*II-*Nco*I fragment located upstream from the *fdhA* gene (Fig. 1, fragment IV) and that transcription extended further downstream of the *fdhB* gene. Interestingly, the relative amounts of the smaller messages conformed to the pattern of the larger message, suggesting that synthesis of the smaller messages was also regulated in response to sodium molybdate (Fig. 3).

Northern gel blot analysis indicated that transcription of the *fdh* operon initiated within the 1.6-kb *Bgl*II-*Nco*I fragment (fragment IV) shown in Fig. 1; thus, the nucleotide sequence of the fragment was determined. The insert contained in plasmid pAR21 used for sequencing and the sequencing strategy are shown in Fig. 4. The deletion clones were generated as described in Materials and Methods.

Primer extension analyses were used to locate the start site(s) of transcription to a single nucleotide within the 1.6-kb *Bgl*II-*Nco*I fragment. Total cellular RNA was isolated from cells grown in media with or without molybdate supplementation (17). The results indicate multiple transcription initiation sites (arrows, Fig. 5). The 5' end of the transcript was also determined by S1 nuclease protection studies, using

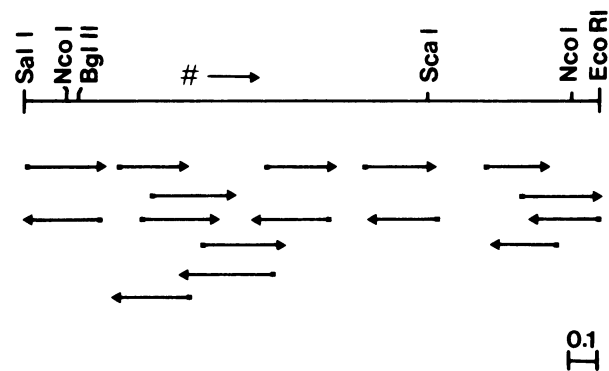


FIG. 4. *M. formicicum* DNA insert in plasmid pAR21. The top line shows a restriction map of the 2.0-kb *Sall*-*Eco*RI fragment cloned into pUC19 to generate pAR21; (#) denotes the location of the transcription start site of the *fdh* operon. The direction of transcription is left to right. The base of the arrows indicate the beginning of the insert in each of the deletion clones obtained from pAR21. The arrow indicates the direction of sequencing and the length of the sequence determined. Bar, 0.1 kb.

total cellular RNA isolated from cells grown in media with or without molybdate supplementation (Fig. 6). In both cases a single major species was protected which indicated that transcription initiation occurred at an adenine on the template strand, one of the predominant start sites identified in the primer extension experiment. S1 nuclease protection studies were also performed to determine whether the transcription initiation occurred within the intergenic region between the open reading frame and gene *fdhA* (Fig. 1). A [³⁵S]dATP-radiolabeled DNA fragment complementary to the mRNA strand, and spanning the region between the open reading frame and *fdhA* (Fig. 1), was used as a probe. The entire length of the probe was protected (data not shown), which, in conjunction with the Northern gel blot results, indicates that transcription initiation did not occur between the open reading frame and *fdhA*.

DISCUSSION

The results show that the *fdhA* and *fdhB* genes of *M. formicicum* are cotranscribed as part of a message approximately 12 kb in length. Partial sequencing has indicated at least one other apparent polypeptide-encoding open reading frame which was transcribed as part of the message (unpublished results). The formation of active formate dehydrogenase is likely to require proteins involved in molybdenum transport and processing and enzymes involved in molybdopterin cofactor synthesis. Although no conclusions can be drawn from the results presented here, the large transcript could potentially code for many or all of these proteins. In the only other study on the regulation of expression of a methanogen gene, Sment and Konisky (26) reported three *hisA*-hybridizing transcripts in aminotriazole-treated cultures of *Methanococcus voltae* that were approximately 1.5, 9, and 10 kb in length. Studies of the bacterio-opsin (*bop*) operon of the archaeobacterium *Halobacterium halobium* has led the authors to speculate that the *bop* mRNA may be processed from a larger mRNA transcript (4). Only one large transcript was detected when DNA probes from within the *fdh* operon from *M. formicicum* were used; however, it cannot be ruled out that smaller transcripts with very short half-lives were processed from the large 12-kb transcript.

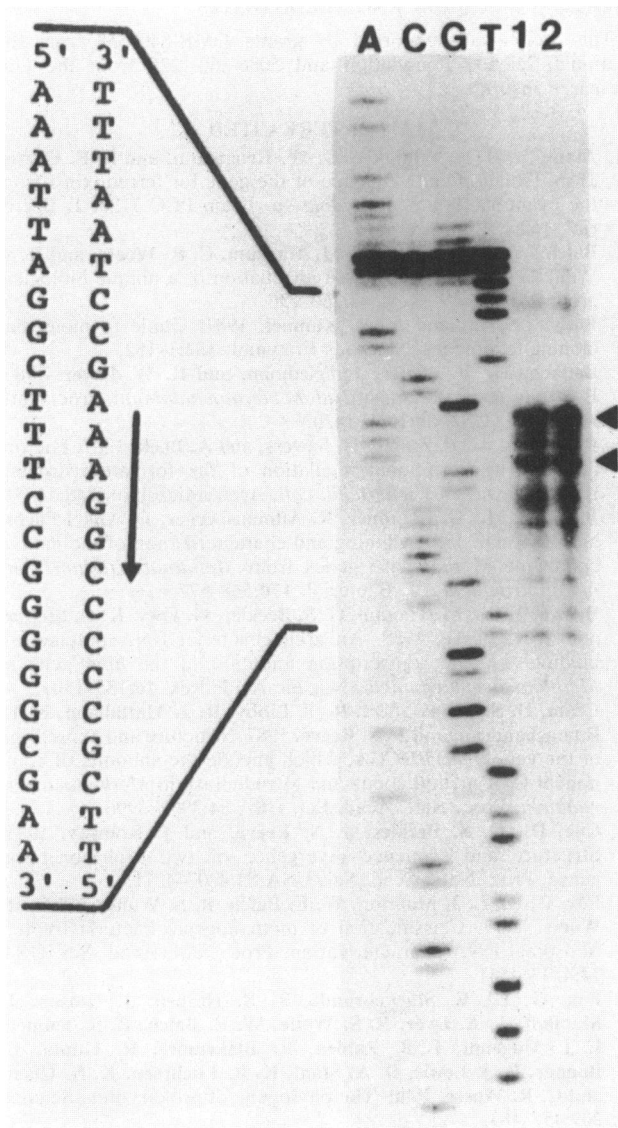


FIG. 5. Primer extension analyses of the transcriptional start of the *fdh* operon of *M. formicicum*. Lanes A, C, G, and T are the sequence ladder. Total cellular RNA isolated from cells (17) containing 2.22 (lane 1) and 0.03 (lane 2) nmol of molybdenum per mg of protein was used as the template. The two predominant transcription start sites are as indicated (▶). The arrow indicates the direction of transcription.

Figure 7 compares the regions upstream of the transcriptional start sites for the *fdh* operon from *M. formicicum*, the *mcr* operon and the *hisA* gene from *Methanococcus vannielii*, and the *mcr* operon from *Methanobacterium thermoautotrophicum* (6, 7, 28). All of the regions contained an A+T-rich sequence 19 to 21 nucleotides preceding the transcriptional start sites. Three of the four upstream regions contained a major conserved feature, C----ATATA--T. The upstream regions compared were from operons encoding functionally diverse enzymes and from organisms representing two of the three methanogen taxonomic orders; thus, a derived sequence is presented in Fig. 7 which may represent general features of promoters for structural genes from the methanogenic archaeobacteria. This sequence also has notable identity with box A of the consensus sequence proposed

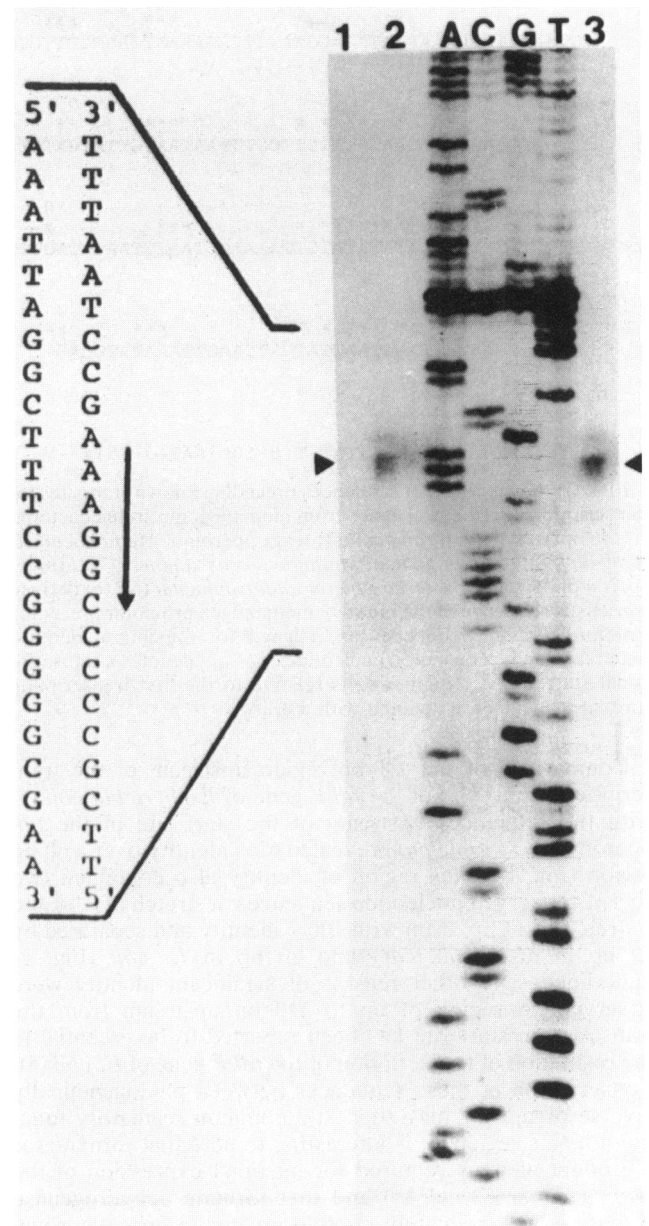


FIG. 6. S1 nuclease mapping of the 5' end of the *fdh* mRNA of *M. formicicum*. Lane 1, No RNA; lane 2, RNA isolated from cells (17) containing 2.22 nmol of molybdenum per mg of protein; lane 3, RNA isolated from cells containing 0.03 nmol of molybdenum per mg of protein. Lanes A, C, G, and T are the sequencing ladder. The first nucleotide in the protected fragment is as indicated (▶). The direction of transcription is indicated by the arrow.

for stable RNA genes from *Methanococcus* spp. (32-34), ACCGAAA-TTTATATA-TA (box A)----18-19 bp----TG CaaagT (box B) where transcription initiation begins at G in box B. The upstream regions shown in Fig. 7 do not have a sequence similar to box B; however, a trinucleotide sequence (5'-AAT-3') occurs immediately preceding the transcription initiation site in all four sequences. The nucleotide sequences of more regions upstream of known transcriptional start sites will be necessary to formulate a consensus sequence for promoters preceding polypeptide-encoding genes of methanogenic archaeobacteria.

15. Krol, A. J. M., J. G. J. Hontelez, B. Roozendaal, and A. van Kammen. 1982. On the operon structure of the nitrogenase genes of *Rhizobium leguminosarium* and *Azotobacter vinelandii*. *Nucleic Acids Res.* **10**:4147-4157.
16. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. May, H. D., P. S. Patel, and J. G. Ferry. 1988. Effect of molybdenum and tungsten on synthesis and composition of formate dehydrogenase in *Methanobacterium formicicum*. *J. Bacteriol.* **170**:3384-3389.
18. McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74**:4835-4838.
19. Mierendorf, R. C., and D. Pfeffer. 1987. Direct sequencing of denatured plasmid DNA. *Methods Enzymol.* **152**:556-562.
20. Reeve, J. N. 1987. Gene structure in methanogenic bacteria. *Poult. Sci.* **66**:927-933.
21. Reeve, J. N., P. T. Hamilton, G. S. Beckler, C. J. Morris, and C. H. Clarke. 1985. Structure of methanogen genes. *Syst. Appl. Microbiol.* **7**:5-12.
22. Schauer, N. L., and J. G. Ferry. 1980. Metabolism of formate in *Methanobacterium formicicum*. *J. Bacteriol.* **142**:800-807.
23. Schauer, N. L., and J. G. Ferry. 1986. Composition of the coenzyme F₄₂₀-dependent formate dehydrogenase from *Methanobacterium formicicum*. *J. Bacteriol.* **165**:405-411.
24. Schauer, N. L., J. G. Ferry, J. F. Honeck, W. H. Orme-Johnson, and C. Walsh. 1986. Mechanistic studies of the coenzyme F₄₂₀-dependent formate dehydrogenase from *Methanobacterium formicicum*. *Biochemistry* **25**:7163-7168.
25. Schuber, A. P., E. C. Orr, M. A. Recny, P. F. Schendel, H. D. May, N. L. Schauer, and J. G. Ferry. 1986. Cloning, expression, and nucleotide sequence of the formate dehydrogenase genes from *Methanobacterium formicicum*. *J. Biol. Chem.* **261**:12942-12947.
26. Sment, K. A., and J. Konisky. 1986. Regulated expression of the *Methanococcus voltae* *hisA* gene. *Syst. Appl. Microbiol.* **7**:90-94.
27. Thomm, M., J. Madon, and K. O. Stetter. 1986. DNA-dependent RNA polymerases of the three orders of methanogens. *Biol. Chem. Hoppe-Seyler* **367**:473-481.
28. Thomm, M., B. A. Sherf, and J. N. Reeve. 1988. The RNA polymerase binding and transcription initiation sites upstream of the methyl reductase operon of *Methanococcus vannielii*. *J. Bacteriol.* **170**:1958-1961.
29. Thomm, M., and G. Wich. 1988. An archaeobacterial promoter element for stable RNA genes with homology to the TATA box of higher eukaryotes. *Nucleic Acids Res.* **16**:151-163.
30. Viera, J., and J. Messing. 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing using universal primers. *Gene* **19**:259-268.
31. Wang, S.-Z., J.-S. Chen, and J. L. Johnson. 1988. The presence of five *nifH* life sequences in *Clostridium pasteurianum*: sequence divergence and transcription properties. *Nucleic Acids Res.* **16**:439-454.
32. Wich, G., H. Hummel, M. Jarsch, U. Bar, and A. Bock. 1986. Transcription signals for stable RNA genes in *Methanococcus*. *Nucleic Acids Res.* **14**:2459-2479.
33. Wich, G., M. Jarsch, and A. Bock. 1984. Apparent operon for a 5S ribosomal RNA gene and for tRNA genes in the archaeobacterium *Methanococcus vannielii*. *Mol. Gen. Genet.* **196**:146-151.
34. Wich, G., L. Sibold, and A. Bock. 1986. Genes for tRNA and their putative expression signals in *Methanococcus*. *Syst. Appl. Microbiol.* **7**:18-25.