Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 4207–4211, June 1985
Genetics

Structure and sequence divergence of two archaebacterial genes

(*methanococcus*/archaebacteria/promoter/homology/translation/evolution)

DAVID CUE*, GREGORY S. BECKLER†, JOHN N. REEVE†‡, AND JORDAN KONISKY*§

*Department of Microbiology, University of Illinois, Urbana, IL 61801; and ‡Molecular, Cellular and Developmental Biology Program, and †Department of Microbiology, The Ohio State University, Columbus, OH 43210

Communicated by Ralph S. Wolfe, March 4, 1985

ABSTRACT **The DNA sequences of a region that includes the *hisA* gene of two related methanogenic archaebacteria, *Methanococcus voltae* and *Methanococcus vannielii*, have been compared. Both organisms show a similar genome organization in this region, displaying three open reading frames (ORFs) separated by regions of very high A+T content. Two of the ORFs, including ORFHisA, show significant DNA sequence homology. As might be expected for organisms having a genome that is A+T-rich, there is a high preference for A and U as the third base in codons. Although the regions upstream of the structural genes contain prokaryotic-like promoter sequences, it is not known whether they are recognized as promoters in these archaebacterial cells. A ribosome binding site, G-G-T-G, is located 6 base pairs preceding the ATG translation initiation sequence of both *hisA* genes. The sequences upstream of the two *hisA* genes show only limited sequence homology. The *M. voltae* intergenic region contains four tandemly arranged repetitions of an 11-base-pair sequence, whereas the *M. vannielii* sequence contains both direct and inverted repetitive sequences. Based on the degree of *hisA* sequence homology, we conclude that *M. voltae* and *M. vannielii* are less closely related taxonomically than are members of the enteric group of eubacteria.**

Methanogens are members of the archaebacteria, a group of prokaryotic organisms exhibiting several properties that are distinct from those seen in either eubacteria or eukaryotes (1, 2). Therefore, it was somewhat surprising when it was found that DNA cloned from several methanogenic species could complement auxotrophic mutations in *Escherichia coli, Bacillus subtilis* (3–6), and *Salmonella typhimurium* (unpublished results). In the case of two methanogens of the genus *Methanococcus* (*vannielii* and *voltae*), it was possible to clone methanogen DNA that complemented the *hisA* mutation of *E. coli* (4, 6). Since the *hisA*-complementing activity was efficiently expressed regardless of the orientation of the cloned DNA fragment within the vector, it seemed likely that the methanogen DNA harbored sequences that were recognized as signals for both transcription and translation. That methanococcal DNA would contain signals for initiation of an early step in translation in *E. coli* was expected because these methanococci have the sequence A-U-C-A-C-C-U-C-C at the 3′ end of their 16S rRNA (1), which is similar to that found at the 3′ end of *E. coli* 16S rRNA (7). This sequence is believed to be important for initial binding of ribosomes to mRNA molecules (8).

Since both *M. voltae* and *M. vannielii* have an overall base composition of ≈70% A+T (1), they should statistically contain those high A+T-rich sequences that often function as promoters in prokaryotes. Whether similar A+T-rich sequences are used as promoters for recognition by methanogen DNA-dependent RNA polymerases is currently unknown. The subunit composition of RNA polymerases from archaebacteria more closely resembles eukaryotic RNA polymerase II than eubacterial RNA polymerase (9), and thus archaebacterial promoters may well be eukaryotic rather than prokaryotic in their sequence organization.

To better understand the molecular basis for expression of the *hisA*-complementing methanogen genes in *E. coli* and to provide basic information for evaluating possible mechanisms of gene expression in methanogens, we have sequenced the *hisA*-complementing genes of both *M. voltae* and *M. vannielii*. The sequences obtained indicate evolutionary divergence of archaebacterial structural genes, support the prediction of ribosomal binding sites, and demonstrate repetitive elements within the intergenic regions.

MATERIALS AND METHODS

Bacterial Strains and Phage. *E. coli* K-12 strain X760, which is a *hisA* auxotroph (4), was used to determine if plasmid derivatives carried *hisA*-complementing DNA.

Enzymes and Chemicals. Restriction enzymes were purchased from Bethesda Research Laboratories, Boehringer Mannheim, or New England Biolabs and were used in accordance with the supplier's directions. Klenow DNA polymerase, T4 DNA ligase, and proteinase K were purchased from Boehringer Mannheim. Deoxy and dideoxynucleotide triphosphates were purchased from Pharmacia and P-L Biochemicals.

DNA Sequencing. For *M. voltae*, purified DNA restriction fragments derived from the plasmid pURB2 [formerly pAW2, (4)] were cloned into the bacteriophage vectors M13 mp8 and M13 mp9 as described by Messing (10) with minor modifications and sequenced using the method of Sanger *et al.* (11). In the case of *M. vannielii*, DNA derived from the plasmid pET805 (6) was sequenced by the Maxam and Gilbert technique (12).

RESULTS

Organization of *his*-Complementing Regions. Analysis of the nucleotide sequence of *M. voltae* and *M. vannielii* complementing DNA revealed a similar organization (Fig. 1). Each contains a central open reading frame (ORFHisA) of a size appropriate to that required to encode the polypeptides that mediate *hisA* complementation (ref. 4; unpublished results). For *M. voltae* this ORF encodes 242 amino acids (M_r 26,347), while in *M. vannielii* it encodes 238 amino acid residues (M_r 25,694). The ORFHisA in *M. voltae* is followed by a 108-base-pair (bp) noncoding region (i.e., containing no ORF) that precedes an ORF (ORF3 in Fig. 1) of 274 bp. Directly preceding the ORFHisA is a 184-bp noncoding sequence that separates ORFHisA from an ORF of 904 bp. In

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); ORF, open reading frame.
§To whom reprint requests should be addressed.

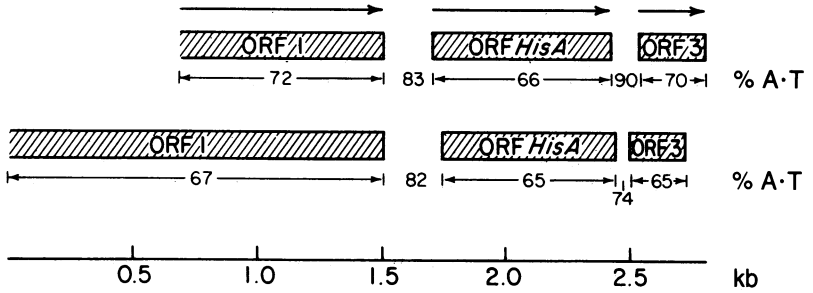

FIG. 1. Organization of the HisA region. The arrows indicate the direction of transcription. The regions sequenced consisted of 2290 bp for *M. voltae* and 2900 bp for *M. vannielii*.

M. vannielii, there is a noncoding sequence of 35 bp immediately downstream from the ORFHisA that is followed by an ORF3 of 230 bp. Upstream of the ORFHisA in *M. vannielii* is a 224-bp noncoding sequence that follows a termination codon for a 1510-bp sequence of in-frame sense codons. An analysis of the opposite DNA strand revealed several ORFs, but these were judged unlikely to be significant, based on their small sizes and lack of ribosome binding sequences. An interesting feature of these sequences is the nonuniform distribution of A·T base pairs. While the overall A+T percentage compositions are 69.3% and 68.9% for *M. voltae* and *M. vannielii* sequences, respectively (1), the regions between the ORFs are even more highly enriched in A·T base pairs (see Fig. 1).

That the ORFs designated ORFHisA (i.e., encoding the *hisA*-complementing activities) do encode this function was established by examining complementation using subclones and deletion derivatives of this region (data not shown). It was found that deletion of *M. voltae* DNA sequences within the ORFHisA region led to loss of complementing activity. In contrast, loss of sequences within either the upstream or downstream ORFs had no effect. Furthermore, insertion of the transposon Tn*5* into this region of the *M. vannielii* DNA eliminated complementing activity. Thus, the indicated ORFHisA is both necessary and sufficient for *hisA*-complementing activity. Because we were able to measure *M. voltae* DNA-dependent *hisA* enzymatic activity in a *S. typhimurium* mutant carrying a deletion in the entire histidine operon (unpublished results), we conclude that the cloned gene encodes an analogous histidine biosynthetic enzyme. Therefore, we believe it is legitimate to designate the methanogen gene, *hisA*.

HisA Sequence Homology. The nucleotide sequences of the *M. voltae* and *M. vannielii* ORFHisAs are compared in Fig. 2. Both are preceded by identical potential ribosomal binding sites, G-G-T-G, at identical positions upstream from the initiation ATG sequence. The *M. voltae* ORF is terminated by tandem ochre (UAA) codons, and the *M. vannielii* frame closes with tandem amber (UAG) codons. The nucleotide sequences are 66% homologous (Fig. 2), which is reflected in conserved amino acid sequences (Fig. 3). In those cases in which amino acid residues differ, there is a strong tendency to preserve the character of the amino acid. Thus, whereas the polypeptides show an absolute amino acid homology of 67%, there is a 76% homology if substitutions that maintain charge, polarity, hydrophobicity, and approximate size are considered to be homologous.

ORF1 and ORF3 Sequence Comparisons. The nucleotide sequences of the ORF1s show extensive regions of homology (Figs. 1 and 4). Determination of additional sequences of the *M. vannielii* DNA in pET805 (6) showed that the ORF extends upstream without interrupting termination codons

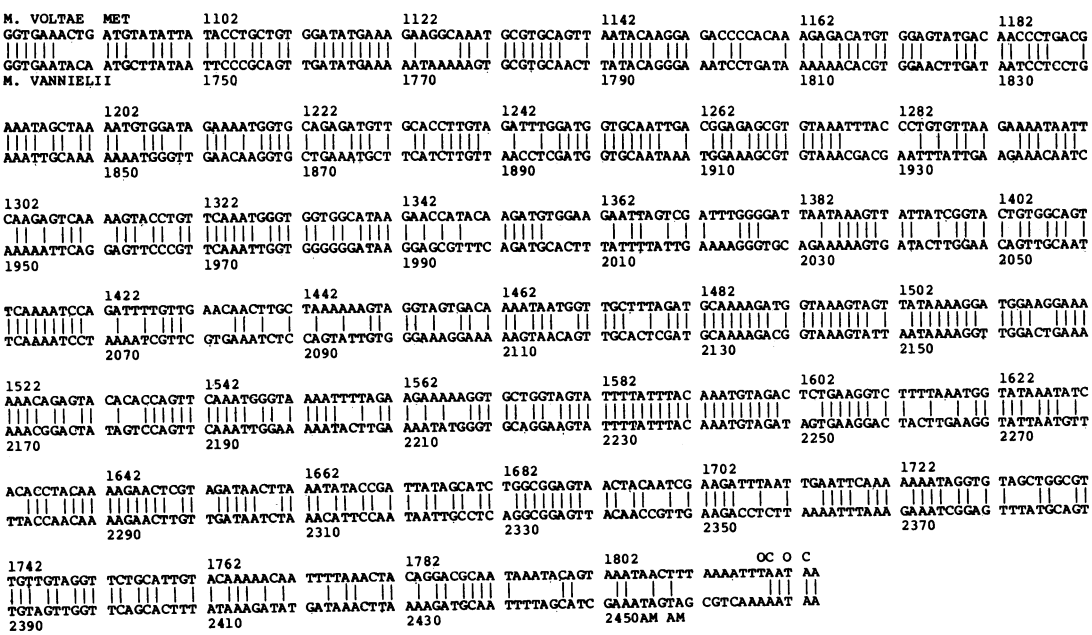

FIG. 2. Comparison of the DNA sequence of the *M. voltae* and *M. vannielii hisA* structural gene. The nucleotides are numbered based on the total DNA sequenced (see Fig. 1). Nucleotides that are identical in the two sequences are connected by a vertical line. The initiation ATG sequences are labeled "MET" and termination codons are labeled "OC" and "AM."

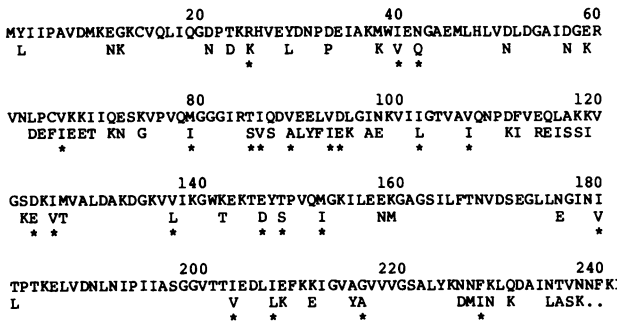

FIG. 3. Amino acid sequence homology of the *M. voltae* (upper row in each pair) and *M. vannielii* (lower row) HisA proteins. The single-letter amino acid code is used. Amino acid assignments not shown for *M. vannielii* indicate identity with the corresponding *M. voltae* residue. Asterisks indicate conservative differences (13).

for an additional 648 nucleotides to the site at which the methanogen DNA was ligated to the pUC8 vector (unpublished results). Thus, at present we are unable to ascertain the total size of ORF1. Sequence information is not available for this region of *M. voltae* DNA.

To achieve maximum homology in the ORF1 sequences, it was necessary to include several gaps in either one or the other sequence. Such gaps may represent additions or deletions in the evolutionary divergence from a common sequence. A more speculative possibility is that they represent introns or vestiges of introns which, if spliced out, would generate more similar mRNA species. For example the *M. voltae* sequence contains an additional 30 bp (671–700; Fig. 4) opposite a gap (position 1306) in the *M. vannielii* sequence. The *M. voltae* sequence A-T-G-G-T-T-A-G-T (positions 644–652) just upstream of the additional 30 bp is reminiscent of the 5′ consensus nuclear splice site sequence $\frac{A}{C}$-A-G-G-T-$\frac{A}{G}$-G-T (13) for nuclear encoded mRNAs. Several sequences (Fig. 4) can be found in the *M. voltae* sequence, downstream from the gap in the *M. vannielii* sequence, that show partial homology to the proposed 3′ consensus splice sequence Y_6-C-A-G-G-$\frac{G}{T}$, in which Y is a pyrimidine (14). Removal of this *M. voltae* "intervening mRNA sequence" would result in greater similarity in the polypeptides encoded by the ORF1s of *M. voltae* and *M. vannielii*.

In contrast to the ORF1 and *hisA* structural gene, the right-most ORF sequences (ORF3) show very little nucleotide or amino acid sequence homology (data not shown).

Codon Usage. As expected from the DNA sequence analysis, there is a strong preference (76% for *M. voltae* and 81% for *M. vannielii*) for codons ending in A or U. This is reflected in a pattern of codon usage (Table 1) that is quite distinct from that found in *E. coli* (15), which has a G+C content of 51% (16). For example, whereas *E. coli* prefers CUG to specify leucine, this codon is used little in either methanococcal ORF. The ACA triplet codon, which is the least-used threonine codon in *E. coli*, is the most frequent one used in the methanogen genes. Other notable differences are observed in the usage of AUA, AGA, CCG, CAG, AAC, and CGC codons. The dinucleotide C-G is rarely used in the methanococcal genes (found in only 17 codons of 1276 analyzed in Table 1).

Transcription Signals and Repeat Sequences. The nucleotide sequences 5′ to both *hisA* genes were compared in the hope of identifying common sequences that might represent promoters or other signals necessary for transcription (Figs. 5 and 6). Although cloning and subcloning experiments indicated that transcription must initiate in this region in *E. coli*, we have no information as to whether transcription is actually initiated in this region in methanococcal cells. Examination of the *M. voltae* DNA did reveal, however, sequences that resemble prokaryotic promoters. The *M. voltae* sequences beginning at nucleotides −84 (T-T-G-T-A-T..N_{17}..T-A-T-A-A-T) and −77 (T-T-G-A-T-A..N_{21}..T-A-T-A-A-T) are similar to the consensus sequence of *E. coli* promoters: T-T-G-A-C-A..N_{15-21}..T-A-T-A-A-T (refs. 17 and 18; N = any nucleotide). Because we found that a 334-bp restriction fragment that includes the entire region between ORF1 and *hisA* is capable of promoting expression of an *E. coli* gene (*galK*) when inserted in the appropriate orientation (−35 consensus, −10 consensus, *galK* structural gene) in the pK01 promoter–detection vector system of Rosenberg *et al.* (19) (unpublished observation), it is possible that these sequences may be responsible for the expression of the methanogen *hisA* gene in *E. coli*. In the case of *M. vannielii*,

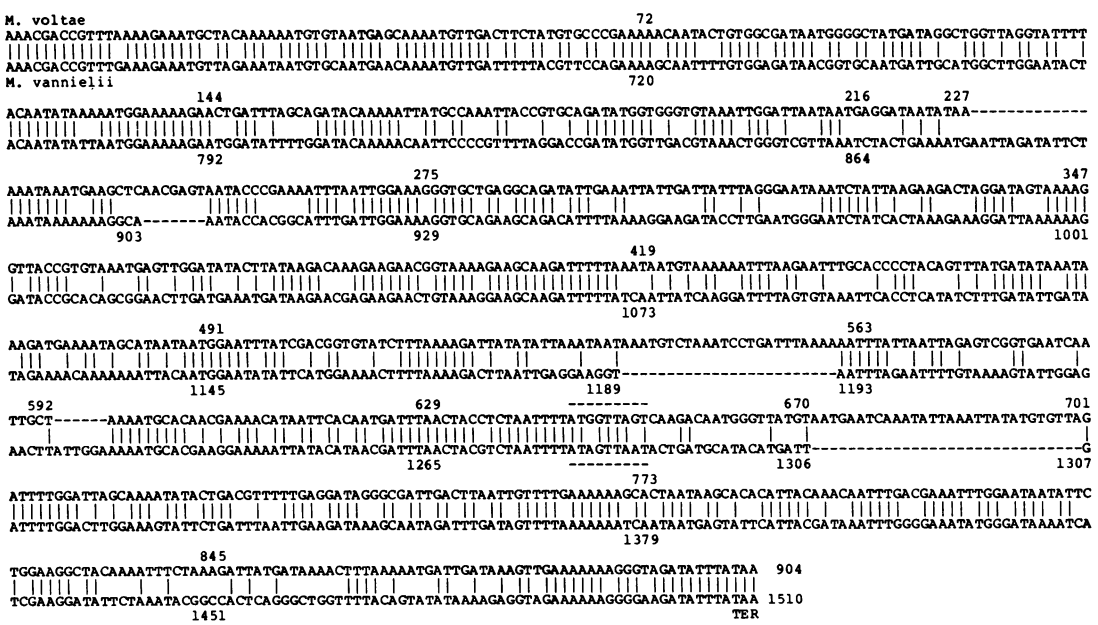

FIG. 4. DNA sequence comparison of the ORF1 regions. The nucleotide sequences are aligned for maximum homology (indicated by vertical lines) by introduction of gaps (horizontal dashed lines). A potential splice site (14) is indicated by the overlining and underlining of the putative nucleotide sequences.

Table 1. Total codon usage in *M. voltae* and *M. vannielii* ORFHisA and ORF1 genes compared to codon usage in *E. coli*

Residue and codon	*E. coli** Total codons	*E. coli** % synonym use	*M. voltae*† Total codons	*M. voltae*† % synonym use	*M. vannielii*† Total codons	*M. vannielii*† % synonym use	Residue and codon	*E. coli** Total codons	*E. coli** % synonym use	*M. voltae*† Total codons	*M. voltae*† % synonym use	*M. vannielli*† Total codons	*M. vannielli*† % synonym use
Phe UUU	104	44	10	83	21	100	Tyr UAU	69	41	11	61	16	67
Phe UUC	135	56	2	17	0	0	Tyr UAC	101	59	7	39	8	33
Leu UUA	36	6	29	64	29	40	Ter UAA	22	88	2	100	1	50
Leu UUG	51	8	6	13	6	8	Ter UAG	1	4	0	0	0	0
Leu CUU	54	9	4	9	27	37	Ter UGA	2	8	0	0	1	50
Leu CUC	41	7	1	2	4	5							
Leu CUA	11	2	4	9	6	8	His CAU	42	39	2	40	7	47
Leu CUG	*432*	*69*	*1*	*2*	*1*	*4*	His CAC	66	61	3	60	8	53
Ile AUU	151	37	29	48	40	53	Gln CAA	75	27	15	88	12	75
Ile AUC	252	62	4	7	10	13	*Gln CAG*	*207*	*73*	*2*	*12*	*4*	*25*
ILE AUA	*2*	*1*	*27*	*45*	*25*	*34*							
							Asn AAU	57	24	40	87	23	66
Met AUG	189	—	16	—	19	—	*Asn AAC*	*179*	*76*	*6*	*13*	*12*	*34*
Val GUU	182	38	16	36	35	64	Lys AAA	296	77	47	87	63	87
Val GUC	62	13	2	4	1	2	Lys AAG	90	23	7	13	9	13
Val GUA	111	23	18	41	14	25							
Val GUG	130	27	8	18	5	9	Asp GAU	175	51	28	68	36	78
							Asp GAC	168	49	13	32	10	22
Ser UCU	86	27	7	46	6	15							
Ser UCC	83	26	0	0	1	2.5	Glu GAA	328	73	26	72	53	96
Ser UCA	27	8	3	20	20	51	Glu GAG	119	27	10	28	2	4
Ser UCG	37	11	0	0	1	2.5							
Ser AGU	21	6	3	20	9	23	Cys UGU	21	42	5	71	4	50
Ser AGC	70	22	2	14	2	5	Cys UGC	29	58	2	29	4	50
Pro CCU	24	9	6	43	7	32	Trp UGG	48	—	5	—	6	—
Pro CCC	16	6	4	28	6	27							
Pro CCA	53	20	3	22	9	41	Arg CGU	201	58	3	19	5	22
Pro CCG	*174*	*65*	*1*	*7*	*0*	*0*	*Arg CGC*	*121*	*35*	*0*	*0*	*1*	*4*
							Arg CGA	8	2	1	6	0	0
Thr AGU	76	24	7	32	11	31	Arg CGG	11	3	0	0	1	4
Thr ACC	162	51	3	14	3	8	*Arg AGA*	*4*	*1*	*10*	*62*	*8*	*35*
Thr ACA	*19*	*6*	*11*	*50*	*15*	*42*	Arg AGG	1	0.25	2	13	8	35
Thr ACG	63	20	1	4	7	19							
Ala GCU	202	28	10	42	5	13	Gly GGU	231	48	21	54	18	29
Ala GCC	136	19	0	0	3	8	Gly GGC	197	41	7	18	4	6
Ala GCA	166	23	13	54	30	77	Gly GGA	22	5	7	18	32	52
Ala GCG	*221*	*30*	*1*	*4*	*1*	*2*	Gly GGG	33	7	4	10	7	13

E. coli codon usages are taken from ref. 15. Notable differences in codon usage from that of *E. coli* are italicized.
†Codon usages are a summation of ORF1 and ORFHisA, which have a very similar usage.

one finds several sequences that could function as promoters in *E. coli* (Fig. 6). However, since *M. vannielii* DNA subcloning starting at the *Kpn* I site at position −108 complements *hisA* mutations in *E. coli*, transcription must be initiated in *E. coli* between the *Kpn* I site and the ATG initiation sequence. The most probable *M. vannielii* sequence used as a promoter in *E. coli* is therefore −81 (T-T-A-A-C-A..N_{15}..T-T-A-A-A-T).

The intergenic regions between ORF1 and ORFHisA contain striking repetitive sequences. The *M. voltae* region contains three perfect repeats and one imperfect repeat T-A-T-A-A-T-A-G-A-A-T starting at position −61 (Fig. 5). This sequence is not found in the *M. vannielii* intergenic region nor in other *M. voltae* sequences determined so far. However, the *M. vannielii* intergenic region preceding *hisA* does contain repetitive sequences (Fig. 6). The sequence T-T-T-T-A-A-A-T-G is tandemly repeated starting at positions −72 and −62; the sequence T-C-C-T-A-C-T-T-T-A-C starting at position −43 is repeated in the opposite direction at position −22. An RNA transcript of the region −43 to −22 could form a perfect 10-bp loop. An eight out of nine match to the inverted repeat sequence is also found starting at position −87, offering an additional opportunity for formation of RNA secondary structure.

We also compared the sequences in the region upstream from the ORF3s described in Fig. 1 (data not shown). In the case of *M. voltae*, a good promoter fit (T-T-G-A-T-A..N_{22}..T-A-T-A-A-T) was found no but ribosome binding sequence. In the case of *M. vannielii*, there is a ribosome binding sequence site (G-G-T-G) 5 bp upstream from the initiating ATG and a T-A-T-A-A-T at position −14 but no sequence reminiscent of a −35 consensus. The regions 5′ to both the *M. voltae* ORFHisA and ORF3 contain the 10-bp sequence A-T-T-T-T-G-A-T-A-T. The region preceding ORF3 also contains the

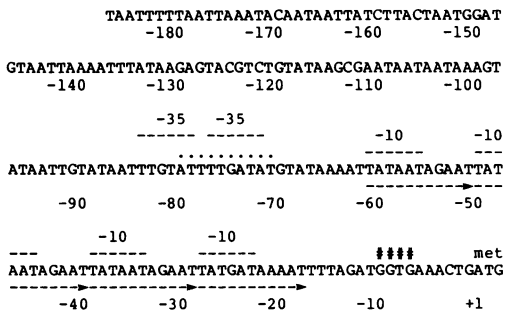

FIG. 5. DNA sequence upstream of the *M. voltae hisA* gene. The initiating methionine codon begins at +1. The ribosome binding sequence is indicated (####). Putative promoter −35 and −10 regions (17, 18) are indicated by overlining. The dots overlay a sequence that is repeated upstream from ORF3 (not shown). The direct repeats are underlined, with orientation indicated by the arrowheads.

related sequence A-T-T-T-T-A-A-T-A-T. Although this sequence does not occur in the *M. vannielii* DNA so far sequenced, it does occur in the intergenic regions between several ORFs, including the *purE* gene, of DNA sequenced from *Methanobrevibacter smithii*, a methanogen having a G+C content of 30.6% (unpublished results).

DISCUSSION

The taxonomy of methanogens and other archaebacteria is, of necessity, still in its infancy and is based for the most part on a binary comparison of 16S RNA oligonucleotide catalogs (1, 20). This study provides the opportunity to estimate the evolutionary distance between two related archaebacterial species by using a comparison of the nucleotide sequence of protein-encoding genes. With the enteric group for comparison, the methanococcal *hisA* (67% homologous) genes would seem to have diverged to a greater extent than genes among this eubacterial group. When comparisons are made of the DNA sequences of all five *trp* genes in the *trp* operons of *E. coli* and *S. typhimurium*, their homologies range from 75% to 84% (21). The *trpA* gene of *Klebsiella aerogenes* is 77% and 75% homologous in nucleotide sequence to *E. coli* and *S. typhimurium*, respectively (22). However, the methanococcal *hisA* genes are more closely related than are the *E. coli* and *B. subtilis trpD* genes (44% homologous). As a further example, the *E. coli ompA* gene nucleotide sequence is highly related to that of *S. typhimurium* (90%), *Enterococcus aerogenes* (85%), *Serratia marcescens* (78%), and *Shigella*

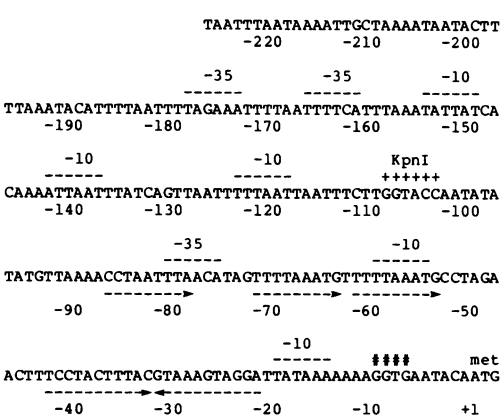

FIG. 6. DNA sequence upstream of the *M. vannielii hisA* gene. The initiating methionine codon is at +1. The putative −35 and −10 regions are indicated by the overlining. The ribosome binding site is indicated (####). Repeats are underlined, with orientation indicated by the arrowheads.

dysenteriae (97%) (23). Although it is possible that divergence of this outer membrane protein is limited by its functional role, its homology throughout the enteric group is striking when compared to the methanogen genes.

We conclude that the enterics are more closely related as a group than are *M. voltae* and *M. vannielii* but that these methanococci are more closely related to each other than are *E. coli* and *B. subtilis*. This agrees with evolutionary distances based on 5S and 16S rRNA sequence comparisons (1, 24). Although these results raise the question as to whether *M. voltae* and *M. vannielii* should be considered as organisms of distinct genera, any reassessment will have to await further comparisons with each other, related methanococci, and other methanogens.

We readily acknowledge the help provided by P. T. Hamilton in obtaining and analyzing the *M. vannielii* sequences and thank him for permission to cite his unpublished *M. smithii* sequences. This research was supported by Department of Energy Grants DE-AC02-81ER10945 (to J.N.R.) and DE-FG02-84ER13241 (to J.K.) by National Science Foundation Grant PCM 83-13901 (to J.K.), and by Contracts 5083-260-0895 and CR810340 from the Gas Research Institute and Environmental Protection Agency, respectively (to J.N.R.). J.N.R. is the recipient of Research Career Development Award IK04AG00108 from the National Institute of Aging.

1. Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R. & Wolfe, R. (1979) *Microbiol. Rev.* **43,** 260–296.
2. Woese, C. R., Magrum, L. J. & Fox, G. E. (1978) *J. Mol. Evol.* **11,** 245–252.
3. Reeve, J. N., Trun, N. J. & Hamilton, P. T. (1982) in *Genetic Engineering of Micro-organisms for Chemicals*, eds. Hollaender, A., DeMoss, R. D., Kaplan, S., Konisky, J., Savage, D. & Wolfe, R. S. (Plenum, New York), pp. 233–244.
4. Wood, A., Redborg, A., Cue, D. & Konisky, J. (1983) *J. Bacteriol.* **156,** 19–29.
5. Morris, C. J. & Reeve, J. N. (1984) in *Microbial Growth on C1 Compounds*, eds. Crawford, R. L. & Hanson, R. S. (Am. Soc. Microbiol., Washington, DC), pp. 205–209.
6. Meile, L. & Reeve, J. N. (1985) *Biotechnology* **3,** 69–72.
7. Magrum, L. J., Luehrsen, K. R. & Woese, C. R. (1978) *J. Mol. Evol.* **11,** 1–8.
8. Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 1342–1346.
9. Zillig, W., Stetter, K. O., Schnabel, R., Madon, J. & Giere, A. (1982) *Zentralbl. Bakteriol. Abt. 1: Orig. Reihe C* **3,** 218–227.
10. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
13. Miyata, T., Miyazawa, S. & Yasunga, T. (1979) *J. Mol. Evol.* **12,** 219–236.
14. Cech, T. R. (1983) *Cell* **34,** 713–716.
15. Konigsberg, W. & Godson, G. N. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 687–691.
16. Johnson, J. L. (1981) in *Manual of Methods for General Microbiology*, eds. Gerhardt, P. & Krieg, N. A. (Am. Soc. Microbiol., Washington, DC), pp. 450–472.
17. Pribnow, D. (1975) *J. Mol. Biol.* **99,** 419–443.
18. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13,** 319–353.
19. Rosenberg, M., Chepelinsky, M. & McKenney, K. (1983) *Science* **222,** 734–739.
20. Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N. & Woese, C. R. (1980) *Science* **209,** 457–463.
21. Horowitz, H., van Arsdell, J. & Platt, T. (1983) *J. Mol. Biol.* **169,** 775–797.
22. Nichols, B. P., Blumenberg, M. & Yanofsky, C. (1981) *Nucleic Acids Res.* **9,** 1743–1755.
23. Braun, G. & Cole, S. T. (1984) *Mol. Gen. Genet.* **195,** 321–328.
24. Hori, H. & Osawa, S. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 381–385.