Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 3992–3996, June 1987
Biochemistry

Structure and expression of the genes, *mcrBDCGA*, which encode the subunits of component C of methyl coenzyme M reductase in *Methanococcus vannielii*

(methyl reductase/archaebacteria/DNA sequence/codon usage/translation regulation)

D. S. CRAM*†, B. A. SHERF*, R. T. LIBBY*‡, R. J. MATTALIANO§, K. L. RAMACHANDRAN§, AND J. N. REEVE*¶

*Department of Microbiology, The Ohio State University, Columbus, OH 43210; and §Biogen Research Corporation, 14 Cambridge Center, Cambridge, MA 02142

Communicated by Leo A. Paquette, February 13, 1987 (received for review January 16, 1987)

ABSTRACT The genes that encode the α, β, and γ subunits of component C of methyl coenzyme M reductase (*mcrA*, *mcrB*, and *mcrG*) in *Methanococcus vannielii* have been cloned and sequenced, and their expression in *Escherichia coli* has been demonstrated. These genes are organized into a five-gene cluster, *mcrBDCGA*, which contains two genes, designated *mcrC* and *mcrD*, with unknown functions. The *mcr* genes are separated by very short intergenic regions that contain multiple translation stop codons and strong ribosome-binding sequences. Although the genome of *M. vannielii* is 69 mol % A+T, there is a very strong preference in the *mcrA*, *mcrB*, and *mcrG* genes for the codon with a C in the wobble position in the codon pairs $\mathrm{AA}^{\mathrm{C}}_{\mathrm{U}}$ (asparagine), $\mathrm{GA}^{\mathrm{C}}_{\mathrm{U}}$ (aspartic acid), $\mathrm{CA}^{\mathrm{C}}_{\mathrm{U}}$ (histidine), $\mathrm{AU}^{\mathrm{C}}_{\mathrm{U}}$ (isoleucine), $\mathrm{UU}^{\mathrm{C}}_{\mathrm{U}}$ (phenylalanine), and $\mathrm{UA}^{\mathrm{C}}_{\mathrm{U}}$ (tyrosine). The *mcrC* and *mcrD* genes do not show this codon preference and frequently have U or A in the wobble position. As the codon pairs listed above are likely to be translated by the same tRNA with a G in the first anticodon position, the presence of C in the wobble position might ensure maximum efficiency of translation of transcripts of these very highly expressed genes.

Methanogenesis, the terminal step in the anaerobic biodegradation of biomass, is a process of major importance in the global cycling of carbon. Although the substrate may be CO_2, formate, methanol, methylamines, or acetate, the final reaction resulting in the release of CH_4 is, in all methanogens, the reduction of the CH_3 moiety of 2-(methylthio)ethanesulfonic acid. This reaction is catalyzed by a complex enzyme system, known as methyl coenzyme M reductase (methyl reductase), which constitutes 5–12% of the cellular protein (1–7). Methyl reductase has been resolved by purification into protein components A and C and a small cofactor, component B (3, 4). Component C is thought to be the actual site of methyl group reduction. Here we report the cloning, DNA sequence determination, and expression in *Escherichia coli* of the genes (*mcr*) that encode the α, β, and γ subunits of component C of *m*ethyl *c*oenzyme M *r*eductase in *Methanococcus vannielii*. Cloning of *mcr* genes from the marine *Methanococcus voltae* has been reported (5, 8, 9) but without detailed DNA sequence data.

MATERIALS AND METHODS

Purification of Methyl Reductase and Production of Antisera. Methyl reductase component C was purified from *M. vannielii* by the procedure of Ellefson and Wolfe (4). Purified holoenzyme was used to raise rabbit anti-methyl reductase antibodies. Antisera preparations were shown by immunoblot analyses to contain antibodies that bound to the α, β, and γ subunits of component C.

Cloning and Subcloning of the *M. vannielii mcr* Genes. A gene library was constructed by ligation of an *Eco*RI partial digest of *M. vannielii* genomic DNA into λgt11 (10). Recombinant λgt11 clones were screened for their ability to direct the synthesis of antigens in infected cells of *E. coli* that bound the rabbit anti-methyl reductase antibodies. Peroxidase-labeled anti-rabbit goat immunoglobin was used to detect the antigen–rabbit antibody complexes (10). DNA was prepared from peroxidase-positive recombinant clones (λgt11MR), and restriction enzyme maps of the *M. vannielii* DNAs cloned in these phages were constructed.

Two different oligonucleotide probes were synthesized for each *mcr* gene designed to hybridize to adjacent but nonoverlapping regions of the α-, β-, and γ-encoding genes (Fig. 1). DNA preparations from the λgt11MR phages were screened by Southern hybridization analysis (11) to identify phages that contained DNA sequences complementary to the oligonucleotide probes. Hybridization of DNA from two such phages, λgt11MR7A (α- and γ-probe positive) and λgt11-MR8A (β-probe positive), to digests of *M. vannielii* genomic DNA indicated that DNA from both phages hybridized to a common 4.3-kilobase-pair (kbp) *Hin*dIII fragment that was subsequently cloned into pUC18 (12) to produce plasmid pET4300 (Fig. 1).

DNA Sequencing. The *M. vannielii* DNA cloned in λgt11MR7A, λgt11MR8A, and pET4300 and subcloned into plasmids, as indicated in Fig. 2, was sequenced by chemical cleavage methods (13). A detailed sequencing strategy is available upon request.

Labeling of Polypeptides Synthesized in *E. coli* Minicells. Polypeptides encoded by plasmids were labeled and identified by their synthesis in minicells incubated in the presence of L-[^{35}S]methionine or a mixture of ^{14}C-labeled amino acids (14). A Zeineh soft laser scanning densitometer was used to quantitate the exposure of the fluorograms.

RESULTS

Purification of Component C of Methyl Reductase and Design of Oligonucleotide Probes. The α-, β-, and γ-subunit polypeptides of purified component C were eluted from gels to provide material for determination of amino-terminal

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.

†Present address: Department of Microbiology, Monash University, Melbourne, Australia 3168.
‡Present address: Department of Genetics, University of Washington, Seattle, WA 98195.
¶To whom reprint requests should be addressed.

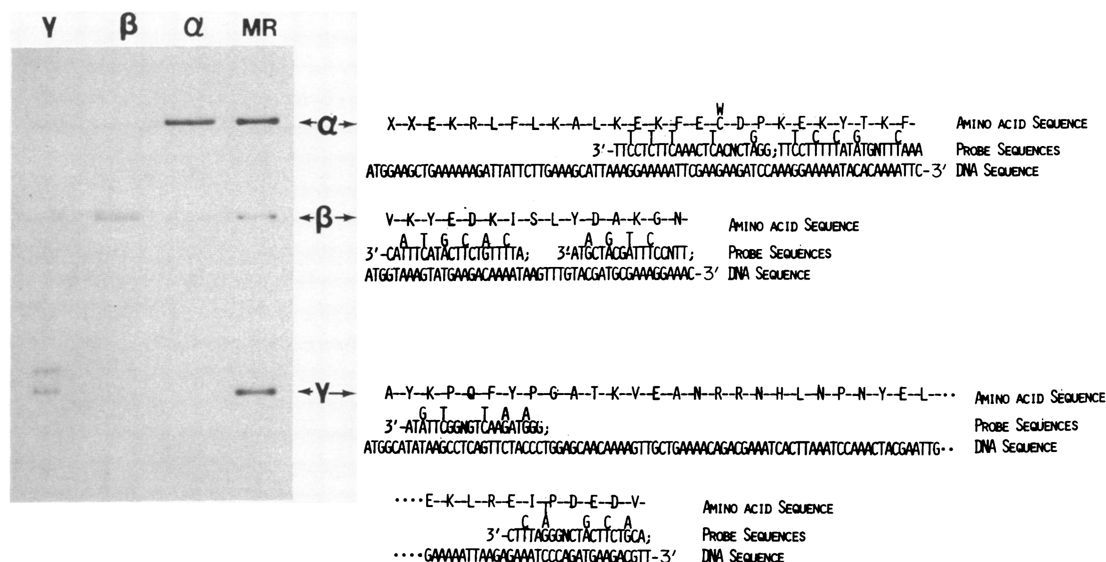

FIG. 1. Preparation of the subunits of component C and design of the oligonucleotide probes. Methyl reductase component C (MR) was purified (4) from *M. vannielii*. The α, β, and γ subunits, separated by $NaDodSO_4$/PAGE, had electrophoretic mobilities indicative of molecular weights of 63,000, 48,000, and 29,500, respectively. Nonoverlapping mixed oligonucleotide probes, with the sequences shown, were designed as indicated from two regions of each amino acid sequence. The DNA sequences obtained for the corresponding regions of *mcrA* (α-encoding), *mcrB* (β-encoding), and *mcrG* (γ-encoding) are shown as the noncoding strand aligned below the encoded amino acids (identified by the single-letter code).

amino acid sequences. Eluted γ subunits always migrated in subsequent gels to form two bands; however, the polypeptides in these two bands were found to have identical amino-terminal amino acid squences (Fig. 1). The amino acid sequences obtained are aligned in Fig. 1 with the DNA sequences synthesized as oligonucleotide probes and with the DNA sequences eventually determined for the corresponding regions of the *mcr* genes. The β and γ subunits did not have amino-terminal methionine residues; a definitive identification of the first two amino-terminal residues of the α subunit was not possible. A mistake in the design of the oligonucleotides was the use of only UUU at positions designating incorporation of phenylalanine residues. Occurrence of the UUC codon had been found to be extremely rare in *M. vannielii* genes (15–17); however, the preferred phenylalanine-encoding codon in the α-, β-, and γ-encoding genes is, in fact, UUC (see Table 2).

Subcloning, Physical Organization, and DNA Sequences of the *mcr* Genes. The *M. vannielii* DNA, cloned in λgt11MR7A, λgt11MR8A, and pET4300, was shown by restriction enzyme analyses and Southern hybridizations to be organized as shown in Fig. 2. The DNA sequence is presented in Fig. 3. The genes encoding the α, β, and γ subunits of component C have been designated *mcrA*, *mcrB*, and *mcrG*, respectively, and encode polypeptides with the properties listed in Table 1. A most unexpected finding was that the *mcrB* and *mcrG* genes are separated by two additional genes, designated *mcrD* and *mcrC*, which if expressed would direct the synthesis of polypeptides with the properties listed in Table 1. It appears likely that *mcrBDCGA* is a single transcriptional unit, since the intergenic regions are very short containing both multiple translation termination codons and strong ribosome-binding sequences (17). The bases that surround the ATG initiation codons are conserved, giving a consensus translation initiation sequence of 5′ C$^{C}_{T}$TATGG for *mcrA*, *mcrB*, and *mcrG*.

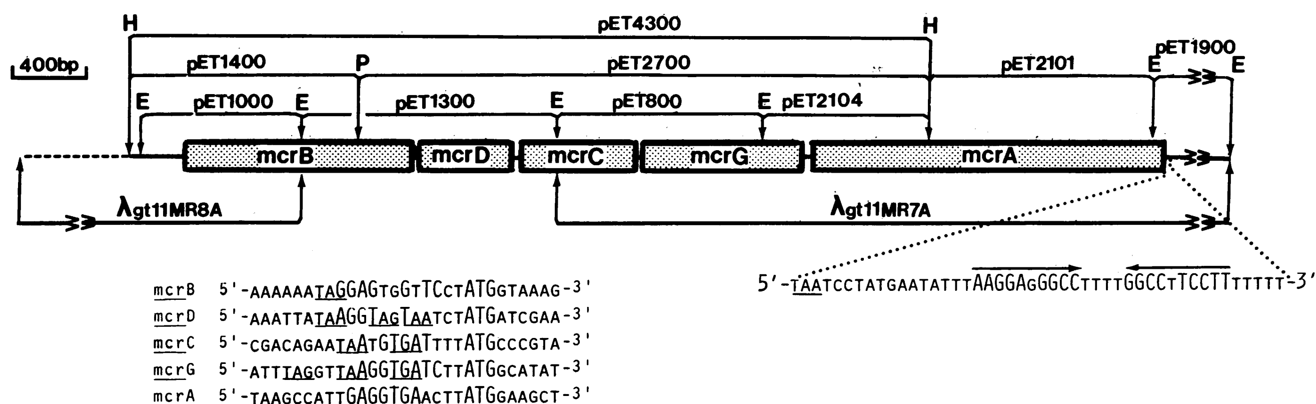

FIG. 2. Organization and subcloning of the *mcrBDCGA* genes. The *M. vannielii* genomic DNA cloned in λgt11MR8A and λgt11MR7A extends 3.5 kbp and 1.7 kbp to the left and right, respectively, of the *mcrBDCGA* gene cluster. The DNAs subcloned into pUC18 (12) to produce the pET plasmids shown are bracketed by arrows, which indicate the sites and the restriction enzymes used in subcloning (H, *Hin*dIII; P, *Pst* I; E, *Eco*RI). There is a second very small *Eco*RI fragment (39 bp; see Fig. 3) in *mcrC*, which is not indicated in the figure but is cloned in pET800. The DNA sequences immediately upstream and surrounding the initiation ATG codon of each *mcr* gene are listed below the diagram. The bases in larger print, in addition to the ATG codons, indicate the proposed ribosome-binding sites and translation termination codons are underlined. The sequence extending 3′ from the termination codon of *mcrA* is also shown.

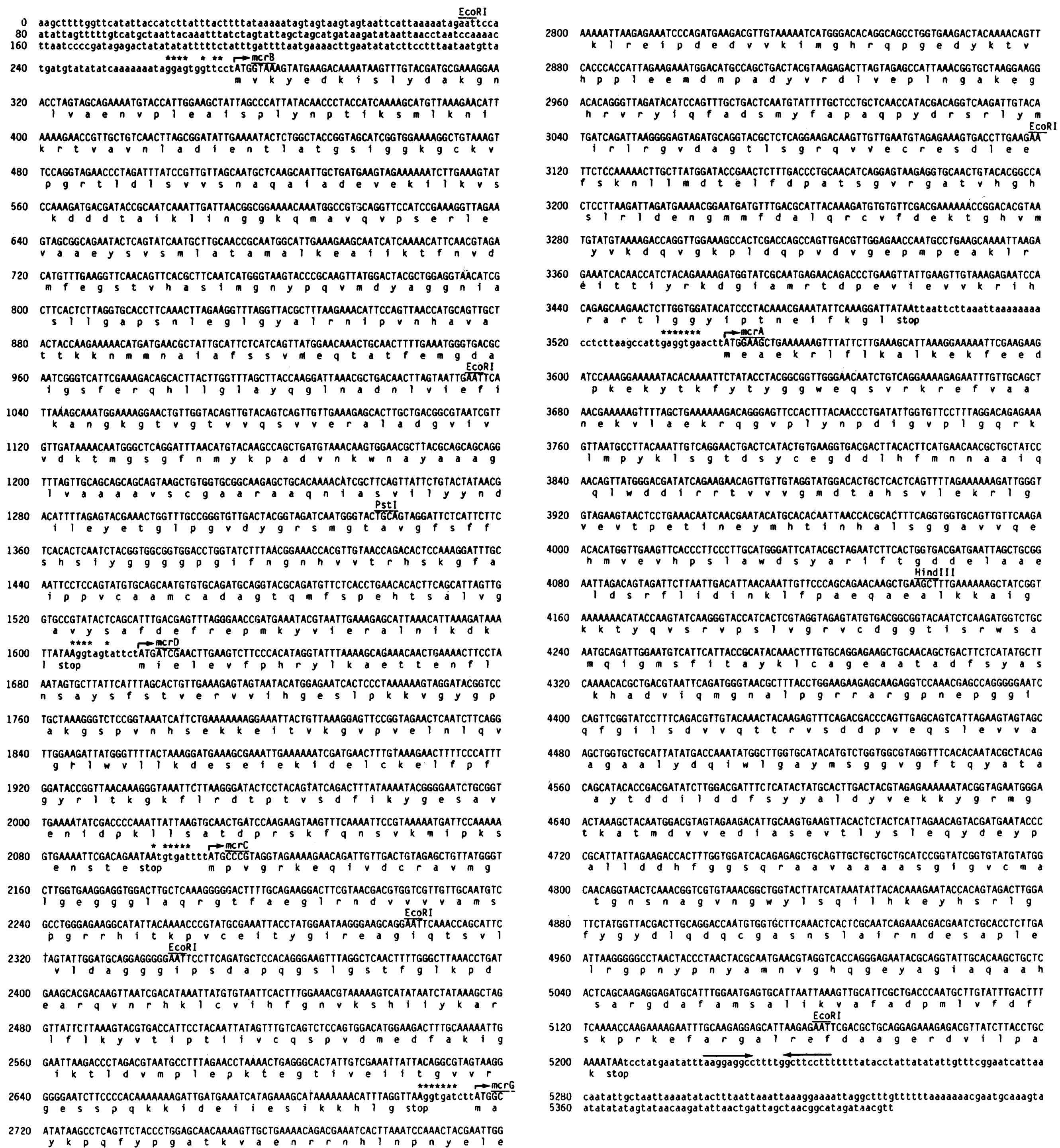

FIG. 3. The DNA and amino acid sequences (identified by the single-letter code) of the *mcrBDCGA* genes. The DNA sequence of the noncoding strand is shown with the first base of each codon directly above the encoded amino acid. Ribosome-binding sites are indicated by asterisks. Bases in intergenic regions are indicated by lowercase letters. Restriction sites used in subcloning (see Fig. 2) are indicated.

We have also cloned and sequenced the DNA regions that flank the *mcrBDCGA* gene cluster (Figs. 2 and 3), but there is no indication of additional closely adjacent open-reading frames. Consistent with previous observations (15–17), these flanking intergenic regions have very high A+T contents; however, there is an unusually G+C-rich sequence immediately 3′ to the *mcrA* gene. As RNA, this region could form a hairpin loop followed by a sequence of five U residues (Fig. 2), a sequence that should function as a transcription terminator in eubacteria (18).

Codon Usage. Codon usage in the *mcrA*, *mcrB*, and *mcrG* genes shows unanticipated preferences. To accommodate the very high 69 mol % A+T in the genome of *M. vannielii*, it was anticipated that whenever possible the third position in a codon would be occupied by an A or a U residue (17). Surprisingly, however, in the *mcrA*, *mcrB*, and *mcrG* genes it is the codon with C in the wobble position in the codon pairs AA^{C}_{U} (asparagine); GA^{C}_{U} (aspartic acid); CA^{C}_{U} (histidine); AU^{C}_{U} (isoleucine); UU^{C}_{U} (phenylalanine), and UA^{C}_{U} (tyro-

Table 1. Properties of the polypeptides encoded by the *mcr* genes

Gene	A+T,* mol %	Molecular weight†	Amino acids†‡					
			Basic	Acidic	Hydro-phobic	Polar	Methio-nines	Total
mcrA	58.1	60,697	55	75	199	128	15	552
mcrB	58.6	46,743	36	43	174	109	17	442
mcrC	59.6	21,222	25	22	69	38	4	197
mcrD	65.2	17,973	23	24	51	38	1	159
mcrG	58.5	29,637	36	43	75	49	11	259

*The mol % A+T of the polypeptide-encoding DNA sequence.
†Amino-terminal methionine residues are assumed not to be present in the mature polypeptides (Fig. 1) (2).
‡Amino acids (single-letter code) in each gene product are tabulated as K, R (basic); D, E (acidic); A, I, L, F, W, V (hydrophobic); and N, C, W, S, T, Y (polar).

sine), which is used most frequently. This increased use of C results in a significant increase in the overall G+C content of the *mcrA*, *mcrB*, and *mcrG* genes (Table 1). In contrast to the *mcrA*, *mcrB*, and *mcrG* genes, codon usage in the *mcrC* and *mcrD* genes follows the pattern previously observed (15–17) of consistent use of A and U in the wobble position (Table 2).

Expression of *mcr* Genes in *E. coli*. Expression of the cloned *mcr* genes was demonstrated in minicells of *E. coli* DS410 (14). The products of the *mcrB* (gp*mcrB*—i.e., the β subunit), *mcrD* (gp*mcrD*), and *mcrG* (gp*mcrG*—i.e., the γ subunit) were identified on the basis of their apparent molecular weights (Fig. 4) and on their synthesis or nonsynthesis in minicells containing pET4300, pET1400, pET2700, and pET1300 as predicted by the organization of *mcr* genes shown in Fig. 2. The products of the *mcrB* and *mcrG* genes were also shown to comigrate with the β- and γ-subunit polypeptides, respectively, of component C purified directly from *M. vannielii*. Expression of *mcrC* in minicells was less certain. Minicells containing pET4300 synthesized a polypeptide, indicated (?) in Fig. 4, which has an electrophoretic mobility indicative of a molecular weight of ≈25,000. The molecular weight of gp*mcrC* calculated from the DNA sequence would be 21,222. The hesitancy in making the gene product identification, in addition to the discrepancy between the observed and calculated molecular weights, is that (?) is not synthesized in detectable amounts in minicells containing pET2700 (Fig. 4). An alternative is that (?) is a truncated gp*mcrA*. Plasmid pET4300 encodes the amino-terminal 199 amino acids of gp*mcrA*, which, if expressed, would result in the synthesis of a polypeptide with a molecular weight of 22,592; however, as with gp*mcrC*, a truncated gp*mcrA* should also be synthesized in minicells containing pET2700. Synthesis of gp*mcrA* (and gp*mcrG*) was observed in minicells infected with λgt11MR7A (data not shown).

DISCUSSION

The sequence of the cloned *mcr* genes has not only added support to the conclusion that the frequency of translation initiation in methanogens is regulated by ribosome-binding sites (17) but has also suggested an additional mechanism to regulate the amounts of polypeptide synthesis. The *mcrA*, *mcrB*, and *mcrG* gene products, the α, β, and γ subunits of component C, must be synthesized in very large amounts, and we have shown that in these genes there are many examples of codons with C in the wobble position being used much more frequently than the synonymous codon with a U in this position (Table 2). These pairs of synonymous codons are translated in *E. coli* (20) and *Saccharomyces cerevisiae* (21) by a single tRNA with a G residue in the first anticodon position and, from the 19 *M. vannielii* tRNA sequences so far available (22–24), this would appear also to be the case in *M. vannielii*. Use of codons with a C in the wobble position maximizes the number of G·C pairs at the sites of codon–anticodon interactions. This could therefore be a mechanism evolved to maximize the rate and/or efficiency of translation (25). This argument was presented previously to explain the preferred use of C-containing codons in highly expressed genes in *S. cerevisiae*, an organism that also has a genome with a very high A+T content (21). It was also pointed out that, although the cysteine-encoding codons UGC and UGU are translated by a single tRNA with G in the first anticodon position, the preferred cysteine codon in *S. cerevisiae* is UGU. This ensures conformity with the overall pattern in *S.*

Table 2. Frequency of occurrence of codons in *M. vannielii* genes

Gene	Codons used									
	AAC/ AAU	GAC/ GAU	CAC/ CAU	AUC/ AUU	UUC/ UUU	UAC/ UAU	UGC/ UGU	AAA/ AAG	GGU + GGA/ GGC + GGG	GUU + GUA/ GUC + GUG
mcrA	17:0	29:9	12:0	12:10	15:6	24:5	0:5	25:3	42:6	37:1
mcrB	23:4	10:10	6:1	13:14	8:7	15:2	0:4	21:5	35:6	39:2
mcrG	6:2	15:5	8:0	9:1	4:4	9:3	0:2	10:4	18:1	20:1
Total	46:6	54:24	26:1	34:25	27:17	48:10	0:9	56:12	95:13	96:4
mcrC	2:1	5:4	1:4	2:14	1:4	1:2	1:3	11:3	15:8	15:5
mcrD	1:6	2:4	0:3	3:3	4:4	3:2	0:1	13:4	7:2	13:1
Total	3:7	7:8	1:7	5:17	5:8	4:4	1:4	24:7	22:10	28:6
hisA	4:10	4:10	1:1	4:13	0:4	0:4	1:0	21:4	19:3	22:4
hisI	3:5	3:10	0:3	3:6	1:6	0:3	0:3	15:3	7:2	7:1
ORF502	8:13	6:26	7:6	6:27	0:17	8:12	3:4	42:5	31:8	27:2
ORF634	7:16	9:24	1:4	8:41	2:19	7:15	3:3	54:9	36:10	54:5
Total	22:44	22:70	9:14	21:87	3:46	15:34	7:10	132:21	93:23	110:12

The DNA sequences are provided in Fig. 3 and refs. 15 and 16. The figures in the two far right columns are the frequency of occurrence of GGU + GGA compared with GGC + GGG and GUU + GUA compared with GUC + GUG. In such groups of four synonymous codons—i.e., GGN, GUN, etc.—there is always very pronounced preferential use of the codons with A or U in the wobble position in all *M. vannielii* genes.

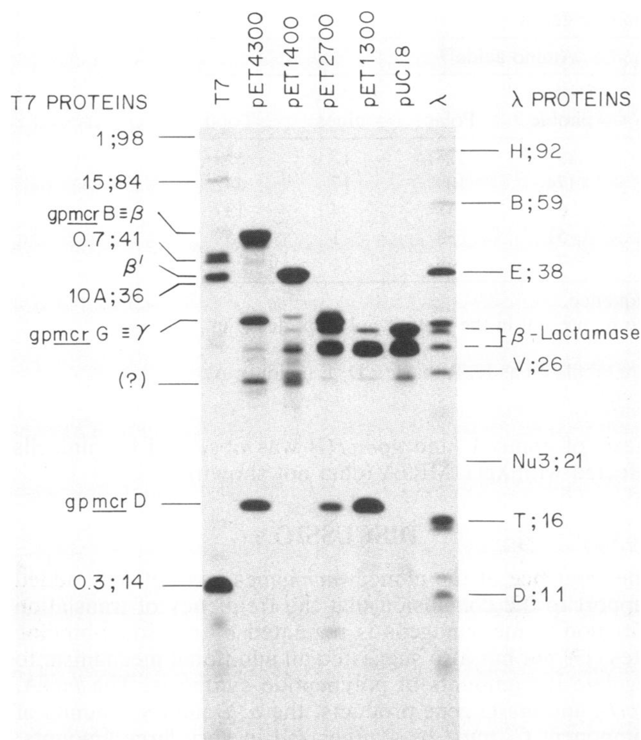

FIG. 4. Synthesis of *mcr* gene products (gp*mcr*) in minicells of *E. coli*. The *mcr* gene products synthesized in minicells containing the plasmid listed above each lane are indicated to the left of the figure. Plasmid pET1400 (see Fig. 2) contains the amino-terminal 1060 bp of *mcrB* and directs the synthesis of a truncated gp*mcrB* (calculated molecular weight, 37,474), indicated as β'. The identity of the polypeptide indicated by (?) remains uncertain (see text). Minicells infected with coliphages T7 or λ and incubated in the presence of [^{35}S]methionine synthesized the phage-encoded gene products used as molecular weight standards (shown here $\times 10^{-3}$). The same patterns of radioactively labeled polypeptides were detected using either a mixture of ^{14}C-labeled amino acids or [^{35}S]methionine as the radioactive precursor. The precursor and mature forms of β-lactamase (19), encoded by the pUC18 vector DNA (12), are indicated to the right of the figure. Based on methionine content and scanning densitometry, minicells containing pET4300 synthesized molecules of gp*mcrB*, gp*mcrG*, gp*mcrD*, and β-lactamase (19) in the ratio 6.5:3:45:1.

cerevisiae DNA of avoidance of the dinucleotide GC (21). Consistent with the *S. cerevisiae* observations, the preferred cysteine codon in almost all *M. vannielii* genes, including *mcrA*, *mcrB*, and *mcrG*, is also UGU (Table 2). One notable difference in preferred codon usage in genes of *S. cerevisiae* as compared with genes of *M. vannielii* is that single-base codons such as AAA are avoided in yeast genes (21), whereas this is the preferred lysine codon in *M. vannielii* genes (Table 2).

If the above arguments for the use of C-containing codons to facilitate high levels of gene expression in *M. vannielii* are valid, then we must predict that the products of the *mcrC* and *mcrD* genes are not synthesized in such large amounts in *M. vannielii*. The observed synthesis of relatively large amounts of gp*mcrD* in *E. coli* minicells may, in this regard, be misleading (Fig. 4). Preferred codon usage in *E. coli* genes (20) is very different from the preferred codon usage in *M. vannielii* genes. Although the amounts of gp*mcrC* and gp*mcrD* may not be very large in *M. vannielii*, it is clearly important to determine their roles, if any, in methanogenesis. The rates of methanogenesis obtained with *in vitro* preparations of methyl reductase are usually very low when compared with *in vivo* rates (1, 2, 6, 7). The possibility must therefore now exist that preparations of methyl reductase, obtained by the standard procedures (2–7), may have low activities because the functions of gp*mcrD* and gp*mcrC* have been lost during purification.

Determination of the amino acid sequences and synthesis of the oligonucleotide probes were performed in the laboratories of Biogen Research Corporation. This work was supported in part by Department of Energy Grant DE-AC02-81ER10945 and by Environmental Protection Agency Contract CR810340 awarded to J.N.R.

1. Daniels, L., Sparling, R. & Sprott, E. D. (1984) *Biochim. Biophys. Acta* **768**, 113–163.
2. Hartzell, P. L. & Wolfe, R. S. (1986) *Syst. Appl. Microbiol.* **7**, 376–382.
3. Gunsalus, R. P. & Wolfe, R. S. (1980) *J. Biol. Chem.* **255**, 1891–1895.
4. Ellefson, W. L. & Wolfe, R. S. (1981) *J. Biol. Chem.* **256**, 4259–4262.
5. Konheiser, U., Pasti, G., Bollschweiler, C. & Klein, A. (1984) *Mol. Gen. Genet.* **198**, 146–152.
6. Ankel-Fuchs, D., Huster, R., Morschel, E., Albracht, S. P. J. & Thauer, R. K. (1986) *Syst. Appl. Microbiol.* **7**, 383–387.
7. Wackett, L. P., Hartwieg, E. A., King, J. A., Orme-Johnson, W. H. & Walsh, C. T. (1987) *J. Bacteriol.* **169**, 718–727.
8. Allmansberger, R., Bollschweiler, C., Konheiser, U., Muller, B., Muth, E., Pasti, G. & Klein, A. (1986) *Syst. Appl. Microbiol.* **7**, 13–17.
9. Schallenberg, J., Thomm, M., Stetter, K. O., Moes, M., Truss, M., Allmansberger, R., Bokranz, M., Muth, E. & Klein, A. (1986) in *Biology of Anaerobic Bacteria*, ed. Dubourguier, H. C. (Elsevier, Amsterdam), pp. 182–192.
10. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
11. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
12. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–109.
13. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–580.
14. Reeve, J. N. (1979) *Methods Enzymol.* **68**, 493–511.
15. Cue, D., Beckler, G. S., Reeve, J. N. & Konisky, J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4207–4211.
16. Beckler, G. S. & Reeve, J. N. (1986) *Mol. Gen. Genet.* **204**, 133–140.
17. Reeve, J. N., Hamilton, P. T., Beckler, G. S., Morris, C. J. & Clarke, C. H. (1986) *Syst. Appl. Microbiol.* **7**, 5–12.
18. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
19. Sutcliffe, J. G. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
20. Ikemura, T. (1981) *J. Mol. Biol.* **146**, 1–21.
21. Bennetzen, J. L. & Hall, B. D. (1982) *J. Biol. Chem.* **257**, 3026–3031.
22. Wich, G., Jarsh, M. & Böck, A. (1984) *Mol. Gen. Genet.* **196**, 146–151.
23. Wich, G., Sibold, L. & Böck, A. (1986) *Syst. Appl. Microbiol.* **7**, 18–25.
24. Wich, G., Hummel, H., Jarsch, M., Bär, U. & Böck, A. (1986) *Nucleic Acids Res.* **14**, 2459–2479.
25. Grosjean, H. & Fiers, W. (1982) *Gene* **18**, 199–209.