Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3031–3035, May 1989
Biochemistry

A hydrogenase-linked gene in *Methanobacterium thermoautotrophicum* strain ΔH encodes a polyferredoxin

(methyl viologen hydrogenase/*mvhDGAB*/archaebacteria)

JOHN N. REEVE[†‡], GREGORY S. BECKLER[†], DAVID S. CRAM[†§], PAUL T. HAMILTON[†¶], JAMES W. BROWN[‡‖], JOSEPH A. KRZYCKI[†], ANDREW F. KOLODZIEJ[††], LISA ALEX[‡‡], WILLIAM H. ORME-JOHNSON[††], AND CHRISTOPHER T. WALSH[‡‡]

[†]Department of Microbiology and [‡]Molecular, Cellular Developmental Biology Program, The Ohio State University, Columbus, OH 43210; [††]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; and [‡‡]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Cambridge, MA 02115

Communicated by Ralph S. Wolfe, December 30, 1988 (received for review November 5, 1988)

ABSTRACT **The genes *mvhDGA*, which encode the subunit polypeptides of the methyl viologen-reducing hydrogenase in *Methanobacterium thermoautotrophicum* strain ΔH, have been cloned and sequenced. These genes, together with a fourth open reading frame designated *mvhB*, are tightly linked and appear to form an operon that is transcribed starting 42 base pairs upstream of *mvhD*. The organization and sequences of the *mvhG* and *mvhA* genes indicate a common evolutionary ancestry with genes encoding the small and large subunits of hydrogenases in eubacterial species. The product of the *mvhB* gene is predicted to contain six tandomly repeated bacterial-ferredoxin-like domains and, therefore, is predicted to be a polyferredoxin that could contain as many as 48 iron atoms in 12 Fe_4S_4 clusters.**

Methanobacterium thermoautotrophicum reduces CO_2 to CH_4 using H_2 as the reductant. Therefore, hydrogenase activity is essential for methanogenesis in this species, and two hydrogenases have been purified and characterized from extracts of *M. thermoautotrophicum* (1–3). In this report we describe the organization and structure of the clustered genes (*mvhDGA*) that encode subunits of the hydrogenase that does not reduce cofactor F_{420}, the enzyme conventionally designated as the methyl viologen-reducing hydrogenase (MV hydrogenase). The results obtained indicate that this archaebacterial hydrogenase and several eubacterial hydrogenases (4–8) have evolved from a common ancestor and that a tightly-linked gene, *mvhB*, encodes a polyferredoxin.

MATERIALS AND METHODS

Cloning, Subcloning, and Sequencing of the Cloned *mvh* Genes. Gene libraries were constructed by ligation of *Sau*3A partial digests of *M. thermoautotrophicum* strains ΔH and Winter genomic DNAs into *Bam*HI-digested phage λ Charon35. Using binding of ^{125}I-labeled protein A, we identified desired recombinant clones by their ability to direct the synthesis of antigens in *Escherichia coli* that bound rabbit antibodies raised against the α subunit of the F_{420}-reducing hydrogenase, purified as previously described from *M. thermoautotrophicum* ΔH (2). DNA prepared from positive clones was subcloned into pUC8 (11) and sequenced (Fig. 1).[§§]

Determination of Amino Acid Sequences. The amino-terminal sequences of the subunit polypeptides of MV-hydrogenase purified from *M. thermoautotrophicum* ΔH and separated by sodium dodecyl sulfate/polyacrylamide gel electrophoresis were determined by using an Applied Biosystems 470A gas-phase protein sequencer. The subunits did not have N-terminal methionyl residues. The N-terminal amino acid sequence of the α subunit was found to be ≈40% identical to the N-terminal sequence of the α subunit of the F_{420}-reducing hydrogenase (2). This conservation of amino acid residues presumably accounts for the immunological cross-reactivity of the two polypeptides.

RNA Preparation, Primer Extension, and RNA Sequencing. Total cellular RNA was prepared from lysates of *M. thermoautotrophicum* strain ΔH. ^{32}P-labeled oligonucleotide primers were synthesized and hybridized to the *M. thermoautotrophicum* ΔH RNA, and the hybrid molecules were used in primer extension procedures to determine the 5′ end of the mvh transcript (9) and to sequence the transcript of the *mvhG* gene (12) in the region of the cloned TGA codon (see *Results*).

Primer-Directed Amplification and Sequencing of *M. thermoautotrophicum* ΔH Genomic DNA. Oligonucleotide primers (24 mers) were synthesized complementary to the sequences located 42 bp 5′ and 42 bp 3′ from the position at which the TGA codon had been detected in the cloned *mvhG* gene from *M. thermoautotrophicum* ΔH (see *Results*) and used in polymerase chain reactions (13) to amplify the region of the *M. thermoautotrophicum* ΔH genome between the primers. The amplified DNA was sequenced.

RESULTS

Physical Organization and DNA Sequences of the *mvh* Genes. The organization, sequences, and putative regulatory signals of the *mvhDGAB* genes cloned from *M. thermoautotrophicum* ΔH and *M. thermoautotrophicum* Winter are shown in Fig. 1. The genes identified as encoding the α (*mvhA*), δ (*mvhD*), and γ (*mvhG*) subunits, by correlation with N-terminal analyses of the subunit polypeptides from the purified enzyme, encode polypeptides with calculated molecular masses of 53, 15.8, and 33 kDa, respectively. The product of the *mvhB* gene is calculated to have a molecular mass of 44 kDa; however, all polypeptides of this apparent size dissociate from the MV-hydrogenase activity during the final steps of enzyme purification. Therefore this polypeptide does not appear to be essential for enzyme activity *in vitro*.

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.

Abbreviation: MV hydrogenase, methyl viologen-reducing hydrogenase.

[§]Present address: Burnet Clinical Research Unit, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia.
[¶]Present address: Becton Dickinson Research Center, P.O. Box 12016, Research Triangle Park, NC 27709.
[‖]Present address: Department of Biology, Indiana University, Bloomington, IN 47405.

[§§]The sequence reported in this paper is being deposited in the EMBL/GenBank data base (accession no. J04540).

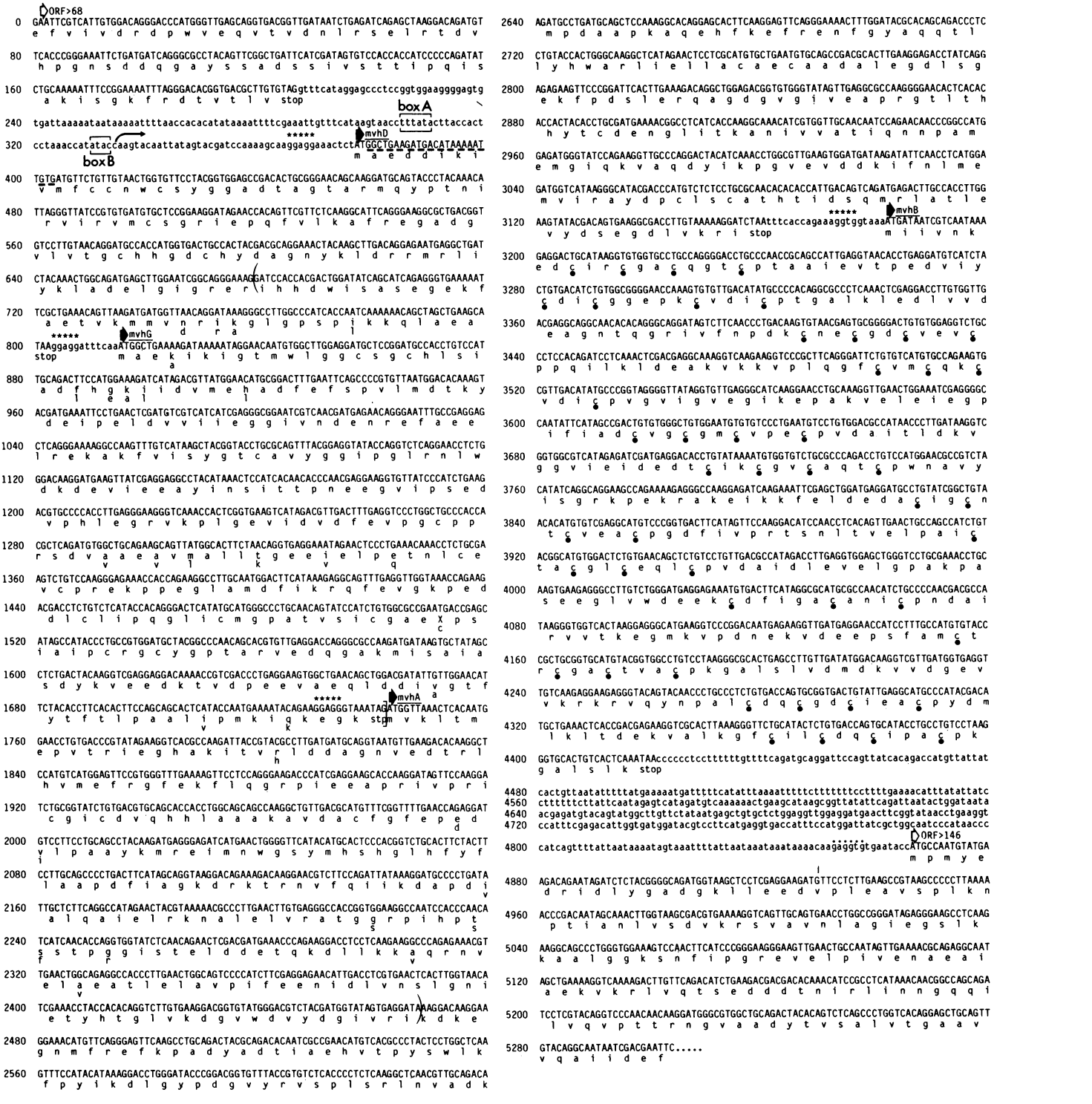

FIG. 1. Organization and sequences of the cloned *mvhDGAB* genes and encoded polypeptides in *M. thermoautotrophicum* strains ΔH and Winter. The DNA sequence of the noncoding strand is shown with the first base of each codon directly above the encoded amino acid. Intergenic regions are shown in lowercase letters. Differences in the amino acid sequence in strain Winter are indicated by listing the amino acid found in strain Winter below the amino acid it replaces in strain ΔH. The limits of the DNA sequenced from strain Winter (nucleotide positions 680–2467) are indicated by the curved brackets. The TGA codon is included in the ΔH sequence, and X is used to identify the "encoded" amino acid residue. Sequences that conform to the consensus for archaebacterial promoters are designated "boxA" and "boxB" (9, 10); the site of transcription initiation, by a curved arrow; potential ribosome-binding sequences, by asterisks; and the ATG translation initiation codons for the *mvh* genes, by short heavy arrows. The 47 cysteinyl residues in *mvhB* are highlighted by dots. The 25-mer used in primer extension experiments (9) to determine the site of transcription initiation was complementary to a transcript of the bases in *mvhG* indicated by the broken underlining (positions 379–404). Truncated flanking ORFs capable of encoding >68 and >146 amino acids are indicated by open arrows.

Comparison of Hydrogenase Sequences and the TGA Codon. The sequence obtained for the *mvhG* gene cloned from *M. thermoautotrophicum* ΔH contains an in-frame TGA codon at position 233, which, it seemed possible, might direct the incorporation of a selenocysteinyl residue (5, 14, 15), but this does not appear to be the case. Sequencing of mRNA isolated

directly from *M. thermoautotrophicum* ΔH (12) and sequencing of *M. thermoautotrophicum* ΔH genomic DNA without cloning after its amplification *in vitro* (13) showed TGC (cysteine) codons in the transcript and in the genomic DNA at the location at which the TGA codon occurs in the cloned DNA. The simplest interpretation of these results is that a TGC → TGA mutation occurred during cloning of the *mvhG* gene from *M. thermoautotrophicum* ΔH.

The genes encoding the small and large subunits of hydrogenases in the eubacterial species *Desulfovibrio vulgaris, Desulfovibrio gigas, Desulfovibrio baculatus, Bradyrhizobium japonicum, Rhodobacter capsulatus* (4–8), *Alcaligenes eutrophus* (B. Friedrich, personal communication), and *E. coli* (A. Przybyla, N. K. Menon, E. S. Choi, C. Chatelus, A. L. Menon, R. Robson, J. LeGall, and H. D. Peck, Jr., personal communication) appear to have a common ancestry with *mvhG* and *mvhA*, respectively. The large subunits of the eubacterial enzymes and the α subunit of the MV-hydrogenase have several regions with well-conserved primary amino acid sequences and also two pairs of conserved cysteinyl residues. The small subunits of the eubacterial enzymes and the γ polypeptide of the MV-hydrogenase have, in contrast, only limited conservation in their primary amino acid sequences, but all do contain 10 cysteinyl residues at similar locations and with conserved spacings (Fig. 2). Unlike the eubacterial small subunits, the γ polypeptide of the MV-hydrogenase does not appear to be synthesized with an N-terminal signal sequence.

***mvhB* Encodes a Polyferredoxin.** The DNA sequence of *mvhB* predicts an encoded polypeptide with a molecular mass of 44 kDa that contains 47 cysteinyl residues organized within six tandomly repeated bacterial-ferredoxin-like domains (Figs. 1 and 3). Five groups of bacterial ferredoxins, including an archaebacterial grouping, have been defined (17), and in Table 1 the six ferredoxin-like domains encoded by *mvhB* are aligned with representative sequences of the different ferredoxin groups. Domain 1 also contains the amino acid sequence Pro-Thr-Ala-Ala-Ile immediately following the Cys-IV* residue (Table 1), which is found at the same location within the ferredoxin-like domain of the large subunit of the Fe-containing hydrogenase of *D. vulgaris* (Table 1; ref. 19). The amino acid sequences in domains 2, 3, 4, and 5 of *mvhB* conform well to the pattern established for group 4 (archaebacterial) ferredoxins. There are 18 amino acid residues within a 30-amino acid sequence in domain 3 that are identical to the sequence of amino acids in a ferredoxin purified from *M. barkeri* DSM800 (20). The isoleucyl residue at the position designated as Cys-II in domain 5 is unusual but is still within the variability of bacterial ferredoxins (17). The absolute conservation of the heptapeptide Pro-Lys-Gly-Ala-Leu-Ser-Leu immediately following the Cys-IV residue in domains 5 and 6 provides strong circumstantial evidence for a gene duplication event. The amino acid sequences designated "Spacer" in Table 1 are predicted by the DNASTAR and SURFACEPLOT 1.2 programs (16) to be charged and hydrophilic and to form surface-located α-helical structures. A

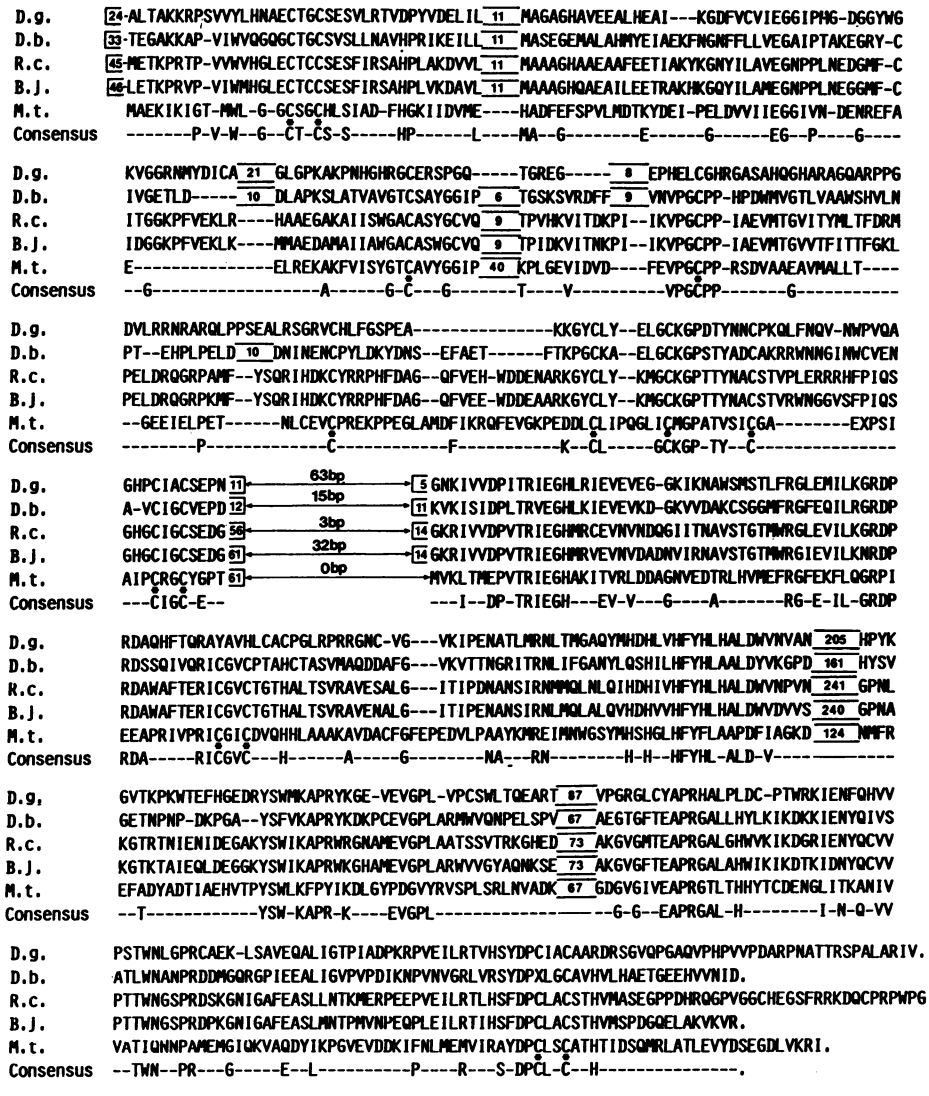

FIG. 2. Comparison of hydrogenases. The amino acid sequences of the small and large subunits of the hydrogenases from *D. gigas* (row D.g.), *D. baculatus* (row D.b.), *R. capsulatus* (row R.c.), and *B. japonicum* (row B.j.) (5–8) are aligned above the sequences of the polypeptides encoded by *mvhG* and *mvhA* in *M. thermoautotrophicum* ΔH (row M.t.). The "Consensus" sequence indicates the locations at which the same amino acid residue occurs in at least four of the sequences, and asterisks indicate the highly conserved cysteinyl residues. The figures in boxes are the number of amino acids in these regions where little or no conservation of sequences can be detected. A dash within the five sequences indicates the absence of an amino-acid residue at that position in that sequence. The number of base pairs (bp) that form the intergenic regions separating the genes that encode the small and large subunits are given. There is no intergenic region between *mvhG* and *mvhA* (Fig. 1).

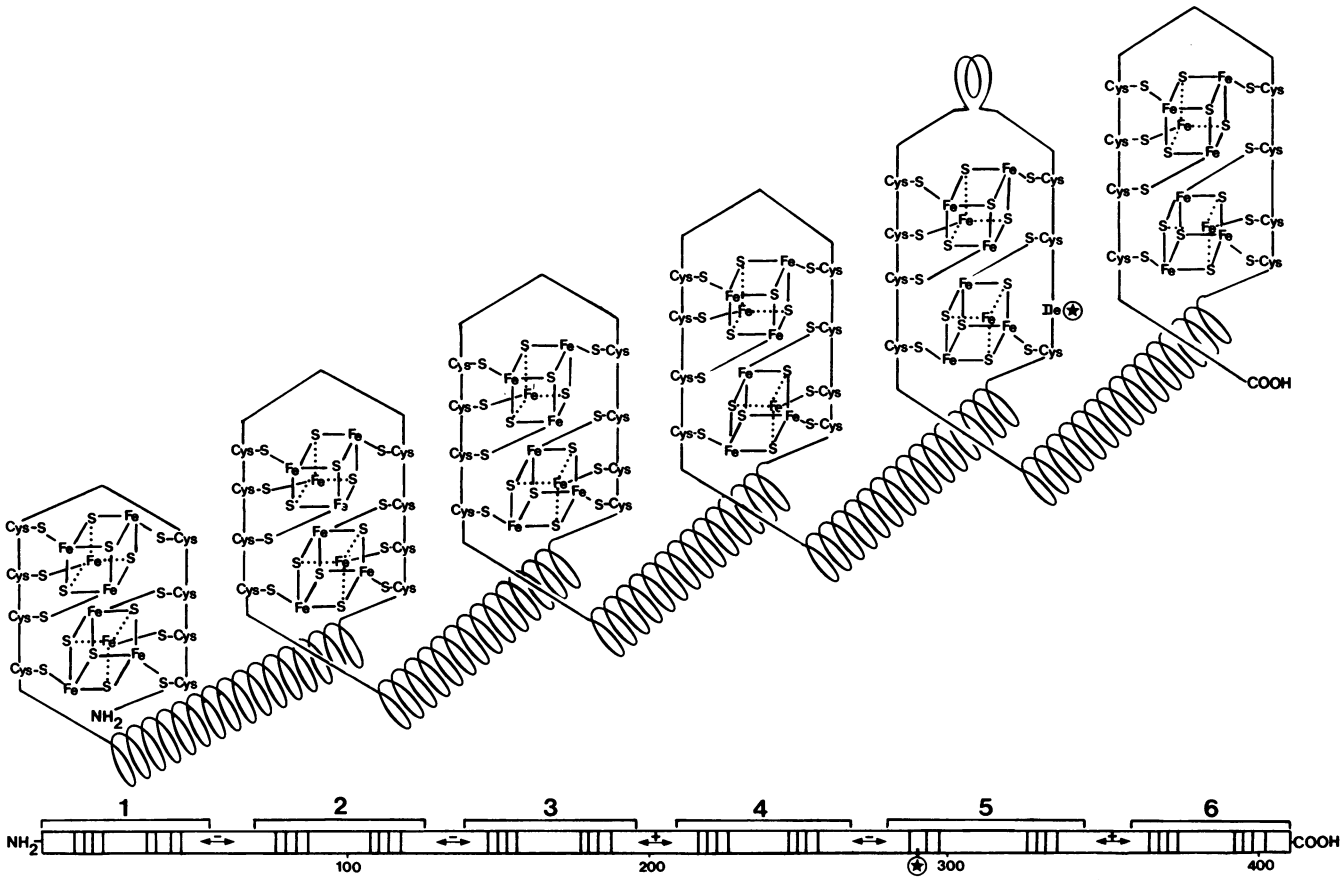

FIG. 3. Schematic representation of the structure of the product of the *mvhB* gene. The box at the bottom of the figure represents the polypeptide chain shown as a linear numbered array of amino acid residues with the locations of cysteinyl residues indicated by vertical lines. The Cys-II position (Table 1), which contains an isoleucyl residue, is starred, and the six ferredoxin-like domains are bracketed. The diagram above the box depicts the polyferredoxin with all of the potential Fe_4S_4 centers shown in detail. The helical charged spacer regions, predicted to be surface-located or to form hydrophilic pores by the SURFACEPLOT 1.2 program (16), are indicated as coiled lines. Domain 1 does not have a connector region, and domain 5 has an extended, negatively charged, and hydrophilic connector region.

polypeptide with the structure suggested in Fig. 3 could bind as many as 48 iron atoms in $12Fe_4S_4$ clusters and would have the capacity to store 12 electrons per molecule.

DISCUSSION

The *mvhDGAB* genes of *M. thermoautotrophicum* strain ΔH are tightly linked with features that indicate an operon organization. A single site of transcription initiation has been identified 42 bases upstream of the ATG translation initiation codon of *mvhD* (Fig. 1). Sequences conforming to the boxA and boxB consensus sequences for archaebacterial promoters (10) are found at the expected locations, the intergenic regions between the *mvh* genes are very short (12, 0 and 21 bases), and each gene is preceded by a potential ribosome-binding sequence. The *mvh* DNA sequences obtained from both *M. thermoautotrophicum* strain ΔH and strain Winter (1.8 kbp; Fig. 1) contain 88% identical bases and encode polypeptides with 90% identical amino acid residues. As the archaebacterial *mvhG* and *mvhA* genes appear to have evolved from the same ancestral sequences as genes that encode the small and large subunits of hydrogenases in both aerobic and anaerobic eubacteria (Fig. 2), the amino acid residues (especially cysteinyl residues) that have been conserved in all these enzymes presumably over a long evolutionary period must be very important in enzymatic activity and ligand binding.

A most intriguing result is the discovery that the *mvhB* gene of *M. thermoautotrophicum* ΔH apparently encodes a polypeptide that we term a polyferredoxin. To our knowledge this is the only example described so far of a polypeptide containing multiple, tandemly-repeated bacterial-ferredoxin-like domains (Table 1). The reduction of CO_2 to CH_4 ultimately requires the hydrogenase-mediated transfer of eight electrons from H_2 to the reaction center(s). Therefore, it seems probable that the polyferredoxin participates in these reaction(s) as an electron-transport protein associated with hydrogenase activity (1, 21). The polymeric nature of the polypeptide suggests that it could act as an electron conduit, transferring electrons from one Fe_4S_4 center of the molecule to the next (Fig. 3), possibly passing electrons through a membrane or into the complex subcellular structures known to house the enzymes responsible for methanogenesis (22, 23). An association of hydrogenase activities with membranes has been observed in *M. thermoautotrophicum* ΔH (1–3). Different domains of the polyferredoxin might supply electrons to different steps in the reduction of CO_2 to CH_4. Alternatively, the polyferredoxin molecule might have evolved for protective purposes. Methanogens require a very reduced environment, and a battery-like protein, storing reducing equivalents, could allow methanogens to survive transient exposures to oxidizing agents.

We thank the investigators who provided us with hydrogenase sequences prior to publication and V. Steigerwald and D. Livingston for help with primer-extension experiments and antisera production, respectively. Research at the Ohio State University was supported by Contract CR812774 from the Environmental Protection Agency and Grant DE-FG02-87ER13731-A001 from the Department of Energy. Research at Harvard University and at Massachusetts Institute of Technology was supported by National Institutes of Health Grant GM31574.

Table 1. Comparison of bacterial ferredoxins with domains of the polypeptide encoded by *mvhB*

Ferredoxin		First unit†		Connector‡		Second unit	Spacer
Cysteines		———I—II—III—IV*———				———I*—II*—III*—IV———	
Domains of *mvhB*						GEP	
1		MIIVNKEDCIRCGACQGTCPTAAIEVT				-PEDVIYCDICGKCVDICPTGALKLE	DLVVDEAGNTQGR
2		IVFNPDKCNECGDCVEVCPPQILKLD		EA	KVK	KVPLQGFCVMCQKCVDICPVGVIGVE	GIKEPAKVELEIE
3		GPIFIADCVGCGMCVPECPVDAITLD		KV	GGV	IEIDEDTCIKCGVCAQTCPWNAVYIS	GRKPEKRAKEIKK
4		FELDEDACIGCNTCVEACPGDFIVPR		T	SNL	TVELPAICTACGLCEQLCPVDAIDLE	VELGPAKPASEEG
5		LVWDEEKCDFIGACANICPNDAIRVV		11	KVD	EEPSFAMCTRCGACTVACPKGALSLV	DMDKVVDGEVVKRKR
6		VQYNPALCDQCGDCIEACPYDMLKLT			DEK	-VALKGFCILCDQCIPACPKGALSLK	
Group 4 M.b.		PATVNADECSGCGTCVDECPNDAITLD		EE	KGI	AVVDNDECVECGACEEACPNQAIKVE	E
M.t.		PALVNADECSGCGSCVDECPSEAITL*		EE	KGI	AVV*Q*E	
T.a.	60	VAVDWDCCIADGACMDVCPVNLYEWN	26	DKC		DPVRESDCIFCMACESVCPVRAIKIT	P
S.a.	37	VGVDFDLCIADGSCITACPVNVFQWY	9	KKA		DPVNEQACIFCMACVNVCPVAAIDVK	PP
Group 1 C.b.		AFVINDSCVSCGACAGECPVSAITQG			DTQ	FVIDADTCIDCGNCANVCPVGAPNQE	
						GYYDEP	
Group 2 C.l.		AHRITEECTYCAACEPECPVNAISAG			DEI	YIVDESVCTDCEACVAVCPVDCIIKV	
		KD					
Group 3 T.t.		PHVICQPCIGVQSCVEVCPVECIYDG			GDQ	FYIHPEECIDCGACVPACPVNAIYPE	48
						E	
Group 5 D.g.		PIEVNDDCMACEACVEICPDVFEMNE		EG	DKA	VVINPDSDLDCVEAIDSCPAEAIRS-	
D.v.H_2ase	27	VQIDEAKCIGCDTCSQYCPTAAIFGE		MG	EPH	SIPHIEACINCGQCLTHCPENAIYEA	-H_2ase

Sequences of ferredoxins and definitions of the groupings have been described (17). Ferredoxins compared are the six domains (1–6) of *mvhB*, *Methanosarcina barkeri* (row M.b.), a partial sequence of *Methanosarcina thermophila* (row M.t.) in which stars indicate uncertainties (18), *Thermoplasma acidophilum* (row T.a.), *Sulfolobus acidocaldarius* (row S.a.), *Clostridium butyricum* (row C.b.), *Chlorovium limicola* (row C.l.); *Thermus thermophilus* (row T.t.); *D. gigas* (row D.g.), and the ferredoxin domain of the hydrogenase of *D. vulgaris*, (row D.v.H_2ase) (19). The remaining 337 amino acids of this hydrogenase are not shown but are indicated under Spacer as H_2ase.

†Bacterial ferredoxins contain two units, each of which conforms to the consensus 7aa-Cys-2aa-Cys-2aa-Cys-3aa-Cys-8aa (aa = amino acid residues) separated by a connector region which is usually 3aa in length (17). The cysteinyl residues designated I, II, III, and IV and I*, II*, III*, and IV* cooperate to form two Fe_4S_4 centers in each ferredoxin molecule (as shown in Fig. 3).

‡*mvhB* domain 5, T.a., and S.a. have 11, 26, and 9 additional amino acid residues in the connector region, respectively.

1. Walsh, C. T. & Orme-Johnson, W. H. (1987) *Biochemistry* **26**, 4901–4906.
2. Fox, J. A., Livingston, D. J., Orme-Johnson, W. H. & Walsh, C. T. (1987) *Biochemistry* **26**, 4219–4227.
3. Kojima, N., Fox, J. A., Hausinger, R. P., Daniels, L., Orme-Johnson, W. H. & Walsh, C. T. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 378–382.
4. Prickril, B. C., Czechowski, M. H., Przybyla, A. E., Peck, H. D. & LeGall, J. (1986) *J. Bacteriol.* **167**, 722–725.
5. Menon, N. K., Peck, H. D., LeGall, J. & Przybyla, A. E. (1987) *J. Bacteriol.* **169**, 5401–5407, and correction (1988) **170**, 4429.
6. Li, C., Peck, H. D., LeGall, J. & Przybyla, A. E. (1987) *DNA* **6**, 539–551.
7. Leclerc, M., Colbeau, A., Cauvin, B. & Vignais, P. M. (1988) *Mol. Gen. Genet.* **214**, 97–107.
8. Sayavedra-Soto, L. A., Powell, G. K., Evans, H. J. & Morris, R. O. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8395–8399.
9. Brown, J. W., Thomm, M., Beckler, G. S., Frey, G., Stetter, K. O. & Reeve, J. N. (1988) *Nucleic Acids Res.* **16**, 135–150.
10. Zillig, W., Palm, P., Reiter, W. D., Gropp, F., Pühler, G. & Klenk, H. P. (1988) *Eur. J. Biochem.* **173**, 473–482.
11. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
12. Geliebter, J., Zeff, R. A., Melvold, R. W. & Nathenson, S. G. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3371–3375.
13. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
14. Chambers, I. & Harrison, P. R. (1987) *Trends Biochem. Sci.* **12**, 255–256.
15. Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.-A. & Böck, A. (1988) *Nature (London)* **331**, 723–725.
16. Parker, J. M. R., Guo, D. & Hodges, R. S. (1986) *Biochemistry* **25**, 5425–5431.
17. Otaka, E. & Ooi, T. (1987) *J. Mol. Evol.* **26**, 257–267.
18. Terlesky, K. C. & Ferry, J. G. (1988) *J. Biol. Chem.* **263**, 4080–4082.
19. Voordouw, G. & Brenner, S. (1985) *Eur. J. Biochem.* **148**, 515–520.
20. Hausinger, R. P., Moura, I., Moura, J. J. G., Xavier, A. V., Santos, M. H., LeGall, J. & Howard, J. B. (1982) *J. Biol. Chem.* **257**, 14192–14197.
21. Rouviere, P. E., Bobik, T. A. & Wolfe, R. S. (1988) *J. Bacteriol.* **170**, 3946–3952.
22. Mayer, F., Rohde, M., Salzmann, M., Jussofie, A. & Gottschalk, J. G. (1988) *J. Bacteriol.* **170**, 1438–1444.
23. Wackett, L. P., Hartweig, E. A., King, J. A., Orme-Johnson, W. H. & Walsh, C. T. (1987) *J. Bacteriol.* **169**, 718–727.