Gene, 90 (1990) 173–176
Elsevier

R22

52

GENE 03552

Nucleotide sequence of the *Methylobacterium extorquens* AM1 *moxF* and *moxJ* genes involved in methanol oxidation

(Methanol dehydrogenase; methylotroph; signal peptide; codon usage; recombinant DNA)

Dawn J. Anderson[a]*, Christina J. Morris[b], David N. Nunn[c]*, Christopher Anthony[c] and Mary E. Lidstrom[b]

[a] Department of Microbiology, SC-42, University of Washington, Seattle, WA 98195 (U.S.A.) Tel. (206)543-8440; [b] Keck Laboratories, 138-78, California Institute of Technology, Pasadena, CA 91125 (U.S.A.), and [c] Department of Biochemistry, University of Southampton, Southampton S09 3TU (U.K.) Tel. 703-559122

Received by M. Bagdasarian: 5 December 1989
Revised: 17 February 1990
Accepted: 20 February 1990

SUMMARY

The nucleotide sequence has been determined for two genes involved in methanol oxidation in the facultative methylotroph, *Methylobacterium extorquens* AM1. The two genes are *moxF*, encoding the 66-kDa subunit of the methanol dehydrogenase and *moxJ*, located immediately downstream from *moxF*, which encodes a 30-kDa protein with unknown function. This information completes the sequence of the 5.86-kb *Xho*I-*Sal*I fragment containing the *moxFJGI* region in *M. extorquens* AM1, and the structure of this gene cluster is presented. Evidence is presented that *moxJ* is also present in *Paracoccus denitrificans*. The aa sequence of MoxJ has provided little information concerning its function, but it does appear to contain a signal sequence suggesting a periplasmic location.

INTRODUCTION

Methylotrophic bacteria grow on reduced single-carbon compounds as sole sources of carbon and energy. *M. extorquens* AM1 can utilize one-carbon compounds such as methanol and methylamine, as well as many multi-carbon compounds. Growth of *M. extorquens* AM1 on methanol requires the periplasmic enzyme MDH to oxidize the substrate to formaldehyde (Anthony, 1982). The formaldehyde is then further oxidized to carbon dioxide, or assimilated into cell carbon through the serine cycle (Anthony, 1982).

Studies of the genetics of methanol oxidation in *M. extorquens* AM1 have revealed that at least seventeen genes play a role in methanol oxidation (Nunn and Lidstrom, 1986a,b; Lidstrom, 1990; Anderson and Lidstrom, 1988). Three of these genes encode the structural proteins of the methanol oxidation complex: *moxF* encodes the MDH large subunit, *moxI* encodes the MDH small subunit, and *moxG* encodes the cytochrome c_L (the electron acceptor of MDH) (Nunn and Lidstrom, 1986a,b; Anderson and Lidstrom, 1988; Nunn et al., 1989). *moxF*, *moxG* and *moxI* have been localized to an 8.6-kb *Hin*dIII fragment from *M. extorquens* AM1 (*Hin*dIII-FG), along with *moxJ*, encoding a 30-kDa polypeptide of unknown function. Expression studies have shown that these four genes are transcribed in the same direction in the order *moxF*, *moxJ*, *moxG*, *moxI* (Anderson and Lidstrom, 1988),

Correspondence to: Dr. M.E. Lidstrom, Keck Laboratories, 138-78, California Institute of Technology, Pasadena, CA 91125 (U.S.A.) Tel. (818)356-2132; Fax (818)356-2940.

* Present addresses: (D.J.A.) Fred Hutchinson Cancer Research Center, 1124 Columbia St., Seattle, WA 98104 (U.S.A.) Tel. (206)467-5000; (D.N.N.) Department of Microbiology, SC-42, University of Washington, Seattle, WA 98195 (U.S.A.) Tel. (206)545-7446.

Abbreviations: aa, amino acid(s); bp, base pair(s); kb, kilobase(s) or 1000 bp; *M.*, *Methylobacterium*; MDH, methanol dehydrogenase; Mox, methanol oxidation; nt, nucleotide(s); ORF, open reading frame; *P.*, *Paracoccus*; PQQ, pyrroloquinoline quinone; *tsp*, transcription start point(s).

0378-1119/90/$03.50 © 1990 Elsevier Science Publishers B.V. (Biomedical Division)

and data from Tn*5* insertion analysis (Nunn and Lidstrom, 1986a) suggest that these four genes may be cotranscribed. *moxF*, *moxG* and *moxI* have been identified in other methylotrophs (Harms et al., 1987; Machlin et al., 1988; Bastien et al., 1989) but the presence of *moxJ* was not determined in these studies.

The sequences of the two downstream genes of the *moxFJGI* cluster have been published (Nunn and Anthony, 1988; Nunn et al., 1989). The aim of the present study is to determine the sequence of *moxF* and *moxJ*, which would complete the sequence of a 5.86-kb *Xho*I-*Sal*I subfragment of *Hin*dIII-FG encoding the entire *moxFJGI* region.

EXPERIMENTAL AND DISCUSSION

(a) Nucleotide and protein sequence of *moxF* and *moxJ*

The nt sequence of the 4.48-kb region containing *moxF*, *moxJ* and 1.29 kb 5′ to these genes was determined (GenBank Accession Number M31108). Two ORFs were identified, both transcribed in the same direction and frame and both preceded by putative ribosome-binding sites. The first ORF extended from nt 1290–3170, which is within the region known to encode *moxF* (Anderson and Lidstrom, 1988). The translated sequence encodes 626 aa, which would produce a polypeptide of M_r 68434. A stretch of 26 aa, previously determined to be the N terminus of mature MDH (Nunn et al., 1989), corresponded exactly to the sequence of residues beginning at position Asn^{28}, and confirms this polypeptide to be the *moxF* product. A calculated M_r 65795 for mature MoxF is slightly larger than the value for MoxF reported in the literature (60000–64000; Anthony, 1982).

A second ORF was located 261 bp downstream from *moxF*, between nt 3232–4334, within the region known to encode *moxJ* (Anderson and Lidstrom, 1988). The translated sequence for this second ORF encodes 300 aa, which would produce a polypeptide of 32425. Assuming cleavage at the putative signal sequence (see next section), the mature polypeptide would be 275 aa in length with a M_r of 29929, which corresponds to the observed size for MoxJ on SDS–PAGE gels (30000; Anderson and Lidstrom, 1988). This, together with the ORF location, direction of transcription and lack of other ORFs with ribosome-binding sites, suggests this ORF corresponds to *moxJ*.

MoxF is known to have a periplasmic location, whereas the cellular location of MoxJ is unknown. However, both translated polypeptides contain a putative signal sequence at the N terminus, suggesting that both MoxF and MoxJ are periplasmic. The signal sequences are 27 aa and 25 aa in length for MoxF and MoxJ, respectively, and exhibit a typical Gram$^-$ signal sequence structure (von Heijne, 1985). The cleavage site occurs after the sequence Ala-Xxx-Ala, a feature conserved in all of the signal peptides for periplasmic Mox polypeptides.

In common with other bacteria with high G + C DNA (Machlin and Hanson, 1988; Harms et al., 1987; West and Iglewski, 1988) a strong preference was found for G and C in the third position for *moxF* and *moxJ*. Of the total codons used in the four sequenced *M. extorquens* AM1 *mox* genes, 91% have either G or C in their third position and some codons were never used [UUA (Leu), AUA (Ile), GUU (Val), UCU (Ser), CCU (Pro), CCA (Pro), AGA (Arg)].

(b) Similarity comparisons

The *moxF* from *P. denitrificans* and *M. organophilum* XX have previously been sequenced (Harms et al., 1987; Machlin and Hanson, 1988). A comparison of these genes with *moxF* described here reveals surprising similarities. The encoded mature polypeptides in all three cases consist of 599 aa, and strong conservation at both the aa and nt levels is observed in the coding regions. The *M. extorquens* AM1 mature MoxF shows 76% identity at the aa level with the *P. denitrificans* mature MoxF and 96% identity at the aa level with the *M. organophilum* XX mature MoxF. At the nt sequence level, the identities are 76% and 92%, respectively. In addition, the PQQ-linked alcohol dehydrogenase from *Acetobacter aceti* (Inoue et al., 1989) shows some identity to the complete *M. extorquens* AM1 MoxF sequence, with 30% at the aa level and 43% at the nt level. The high similarity between *moxF* for the two *Methylobacterium* strains is surprising, since recent taxonomic studies suggest that these two strains are distinct (Tsuji et al., 1990) and this idea is supported by the fact that the similarity decreases outside the coding regions. In the case of the *M. extorquens* AM1 MoxF as with the other two MoxF sequences (Harms et al., 1987; Machlin and Hanson, 1988), a homology search of GenBank and EMBL DNA and protein sequence data bases revealed no significant similarity to other sequences.

Although no other *moxJ* has been identified, the sequence of a putative ORF immediately downstream from *moxF* in *P. denitrificans* has been reported in a preliminary communication (Harms et al., 1989). A comparison of this sequence with *moxJ* reported here shows regions of strong identity at the aa level (Fig. 1). An apparent insertion has occurred in the *M. extorquens* AM1 MoxJ between aa 26–49, but a comparison of residues 49–300 shows 13% overall identity with one gap of a single aa. Three stretches are more highly conserved, with identities of 67% (aa 93–141), 58% (aa 209–239) and 56% (aa 277–300) each with no gaps. This suggests that *orf2* of *P. denitrificans* is the homologue of *moxJ*, and that the two genes are relatively well conserved. A search of GenBank and EMBL DNA and protein sequence data bases revealed no significant

```
                                10        20        30        40
AM1 moxJ     MSLVNGRRRTAASVVALTAALTALAALCAPAQAQDTKATSKAAEAAKP
Pd ORF2       MLIDFRQVCGAGAAALALASPALA--------------------

50          60        70        80        90          100
DAGTLRVCAA-EQPPLSMKDGSGLENRIATTVAEAMGRKAQFVWLGKPAIYLVRDGLEKK
DTTNLRVCASTKDAPFSDAQGAGFENKIAQVLADEMGATLDLVMLEKDAIYLVRDGIEKD

110         120       130       140       150         160
TCDVVIGLDADDPRVLTSKPYYRSGYVFLTRADKDLDIKSWSDPRLKEVSHMVVGFGTPG
LCDVLVGVDAGDERLLTTRPYYRSGYAFVTRQDRNFEGDKWQDVDQEGFDTFSYRLHSPA

170         180       190       200       210         220
EAMLKDIGRYEEDMAYLYSLVNFRAPRNQYTQIDPARMVSEVAIGKAEVGGAFGPDVARY
ETILKYTGRYEYNLIYQASLTNFEDRRNKYTQVEASRVITEVADGGADLAIVFAPEAARY

230         240       250       260       270         280
VRDSSTKLRMTPVPDDTQASDGRKMPQSFDQAMGVRKDDTALKAEIDAAALEKAKPKIEAI
VRDSREPLRMTLITNEIERSDGVIIPLQYSQSVGVSKTHPELLGPIEQALQSGKARIDAI

290         300
LKEEGVPVLPVSN
LTEEGIPLLPSS
```

Fig. 1. Comparison of deduced aa sequences for MoxJ. AM1 moxJ, MoxJ from *M. extorquens* AM1; Pd ORF2, ORF2 from *P. denitrificans*. Translation and analyses of DNA and DNA-derived polypeptide sequences were carried out using PC/Gene (Genofit SA., Geneva, Switzerland), Genepro version 4.0 (Riverside Scientific Enterprises, Seattle, WA) and DNA-Master (California Institute of Technology, Pasadena, CA). Last digit of numbers is aligned with corresponding aa. Identical aa are boxed. Dashes represent gaps introduced to maximize alignment.

similarity with either sequence. Therefore, although the sequence suggests that MoxJ acts in the periplasm, its mode of action remains unknown.

(c) Structure of the *moxFJGI* region

The sequence reported here overlaps that reported previously for *M. extorquens* AM1 *moxG* and *moxI* (EMBL Accession Number X07856), thereby completing the sequence for the entire 5.86-kb *Xho*I-*Sal*I fragment containing all four genes. Fig. 2 summarizes the structure and organization of this region. The four genes are closely spaced and transcribed left to right as shown in Fig. 2, in the order *moxF*, *moxJ*, *moxG* and *moxI*. The intergenic regions are 261, 142 and 105 bp, respectively, and although each contains a standard ribosome-binding-site sequence, none contains a sequence similar to known promoters for *E. coli* (Hawley and McClure, 1983), *Pseudomonas* (Mermod et al., 1984) or *Rhodobacter* (Gabellini and Sebald, 1986) strains. Tn*5* insertions throughout this region show polarity for complementation of *moxF* and *moxG*

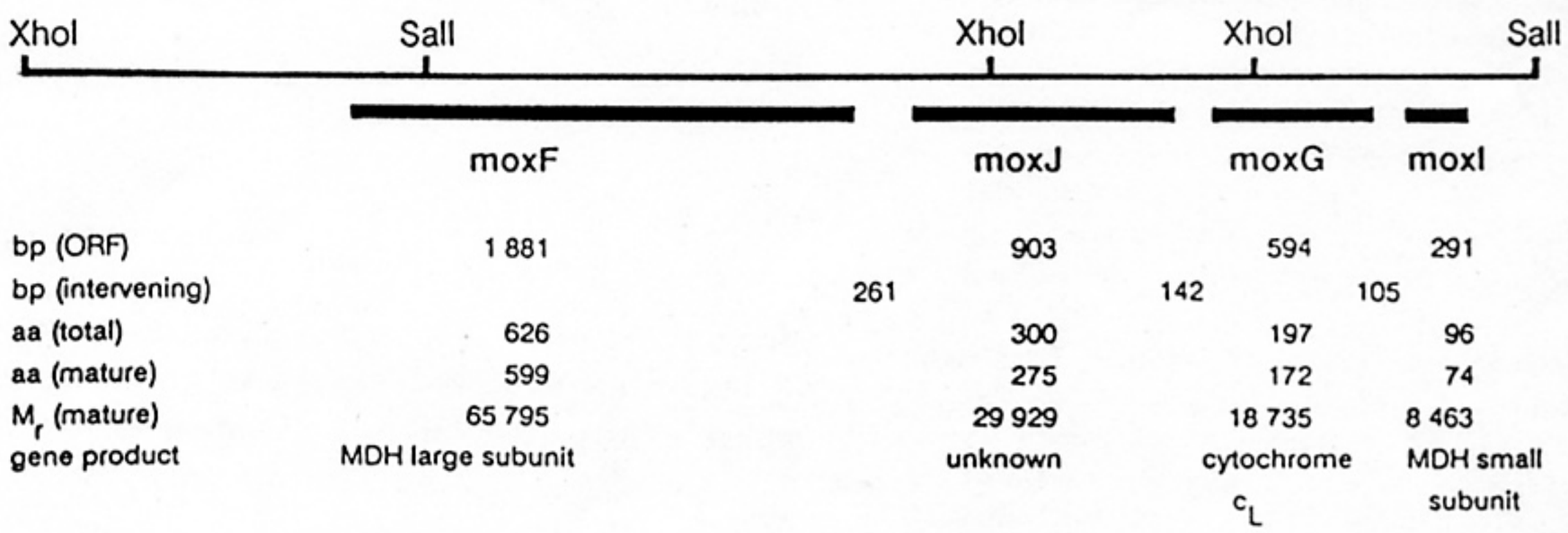


Fig. 2. The *moxFJGI* region of *M. extorquens* AM1. The entire 5.86-kb region shown has been sequenced [this paper (GenBank Accession Number M31108); Nunn and Anthony, 1988; Nunn et al., 1989 (EMBL Accession Number X07856)]. Sequencing was carried out using the dideoxy chain-termination method of Sanger et al. (1977). bp (ORF), length of ORF in bp; bp (intervening), distance between adjacent ORFs in bp; aa (total), total number of aa encoded by ORF; aa (mature), number of aa in mature polypeptide once signal sequence is cleaved; M_r (mature), molecular size of mature polypeptides in kDa.

mutants, suggesting that these genes are cotranscribed (Nunn and Lidstrom, 1986a). In *M. organophilum* XX, mapping data show that functionally homologous *moxF*, *moxG* and *moxI* are present and in the same order as in *M. extorquens* AM1, but it is not known whether a homologous *moxJ* is present (Machlin et al., 1988). Genes *moxF* and *moxI* are apparently transcribed separately in *M. organophilum* XX (Machlin and Hanson, 1988; Bastien et al., 1989).

The *tsp* for *M. extorquens* AM1 *moxF* has been mapped to nt 1121 and 1122, 168–167 bp upstream from the *moxF* start codon (D. Anderson, Ph.D. Thesis, University of Washington, 1988), and is within 1 bp of the corresponding *tsp* in *M. organophilum* XX (Machlin and Hanson, 1988). Upstream from both *moxF tsp*, a sequence exists at approx. the −10 and −35 positions that has some similarity to the *E. coli* consensus promoter, but expression in *E. coli* of *moxF* from this sequence has not been observed (Machlin and Hanson, 1988; Anderson and Lidstrom, 1988).

The conservation at the gene and multi-gene level of the *mox* system noted above apparently extends to other methylotrophs. Both type-I and type-II methanotrophs show a similar organization for *moxF* and *moxI* (Bastien et al., 1989; Lidstrom, 1990), and probes for *moxF* hybridize to specific genome fragments in a variety of methylotrophs (Stephens et al., 1988; Bastien et al., 1989).

(d) Conclusions

(1) Gene *moxF* encodes a 65795-Da polypeptide. The aa sequence is surprisingly similar to those reported for two other MoxF.

(2) Gene *moxJ* encodes a 29929-Da polypeptide; the aa sequence is similar to a previously unidentified ORF (ORF2) from *P. denitrificans*. We suggest *orf2* is the *moxJ* homologue and that MoxJ is a periplasmic protein.

(3) In *M. extorquens* AM1, *moxF*, *moxJ*, *moxG* and *moxI* are tightly clustered and are separated by intergenic regions of 261, 142 and 105 bp, respectively. A putative promoter region was identified approx. 180 bp upstream from the *moxF* start codon.

ACKNOWLEDGEMENTS

This work was supported by a grant to M.E.L. from the US Department of Energy (No. DE-FG03-87ER13753). D.N.N. was supported by an NIH Post-doctoral fellowship. C.A. was supported by SERC.

REFERENCES

Anderson, D. and Lidstrom, M.: The *moxFG* region encodes four polypeptides in the methanol-oxidizing bacterium *Methylobacterium* sp. strain AM1. J. Bacteriol. 170 (1988) 2254–2262.

Anthony, C.: The Biochemistry of Methylotrophs. Academic Press, London, 1982.

Bastien, C., Machlin, S., Zhang, Y., Donaldson, K. and Hanson, R.: The organization of genes required for the oxidation of methanol to formaldehyde in three type II methylotrophs. Appl. Environ. Microbiol. 55 (1989) 3124–3130.

Gabellini, N. and Sebald, W.: Nucleotide sequence and transcription of the *fbc* operon from *Rhodopseudomonas sphaeroides*. Eur. J. Biochem. 154 (1986) 569–579.

Harms, N., De Vries, G., Maurer, K., Hoogendijk, J. and Stouthamer, A.: Isolation and nucleotide sequence of the methanol dehydrogenase structural gene from *Paracoccus denitrificans*. J. Bacteriol. 169 (1987) 3969–3975.

Harms, N., VanSpanning, R., Oltmann, L. and Stouthamer, A.: Regulation of methanol dehydrogenase synthesis in *Paracoccus denitrificans*. Ant. v. Leeuwenhoek 56 (1989) 47–50.

Hawley, D. and McClure, W.: Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. 11 (1983) 2237–2255.

Inoue, T., Sunagawa, M., Mori, A., Imai, C., Fukuda, M., Takagi, M. and Yano, K.: Cloning and sequencing of the gene encoding the 72-kilodalton dehydrogenase subunit of alcohol dehydrogenase from *Acetobacter aceti*. J. Bacteriol. 171 (1989) 3115–3122.

Lidstrom, M.: Genetics of carbon metabolism in methylotrophic bacteria. FEMS Microbiol. Rev. (1990) in press.

Machlin, S. and Hanson, R.: Nucleotide sequence and transcriptional start site of the *Methylobacterium organophilum* XX methanol dehydrogenase structural gene. J. Bacteriol. 170 (1988) 4739–4747.

Machlin, S., Tam, P., Bastien, C. and Hanson, R.: Genetic and physical analyses of *Methylobacterium organophilum* XX genes encoding methanol oxidation. J. Bacteriol. 170 (1988) 141–148.

Mermod, N., Lehrbach, P., Reineke, W. and Timmis, K.: Transcription of the TOL plasmid toluate catabolic pathway operon of *Pseudomonas putida* is determined by a pair of coordinately and positively regulated overlapping promoters. EMBO J. 3 (1984) 2461–2466.

Nunn, D. and Anthony, C.: The nucleotide sequence and deduced amino acid sequence of the cytochrome c_L gene of *Methylobacterium extorquens* AM1, a novel class of *c*-type cytochrome. Biochem. J. 256 (1988) 673–676.

Nunn, D. and Lidstrom, M.: Isolation and complementation analysis of 10 methanol oxidation mutant classes and identification of the methanol dehydrogenase structural gene of *Methylobacterium* sp. strain AM1. J. Bacteriol. 166 (1986a) 581–590.

Nunn, D. and Lidstrom, M.: Phenotypic characterization of 10 methanol oxidation mutant classes in *Methylobacterium* sp. strain AM1. J. Bacteriol. 166 (1986b) 591–598.

Nunn, D., Day, D. and Anthony, C.: The second subunit of methanol dehydrogenase of *Methylobacterium extorquens* AM1. Biochem. J. 260 (1989) 857–862.

Sanger, F., Nicklen, S. and Coulson, A.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Stephens, R., Haygood, M. and Lidstrom, M.: Identification of putative methanol dehydrogenase (*moxF*) structural genes in methylotrophs and cloning of moxF genes from *Methylococcus capsulatus* Bath and *Methylomonas albus* BG8. J. Bacteriol. 170 (1988) 2063–2069.

Tsuji, K., Tsien, H., Hanson, R., DePalma, S., Scholtz, R. and LaRoche, S.: 16S ribosomal RNA sequence analysis for determination of phylogenetic relationship among methylotrophs. J. Gen. Microbiol. 136 (1990) 1–10.

von Heijne, G.: Signal sequences. The limits of variation. J. Mol. Biol. 184 (1985) 99–105.

West, S.E.H. and Iglewski, B.H.: Codon usage in *Pseudomonas aeruginosa*. Nucleic Acids Res. 16 (1988) 9323–9335.