ABSTRACT

The two genes encoding the class IIS restriction-modification system Mboll from Moraxella bovis were cloned separately in two compatible plasmids and expressed in E. coli RR1 AM15. The nucleotide sequences of the Mboll endonuclease (R. Mboll) and methylase (M. Mboll) genes were determined and the putative start codon of R. Mboll was confirmed by amino acid sequence analysis. The mbollR gene specifies a protein of 416 amino acids (MW: 48,617) while the mbollM gene codes for a putative 260-residue polypeptide (MW: 30,077). Both genes are aligned in the same orientation. The coding region of the methylase gene ends 11 bp upstream of the start codon of the restrictase gene. Comparing the amino acid sequence of M. Mboll with sequences of other N6-adenine methyltransferases reveals a significant homology to M. RsrI, M. HinfI and M. DpnA. Furthermore, M. Mboll shows homology to the N4-cytosine methyltransferase BamHI.

INTRODUCTION

Class IIS restriction endonucleases (ENases) and methyltransferases (MTases) are enzymes that recognize specific, mostly asymmetric DNA sequences 4–6 bp in length (1). ENases cleave within both DNA strands at fixed positions 1 to 20 bases to the right of their target site. MTases modify DNA by the addition of a methyl group either at adenines or at cytosines in the recognition sequence (2). Like the enzymes of class II R-M systems, ENases require only Mg++ ions, and MTases need only S-adenosylmethionine (AdoMet) as sole cofactor. A recent compilation of R-M systems listed 28 enzymes belonging to class IIS (3).

Class IIS enzymes appear especially suited as model proteins to study protein-DNA interaction, because the spatial separation of their DNA recognition and cleavage sites suggests a two domain structure: one being necessary for DNA recognition, the other bearing the catalytic center to introduce a double strand break into the DNA. If so, the different functions could be uncoupled. DNA specificity may then be altered by switching domains as has been achieved in phage repressors (4), whereas abolishing catalytic activity should not affect the sequence specificity of binding. Moreover, a thorough examination of the structural and functional features of class IIS ENases may help to improve our knowledge about the mode of endonucleolytic activity. According to the prevalent model of 'allosteric activation' which emerged from a combination of crystallographic, biochemical and genetic analyses of the EcoRI system (5, 6, 7), it is evident that upon binding to its target sequence, this ENase isomerizes to an active conformation which allows for DNA strand cleavage. To elucidate the molecular mechanisms directing the mode of action of class IIS ENases the cloning and expression of their genes is required. To date, only two class IIS R-M systems, HgaI from Haemophilus gallinarum and FokI from Flavobacterium okeanokoites, have been cloned completely (8, 9).

Mboll ENase recognizes a pentanucleotide sequence, 5'-GAAGA-3', and cleaves the DNA 8 and 7 nucleotides, respectively, downstream, leaving a single 3' protruding nucleotide (10). Furthermore, the Mboll target sequence displays a peculiar kind of asymmetry in that one strand is composed of only purines and the other contains only pyrimidine nucleotides. In this respect, Mboll is related to Ksp632I and MnlI which recognize the hexanucleotide 5'-CTCTTC-3' and the tetranucleotide 5'-CCTC-3', respectively (11). With respect to the cleavage position, HphI is the closest class IIS enzyme to Mboll. It also cuts to the right of the recognition sequence, 8 nucleotides away on the top and 7 on the bottom strand (12).

Mboll MTase methylates the 3' adenine, producing 5'-GAAGMeA-3' (13). After replication, however, one daughter DNA molecule will be unmethylated and therefore sensitive to restriction. This is in contrast to MTases with palindromic targets where both daughter molecules are hemimethylated, thus providing sufficient protection against the cognate ENase. It is unclear how M. bovis DNA is protected against Mboll ENase activity during replication. Whether the cytosine residues within the Mboll complementary strand, 5'-TCTTC-3', are modified by an alternative activity of the Mboll methylase or perhaps by a completely different methylase is still a matter in question. Our report about the cloning, sequencing and expression of the genes encoding the Mboll R-M system does suggest the existence of a gene coding for a Mboll site-specific methylase but does not provide a comprehensive answer how methylation functions in vivo.

* To whom correspondence should be addressed
MATERIALS AND METHODS

Bacterial strains and plasmids

E. coli strain RRIΔM15: leu, pro, thi, strA, hsd r-m-, lacZΔM15 F' lac' ZΔM15 pro+ (14), and plasmids pUC9 (15), pACYC184 (16) and pMO320 (17) were used for cloning and expression. *Moraxella bovis* (ATCC 10900) was purchased from American Type Culture Collection and cultivated in dYT Medium at 37°C under aerobic conditions (18).

Enzymes and chemicals

Restriction endonucleases were purchased either from New England Biolabs or Boehringer Mannheim. T4-DNA Ligase was from BRL. Terminal deoxynucleotidyltransferase was from Boehringer Mannheim. Taq-Polymerase was from Perkin-Elmer-Cetus. Proteinase K and RNase A were from Sigma Chemicals. IPTG was from Bachem Biochemicals. All radiochemicals were obtained from Amersham Buchler.

Purification of *MboII* ENase

About 250g of *M. bovis* cells were suspended in 300ml of 10mM Tris-HCl buffer (pH 7.5) containing 10mM 2-mercaptoethanol and disrupted by French press. The cell debris were removed by centrifugation (100,000g for 3h). To the supernatant a 25% solution of streptomycin sulfate was added to a final concentration of 5%. The precipitate was removed by centrifugation and solid ammonium sulfate was added to 50% saturation. After removal of the precipitate solid ammonium sulfate was added to a final concentration of 70%. The resulting precipitate was collected by centrifugation, dissolved in 10mM K-phosphate buffer (pH 7.4) containing 10mM 2-mercaptoethanol, 10% glycerol (buffer 1), and dialysed against the same buffer. The dialysate was loaded onto a phophocellulose column (3 x 29cm; Whatman P-11) and chromatographed with 1000ml of a linear KCl gradient from 0.0—0.8M in buffer 1. The R.*MboII* activity was eluted at 0.3—0.4M KCl. The active fraction was dialysed against buffer 1, loaded onto a Hydroxyapatite column (2 x 27cm; CalbioChem), and eluted with 400 ml of a linear gradient of K-phosphate (pH 6.8) from 0.1—0.5M. The active fraction was dialysed against Heparin-Agarose buffer (10mM Na-phosphate buffer (pH 6.8) containing 10mM 2-mercaptoethanol), loaded on a Heparin-Agarose column (2 x 13cm), and eluted with 400 ml of a linear gradient of NaCl from 0.0—0.6M. R.*MboII* activity was eluted at 0.2—0.3M NaCl. The active fractions were washed and concentrated by ultrafiltration with Centricon tubes (Amicon Corp.). The final recovery was about 2mg of homogenous R.*MboII*. Running aliquots of purified R.*MboII* on a 12% SDS-PAGE gel gave rise to a unique band of 49 kD. Gel filtration on a Superose 12 HR 10/30 (Pharmacia) FPLC column revealed that *MboII* ENase is a monomer under native conditions (data not shown).

Analysis of the N-terminal amino acid sequence of R.*MboII*

500 pMol of homogenous R.*MboII* were loaded onto an Applied Biosystems 470A gas-phase protein sequencer. The phenylthiohydantion (PTH) derivates of the amino acids were identified with an Applied Biosystems 120A PTH-analysyer connected to the sequencer. The first 19 PTH-amino acids, except the PTH-Serines in position 6, 13 and 14, were unambiguously identified.

Preparation of genomic DNA fragments with DNasel

Total cellular DNA from 1g of frozen *M. bovis* cells was purified using Proteinase K/SDS, phenol/chloroform extraction, and RNase A digestion, according to methods previously described (18, 19). 5µg *M. bovis* DNA were incubated in 50µl of 33mM Tris-HCl (pH 7.6), 10mM MnCl2 containing 3ng of DNasel (Boehringer Mannheim) at 37°C. After 2min the reaction was stopped by adding 5µl of 0.1M EDTA at pH 8.0. The digested DNA was purified on a 5 % polyacrylamide gel. DNA of the required size was cut out and eluted overnight in 333 buffer (300mM NaCl, 30mM Tris-HCl pH 7.6, 3mM EDTA). The eluted DNA was precipitated with ethanol, dried and dissolved in 30µl TE-buffer (20mM Tris-HCl pH 7.6, 1mM EDTA).

Selection of clones expressing the *MboII* M phenotype

DNA of *M. bovis* was digested by DNasel to produce fragments with an average size of 2.5Kb. 10pMol digested DNA were tailed by adding poly dC residues and then annealed into the *PstI* cut, poly dG tailed expression vector pUC9. Annealed DNA was transformed into competent *E. coli* RRIΔM15. 22,000 transformants were collected from Yeast-Tryptone(YT), ampicillin plates by rinsing with 5ml saline. 2ml of this mixture were inoculated into 400ml dYT-Medium containing 200µg/ml ampicillin and grown to saturation. 5µg of the plasmid population purified by CsCl/EtBr centrifugation were incubated with an at

Fig. 1. Plasmid pMboM1.1 carrying the *MboII* MTase gene is resistant to digestion with *MboII* restrictase. 1, pUC9 digested with *PstI*. 2, pMboM1.1 digested with *PstI*. 3, pUC9 digested with *MboII*. 4, pMboM1.1 and pUC9 digested with *MboII*. 5, pMboM1.1 digested with *MboII*. 6, pMboM1.1 undigested. M. BRL 1Kb-ladder.

Fig. 2. Schematic map of the *MboII* R-M region. Bold arrows indicate extent and orientation of the *MboII* genes. Below the genes, the size of DNasel fragments inserted in pUC9 are shown. MboR1.3 represents a fragment that was isolated from chromosomal *M. bovis* DNA by the polymerase chain reaction (PCR). — signifies N- and C-terminal specific primers. To the left of the fragments, the *MboII* activities specified by the clones are summarized. The M column refers to the state of *MboII* modification of the plasmids, as measured by their resistance to *MboII* digestion 'in vitro'. The R column refers to whether or not *MboII* activity was detected in sonicated cell extracts. Symbols: +, presence of ENase/MTase; —, absence of ENase/MTase. Details of subcloning of MboR1.3 are outlined in Fig. 4B.
least ten fold surplus of MboII restrictase in a total volume of 100μl for 3h at 37°C. Transformation of this digest into E. coli RR1ΔM15 gave rise to 450 ampR colonies. Plasmid DNA from 12 individual clones was prepared and incubated with MboII as described above.

DNA sequence analysis

DNA was sequenced either by the Maxam-Gilbert method (20) or by the chain termination method (21) with the T7 polymerase sequencing kit provided by Pharmacia. Oligo primers were synthesized with a 380A Applied Biosystems DNA synthesizer using the cyanethyl phosphoramidite chemistry.

Southern hybridisation and colony screening

Southern blots were performed as described (22). Colony screening was done according to standard protocols (22), except that filters were washed only once in 6× SSC for 15min after hybridisation.

Fig. 3. Nucleotide sequence of the mboII R-M genes and the deduced amino acid sequence of the MboII MTase and ENase. Shine-Dalgarno sequences are marked **. The N-terminal amino acids of the ENase that were determined by protein sequencing are boxed. The positions of the MboII recognition sites in the ENase gene are shown. A potential -10/-35 region referring to the gene is indicated in the 3' end of the MTase gene.

<table>
<thead>
<tr>
<th>SD</th>
<th>Nucleic Acids Research, Vol. 19, No. 5 1009</th>
</tr>
</thead>
</table>

**Fig. 3.** Nucleotide sequence of the mboII R-M genes and the deduced amino acid sequence of the MboII MTase and ENase. Shine-Dalgarno sequences are marked **. The N-terminal amino acids of the ENase that were determined by protein sequencing are boxed. The positions of the MboII recognition sites in the ENase gene are shown. A potential -10/-35 region referring to the gene is indicated in the 3' end of the MTase gene. DPPY and FXGXG sequences of M.AfrtoIl are underlined. The sequence of an N-terminal tetrapeptide that would be translated if the ATG codon at position 100 was used as start codon is shown in brackets, with lower case letters.
Polymerase chain reaction

Total DNA from *M. bovis* was used as template. PCR reactions (100 µl) contained 200 pmol of each primer, 500 ng of template DNA, 10 mM of each dNTP, 67 mM Tris-Cl (pH 8.8), 16.6 mM ammonium sulfate, 6.7 mM MgCl₂, 6.7 mM EDTA, 10 mM 2-mercaptoethanol, 5 units Taq DNA polymerase. Reaction mixtures were overlaid with paraffin and reactions were carried out using the Biozym DNA incubator. The cycle program was programmed as follows: 1 min at 50°C (annealing), 5 min at 70°C (extension) and 1 min at 95°C (denaturation). This profile was repeated for 35 cycles. The aqueous layer of the reaction mixtures were pooled and put onto a 1% agarose gel. The amplified DNA band was cut out, eluted and resuspended in TE-buffer.

RESULTS AND DISCUSSION

Cloning of the *mboII* M gene

Recombinants expressing a specific MTase activity will survive digestion by the corresponding ENase, and hence can be recovered by transformation back into an *E. coli* strain, if the formation of a new methylation pattern does not exert any damaging effects to the host cell. *E. coli* K12 is known to possess restriction systems that degrade DNA specifically when it is methylated (23-26). Because the *MboII* MTase modified sequence, 5'-GAAGmeA-3', overlaps with the suggested consensus sequence of the *mrr* restriction system, 5'-GmeAC-3' or 5'-CmeAG-3', we used *E. coli* RR1M15, a derivative of RR1, which is known to carry a deletion spanning the entire *mrr* region (26), as the host in constructing a genomic *M. bovis* library. *M. bovis* DNA was partially digested with DNasel and inserted into pUC9. After transformation into RR1M15 recombinants carrying the *MboII* MTase gene will be resistant to cutting by the *MboII* ENase. Pooled plasmid DNA was recovered, cut with *MboII* and retransformed into RR1M15.

From 450 clones obtained after the first round of selection with *MboII* digestion, three, out of 12 recombinants tested, were found to be protected against the ENase attack. Restriction mapping revealed that two of the *MboII* resistant clones had an identical insert of 2.3 Kb, while the third contained a fragment of 1.1 Kb. The latter, designated *pMboM1.1* (Fig. 1), was chosen for DNA sequence analysis.

DNA sequence analysis and gene organization

The nucleotide sequence of the 1.1 Kb DNasel fragment of plasmid *pMboM1.1* (Fig. 2) was determined on both strands (Fig. 3). Only one long open reading frame was found that could code for a protein of a molecular weight corresponding to the average size of N6A-MTases (27). Potential ATG start codons were detected at nucleotide positions 100, 112 and 139, and the termination codon UAA at nucleotide position 892 (Fig. 3). We propose the ATG codon at 112 to be the initiation site for translation of the MTase gene since only this one is preceded by a putative Shine-Dalgarno sequence, ATGG, in the appropriate position. The corresponding reading frame predicts a protein of 260 amino acids (MW: 30,077), displaying the motif DPPY which is characteristic of adenine-specific MTases (27).

Examination of the region downstream of the MTase gene revealed on the same strand, only 11 bp apart, the start of a second open reading frame. The assignment of the *MboII* ENase gene to this coding region (Fig. 3) was achieved by N-terminal amino acid sequence analysis of *R.MboII* (s. Materials and Methods). A Shine-Dalgarno sequence, AAGG, positioned 6 nucleotides 5' to the start codon supported the assignment. The appropriate reading frame, however, covered only a part of the *mboII* gene coding for the first 59 amino acids. In order to isolate clones carrying the rest of the *mboII* gene, we screened our DNasel library of the *M. bovis* genome with a 21mer oligonucleotide (5'-CATAACCGGTATGGTTCAC-3') derived from the 3' end of the *pMboM1.1* insert. Ten colonies, among 50000 tested, appeared positive. The plasmid *pMboR1.5* (Fig. 2) contains a 1.5 Kb insert covering by far the largest section of the *mboII* gene (3' end at nucleotide position 1990). Since no in-frame stop codon could be discovered on this fragment, we performed a second screening, now using a 20mer oligonucleotide (5'-GTGGATGAATATCGTTATTT-3') derived from the 3' end of the *mboII* part on *pMboR1.5*. Testing of 2000 transformants gave rise to the plasmid *pMboR2.4* (Fig. 2), possessing an insert of 2.4 Kb on which we localized the UAA stop codon of *mboII* at nucleotide position 2151 (Fig. 3). By this means, we established the coding region of *mboII* as being 1248 bp long, corresponding to a protein of 416 amino acid residues (MW: 48,617).
The base composition of the MboII genes is characterized by a high A+T content: 68% for the ENase and 65% for the MTase. This is higher than the average A+T content of M. bovis (57%, 28). The base composition is reflected in the codon usage: A and U are strongly preferred nucleotides in the third position, less so at the second and first position.

The operon-like arrangement in the MboII system appears logical. Close linkage and cotranscription of the two genes in the order M to R assures that MTase will always be coproduced with restriction activity and the cellular DNA will be modified and protected, whereas independent expression of the ENase gene should result in lethality. In addition to the MboII system, the organization of only one other class II R-M system (FokI) has been determined (29). Different gene arrangements of 11 class II R-M systems (BsuRI, Ddel, DpnII, EcoRI, EcoRV, HhaII, HinfI, MspI, PaeR71, PstI, TaqI) have been recently reviewed (30). The arrangement of the MboII genes is similar to the FokI, HhaII, HinfI, PaeR71 and TaqI genes, but the length of the intergenic region is different ranging from a minimum of 11 bp between the MboII genes to a maximum of 132 bp between the TaqI genes.

Expression of the mboIIR gene

Since the length of the MboII R-M system did not exceed the average size of 2.5 Kb of DnaseI generated fragments used to establish the M. bovis library, we tried to isolate recombinants possibly carrying both genes. We analysed 450 ampR transformants which were recovered from the first round of selection for MboII M activity (s. Materials and Methods). Attempts, however, to identify clones expressing both, methylase and restrictase activity, either by infecting phage restriction or by in vitro assays of cell extracts (30) were unsuccessful. A plausible explanation of this result may be that the E. coli-like promoter sequence we identified at the 3' end of the MTase gene (Fig. 3) does promote the transcription of the ENase in E. coli, while it is possibly without function in M. bovis. It is reasonable to suppose that an uncoordinated expression of the MboII genes will affect the viability of their host strain. Furthermore, the phenomenon that an endogenous promoter for ENase gene expression is located within its cognate MTase gene may possibly account for the negative outcome of various approaches to clone R-M systems as a whole. In the case of some Bacillus species only MTase genes have been cloned, but none has yet been associated with a functioning restriction gene (31, 32).

A different approach was therefore used to clone the ENase gene. First, the PCR-generated fragment MboR1.3, carrying the mboIIR gene alone was ligated into the unique SmaI site of pMO320 (17). pMO320, a plasmid bearing the col El origin of replication, contains a weak, synthetic promoter [specific β-galactosidase activity (33): 390 U (+IPTG), 0.22 U (-IPTG); for comparison, wt lac promoter: 1000 U (+IPTG), 1 U (-IPTG)], two copies of wt lac operator and the lacI gene.

---

Fig. 5. Dot-plot representation of homologies between M.MbolI and four other MTases. A: M.RsiI; B: M.Hinfl; C: M.DpnA; D: M.BamHI The dot matrixes were generated by the program COMPARE. Diagonal lines indicate regions of strong homology. For all comparisons the window size and the stringency values were 30 and 14, respectively.
possessing the 1st promoter (Fig. 4). Under noninducing conditions lac repressor prevents the transcription of DNA sequences which are cloned to the Smal restriction site located 3' to the operator elements. The general applicability of such a 'conditional-lethal' vector system had been demonstrated previously by the successful cloning of the PstIIR gene (17). Under these conditions, it became possible to clone the mboU gene despite the absence of any MboII MTase activity. The recombinant plasmid, named pMboR1.3 (Fig. 4), was digested with EcoRI. A 3 Kb fragment covering lacI, mboU and the lac control region was prepared and subcloned into the EcoRI site of pACYC184 whose origin of replication is different from that of pMboM1.1. The resulting plasmid, pMboR3.0, was transformed into RR1AM15 harboring the plasmid pMboM1.1 where MboII MTase was constitutively expressed. When examining the new strain, E. coli RR1 M15-Mbo, we found that the steady state level of R.MboII activity was at least a factor of 20 lower than in Moraxella bovis. As yet, the reason for this low activity is not clear.

**Location of the MboII R-M system in M. bovis**

To determine whether the 7 Kb plasmid present in M. bovis carries the MboII R-M system, we performed a Southern-blot, using as a labeled probe both the mboIIM gene and the PCR fragment MboR1.3. The mboIIM gene as well as the mboU gene hybridize to a >20 Kb HindIII fragment of total DNA, but not to plasmid DNA (data not shown). Thus, it is likely that the MboII R-M system is located on the chromosome of M. bovis.

**Protein sequence comparisons**

The amino acid sequences of the MboII proteins were compared with other known ENase and MTase sequences (EMBL protein library) by using the GAP program of UWCGG which calculates the degree of similarity between two protein sequences. As anticipated, comparing R.MboII and MboII revealed no significant homology. Similarly, no homology could be detected between MboII and other sequenced ENases.

MboII is an adenine-specific MTase generating N6mA. All N6mA-MTases analysed to date share two common features: the tetrapeptide DPPY and its variants involved in the methylation of the exocyclic amino group and the FXGXG motif suggested to be the binding site of AdoMet (34). The two conserved sequence motifs are found within the reading frame of the mboU reading frame defined by the ATG at nucleotide position 112: DPPY at codons 27-30 and FXGXG at codons 220-224. Global alignment of MboII to all available amino acid sequences of adenine-specific MTases demonstrated that the MboII MTase sequence is closely related to M.RsrI from Rhodobacter sphaeroides. We found an overall identity of 38% and a further 19.6% if conservative replacements are permitted. The extent of nucleotide identity is 44%. Prior to this study, only the sequence comparisons between M.NalIII and M.FokI had disclosed a comparable high degree of homology (35). By then, the maximal amount of amino acid sequence identity found among N6mA-MTases was about 30% (36). As shown in Fig. 5(A-C) the homologous regions are arranged in at least three clusters: DPPY at the N-terminus, FXGXG at the C-terminus and an as yet functionally not characterized region between them. The evident homology between MboII and M.Hinfl (protein identity: 32%), M.DpnA (protein identity: 38.5%) as well as M.BamHI (protein identity: 31.4%) (Fig. 5 D) is expected because similar degrees of homology have been reported for M.RsrI (37). Furthermore, the relatedness between M.MboII and M.BamHI, a N4-methylcytosine MTase, represents an additional example to confirm the notion that N4mC-MTases are more related to N6mA-MTases than to 5-methylcytosine (5mC) MTases. MboII has no significant homology to any of the known Smc-MTases.

**ACKNOWLEDGEMENTS**

We thank K. Beyreuther for providing protein sequencing facilities, K. Otto for synthesizing the oligonucleotides, B. Kisters-Woike for help with the computer work, S. Corbach for technical assistance and B. Jack for critically reading the manuscript. This work was supported by a fellowship of the Studienstiftung des deutschen Volkes to H.B. and by the Deutsche Forschungsgemeinschaft through 'Schwerpunkt Gentechnologie' to B.M.-H.

**REFERENCES**