Purification, cloning and sequence analysis of *RsrI* DNA methyltransferase: lack of homology between two enzymes, *RsrI* and *EcoRI*, that methylate the same nucleotide in identical recognition sequences

Wiweka Kaszubska, Christopher Aiken, C.David O'Connor¹, and Richard I.Gumport*

Department of Biochemistry and College of Medicine, University of Illinois, Urbana, IL 61801, USA and ¹Department of Biochemistry, University of Southampton, Southampton SO9 3TU, UK

Received September 26, 1989; Revised and Accepted October 24, 1989

EMBL accession no. X16456

ABSTRACT

RsrI DNA methyltransferase (M-*RsrI*) from *Rhodobacter sphaeroides* has been purified to homogeneity, and its gene cloned and sequenced. This enzyme catalyzes methylation of the same central adenine residue in the duplex recognition sequence d(GAATTC) as does M-*EcoRI*. The reduced and denatured molecular weight of the *RsrI* methyltransferase (MTase) is 33,600 Da. A fragment of *R. sphaeroides* chromosomal DNA exhibited M-*RsrI* activity in *E. coli* and was used to sequence the *rsrIM* gene. The deduced amino acid sequence of M-*RsrI* shows partial homology to those of the type II adenine MTases *HinfI* and *DpnA* and N⁴-cytosine MTases *BamHI* and *PvuII*, and to the type III adenine MTases *EcoP1* and *EcoP15*. In contrast to their corresponding isoschizomeric endonucleases, the deduced amino acid sequences of the *RsrI* and *EcoRI* MTases show very little homology. Either the *EcoRI* and *RsrI* restriction-modification systems assembled independently from closely related endonuclease and more distantly related MTase genes, or the MTase genes diverged more than their partner endonuclease genes. The *rsrIM* gene sequence has also been determined by Stephenson and Greene (*Nucl. Acids Res.* (1989) 17, this issue).

INTRODUCTION

DNA methylation is involved in diverse biological functions ranging from methyl-directed mismatch repair and the restriction-modification (R-M) phenomenon to the regulation of gene expression (see ref. 1 for a review). The methyltransferases (MTases) of the R-M systems recognize specific DNA sequences and deposit methyl groups on target nucleotides rendering the modified sequences resistant to cleavage by the corresponding endonucleases. MTases use S-adenosylmethionine (AdoMet) as the methyl donor, and the N⁶-amino group of adenine, the N⁴-amino group of cytosine, or the C5 carbon of cytosine as methyl acceptors. The R-M systems have been classified as types I, II, and III according to their cofactor requirements and polypeptide subunit compositions. See *Gene* (1988) 74, No. 1 for a compendium of recent research on the DNA MTases.

The type II R-M enzymes provide attractive systems for the study of sequence-specific protein-DNA interactions, because they are structurally simple and can use oligodeoxyribonucleotides as substrates. They also allow for comparative biochemical and evolutionary studies, because some R-M systems, although isolated from different

organisms, have the same DNA target sequences and catalyze identical reactions. For these reasons, we and others have been studying the *RsrI* and *EcoRI* R-M enzymes. The two endonucleases cleave duplex d(GAATTC) between the G and A residues (2-4), and the MTases methylate the central adenine (5, this work). We report here the purification to homogeneity and partial characterization of M-*RsrI*. In addition, we have cloned and sequenced the *rsrIM* gene and compared its predicted amino acid sequence to other methyltransferases.

The nucleotide sequences of numerous MTases and some of their accompanying endonucleases have been determined (6). Comparisons of the protein sequences encoded by these genes have revealed several general relationships (7-10). There is little sequence similarity between a given methyltransferase and its corresponding endonuclease even though they recognize the same DNA sequence. There are, however, significant similarities among the sequences of the MTases that methylate cytosine residues at C5 or among those that methylate adenine residues at N⁶. Recently, similarities between N⁴-cytosine and N⁶-adenine MTases have also been detected (9, J. Brooks and D. Landry personal communication, 11). In addition, five families of methyltransferases have been classified according to their DNA recognition sequences and amino acid sequence homologies (7). We report here that M-*RsrI* shares homology with the adenine MTases *HinfI*, *DpnA*, *EcoP1*, and *EcoP15*. Furthermore, the enzyme is homologous to the N⁴-cytosine MTases *BamHI* and *PvuII*.

Most surprisingly, M-*RsrI* lacks homology to its functional analogue M-*EcoRI*. Stephenson *et al.* (12) showed that the *RsrI* endonuclease (R-*RsrI*) is highly homologous to the *EcoRI* endonuclease (R-*EcoRI*) and suggested that these two genes evolved from a common ancestor. We anticipated a similar relationship between the *RsrI* and the *EcoRI* methyltransferases that are paired with these endonucleases. However, the lack of homology between the two MTases suggests that they may have evolved from different ancestors and joined their partner endonucleases in separate events to form the two distinct but catalytically-identical R-M systems that reside in *R. sphaeroides* and in *E. coli*. Divergence of the MTases with more stringent conservation of the paired endonucleases could also explain the dissimilarity of the two MTases. An initial announcement of the purification of M-*RsrI* has appeared (13) and the sequence of the *rsrIM* gene has been independently determined by Stephenson and Greene (*Nucl. Acids Res.* (1989) 17, this issue).

MATERIALS AND METHODS

a) Bacterial Strains, Phage, Plasmids

R. sphaeroides strain 630 was obtained from S. Kaplan and cultivated in YEP medium (Bacto-peptone, 3 g; Bacto-yeast extract, 2 g per liter) for the construction of the

genomic library, and in Sistrom's minimal medium (14) for the purification of M-RsrI. Bacteriophage vector λ L47.1 and its host strain *E. coli* WL95 *hsdR_k⁻ metB trpR tonA Sull SullI (P2) lysogen* were obtained from W. Brammar (15). *E. coli* C600 *hsdR_k⁺ M_k⁺ mcrA⁻ B⁺ supE44 thi-1 leuB6 lacY1 tonA21 λ^-* and its derivative carrying the *ecoRI* R-M genes on plasmid NTP14 (16) were used to enrich for phage containing the *rsrIM* gene. *E. coli* RRI *hsdS20R_b⁻ M_b⁻ mcrA⁺ B⁻ mrr⁻ supE44 leuB6 ara-14 proA2 lacY1* was used to propagate vector pBR328, both of which were gifts from J. Brooks. *E. coli* TB1 *hsdR17R_k⁻ M_k⁺ ara Δ (lac-proAB) rpsL (ϕ 80 lacZ Δ M15), obtained from M. Schuler, was used to propagate plasmid pUC18. Vector pTZ18U, a gift from V. Bankaitis, was propagated in *E. coli* NM522 *hsd5R⁻ M⁻ supE thi-1 Δ (lac-proAB) /F⁺proAB, lacI^qZ Δ M15* obtained from B. Kemper. Phagemid pTZ18U contains both the f1 and ColE1 origins of replication, a T7 RNA polymerase promoter in phase with the *lacZ'* and *Amp^r* genes, and the multiple cloning site from pUC18. Strains RRI, TB1, NM522 were grown in LB medium (Bacto-tryptone, 10 g; Bacto-yeast extract, 5 g; NaCl, 5 g per liter) with 50 μ g/ml ampicillin when transformed with plasmids. A plasmid containing 19 *EcoRI* sites, pRK112-8, was a gift from S.-H. Kim. Genetic nomenclature follows the recommendations of Szybalski *et al.* (17).*

b) Enzymes, Radioisotopes, and Chemicals

All restriction enzymes were purchased from BRL Life Technologies, Inc. or New England BioLabs, Inc. and were used with buffers made according to the manufacturer's specifications. T4 polynucleotide kinase was a gift from O. Uhlenbeck. Snake venom phosphodiesterase was purchased from Worthington Biomedical Corp., and T4 DNA ligase from BRL Life Technologies, Inc. Exonuclease III and S1 nuclease were supplied in a Nested Deletions Kit obtained from Pharmacia LKB Biotechnology, Inc. The Sequenase[®] kit was purchased from United States Biochemical Corp. S-adenosyl-L-[methyl-³H]methionine (11 Ci/mmol) was obtained from ICN Biomedicals, Inc., [γ -³²P]ATP (>6000 Ci/mmol) from Du Pont Co., and [α -³⁵S]thio-dATP (>1000 Ci/mmol) from Amersham Corp. Enzyme grade ammonium sulfate was from ICN Biomedicals, Inc. Homogeneous M-*EcoRI* and HPLC-purified AdoMet were gifts from P. Modrich.

c) DNA Manipulations

Total DNA from *R. sphaeroides* strain 630 was prepared essentially by the method of Marmur (18). To aid lysis, cells were pretreated with 0.1% (w/v) sodium dodecyl sulfate (SDS) prior to incubation with lysozyme. Bacteriophage λ DNA and plasmid DNA were prepared as described (19). Recombinant phage λ L47.1 were packaged *in vitro* by the method of Hohn and Murray (20). Ligation reactions during subcloning were carried out in agarose slices as described by Struhl (21). Cells were rendered competent and transformed according to the procedures in Maniatis *et al.* (19).

Colonies transformed with pBR328 clones were screened for sensitivity to tetracycline, and those transformed with pUC18 or pTZ18U were selected on 5-bromo-4-chloro-3-indolyl- β -D-galactoside, ampicillin plates. Double-strand templates for DNA sequencing were prepared by the method of G. Chan (University of Illinois; personal communication), which involved plasmid isolation by a modified rapid-boiling procedure (19), purification using GeneClean™ (Bio 101, Inc.), and denaturation by incubation with NaOH.

d) Gel Electrophoresis

The reduced and denatured molecular weight of the enzyme was determined by polyacrylamide-gel electrophoresis in the presence of SDS. Slab gels were prepared as described (22), and the proteins were visualized by silver-staining (23) or by staining with Coomassie blue. Protein standards were from Bio-Rad Laboratories. Isoelectric focusing in thin-layer polyacrylamide gels was carried out using Ampholine® 3.5-9.5 plates with a Multiphor electrophoresis apparatus from LKB. Samples of enzyme were desalted and concentrated using a Centricon™-10 microconcentrator from Amicon Corp. and drop-dialyzed into the Ampholine® solution. Electrophoresis was performed at 15 W constant power, at 15°C, for about 2.5 h with final voltage being 1500 V. Slices (1 cm) were excised from the gel, placed in 2 ml of water overnight, and the pH values determined. The intact portion of the gel containing the standards (Pharmacia LKB Biotechnology, Inc.) and a sample of enzyme was fixed with trichloroacetic acid and stained with Coomassie blue. Products of DNA restriction digests were separated by electrophoresis on 0.9% agarose gels in 40 mM Tris-acetate (pH 8.3), 1 mM EDTA. DNA was visualized by staining with ethidium bromide, and photographed under UV light. The products of the DNA-sequencing reactions were separated on 8% polyacrylamide gels containing urea. Electrophoresis was performed at 1800 V, and the gels were subsequently soaked in a 10% acetic acid, 12% methanol solution, dried, and subjected to autoradiography.

e) *RsrI* Methyltransferase Assays

Qualitative. Aliquots (1 μ l) of *M-RsrI* were incubated with 1 μ g bacteriophage λ DNA in a total of 10 μ l of 100 mM Tris-HCl (pH 8), 5 mM EDTA, 0.4 mg/ml bovine serum albumin (BSA), 5 mM dithiothreitol (DTT), 0.1 mM AdoMet at 30°C or 37°C for 30 min. *R-EcoRI* was diluted into REact® Buffer 3 (BRL) to a concentration of 125 U/ml, and an aliquot of 40 μ l was added to the methylation reaction mixture. The incubation was continued for another 30 min. at 37°C, and the reaction terminated by adding 10 μ l of gel-loading dye (19) and heating for 10 min. at 65°C prior to agarose gel electrophoresis. Reactions in which *M-RsrI* and *R-EcoRI* were omitted, and reactions in which only *M-RsrI* was omitted, were included as controls. Phage λ DNA that had been methylated by *M-RsrI* remained intact, and unmethylated λ DNA was completely

cleaved by *R·EcoRI* under these conditions. Column fractions containing *M·RsrI* activity were detected by their ability to confer resistance to *R·EcoRI* cleavage upon λ DNA *in vitro*.

Quantitative. Samples of *M·RsrI* preparations (volumes up to 5 μ l) were incubated with 20 nM (5 μ g) pRK112-8 in a total of 50 μ l of 100 mM Tris-HCl (pH 8), 5 mM EDTA, 0.4 mg/ml BSA, 2.5 mM DTT, 1.8 mM [3 H]AdoMet (11 Ci/mmol) for 15 min. at 30°C or 37°C. Samples of 45 μ l were spotted on DE81 filters (2.3 cm circles; Whatman BioSystems, Inc.), washed 3 times with 0.1 M NH_4HCO_3 to remove unincorporated AdoMet, 2 times with 95% ethanol, once with diethylether, and dried. The radioactivity was quantified by liquid scintillation counting. A unit (U) of *M·RsrI* activity is defined as 1 pmol of methyl groups transferred per min. and corresponds to approximately 13,500 cpm (above a background of less than 0.1%). The assay is linear over 20 min. with up to 30 nM (0.4 U) enzyme under these reaction conditions.

f) Reaction Optimization

To optimize the *M·RsrI* reaction conditions, the quantitative assay was used with various pH values, temperatures of incubation, and NaCl concentrations. The optimal pH was found by preparing Tris-HCl buffers at pH values ranging from 6.0 to 9.5 measured at 30°C. The optimal temperature was determined in the range of 4°C to 50°C using buffers that had been adjusted to pH 7.5 at each temperature. Ionic strength dependence was tested by adding up to 160 mM NaCl to the reaction mixtures at the optimal pH and temperature (the reaction mixtures contained 1.5 mM KCl from the enzyme aliquot). The effects of N-ethylmaleimide (NEM) were determined by adding to the reaction mixtures a 50% ethanol solution of NEM to final concentrations of 5 and 10 mM. Controls with equal additions of only 50% ethanol did not inhibit the enzyme. The effects of MgSO_4 or MgCl_2 were tested at concentrations up to 100 mM.

g) Identification of Methylated Nucleotide

The position of methylation by *M·RsrI* was determined essentially as described (24, 25). The oligodeoxyribonucleotide d(pCTGAATTCAG), 1 μ M, was annealed and incubated in 50 μ l of the quantitative assay reaction buffer with 3.6 mM [3 H]AdoMet (11 Ci/mmol) and 1.5 U of *M·RsrI* for 60 min. at room temperature. The incompletely-labeled decanucleotide was purified by centrifugation on a G-50 column, and phosphorylated using T4 polynucleotide kinase and [γ - 32 P]ATP (75 Ci/mmol). The doubly-labeled decanucleotide was partially digested with snake venom phosphodiesterase (VPD). Aliquots, 5 μ l, were removed at intervals up to 180 min. By combining appropriate samples, a mixture containing all the partial digestion fragments of the decanucleotide was obtained. The fragments were separated by homochromatography and located by autoradiography. The areas containing ^{32}P

were excised and the ^3H and ^{32}P in each fragment determined by liquid scintillation spectrometry.

h) Sequence Comparisons

The software package DNASTAR (Madison, Wisconsin) was used for analyses of the DNA and amino acid sequences. The amino acid sequence of M-RsrI, derived from the nucleotide sequence, was compared to other methyltransferases. The following DNA adenine MTases were examined: CviBIII (26), *dam* (27), DpnA (28), DpnII (29), EcoP1 (30), EcoP15 (30), EcoRI (31, 32), EcoRV (33), HhaII (34), HinfI (35), PaeR7 (36), PstI (37), T4dam (38), and TaqI (39). Two N^4 -cytosine MTases, BamHI (J. Brooks and D. Landry, personal communications) and PvuII (11) were also compared to M-RsrI. Pairwise comparisons between the MTases were carried out using the DNASTAR program Aalign. Global and partial alignments were conducted with the gap penalties of 2 to 4, and deletion penalties of 6 to 12, using the PAM250 amino acid similarity matrix (40). Further refinements were carried out manually. The nucleotide sequences of M-RsrI and M-EcoRI were compared using the program Align with k-tuple values from 3 to 6, a range of 20, and gap penalties from 3 to 15. The program DotPlot was used to generate dot-matrix comparisons with a window of 20 amino acids with 35% residues matching. The NBRF-PIR database, release 20, was used for homology searches. A. Bhagwat (Wayne State University) and G. Wilson (New England BioLabs) kindly compared the M-RsrI to the following C5 MTases: BepI, Bsp/BsuRI, BsuFI, dcm, DdeI, EcoRII, ϕ 3T, HaeIII, HhaI, HpaII, MspI, NgoPII, p11s, SinI, SPR, SspMQI.

i) Buffers

All buffer pH values were determined at the temperatures and concentrations at which they were used. Buffers A and B were as previously described (3), except that 20 mM β -mercaptoethanol was used. Buffer C: 20 mM potassium phosphate (KPi) pH 7.5, 0.1 M NaCl, 10% glycerol, 20 mM β -mercaptoethanol. Buffer D: same as buffer C, but with NaCl omitted.

j) Other Methods

All purification steps, except HPLC, were carried out at 0-4°C. HPLC was at room temperature using a Beckman model 420 system. The KB Hydroxylapatite HPLC column was purchased from Regis Chemical Co., and the WCX HPLC (Zorbax Bio-series) from Du Pont Co. All centrifugations were at 1200xg for 30 min. Size-exclusion chromatography was performed on a TSK G3000SW column purchased from The Anspec Company, Inc. Proteins were eluted with Buffer C in which KCl was substituted for NaCl, and detected by their absorption at 280 nm. Gel-filtration protein standards were from Bio-Rad Laboratories and United States Biochemical Corp. Prior to N-terminal amino acid sequence analysis, proteins were electroblotted onto a GF-F

glass filter (Whatman BioSystems Inc.) or a polyvinylidene difluoride (PVDF) membrane (Immobilon™; Millipore Corp.). N-terminal amino acid sequence determinations and oligonucleotide syntheses were performed using an Applied Biosystems 470A Protein Sequenator and 380A DNA synthesizer, respectively (Biotechnology Center, University of Illinois). The oligonucleotides were purified on NAP™-5 (Pharmacia LKB Biotechnology, Inc.) columns. Protein concentrations were determined by the method of Bradford (41) using a kit from Bio-Rad Laboratories. Lysozyme was used as standard for protein assays and its concentration was determined spectrophotometrically using $E^{0.1\%}_{281.5\text{ nm}} = 2.64$ in 0.1 M KCl (42).

RESULTS AND DISCUSSION

a) Purification of the *RsrI* Methyltransferase

Since the *RsrI* methyltransferase and endonuclease partially copurify, the crude extract, polyethyleneimine (polymin P) and ammonium sulfate precipitations, DEAE and phosphocellulose chromatography steps were carried out as previously described (3). *M-RsrI* elutes on the phosphocellulose column at 0.3 M KCl in buffer B and is separated from the *R-RsrI* activity. *R-RsrI* often coelutes with *M-RsrI* at this step.

HPLC Hydroxylapatite Chromatography. Solid $(\text{NH}_4)_2\text{SO}_4$ was added to the phosphocellulose pool to 80% saturation (516 mg/ml) over 30 min. The mixture was stirred for an additional 1 h and centrifuged. The pellet was dissolved in 1 ml of buffer C and dialyzed against 1 l of the same buffer. The sample was applied at a flow rate of 1 ml/min to a hydroxylapatite column (0.6 cm ID x 10 cm) previously equilibrated in buffer C. The column was washed for 5 min., and eluted with a 40 ml linear gradient of 20 to 300 mM KPi (pH 7.5) in buffer C. Fractions containing *M-RsrI* activity, which eluted at approximately 25 mM KPi , were pooled.

HPLC Weak Cation Exchange (WCX) Chromatography. The hydroxylapatite pool was concentrated and desalted using a Centriprep™-10 concentrator. The sample (0.4 ml) was applied at a flow rate of 2 ml/min to a WCX column (0.62 cm ID x 8 cm) previously equilibrated in buffer D. The column was washed for 5 min., and eluted with a 40 ml linear gradient of 0 to 1 M KCl in buffer D. *M-RsrI* eluted at 0.3 M KCl. Fractions, which were >95% pure as judged by SDS-polyacrylamide gel electrophoresis followed by silver staining, were pooled, concentrated, and made 50% (v/v) in glycerol. The purified enzyme (0.5 ml) was stored at -20°C and retained 50% activity after one year.

Purification Summary. The purification uses two conventional and two HPLC chromatographic steps (Table I). Several resins often used for the purifications of other MTases were tested using conventional liquid chromatography. These included

Table I. Purification of *RsrI* Methyltransferase^a

Fraction	Total Protein (mg)	Total Activity (Units) ^c	Specific Activity (U/mg)	Yield (%)	Fold Purification
Crude ^b	6,910	--	--	--	--
Polymin P ^b	4,620	--	--	--	--
Ammonium Sulfate	5,780	22,000	4	(100)	--
DEAE-Cellulose	1,160	15,000	13	68	3
Phosphocellulose	11	3,100	280	14	80
Hydroxylapatite	0.312	538	1,720	2.4	430
Weak Cation Exchange	0.050	114	2,280	0.5	570

^a100 g of *R. sphaeroides* cell paste was used.

^bReliable measurements of M-*RsrI* activity cannot be obtained in these fractions.

^cAll reported measurements were done on samples after dialysis and concentration using the quantitative assay (Materials and Methods).

phenyl sepharose, S-adenosyl-homocysteine sepharose, heparin sepharose, and Affi-Gel® Blue. We found that only hydroxylapatite HPLC and WCX HPLC columns gave the additional purification needed to achieve homogeneity. The purification yields from 50 to 100 µg of pure M-*RsrI* per 100 g of cell paste. The biggest loss of activity (70%) occurs during the phosphocellulose chromatography. We have recently observed that adding 50 mM KCl to the phosphocellulose buffer halves this loss. Each of the remaining three columns gives a recovery of at least 50%, but substantial losses are

Figure 1. Electrophoresis of purified *RsrI* and *EcoRI* MTases under denaturing conditions. M-*RsrI* (2 µg; lane B), M-*EcoRI* (0.7 µg; lane C), and protein markers (lane A) were subjected to electrophoresis on 15% polyacrylamide gel containing SDS. The proteins were visualized by silver staining (Materials and Methods).

sustained during concentration and dialysis steps. A doublet was sometimes observed upon SDS-polyacrylamide gel electrophoresis of the purified *M-RsrI*. We attribute this apparent heterogeneity to incomplete reduction of a disulfide bond prior to electrophoresis. When higher concentrations of β -mercaptoethanol were added to the sample before electrophoresis the doublet band corresponding to the higher apparent molecular weight species decreased in intensity.

b) Physical Properties

Molecular Weight. The reduced and denatured molecular weight of *M-RsrI* was determined by electrophoresis on 12.5% and 15% SDS-polyacrylamide slab gels. A single band of $M_r = 33,600 \pm 2,000$ was observed upon silver staining. The *RsrI* and *EcoRI* MTases migrate differently on SDS gels (Figure 1). The apparent molecular weight of *M-EcoRI* in our hands is $38,000 \pm 1,000$ Da, a value in agreement with the reported 39,000 Da (43).

The molecular weight of *M-RsrI* under nondenaturing conditions was examined using high performance, size-exclusion chromatography (Materials and Methods). A *M-RsrI* sample, approximately 10 μ g of the WCX pool in 50% glycerol, was applied to a TSK G3000SW column (0.75 cm ID x 30 cm). The activity eluted in a broad peak in the volumes expected for a globular protein with molecular weights from 71,000 Da to approximately 30,000 Da. *M-RsrI* is partially dimerized under these conditions, in contrast to *M-EcoRI* and most other MTases which are monomers (10). We will investigate the native molecular weight of *M-RsrI* under conditions more nearly approximating those of the assay.

N-terminal Amino Acid Sequence. Two independent determinations of the sequence of the amino terminus of *M-RsrI*, which had been electroblotted onto a solid support, were performed. One sample of pure *M-RsrI*, 7 μ g (200 pmol), was run on a 10% SDS-polyacrylamide gel. The protein was electroblotted onto a PVDF membrane (44), stained with Coomassie blue, and the band excised. A second sample, 50 μ g, of partially purified *M-RsrI* (hydroxylapatite fraction) was treated as above, except that separated proteins were blotted onto an activated glass filter (45) and the band with the mobility of *M-RsrI* was excised. In the first determination 6 of 8 amino acids were identified, and in the second 11 of 16. The sequences agreed with each other in the positions that were determined. The identified N-terminal amino acids of *M-RsrI* are double-underlined in Figure 3. The N-terminal methionine is absent in the purified *M-RsrI*, as it is in the case with *R-RsrI* (3, 4) and both *EcoRI* enzymes (31, 32).

Isoelectric Point. Samples of purified *M-RsrI* and *M-EcoRI*, 6 μ g each, were focused on a thin-layer polyacrylamide gel. Following staining, one band was observed for each enzyme. The pI value of *M-RsrI* was 6.8, and of *M-EcoRI* 9.0, a value approximating the reported pI of 8.7 (43).

Table II. Identification of the Base Methylated by *M·RsrI*^a

Oligodeoxyribonucleotide	Radioactivity (cpm)		Ratio ³ H/ ³² P
	³ H	³² P	
d(pCTGAATTCAG)	772	4350	0.18
d(pCTGAATTCA)	1296	4870	0.26
d(pCTGAATTC)	1125	4248	0.26
d(pCTGAATT)	977	3358	0.29
d(pCTGAAT)	570	2249	0.25
d(pCTGAA)	1911	2600	0.73
d(pCTGA)	180	3080	0.06
d(pCTG)	179	4511	0.04

^aSee Materials and Methods. ³H is in *N*⁶mA and ³²P is in the 5'-phosphomonoester terminus of the oligodeoxyribonucleotide.

c) Catalytic Properties

Identification of the Methylated Base. The central adenine residue of duplex d(pCTGAATTCAG) is methylated by *M·RsrI*, as indicated by the presence of ³H in only the oligonucleotides containing this base (Table II). The increase in the ³H/³²P ratio in d(CTGA³mA) has been observed before (25) and is due to the faster rate of cleavage by VPD of the d(CTGA³AA) contained in this incompletely methylated product. In another experiment total hydrolysis of d(GTAGAATTCTAC), that had been incubated with [³H]AdoMet and *M·RsrI*, followed by HPLC analysis showed label in *N*⁶mA only (Webb and Gumpert, unpublished observations). Thus, *M·RsrI* catalyzes the transfer of methyl groups from AdoMet to the same position on central adenine residue in the recognition sequence as does *M·EcoRI* (5).

Reaction Optimization. *M·RsrI* is most active at pH 7.5, and retains 80% of its activity at pH 7.1 and 8.3. The enzyme is maximally active at 37°C, and 80% as active at 33°C and at 40°C. *M·RsrI* is irreversibly inactivated at 65°C. Addition of NaCl to the reaction mixture inhibits the activity. At 20 mM NaCl 80% of the maximal activity was observed, and only 5% of the activity remained at 160 mM NaCl. The *RsrI* methyltransferase and endonuclease activities show similar responses to pH and NaCl, however they differ in their temperature optima. Whereas *R·RsrI* is maximally active at 23°C and inactive at 37°C (3, 4) *M·RsrI* is maximally active at 37°C with only 30% activity at 23°C. Both enzymes retain approximately 60% of their maximal activities at 30°C, the optimal growth temperature for *R. sphaeroides*. *M·RsrI* loses 95% of its activity upon incubation with 5 mM NEM showing that at least one thiol group is required for activity. NEM has the same effect on *M·EcoRI* (43) and *R·RsrI*

Figure 2. *R. sphaeroides* clones and sequenced regions. The recombinant vectors are labeled on the left. All inserts (thin bars) are drawn to scale, whereas the cloning vectors (shaded bars) are not. The scale on the bottom of the figure refers only to the sequenced region. The sequence obtained from an individual sequencing reaction is indicated by an arrow. The plasmid pTZ-RsrM3.2 contains only 235 3'-end base pairs of *rsrIR* and thus does not produce active endonuclease.

(3), but not on R-*EcoRI* (43). Whereas Mg^{2+} is required for the endonuclease activity, 20 mM $MgCl_2$ or $MgSO_4$ inactivates M-*RsrI*.

d) Construction and Screening of *R. sphaeroides* DNA Library

Several studies have demonstrated that it is possible to isolate recombinant clones expressing methyltransferase genes by selecting DNA molecules that are resistant to the action of a cognate endonuclease *in vitro* (46). Since there is an *E. coli* enzyme that is a functional analogue of R-*RsrI*, we were able to select clones expressing the *rsrIM* gene on the basis of their immunity to R-*EcoRI* restriction *in vivo*. Total DNA from *R. sphaeroides* strain 630 was partially digested with *Sau3AI* and fragments between approximately 7 and 18 kb were ligated to λ L47.1 DNA previously cut with *BamHI*. The 2×10^4 phage obtained constituted the primary genomic library. The *EcoRI* restricting strain *E. coli* C600(NTP14) was infected with the recombinant phage at a multiplicity of one to enrich for phage carrying and expressing the *rsrIM* gene. It was assumed that any modification of *EcoRI* sites within the λ phage genome by M-*RsrI* in addition to that by the M-*EcoRI* encoded on the NPT14 plasmid would

enhance protection from *in vivo* restriction. The recombinant phage plated with a 10^4 -fold reduced titer on this strain relative to *E. coli* C600. Surviving phage were harvested and propagated in *E. coli* C600 before reinfecting C600(NTP14). Following this second cycle of enrichment, two phage, designated λ DOC 1.1 and λ DOC 1.2 plated with equal efficiencies on host strains C600 and C600(NTP14). In contrast to their *in vivo* properties, however, DNA purified from either phage could be partially digested with R-EcoRI. Discrepancies between the *in vivo* and *in vitro* properties of cloned R-M enzymes have been noted in other systems (for example, ref. 47). The λ DOC 1.1 phage was taken for further characterization, and upon subcloning the DNA became completely resistant to cleavage by R-EcoRI *in vitro* (see below).

e) Construction of the *RsrI* Methyltransferase Clones

Restriction digests of phage λ DOC 1.1 showed it contained an 11 kb *R. sphaeroides* insert (Figure 2). DNA from λ DOC 1.1 was cut with *HindIII* and *Sall*, and the resulting 7 kb fragment, originating almost entirely from within the insert, was ligated to pBR328 that had been treated with the same restriction enzymes. A number of *AvaI* sites were identified in the recombinant plasmid pBR-*RsrM*7.0. Although *AvaI* cleaves degenerate DNA sequences, a 3.2 kb fragment (Figure 2) was successfully ligated to pUC18 previously cut with *AvaI*. In preparation for DNA sequencing, the 3.2 kb fragment carrying the *rsrIM* gene was ligated to pTZ18U using the *SstI* and *XbaI* sites within the polylinker to yield pTZ-*RsrM*3.2. The isolated plasmids, pBR-*RsrM*7.0, pUC-*RsrM*3.2, and pTZ-*RsrM*3.2, were totally resistant to cleavage by R-EcoRI, indicating that the *rsrIM* gene was expressed in *E. coli*.

f) Nucleotide Sequence of the *rsrIM* Gene

Sequence Determination. The nucleotide sequence of the *rsrIM* gene was determined by the dideoxynucleotide chain-termination method of Sanger *et al.* (48). A set of double-strand, nested deletions was generated in pTZ-*RsrM*3.2 using exonuclease III and S1 nuclease. The longest deletion still carrying the intact *rsrIM* gene and nine successive deletions, differing from one another by approximately 150 bp, provided the DNA templates for sequencing one strand using a universal primer. The opposite strand of the *rsrIM* gene was sequenced by the same methodology using six oligodeoxyribonucleotide primers complementary to the strand initially sequenced. The sequenced region and some of its restriction sites are shown in Figure 2. The identity of nearly every nucleotide within the *rsrIM* gene was determined in at least three independent reactions, and up to six times in regions with ambiguities. A single open reading frame of sufficient length to specify a protein of the size of M-*RsrI* was found. It consists of 960 bp (terminator codon included) and encodes a 319 residue polypeptide with a predicted molecular weight of 35,660 Da (including the N-terminal methionine). This is in agreement with the

CATATGGACTTTTCGCCGAAGAAGCGTGCATGGTCCACAGGTCAAGCCAAAGCACAGAACAAGGAGTAGCACCACGA
 1 ATGSCAAACCGATCTCACCACAATGCGGGCCACCGAGCGATGAACGCTCTCCGCAAGTCAGGCCAAAAGC
 1 M A N N R S H H N A G H R A M N A L R K S G Q K H
 71 ATTCGTCGAGTCTCAGTTGGGATCTTCGGAGATCGGAACACACGACATGTGTATGACGTTTTCGGATTG
 25 S S E S Q L G S S E I G T T R H V Y D V C D C
 141 CCTCGACACCTTGGCGAAGCTGCCGACGACTCCGTCCAGTTGATCATTGTGTATCCTCCATACACATC
 48 L D T L A K L P D D S V Q L I I C D P P Y N I
 211 ATGCTGGCGAAGTGGGATGATCAGTGGACTACATCGGCTGGGCGAAGCGATGGCTAGCCGAAGCCGAGC
 71 M L A D W D D H M D Y I G W A K R W L A E A E R
 281 GTGTTCTATCCCAACGGGACGATAGCGATTTTCGGCGGCTACAATACCAAGGGGAGGCTGGTTTCAGG
 95 V L S P T G S I A I F G G L Q Y Q G E A C C S G
 351 TGATCTGATCTCAATCATTTTCGCACATGAGACAAAACGCAAAATGCTGCTCGCAACCTTATCATCTGG
 118 D L I S I S I S H M R Q N S K M L L A N L I I W
 421 AACTATCCGAACGGCATGAGTGCAGCGCTTCTTTGCAATAGGCACGAGGAAATCGCGTGGTTTCGCCA
 141 N Y P N G M S A Q R F F A N R H E E I A W F A K
 491 AGACCAAGAAGTATTTTCTCGATCTTGACGCTGTGCGGGAACCATATGACGAGGAACCAAAAGCCGCTTA
 165 T K K Y F F D L D A V R E P Y D E E T K A A Y
 561 CATGAAGGACAAGCGGTTGAACCTGAGTCGGTTGAAAAGGGACGCAACCCAAACACGTTGTCGGGATG
 188 M K D K R L N P E S V E K G R N P T N V W R M
 631 TCTCGGCTTAACGGCAATTCCTTGAACGTGTGCGCCACCCACGCAAAAACCGCCGAGTGATTGAGA
 211 S R L N G N S L E R V G H P T Q K P A A V I E R
 701 GACTGGTTTCGGCGTGTTCGCATCCGGTTCAACGGTCTCTCGATTTTTCGTTGGAAGTGGGGTAACGGC
 235 L V R A L S H P G S T V L D F F A G S G V T A
 771 GCGCGTTGCAATTCAAGAGGGACGGAACAGCATTTCGACTGACGCGCGCTGTATTCAAGAATATTAC
 258 R V A I Q E G R N S I C T D A A P V F K E Y Y
 841 CAGAAGCAGTAACTTCTTCAAGATGATGACTGATGACAAAGCGCATATGAATTTGTTGAAG
 281 Q K Q L T F L Q D D G L I D K A R S Y E I V E G
 911 GGGCTCGAAGCTTGTGTCAGCATTGCAGCGGGGAGATGTTGCTTCAATATGGCTGTTGCTGTCTGCGC
 305 A A N F G A A L Q R G D V A S
 980 ATAG

Figure 3. The nucleotide and deduced amino acid sequences of *M-RsrI*. The nucleotide sequence starts with the *NdeI* site (see Figure 2). Dotted-underlined nucleotides represent putative promoters (the consensus sequence is GGCN₇₋₁₁CGCC, see text). Underlined nucleotides represent a possible ribosome-binding site. Double-underlined amino acids were identified by sequencing the N-terminus of purified *RsrI* MTase, the identities of single-underlined amino acids were not determined. The start and stop codons of *rsrIM* are italicized. The DPPY sequence is in bold letters. Nucleotides and amino acids are numbered on the left.

observed value of $33,600 \pm 2,000$ Da for purified *M-RsrI*. The calculated isoelectric point of the predicted polypeptide is 6.9, approximating the experimentally determined value of 6.8. The pI calculated from the *M-EcoRI* predicted amino acid sequence is 8.4 also in agreement with the reported value of 8.7 (43). The N-terminal sequence of the purified MTase agrees with the deduced amino acid sequence (Figure 3). The lengths of several restriction fragments predicted by the *rsrIM* sequence were confirmed by electrophoresis (data not shown). *M-RsrI* activity was quantified (Materials and Methods) in crude extracts of *E. coli* bearing the 3.2 kb *R. sphaeroides* fragment on the pUC18 plasmid and had 50% of the specific activity of *M-RsrI* found in *R. sphaeroides* cells. We never observed more than one peak of MTase activity that was specific for duplex d(GAATTC) during the chromatographic steps of the purification, leading us to believe that there is but one such activity in these cells. The identical nucleotide sequence of *rsrIM* was obtained from an independently isolated clone by Stephenson and Greene (*Nucl. Acids Res.* (1989) 17, this issue).

Relative Locations of *rsrIM* and *rsrIR* Genes. In common with other type II R-M genes the *rsrIM* and *rsrIR* genes are in proximity (9). The C-terminus of the *RsrI* endonuclease is located 83 bases upstream from the initiation codon of the methyltransferase. The two enzymes are encoded on the same strand of DNA and their reading frames are shifted by -1 nucleotide with respect to one another. The relative positions and orientations of the *ecoRI* R-M genes, as well as the -1 shifted reading frame, are identical to the *rsrI* genes. However, the *ecoRI* genes are separated by only 29 base pairs (31, 32). Another difference between the *E. coli* and *R. sphaeroides* restriction-modification systems is that the *ecoRI* genes are plasmid-borne (49, 50), whereas the *rsrI* genes are located on the chromosome. Degenerate heptadecadeoxyribonucleotides complementary to the 5'-end of the *rsrIR* gene were used as probes of Southern blots of total *R. sphaeroides* 630 chromosomal DNA and plasmid DNA. No hybridization was observed with the two plasmids present in these cells (51), whereas a distinct band was observed with the chromosomal DNA (data not shown). Stephenson and Greene have also determined that the *rsrI* genes are chromosomally encoded (*Nucl. Acids Res.* (1989) 17, this issue).

Regulatory Sequences. A putative *R. sphaeroides* promoter consensus sequence, GGCN₇₋₁₁CGCC (R. Gould and S. Kaplan, personal communication) is located 68 nucleotides upstream from the start codon of the *rsrIR* gene. In addition, less perfect matches to this sequence are 45 (GGTN₉AGCC) and 71 (GGAN₆CGCG) base pairs upstream of the initiator codon of the *rsrIM* gene in the intercistronic region (Figure 3). A potential ribosome binding site (52), AGAAGG, that is complementary to five of six nucleotides near the 3' end of *R. sphaeroides* 16S rRNA (S. Kaplan, personal communication) is located 17 base pairs upstream of the *rsrIM* gene (Figure 3). Although it is possible that the imperfect consensus promoters occasionally serve to initiate transcription of only the *rsrIM* gene, it seems more likely that both the endonuclease and methyltransferase genes would be transcribed from the promoter with the better consensus sequence upstream of the *rsrIR* gene in *R. sphaeroides* cells, and that the Shine-Dalgarno sequence preceding the *rsrIM* gene is used to reinitiate translation of the MTase. It has been suggested that the *ecoRI* genes may be coordinately regulated (32), and existence of a separate promoter for the methylase gene has also been reported (53). We have observed M-*RsrI* activity in *E. coli* cells bearing *R. sphaeroides* fragments containing the MTase gene but lacking the intact endonuclease gene. Since there are no significant similarities to *E. coli* consensus promoter sequences upstream of the *rsrIM* gene in the cloned fragment, it is probably transcribed from a vector promoter.

We find no putative *R. sphaeroides* transcription-termination structures (54) within the 150 bases distal to *rsrIM*. Another gene may be located immediately

Table III. Codon Usage in *rsrIM* and *ecoRIM*

Amino Acid	Codon	<i>rsrIM</i> ^a	<i>ecoRIM</i> ^a	Amino Acid	Codon	<i>rsrIM</i>	<i>ecoRIM</i>	Amino Acid	Codon	<i>rsrIM</i>	<i>ecoRIM</i>
Ala	GCA	7	3	Gly	GGA	7	3	Pro	CCA*	5	5
	GCC	6	1		GGC	9	3		CCC	2	0
	GCG	13	2		GGG	3	5		CCG	2	0
	GCU	7	4		GGU	4	6		CCU	3	6
		10.3 ^b	3.1 ^b			7.2	5.2			3.8	3.4
Arg	AGA*	2	10	His	CAC	7	2	Ser	AGC	3	4
	AGG*	1	1		CAU	3	5		AGU*	2	4
	CGA	5	1	Ile	AUA*	1	7		UCA*	6	3
	CGC	5	1		AUC	10	2		UCC	3	0
	CGG	6	0		AUU	7	14		UCG	5	4
	CGU	2	0			5.6	7.1		UCU	5	8
		6.6	4.0							7.5	7.1
Asn	AAC	13	9	Leu	CUA*	4	5	Thr	ACA*	2	5
	AAU	3	19		CUC	4	1		ACC	4	0
		5.0	8.6		CUG	6	3		ACG	4	1
Asp	GAC	11	2		CUU	5	5		ACU*	2	3
	GAU	10	21		UUA*	0	9	Trp	UGG	6	2
		6.6	7.1		UUG*	6	5			1.9	0.6
						7.8	8.6				
Cys	UGC	3	4	Lys	AAA	5	22	Tyr	UAC	5	5
	UGU*	1	3		AAG	11	13		UAU	6	16
		1.2	2.1			5.0	10.7			3.4	6.4
Gln	CAA	7	4	Met	AUG	9	2	Val	GUA*	2	3
	CAG	6	1			2.8	0.6		GUC	3	1
		4.1	1.5	Phe	UUC	7	7		GUG	4	2
Glu	GAA	9	10		UUU*	4	16		GUU	7	13
	GAG	9	10			3.4	7.1			5.0	5.8
		5.6	6.1								

^aThe number indicates the times the codon is used in the 319 amino acids of *M-RsrI* and the 326 amino acids of *M-EcoRI*.

^bPercent amino acid composition.

*Codons used 5% or less in the sequenced genes of *R. sphaeroides* (R. Gould and S. Kaplan, personal communication).

downstream, because an imperfect *R. sphaeroides* promoter consensus sequence (GGC_NTTGCC) is found one base pair beyond the MTase stop codon (Figure 3), and is followed by a potential Shine-Dalgarno sequence and a start codon (data not shown).

g) Comparisons of the *RsrI* and *EcoRI* Methyltransferase Sequences

Nucleotide Sequences. Although they carry out identical reactions on the same DNA sequence, neither the nucleotide nor the deduced amino acid sequences of *M-RsrI* and *M-EcoRI* exhibit significant similarities. The nucleotide sequences of *rsrIM* and *ecoRIM* differ in length and total base composition. The *E. coli* gene is encoded by 978 base pairs and the *R. sphaeroides* gene by 960. Optimal alignment of the two sequences indicates 24% nucleotide identity. The *rsrIM* is 52% G + C compared to the approximately 68% G + C composition of the *R. sphaeroides* genomic DNA (55), and the *ecoRIM* gene is 32% G + C (31, 32) compared to an *E. coli* genome of 50% G + C

(56). Thus, *rsrIM*, like *ecoRIM* and each of their corresponding endonuclease genes (12), is A + T rich with respect to the genome of the bacterium in which it resides.

Codon Usages. The codon biases of *rsrIM* and *ecoRIM* are also markedly different (Table III). The most extreme dissimilarities are illustrated by six codons used to synthesize M-*RsrI*, CGU, CGG, UCC, ACC, CCG, and CCC, which are unused in the translation of M-*EcoRI*. Conversely, UUA encodes nine leucines in the *E. coli* MTase and none in the *R. sphaeroides* MTase. Comparing the overall biases in the third positions of the codons, those of *rsrIM* are 58% G + C and those of *ecoRIM* are 28% G + C. A similar comparison between *rsrIR* and *ecoRIR* showed a larger difference with the wobble-position compositions of 62% and 22%, respectively (12). Since the endonucleases are highly homologous (12), it is clear that third-position differences are not necessarily related to amino acid sequence homologies. These biases do reflect the relative G + C richness of the genomes of the two organisms.

It is interesting to note that there are also differences in the codon usages between *rsrIM* and *rsrIR* (12) themselves. For example, CCA occurs five times in *rsrIM* and once in *rsrIR*, and UUG is used six and three times in the respective genes. ACA and GUA occur twice in the methyltransferase and are unused in the endonuclease gene. All of these codons are rarely used in *R. sphaeroides* genes (Table III), and the occurrence of rare codons in a gene has been correlated with low-level expression of the encoded protein in some organisms (57). Given the possible coordinated production of M-*RsrI* and R-*RsrI* (see above), the meaning of the rare-codon usage differences is unclear.

Amino Acid Sequences. As predicted by the differences in the molecular weights and isoelectric points of M-*RsrI* and M-*EcoRI*, the deduced amino acid sequences of these two proteins differ in length and composition (Table III). An optimal alignment of the deduced amino acid sequences of these MTases shows only 15.6% identity. M-*RsrI* contains the tetrapeptide DPPY, a sequence that is found in adenine MTases (7, 9, 10), whereas the M-*EcoRI* contains a variant of the sequence, NPPF (31, 32). These tetrapeptide motifs are located in different regions of the corresponding genes (Figure 6). To the extent that the similarity between the two MTases exists, it is evident in a sequence designated region III by Hattman *et al.* (58). This region, postulated to be the DNA-recognition domain in certain adenine MTases (59), shows 23% amino acid identity and extends from residue 162 to 195 of M-*RsrI* and from residue 297 to 326 of M-*EcoRI*. Another short area of homology between M-*RsrI* and M-*EcoRI*, in which seven of seventeen amino acids are identical, is found in the N-terminal regions of these proteins (Figure 4A). The longest identical stretch of amino acids in the *RsrI* and *EcoRI* MTases is the tetrapeptide SSES found at residue 124 of M-*EcoRI* and residue 25 of M-*RsrI*. A subsequence, SSE, is also found in different

Figure 4. Dot-matrix comparisons between the *RsrI* and *EcoRI* MTases (A) and ENases (B). The DNASTAR program DotPlot was used (Materials and Methods). The amino acid positions are shown as tenths of their actual values.

regions of both methyltransferases. Although there are some conserved amino acid residues in *M·RsrI* and *M·EcoRI*, they are located so differently in the respective proteins that their significance is unclear.

The lack of homology between the two methyltransferases is in striking contrast to the *RsrI* and *EcoRI* endonucleases which show 50% amino acid identity and 52% nucleotide identity (12). A dot-plot comparison of the two endonucleases and methyltransferases vividly illustrates this difference (Figure 4). The extensive sequence homology presumably reflects a common ancestor for the two endonuclease genes (12). Conversely, the lack of sequence similarity between the MTase genes suggests evolution from different or more distantly related progenitor genes. If these two MTases are themselves derived from different ancestors, the joining of related endonucleases and unrelated methyltransferases to form their respective R-M systems must have occurred as separate events. The apparently independent assembly of these R-M systems with one homologous and one nonhomologous element only adds to the puzzle of how the genes of methyltransferases and endonucleases of a particular specificity become associated with one another in an organism (8).

Alternatively, these R-M systems may have had a common ancestor, and their MTases diverged while changes in the endonucleases were selected against. The structure of the *R·EcoRI*-DNA co-crystal reveals a complex set of interactions between the two components with an intimately interdigitated DNA recognition and catalytic center (60). Mutational studies of *R·EcoRI* indicate that amino acid changes in regions both near and far from the protein-DNA interface can disrupt function

Figure 5. Dot-matrix comparisons between *RsrI* and seven other MTases. The comparisons are arranged in the decreasing order of homology as judged by the optimal scores (left to right and with *EcoP1* preceding *EcoP15*; see text). The DNASTAR program DotPlot was used (Materials and Methods). The amino acid positions are shown as hundredths of their actual values.

(61, 62). These findings suggest that the endonuclease might be relatively refractory to amino acid substitutions without loss of function.

On the other hand, studies of the C5-cytosine methyltransferases indicate that a subset of these enzymes may be more malleable. The target recognition domains and catalytic cores (8, 63) have been interchanged between different, but closely related *B. subtilis* bacteriophage MTases, and the enzymes retain activity (64-66). To the extent that *M·RsrI* and *M·EcoRI* may be similar to these cytosine MTases in having a modular structure, they might have accumulated stochastic changes because they were less subject to a tight selection for a conserved structure than were the corresponding endonucleases. Arguing against this hypothesis and suggesting that the methyltransferases ought to be more conserved than the endonucleases is the finding in several R-M systems, including the *EcoRI* combination, that the R^+M^- phenotype is lethal whereas R^-M^+ cells are viable. The different evolutionary histories of the two R-M systems may be reflected in their intracellular locations where the *R. sphaeroides* genes are on the chromosome and the *E. coli* genes are on a plasmid (see above).

h) Comparison of the *RsrI* MTase to other Methyltransferases

The type II MTases can generate one of the following modified bases: N^6 -methyladenine (N^6 mA), N^4 -methylcytosine (N^4 mC), and 5-methylcytosine (5mC). *M·RsrI* is most similar to two closely related (67) type II adenine methyltransferases *HinfI* and *DpnA*. Interestingly, *M·RsrI* also exhibits significant sequence similarities to two type II N^4 mC methyltransferases, *PvuII* and *BamHI*. Blumenthal *et al.* (11) have demonstrated that *M·PvuII* has substantial similarity to two other N^6 mA MTases (*M·EcoRI* and *M·PstI*), and suggested that the N^4 mC and N^6 mA MTases may use the same mechanism of methylation. In addition, *BamHI* is also homologous to some other N^6 mA methyltransferases (J. Brooks and D. Landry personal communication, 9). Our findings extend the observed interrelationships between enzymes that methylate an exocyclic amino group of a nucleobase and lend further support to the suggestion of a common mechanism. The R-M systems are of three types (I, II, and III) classified according to their subunit composition and cofactor requirements. *M·RsrI* shows sequence similarities to two allelic type III N^6 mA MTases, *EcoP1* and *EcoP15*. The sequence similarities between type II and type I methyltransferases reported previously (68), and the similarities between type II and type III reported here suggest that these three mechanistically and structurally distinct classes may all share parts of common primordial ancestors (10).

Global alignments of the six methyltransferases homologous to *M·RsrI*, using the Aalign program with a gap penalty of 4 and a deletion penalty of 12, gave the following optimal scores and percentages of identity in overlapping amino acids, respectively: *HinfI* (138, 27%), *DpnA* (132, 28%), *BamHI* (129, 34%), *PvuII* (105, 27%), *EcoP1* (86, 46%), and *EcoP15* (72, 39%). Using the same parameters, an alignment with *M·EcoRI* gave an optimal score of 34 and 15.6% amino acid identity. For comparison, an average score of 22 was obtained when the *M·RsrI* amino acid sequence was scanned against the complete contents of the PIR database. A maximum score of 1560 results from the alignment of *M·RsrI* with itself. The overall homologies between *M·RsrI* and the six methyltransferases listed above are illustrated in the dot-matrix comparisons of Figure 5, and the sequence alignments of two regions partially responsible for this homology are presented in Figure 6. One area of substantial similarity encompasses the sequence near the tetrapeptide DPPY and its variants commonly found in adenine MTases (Figure 6A). *M·BamHI* (data not shown) has the highest amino acid identity with *M·RsrI* in this region followed by *M·HinfI*. The low level of homology with *M·EcoRI* is also evident here. The second area of homology, which is not found in *M·EcoRI*, contains nine consecutive and identical amino acids in *M·RsrI* and *M·EcoP1* as well as *M·EcoP15* (Figure 6B). A similar motif has been identified in other type II adenine and cytosine MTases (10). As anticipated, *M·RsrI*

Figure 6. Alignments of amino acid sequences from two regions of homology between M·*RsrI* and other methyltransferases. Shaded boxes indicate invariant amino acids. The underlined nucleotide within the recognition sequence of each MTase represents the position of methylation as reported by Wilson (6). Gaps (indicated by dashes) and loops were introduced for optimal alignment. Panel A, shows sequence similarities in the DPPY region. Panel B, shows another region with marked sequence similarities.

has no significant homology to any of the 5mC methyltransferases (A. Bhagwat and G. Wilson, personal communications).

CONCLUSIONS

We have purified the *RsrI* methyltransferase to apparent homogeneity and examined some of its properties. M·*RsrI* is a catalytic analogue of M·*EcoRI*. These enzymes methylate the central adenine residue of the recognition sequence d(GAATTC) at the N⁶ position. The enzymes share similar optimal reaction conditions,

but differ in their molecular weights and isoelectric points. We plan to overproduce the *RsrI* MTase and to further characterize it.

We compared the deduced amino acid sequence of *M-RsrI* to those of several DNA methyltransferases. The enzyme is most homologous to two other type II adenine MTases, *HinI* and *DpnA*. *M-RsrI* is also significantly homologous to two type II MTases that transfer methyl groups to the N^4 position of cytosine residues, *BamHI* and *PvuII*. Some of these homologies may reflect the reaction mechanism required to methylate the exocyclic amino group of a nucleotide (69). Finally, *M-RsrI* also shows sequence similarities to the type III enzymes, *EcoPI* and *EcoP15*. Although structurally and catalytically more complex than the type II MTases, these enzymes also methylate adenine residues. None of the enzymes that share homology with *M-RsrI* have identical DNA recognition sequences (Figure 6).

The one methyltransferase, *M-EcoRI*, that does have the same DNA target and forms the same product lacks homology with *M-RsrI*. Since the *EcoRI* and *RsrI* endonucleases are highly homologous (12), the methyltransferases and endonucleases of these functionally equivalent R-M systems may have either joined one another independently during their evolution or the methyltransferases diverged in amino acid sequence while their linked endonucleases did not.

ACKNOWLEDGEMENTS

We would like to thank A. Bhagwat, J. Brooks, J. Gardner, P. Greene, S. Maloy, and G. Wilson for advice and critical reading of the manuscript. Special thanks to R. Gould and S. Kaplan for information on *R.sphaeroides* regulatory sequences, to A. Bhagwat and G. Wilson for performing sequence comparisons, to R. Roberts for sharing his collection of sequences of the adenine MTases, to J. Brooks, D. Landry, F. Stephenson, and P. Greene for sharing data prior to publication, and to A. Warinner for valuable assistance. This work was supported in part by NIH Grant GM25621.

*To whom correspondence should be addressed at University of Illinois, 190 Medical Sciences Building, 506 S. Mathews St., Urbana, IL 61801, USA

REFERENCES

1. Razin, A., Cedar, H., and Riggs, A. D. (eds.) (1984) DNA Methylation, Biochemistry and Biological Significance, Springer-Verlag, New York.
2. Hedgpeth, J., Goodman, H. M., and Boyer, H. W. (1972) Proc. Natl. Acad. Sci. USA **69**, 3448-3452.
3. Aiken, C., and Gumport, R. I. (1988) Nucl. Acids Res. **16**, 7901-7916.
4. Greene, P. J., Ballard, B. T., Stephenson, F., Kohr, W. J., Rodriguez, H., Rosenberg, J. M., Boyer, H. W. (1988) Gene **68**, 43-52.

5. Dugaiczky, A., Hedgpeth, J., Boyer, H. W., and Goodman, H. M. (1974) *Biochemistry* **13**, 503-512.
6. Wilson, G. G. (1988) *Gene* **74**, 281-289.
7. Chandrasegaran, S., and Smith, H. O. (1988) In Sarma, R. H., and Sarma, M. H. (eds.), *Structure and Expression Volume 1: From Proteins to Ribosomes* Adenine Press, Guilderland, NY, pp. 149-156.
8. Posfai, J., Bhagwat, A. S., Posfai, G., Roberts, R. J. (1989) *Nucl. Acids Res.* **17**, 2421-2435.
9. Wilson, G. G. (1988) *Trends in Genet.* **4**, 314-318.
10. Lauster, R. (1989) *J. Mol. Biol.* **206**, 313-321.
11. Tao, T., Walter, J., Brennan, K. J., Cotterman, M. M., and Blumenthal, R. M. (1989) *Nucl. Acids Res.* **17**, 4161-4175.
12. Stephenson, F. H., Ballard, B. T., Boyer, H. W., Rosenberg, J. M., and Greene, P. J. (1989) *Gene*, **85**, 1-12.
13. Kaszubska, W., Aiken, C., Gumpert, R. I. (1988) *Gene* **74**, 83-84.
14. Siström, W. R. (1982) *J. Gen. Microbiol.* **28**, 607-616.
15. Loenen, W. A. M., and Brammar, W. J. (1980) *Gene* **20**, 249-259.
16. Humphreys, G. O., Willshaw, G. A., Smith, H. R., and Anderson, E. S. (1976) *Mol. Gen. Genet.* **145**, 101-108.
17. Szybalski, W., Blumenthal, R. M., Brooks, J. E., Hattman, S., and Raleigh, E. A. (1988) *Gene* **74**, 279-280.
18. Marmur, J. (1961) *J. Mol. Biol.* **3**, 208-218.
19. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, NY.
20. Hohn, B., and Murray, K. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3259-3263.
21. Struhl, K. (1985) *Biotechniques* **3**, 452-453.
22. Hames, B. D., and Richwood, D. (eds.) (1981) *Gel Electrophoresis of Proteins: A Practical Approach*, IRL Press Limited, London and Washington, DC.
23. Merril, C. R., Goldman, D., Sedman, S. A., and Ebert, M. H. (1981) *Science* **211**, 1437-1438.
24. Greene, P. J., Poonian, M. S., Nussbaum, A. L., Tobias, L., Garfin, D. E., Boyer, H. W., and Goodman, H. M. (1975) *J. Mol. Biol.* **99**, 237-261.
25. Brennan, C. A., Van Cleve, M. D., and Gumpert, R. I. (1986) *J. Biol. Chem.* **261**, 7279-7286.
26. Narva, K. E., Wendell, D. L., Skrdla, M. P., and Van Etten, J. L. (1987) *Nucl. Acids Res.* **15**, 9807-9823.
27. Brooks, J. E., Blumenthal, R. M., and Gingeras T. R. (1983) *Nucl. Acids Res.* **11**, 837-851.
28. Lacks, S. A., Manarelli, B. M., Springhorn, S. S., and Greenberg, B. (1986) *Cell* **46**, 993-1000.
29. Mannarelli, B. M., Balganes, T. S., Greenberg, B., Springhorn, S. S., and Lacks, S. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4468-4472.
30. Humbelin, M., Suri, B., Rao, D. N., Hornby, D. P., Eberle, H., Pripfl, T., Kenel, S., and Bickle, T. A. (1988) *J. Mol. Biol.* **200**, 23-29.
31. Newman, A. K., Rubin, R. A., Kim, S.-H., and Modrich, P. (1981) *J. Biol. Chem.* **256**, 2131-2139.
32. Greene, P. J., Gupta, M., Boyer, H. W., Brown, W. E., and Rosenberg, J. M. (1981) *J. Biol. Chem.* **256**, 2143-2153.
33. Bougueleret, L., Schwarzstein, M., Tsugita, A., and Zabeau, M. (1984) *Nucl. Acids Res.* **12**, 3659-3676.
34. Schoner, B., Kelly, S., and Smith, H. O. (1983) *Gene* **24**, 227-236.
35. Chandrasegaran, S., Lunnen, K. D., Smith, H. O., and Wilson, G. G. (1988) *Gene* **70**, 387-392.
36. Theriault, G., Roy, P. H., Howard, K. A., Benner, J. S., Brooks, J. E., Waters, A. F., and Gingeras, T. R. (1985) *Nucl. Acids Res.* **13**, 8441-8461.

-
37. Walder, R. Y., Walder, J. A., and Donelson, J. E. (1984) *J. Biol. Chem.* **259**, 8015-8026.
 38. MacDonald, P. M., and Mosig, G. (1984) *EMBO J.* **3**, 2865-2871.
 39. Slatko, B. E., Benner, J. S., Jager-Quinton, T., Moran, L. S., Simcox, T. G., Van Cott, E. M., and Wilson, G. G. (1987) *Nucl. Acids Res.* **15**, 9781-9796.
 40. Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* **5**, 89-110.
 41. Bradford, M. M. (1978) *Anal. Biochem.* **86**, 142-146.
 42. Aune, K. C., and Tanford, C. (1969) *Biochemistry* **8**, 4579-4585.
 43. Rubin, R. A., and Modrich, P. (1977) *J. Biol. Chem.* **252**, 7265-7272.
 44. Yuen, S. W., Chui, A. H., Wilson, K. J., and Yuan, P. M. (1988) *Protein Sequencer User Bulletin* **36**, Applied Biosystems.
 45. Yuen, S., Hunkapiller, M. W., Wilson, K. J., and Yuan, P. M. (1986), *Protein Sequencer User Bulletin* **25**, Applied Biosystems.
 46. Lunnen, K. D., Barsomian, J. M., Camp, R. R., Card, C. O., Chen, S.-Z., Croft, R., Looney, M. C., Meda, M. M., Moran, L. S., Nwankwo, D. O., Slatko, B. E., Van Cott, E. M., and Wilson, G. G. (1988) *Gene* **74**, 25-32.
 47. Gingeras, T. R., and Brooks, J. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 402-406.
 48. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
 49. Betlach, M., Herschfield, V., Chow, L., Brown, W., Goodman, H. M., and Boyer, H. W. (1976) *Fed. Proc.* **35**, 2037-2043.
 50. Smith, H. R., Humphreys, G. O., Willshaw, G. A., and Anderson, E. S. (1976) *Molec. Gen. Genet.* **143**, 319-325.
 51. Fornari, C. S., Watkins, M., and Kaplan S. (1984) *Plasmid* **11**, 39-47.
 52. Shine J., and Dalgarno L. (1975) *Nature* **254**, 34-38.
 53. O'Connor, C. D., and Humphreys, G. O. (1982) *Gene* **20**, 219-229.
 54. DeHoff, B. S., Lee, J. K., Donohue, T. J., Gumpert, R. I., and Kaplan S. (1988) *J. Bacteriol.* **170**, 4681-4692.
 55. Krieg, N. R., and Holt, J. G. (eds.) (1984) *Bergey's Manual of Systematic Bacteriology* **1**, Williams & Wilkins, Baltimore, MD.
 56. Shapiro, H. S. (1970) In Sober, H. A. (ed.), *CRC Handbook of Biochemistry Selected Data For Molecular Biology*, The Chemical Rubber Co., Cleveland, 2nd Ed, pp. H80-H88.
 57. Sharp, P. M., and Li, W.-H. (1986) *Nucl. Acids Res.* **14**, 7737-7749.
 58. Hattman, S., Wilkinson, J., Swinton, D., Schlagman, S., MacDonald, P. M., and Mosig, G. (1985) *J. Bacteriol.* **164**, 932-937.
 59. Guschlbauer, W. (1988) *Gene* **74**, 211-214.
 60. McClarin, J. A., Frederick, C. A., Wang, B. C., Greene, P. J., Boyer, H. W., Grable, J., and Rosenberg, J. M. (1986) *Science* **234**, 1526-1541.
 61. Yanofsky, S. D., Love, R., McClarin, J. A., Rosenberg, J. M., Boyer, H. W., and Greene, P. J. (1987) *Proteins* **2**, 273-282.
 62. King, K., Benkovic, S. J., and Modrich, P. (1989) *J. Biol. Chem.* **264**, 11807-11815.
 63. Lauster, R., Trautner, T. A., and Noyer-Weidner, M. (1989) *J. Mol. Biol.* **206**, 305-312.
 64. Balganesch, T. S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K., and Trautner, T. A. (1987) *EMBO J.* **6**, 3543-3549.
 65. Trautner, T. A., Balganesch, T. S., and Pawlek, B. (1988) *Nucl. Acids Res.* **16**, 6649-6658.
 66. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., and Trautner, T. A. (1988) *EMBO J.* **7**, 2601-2609.
 67. Lauster, R. (1989) *Nucl. Acids Res.* **17**, 4402.
 68. Price, C., Lingner, J., Bickle, T. A., Firman, K., and Glover, S. W. (1989) *J. Mol. Biol.* **205**, 115-125.
 69. Pogolotti Jr., A. L., Ono, A., Subramaniam, R., and Santi, D. V. (1988) *J. Biol. Chem.* **263**, 7461-7464.
-