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

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

Department of Biochemistry and College of Medicine, University of Illinois, Urbana, IL 61801, USA and Department of Biochemistry, University of Southampton, Southampton S09 3TU, UK

Received September 26, 1989; Revised and Accepted October 24, 1989 EMBL accession no. X16456

ABSTRACT

Rsrl DNA methyltransferase (M-Rsrl) 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 Rsrl methyltransferase (MTase) is 33,600 Da. A fragment of R. sphaeroides chromosomal DNA exhibited M-Rsrl activity in E. coli and was used to sequence the rsrIM gene. The deduced amino acid sequence of M-Rsrl shows partial homology to those of the type II adenine MTases HinfI and Dpna and N^4-cytosine MTases BamHI and Pvull, and to the type III adenine MTases EcoP1 and EcoP15. In contrast to their corresponding isoschizomeric endonucleases, the deduced amino acid sequences of the Rsrl and EcoRI MTases show very little homology. Either the EcoRI and Rsrl 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 N^6-amino group of adenine, the N^4-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 Rsrl 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-Rsrl. In addition, we have cloned and sequenced the rsrlM 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-Rsrl shares homology with the adenine MTases Hinfl, DpnA, EcoP1, and EcoP15. Furthermore, the enzyme is homologous to the N⁴-cytosine MTases BamHI and PvuII.

Most surprisingly, M-Rsrl lacks homology to its functional analogue M-EcoRI. Stephenson et al. (12) showed that the Rsrl endonuclease (R-Rsrl) 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 Rsrl 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-Rsrl has appeared (13) and the sequence of the rsrlM 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 LA7.1 and its host strain E. coli WL95 hsdRk-redB trpR tonA SulI Sull (P2) lysogen were obtained from W. Brammar (15). E. coli C600 hsdRk+Mk+ mcrA+B+ supE44 thi-I leuB6 lacYI 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 hsdS20Rb-Mb- mcrA+B- mrr- supE44 leuB6 ara-I4 proA2 lacYI was used to propagate vector pBR328, both of which were gifts from J. Brooks. E. coli TB1 hsdR17Rk+Mk+ara Δ(lac-proAB) rpsL (φ80 lacZΔM15), obtained from M. Schuler, was used to propagate plasmid pUC18. Vector pTZ18U, a gift from V. Bankaitis, was propagated in E. coli NM522 hsdSR-Mb+ supE thi-I Δ(lac-proAB) IF'proAB, lacI9ZΔM15 obtained from B. Kemper. Phagemid pTZ18U contains both the fl and ColEl origins of replication, a T7 RNA polymerase promoter in phase with the lacZ' and Amp' 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 μg/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 LA7.1 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) Rsrl Methyltransferase Assays

Qualitative. Aliquots (1 µl) of M-Rsrl were incubated with 1 µg bacteriophage λ DNA in a total of 10 µl 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 µl 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 µl of gel-loading dye (19) and heating for 10 min. at 65°C prior to agarose gel electrophoresis. Reactions in which M-Rsrl and R-EcoRI were omitted, and reactions in which only M-Rsrl was omitted, were included as controls. Phage λ DNA that had been methylated by M-Rsrl 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 µl) were incubated with 20 nM (5 µg) pRK112-8 in a total of 50 µl of 100 mM Tris-HCl (pH 8), 5 mM EDTA, 0.4 mg/ml BSA, 2.5 mM DTT, 1.8 mM [3H]AdoMet (11 Ci/mmol) for 15 min. at 30°C or 37°C. Samples of 45 µl were spotted on DE81 filters (2.3 cm circles; Whatman BioSystems, Inc.), washed 3 times with 0.1 M NH4HCO3 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 assays 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 MgSO4 or MgCl2 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(pCTGAAATTCAG), 1 µM, was annealed and incubated in 50 µl of the quantitative assay reaction buffer with 3.6 mM [3H]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-32P]ATP (75 Ci/mmol). The doubly-labeled decanucleotide was partially digested with snake venom phosphodiesterase (VPD). Aliquots, 5 µl, 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 32P
were excised and the $^3$H 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-Rsrl, derived from the nucleotide sequence, was compared to other methyltransferases. The following DNA adenine MTases were examined: CviIII (26), dam (27), DpnA (28), DpnII (29), EcoP1 (30), EcoP15 (30), EcoRI (31, 32), EcoRV (33), HhaII (34), HinfI (35), PaeR7 (36), PstI (37), TaqI (38), and TaqII (39). Two $\mathrm{N}^4$-cytosine MTases, BamHI (J. Brooks and D. Landry, personal communications) and PvuII (11) were also compared to M-Rsrl. 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-Rsrl 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-Rsrl to the following C5 MTases: BpiI, Bsp/BsuRI, BsuFl, dcm, Ddel, EcoRII, f3T, HaeIII, HhaI, HpalI, Mspl, NgoPII, pII1s, 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 Bioseries) 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 Biotechnolgy, 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 $\epsilon_{281.5} = 2.64$ in 0.1 M KCl (42).

RESULTS AND DISCUSSION

a) Purification of the Rsrl Methyltransferase

Since the Rsrl 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-Rsrl elutes on the phosphocellulose column at 0.3 M KCl in buffer B and is separated from the R-Rsrl activity. R-RsrlII often coelutes with M-Rsrl at this step.

HPLC Hydroxylapatite Chromatography. Solid (NH$_4$)$_2$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-Rsrl 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-Rsrl 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 Rsrl Methyltransferase

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

*100 g of R. sphaeroides cell paste was used.

b Reliable measurements of M-Rsrl activity cannot be obtained in these fractions.

c All 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-Rsrl 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 Rsrl and EcoRI MTases under denaturing conditions. M-Rsrl (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·Rsrl. 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·Rsrl 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 Rsrl 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·Rsrl under non-denaturing conditions was examined using high performance, size-exclusion chromatography (Materials and Methods). A M·Rsrl sample, approximately 10 μg 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·Rsrl 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·Rsrl 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·Rsrl, which had been electroblotted onto a solid support, were performed. One sample of pure M·Rsrl, 7 μg (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 μg, of partially purified M·Rsrl (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·Rsrl 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·Rsrl are double-underlined in Figure 3. The N-terminal methionine is absent in the purified M·Rsrl, as it is in the case with R·Rsrl (3, 4) and both EcoRI enzymes (31, 32).

Isoelectric Point. Samples of purified M·Rsrl and M·EcoRI, 6 μg each, were focused on a thin-layer polyacrylamide gel. Following staining, one band was observed for each enzyme. The pI value of M·Rsrl 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 \textit{M-Rsrl}$^a$

<table>
<thead>
<tr>
<th>Oligodeoxyribonucleotide</th>
<th>Radioactivity (cpm)</th>
<th>Ratio $^{3}\text{H}/^{32}\text{P}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d(p\text{CTGAATTCA})$</td>
<td>772</td>
<td>4350</td>
</tr>
<tr>
<td>$d(p\text{CTGAATTCA})$</td>
<td>1296</td>
<td>4870</td>
</tr>
<tr>
<td>$d(p\text{CTGAATTCC})$</td>
<td>1125</td>
<td>4248</td>
</tr>
<tr>
<td>$d(p\text{CTGAATT})$</td>
<td>977</td>
<td>3358</td>
</tr>
<tr>
<td>$d(p\text{CTGAAT})$</td>
<td>570</td>
<td>2249</td>
</tr>
<tr>
<td>$d(p\text{CTGA})$</td>
<td>1911</td>
<td>2600</td>
</tr>
<tr>
<td>$d(p\text{CTG})$</td>
<td>179</td>
<td>4511</td>
</tr>
</tbody>
</table>

$^a$See Materials and Methods. $^3\text{H}$ is in $N^6\text{mA}$ and $^{32}\text{P}$ is in the 5'-phosphomonoester terminus of the oligodeoxyribonucleotide.

c) Catalytic Properties

Identification of the Methylated Base. The central adenine residue of duplex $d(p\text{CTGAATTCA})$ is methylated by \textit{M-Rsrl}, as indicated by the presence of $^3\text{H}$ in only the oligonucleotides containing this base (Table II). The increase in the $^3\text{H}/^{32}\text{P}$ ratio in $d(\text{CTGAmA})$ has been observed before (25) and is due to the faster rate of cleavage by VPD of the $d(\text{CTGAA})$ contained in this incompletely methylated product. In another experiment total hydrolysis of $d(\text{GTGAAATTCTAC})$, that had been incubated with $[^3\text{H}]\text{AdoMet}$ and \textit{M-Rsrl}, followed by HPLC analysis showed label in $N^6\text{mA}$ only (Webb and Gumport, unpublished observations). Thus, \textit{M-Rsrl} catalyzes the transfer of methyl groups from AdoMet to the same position on central adenine residue in the recognition sequence as does \textit{M-EcoRI} (5).

Reaction Optimization. \textit{M-Rsrl} 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. \textit{M-Rsrl} 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 \textit{Rsrl} methyltransferase and endonuclease activities show similar responses to pH and NaCl, however they differ in their temperature optima. Whereas \textit{R-Rsrl} is maximally active at 23°C and inactive at 37°C (3, 4) \textit{M-Rsrl} 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 \textit{R. sphaeroides}. \textit{M-Rsrl} 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 \textit{M-EcoRI} (43) and \textit{R-Rsrl}.
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 Mg2+ is required for the endonuclease activity, 20 mM MgCl2 or MgSO4 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
Nucleic Acids Research

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 Rsrl 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 SalI, 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 Aval sites were identified in the recombinant plasmid pBR-RsrtM7.0. Although Aval cleaves degenerate DNA sequences, a 3.2 kb fragment (Figure 2) was successfully ligated to pUC18 previously cut with Aval. In preparation for DNA sequencing, the 3.2 kb fragment carrying the rsrlM gene was ligated to pTZ18U using the SstI and XbaI sites within the polylinker to yield pTZ-RsrlM3.2. The isolated plasmids, pBR-RsrtM7.0, pUC-RsrlM3.2, and pTZ-RsrlM3.2, were totally resistant to cleavage by R-EcoRI, indicating that the rsrlM gene was expressed in E. coli.

f) Nucleotide Sequence of the rsrlM Gene

Sequence Determination. The nucleotide sequence of the rsrlM 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-RsrlM3.2 using exonuclease III and S1 nuclease. The longest deletion still carrying the intact rsrlM 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 rsrlM 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 rsrlM 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-Rsrl 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
The nucleotide and deduced amino acid sequences of M·Rsrl. The nucleotide sequence starts with the Ndel 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 Rsrl MTase, the identities of single-underlined amino acids were not determined. The start and stop codons of rsrlM are italicized. The DPPY sequence is in bold letters.

observed value of 33,600 ± 2,000 Da for purified M·Rsrl. 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 rsrlM sequence were confirmed by electrophoresis (data not shown). M·Rsrl 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·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 rsrlM was obtained from an independently isolated clone by Stephenson and Greene (Nucl. Acids Res. (1989) 17, this issue).
Relative Locations of rsrlM and rsrlR Genes. In common with other type II R-M genes the rsrlM and rsrlR genes are in proximity (9). The C-terminus of the Rsrl 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 rsrl 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 rsrl genes are located on the chromosome. Degenerate heptadecadeoxyribonucleotides complementary to the 5'-end of the rsrlR 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 rsrl 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 rsrlR gene. In addition, less perfect matches to this sequence are 45 (GGTN9AGCC) and 71 (GGAN6CGCG) base pairs upstream of the initiator codon of the rsrlM 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 rsrlM gene (Figure 3). Although it is possible that the imperfect consensus promoters occasionally serve to initiate transcription of only the rsrlM 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 rsrlR gene in R. sphaeroides cells, and that the Shine-Dalgarno sequence preceding the rsrlM 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-Rsrl 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 rsrlM 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 rsrlM. Another gene may be located immediately
## Table III. Codon Usage in rsrIM and ecoRIM

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Codon</th>
<th>rsrIM&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ecoRIM&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Amino Acid</th>
<th>Codon</th>
<th>rsrIM</th>
<th>ecoRIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>GCA</td>
<td>7</td>
<td>3</td>
<td>Gly</td>
<td>GGA</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GCC</td>
<td>6</td>
<td>1</td>
<td></td>
<td>GGC</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>GCG</td>
<td>13</td>
<td>2</td>
<td></td>
<td>GGG</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>GCU</td>
<td>7</td>
<td>4</td>
<td></td>
<td>GGU</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td>7.2</td>
<td>5.2</td>
</tr>
<tr>
<td>Arg</td>
<td>AGA*</td>
<td>2</td>
<td>10</td>
<td>His</td>
<td>CAC</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>AGG*</td>
<td>1</td>
<td>1</td>
<td>CAU</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGA</td>
<td>5</td>
<td>1</td>
<td></td>
<td>3.1</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGC</td>
<td>5</td>
<td>1</td>
<td>UUA*</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGG</td>
<td>6</td>
<td>0</td>
<td>AUC</td>
<td>10</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGU</td>
<td>2</td>
<td>0</td>
<td>AUU</td>
<td>7</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.6</td>
<td>4.0</td>
<td></td>
<td></td>
<td>5.6</td>
<td>7.1</td>
</tr>
<tr>
<td>Asn</td>
<td>AAC</td>
<td>13</td>
<td>9</td>
<td>Leu</td>
<td>CUA*</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>AAU</td>
<td>3</td>
<td>19</td>
<td>CUC</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.0</td>
<td>8.6</td>
<td></td>
<td>CUG</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Asp</td>
<td>GAC</td>
<td>11</td>
<td>2</td>
<td>CUU</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GAU</td>
<td>10</td>
<td>21</td>
<td>UUA*</td>
<td>0</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.6</td>
<td>7.1</td>
<td></td>
<td>UUG*</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Cys</td>
<td>UGC</td>
<td>3</td>
<td>4</td>
<td>Lys</td>
<td>AAA</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>UGU*</td>
<td>1</td>
<td>3</td>
<td>AAG</td>
<td>11</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2</td>
<td>2.1</td>
<td></td>
<td></td>
<td>5.0</td>
<td>10.7</td>
</tr>
<tr>
<td>Gln</td>
<td>CAA</td>
<td>7</td>
<td>4</td>
<td>Met</td>
<td>AUG</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CAG</td>
<td>6</td>
<td>1</td>
<td></td>
<td></td>
<td>2.8</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1</td>
<td>1.5</td>
<td></td>
<td></td>
<td>3.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Glu</td>
<td>GAA</td>
<td>9</td>
<td>10</td>
<td>Phe</td>
<td>UUC</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>GAG</td>
<td>9</td>
<td>10</td>
<td>UUU*</td>
<td>4</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.6</td>
<td>6.1</td>
<td></td>
<td></td>
<td>3.4</td>
<td>7.1</td>
</tr>
</tbody>
</table>

<sup>a</sup>The number indicates the times the codon is used in the 319 amino acids of M-RsrI and the 326 amino acids of M-EcoRI.

<sup>b</sup>Percent amino acid composition.

<sup>*</sup>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 (GGCN<sub>11</sub>TGCC) 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 *EcoRl* 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-EcoRl* 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.

---

10417
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-Rsrl, 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-Rsrl 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-Rsrl 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-Rsrl 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-Rsrl and from residue 297 to 326 of M-EcoRI. Another short area of homology between M-Rsrl 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 Rsrl and EcoRI MTases is the tetrapeptide SSES found at residue 124 of M-EcoRI and residue 25 of M-Rsrl. A subsequence, SSE, is also found in different
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 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.
Figure 5. Dot-matrix comparisons between Rsrl 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 EcoPI 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.

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-Rsrl 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 Rsrl MTase to other Methyltransferases

The type II MTases can generate one of the following modified bases: N^6-methyladenine (N^6mA), N^4-methylcytosine (N^4mC), and 5-methylcytosine (5mC). M-Rsrl is most similar to two closely related (67) type II adenine methyltransferases Hinfl and DpnA. Interestingly, M-Rsrl also exhibits significant sequence similarities to two type II N^4mC methyltransferases, Pvull and BamHI. Blumenthal et al. (11) have demonstrated that M-Pvull has substantial similarity to two other N^6mA MTases (M-EcoRI and M-Psrl), and suggested that the N^4mC and N^6mA MTases may use the same mechanism of methylation. In addition, BamHI is also homologous to some other N^6mA 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-Rsrl shows sequence similarities to two allelic type III N^6mA MTases, EcoPl 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-Rsrl, 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: Hinfl (138, 27%), DpnA (132, 28%), BamHI (129, 34%), Pvull (105, 27%), EcoPl (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-Rsrl amino acid sequence was scanned against the complete contents of the PIR database. A maximum score of 1560 results from the alignment of M-Rsrl with itself. The overall homologies between M-Rsrl 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-Rsrl in this region followed by M-Hinfl. 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-Rsrl and M-EcoPl 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-Rsrl
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 Rsrl MTase and to further characterize it.

We compared the deduced amino acid sequence of M:Rsrl to those of several DNA methyltransferases. The enzyme is most homologous to two other type II adenine MTases, Hinfl and DpnA. M:Rsrl is also significantly homologous to two type II MTases that transfer methyl groups to the N^4 position of cytosine residues, BamHI and Pvull. Some of these homologies may reflect the reaction mechanism required to methylate the exocyclic amino group of a nucleotide (69). Finally, M:Rsrl also shows sequence similarities to the type III enzymes, EcoP1 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:Rsrl 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:Rsrl. Since the EcoRI and Rsrl 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