How Does a DNA Interacting Enzyme Change Its Specificity during Molecular Evolution? A Site-Directed Mutagenesis Study at the DNA Binding Site of the DNA-(Adenine-N^6)-methyltransferase *Eco*RV†

Carsten Beck, Susanne Cranz,‡ Merdol Solmaz, Markus Roth, and Albert Jeltsch*

Institut für Biochemie, Fachbereich 8, Justus-Liebig-Universität, Heinrich-Buff-Ring 58, 35392 Giessen, Germany

Received May 16, 2001; Revised Manuscript Received June 22, 2001

ABSTRACT: The *Eco*RV DNA-(adenine-N6)-methyltransferase (MTase) recognizes GATATC sequences and modifies the first adenine residue within this site. Parts of its DNA interface show high sequence homology to DNA MTases of the *dam* family which recognize and modify GATC sequences. A phylogenetic analysis of M.*Eco*RV and *dam*-MTases suggests that *Eco*RV arose in evolution from a primordial *dam*-MTase in agreement to the finding that M.*Eco*RV also methylates GATC sites albeit at a strongly reduced rate. GATCTC sites that deviate in only one position from the *Eco*RV sequence are preferred over general *dam* sites. We have investigated by site-directed mutagenesis the function of 17 conserved and nonconserved residues within three loops flanking the DNA binding cleft of M.*Eco*RV. M.*Eco*RV contacts the GATATC sequence with two highly cooperative recognition modules. The contacts to the GAT-part of the recognition sequence are formed by residues conserved between *dam* MTases and M.*Eco*RV. Mutations at these positions lead to an increase in the discrimination between GATATC and GATC substrates. Our data show that the change in sequence specificity from *dam* (GATC) to *Eco*RV (GATATC) was accompanied by the generation of a second recognition module that contacts the second half of the target sequence. The new DNA contacts are formed by residues from all three loops that are not conserved between M.*Eco*RV and *dam* MTases. Mutagenesis at important residues within this module leads to variants that show a decreased ability to recognize the TC-part of the GATATC sequence.

DNA methyltransferases are a wide spread and important family of enzymes catalyzing the transfer of a methyl group from S-adenosyl-methionine (AdoMet)[1] to specified nucleobases within defined recognition sites on the DNA (see, for review, refs *1*−*5*). The methyl group is transferred to either the C5 or N4 position of cytosine or the N6 position of adenine, yielding 5-methylcytosine, 4-methylcytosine or 6-methyladenine. Accordingly, DNA MTases can be subdivided into two groups, C-MTases comprising cytosine-C5 MTases (EC 2.1.1.37 and 2.1.1.73) which form a C−C bond and N-MTases comprising Adenine-N6 (EC 2.1.1.72) and cytosine-N4 MTases (EC 2.1.1.113) which form a C−N bond. N-MTases are related in their amino acid sequences (*6*, *7*) and show overlapping specificity (*8*, *9*). All MTases rotate their target base out of the DNA helix prior to catalysis (*10*−*12*), a catalytic mechanism that is also observed for many other enzymes that interact with DNA, like various DNA repair enzymes (see, for review, ref *13*). In prokaryotes, all three types of MTases exist. DNA MTases are required to control DNA replication, they are important for post-replicative mismatch repair and one component of restriction-modification-systems (see, for review, ref *14*−*16*). Recently, it was found that DNA-(adenine-N^6)-MTases are also important for pathogenicity of different human pathogens [*Escherichia coli* (*17*), *Salmonella thyphimurium* (*18*−*20*), *Neisseria meningitidis* (*21*)]. In higher eukaryotes, as far as known today, only C-MTases occur that are involved in epigenetic control of gene expression during development and in processes such as parental imprinting and X-chromosome inactivation (see, for review, refs *5* and *22*).

Here, we investigate the M.*Eco*RV MTase that methylates the first adenine within G**A**TATC sequences (*23*). The enzyme comprises 298 amino acid residues (*24*) and, according to the order of characteristic motifs, it can be assigned to the α-group of N-MTases (*6*, *7*, *25*). M.*Eco*RV is closely related to several other α-type MTases most of which are *dam* (DNA adenine methylation) MTases that recognize GATC sequences (Figure 1). The similarity of the first three positions of the recognition sequences of the *Eco*RV/*dam* family suggests that M.*Eco*RV might be related to the *dam*-MTases. Indeed, it was shown that M.*Eco*RV also methylates GATC sites in vivo (*26*). Moreover, a highly conserved amino acid motif, the so-called RNFP-motif (R128, N130, F135, and P138 in M.*Eco*RV), could be identified within the putative DNA recognition domain of all members of this family (*27*, *28*).

† This work has been supported by the Deutsche Forschungsgemeinschaft (JE 252/2-3).

* To whom correspondence should be addressed. Phone: +49 641 99 35410. Fax: +49 641 99 35409. E-mail: Albert.Jeltsch@chemie.bio.uni-giessen.de.

‡ Present address: Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.

[1] Abbreviations: AdoMet, S-adenosylmethionine; Bt, Biotin; *dam*, DNA adenine methylation; GST, glutathione-S-transferase; ^{m}A, 6-methyladenine; MTase, DNA methyltransferase.

10.1021/bi0155450 CCC: $20.00
Published on Web 08/18/2001

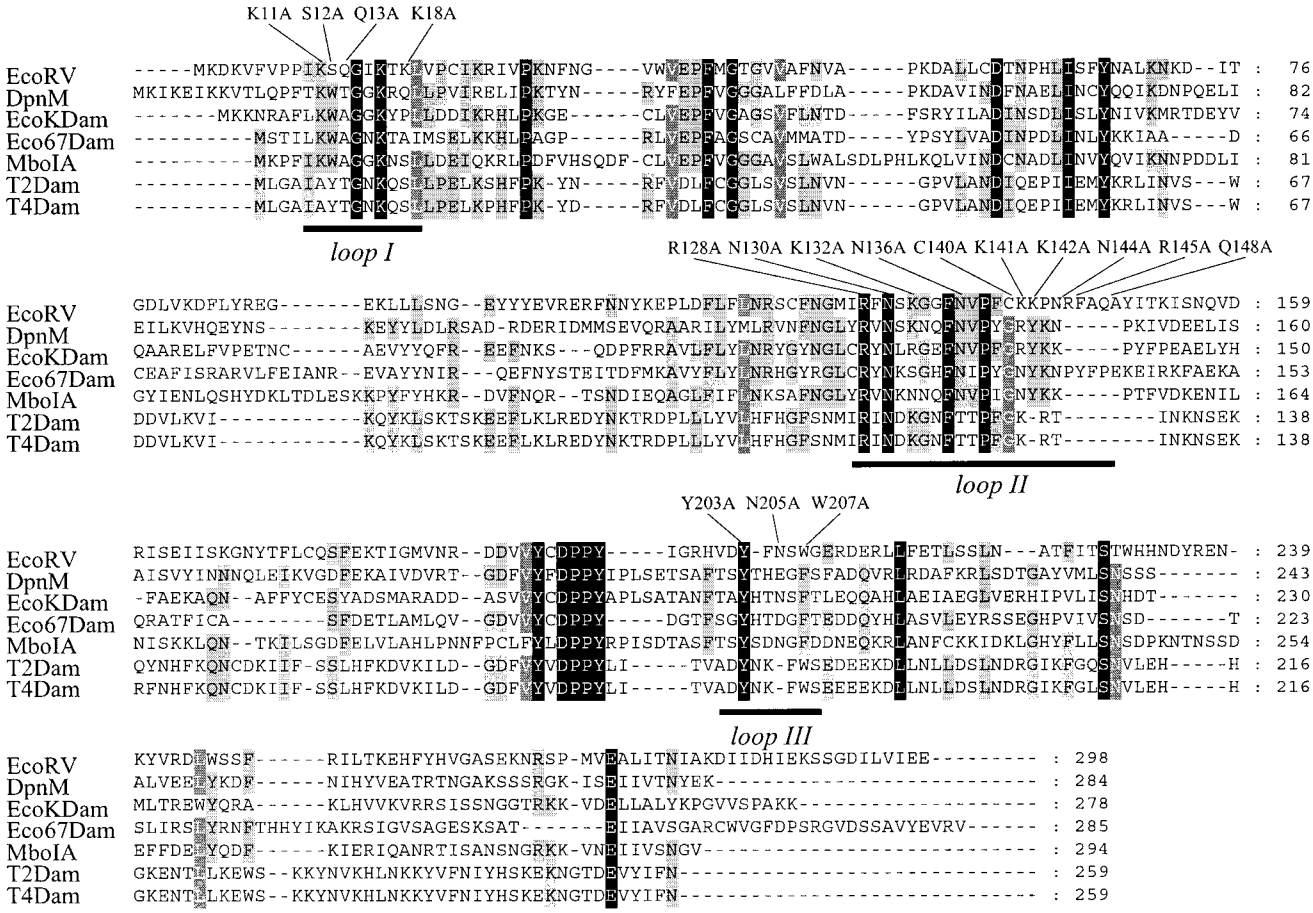

FIGURE 1: Multiple sequence alignment of M.*Eco*RV and representative *dam*-MTases. The proteins included in this alignment are M.*Eco*RV (accession no. P04393), *Dpn*M (AAA88580), the two *dam* MTases from *E. coli* (*Eco*K*Dam*: P00475, *Eco*67*Dam*: P00475), *Mbo*IA (P34720), as well as phage T2 (P12427) and T4 *dam* MTases (P12427). The approximate positions of the loops I, II, and III in M.*Eco*RV as well as the amino acid residues subjected to mutagenesis are indicated.

Given the differences and similarities in the recognition sequences of *dam* MTases and M.*Eco*RV, three groups of DNA contacting residues are expected in M.*Eco*RV: (i) Residues that contact the GATATC-part of the *Eco*RV site should be conserved between M.*Eco*RV and *dam*-MTases. (ii) Residues involved in recognition of the C in the *dam* sequence (GATC) should be conserved among the *dam*-MTases but different between *dam* and M.*Eco*RV, where an A must be recognized at the corresponding position (GATATC). (iii) Residues that are involved in contacts to the GATATC-part of the M.*Eco*RV sequence most likely do not have an important function in the *dam*-MTases and thus should not be conserved among the *dam*-family.[2]

We follow a dual approach to analyze the mechanism of DNA recognition by M.*Eco*RV combining the usage of DNA substrates containing altered *Eco*RV sites and mutagenesis of putative DNA contacting residues. Candidate residues for mutagenesis were selected on the basis of the structure of the *Dpn*M MTase (*29*) which recognizes GATC and so far is the only known structure of an enzyme belonging to the α-group of N-MTases. Although the structure was solved without DNA, the approximate location of the DNA binding cleft and of several residues that might be involved in DNA recognition could be identified. The DNA binding cleft of *Dpn*M is formed by at least three loops of the catalytic domain and the small domain of the enzyme (Figure 2).

Loop 1 comprises the N-terminal region of the enzymes immediately preceding motif X. It contains a highly conserved Lys (Lys16 in M.*Eco*RV) that has been shown to be involved in contacts to the cofactor (*28*, *29*) and to the flipped base (*30*). This residue has not been further analyzed here. Since M.*Eco*RV binds its cofactor and substrate in a cooperative fashion, loop 1 probably comes close to the DNA and other amino acids from this loop might contact the recognition sequence. We therefore investigated Lys11, Ser12, Gln13, and Lys18. Whereas a Lys is present at the position equivalent to Lys11 also in *dam*-MTases, the other residues are dissimilar between M.*Eco*RV and *dam*-MTases. Interestingly, there is an aromatic amino acid conserved at position 12 within *dam*-MTases, whereas M.*Eco*RV has the Ser.

The region of loop 2 has been implicated in DNA recognition first on the basis of a multiple sequence alignment which showed the presence of the highly conserved

[2] Clearly, it would be beneficial for the analysis, if a set of related enzymes recognizing GATATC were available, because under these conditions residues conserved among the GATATC-family but not between the GATATC and GATC enzymes would be good candidates for recognition of the GATATC-part. In the present data set where only one GATATC enzyme is available (M.*Eco*RV) residues that are not conserved between the *dam*-MTases and M.*Eco*RV and also not within the *dam* family could be involved in contacts to the GATATC-part, but they just could also have no important function at all.

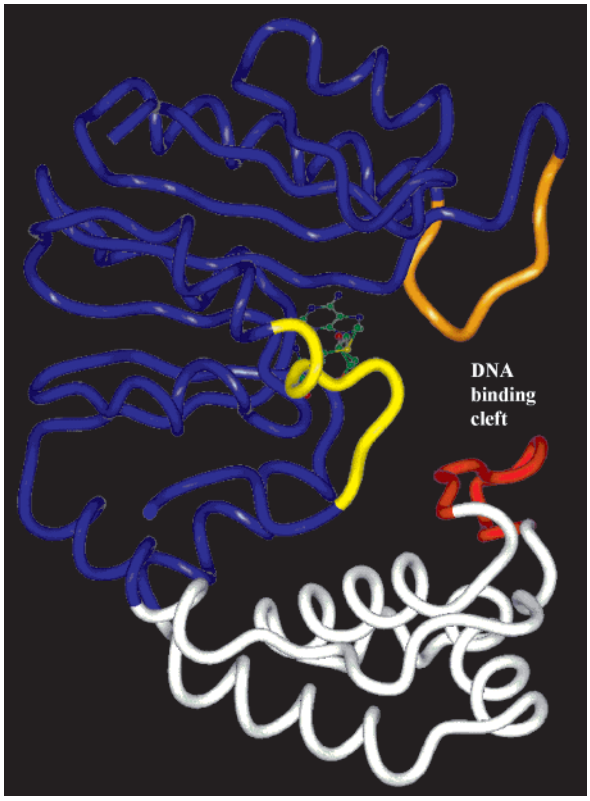

FIGURE 2: Structure of *Dpn*M (PDB code: 2DPM), a *dam*-MTase that is related to M.*Eco*RV. The catalytic domain and small domain of the protein are colored blue and white, respectively. The cofactor (AdoMet) is shown as ball-and-stick model. The regions corresponding to the loops 1, 2, and 3 of M.*Eco*RV are colored yellow, red, and orange, respectively.

RNFP-motif (Arg128−Pro138 in M.*Eco*RV) and other MTases that recognize sequences containing a GAT triplet (*27*). Later, a role in DNA binding has been confirmed for Arg128 and Asn130 (*28*). Here, we have also investigated the residues Lys132, Asn136, Lys141, and Asn144 which show a moderate conservation between M.*Eco*RV and the *dam*-MTases. Following the RNPF-motif, no significant similarity can be found between M.*Eco*RV and the *dam*-MTases, until motif IV appears which contains the catalytic DPPY-tetrapeptide (Asp193−Tyr196 in M.*Eco*RV). Therefore, the amino acid residues following Asn144 as well as some not conserved residues within the conserved motif are candidate residues to be responsible for DNA recognition of the 3′-end of the target site**.** Searching for polar and nonconserved residues, the following positions were selected for site-directed mutagenesis: Cys140 (because a Gly is strongly conserved among the *dam*-MTases at the equivalent position), Lys142, Asn144, Arg145, and Gln148.

The third loop flanking the DNA binding cleft of *Dpn*M consists of residues that immediately follow motif IV. This flexible loop is not ordered in the crystal structure of M.*Dpn*M (*29*). In the corresponding region of M.*Eco*RV, we investigated the residues Tyr203, Asn205, and Trp207 that all are highly to moderately conserved between M.*Eco*RV and *dam*-MTases. Trp207 is the first residue of motif V as defined by Malone et al. (*6*) and an aromatic residue is conserved among all α-MTases at this position. Interestingly, in the crystal structure of the *Taq*I MTase in complex with DNA the region following motif IV is ordered and interacts with the minor groove of the DNA (*11*), suggesting that a DNA interaction of this loop also might be possible in M.*Eco*RV.

Altogether 17 amino acids were exchanged by site-directed mutagenesis to alanine in order to prevent potential interactions with the DNA backbone and/or nucleobases. Every M.*Eco*RV variant was tested for its ability to methylate DNA in vivo, and all were purified and also characterized in vitro with respect to their catalytic properties on different substrates as well as DNA binding affinity.

EXPERIMENTAL PROCEDURES

Plasmids and Site-Directed Mutagenesis by PCR. The plasmid pGEXMRV (*31*) was used for expression of a GST-M.*Eco*RV fusion protein. It contains the *m.ecorv* gene in frame to the 3′ end of the gene for the gluthatione-S-transferase (GCT) that is controlled by a synthetic promotor which consists of a T5 promotor and two *lac*-operators. The vector carries two GATATC sites that can be employed for in vivo activity assays. PCR mutagenesis was performed using a protocol modified after Kirsch and Joly (*32*). All altered genes were completely sequenced to verify the mutations and exclude the presence of additional base exchanges.

In Vivo Activity Assay. Due to basal expression, plasmids encoding a DNA MTase usually are resistant to digestion by the corresponding restriction endonuclease. This allows to obtain information about the catalytic activity of the MTase variants in vivo by a restriction digest of the plasmid. For this assay, pGEXMRV plasmids were grown in *E. coli* ER1992 cells. Plasmid preparations from overnight cultures were carried out using DNA minipreparation kits (Qiagen) according to the instructions of the supplier. A total of 1−2 μg of plasmid DNA was incubated with 100 nM R.*Eco*RV in 10 μL of Tris/HCl, pH 7.5, 10 mM $MgCl_2$, and 50 mM NaCl for 30−60 min. Digestion was analyzed by gel electrophoresis in 1% agarose gels, stained in ethidium bromide solution.

Overexpression and Purification of Wild-Type M.EcoRV and the Variants. The pGEXMRV plasmids encoding wild-type GST-M.*Eco*RV and the variants were transformed into *E. coli* ER1992 cells. The cells were grown at 37 °C until 0.7 OD_{600nm}. Overexpression of the GST-M.*Eco*RV protein was induced by addition of 1 mM IPTG. After 5 h, cells were harvested by centrifugation (15 min, 4000*g*, 4 °C) and the cell pellet was washed with STE buffer (Tris/HCl, pH 8, 100 mM NaCl, 0.1 mM EDTA). After sonification the cells debris was removed by centrifugation (30 min, 40000*g*, 4 °C). Purification was performed at 4 °C by chromatography using 1 mL of GSH−Sepharose (Pharmacia Biotech) under the following buffer conditions: 20 mM HEPES/pH 8, 0.1 mM dithiothreitol, 500 mM NaCl, 1 mM EDTA, 10% (v/v) glycerol. The eluated proteins were concentrated using PEG (20000) and dialyzed against 20 mM HEPES/pH 8, 0.1 mM DTT, 500 mM NaCl, 1 mM EDTA, and 70% glycerol.

DNA Binding Experiments. DNA binding of wild-type GST-M.*Eco*RV and the variants was analyzed by nitrocellulose filter binding experiments as described (*12*, *33*) using a 40-mer oligonucleotide d(ATGCGA GATATC GTGCAT-TCCGACAGCTTCGCCAGTCACT)/ d(AGTGACTGGC-GAAGCTGTCGGAATGCAC GATATC TCGCAT) that

contains one *Eco*RV site. The oligonucleotide substrate was radioactively labeled using γ[^{32}P]ATP (NEN) and T4 polynucleotide kinase (MBI Fermentas). DNA (2.5 nM) was incubated with different concentrations of enzyme (0−500 nM) in 50 μL of binding buffer (50 mM Tris/HCl, pH 7.5, 20 mM NaCl, BSA 50 μg/mL, 250 μM S-adenosylhomocysteine) for 30 min at room temperature. The nitrocellulose membrane (Macherey-Nagel porablot NCP, 0.45 μm porosities) was preincubated in washing buffer (50 mM Tris/HCl, pH 7.5, 20 mM NaCl) for 30 min. Then, the membrane was transferred into a dot blot chamber (Bio-Rad), the slots were washed once with 180 μL of washing buffer. The samples were transferred into the wells of the dot blot chamber and immediately washed twice with 180 μL washing buffer. After drying the membrane, the bound radioactivity was measured using an Instant Imager (Canberra Packard) and the signals were fitted to an equation describing a bimolecular binding equilibrium. Under these conditions, mainly unspecific DNA binding of M.*Eco*RV is observed (*33*).

In Vitro Methylation Assay with Different Biotinylated DNA Substrates. In vitro DNA methylation kinetics were carried out using biotinylated oligonucleotide substrates and [methyl-^{3}H]AdoMet as described (*34*). All reactions were performed under single turnover conditions in 50 μL containing 0.5 μM oligonucleotide, 1 μM enzyme, 0.7 μM [^{3}H]SAM (3.0 × 1051 Bq/mol, NEN), 10 μg/mL BSA in 50 mM HEPES (pH 7.5) and 50 mM NaCl. After defined time intervals, 3 μL of the reaction mixture were transferred into the wells of an avidin coated microplate to immobilize the oligonucleotides. The incorporation of [^{3}H]methyl groups into the DNA was quenched by addition of unlabeled AdoMet to the binding buffer. Unreacted AdoMet and enzyme were removed by washing five times with 180 μL of PBST buffer (140 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM K_2HPO_4, and 0.05% v/v Tween 50, pH 7.2). To release the radioactivity incorporated into the DNA, the wells were incubated with nonspecific endonuclease from *Serratia marcescens* and the radioactivity determined by liquid scintillation counting.

The following oligonucleotides were used as methylation substrates:

L20 d(GATCGTAGATATCGCATCGA) / Bt-d(TCGATGCG^{m}ATATCTACGATC)
dam d(CGCGGCCGGATCCCCGGGC) / Bt-d(GCCCGGGGATCCGGCCGCG)
***dam*TC** d(CGCGGCCGGATCTCCCGGGC) / Bt-d(GCCCGGGG^{m}AGATCCGGCCGCG)
L20_4G d(GATCGTAGATGTCGCATCGA) / Bt-d(CGATGCG^{m}ACATCTACGATC)
L20_5C d(GATCGTAGATACCGCATCGA) / Bt-d(CGATGCGGTATCTACGATC)
L20_6T d(GATCGTAGATATTGCATCGA) / Bt-d(CGATGCA^{m}ATATCTACGATC)
L20_CT d(GATCGTAGATACTGCATCGA) / Bt-d(CGATGCAGTATCTACGATC)

(Bt, Biotin; ^{m}A, 6-methyladenine)

Note that *dam*TC contains a *dam* site followed by a TC dinucleotide, that corresponds to the specificity of M.*Eco*RV at the fifth and sixth position of the recognition sequence. Thus, this substrate contains only one deviation from the canonical *Eco*RV sequence and could also be designated "4C". The *dam*-substrate, in principle, can be methylated in both strands, whereas one strand of all other substrates is already methylated so that only one strand is available for M.*Eco*RV as substrate. Due to this difference the rate of methylation of the *dam* substrate could be up to 2-fold higher than one would observe with a hemimethylated *dam* substrate. This uncertainty does not affect the interpretation of our data as presented in this paper, because we only discuss effects much larger than a factor of 2.

Under the conditions of our experiments, biphasic kinetics in which a fast exponential phase is followed by a slower almost linear phase were observed with fully active variants and good substrates. To derive an initial slope, the data were fitted to eq 1 which describes a single turnover reaction that is followed by a slower linear phase. This form of the reaction progress curve is observed although the concentration of the enzyme is larger than the concentration of the substrate, because under the experimental conditions M.*Eco*RV is not saturated with AdoMet resulting in the formation of many nonproductive enzyme−DNA complexes (*35*).

$$\mathrm{CPM}_{\mathrm{theo}}(t) = f(1 - e^{-k_1 t}) + k_2 t \qquad (1)$$

with k_1, rate constant of the exponential phase of DNA methylation (s^{-1}); k_2, rate constant of DNA methylation during the linear phase (cpm/s); *f*, factor taking into account the fraction of enzyme−DNA complexes that react during the fast exponential phase and the specific activity of the AdoMet (cpm).

To obtain the initial rate constant of DNA methylation (k_{met}) in cpm/s, eq 1 was differentiated at $t = 0$ which results in eq 2. These values can easily converted into turnover numbers per enzyme molecule per second by considering the specific activity of the AdoMet, the volume of the reaction mixture transferred into each well of the microplate and the concentration of the enzyme.

$$k_{\mathrm{met}} = fk_1 + k_2 \qquad (2)$$

In many reactions with less active variants and/or modified substrates a clear linear initial phase without a burst was observed. Then, the initial rate constant of DNA methylation was derived by linear regression of the initial part of the reaction progress curve.

Multiple Sequence Alignments and Preparation of Phylogenetic Trees. Blast searches were carried out at NIH (http://www.ncbi.nlm.nih.gov/BLAST/) using the standard protein blast (*36*) or at Rebase (http://www.neb.com/rebase) (*37*). In a first step, the 10 sequences most related to M.*Eco*RV were retrieved. Subsequently, blast searches were performed with these sequences and always the most related enzymes were introduced into the list of sequences until a self-consistent list of 27 enzyme related to each other was obtained. Multiple sequence alignment was done using the ClustalX program (http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html) (*38*). Phylogenetic trees were constructed using the programs *protdist* and *drawgram* from the Phylip program package (http://evolution.genetics.washington.edu/phylip.html). For the construction of a phylogenetic tree, M.*Nla*III, an α-MTase that recognizes CATG sequences and does not contain an RNFP motif, was included as an outgroup.

RESULTS

The M.*Eco*RV DNA MTase recognizes the sequence G**A**TATC and modifies the first adenine residue within the site. The enzyme is closely related to MTases of the *dam*-family which recognize and modify adenine residues in G**A**TC sequences (Figure 1). It is the main issue of this work

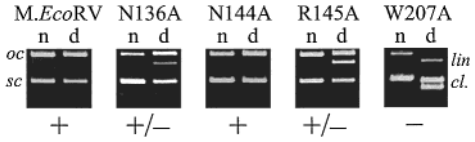

FIGURE 3: Examples of the results of the plasmid protection assays to test for in vivo activity of M.*Eco*RV and the N136A, N144A, R1454A, and W207A variants (n, native plasmid preparations; d, after digestion with R.*Eco*RV). The relative positions of the uncleaved and cleaved DNA in the gel are indicated (sc, superhelical DNA; oc, open circle DNA; lin, linear DNA; cl., cleavage products). The relative activities of the variants are given below the figure [(+) fully protected; (±) >50% protected; (−) <50% protected].

Table 1: In Vivo Activity of the Various M.*Eco*RV Variants and DNA Binding Constants of the Purified Proteins[a]

variant	in vivo activity[b]	DNA binding constant[c] K_{ass} (M^{-1})
M.*Eco*RV	+	1.2×10^7
K11A	+	7.5×10^6
S12A	+	n.a.
Q13A	+	5.7×10^6
K18A	+	7.4×10^6
R128A	−	$<10^5$
N130A	±	$<10^5$
K132A	+	1.2×10^7
N136A	±	1.1×10^7
C140A	+	$<10^5$
K141A	+	2.2×10^7
K142A	+	1.3×10^7
N144A	+	2.5×10^7
R145A	±	6.7×10^6
Q148A	+	2.2×10^7
Y203A	+	6.2×10^6
N205A	+	4.2×10^6
W207A	−	$<10^5$

[a] In vivo data were obtained from at least three independent clones. Binding constants are averages of at least three experiments with errors being smaller than ±30%. [b] (+) Fully protected; (±) >50% protected; (−) <50% protected. [c] na, not analyzed.

to investigate the mechanistic basis that makes M.*Eco*RV recognize a GATATC sequence despite a good homology with many enzymes that recognize GATC. To this end, we have exchanged 17 amino acid residues of M.*Eco*RV to alanine by site-directed mutagenesis and characterized the variants biochemically. In addition, we have investigated the accuracy of the recognition of the GATATC-positions by wild-type M.*Eco*RV and the variants using substrates that contain modified *Eco*RV sequences.

In Vivo DNA Methylation Activity of M.EcoRV and M.EcoRV Variants. Site-directed mutagenesis was carried out by PCR and verified by DNA sequencing. pGEXMRV plasmids encoding GST-M.*Eco*RV fusion proteins were prepared from ER1992 cells. The plasmid carries two GATATC sites which due to the basal expression of M.*Eco*RV become modified in the cell such that the plasmid is protected against R.*Eco*RV cleavage if it encodes an active variant. Therefore, the catalytic activity of the encoded M.*Eco*RV variant in vivo can be determined by a digestion of the pGEXMRV plasmid with R.*Eco*RV. Representative results are shown in Figure 3, all data are compiled in Table 1. As described previously, pGEXMRV encoding wild-type GST-M.*Eco*RV is completely protected against R.*Eco*RV cleavage indicating a complete methylation (*28*, *31*) similarly as observed with the His_6-M.*Eco*RV fusion protein (*28*, *39*). The same result was obtained with several of the variants

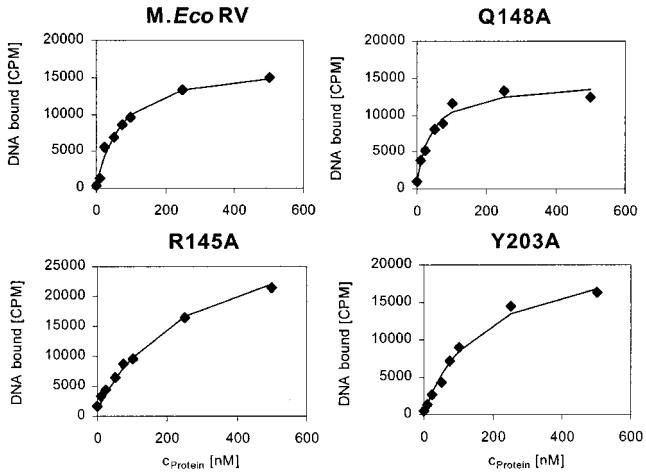

FIGURE 4: Examples of the DNA binding analyses of wild-type M.*Eco*RV and the R145A, Q148A, and Y203A variants. DNA binding was analyzed by nitrocellulose filter binding experiments using a radioactively labeled oligonucleotide. The lines show fits of the data to a bimolecular binding equilibrium. The binding constants derived for all the variants are compiled in Table 1.

(K11A, S12A, Q13A, K18A, K132A, C140A, N144A, Q148A, Y203, and N205A). Other variants such as R128A, N130A, N136A, and R145A showed decreased methylation activities in vivo. W207A turned out to be almost inactive in vivo.

DNA Binding of M.EcoRV and the M.EcoRV Variants. Purification of GST-M.*Eco*RV wild type and the variants was carried out with GSH−Sepharose. All enzymes could be overexpressed and preparations with purities >95% and concentrations between 10 and 136 μM were obtained, with two exceptions namely the variants C140A and W207A that could not be prepared at a concentration higher than 5 μM. To compare the DNA binding constants of GST-M.*EcoRV* and its variants, we carried out nitrocellulose filter binding assays with a 40-mer oligonucleotide containing a canonical GATATC site. Examples of the binding isotherms are shown in Figure 4, the results are summarized in Table 1. Wild-type M.*Eco*RV has a binding constant (K_{ass}) of 1.2×10^7 M^{-1}. Several of the variants showed wild type-like binding properties (K132A, N136A, K141A, K142A, N144A, and Q148A). A second group of variants showed a slightly reduced binding constant (K11A, Q13A, K18A, R145A, Y203A, and N205A). With a third group of variants no binding could be detected (R128A, N130A, C140A, and W207A) indicating a strongly reduced affinity for DNA. This result confirms data obtained in an earlier study with R128A and N130A (*28*). It should be noticed that M.*Eco*RV does not bind specifically to DNA under these conditions (*33*) such that the results obtained in these experiments represent nonspecific DNA binding.

In Vitro DNA Methylation Activity of M.EcoRV with Different Oligonucleotide Substrates. Methylation kinetics using [^{3}H]AdoMet were carried out with the biotin−avidin microplate assay (*34*). To investigate the influence of different nucleobases within the 3′-part of the target sequence on the rate of DNA methylation by wild-type M.*Eco*RV, seven oligonucleotides were employed as substrates for the methylation reactions. These substrates contain a GATATC site or sites differing in one, two, or three positions from

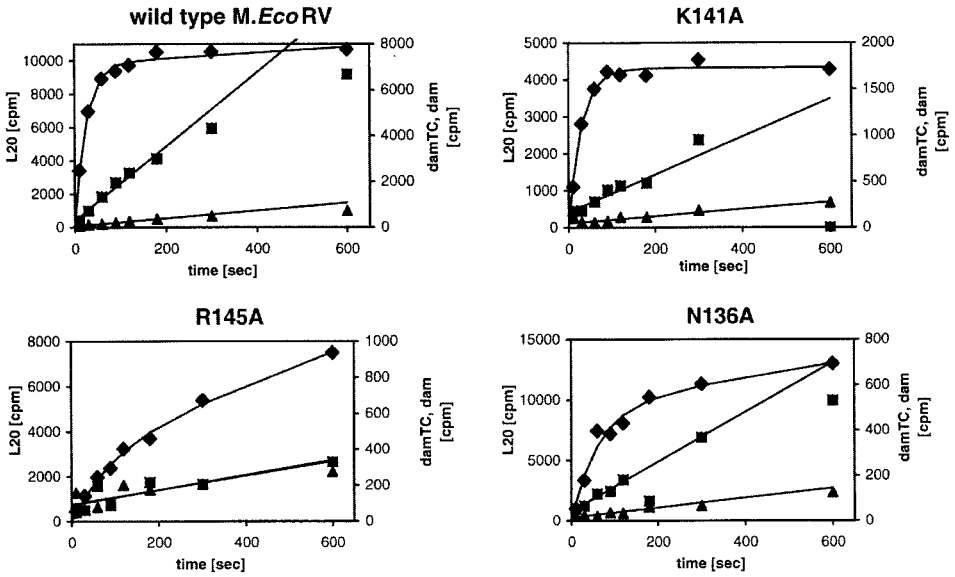

FIGURE 5: Examples of the DNA methylation experiments with wild-type M.*Eco*RV and the M.*Eco*RV variants N136A, K141A, and R145A using different oligonucleotide substrates (L20, diamonds; *dam*TC, squares; *dam*, triangles). The curves for L20 always refer to the left ordinate, those for *dam*TC and *dam* refer to the right ordinate. K141A behaves similar to wild-type M.*Eco*RV. N136A shows an increased discrimination between L20 and the two GATC substrates. R145A does not discriminate between *dam*TC and *dam*. The results obtained with all variants are compiled in Figure 6.

Table 2: Relative Rates of Methylation of Various Substrates by M.*Eco*RV[a]

substrate	sequence	rel. activity (%)
L20	GATATC/G^{m}ATATC	100
dam	GATCCC/GATCCG	0.4
*dam*TC	GATCTC/G^{m}AGATC	4.1
L20_4G	GATGTC/G^{m}ACATC	2.9
L20_5C	GATACC/GGTATC	0.7
L20_6T	GATATT/A^{m}ATATC	2.2
L20_CT	GATACT/AGATATC	0.6

[a] All experiments were carried out at least in triplicate. Relative rates were reproducible within ±20%. The rate of methylation of the L20 substrate was k_{met} = 400 ± 100 cpm/sec which corresponds to 0.1 turnover per min per enzyme molecule.

GATATC. Examples of the results obtained are shown in Figure 5, all data are summarized in Table 2.

The results show that M.*Eco*RV modifies all the substrates, even the *dam*- and L20_CT-substrates in which two bases within the 3′-part of the GATATC sequence are exchanged. Our data confirm the in vivo observation that M.*Eco*RV also modifies *dam*-sites (*26*). If only position 4 or position 6 of the GATATC sequence is exchanged, the catalytic rates decrease 20−40-fold. In contrast, by changing the position 5, activity is reduced almost 200-fold. This result demonstrates that the fifth base of the target sequence (GATATC) plays a more important role in DNA specificity of M.*Eco*RV than the fourth or sixth base.

A comparison of the rates of methylation of *dam*TC (GATCTC) and L20_4G (GATGTC) shows, that M.*Eco*RV does not strongly prefer C as second best base after A at the fourth position of the recognition sequence. This result has two implications: first *dam* sites are not recognized by M.*Eco*RV as GATC sites but most likely as GAT sites. Therefore, the residual activity observed at *dam* sites most likely reflects a general activity at sequences containing GAT trinucleotides. Second, if M.*Eco*RV was derived from a *dam* MTase, the recognition of the cytosine in the GAT*C* sequence has been completely removed.

In Vitro DNA Methylation Activity of M.EcoRV Variants. We used three of the oligonucleotide substrates with different target sites to characterize DNA recognition by the M.*Eco*RV variants: L20 has the canonical GATATC site, "*dam*" carries a GATC site and "*dam*TC" contains a GATCTC sequence. Examples of the results are shown in Figure 5, the data are summarized in Figure 6. In general, the in vitro catalytic activities are in good agreement with the in vivo results. Slight deviations might be due to differences in the in expression levels or stabilities of the variants in vivo. All the variants behaved differently in the in vitro kinetics; however, 16 of the variants can be divided into four groups with respect to their catalytic properties and specificities.

(1) Group I: S12A, K141A, K142A, N144A, and Q148A. This group comprises variants, which show an almost wild type-like behavior with the three substrates L20, *dam*, and *dam*TC. All of the variants show wild-type like DNA binding affinities.

(2) Group II: N130A, K132A, C140A, and N205A. The second group contains variants that have a 10−50-fold reduced catalytic activity, but show a similar specificity as wild-type M.*Eco*RV, because the rates of methylation of *dam* and *dam*TC are reduced 10−100-fold with respect to L20 and *dam*TC is significantly preferred over *dam*, although the preference for *dam*TC over *dam* is reduced when compared with wild-type M.*Eco*RV.

(3) Group III: K11A, Q13A, K18A, R145A, and Y203A. The third group comprises very interesting variants that cannot distinguish between *dam* and *dam*TC indicating that in the context of a GATC-site the fifth and sixth position of the recognition sequence cannot be identified by these variants. The high number of variants identified in this study that display a strongly altered DNA recognition behavior confirms the presumption that the loops 1, 2, and 3 are involved in DNA recognition by M.*Eco*RV. All the variants also show a 10−100-fold reduced catalytic activity. Furthermore, all of them prefer L20 approximately 100-fold over the *dam* substrate indicating that the discrimination between

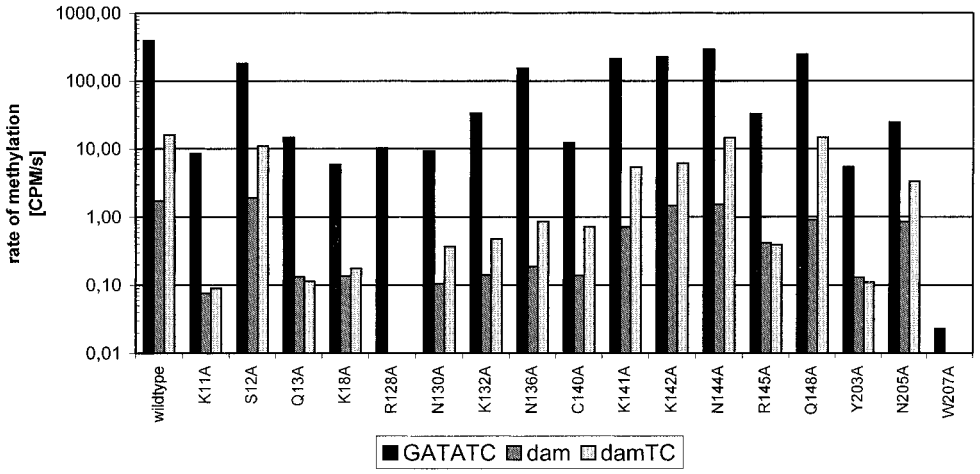

FIGURE 6: Compilation of the rates of DNA methylation of the L20, *dam*TC, and *dam* substrates by all variants investigated in this study. Note that activities are displayed in logarithmic scale. The rate of methylation of the L20 substrate by wild-type GST-M.*Eco*RV (k_{met} = 400 ± 100 cpm/s) corresponds to 0.1 turnover/min/enzyme molecule. All rates are averages obtained from at least three independent experiments with different enzyme preparations. Absolute rates were reproducible within a factor of 2, the errors of ratios of rate constants are below ±30%.

L20 and *dam* is similar as in wild-type M.*Eco*RV. Different from the wild-type enzyme, however, the *dam*TC substrate is modified at the same (slow) rate as the *dam* substrate.

(4) Group IV: R128A and N136A. The catalytic rate of the R128A variant is reduced by 2 orders of magnitude at GATATC sites, but methylation of *dam* sites (either *dam* or *dam*TC) was not detectable, showing that this variant displays an significant increase in specificity. A similar observation was made with the N136A variant that also shows a significantly increased discrimination against *dam* sites although the overall rates of methylation are much less reduced than in the case of R128A.

(5) The W207A variant displays a strongly reduced catalytic activity even with the L20 substrate and is almost inactive in vivo.

DISCUSSION

DNA Recognition by M.EcoRV. To study the accuracy of the recognition of the 3′-part of the *Eco*RV site (GATATC) by the *Eco*RV DNA-(adenine-N6)-methyltransferase, we have carried out methylation experiments using substrates that contain degenerated *Eco*RV recognition sites. The data show that like M.*Fok*I (*40*) and M.*Eco*RI (*41*), M.*Eco*RV has a much lower specificity than restriction endonucleases, because all substrates were modified with a readily detectable rate, a result that is completely dissimilar from what would have been found with a typical restriction enzyme (see, for review, refs *42* and *43*). Two DNA substrates differed from the canonical GATATC site in more than one bp (*dam*, GATCCC/GATCCG; L20_CT, GATACT). The results obtained with these substrates can be compared with those obtained with substrates, where only one position was altered. If different positions of the target site were recognized independently of each other, one would expect that the individual changes of $\Delta G^{\#}$ resulting from the removal of some contacts between the enzyme and the DNA should be additive. This would result in a loss of activity with a substrate containing two changed bases that is the multiplied effect of the loss of activity of the corresponding substrates in which only one base is exchanged.

The *dam* substrate (GATCCC) can be considered as a combination of the *dam*TC (GATCTC) and L20_5C (GATACC) substrates which are modified at 24 and 143-fold reduced rates with respect to the canonical substrate. Thus, theoretically the *dam* substrate should be methylated at a 3400-fold reduced rate which corresponds to 0.03% relative activity if both positions were recognized completely independent of each other. We observe, however, a relative activity of 0.4% which is more than 10-fold higher than one would expect for an independent recognition of the fourth and fifth positions of the *Eco*RV sequence and quite close to the rate of methylation of the L20_5C substrate (0.7%) in which only the fifth base is altered. The same observation was made with the L20_CT substrate which can be considered as a combination of L20_5C and L20_6T: here a relative rate of methylation of 0.02% would be expected much lower than the rate of 0.6% that is observed. Both results suggest that if the fifth position of the GATATC sequence is exchanged, recognition of the fourth and sixth positions is almost impossible. We conclude that the ATC part of the *Eco*RV sequence is recognized in a cooperative manner where the removal of one contact impedes or even prevents the proper formation of all other sequence specific contacts. In this respect M.*Eco*RV behaves similar as M.*Eco*RI that also shows nonadditive effects upon removal of more than one contacts between the enzyme and the DNA (*44*).

The cooperative nature of the interaction of M.*Eco*RV with the DNA is also detectable with some of the M.*Eco*RV variants. Although wild-type M.*Eco*RV significantly prefers the *dam*TC over the *dam* substrate, many of the variants are not able to discriminate between these two substrates indicating that the recognition of the fifth and sixth base pair of the *Eco*RV sequence is not possible if an A → C exchange is present at position four. The mechanistic basis of this cooperativity might be some conformational changes of the enzyme−DNA complex that accompany specific complex formation, because if a structural perturbation (caused by an exchange either of a bp in the recognition sequence or of an amino acid in the protein) interferes with a conformation that is characteristic for the specific enzyme-DNA complex,

the formation of all other specific contacts in the vicinity is prevented.

Group I Variants. Group I (S12A, K141A, K142A, N144A, and Q148A) comprises variants, which show an almost wild-type-like behavior in DNA binding and DNA methylation. None of these residues is highly conserved and according to our results none of them has an important function in DNA recognition or catalysis.

Group II Variants. Group II (N130A, K132A, C140A, and N205A) contains variants that have a significantly reduced catalytic activity, but do not display a change in specificity. We conclude, that these residues have an important role (directly or indirectly) for catalysis. Three of the residues show a reduced DNA binding affinity which supports the model that they are close to the enzyme−DNA interface.

Group III Variants. The variants of group III (K11A, Q13A, K18A, and R145A) do not discriminate between *dam* and *dam*TC suggesting that the mutations interfere with recognition of the last two positions of the GATATC sequence. The rate of methylation of the canonical substrate by all the variants is reduced by about 2 orders of magnitude similar to rates observed for wild-type M.*Eco*RV with target sites that contain an exchange of one base pair in the ATC part of the recognition sequence. Therefore, these variants interact with the canonical substrate with a similar efficiency as the wild-type enzyme with a *dam* substrate. In all the cases, the relative rates of methylation of the GATC substrates are approximately 100-fold lower than the rate of methylation of L20 indicating that the loss of discrimination between *dam* and *dam*TC is due to a relative reduction of the rate of methylation of the *dam*TC substrate which brings it to the level of *dam*. This result indicates that the recognition of the GATATC adenine works properly in all the variants, because L20 and *dam*TC only differ at this position. However, binding of a GATC site leads to two structural perturbation: the mutation in the protein that interferes with recognition of the fifth and sixth position of the target site and the cytosine that is present at the fourth position of the target sequence where an adenine should be situated. These alterations together prevent any recognition of the fifth and sixth base pairs of the *Eco*RV sequence.

On the basis of these data, loop 1 appears very important for the recognition of the TC-part of the GATATC sequence, because three variants in this region belong to group III. Although our data do not allow to define molecular interactions between M.*Eco*RV and the recognition sequence in details, some conclusions can be made:

(1) Lys11 is conserved between M.*Eco*RV and the *dam* MTases. Therefore, this residue is unlikely to be responsible for base contacts to the fifth and sixth position of the *Eco*RV sequence. However, one could imagine that it forms a backbone contact that stabilizes the conformation of loop 1 and/or helps to position the DNA for recognition.

(2) Gln13 and Arg145 are not conserved between M.*Eco*RV and *dam* MTases. Moreover, glutamine and arginine residues are among the amino acid types that most often are involved in specific recognition of DNA. Therefore, Gln13 and Arg145 are our prime candidates for formation of contacts to the bases of the DNA. The importance of Arg145 is further emphasized by the finding that this residue is surrounded by amino acid residues that are not important at all for DNA recognition and catalysis, because this observation make an indirect effect of Arg145 quite unlikely.

(3) Lys18 is also not conserved between M.*Eco*RV and *dam*-MTases. However, lysine residues typically are responsible for backbone contacts to the DNA and not for base contacts so we speculate that the function of Lys18, like Lys11, might be formation of a backbone contact to the DNA that supports recognition.

Group IV Variants. Two variants showed an increased level of discrimination between *Eco*RV and *dam* sites (R128A and N136A). Arg128 is present in all members of the *Eco*RV/*dam* family and N136 is conserved. Both these residues are located in the conserved part of the RNFP-motif, suggesting an interaction with the GAT-part of the recognition sequence. The finding that the variants show an increased level of discrimination can be explained if M.*Eco*RV has two independent modules for the recognition of 5′ and 3′ parts of the *Eco*RV sequence. Both these modules interact with the DNA in a cooperative fashion. If wild-type M.*Eco*RV binds a *dam* substrate, the interaction of the ATC-module is disturbed by the nucleotide exchanges at the 3′ end of the DNA sequence. However, the GAT-module can still interact with the GAT-part of the recognition sequence such that the catalytic activity is reduced only 10−100-fold. If the R128A or N130A variants bind the canonical *Eco*RV sequence, the interaction of the GAT-module with the DNA is disturbed by the amino acid exchange, but the ATC-module can interact with the 3′-part of the target site again leading to a moderate decrease in activity. If, however, the variants interact with a *dam* sequence, both recognition modules are disturbed which leads to a very large loss of catalytic activity.

Role of Loop 3. Three residues of loop 3 were investigated in this study. W207A was among the variants with the most severe phenotype in the whole study. It is almost inactive in vivo and shows a strongly reduced DNA binding affinity and catalytic activity in vitro. Both the Y203A and N205A variants also show a significant reduced catalytic activity. The Y203A variant belongs to group III, because it does not discriminate between *dam* and *dam*TC. Since loop 3 does not harbor residues that differ between M.*Eco*RV and *dam* MTases, a direct contact of Tyr203 or Trp207 to the GATATC bases is not very probable. Therefore, a more indirect role of loop 3 in positioning or bending the DNA and thereby allowing the specific contacts to form is more likely. M.*Eco*RV bends DNA in the specific complex significantly by 60° (*45*) which is correlated with unstacking of the DNA helix at the GATATC position as shown by a strong increase of the fluorescence of substrates containing 2-aminopurine at the fourth position of the GATATC sequence (*35*). One could imagine that one of the conserved aromatic residues of loop 3 may intercalate into the DNA helix, thereby causing DNA bending and the strong increase of the fluorescence observed with GAT(2-aminourine)TC substrates.

Molecular Evolution of M.EcoRV. So far over 700 different DNA MTases have been sequenced which recognize and methylate almost 300 different DNA sequences (see http://www.neb.com/rebase, ref *37*). Thus, DNA methyltransferases comprise a very large group of related enzymes and provide an enormous amount of information to study molecular evolution. In particular, they exemplify how

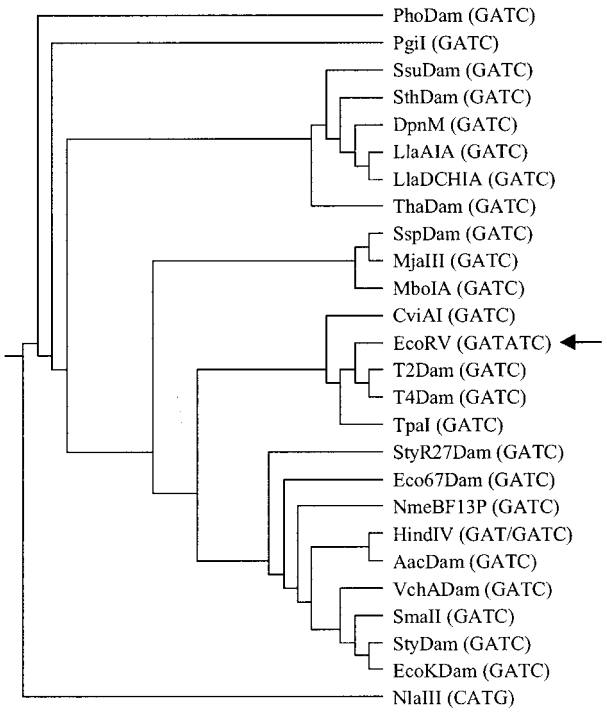

FIGURE 7: Phylogenetic tree of M.*Eco*RV and related *dam*-MTases. The recognition sequences of the different enzymes are given in parentheses, in each case the adenine residue at the second position of the recognition site is modified. The specificity of M.*Hin*dIV has been reported to be GAT (*46*), according to an anonymous reviewer of this paper it is GATC. M.*Nla*III was introduced as an outgroup.

natural evolution has generated many different specificities from one (or few) ancestral proteins and by which pathways the sequence specificity of DNA interacting enzymes has changed during molecular evolution. Previously, we have shown that one domain of the M.*Fok*I DNA MTase has a low specificity and could be considered as an evolutionary intermediate of an enzyme that changes its recognition sequence (*40*). Here, we investigate the *Eco*RV MTase that recognizes G**A**TATC sequences and modifies the first adenine residue within this site. Parts of the DNA interface of this enzyme show high sequence homology to DNA MTases of the *dam* family which recognize and modify G**A**TC sequences. A phylogenetic analysis of this group of enzymes shows that M.*Eco*RV is a clear member of the *dam*-MTase family with all close relatives being *dam*-MTases (Figure 7). This result suggests that M.*Eco*RV has been derived from a *dam*-MTase by changing the recognition sequence from GATC to GATATC.

Our biochemical results demonstrate that M.*Eco*RV was generated in molecular evolution by retaining the recognition module of *dam* MTases that is responsible for the interaction with the GAT sequence and addition of a new recognition module for the recognition of the 3′ part of the recognition site. Vestiges of the evolutionary origin of M.*Eco*RV are still visible, for example the ability of the enzyme to modify also GATC sequences. M.*Eco*RV is not a perfect enzyme, because our data show that the discrimination between GATATC and GATC can be increased by an exchange of a single amino acid residue without a strong loss of catalytic activity (N136A). These data together with the results obtained with M.*Fok*I (*40*) suggest that the evolutionary pathway for changing one specificity to another goes through a stage of relaxed specificity.

REFERENCES

1. Bestor, T. H., and Verdine, G. L. (1994) *Curr. Opin. Cell Biol. 6*, 380−389.
2. Cheng, X. (1995) *Curr. Opin. Struct. Biol. 5*, 4−10.
3. Cheng, X. (1995) *Annu. Rev. Biophys. Biomol. Struct. 24*, 293−318.
4. Cheng, X., and Blumenthal, R. M. (1999) *S-adenosylmethionine-dependent methyltransferases: Structures and functions*, World Scientific Publishing, Singapore.
5. Robertson, K. D., and Wolffe, A. P. (2000) *Nat. Rev. Genet. 1*, 11−19.
6. Malone, T., Blumenthal, R. M., and Cheng, X. (1995) *J. Mol. Biol. 253*, 618−632.
7. Wilson, G. G. (1992) *Methods Enzymol. 216*, 259−279.
8. Jeltsch, A., Christ, F., Fatemi, M., and Roth, M. (1999) *J. Biol. Chem. 274*, 19538−19544.
9. Jeltsch, A. (2001) *Biol. Chem. 382*, 707−710.
10. Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994) *Cell 76*, 357−369.
11. Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J., and Weinhold, E. (2001) *Nat. Struct. Biol. 8*, 121−125.
12. Jeltsch, A., Roth, M., and Friedrich, T. (1999) *J. Mol. Biol. 285*, 1121−1130.
13. Roberts, R. J., and Cheng, X. (1998) *Annu. Rev. Biochem. 67*, 181−198.
14. Messer, W., and Noyer-Weidner, M. (1988) *Cell 54*, 735−737.
15. Noyer-Weidner, M., and Trautner, T. A. (1993) in *DNA Methylation: Molecular Biology and Biological Significance* (Jost, J. P., and Saluz, H. P., Eds.) pp 40−108, Birkhauser Verlag, Basel.
16. Dryden, D. T. F. (1999) in *S-Adenosylmethionine-dependent Methyltransferases: Structures and Functions* (Cheng, X., and Blumenthal, R. M., Eds.) pp 283−340, World Scientific Publishing, Singapore.
17. Hale, W. B., van der Woude, M. W., Braaten, B. A., and Low, D. A. (1998) *Mol. Genet. Metab. 65*, 191−6.
18. Heithoff, D. M., Sinsheimer, R. L., Low, D. A., and Mahan, M. J. (1999) *Science 284*, 967−70.
19. Enserink, M. (1999) *Science 284*, 883.
20. Garcia-Del Portillo, F., Pucciarelli, M. G., and Casadesus, J. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11578−11583.
21. Bucci, C., Lavitola, A., Salvatore, P., Del Giudice, L., Massardo, D. R., Bruni, C. B., and Alifano, P. (1999) *Mol. Cell 3*, 435−445.
22. Bestor, T. H. (2000) *Hum. Mol. Genet. 9*, 2395−2402.
23. Nwosu, V. U., Connolly, B. A., Halford, S. E., and Garnett, J. (1988) *Nucleic Acids Res. 16*, 3705−3720.
24. Bougueleret, L., Schwarzstein, M., Tsugita, A., and Zabeau, M. (1984) *Nucleic Acids Res. 12*, 3659−3677.
25. Jeltsch, A. (1999) *J. Mol. Evol. 49*, 161−164.
26. Taylor, J. D., Goodall, A. J., Vermote, C. L., and Halford, S. E. (1990) *Biochemistry 29*, 10727−10733.
27. Guschlbauer, W. (1988) *Gene 74*, 211−214.
28. Roth, M., Helm-Kruse, S., Friedrich, T., and Jeltsch, A. (1998) *J. Biol. Chem. 273*, 17333−17342.
29. Tran, P. H., Korszun, Z. R., Cerritelli, S., Springhorn, S. S., and Lacks, S. A. (1998) *Structure 6*, 1563−1575.
30. Roth, M., and Jeltsch, A. (2001) *Nucleic Acids Res. 29*, 3137−3144.
31. Friedrich, T., Roth, M., Helm-Kruse, S., and Jeltsch, A. (1998) *Biol. Chem. 379*, 475−480.
32. Kirsch, R. D., and Joly, E. (1998) *Nucleic Acids Res. 26*, 1848−50.
33. Jeltsch, A., Friedrich, T., and Roth, M. (1998) *J. Mol. Biol. 275*, 747−758.
34. Roth, M., and Jeltsch, A. (2000) *Biol. Chem. 381*, 269−272.
35. Gowher, H., and Jeltsch, A. (2000) *J. Mol. Biol. 303*, 93−110.
36. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.

37. Roberts, R. J., and Macelis, D. (2001) *Nucleic Acids Res. 29*, 268−269.
38. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res. 25*, 4876−82.
39. Jeltsch, A., Sobotta, T., and Pingoud, A. (1996) *Protein Eng. 9*, 413−423.
40. Friedrich, T., Fatemi, M., Gowhar, H., Leismann, O., and Jeltsch, A. (2000) *Biochim. Biophys. Acta 1480*, 145−159.
41. Reich, N. O., Olsen, C., Osti, F., and Murphy, J. (1992) *J. Biol. Chem. 267*, 15802−15807.
42. Pingoud, A., and Jeltsch, A. (1997) *Eur. J. Biochem. 246*, 1−22.
43. Pingoud, A., and Jeltsch, A. (2001) *Nucleic Acids Res.* (in press).
44. Reich, N. O., and Danzitz, M. J. (1991) *Nucleic Acids Res. 19*, 6587−6594.
45. Cal, S., and Connolly, B. A. (1996) *J. Biol. Chem. 271*, 1008−1015.
46. Roy, P. H., and Smith, H. O. (1973) *J. Mol. Biol. 81*, 445−459.

BI0155450