Probing the DNA Interface of the *Eco*RV DNA-(Adenine-N6)-methyltransferase by Site-Directed Mutagenesis, Fluorescence Spectroscopy, and UV Cross-Linking†

Carsten Beck and Albert Jeltsch*

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

Received April 17, 2002; Revised Manuscript Received October 11, 2002

ABSTRACT: The *Eco*RV DNA-(adenine-N6)-methyltransferase recognizes GATATC sites and methylates the DNA as indicated. It is related to the large family of dam methyltransferases which modify GATC sites. We have studied the interaction of DNA with M.*Eco*RV and 12 M.*Eco*RV variants using oligonucleotides containing 2-aminopurine as a fluorescence probe in equilibrium and stopped-flow DNA binding studies and 5-iododeoxyuracil for UV cross-linking. M.*Eco*RV binds to DNA in a multistep binding reaction, including two different conformations of the specific enzyme−DNA complex, and induces a strong conformational change of the DNA at the fourth position of the recognition site. Mutations at residues forming contacts to the GAT part of the recognition site reduce the stability of both specific enzyme−DNA complexes. Two enzyme variants which fail to recognize the ATC part do not induce the deformation of the DNA which explains why they cannot interact properly with the recognition site. Other mutations at residues which interact with the ATC part selectively reduce the stability of the second enzyme−DNA complex. These results show that when approaching the DNA M.*Eco*RV first contacts the GAT part of the target site. Since the residues mediating these contacts are conserved among M.*Eco*RV and dam MTases, the kinetics of formation of the enzyme−DNA complex correspond to the evolutionary history of the protein. Whether the observation that evolutionarily conserved contacts are formed early during complex formation is a general rule for DNA interacting enzymes or proteins that change their specificity during evolution remains to be seen.

In prokaryote genomes, methylation of DNA occurs at adenine-N6, cytosine-N4, and cytosine-C5. DNA methylation is involved in postreplicative DNA mismatch repair, control of DNA replication and cell cycle, regulation of gene expression, and protection of bacteria against bacteriophages (see refs *1*−*4* for reviews). DNA methyltransferases (MTases)[1] can be grouped into N- and C-MTases according to the target atom for methylation. The group of N-MTases comprises adenine-N6 and cytosine-N4 MTases, which have similar active sites and show overlapping specificities (*5*, *6*). All DNA MTases consist of two domains (see refs *3* and *4* for reviews). The structure of the large domain of these enzymes is conserved. It harbors the binding sites for the cofactor *S*-adenosyl-L-methionine (AdoMet) and for the target base, which is flipped out of the DNA helix before methyl group transfer. DNA MTases are one of the largest groups of related enzymes, and they thereby represent an enormous reservoir of information for studying DNA recognition by proteins and changes thereof during molecular evolution: more than 2000 different restriction/modification systems are known, and 700 different DNA MTases have been sequenced which recognize and methylate almost 300 different DNA sequences (*46*).

Unfortunately, so far only three structures of DNA MTases in complex with DNA are known, two cytosine-C5 MTases and one adenine-N6 MTase (*7*−*9*). In addition, the dynamics of DNA recognition and base flipping cannot be understood on the basis of structural studies alone. Therefore, we followed a biochemical approach to investigate the mechanism of DNA recognition of the *Eco*RV DNA-(adenine-N6)-MTase from *Escherichia coli* which recognizes the hexanucleotide sequence GATATC and methylates the first adenine residue within this site (*10*, *11*). According to the sequence and order of conserved motifs, the enzyme can be classified as an α-N-MTase (*1*, *12*, *13*). M.*Eco*RV is closely related to dam MTases which recognize GATC sites, and therefore, it is an interesting example of how molecular evolution has changed the specificity of DNA recognition of a DNA-interacting enzyme (*1*, *14*−*17*). In a previous work, we have shown that amino acid residues from three loops of the enzyme contribute to DNA recognition by M.*Eco*RV (Figure 1) which is based on two functional modules, one that recognizes the GAT part of the recognition sequence and one that interacts with the ATC part of the GATATC site (*17*). The first recognition module comprises amino acid residues that are conserved among M.*Eco*RV and dam MTases which are mainly located within a conserved RNFP motif which forms a part of loop II. Substitutions of

† This work has been supported by the Deutsche Forschungsgemeinschaft (Grant 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.

[1] Abbreviations: AdoMet, *S*-adenosyl-L-methionine; AdoHcy, *S*-adenosyl-L-homocysteine; ^{5I}U, 5-iododeoxyuracil; 2AP, 2-aminopurine; MTase, DNA methyltransferase.

10.1021/bi025979a CCC: $22.00
Published on Web 11/09/2002

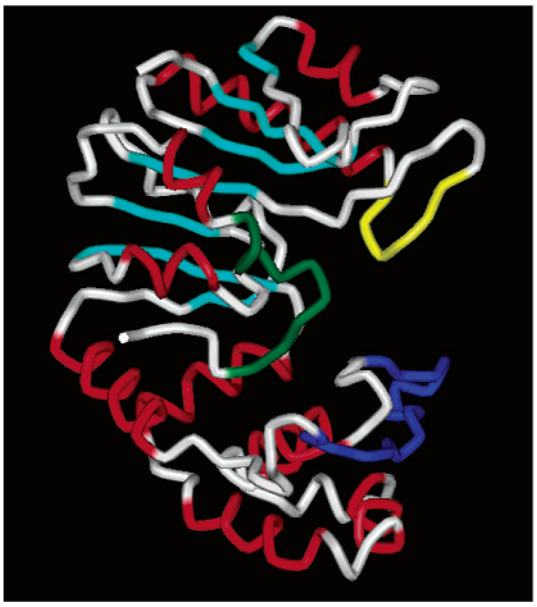

FIGURE 1: Structural model of M.*Eco*RV. The structure of M.*Eco*RV was modeled using *Dpn*M (PDB entry 2DPM) as a template (*17*). α-Helical regions are colored red, and sheets are colored cyan. The DNA binding cleft of M.*Eco*RV is created by three loops (loop 1, Ile10−Leu19; loop 2, Met126−Arg145; and loop 3, Tyr203−Gly208) colored green, blue, and yellow, respectively.

these residues destroy enzyme activity, or if less important residues are affected, they lead to enzymes that are more specific than wild-type M.*Eco*RV, because they are no longer able to methylate dam sites (*15*, *17*) which is possible for the wild-type enzyme (*17*, *18*). The second recognition module comprises nonconserved residues from all three loops. These residues directly or indirectly mediate contacts to the second part of the GATATC site, and mutations at these positions lead to a weakened or even completely lost ability to recognize the fifth and sixth positions of the M.*Eco*RV site (*17*). The specificities of DNA methylation determined with 15 M.*Eco*RV variants carrying single-amino acid exchanges were used as a basis for the study presented here (*17*). These mutants could be divided into four groups, two of which showed significantly altered preferences for noncanonical sites.

R128A and N136A display a significant increase in specificity, because the relative rates of methylation at dam sites were reduced for these variants. This result suggests that these residues are involved in recognition of the first part of the *Eco*RV site.

K11A, Q13A, K18A, R145A, and Y203A cannot distinguish between GATC and GATCTC sites, indicating that the recognition of the fifth and sixth positions of the target sequence is affected by the mutations. Recognition of the forth position of the GATATC site is still possible.

Here, we intended to probe the structure of the DNA interface of M.*Eco*RV in more detail by employing protein−DNA photo-cross-linking and equilibrium as well as stopped-flow fluorescence DNA binding studies. The focus of our efforts has been the recognition of the second half of the recognition sequence since in this part the discrimination between GATATC and GATC sites is achieved.

EXPERIMENTAL PROCEDURES

Oligonucleotides and Proteins. Purified oligonucleotides were obtained from Interactiva (Ulm) and MWG (Eberberg) (for the sequences of the substrates that were used, see Table 1). Complementary oligonucleotides were annealed by heating to 70 °C and cooling to ambient temperature over the course of 30−45 min. M.*Eco*RV and M.*Eco*RV variants were purified as GST-tagged proteins as described previously (*17*). Protein preparations were at least 90% pure.

UV Cross-Linking Studies. The CL-A4, CL-T5, CL-C6, and CL-N+1 oligodeoxynucleotides were used for the UV cross-linking studies (Table 1). The oligonucleotides were labeled with [γ-^{32}P]ATP (Amersham) and T4-PNK (MBI) in one strand prior to annealing. DNA (1 μM), enzyme (1 μM), and *S*-adenosyl-L-homocysteine (100 μM) (AdoHcy, Sigma) were irradiated for 15 min with UV light at 325 nm using a Cd−He laser (OMI-3074-40 M, Omnichrome) in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, and 0.1 mM DTT at ambient temperature. Samples were heated to 95 °C for 2 min and analyzed on denaturing urea gels.

Steady State 2-Aminopurine Fluorescence Studies. The AP-T3, AP-A4, AP-T5, and AP-C6 oligodeoxynucleotides were used for the 2AP fluorescence studies (Table 1). DNA (0.5 μM), enzyme (1 μM), and *S*-adenosyl-L-homocysteine (100 μM) were incubated in 50 mM HEPES (pH 7.5), 50 mM NaCl, and 1 mM EDTA at ambient temperature for 5 min. The fluorescence was determined using a Hitachi F4500 spectrofluorometer. The excitation wavelength was 313 nm (slit, 2.5 nm). The fluorescence was measured between 320 and 440 nm (slit, 5 nm). For quantitative analysis, the fluorescence intensity was integrated between 360 and 390 nm. Reference data were obtained with each oligonucleotide in the absence of protein.

Stopped-Flow 2-Aminopurine Fluorescence Studies. Stopped-flow 2AP fluorescence studies were carried out as described previously (*19*). AP-A4 and enzyme were mixed at equal concentrations (0.125, 0.25, and 0.5 μM) in a stopped-flow device (SFM-3, Bio-Logic, Claix, France) in buffer [50 mM HEPES (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 100 μM AdoHcy]. The fluorescence was excited at 313 nm and detected using a 340 nm cutoff filter. Data were collected between 0.05 and 30 s. Each experiment was repeated 15−25 times, and the results were averaged and fitted to Scheme 1 by nonlinear least-squares fit and numerical integration as described previously (*20*).

RESULTS

UV Cross-Linking Studies. The base analogue 5-iododeoxyuracil (^{5I}U) can be photoactivated with a Cd−He laser at 325 nm which leads to a homolytic cleavage of the C−I bond. The uracil radical cross-links with high yields to several amino acid residues, including Phe, Tyr, Trp, Met, and His (*21*, *22*). The yields of the cross-linking reaction are significantly enhanced if a $\pi-\pi$ stacking interaction exists between the base analogue and an aromatic amino acid residue (*23*), suggesting that cross-linking studies with oligonucleotide substrates in which the target adenine base is replaced with ^{5I}U could be a convenient assay for studying base flipping by DNA MTases. To test this model, we had investigated the cross-linking yields of four substrates containing one ^{5I}U at positions −1 and 1−3 of the *Eco*RV

Table 1: Oligodeoxynucleotide Substrates Used in This Study

substrate	sequence
CL-A4	5′-GATCGTA**GAT-^{5I}U-TC**GCATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′
CL-T5	5′-GATCGTA**GATA-^{5I}U-C**GCATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′
CL-C6	5′-GATCGTA**GATAT-^{5I}U-**GCATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′
CL-N+1	5′-GATCGTA**GATATC-**^{5I}U-CATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′
AP-T3	5′-GATCGTA**GA-2AP-ATC**GCATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′
AP-A4	5′-GATCGTA**GAT-2AP-TC**GCATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′
AP-T5	5′-GATCGTA**GATA-2AP-C**GCATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′
AP-C6	5′-GATCGTA**GATAT-2AP-**GCATCGA-3′/5′-TCGATGC**G^{m}ATATC**TACGATC-3′

Scheme 1

$$E + D \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E10' \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} ED''$$

sequence ($N^{-1}G^{1}A^{2}T^{3}A^{4}T^{5}C^{6}N^{+1}$). In agreement with the expectation, we have shown that the target base (A^{2}) by far shows the highest cross-linking yield (*24*). High cross-linking yields with ^{5I}U located at the position of the target base also have been observed with other DNA MTases, and ^{5I}U cross-linking has been employed to map the binding pocket for the flipped target base of different MTases (*22*, *25*, *26*).

Here we have extended these studies and investigated the photo-cross-linking yields of substrates containing single ^{5I}U bases at positions 4−6 and +1 of the N^{-1} $G^{1}A^{2}T^{3}A^{4}T^{5}C^{6}N^{+1}$ recognition sequence of the M.*Eco*RV MTase. All these substrates are bound by M.*Eco*RV with similar affinities (data not shown). Representative cross-linking data are shown in Figure 2A, and the results obtained with wild-type M.*Eco*RV and five variants are compiled in Figure 2B. The most important result of these experiments is that a ^{5I}U at the position of the second adenine shows a very high cross-linking yield. Additional experiments demonstrated that this yield is even higher than that with a ^{5I}U situated at the position of the target base (data not shown). Since position 4 is not methylated by M.*Eco*RV (*11*, *19*), it is very unlikely that this base is also flipped out of the DNA helix. Therefore, a high cross-linking yield with ^{5I}U does not necessarily correlate to base flipping. This result is similar to our finding that substrates carrying a 2-aminopurine (2AP) at this position also show a large increase in fluorescence upon binding to M.*Eco*RV (*24*), although this effect has been taken as evidence for base flipping before. It argues that a high cross-linking yield with ^{5I}U (like an increase in the fluorescence of 2AP) cannot be taken as absolute evidence of base flipping. The high cross-linking yield suggests that a close contact between an aromatic amino acid residue of M.*Eco*RV and the second adenine residue in the GATATC sequence is formed in the specific complex.

Similar experiments were carried out with all M.*Eco*RV variants which were active in DNA binding and exhibited interesting properties in the previous work (Q13A, K18A, N136A, R145A, and Y203A). As shown in Figure 2, most of the variants exhibited a pattern of cross-linking yields very similar to that of the wild-type enzyme with one striking exception: the N136A variant did not cross-link with good yields to the substrate that carries a ^{5I}U at the fourth position of the recognition sequence which had the best yield with wild-type M.*Eco*RV and all other variants. A straightforward interpretation of this result could be that a contact between the enzyme and this base is not present in the variant. However, this model is very unlikely to be true, because the

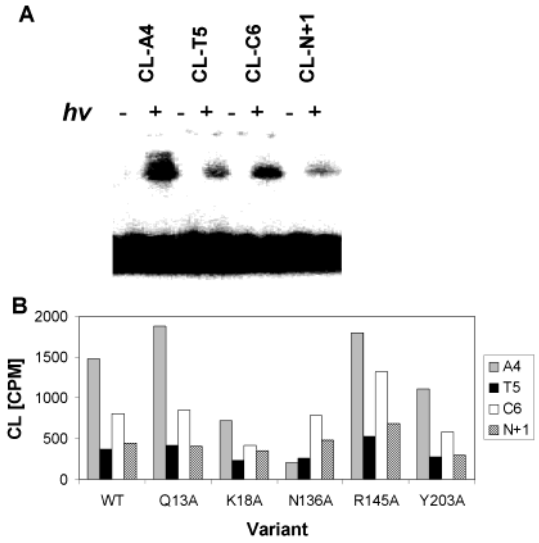

FIGURE 2: UV protein−DNA cross-linking analysis using M.*Eco*RV and M.*Eco*RV variants. In these experiments, oligonucleotides carrying ^{5I}U residues at different positions were used (A4, CL-A4, GAT^{5I}UTC; T5, CL-T5, GATA^{5I}UC; C6, CL-C6, GATAT^{5I}U; and N+1, CL-N+1, GATATC^{5I}U). Cross-linking was carried out using 1 μM DNA, 1 μM enzyme, and 100 μM AdoHcy in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, and 0.1 mM DTT. In panel A, some representative data are shown, and in panel B, the compilation of at least three independent experiments is displayed. The relative cross-linking yield could be reproduced within ±30%.

N136A variant is significantly more specific than the wild-type enzyme (*17*), a result that is not in agreement with the assumption that important contacts are lost. We, therefore, favor an alternative explanation for this result. The substrate used for the cross-linking experiments carries a ^{5I}U (that is very close to T in terms of molecular recognition) at the position of the A4. Consequently, it is a noncanonical substrate. Since the N136A variant shows an increased selectivity, it might discriminate against the GAT^{5I}UAT substrate with high accuracy. Then, it could not form the specific complex, and the cross-link that is characteristic of the conformation of the specific complex cannot be formed.

Equilibrium DNA Binding Studies with Oligonucleotide Substrates Containing 2AP. DNA binding was investigated by fluorescence experiments carried out using oligodeoxynucleotide substrates which contain 2-aminopurine substitutions at positions 4−6 and +1 of the *Eco*RV target sequence $N^{-1}G^{1}A^{2}T^{3}A^{4}T^{5}C^{6}N^{+1}$. This base analogue is fluorescent in solution, but when it is located in the DNA helix, stacking interactions with neighboring base pairs strongly quench the fluorescence (*27*). However, the 2AP fluorescence increases after proteins like DNA MTases (*28*−*35*), RNA polymerases (*36*), DNA polymerases (*37*), helicases (*38*), glycosylases (*39*−*41*), DNA photolyase (*42*), or deaminases (*43*) bind to

Table 2: Increase in the Fluorescence of Substrates Carrying a Single 2AP Group at Different Positions of the GATATC Site upon Binding to M.*Eco*RV[a]

substrate	recognition sequence	$F_{M.EcoRV}/F°$
AP-T3	GA-2AP-ATC	1.3
AP-A4	GAT-2AP-TC	8.8
AP-T5	GATA-2AP-C	1.3
AP-C6	GATAT-2AP	1.4

[a] In these experiments, the 2AP fluorescence was determined using 0.5 μM solutions of the oligonucleotides in the absence ($F°$) or presence of 1 μM M.*Eco*RV ($F_{M.EcoRV}$). Integration of the fluorescence intensity was carried out between 360 and 390 nm.

the DNA or RNA. In general, changes in 2AP fluorescence could be correlated with local unstacking caused by conformational changes of the nucleic acid, like partial melting, bending, or unwinding. In the case of some DNA MTases, a strong increase in fluorescence is observed if the target base for methylation is replaced with 2AP. It has been suggested that this effect is caused by the flipping of the base analogue out of the DNA helix (*28*−*30*). We have shown previously that a strong increase in fluorescence was observed after binding of M.*Eco*RV to the AP-A4 substrate in which 2AP replaces the second adenine residue in the center of the *Eco*RV recognition sequence (*19*). No detectable fluorescence change was observed in similar experiments with substrates carrying the 2AP base analogue at the position of the target base and also at the position 5′ to the recognition sequence. To complement these results, we have investigated here whether substrates containing the 2AP probe in the second part of the recognition sequence exhibit a fluorescence change during M.*Eco*RV binding. However, also in this set of experiments, only AP-A4 produces a strong 2AP signal (Table 2), although all modified substrates were bound by M.*Eco*RV with a similar affinity (data not shown). This result suggests that a significant conformational change in the DNA that is not base flipping takes place at the fourth position of the recognition sequence.

Next we used the AP-A4 substrate to study the DNA interaction of 12 variants of M.*Eco*RV carrying mutations in the DNA binding interface (Figure 3). According to the AP-A4 signal, the mutants could be divided into three groups. K18A, K132A, K141A, N144A, Q148A, and Y203A exhibited a signal similar to that of the wild-type enzyme. Most of these variants exhibited wild-type-like behavior also in their kinetic characterization, confirming that 2AP fluorescence is a tool that is well suited to probing the conformation of the specific M.*Eco*RV−DNA complex. K11A, N130A, N136A, and K142A exhibit a slightly reduced signal which we do not interpret to be significant. However, a strongly reduced fluorescence signal was observed with the Q13A, R128A, C140A, and R145A variants. Of these variants, R128A and C140A showed a strongly reduced DNA binding affinity (*17*) such that the reduced 2AP fluorescence signal is not diagnostic. However, the Q13A and R145A variants which both have a strongly reduced 2AP fluorescence signal do not show a reduced DNA binding affinity. These variants are members of the group with the most significantly altered DNA recognition specificity, because both of them show strongly reduced capabilities in the recognition of the fifth and sixth bases of the GATATC sequence (*17*). The absence of a strong 2AP signal with AP-A4 shown here suggests

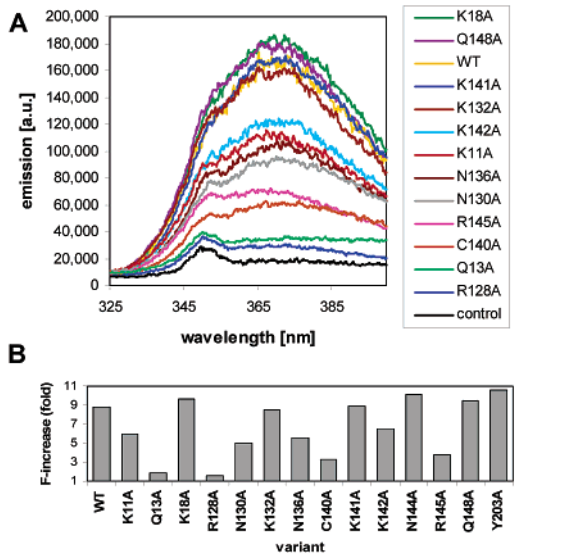

FIGURE 3: Equilibrium DNA binding studies using the increase in fluorescence of AP-A4 after binding to M.*Eco*RV and M.*Eco*RV variants. In each experiment, 0.5 μM AP-A4 and 1 μM enzyme were used in 50 mM HEPES (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 100 μM AdoHcy. In panel A, representative data of binding of AP-A4 to different M.*Eco*RV variants are shown (control denotes a sample with AP-A4 but without protein under identical conditions). Panel B shows a compilation of the relative changes of the fluorescence signal averaged between 360 and 380 nm from at least two independent experiments. Relative fluorescence changes could be reproduced within ±20%.

that the conformation of the DNA in complex with these variants is dramatically different from the conformation of the DNA bound to wild-type M.*Eco*RV in the second half of the recognition sequence. This conformational difference explains why these variants cannot properly recognize the bases in this part of the recognition sequence. However, not all the variants which do not properly recognize the second part of the M.*Eco*RV sequence exhibit a reduced 2AP signal with AP-A4. For example, K11A, K18A, and Y203A showed a similar loss of recognition of the fifth and sixth base pairs of the *Eco*RV site like Q13A and R145A but did not display a reduced AP-A4 signal. This is not surprising since more subtle changes of the complex structure could also interfere with base recognition.

Kinetics of DNA Binding. To investigate the formation of the specific M.*Eco*RV−DNA complex in more detail, we have used the AP-A4 substrate to study DNA binding by M.*Eco*RV in a stopped-flow device. Binding kinetics were carried out using M.*Eco*RV and DNA at concentrations between 0.12 and 0.5 μM. All data obtained here could be fitted to the two-step binding model described by Scheme 1; fits to an easier one-step model were not possible (data not shown). In Scheme 1, two enzyme−DNA complexes are considered which both have different fluorescence signals: ED′ and ED″, different conformations of the specific enzyme−DNA complex.

In a previous work, we have carried out similar experiments with wild-type M.*Eco*RV and showed that a three-step mechanism was better suited to fitting the data (*19*). In this model, the enzyme first binds to the DNA in a nonspecific binding mode that does not lead to a fluorescence signal. Later, two specific complexes are formed that both have different fluorescence properties. However, in that work,

Table 3: Transient Kinetics of Binding of the AP-A4 Substrate to Various M.*Eco*RV Mutants[a]

	k_1 (M^{-1} s^{-1})	k_{-1} (s^{-1})	K_1 (M^{-1})	k_2 (s^{-1})	k_{-2} (s^{-1})	K_2
wild-type M.*Eco*RV	$(7.0 \pm 3.0) \times 10^6$	0.18 ± 0.06	$(3.9 \pm 1.0) \times 10^7$	0.31 ± 0.14	nd[b]	>100
K11A	$(1.3 \pm 0.4) \times 10^7$	0.24 ± 0.10	$(5.4 \pm 2.0) \times 10^7$	0.41 ± 0.19	nd[b]	>100
K132A	$(9.1 \pm 3.0) \times 10^6$	0.16 ± 0.05	$(5.8 \pm 1.5) \times 10^7$	0.34 ± 0.15	nd[b]	>100
N136A	$(3.3 \pm 2.0) \times 10^7$	170 ± 60	$(1.9 \pm 0.4) \times 10^5$	1.2 ± 0.4	1.5 ± 0.5	1.8 ± 0.6
R145A	$(2.0 \pm 1.5) \times 10^7$	0.09 ± 0.06	$(2.8 \pm 2.5) \times 10^8$	$(5.4 \pm 4.1) \times 10^{-4}$	0.16 ± 0.08	$(3.4 \pm 1.5) \times 10^{-3}$
Y203A	$(1.4 \pm 0.6) \times 10^7$	0.01 ± 0.005	$(1.0 \pm 0.8) \times 10^9$	0.058 ± 0.025	0.07 ± 0.03	0.07 ± 0.04

[a] Results of the stopped-flow experiments carried out to investigate binding of M.*Eco*RV and M.*Eco*RV variants to AP-A4. Data were fitted to a two-step binding model (Scheme 1). [b] Not detectable.

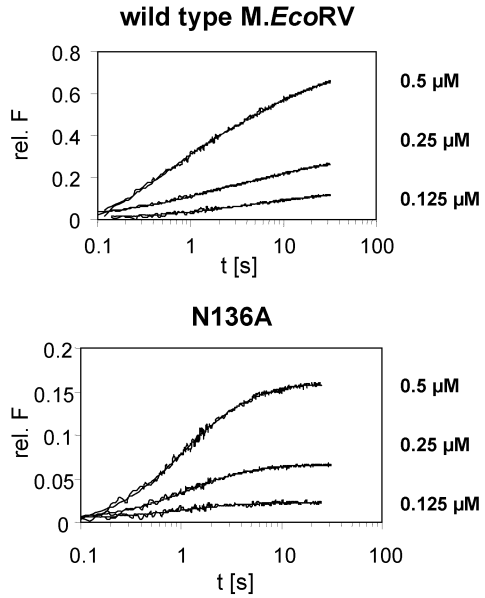

FIGURE 4: Examples of transient kinetics of DNA binding of M.*Eco*RV and M.*Eco*RV variants to AP-A4. In each experiment, 0.5, 0.25, and 0.125 μM AP-A4 and enzyme were mixed in a stopped-flow device in 50 mM HEPES (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 100 μM AdoHcy and the fluorescence was determined over the course of 30 s. In each case, the scattered curve shows the experimental data and the smooth curve the fit of the data to Scheme 1.

we used concentrations of enzyme and substrate of up to 700 nM and also analyzed the kinetics of dissociation of the enzyme−DNA complex, both of which were not possible with the mutant proteins. This difference in the experimental setup explains why in this work only two of the three steps could be resolved. Since both ED complexes in Scheme 1 exhibit a fluorescence signal, they must correspond to the specific complexes observed in the earlier work. Therefore, in the current study, the initial nonspecific DNA binding step is not detected. To compare the data obtained with the mutants to data for the wild-type enzyme, we also carried out experiments with wild-type M.*Eco*RV under exactly the same conditions as with the variants. The rate constants obtained from the experiments with the wild-type enzyme as well as K11A, K132A, N136A, R145A, and Y203A (the Q13A variant could not be investigated because its fluorescence signal was too weak) are summarized in Table 3.

The rate constants obtained with wild-type M.*Eco*RV (Figure 4 and Table 3) confirm our earlier conclusion that DNA binding is a multistep process in which two specific complexes are in a slow equilibrium (*19*, *44*). Under the conditions used here, formation of the second complex was practically irreversible in the time scale of our experiments such that we could not assign a rate constant to the dissociation of the second enzyme−DNA complex. This observation is in perfect agreement with earlier findings where a very slow dissociation of the specific complex was observed (*19*, *44*). The K132A variant was investigated in the stopped-flow experiments, because it behaved in a manner very similar to that of the wild-type enzyme in all previous experiments (see above and ref *17*). As shown in Table 3, it also did not show any significant difference in the kinetics of AP-A4 binding. The same is true for the K11A variant, although it cannot properly recognize the second part of the recognition sequence. However, the kinetics of the DNA interaction of the N136A, R145A, and Y203A variants were dramatically different from those of wild-type M.*Eco*RV.

The stability of both DNA complexes of the N136A is strongly reduced (Figure 4 and Table 3) which is due to a reduced rate of the first complex formation (k_1) in combination with an increased rate of the first complex dissociation (k_{-1}) and a strongly increased rate of the second complex dissociation (k_{-2}). This result demonstrates that the exchange of Asn136 with Ala weakens contacts between the enzyme and the DNA that are formed in early stages of complex formation and persist in later phases.

In contrast to those of the N136A variant, the rate of formation and the stability of the first enzyme−DNA complex were not reduced with the R145A and Y203A variants; in fact, the stability of the first enzyme−DNA complex of these variants is slightly increased (Table 3). However, in both the cases, the rates of dissociation of the second complex were strongly enhanced, with rate constants in the range of 0.1 s^{-1} which are at least 100-fold faster that with the wild-type enzyme and the wild-type-like variants, where this reaction was not detectable at all. The increase in the k_{-2} values of R145A and Y203A leads to a significantly reduced stability of the second complex. This result indicates that the contacts between the enzyme and the DNA that are affected in the R145A and Y203A variants are only important in later stages of specific complex formation.

DISCUSSION

In this study, we extend results from earlier experiments where we have described 15 variants of M.*Eco*RV which carry single-amino acid exchanges in the DNA binding loops of M.*Eco*RV and investigated the specificity of DNA methylation by these variants (*17*). In that paper, we demonstrated that two of these variants (R128A and N136A) show an increased specificity in the recognition of the fourth position of the GATATC site. Five variants showed a strongly reduced ability to recognize the fifth and sixth bases of the GATATC site (K11A, Q13A, K18A, R145A, and Y203A) (*17*).

Here, we have studied the interaction of the M.*Eco*RV DNA methyltransferase with the bases in the second half of the GATATC sequence by performing cross-linking studies with ^{5I}U and fluorescence studies with 2AP. Both data sets agree in the finding that a very strong signal was observed with substrates carrying the probe at the fourth position of the recognition sequence GATXTC. However, even transient flipping of this base into the target base binding pocket is very unlikely, because methylation of this base is not detectable, whereas methylation of a cytosine residue located at the target position of M.*Eco*RV was (*5*, *19*). In addition, complete melting of the whole target region, including A2 and A4, is also unlikely. Therefore, neither high cross-linking reactivity nor a strong 2AP fluorescence signal is an unequivocal proof of base flipping. The 2AP result suggests that this base is significantly unstacked in the specific complex. The ^{5I}U signal suggests that an aromatic amino acid residue of the enzyme closely approaches the nucleotide residue at this position. M.*Eco*RV is known to bend the DNA by 60° in the specific complex (*45*). Since our data show that one of the biggest conformational changes of the DNA occurs at A4, it is likely that the center of DNA bending is at this position. The results obtained with the ^{5I}U and 2AP substrates suggest that an aromatic amino acid residue might intercalate into the DNA next to A4. Therefore, the most likely scenario that explains our data is that DNA bending and intercalation of an aromatic amino acid residue into the DNA next to A4 occur, which leads to an overall unstacking at A4.

In the previous study, we have identified two interesting groups of M.*Eco*RV variants. One group (R128A and N136A) comprises variants that are more specific than the wild-type enzyme, because they show a weakened ability to modify GATC sites. An increase in the specificity of the N136A variant was also confirmed in this work, since this variant shows much better discrimination against substrates carrying a ^{5I}U residue at position 4 of the recognition sequence than the wild-type enzyme. These results can be interpreted if the exchange of Arg128 or Asn136 with alanine affects contacts of the enzyme with the GAT part of the *Eco*RV sequence (GATATC). Then, contacts with the second part of the recognition sequence become relatively more important for inducing specific complex formation with all the conformational changes required to activate the catalytic center. This model explains the increase in the level of discrimination against GATC and GAT^{5I}UTC, because with the altered substrates some of the contacts with the ATC part of the *Eco*RV site are also lost. It is supported by the observation that R128 and N136 are conserved among M.*Eco*RV and dam MTases which recognize GATC sites. The R128A variant shows a strongly reduced DNA binding affinity and, therefore, was not suited to the experiments carried out in this work.

A second group of variants identified in the previous work contained amino acid exchanges in all three loops of M.*Eco*RV that create the DNA binding cleft of the enzyme (K11A, Q13A, K18A, R145A, and Y203A) (*17*). These variants show a strongly reduced ability to recognize the fifth and sixth base pairs of the *Eco*RV site (GATATC). Here, we show that the structure of the specific M.*Eco*RV−DNA complex of two of these variants (Q13A and R145A) differs from wild-type M.*Eco*RV, because the fluorescence signal that accompanies DNA binding and bending of GAT-2AP-TC substrates is almost completely lost (Q13A) or significantly reduced (R145A) in these variants. This finding is strong evidence that these two residues are directly involved in contacts with the bases of the ATC part of the *Eco*RV sequence. In the case of the R145A variant, the residual fluorescence signal was sufficient for an analysis of the kinetics of DNA binding by this variant. M.*Eco*RV binds to DNA in a multistep binding reaction, including at least one nonspecific and two specific enzyme−DNA complexes. The kinetic analysis of the R145A variant showed that the stability of the second enzyme−DNA complex is markedly reduced when compared with that of wild-type M.*Eco*RV. A similar result was observed with the Y203A variant, indicating also that alterations in the structure of the third DNA interaction loop can influence contacts with the ATC part of the *Eco*RV site. After all, we had just two variants (K11A and K18A) that exhibited considerably altered specificity of DNA recognition in the previous work, but did not show any significant alteration of the conformation and rate of formation of the specific DNA complex. We conclude that these lysine residues most likely do not interact with the bases of the recognition sequence but rather with the phosphodiester backbone. Such interactions with the phosphate groups are very important for processes such as base flipping and DNA bending, and the amino acid residues might serve to couple the conformational changes of the DNA to the activation of the catalytic center. This model would explain why pronounced changes in specificity are observed, although the energetics and kinetics of DNA interaction are not significantly altered.

The transient kinetics of DNA binding performed in this work confirm earlier studies that DNA binding by M.*Eco*RV is a multistep process where at least one nonspecific and at least two specific complexes are formed which are in equilibrium with each other. The stability of both specific enzyme−DNA complexes of the N136A variant which interacts with the GAT part of the *Eco*RV site is strongly reduced. In contrast, the first complex formation is not affected by exchanges of residues that mediate interactions to the second part of the recognition sequence (R145A and Y203A). This result demonstrates that during specific complex formation, M.*Eco*RV interacts with the GAT part of the recognition sequence first. Multiple sequence alignments show that M.*Eco*RV has evolved from a GATC recognizing dam MTase and that the contacts to the GAT parts of the recognition sequences are conserved between M.*Eco*RV and dam MTases (*1*, *14*, *15*, *17*). According to the data presented here, these evolutionarily conserved interactions are the first contacts formed during DNA recognition by M.*Eco*RV. It is interesting to note that when the precursor MTase changed its DNA recognition specificity from GATC to GATATC, the contacts with the GAT part of the site that are formed early are maintained. In addition, new contacts with the ATC part are created that mediate the increased specificity. However, these new contacts form only in later stages of complex formation. Therefore, it appears that the order of events in enzyme−DNA complex formation reflects the evolutionary history of the enzyme. It will be interesting to see if this is a general mechanism of molecular evolution.

Finally, one should comment on the surprising observation that a couple of protein variants with dramatically altered DNA recognition specificity show a very similar "fingerprint" of relative cross-linking yields that is comparable to that of the wild-type enzyme. For example, why do variants such as Q13A, K18A, R145A, and Y203A which are not able to recognize the fifth and sixth positions of the *Eco*RV site still form cross-links with substrates carrying a ^{5}IU at this position, and even more so, why are the relative yields of all these cross-links so comparable? One solution to this discrepancy could be that the cross-link only occurs in a very special and rarely populated conformation of the enzyme−DNA complex that is of no relevance for the biochemical properties. Since the cross-linking reaction is irreversible, during the time of incubation a detectable amount of cross-link can form even if the fractional occupation of the reactive complex conformation is rather low. This hypothetical conformation could be quite different in structure from the transition state of enzymatic catalysis. Then, the results of the cross-linking studies are not highly relevant for the analysis of enzyme catalysis or the mechanism of DNA interaction. This observation argues for caution in the interpretation of cross-linking studies in investigating the structure and function of macromolecular interactions. This problem does not exist with the fluorescence studies, because here all fluorophores are always observed and not just a very small subset that is in a very special conformation, and no irreversible conversions take place that could bias the observed signal.

CONCLUSIONS

M.*Eco*RV recognizes GATATC sites. It has evolved from dam MTases which recognize GATC sites. We show here that contacts with the first part of the recognition sequence of M.*Eco*RV, which are evolutionarily old, are formed in an early phase of recognition complex formation, whereas the contacts with the second part which discriminate between GATATC and GATC are formed later. Whether the observation that the order of events in complex formation corresponds to the evolutionary origin of the enzyme is a general rule for DNA interacting enzymes or proteins that change their specificity during evolution remains to be seen. We show that M.*Eco*RV induces a strong conformational change at the fourth position of the recognition sequence that is likely to include unstacking, intercalation of aromatic side chains into the DNA, and DNA bending. This distortion of the DNA is a prerequisite for the recognition of the second half of the target sequence. We have also demonstrated that DNA bound to two M.*Eco*RV variants that do not properly recognize the fifth and sixth positions of the *Eco*RV sequence (Q13A and R145A) does not show strong conformational changes. This result suggests that these residues are involved in sequence specific contacts with the DNA. Other variants interfere with DNA recognition in a more subtle manner and only change the relative energetic levels of different conformations of the specific complex, thereby favoring or disfavoring the catalytically competent form. Overall, this leads to a reduced or increased specificity of DNA recognition.

ACKNOWLEDGMENT

Thanks are due to M. Roth and A. Pingoud for helpful discussions.

REFERENCES

1. Malone, T., Blumenthal, R. M., and Cheng, X. (1995) *J. Mol. Biol. 253*, 618−632.
2. 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.
3. Cheng, X., and Roberts, R. J. (2001) *Nucleic Acids Res. 29*, 3784−3795.
4. Jeltsch, A. (2002) *ChemBioChem 3*, 274−293.
5. Jeltsch, A., Christ, F., Fatemi, M., and Roth, M. (1999) *J. Biol. Chem. 274*, 19538−19544.
6. Jeltsch, A. (2001) *Biol. Chem. 382*, 707−710.
7. Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994) *Cell 76*, 357−369.
8. Reinisch, K. M., Chen, L., Verdine, G. L., and Lipscomb, W. N. (1995) *Cell 82*, 143−153.
9. Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J., and Weinhold, E. (2001) *Nat. Struct. Biol. 8*, 121−125.
10. Bougueleret, L., Schwarzstein, M., Tsugita, A., and Zabeau, M. (1984) *Nucleic Acids Res. 12*, 3659−3677.
11. Nwosu, V. U., Connolly, B. A., Halford, S. E., and Garnett, J. (1988) *Nucleic Acids Res. 16*, 3705−3720.
12. Wilson, G. G. (1992) *Methods Enzymol. 216*, 259−279.
13. Jeltsch, A. (1999) *J. Mol. Evol. 49*, 161−164.
14. Lauster, R., Kriebardis, A., and Guschlbauer, W. (1987) *FEBS Lett. 220*, 167−176.
15. Roth, M., Helm-Kruse, S., Friedrich, T., and Jeltsch, A. (1998) *J. Biol. Chem. 273*, 17333−17342.
16. Tran, P. H., Korszun, Z. R., Cerritelli, S., Springhorn, S. S., and Lacks, S. A. (1998) *Structure 6*, 1563−1575.
17. Beck, C., Cranz, S., Solmaz, M., Roth, M., and Jeltsch, A. (2001) *Biochemistry 40*, 10956−10965.
18. Taylor, J. D., Goodall, A. J., Vermote, C. L., and Halford, S. E. (1990) *Biochemistry 29*, 10727−10733.
19. Gowher, H., and Jeltsch, A. (2000) *J. Mol. Biol. 303*, 93−110.
20. Pingoud, A., Urbanke, C., Hogget, J., and Jeltsch, A. (2002) in *Biochemical Methods*, Wiley-VCH, Weinheim, Germany.
21. Willis, M. C., Hicke, B. J., Uhlenbeck, O. C., Cech, T. R., and Koch, T. H. (1993) *Science 262*, 1255−1257.
22. Wong, D. L., Pavlovich, J. G., and Reich, N. O. (1998) *Nucleic Acids Res. 26*, 645−649.
23. Norris, C. L., Meisenheimer, P. L., and Koch, T. H. (1996) *J. Am. Chem. Soc. 118*, 5796−5803.
24. Jeltsch, A., Roth, M., and Friedrich, T. (1999) *J. Mol. Biol. 285*, 1121−1130.
25. Holz, B., Dank, N., Eickhoff, J. E., Lipps, G., Krauss, G., and Weinhold, E. (1999) *J. Biol. Chem. 274*, 15066−15072.
26. Wong, D. L., and Reich, N. O. (2000) *Biochemistry 39*, 15410−15417.
27. Ward, D. C., Reich, E., and Stryer, L. (1969) *J. Biol. Chem. 244*, 1228−1237.
28. Allan, B. W., and Reich, N. O. (1996) *Biochemistry 35*, 14757−14762.
29. Allan, B. W., Beechem, J. M., Lindstrom, W. M., and Reich, N. O. (1998) *J. Biol. Chem. 273*, 2368−2373.
30. Holz, B., Klimasauskas, S., Serva, S., and Weinhold, E. (1998) *Nucleic Acids Res. 26*, 1076−1083.
31. Reddy, Y. V., and Rao, D. N. (2000) *J. Mol. Biol. 298*, 597−610.
32. Szegedi, S. S., Reich, N. O., and Gumport, R. I. (2000) *Nucleic Acids Res. 28*, 3962−3971.
33. Allan, B. W., Reich, N. O., and Beechem, J. M. (1999) *Biochemistry 38*, 5308−5314.
34. Malygin, E. G., Zinoviev, V. V., Petrov, N. A., Evdokimov, A. A., Jen-Jacobson, L., Kossykh, V. G., and Hattman, S. (1999) *Nucleic Acids Res. 27*, 1135−1144.
35. Pues, H., Bleimling, N., Holz, B., Wolcke, J., and Weinhold, E. (1999) *Biochemistry 38*, 1426−1434.
36. Jia, Y., Kumar, A., and Patel, S. S. (1996) *J. Biol. Chem. 271*, 30451−30458.
37. Frey, M. W., Sowers, L. C., Millar, D. P., and Benkovic, S. J. (1995) *Biochemistry 34*, 9185−9192.
38. Raney, K. D., Sowers, L. C., Millar, D. P., and Benkovic, S. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6644−6648.
39. Stivers, J. T., Pankiewicz, K. W., and Watanabe, K. A. (1999) *Biochemistry 38*, 952−963.
40. Wong, I., Lundquist, A. J., Bernards, A. S., and Mosbaugh, D. W. (2002) *J. Biol. Chem. 277*, 19424−19432.

41. Bernards, A. S., Miller, J. K., Bao, K. K., and Wong, I. (2002) *J. Biol. Chem. 277*, 20960−20964.
42. Christine, K. S., MacFarlane, I. A., Yang, K., and Stanley, R. J. (2002) *J. Biol. Chem. 277*, 38339−38344.
43. Yi-Brunozzi, H. Y., Stephens, O. M., and Beal, P. A. (2001) *J. Biol. Chem. 276*, 37827−37833.
44. Jeltsch, A., Friedrich, T., and Roth, M. (1998) *J. Mol. Biol. 275*, 747−758.
45. Cal, S., and Connolly, B. A. (1996) *J. Biol. Chem. 271*, 1008−1015.
46. Roberts, R. J., Macelis, D. (2001) *Nucleic Acids Res. 29*, 286−289.

BI025979A