The EMBO Journal vol.15 no.11 pp.2870–2882, 1996

# Structural adaptations in the interaction of *Eco*RI endonuclease with methylated GAATTC sites

**Linda Jen-Jacobson[1], Lisa E.Engler, David R.Lesser[2], Michael R.Kurpiewski, Cynthia Yee and Brian McVerry**

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, USA

[2]Present address: US Navy Drug Screening Laboratory, Oakland, CA 94627, USA

[1]Corresponding author

**We have studied the interaction of *Eco*RI endonuclease with oligonucleotides containing GAATTC sites bearing one or two adenine-$N^6$-methyl groups, which would be in steric conflict with key protein side chains involved in recognition and/or catalysis in the canonical complex. Single-strand methylation of either adenine produces small penalties in binding free energy ($\Delta\Delta G^o_S$ ~ +1.4 kcal/mol), but elicits asymmetric structural adaptations in the complex, such that cleavage rate constants are strongly inhibited and unequal in the two DNA strands. The dependences of cleavage rate constants on the concentration of the $Mg^{2+}$ cofactor are unaltered. When either adenine is methylated on both DNA strands, $\Delta\Delta G^o_S$(~+4 kcal/mol) is larger than the expected sum of the $\Delta\Delta G^o_S$ values for the single-strand methylations, because the asymmetric adaptations cannot occur. Cleavage rate constants are reduced by 600 000-fold for the biologically relevant GA$^m$ATTC/CTT$^m$AAG site, but the G$^m$AATTC/CTTA$^m$AG site forms only a non-specific complex that cannot be cleaved. These observations provide a detailed thermodynamic and kinetic explanation of how single-strand and double-strand methylation protect against endonuclease cleavage *in vivo*. We propose that non-additive effects on binding and structural 'adaptations' are important in understanding how DNA methylation modulates the biological activities of non-catalytic DNA binding proteins.**

*Keywords*: DNA methylation/*Eco*RI endonuclease/protein–DNA interactions/restriction endonuclease/thermodynamics

## Introduction

DNA methylation is known to be involved in a wide variety of biological phenomena. In prokaryotes, DNA methylation controls (positively or negatively) site-specific DNA cleavage by restriction endonucleases (Raleigh and Wilson, 1986; McClelland *et al.*, 1994), is required for efficient initiation of DNA replication (Messer *et al.*, 1985; Smith *et al.*, 1985), is required for differentiation between template and daughter DNA strands in mismatch repair (Lahue *et al.*, 1989) and may be involved in the regulation of gene expression (Barras and Marinus, 1989; Braaten *et al.*, 1994). In eukaryotes, DNA methylation appears to be involved in maintaining X-chromosome inactivation (Riggs and Pfeifer, 1992; Migeon, 1994), in genomic imprinting (Chaillet, 1994), in modulating recombination of immunoglobulin genes (Hsieh and Lieber, 1992) and in controlling gene expression (Razin and Cedar, 1991; Bird, 1992; Bestor and Coxon, 1993). It is important to note that a causal role for DNA methylation has not been established conclusively for all these phenomena. In mice, a methyltransferase mutation that prevents DNA methylation at cytidines is embryonic lethal (Li *et al.*, 1992).

For those phenomena where DNA methylation has (or may ultimately be shown to have) a causal role, it seems likely that this role will involve the positive or negative modulation of protein interaction with one or more specific DNA sites. Even in cases where biological function involves a multiprotein complex bound to DNA, some direct DNA–protein interface must be affected. At present, there is a lack of detailed structural and thermodynamic information to shed light on the key questions of how DNA methylation influences site-specific protein–DNA interactions. What structural features of a DNA–protein interface might permit one or a few methyl groups in a DNA recognition site to exert a very large effect (positive or negative) on the ability of a protein to interact with that site? How (in both energetic and structural terms) does the position or clustering of DNA methyl groups affect protein interactions with methylated DNA sites?

It is reasonable to suppose that the answers to these questions will be based on the common structural themes found in many DNA–protein interfaces, including direct hydrogen bonding of protein to DNA bases and phosphates and elaborate 'networks' that interconnect the interacting elements, but will differ in detail for each protein. A well-characterized DNA binding protein, the *Eco*RI restriction endonuclease, provides an immediate opportunity to adduce some of the relevant structural and thermodynamic principles. This endonuclease and its corresponding *Eco*RI methyltransferase form the RI restriction–modification system. Expression of the endonuclease *in vivo* in the absence of the methyltransferase leads to lethality as the result of suicidal DNA cleavage (Kuhn *et al.*, 1986; Heitman *et al.*, 1989). The crystal structure of the endonuclease in complex with its DNA recognition site has been determined at 2.5 Å resolution (Kim *et al.*, 1990; Rosenberg, 1991), and extensive thermodynamic and kinetic studies (Lesser *et al.*, 1990, 1992, 1993; Jen-Jacobson, 1995) have led to a reasonably detailed understanding of how particular structural features of the complex relate to the efficacy and specificity of DNA binding and cleavage.

Here, we examine the interaction of *Eco*RI endonuclease

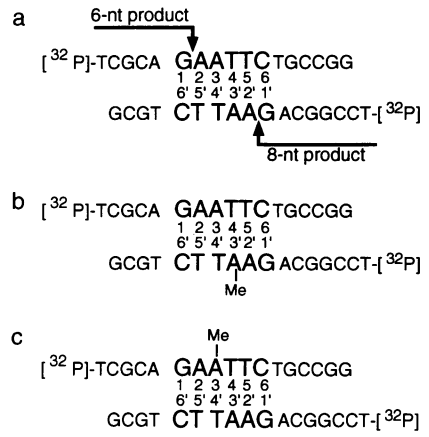



**Fig. 1.** Design of oligonucleotide substrates. Upper sequence (a) shows the canonical duplex 17 nucleotide oligomer. *Eco*RI endonuclease cleavage at GpA in the upper strand produces an end-labeled six nucleotide product; cleavage in the lower strand produces an eight nucleotide product (see Figure 3). In all experiments, each hemimethylated site was tested in both orientations, for example, $N^6$-$^m$A in positions 3′ (sequence b) and 3 (sequence c).

with GAATTC recognition sites modified with one or two $N^6$-methylated adenine ($^m$A) bases. We find that one methyl group, regardless of position, evokes only a modest penalty in binding free energy, but triggers an adaptation that alters protein contacts with DNA phosphates and preferentially inhibits cleavage in the methylated DNA strand. When two methyl groups are placed symmetrically in both DNA strands, the penalties in both binding free energy and transition state interaction free energy show dramatic non-additivity, being much greater than twice those for single methyl groups. The site methylated at both inner adenines (the biologically relevant form) can be cleaved, albeit very slowly, but the site methylated at both outer adenines permits the endonuclease to form only a non-specific, cleavage-incompetent complex. These observations can be understood by examining the interaction of adenine $N^6$-methyl groups with particular protein side chains that have known roles in binding and/or catalysis.

## Results

### *Methylation inhibits endonuclease binding*

The interface between *Eco*RI endonuclease and its DNA recognition site is extremely crowded, in that elements of both protein subunits interdigitate to form intimate contacts to every major-groove functional group (except the O4 atoms of the inner thymines) on both purine and pyrimidine bases (Rosenberg, 1991). Thus, it has been proposed (McClarin *et al.*, 1986; Frederick *et al.*, 1988) that the introduction of one or more bulky methyl groups on adenine amino groups would completely prohibit endonuclease binding. This prediction is not borne out by our data.

The oligonucleotide substrates used in these studies are based upon the duplex 17 nucleotide oligomer shown in Figure 1. This series retains the same flanking sequence used in our earlier thermodynamic and kinetic studies (Lesser *et al.*, 1990, 1992, 1993). In all experiments discussed below, each site was tested in both orientations (Figure 1).

Introduction of a single $N^6$-methyl group on any adenine has only a small effect (Table I) on equilibrium binding of *Eco*RI endonuclease ($\Delta\Delta G^o_S \sim +1.4$ kcal/mol). Based upon comparison with previous work (Lesser *et al.*, 1990, 1993) in which we systematically deleted each of the major-groove functional groups, this is approximately the penalty to be expected for the loss of one protein–base interaction, which can be assessed directly only for G$^7$CAATTC and GA$^7$CATTC sites. The $\Delta\Delta G^o_S$ value (–1.0 to +1.7 kcal/mol) for deletion of any other contacted functional group includes additional contributions from favorable or unfavorable effects on DNA conformation and/or structural adaptations in the complex (Lesser *et al.*, 1993). Note in particular that it is not valid to compare sites with adenine-$N^6$-methyl groups with those in which the $N^6$-amino group is deleted from position 2 (GPATTC site, where P = purine) or position 3 (GAPTTC site), since deletion of either adenine amino group also affects the energy required to distort the DNA (Lesser *et al.*, 1993). As a result, the net $\Delta\Delta G^o_S$ (–1 kcal/mol) for GAPTTC shows that binding is more favorable than for the unmodified site, despite the unfavorable deletion of a protein–base hydrogen bond (Lesser *et al.*, 1993).

We show below that the observed penalties for methylated sites (Table I) are not simply attributable to blocking hydrogen bonding between adenine-$N^6$ and Asn141 of the endonuclease. Note that the penalty in binding free energy for these hemimethylated sites is independent of the position of the $N^6$-methyl group in the recognition site. By contrast, binding is not affected (Table I) by methylation of an adenine just outside the recognition site, in a position where the endonuclease makes no protein–base contacts.

Symmetrical methylation of two adenines (i.e. 2,2′-di-$^m$A or 3,3′-di-$^m$A) in a site produces binding penalties (Table I) significantly larger than would be expected from the sum of the separate penalties for the two corresponding single-strand methylations. The observed equilibrium binding constants ($K_A$) approach those observed for 'non-specific' endonuclease–DNA binding (Lesser *et al.*, 1990), and we show below that the 2,2′-di-$^m$A site in fact forms only a 'non-specific' complex which does not permit cleavage.

### *Adaptations in the methylated site complexes*

We have shown previously (Becker *et al.*, 1988; Lesser *et al.*, 1990) that structural adaptations in endonuclease complexes with non-canonical DNA sites can be detected by ethylation-interference footprinting, which measures how single ethylphosphotriester groups affect binding.

In the canonical *Eco*RI endonuclease–GAATTC complex, there is a set of six symmetrically disposed DNA phosphates (Figure 2a and b) where the protein makes constraining hydrogen bonds from polypeptide main chain amides or polar side chains to stabilize the distorted DNA conformation and to anchor the recognition elements. Ethylation of any one of these phosphates interferes strongly with binding. This set includes a pair of 'primary clamps' (at pGAATTC), which show strong interference in every endonuclease–DNA complex except the 'non-specific complex', and four 'supplementary clamps' (at pAGAApTTC) which often show adaptive changes in

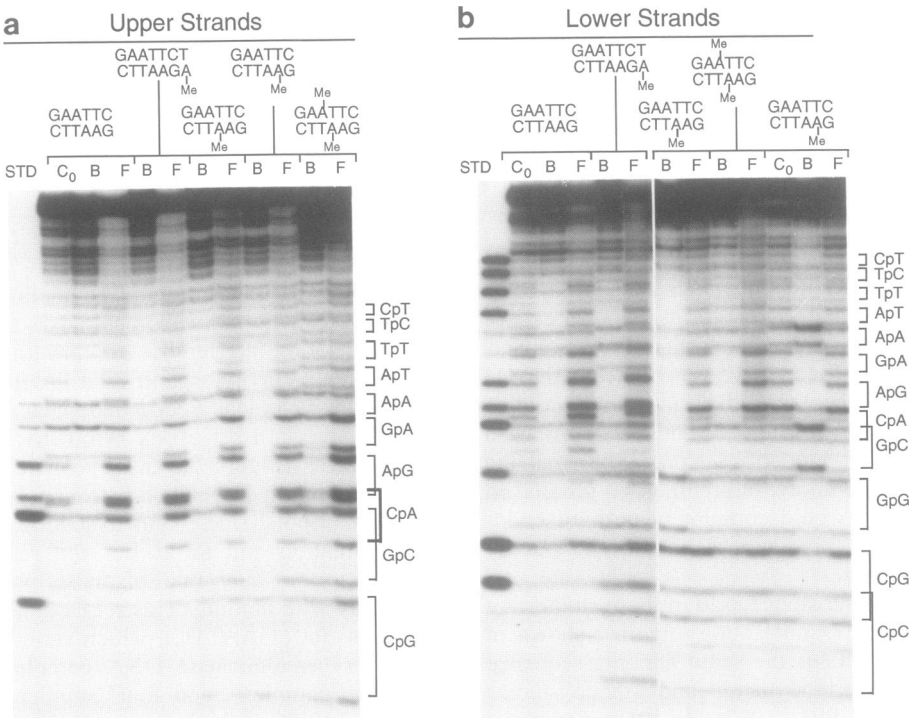
complexes with non-canonical DNA sites (Lesser *et al.*, 1990). In addition, non-canonical complexes often show *increased* interference at other phosphates which show only weak interference in the canonical complex. The characteristic interference footprint was unaffected (Figure 2a and b) by DNA methylation outside the recognition site.

As shown in Figure 2f, the hemimethylated GA$^{m}$ATTC complex shows retention of interference at all 'primary clamp' and 'supplementary clamp' positions, but abnormally strong interferences at all phosphates in the methylated half-site, including one position (GAp$^{m}$ATTC) where there is absolutely no interference in the unmodified complex. Note also the abnormally strong interference at the scissile GpA bond in the methylated half-site; in all

**Table I.** Binding and cleavage of methylated DNA sites by *Eco*RI endonuclease

| Site | $K_A$ ($M^{-1}$) | $\Delta\Delta G^{o}_S$ (kcal/mol) | $k_1$ ($s^{-1}$) | Rel. $k_1 x K_A$ ($M^{-1}\ s^{-1}$) | $\Delta\Delta G^{o}_{I}{}^{\ddagger}$ (kcal/mol) | $k_2$ ($s^{-1}$) | Asymmetry $k_1/k_2$ |
|---|---|---|---|---|---|---|---|
| GAATTC<br>CTTAAG | 2.9 (±0.4)×$10^{10}$ | 0 | 7.9 (±1.4)×$10^{-1}$ | 1 | 0 | 8.0 (±1.6)×$10^{-1}$ | 1 |
| GA$^{m}$ATTC<br>CT TAAG | 2.6 (±0.3)×$10^{9}$ | +1.4 ± 0.1 | 2.6 (±0.4)×$10^{-2}$ | 3.0×$10^{-3}$ | +3.4 ± 0.3 | 2.1 (±0.7)×$10^{-3}$ | 12 |
| G$^{m}$AATTC<br>C TTAAG | 2.2 (±0.4)×$10^{9}$ | +1.5 ± 0.1 | 4.1 (±0.5)×$10^{-4}$ | 3.9×$10^{-5}$ | +6.0 ± 0.3 | 8.8 (±0.8)×$10^{-5}$ | 5 |
| GA$^{m}$A TTC<br>CT T$^{m}$AAG | 2.2 (±0.4)×$10^{7}$ | +4.2 ± 0.1 | 1.3 (±0.2)×$10^{-6}$ | 1.3×$10^{-9}$ | +12.1 ± 0.3 | 1.4 (±0.1)×$10^{-6}$ | 1 |
| G$^{m}$AAT TC<br>C TTA$^{m}$AG | 2.0 (±0.4)×$10^{7}$ | +4.3 ± 0.1 | 0 | 0 | ∞ | 0 | NMF |
| $^{m}$AGAATTC<br>TCTTAAG | 2.6 (±0.5)×$10^{10}$ | ~0 | 8.3 (±1.2)×$10^{-1}$ | 1 | 0 | 8.3 (±1.2)×$10^{-1}$ | 1 |

Each site, embedded in a 17 nucleotide oligomer was tested in both orientations, e.g. $N^6$-methyladenine ($^{m}$A) in positions 3′ or 3 (Figure 1). Values of $K_A$ for both orientations were in close agreement; means ± SD of three determinations for one orientation are shown. Binding buffer was 10 mM bis-tris-propane, 120 mM NaCl, 1 mM $Na_2EDTA$, 100 μg/ml bovine serum albumin, 5 μM dithiothreitol (pH 7.5, 25°C). Equilibrium association constants $K_A$ are expressed as moles of duplex 17 nucleotide oligomer, and endonuclease dimers as the active species. The difference in standard binding free energy between the unmodified site and each modified site is: $\Delta\Delta G^{o}_S = -RT \ln (K_A/K_{A\ canonical})$. In sites with single $^{m}$A substitutions, $k_1$ represents cleavage in the unmodified DNA half-site, and $k_2$ represents cleavage in the modified DNA half-site. For the 3,3′-di-mA site, the designation of $k_1$ and $k_2$ is arbitrary. Values of $k_1$ and $k_2$ (obtained at 5 mM $Mg^{2+}$) for both orientations were in close agreement; means ± SD of the two orientations (8–11 determinations) are presented, including data not in the $Mg^{2+}$ concentration-dependent series of Figure 4. The difference (between methylated and unmethylated DNA) in the standard free energy for converting free enzyme and DNA to the transition state for the first bond-breaking step is: $\Delta\Delta G^{o}_{I}{}^{\ddagger} = -RT \ln [(k_1 \times K_A)_{\ modified}/(k_1 \times K_A)_{canonical}]$. Calculations were made with $k_1$ measured at 72 mM NaCl (see Figure 3) and values of $K_A$ measured at 120 mM NaCl; values of $k_1$ are independent of salt concentration in this range for the canonical site (Lesser *et al.*, 1990) and for the methylated sites (data not shown).



other cases where we have seen increased interference at this phosphate, there has been a correlation with decreased cleavage rates (Lesser *et al.*, 1990, 1992, 1993) in that half-site.

For the non-biological hemimethylation at G$^{m}$AATTC, the complex shows a more complicated pattern of adaptive changes in phosphate contacts: there are increased interferences throughout the unmethylated half-site, accompanied by loss of the supplementary clamp contacts in the methylated half-site (Figure 2e). At two phosphates pCAG$^{m}$ApATTC we observed that the ethylated DNA is overrepresented in the protein-bound fraction, implying that ethylation of these phosphates *enhances* endonuclease binding. This pattern closely resembles those observed (Lesser *et al.*, 1990) for sites with one incorrect base pair (so-called *Eco*RI* sites).

Thus, the two distinct hemimethylated sites exhibit different 'adaptive' changes in phosphate contacts, despite the fact that the changes in binding free energy are quite similar (Table I). These cases illustrate the general conclusion that enzyme–substrate complexes can adapt to optimize binding free energy. The different adaptations

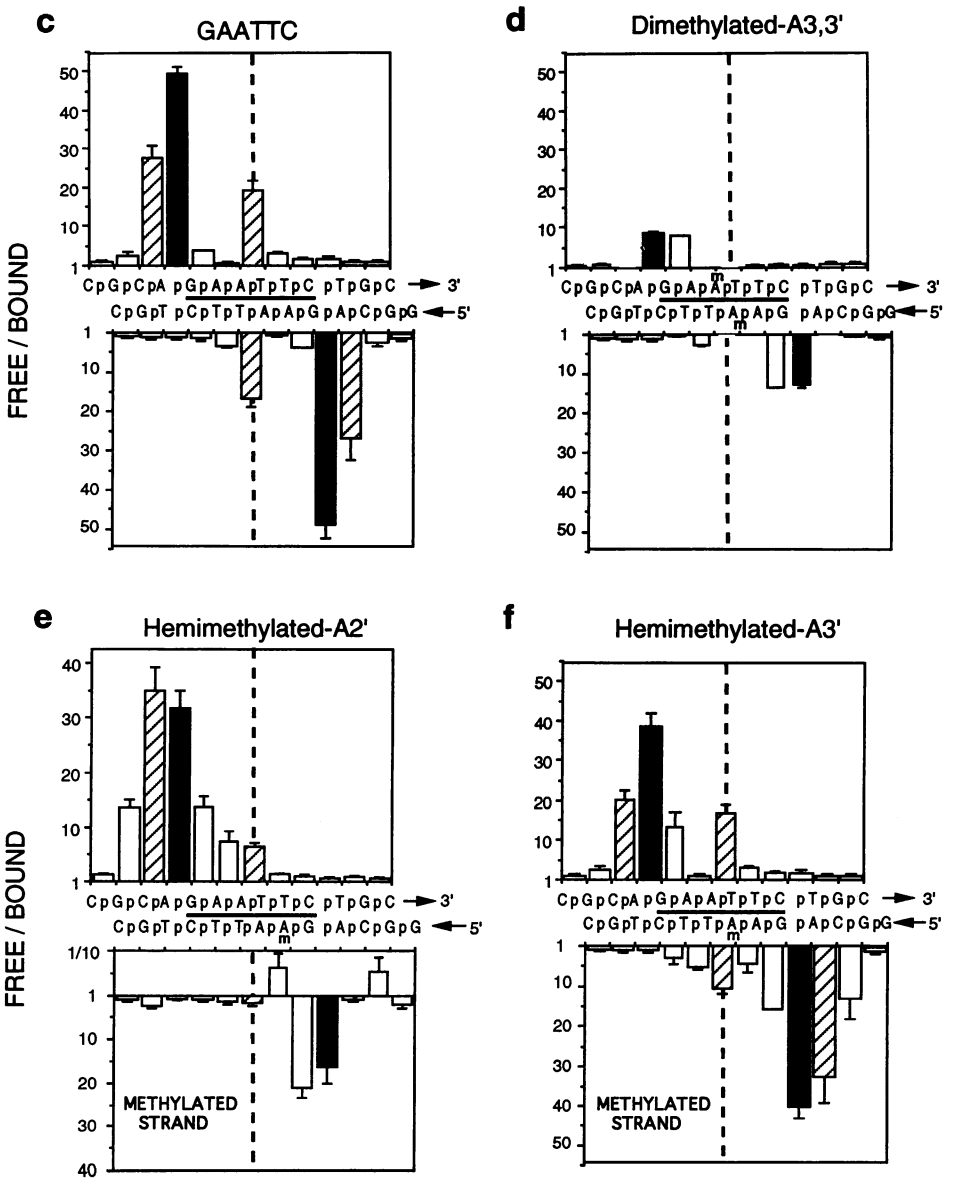



**Fig. 2.** Ethylation-interference footprints made by *Eco*RI endonuclease on canonical and methylated sites. Each DNA strand was ethylated separately, then 5′-end labeled and annealed to its unlabeled, unethylated complement. Separate lanes show ethylated DNA alone ($C_0$), and the bound (B) and free DNA (F) fractions. Alkaline cleavage at each phosphotriester yields two bands, corresponding to 3′-OH and 3′-P(OEt) groups, identified by comparison with labeled synthetic oligonucleotide standards (STD) terminating in 3′-OH. Note that each product containing $^{m}$A is slightly displaced from its unmethylated counterpart. (**a**) Alkaline cleavage products from the labeled, ethylated top strand (see Figure 1) for each site. (**b**) Products from the labeled, ethylated bottom strand for each site. (c–f) Quantitative data obtained by densitometry of autoradiograms similar to those in (a) and (b). Each bar (mean ± SD of 4–5 independent experiments) shows the degree of interference (for each doublet, the ratio of intensity in the free DNA fraction to that in the endonuclease-bound fraction). The interference axis (ordinate) is positive in both directions. A value >1 indicates that ethylation interferes with binding; a value <1 indicates that ethylation increases binding. There was no interference at phosphates not shown. (**c**) Canonical GAATTC site. The 'primary clamp' (black) and 'supplementary clamp' phosphates (striped) are indicated for reference. Vertical dashed lines indicate the center of symmetry in the canonical site. (**d**) 3,3′-di-$^{m}$A site; (**e**) 2′-$^{m}$A site; (**f**) 3′-$^{m}$A site. Footprints on 2-$^{m}$A and 3-$^{m}$A sites (not shown) were the mirror images of those in (e) and (f), respectively.

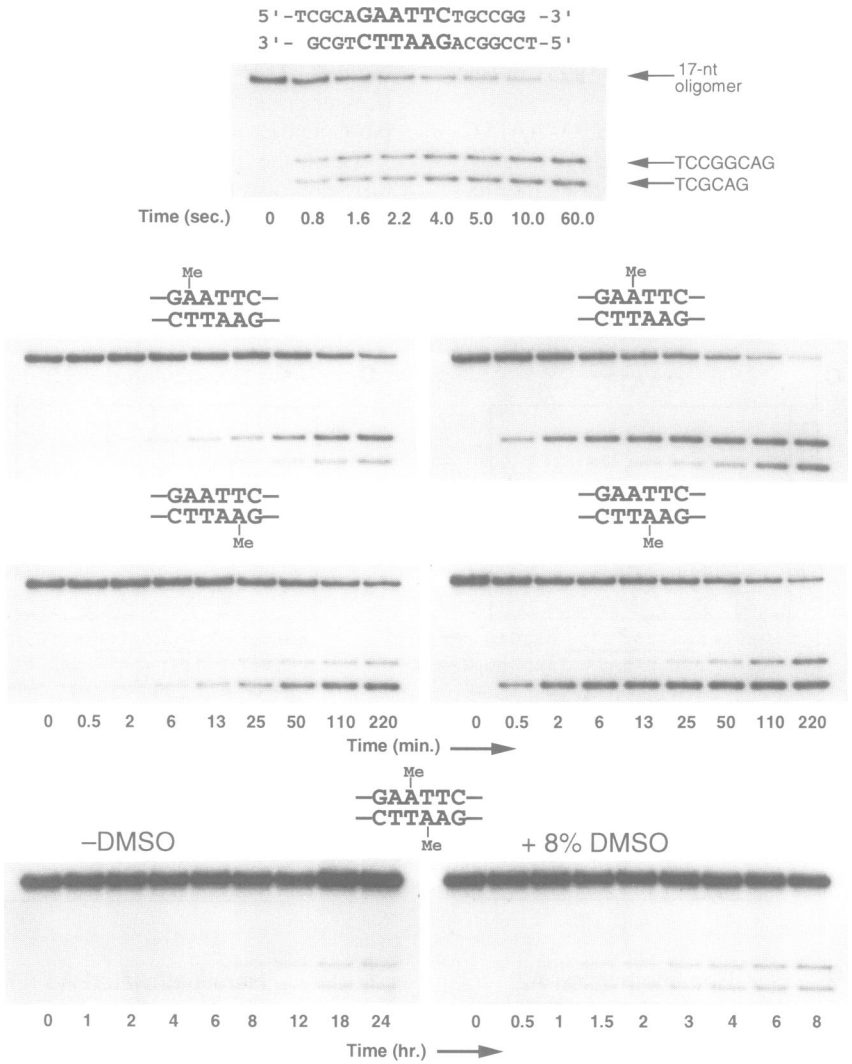



**Fig. 3.** Single-turnover cleavage reactions for canonical and methylated DNA sites. Methylated DNA sites were embedded in the 17 nucleotide oligomer shown at the top. A representative time course for each orientation of a site methylated in one strand is shown. Enzyme (0.72 μM dimer) and DNA (0.54 μM) were pre-equilibrated in 10 mM bis-tris-propane, pH 7.5, 72 mM NaCl, 100 μg/ml bovine serum albumin, 5 μM dithiothreitol for 30 min at 25°C such that all DNA was enzyme bound. Reactions were initiated by the addition of $MgCl_2$ to a final concentration of 5 mM, quenched at the indicated times and the cleavage products separated as described (Lesser *et al.*, 1990, 1992).

seen in these two complexes are reflected in differential effects on cleavage rates in the two DNA half-sites (see below).

Although methylation in one DNA strand is accommodated by adaptation, this tightly packed interface evidently finds difficulty in making two such adaptations simultaneously. In the biologically relevant 3,3′-di-$^{m}$A site, the disruption of the interface is severe enough so that the footprint shows loss of the 'supplementary clamp' contacts in both DNA half-sites (Figure 2d). This further loss of phosphate contacts may, in part, account for the observation that the effect of double methylation on binding (Table I) is more than twice the effect of single methylation.

Nonetheless, the 3,3′-di-$^{m}$A site shows a discrete, localized footprint and the 'primary clamp' contacts at pGA$^{m}$A-TTC remain in place. By contrast, for the 2,2′- di-$^{m}$A site, the overall unfavorable effect is too large to be overcome, such that only a 'non-specific' complex is formed. This is reflected in the complete absence of a discrete ethylation-interference footprint on this site, and in the fact that this complex is incompetent to catalyze a cleavage reaction (see below).

### *Endonuclease cleavage of methylated DNA*

Methylation of the *Eco*RI recognition site confers protection against *Eco*RI endonuclease *in vivo* (Kuhn *et al.*, 1986; Heitman *et al.*, 1989), and *in vitro* at low endonuclease concentration (Rubin and Modrich, 1977). Although it is generally believed that methylation completely prevents cleavage (McClelland *et al.*, 1994), this belief is not supported by the facts. The data indicate a more complicated situation.

*Eco*RI sites bearing one methyl group on either adenine-$N^{6}$, or two methyl groups on A3 and A3′ (3,3′-di-$^{m}$A), are cleaved by *Eco*RI endonuclease *in vitro* (Table I and Figure 3), although first-order (single-turnover) cleavage rate constants are significantly reduced relative to those for the unmethylated DNA. Only the 2,2′-di-$^{m}$A site is completely protected against cleavage.

We consider separately the cleavage rates in the two DNA 'half-sites', since each endonuclease–DNA complex has access to two distinct transition states corresponding to cleavage in each of the DNA strands, and cleavage is thus a parallel-sequential process (Lesser *et al.*, 1990). For recognition sites containing a modified base in one half-site, we use the convention of designating rate constants $k_1$ for cleavage in the unmodified half-site and $k_2$ for the modified half-site. For sites with symmetrical modification in both strands, we always observe equal cleavage rates in the two DNA strands, and the designation of $k_1$ and $k_2$ is arbitrary.

We have pointed out previously (Lesser *et al.*, 1990) that since $k_1$ represents a catalytic event at the GpA bond remote from the locus of DNA modification, changes in $k_1$ report on adaptive changes in the protein–DNA interface over the whole complex. By contrast, $k_2$ is more strongly influenced by local changes in the modified half-site. To a considerable extent, changes in these rate constants can be correlated with changes in the ethylation-interference footprints (Lesser *et al.*, 1990).

Single-strand methylation at A3 depresses $k_1$ by 30-fold (GA$^{m}$ATTC, Table I). This inhibition does not reflect simply the deletion of a protein–base hydrogen bond, since deletion of one protein–base hydrogen bond by substitution of $^{7C}$A for A2 or A3 (M.R.Kurpiewski, D.R.Lesser and L.Jen-Jacobson, in preparation) or deletion of two protein–base hydrogen bonds by double-strand substitution of purine for A (Lesser *et al.*, 1993) have little or no effect on $k_1$. Thus, in such cases, the deletion of one or two hydrogen bonds equally destabilizes the unactivated enzyme–DNA complexes and the transition state complexes [ES$^{\ddagger}$], such that $\Delta\Delta G^{o}{}_{1}{}^{\ddagger} \sim \Delta\Delta G^{o}{}_{S}$ (Lesser *et al.*, 1992). On the other hand, the 30-fold inhibition of $k_1$ for GA$^{m}$ATTC implies that methylation preferentially destabilizes the transition state complex. This preferential destabilization of the [ES$^{\ddagger}$] complex is even more evident for G$^{m}$AATTC, where $k_1$ is inhibited by 1900-fold relative to that for the unmodified GAATTC site.

For modified half-sites, the simplest case is that of the G$^{7C}$AATTC site, where not only is $k_1$ unaffected, but also there is no inhibition of $k_2$ (M.R.Kurpiewski, D.R.Lesser and L.Jen-Jacobson, in preparation). Thus, the transition state for cleavage in this modified half-site is destabilized to approximately the same extent as is the unactivated E′S′ enzyme–DNA complex; i.e. there is no additional transition state destabilization beyond what is expressed in the E′S′ complex. [The designation E′S′ implies that this is not an initial collision complex, see Discussion.] This destabilization presumably reflects the effect of deleting a single protein–base hydrogen bond.

By contrast, for other modified sites (GA$^{7C}$ATTC, G$^{m}$AATTC, GA$^{m}$ATTC), $k_2$ is inhibited to a greater degree than is $k_1$, such that there is a 5- to 12-fold preference for cleavage in the unmodified half-sites. This preference for cleavage in the unmodified half-site does not depend on the orientation of the eccentrically located site in the parent oligonucleotide (Figure 3). Among these sites, the highest $k_1/k_2$ ratio (Table I) is observed for the 'natural' hemimethylated site GA$^{m}$ATTC. This fact may be relevant to the protection of hemimethylated sites *in vivo* (see Discussion).

Methylation in both strands inhibits cleavage more profoundly. The 'natural' 3,3′-di-$^{m}$A site is cleaved in both DNA strands at extremely slow but approximately equal rates in both DNA strands (Figure 3, Table I). The overall penalty in transition state interaction free energy $\Delta\Delta G^{o}{}_{1}{}^{\ddagger}$ (+12.1 kcal/mol, Table I) is similar to that for the worst *Eco*RI* sites (Lesser *et al.*, 1990). The cleavage rate for the 3,3′-di-$^{m}$A DNA is dramatically increased (~10-fold, Figure 3) by the addition of 8% dimethylsulfoxide (DMSO), which is also known to promote cleavage at *Eco*RI* sites with one incorrect base pair (Malyguine *et al.*, 1980; Robinson and Sligar, 1993; C.Yee, D.R.Lesser, M.Kurpiewski and L.Jen-Jacobson, in preparation). We observed no cleavage of the 2,2′-di-$^{m}$A site (Figure 3, Table I). Taken with the absence of a localized interference footprint (see above), this implies that this site forms only a 'non-specific' complex with the endonuclease (Lesser *et al.*, 1990). DMSO does not promote cleavage of the 2,2′-di-$^{m}$A DNA.

To show that the measured changes in cleavage rate constants for the structurally adapted complexes with methylated DNA sites were not influenced by an altered affinity for the $Mg^{2+}$ cofactor, we measured the first-order cleavage rate constants as a function of $Mg^{2+}$ concentration. We found (Figure 4) that cleavage rate constants $k_1$ and $k_2$ for both the hemimethylated G$^{m}$AATTC and GA$^{m}$ATTC sites showed hyperbolic (saturable) dependences on $Mg^{2+}$ concentration that were not significantly different from that for the canonical GAATTC site. The best-fit values of the equilibrium dissociation constants $K_D$ for $Mg^{2+}$ were: GAATTC, 3.0 ± 1.1 and 3.3 ± 0.8 mM (for $k_1$ and $k_2$, respectively); GA$^{m}$ATTC, 2.0 ± 0.8 and 2.2 ± 1.4 mM; G$^{m}$AATTC, 1.5 ± 0.5 and 2.4 ± 0.5 mM. It is especially noteworthy that the $K_D$ values for $Mg^{2+}$ do not differ significantly for cleavage in the normal half-site ($k_1$) and the abnormal half-site ($k_2$) for any substrate, despite the fact that the 'adaptations' in the hemimethylated complexes (Figure 2) produce marked structural asymmetries.

Most of the data in Figure 4 were obtained by adding $Mg^{2+}$ to pre-formed endonuclease–DNA complexes, but when we added endonuclease to a pre-equilibrated mixture of $Mg^{2+}$ and DNA (Figure 4A) we obtained similar values for both the $Mg^{2+}$ $K_D$ (for GAATTC, 2.8 ± 0.9 and 2.8 ± 1.2 mM for $k_1$ and $k_2$, respectively) and the calculated values of $k_1$ and $k_2$ extrapolated to saturating [$Mg^{2+}$]. This is reasonable because the rate constants for endonuclease–DNA association (Jen-Jacobson *et al.*, 1986) are much larger than those for strand scission, so that the reactions proceed under single-turnover conditions so long as endonuclease is in excess over DNA and concentrations of endonuclease and DNA are well above the protein–DNA binding constants.

Note that our 'standard' condition for cleavage reactions (Figure 3, Table I) uses 5 mM $Mg^{2+}$, which is above the apparent $K_D$ for $Mg^{2+}$, and only slightly below saturation. This concentration was chosen to approximate the [$Mg^{2+}$] found *in vivo* in *Escherichia coli*. Our data, however, indicate that differences in affinities for $Mg^{2+}$ do not play a significant role in the observed differences in cleavage rates or transition state interaction free energies among methylated and unmethylated complexes.

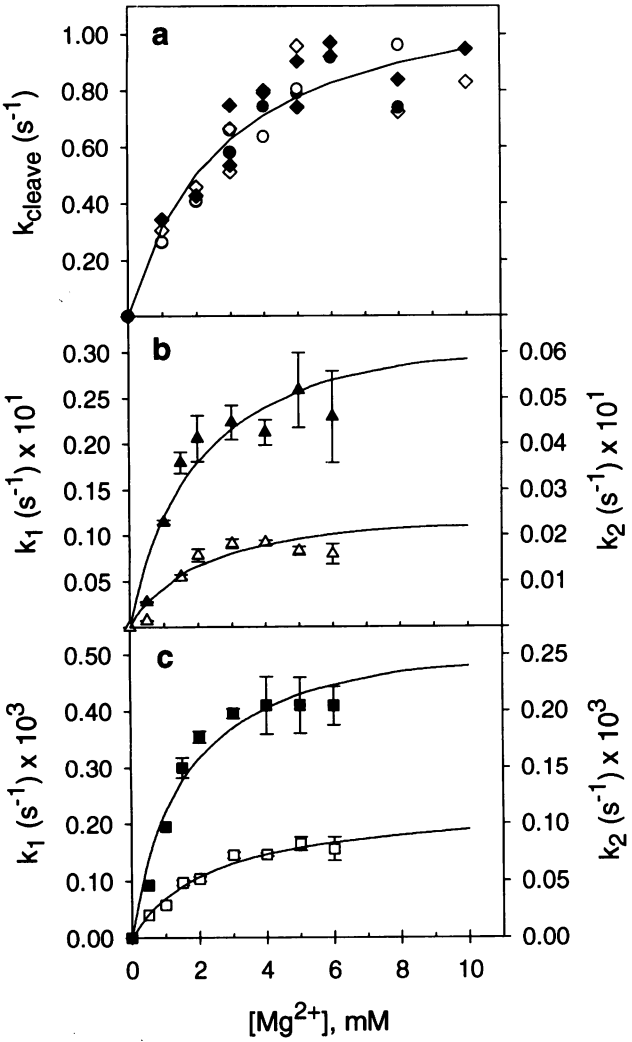



**Fig. 4.** Dependence of first-order cleavage-rate constants on [$Mg^{2+}$]. Reactions were conducted and analyzed as described in Materials and methods, except that the concentration of $MgCl_2$ was varied as indicated. For each reaction, the ionic strength of the buffer was adjusted with NaCl to compensate for differences in [$Mg^{2+}$]; the same rate constants were obtained whether compensating NaCl was added to the equilibrated protein–DNA mixture, or at the same time as $Mg^{2+}$. In all panels, filled symbols show $k_1$ and open symbols show $k_2$. Lines show non-linear least-squares regression fits of each data set to the equation: $k_{app} = k_{max}$ [$Mg^{2+}$]/{$K_D$ + [$Mg^{2+}$]}, where $K_D$ is the apparent equilibrium dissociation constant for $Mg^{2+}$ that pertains to each first-order cleavage rate constant. Note the different exponential multipliers in the left ordinates of (a), (b) and (c). (**a**) Unmodified GAATTC site. Reactions were initiated by adding $Mg^{2+}$ to pre-equilibrated endonuclease and DNA (●, ○) or by adding endonuclease to pre-equilibrated DNA and $Mg^{2+}$ (◆, ◇). (**b**) Hemimethylated GA$^m$ATTC site. Note that the right ordinate ($k_2$) is magnified 5-fold relative to the left ordinate ($k_1$). (**c**) Hemimethylated G$^m$AATTC site. The right ordinate ($k_2$) is magnified 2-fold relative to the left ordinate ($k_1$). Data in (b) and (c) show means ± SD of three independent measurements in which $Mg^{2+}$ was added to pre-equilibrated protein and DNA.

## Discussion

### *Key residues in the endonuclease–DNA interface*

Much of our discussion of the effects of DNA methylation will center on the interaction of adenine $N^6$-methyl groups with the side chains of Asn141 and Arg145 (see Figure 5), so we first consider the roles of these side chains in the canonical interface (Kim *et al.*, 1990, 1994; Rosenberg 1991). The canonical interface is fully symmetrical, with identical protein–DNA contacts in the two half-sites; we describe only one half-site. Note especially that Asn141 and Arg145 making the described contacts in one DNA half-site derive from different subunits of the endonuclease dimer.

The side chain carbonyl group of Asn141 lies between the $N^6$-amino groups of the two adenines in one half-site, receiving hydrogen bonds from both, while the side chain amide of Asn141 donates a hydrogen bond to the imidazole-$N^7$ of adenine-2. Thus, Asn141 is responsible for making three protein–base hydrogen bonds in each half-site. The characteristic DNA distortion in the *Eco*RI endonuclease complex ('kinking'; Frederick *et al.*, 1984; Kim *et al.*, 1994) includes an abnormal tip angle of adenine-3; this is necessary to bring adenine-2 and adenine-3 into precise positions for hydrogen bond 'bridging' by Asn141. It should also be noted that Asn141 lies near the C-terminal end of the 'extended chain motif', which is responsible for many of the recognition contacts with the pyrimidine bases (Rosenberg, 1991).

The $N^\delta$ of Arg145 donates a hydrogen bond to the imidazole-$N^7$ of adenine-3, while the guanidino-$N^\varepsilon$ donates a hydrogen bond to a phosphoryl oxygen at the scissile GpA bond. The latter (charge–charge) interaction probably helps polarize the scissile GpA phosphate for nucleophilic attack on phosphorus during catalysis. This base–phosphate bridging, like the base–base bridging by Asn141, would be impossible without the particular DNA distortion characteristic of the endonuclease–DNA complex.

### *DNA methyl groups interfere with key recognition and catalytic side chains*

An $N^6$-methyl group at either adenine obviously prevents hydrogen bonding between adenine-$N^6$ and the side chain carbonyl of Asn141, but the effects of methylation on the structure of the complex are far more profound than a simple blockage of a protein–base hydrogen bond. To visualize how adenine-$N^6$-methyl groups might fit into the interface, we simulated (Figure 5) the *Eco*RI endonuclease–DNA interface with methyl groups added. These views were constructed from the atomic coordinates of the canonical (unmethylated) complex, with methyl groups added by computer to adenine-2 (Figure 5a) or adenine-3 (Figure 5b) using the bond lengths, angles and rotations taken from the coordinates of Frederick *et al.* (1988) for free duplex CGCGA$^m$ATTCGCG.

In the case of $N^6$-methylation of adenine-2 (Figure 5a), it is evident that the methyl group conflicts with both the $C^\gamma$ and the carbonyl of Asn141, such that the entire side chain of Asn141 must be displaced. It is hard to guess at the nature of this displacement. If the polypeptide backbone remained in a fixed position and if Asn141 were to rotate around the Cα–Cβ bond to the right (i.e. toward adenine-3) enough to relieve the conflict with the methyl group on adenine-2, Asn141 would come into steric conflict with the side chain of Arg145. Conversely, rotation of Asn141 to the left (i.e. toward guanine-1) would produce conflict with the Arg203 side chain (not shown) that recognizes guanine-1. We therefore surmise that resolution of the steric clash between Asn141 and $N^6$-methyl-adenine-2 requires some displacement of the polypeptide backbone in

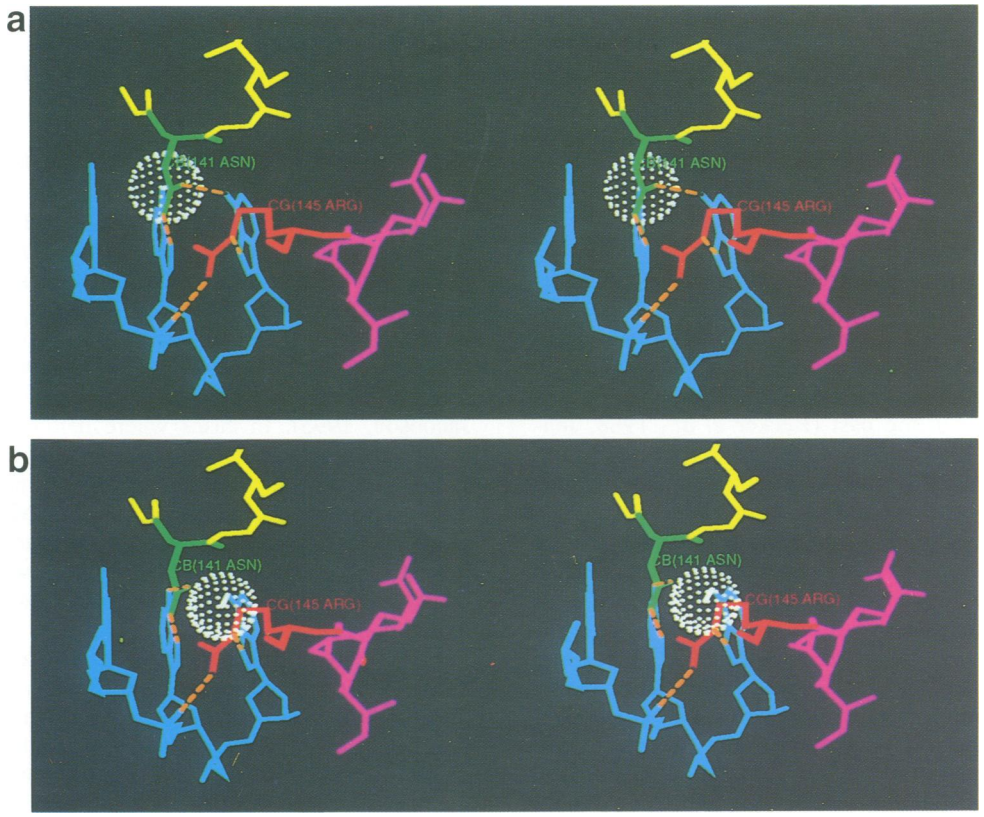



**Fig. 5.** Computer simulation of methyl group positions superimposed on the canonical *Eco*RI endonuclease–DNA interface. Only one half-site of each complex is shown with the GAA nucleotides of the DNA (light blue) left to right. Models were drawn from the refined atomic coordinates of the *Eco*RI endonuclease–TCGCGAATTCGCG complex (Brookhaven Protein Data Bank accession No. 1eri). Methyl groups (white van der Waals spheres) were added to adenine-2 (**a**) or adenine-3 (**b**) using the bond lengths, angles and rotations determined crystallographically (accession No. 4dnb) by Frederick *et al.* (1988) for the free homoduplex CGCGA$^{m}$ATTCGCG. Note that Arg145 (red) derives from a different endonuclease subunit (magenta polypeptide segment) than the subunit (yellow polypeptide segment) that provides Asn141 (green). Dashed lines indicate hydrogen bonds in the canonical complex.

the vicinity of Asn141. We have no basis for estimating the exact nature or magnitude of such movement, but the ethylation-interference footprints (Figure 2; see below) imply a net adjustment of the interface in the direction of the unmodified half-site. Displacement of the region around Asn141 may have consequences for the position of the 'extended chain motif' (residues Met137–Ala142) that contributes recognition contacts with the pyrimidine bases.

An $N^6$-methyl group on adenine-3 (Figure 5b) comes into steric conflict with $C^{\delta}$ and $N^{\delta}$ of Arg145, preventing hydrogen bonding between $N^{\delta}$ and the imidazole-$N^7$ of adenine-3. This conflict might be resolved, without movement of the polypeptide backbone, by any of several relatively unrestricted rotations of the Arg145 side chain. Many of these would slightly increase the distance between the guanidino-$N^{\varepsilon}$ and the phosphoryl oxygen at the scissile GpA bond.

The $N^6$-methyl group on adenine-3 also conflicts with the side chain carbonyl of Asn141, but this can be accommodated relatively easily by rotation of the Asn141 side chain away from adenine-3, while still retaining good hydrogen bonding distances to $N^6$ and $N^7$ of adenine-2.

### *Endonuclease 'adaptations' to methylated DNA sites minimize binding free energy*

Considering the steric conflicts between adenine methyl groups and 'recognition' side chains, it is clear that even one adenine-$N^6$-methyl group must prevent formation of more than one of the protein–base hydrogen bonds that are normally found in the canonical GAATTC complex. In addition, the G$^{m}$AATTC complex lacks two of the 'supplementary clamp' phosphate contacts. It is therefore surprising that hemimethylation has so little effect on binding free energy ($\Delta\Delta G^{o}_{S} \sim +1.4$ kcal/mol; Table I), about the same as the penalty for deletion of a single protein–base hydrogen bond (Lesser *et al.*, 1993). Thus, the energetic penalty for loss of protein–base hydrogen bonds and protein–phosphate contacts in the hemimethylated complexes must be partially compensated by favorable factors not present in the canonical complex. There are several possibilities, not mutually exclusive, for the origin of such favorable contributions to $\Delta\Delta G^{o}_{S}$.

(i) The observed binding free energy change $\Delta\Delta G^{o}_{S}$ (Table I) measures the difference in free energies between the free and bound molecules. In forming the canonical complex, there may be an unfavorable contribution from desolvating the $N^6$-amino groups of the adenines, which had been solvated in the free DNA. An $N^6$-methyl group might prevent such solvation in the free DNA, so the contribution from this unfavorable term would be reduced for a hemimethylated DNA site. Conversely, if $N^6$-methyl groups are solvated in the free DNA, release of bound water from this non-polar surface would make a favorable contribution that has no counterpart in the canonical

complex. The magnitude of such energetic contribution(s) is presently unknown.

(ii) After rotation or other movement to avoid steric clash, the side chain of Arg145 or Asn141 may form non-polar (van der Waals) interactions with the $N^6$-methyl group. We have estimated previously (Lesser *et al.*, 1990), for example, that a non-polar interaction between a protein side chain and a thymine methyl group contributes about −1 kcal/mol to binding free energy.

(iii) The ethylation-interference footprints (Figure 2) show that there are structural differences between the canonical complex and complexes with hemimethylated DNA sites, reflected in changes in the degree to which ethylation at particular phosphates inhibits binding. We have been careful in the past (Becker *et al.*, 1988; Jen-Jacobson *et al.*, 1991; Lesser *et al.*, 1992; Jen-Jacobson, 1995) to point out that such differences in interference cannot always be interpreted in simple structural terms, because interference may result either from direct conflict of an ethylphosphotriester with protein, or from an effect of ethylation on the ability of the DNA to assume a particular distorted conformation. Nevertheless, changes of the magnitude we observe here indicate that the relative positions of protein and DNA must be different in the canonical and hemimethylated complexes. This adjustment of relative positions is what we have termed 'adaptation' in the complex (Lesser *et al.*, 1990). Adaptation presumably serves to optimize binding free energy, and it is possible that the 'adapted' complexes include protein–DNA interactions (including, perhaps, new protein–phosphate contacts) that had no counterpart in the canonical complex.

The ethylation-interference footprints of both hemimethylated complexes, although differing in detail, give the impression that the 'adaptation' involves an asymmetric adjustment in the position of the endonuclease dimer relative to the DNA. For the G$^m$AATTC site, this adjustment appears to be in the direction of the unmethylated half-site, at the expense of phosphate contacts (especially the 'supplementary clamp' contact at pAG$^m$AATTC) in the methylated half-site. This impression is reinforced by the observation that ethylation at the upstream phosphate pCAG$^m$AATTC actually enhances endonuclease binding. The remarkable similarity of the adaptations in the G$^m$AATTC site to those in *Eco*RI* sites (Lesser *et al.*, 1990) leads us to suspect that *Eco*RI endonuclease has a limited repertoire of alternative structural adaptations to minimize binding free energy.

For the GA$^m$ATTC site, the asymmetric adjustment is more subtle and results primarily in increased interferences in the methylated half-site, but without loss of the 'supplementary clamp' phosphate contacts in either half-site.

If adaptation to a hemimethylated site involves asymmetric adjustments of the protein–DNA interface to avoid steric conflict and optimize binding free energy, it should not be surprising that the endonuclease cannot achieve such adaptations when both half-sites are methylated. That is, the endonuclease dimer cannot adjust in two opposing directions simultaneously. In the case of the 2,2′-di-$^m$A site, no accommodating adaptation is evidently possible, such that only a non-specific (cleavage-incompetent) complex can be formed. The complex with the 3,3′-di-$^m$A site shows fully symmetric interferences (Figure 2), but the 'supplementary clamp' contacts are lost in both half-sites. The 'primary clamp' contacts at pGA$^m$ATTC remain as points of strong interference; this is an invariant feature of all site-localized complexes of *Eco*RI endonuclease (Lesser *et al.*, 1990).

### Adaptation and the non-additivity of binding free energy

When a DNA modification produces a change in binding free energy ($\Delta\Delta G^\circ_S$), the effect of two modifications may or may not be the sum of the $\Delta\Delta G^\circ_S$ values for the separate modifications. The conceptual basis for additive (Lesser *et al.*, 1993), superadditive (Lesser *et al.*, 1990) and subadditive (Lesser *et al.*, 1992) behaviors of binding free energy changes has been developed elsewhere (Jen-Jacobson, 1995). In cases where neither the single nor double modifications elicit adaptive changes in the interface (e.g. replacement of adenine by purine, Lesser *et al.*, 1993; replacements of adenine by $^7$CA, M.R.Kurpiewski, D.R.Lesser and L.Jen-Jacobson, in preparation), we have generally observed that $\Delta\Delta G^\circ_S$ values (whether positive or negative) are additive for multiple substitutions, within experimental error.

In the case of the hemimethylated sites, the observed values of $\Delta\Delta G^\circ_S > 0$ include (presumably favorable) contributions from the asymmetric adaptive changes in the interface. Because these asymmetric adaptations cannot be made in a double-methylated site, the favorable energetic factors associated with the adaptations do not contribute (or contribute less) to the observed $\Delta\Delta G^\circ_S$ for the double-methylated site, with the result that $\Delta\Delta G^\circ_S$ (double) $> 2\times\Delta\Delta G^\circ_S$ (single). This observation illustrates the important general principle that additivity of binding free energy changes is generally impossible unless the complexes being compared are (approximately) structurally isomorphous.

Our observations also show why clustering of methyl groups may be important in controlling the interaction of DNA binding proteins with their recognition sites. The disruption of protein–DNA contacts by a single DNA methylation may be partially compensated by adaptation of the DNA–protein interface, but two (or more) methyl groups in a recognition site may make the compensating adaptation impossible. This has the effect of inhibiting binding by far more than the linear sum of the effects of the individual methylations.

### Structural adaptations destabilize transition states

The structural requirements to achieve the transition state [ES$^‡$] complex leading to catalysis are much more precise than those of ground state enzyme–DNA binding, since the active site residues must be aligned very precisely in the transition state for a successful catalytic event. For *Eco*RI endonuclease, the complex that appears in the crystal structure and is detected in our binding and footprinting experiments is almost certainly not an initial collision complex ES, but an E′S′ complex which has undergone significant conformational change subsequent to binding (Jen-Jacobson *et al.*, 1986). Both crystallographic and biochemical evidence indicate that this unactivated E′S′ complex bears a very strong structural resemblance to the transition state complex [ES$^‡$], such that only very small adjustments of position are required prior to catalysis. Y.Kim, J.Choi and J.M.Rosenberg (in

preparation) have found that $Mg^{2+}$ can be diffused into endonuclease–DNA crystals and catalysis can occur without disrupting the crystals, implying that no major movements of protein or DNA elements are required to reach the transition state. The enzyme–product complex remains approximately isomorphous with the enzyme–substrate complex, except in the immediate vicinity of the cleaved GpA bond.

The close resemblance between the E′S′ complex and the transition state complex serves to minimize the energy of activation $\Delta G^{o\ddagger}$ (i.e. to maximize the first-order cleavage rate constants). This permits the full utilization of binding energy to provide energy of activation (Lesser *et al.*, 1992). In the fully symmetrical canonical complex, the probabilities of achieving each of the two distinct transition states (corresponding to cleavage events in each of the two DNA strands) are precisely equal, such that $k_1 = k_2$. However, any structural alteration that perturbs symmetry in the E′S′ complex will have unequal effects on the $\Delta G^{o\ddagger}$ required to reach each of the two transition state complexes, such that $k_1 \neq k_2$. Any changes in the positions of protein side chains with *direct* roles in the catalytic mechanism will produce additional inhibitory effects on individual rate constants.

Some individual protein–DNA interactions contribute precisely the same stabilization in the E′S′ and transition state complexes (Lesser *et al.*, 1992, 1993), confirming the close resemblance between E′S′ and [ES‡]. For example, we have found (M.R.Kurpiewski, D.R.Lesser and L.Jen-Jacobson, in preparation) that deletion of single hydrogen bonding groups in G$^{7C}$AATTC and GA$^{7C}$ATTC sites has no effect on the rate constant $k_1$ for cleavage in the unmodified half-site; i.e. $\Delta G^{o\ddagger} = 0$. In such cases, there is no confounding effect on the energetics of DNA distortion and no sign of any 'adaptation' in the complexes. The deletion of one hydrogen bond with Asn141 in the G$^{7C}$AATTC site does not affect $k_2$ (cleavage in the modified half-site), but the deletion of the hydrogen bond with Arg145-$N^{\delta}$ (in the GA$^{7C}$ATTC site) inhibits $k_2$, presumably because of a change in the positioning of Arg145 to polarize the scissile phosphate at GpA.

DNA methylation, by contrast, produces various degrees of adaptive alterations in the complexes, and these are reflected in changes in the first-order cleavage rate constants (Table I). The ethylation-interference footprints (Figure 2) imply that the structural adaptation in the G$^{m}$AATTC complex is more profound than that in the GA$^{m}$ATTC complex. Correspondingly, we observe (Table I) that $k_1$ is inhibited ~1900-fold for G$^{m}$AATTC, but ~30-fold for GA$^{m}$ATTC.

These adaptations also produce asymmetries in the complexes, as described above. For the G$^{m}$AATTC site, there is no direct effect on a side chain involved in catalysis, but the loss of protein–phosphate contacts in the methylated half-site apparently destabilizes the transition state for cleavage in this half-site, such that $k_2$ is inhibited ~9100-fold. In the case of the GA$^{m}$ATTC site, it seems certain that the position of the Arg145 side chain must change (see above). This would affect the positioning of Arg145 to polarize the scissile phosphate at GpA and, consequently, $k_2$ is inhibited by ~380-fold. For both these hemimethylated sites, the asymmetric changes in the complex thus lead to differential effects on the stabilities of the two distinct transition states and $k_1 > k_2$ (Table I).

When both inner adenines are methylated (the 3,3′-di-$^{m}$A site) the complex shows pronounced adaptive changes (loss of the 'supplementary clamp' phosphate contacts) in both half-sites (Figure 2). The loss of these phosphate contacts in other modified complexes is always associated with large energetic penalties in the transition state (Lesser *et al.*, 1990). This complex is symmetrical and we presume that the position of the Arg145 side chains must be altered in both half-sites to avoid steric conflict with the methyl groups. The result is that both $k_1$ and $k_2$ are inhibited ~600 000-fold, but the symmetry of the complex dictates that $k_1 = k_2$.

### *The molecular basis for methylation protection in vivo*

It has been believed generally (Rubin and Modrich, 1977; McClelland *et al.*, 1994) that *Eco*RI endonuclease cannot cleave its DNA recognition site when the site is protected by methylation. Structural explanations have been advanced to rationalize this belief. For example, Frederick *et al.* (1988) showed that the DNA conformation in the 3,3′-di-$^{m}$A site is essentially identical to that of the unmethylated DNA, and proposed that the closely apposed methyl groups in the major groove of the 3,3′-di-$^{m}$A site (Figure 6a) would completely block recognition and binding by *Eco*RI endonuclease. Our data (Table I) show that binding to this site is possible (although $K_A$ is inhibited ~1300-fold). It is interesting to note that the methyl groups would in fact be well separated if the 3,3′-di-$^{m}$A DNA is distorted in the same way as the unmodified DNA when bound to *Eco*RI endonuclease (Figure 6b).

Our data show that methylated GAATTC sites (except the non-biological 2,2′-di-$^{m}$A site) can be cleaved, albeit slowly, *in vitro*. On the other hand, it is clear (Kuhn *et al.*, 1986; Heitman *et al.*, 1989) that expression of the *Eco*RI methylase gene is sufficient to protect against the otherwise lethal effect of *Eco*RI endonuclease *in vivo*. We therefore pose two questions: (i) Why is the double-methylated DNA not cleaved *in vivo*? (ii) Immediately after a DNA replication fork passes a double-strand methylated *Eco*RI site, a hemimethylated GA$^{m}$ATTC site exists transiently until the newly synthesized strand is methylated by *Eco*RI methylase. Why is this site not cleaved by the endonuclease?

The best measure of the relative probability of a successful cleavage event is the product $k_{cleave} \times K_A$, which takes into account both binding and catalysis. Our data (Table I) show that the probability of first-strand cleavage (i.e. $k_1 \times K_A$) is reduced by $\sim 8 \times 10^8$-fold for the double-methylated GA$^{m}$ATTC site *in vitro*. This, in itself, may be an adequate degree of protection *in vivo*. However, the strong inhibition of $K_A$ (1300-fold) means that binding is not much stronger than to non-specific DNA ($K_A$ ~5× $10^7$/M; Lesser *et al.*, 1990), so that the large excess of non-specific DNA sites *in vivo* provides effective competitive inhibition of binding to 3,3′-di-$^{m}$A sites. Even more protection is afforded by inhibition of the rate constant (too slow to be measured) for cleavage of the nicked intermediate. Thus, the probability of lethal double-strand cleavage of a double-methylated site *in vivo* is vanishingly small.

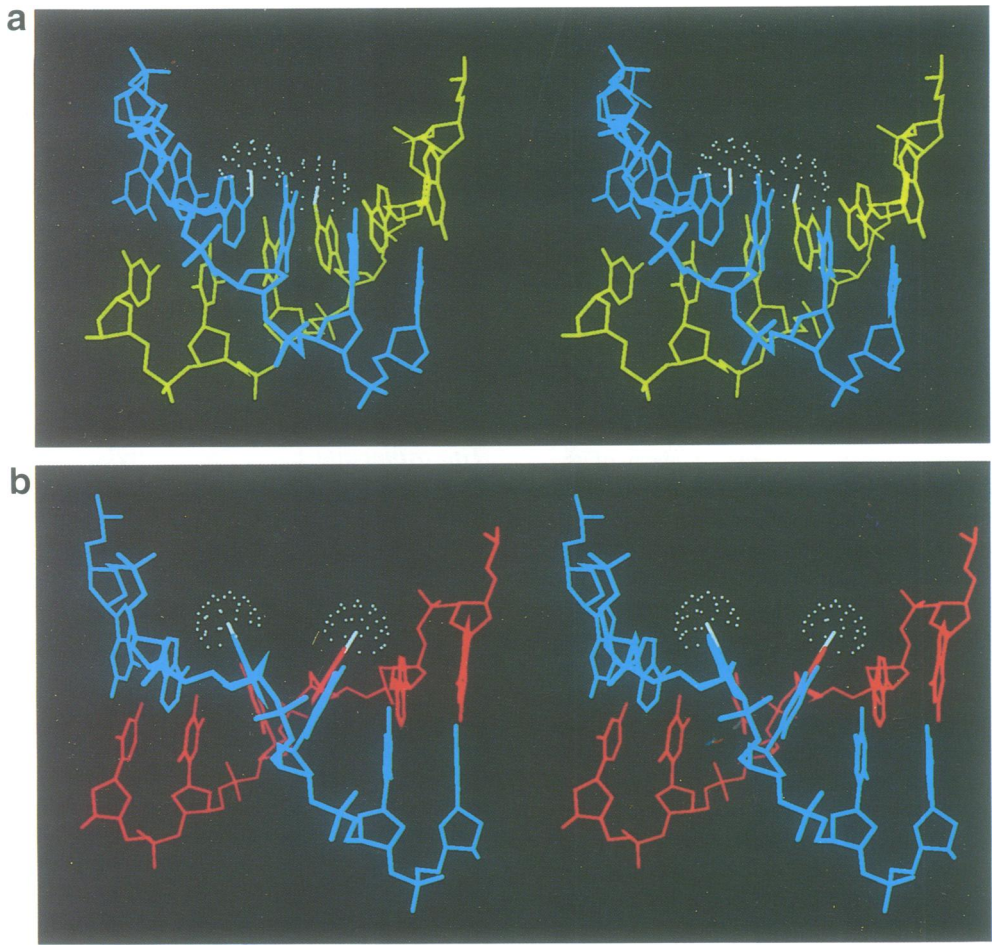



**Fig. 6.** Comparison of the positions of methyl groups in free and endonuclease-bound 3,3′-di-$^{m}$A DNA. The free DNA (**a**) is from the coordinates of Frederick *et al.* (1988). *Eco*RI endonuclease-bound 3,3′-di-$^{m}$A DNA (**b**) was simulated by computer as described in the legend to Figure 5. The van der Waals spheres of the methyl groups are in white.

By contrast, our data for the hemimethylated GA$^{m}$ATTC site show that $k_1 \times K_A$ is reduced only ~330-fold relative to an unmodified GAATTC site. Although this accounts for the previous failure to observe *in vitro* cleavage of hemimethylated plasmid or bacteriophage DNA at low endonuclease concentration (Rubin and Modrich, 1977), this would seem to be an inadequate degree of protection, taken alone, during the transient existence of such hemimethylated sites *in vivo* after DNA replication. Even if a hemimethylated site were to suffer a single-strand nick in the unmethylated DNA strand, however, it is very unlikely that this would progress to double-strand cleavage because the second-order rate constant ($k_2 \times K_A$) for the second cut (in the methylated strand) is reduced by 4200-fold. This rate is so slow that the endonuclease would dissociate from the nicked DNA before making such a second-strand cut, and the single-strand nick would then be repaired by DNA ligase. This 'kinetic proofreading' mechanism for protection is very similar to that which we have previously demonstrated for *Eco*RI* sites (Lesser *et al.*, 1990).

### *DNA methylation and biological function of DNA binding proteins*

This well-characterized case demonstrates some of the structural and thermodynamic factors involved in protein interaction with methylated DNA recognition sites. Because restriction endonucleases exercise more stringent sequence discrimination than some other classes of DNA binding proteins (Lesser *et al.*, 1990), we urge caution in extrapolating detailed mechanisms from *Eco*RI endonuclease to other proteins whose biological functions are modulated by DNA methylation. Nevertheless, we infer some potentially useful principles to guide future studies of how DNA methylation affects the interactions and biological functions of site-specific DNA binding proteins, including non-catalytic DNA binding proteins.

(i) Methylation of DNA bases may affect the energetics of DNA–protein interactions in a variety of ways, of which the simplest conceptually is direct steric conflict with recognition residues on the protein. This direct conflict may, in turn, affect interactions with unmodified bases and/or with DNA phosphates. Even if the methylated base is not normally recognized by protein, DNA methylation may change the energetic contribution from release of bound water upon protein–DNA association, or may affect either the conformation of the free DNA or the energy required to achieve a particular DNA conformation in the protein–DNA complex (Lesser *et al.*, 1990, 1993).

(ii) When DNA methyl group(s) are in direct steric conflict with protein elements crucial to recognition, the net effect on binding free energy ($\Delta\Delta G^{o}_{S}{}^{obs}$) can be unexpectedly small because the protein–DNA interface

adapts to minimize binding free energy. This adaptation may produce favorable contributions to $\Delta\Delta G^{o}_{S}$ that have no counterpart(s) in the canonical complex.

(iii) Energy-minimizing adaptations may become impossible when two or more DNA methyl groups lie in reasonable proximity to each other in the protein binding site. As a result, penalties in $\Delta\Delta G^{o}_{S}$ are superadditive; i.e. $\Delta\Delta G^{o}_{S}$ is more unfavorable than would be predicted from the linear sum of the binding penalties for the constituent methylation events.

(iv) Adaptation alters the structural relationship between the DNA and bound protein. For catalytic DNA binding proteins, such as restriction endonucleases, this may greatly increase the energy ($\Delta G^{o\ddagger}$) required to reach the extremely precise geometry of the transition state, such that catalytic function is compromised more than might be guessed from the effect on binding alone. For non-catalytic DNA binding proteins, such as transcription factors, biological function often requires the DNA-bound protein to provide a site for interaction with one or more other proteins. If such protein–protein interactions also require highly precise geometries (perhaps in precise geometric relationship to the DNA) for proper function, then even seemingly minor 'adaptations' of a protein bound to methylated DNA might produce major effects on the biological function of the multiprotein–DNA complex. In other words, a non-catalytic protein may bind to a modified recognition site in a 'non-productive' complex. At present, there is no non-catalytic DNA binding protein for which we know the structural details of geometrically constrained protein–protein–DNA complexes. Definition of such details might provide a useful conceptual analog to the 'transition state' in catalytic proteins.

(v) Thus, the observation of a relatively small effect of DNA methylation on binding free energy may not be a very useful indicator of the overall effect on the biological function of the protein–DNA complex.

## Materials and methods

*Eco*RI endonuclease was prepared as described previously (Jen-Jacobson *et al.* 1983). All experiments used duplex 17 nucleotide oligomers (Figure 1) with the same flanking sequence used in our earlier studies (Lesser *et al.*, 1990, 1992, 1993). Oligomers containing $N^6$-methyl-2′-deoxyadenosine (monomer phosphoramidite purchased from Glen Research Co.) were synthesized by the DNA Synthesis Facility, Department of Biological Sciences, University of Pittsburgh. Purification, annealing and end-labeling of oligonucleotides were as described previously (Jen-Jacobson *et al.*, 1986; Lesser *et al.*, 1990, 1992). Equilibrium association constants ($K_A$) were measured by the rate competition method (Lin and Riggs, 1972) as modified by Lesser *et al.* (1990). Values for $K_A$ were confirmed independently by direct binding measurements for the canonical 17 nucleotide oligomer and sites methylated in only one DNA strand. Values of $K_A$ (rate competition) were obtained at pH 7.5, 25°C at six different NaCl concentrations. Only $K_A$ values determined at 0.12 M NaCl are reported. Single-turnover cleavage reactions were performed and analyzed as described (Lesser *et al.*, 1990, 1992). Ethylation-interference footprinting was done as described (Becker *et al.*, 1988; Lesser *et al.*, 1990).

## Acknowledgements

We thank J.M.Rosenberg for providing the refined coordinates for the canonical *Eco*RI endonuclease–DNA complex before their deposition, and L.Jacobson for assistance with computer graphics. This work was supported by a grant (GM-29207) from the National Institutes of Health to L.J.-J.

## References

Barras,F. and Marinus,M.G. (1989) The great GATC: DNA methylation in *E.coli*. *Trends Genet.*, **5**, 139–143.

Becker,M.M., Lesser,D., Kurpiewski,M., Baranger,A. and Jen-Jacobson,L. (1988) 'Ultraviolet footprinting' accurately maps sequence-specific contacts and DNA kinking in the *Eco*RI endonuclease–DNA complex. *Proc. Natl Acad. Sci. USA*, **85**, 6247–6251.

Bestor,T.H. and Coxon,A. (1993) The pros and cons of DNA methylation. *Curr. Biol.*, **3**, 384–386.

Bird,A. (1992) The essentials of DNA methylation. *Cell*, **70**, 5–8.

Braaten,B.A., Nou,X., Kaltenbach,L.S. and Low,D.A. (1994) Methylation patterns in *pap* regulatory DNA control pyelonephritis-associated pili phase variation in *E.coli*. *Cell*, **76**, 577–588.

Chaillet,J.R. (1994) Genomic imprinting: lessons from mouse transgenes. *Mutat. Res.*, **307**, 441–449.

Frederick,C.A., Grable,J., Melia,M., Samudzi,C., Jen-Jacobson,L., Wang,B.-C., Greene,P., Boyer,H.W. and Rosenberg,J.M. (1984) Kinked DNA in crystalline complex with *Eco*RI endonuclease. *Nature*, **309**, 327–331.

Frederick,C.A., Quigley,G.A., van der Marel,G.A., van Boom,J.H., Wang,A.J.-H. and Rich,A. (1988) Methylation of the *Eco*RI recognition site does not alter DNA conformation: the crystal structure of d(CGCGAm$^6$ATTCGCG) at 2.0-Å resolution. *J. Biol. Chem.*, **263**, 17872–17879.

Heitman,J., Zinder,N.D. and Model,P. (1989) Repair of the *Escherichia coli* chromosome after *in vivo* scission by the *Eco*RI endonuclease. *Proc. Natl Acad. Sci. USA*, **86**, 2281–2285.

Hsieh,C.-L. and Lieber,M.R. (1992) CpG methylated minichromosomes become inaccessible for V(D)J recombination after undergoing replication. *EMBO J.*, **11**, 315–325.

Jen-Jacobson,L. (1995) Structural-perturbation approaches to thermodynamics of site-specific protein–DNA interactions. *Methods Enzymol.*, **259**, 305–344.

Jen-Jacobson,L., Kurpiewski,M., Lesser,D., Grable,J., Boyer,H.W., Rosenberg,J.M. and Greene,P.J. (1983) Coordinate ion pair formation between *Eco*RI endonuclease and DNA. *J. Biol. Chem.*, **258**, 14638–14646.

Jen-Jacobson,L., Lesser,D. and Kurpiewski,M. (1986) The enfolding arms of *Eco*RI endonuclease: role in DNA binding and cleavage. *Cell*, **45**, 619–629.

Jen-Jacobson,L., Lesser,D.R. and Kurpiewski,M.R. (1991) DNA sequence discrimination by *Eco*RI endonuclease. In Eckstein,F. and Lilley,D.M.J. (eds), *Nucleic Acids and Molecular Biology*. Springer-Verlag, Berlin, Vol. 5, pp. 141–170.

Kim,Y., Grable,J.C., Love,R., Greene,P.J. and Rosenberg,J.M. (1990) Refinement of *Eco*RI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.

Kim,Y., Choi,J., Grable,J.C., Greene,P., Hager,P. and Rosenberg,J.M. (1994) Studies on the canonical DNA–*Eco*RI endonuclease complex and the *Eco*RI kink. In Sarma,R.H. and Sarma,M.H. (eds), *Structural Biology: The State of the Art, Proceedings of the Eighth Conversation State University of New York, Albany, NY 1993*. Adenine Press, Albany, NY, pp. 225–246.

Kuhn,I., Stephenson,F.H., Boyer,H.W. and Greene,P.J. (1986) Positive-selection vectors utilizing lethality of the *Eco*RI endonuclease. *Gene*, **42**, 253–263.

Lahue,R.S., Au,K.G. and Modrich,P. (1989) DNA mismatch correction in a defined system. *Science*, **245**, 160–164.

Lesser,D.R., Kurpiewski,M.R. and Jen-Jacobson,L. (1990) The energetic basis of specificity in the *Eco*RI endonuclease–DNA interaction. *Science*, **250**, 776–786.

Lesser,D.R., Grajkowski,A., Kurpiewski,M.R., Koziolkiewicz,M., Stec,W.J. and Jen-Jacobson,L. (1992) Stereoselective interaction with chiral phosphorothioates at the central DNA kink of the *Eco*RI endonuclease–GAATTC complex. *J. Biol. Chem.*, **267**, 24810–24818.

Lesser,D.R., Kurpiewski,M.R., Waters,T., Connolly,B.A. and Jen-Jacobson,L. (1993) Facilitated distortion of the DNA site enhances *Eco*RI endonuclease–DNA recognition. *Proc. Natl Acad. Sci.* USA, **90**, 7548–7552.

Li,E., Bestor,T.H. and Jaenisch,R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.

Lin,S.-Y. and Riggs,A.D. (1972) *lac* repressor binding to non-operator DNA: detailed studies and a comparison of equilibrium and rate competition methods. *J. Mol. Biol.*, **72**, 671–690.

Malyguine,E., Vannier,P. and Yot,P. (1980) Alteration of the specificity of restriction endonucleases in the presence of organic solvents. *Gene*, **8**, 163–177.

McClarin,J.A., Frederick,C.A., Wang,B.-C., Greene,P., Boyer,H.W., Grable,J. and Rosenberg,J.M. (1986) Structure of the DNA–*Eco*RI endonuclease recognition complex at 3 Å resolution. *Science*, **234**, 1526–1541.

McClelland,M., Nelson,M. and Raschke,E. (1994) Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Res.*, **22**, 3640–3659.

Messer,W., Bellekes,U. and Lother,H. (1985) Effect of *dam* methylation on the activity of the *E.coli* replication origin, *oriC*. *EMBO J.*, **4**, 1327–1332.

Migeon,B. (1994) X-chromosome inactivation: molecular mechanisms and genetic consequences. *Trends Genet.*, **10**, 230–235.

Raleigh,E.A. and Wilson,G. (1986) *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl Acad. Sci. USA*, **83**, 9070–9074.

Razin,A., and Cedar,H. (1991) DNA methylation and gene expression. *Microbiol. Rev.*, **55**, 451–458.

Riggs,A.D. and Pfeifer,G.P. (1992) X-chromosome inactivation and cell memory. *Trends Genet.*, **8**, 169–174.

Robinson,C.R. and Sligar,S. (1993) Molecular recognition mediated by bound water. A mechanism for star activity of the restriction endonuclease *Eco*RI. *J. Mol. Biol.*, **234**, 302–306.

Rosenberg,J.M. (1991) Structure and function of restriction endonucleases. *Curr. Opin. Struct. Biol.*, **1**, 114–122.

Rubin,R.A. and Modrich,P. (1977) *Eco*RI methylase. Physical and catalytic properties of the homogeneous enzyme. *J. Biol. Chem.*, **252**, 7265–7272.

Smith,D.W., Garland,A.M., Herman,G., Enns,R.E., Baker,T.A. and Zyskind,J.W. (1985) Importance of state of methylation of *oriC* GATC sites in initiation of DNA replication in *Escherichia coli*. *EMBO J.*, **4**, 1319–1326.

*Received on August 21, 1995; revised on January 5, 1996*