Published online 22 November 2011 *Nucleic Acids Research, 2012, Vol. 40, No. 6* ***2747–2758***
doi:10.1093/nar/gkr1057

Recognition of methylated DNA through methyl-CpG binding domain proteins

Xueqing Zou[1], Wen Ma[1,2], Ilia A. Solov'yov[1], Christophe Chipot[1,3] and Klaus Schulten[1,2,4,*]

[1]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, [2]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA, [3]Equipe de Dynamique des Assemblages Membranaires, UMR Centre National de la Recherche Scientifique/UHP 7565, Nancy Universite BP 239, Nancy, France and [4]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Received August 15, 2011; Revised October 25, 2011; Accepted October 26, 2011

ABSTRACT

DNA methylation is a key regulatory control route in epigenetics, involving gene silencing and chromosome inactivation. It has been recognized that methyl-CpG binding domain (MBD) proteins play an important role in interpreting the genetic information encoded by methylated DNA (mDNA). Although the function of MBD proteins has attracted considerable attention and is well characterized, the mechanism underlying mDNA recognition by MBD proteins is still poorly understood. In this article, we demonstrate that the methyl-CpG dinucleotides are recognized at the MBD–mDNA interface by two MBD arginines through an interplay of hydrogen bonding and cation-π interaction. Through molecular dynamics and quantum-chemistry calculations we investigate the methyl-cytosine recognition process and demonstrate that methylation enhances MBD–mDNA binding by increasing the hydrophobic interfacial area and by strengthening the interaction between mDNA and MBD proteins. Free-energy perturbation calculations also show that methylation yields favorable contribution to the binding free energy for MBD–mDNA complex.

INTRODUCTION

DNA methylation is an epigenetic modification of eukaryotic genomes, including plant genomes, involving the addition of a methyl group at the 5-position of cytosine in CG dinucleotides (1–4). It is now well established that DNA methylation plays an essential role in organismal development, e.g. controlling gene expression, X-chromosome inactivation and gene imprinting (5,6). In the cell, DNA methylation exerts its biological function in at least two ways. First, DNA methylation affects promoter regions to impede binding of transcriptional factors to the gene and, hence, directly interferes with gene activation (7–9). Second, DNA methylation attracts proteins that bind to the methyl-CpG (mCpG) steps and leads to the formation of silenced states of chromatin, which influence gene activity (10–15). A family of proteins that recognize mCpG steps in DNA, known as methyl-CpG binding domain (MBD) proteins, recruit co-repressor complexes that alter chromatin structure, thereby repressing transcription from non-methylated CpG promoters (16–19).

Biological functions of the MBD proteins in epigenetic regulation have been studied over several decades. At present, five proteins with mCpG-binding motifs have been identified in the MBD family, namely MBD1 (20,21), MBD2 (22,23), MBD3 (24), MBD4 (25) and MeCP2 (26–28). In mammals, MBD1, MBD2 and MeCP2 have been observed to bind to mCpG steps and to repress gene transcription. MBD4 binds to T:G mismatches caused by deamination of the methylated cytosine at mCpG sites and is involved in DNA repair (29). Mammalian MBD3 does not bind to the methylated DNA (mDNA) due to the lack of several important amino acids that are conserved in other MBD proteins (16).

All MBD proteins share a similar MBD of about 75 amino acid residues. Sequence alignment of MBD proteins reveals several highly conserved residues on the interface between MBD and methylated DNA (mDNA). Figure 1 shows that the most conserved region in the primary structure of the MBD is located between residues ARG22 and LYS46, including loop L1, strand β3, strand β4 and loop L2. The conserved residues of the MBD form the interfacial surface contacting mDNA and are responsible for recognition of the methyl-CpG steps.

*To whom correspondence should be addressed. Tel: +1 217 244 1604; Fax: +1 217 244 6078; Email: kschulte@ks.uiuc.edu

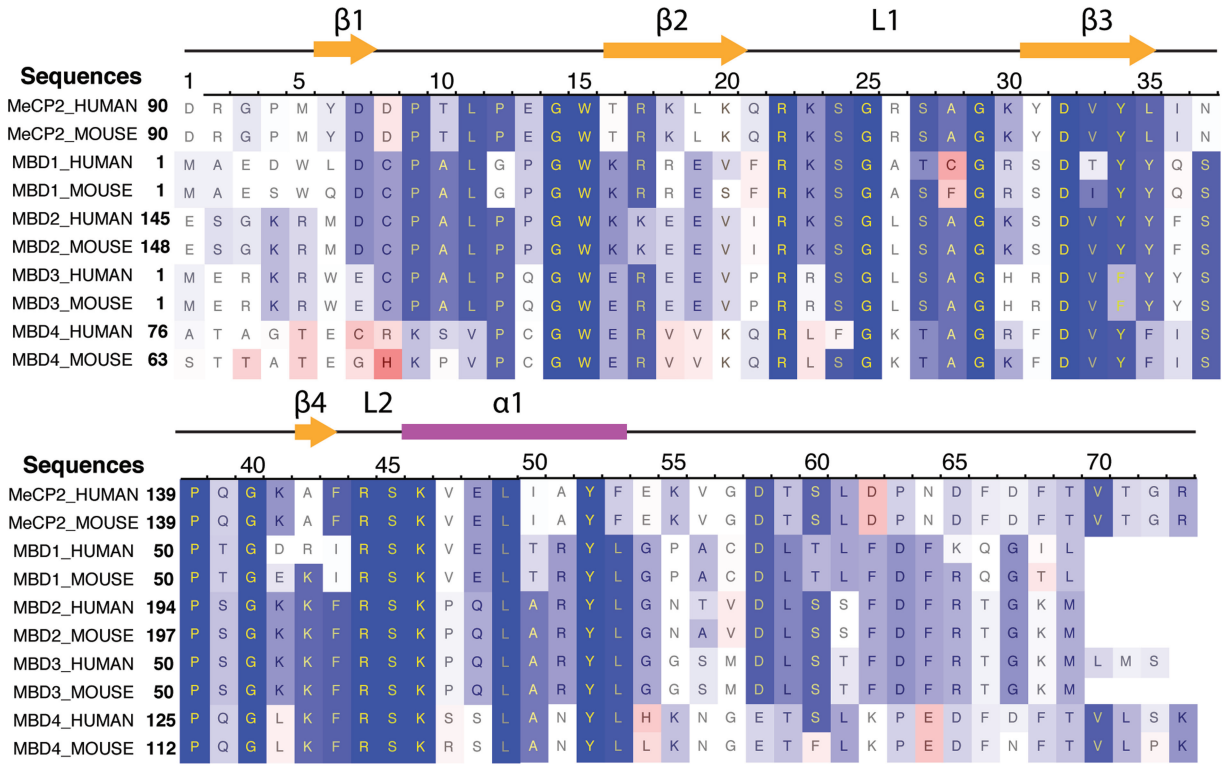

Figure 1. Sequence alignment of five methyl-CpG binding domain (MBD) proteins from human and mouse. A red-white-blue color scale is used to indicate sequence similarity. The most conserved residues are shown in dark blue, while the less conserved residues are indicated in red.

The NMR structures of MBD1, MBD2 and the X-ray structure of MeCP2 proteins bound to the mDNA provide useful information on how residues in the MBD contact mDNA (21,23,28). Figure 2 illustrates the MBD interaction with the mCpG steps in the major groove of mDNA. The methyl groups of methyl-cytosine are buried in the mDNA–MBD interface. The principal mCpG binding interface composed of hydrophilic and hydrophobic residues involves two types of interaction with mCpG steps (30): the polar atoms in MBD form hydrogen bonds with the polar atoms in the DNA, while the non-polar atoms in MBD form a hydrophobic patch around the methyl groups. However, a detailed picture of how the MBD recognizes mCpG steps over CpG steps, e.g. the influence of methyl groups on the thermodynamics of MBD–DNA binding, is still poorly characterized.

Binding of a protein to a specific DNA sequence involves complex binding networks of electrostatic and hydrophobic interactions. Earlier investigations suggested that the pyrimidine-guanine dinucleotide step is recognized by an interaction network in which an arginine is featured prominently through a coupled hydrogen bonding/cation-π interaction (31–34). The conformation composed of pyrimidine, guanine and arginine is usually referred to as the PYR ∴ ARG ∨ GUA stair motif (∴ denotes the cation-π interaction and ∨ denotes the hydrogen bonding), which includes three interactions: nucleobase stacking (arising between pyrimidine and guanine), hydrogen bonding (arising between guanine and arginine) and cation-π interaction (arising between pyrimidine and arginine) (35,36). The stair motif was observed in the structure of MBD2–mDNA and MeCP2–mDNA complexes, in which the arginine amino acids form hydrogen bonds with the guanine bases and concomitantly stack with the methyl-cytosines (23,28). Since the two arginines involved in the stair motifs are conserved in all MBD proteins (see Figure 1), it is likely that recognition of the mCpG dinucleotide by MBD proteins is mediated by stair motifs which are stabilized through methyl groups. Although previous studies have reported the two arginine forming hydrogen bonds with the guanines, the stacking interaction between the methyl-cytosines and the arginines has always been neglected.

In the present study, we performed molecular dynamics (MD) simulations (37) of the MBD1–DNA, MBD2–DNA and MeCP2–DNA complexes in order to reveal the general mechanism whereby the MBD recognizes the mCpG steps. The simulations demonstrate formation of the stable stair motif at the MBD–mDNA binding interface. Comparison between MBD–nDNA (nDNA denotes non-methylated DNA) and MBD–mDNA interactions shows that DNA methylation stabilizes the stair motif, thereby facilitating mCpG recognition by the MBD. The strong effect of methylation on reducing the solvent accessible surface at the MBD–mDNA interface is observed through calculation of the contact area between MBD protein and DNA. Quantum-chemistry analysis

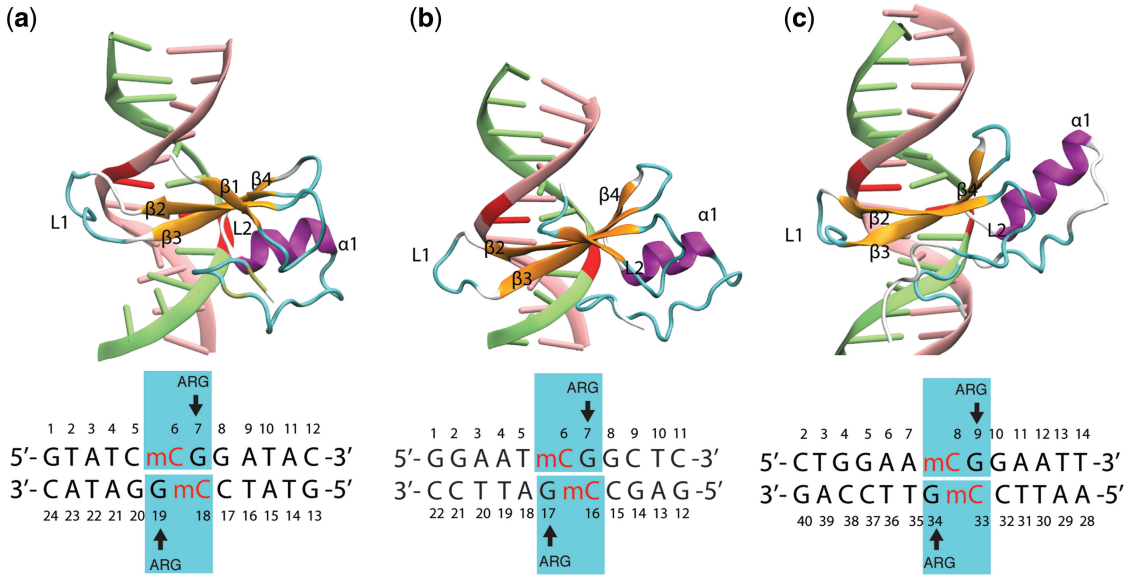

Figure 2. MBD proteins binding to mDNA. Shown are the structures of (a) MBD1-mDNA, (b) MBD2-mDNA and (c) MeCP2–mDNA complexes. The methyl-cytosines are indicated in red. The DNA sequence for each complex is shown below the corresponding image. The stair motif mCYT ∴ ARG ν GUA is highlighted (cyan rectangle) and the important motifs of the secondary structure, i.e. β-strands (orange), α-helix (purple) and loops (cyan), are labeled. The structures (**a**)–(**c**) are compared in Supplementary movies S1–S3.

characterizes quantitatively the influence of methylation on the stability of the MBD–mDNA complex. Free-energy perturbation (FEP) calculations determine the overall methylation-induced free-energy change in MBD–mDNA binding. Finally, MD simulations on mutants of MBD1 bound with mDNA reveal the important role of other residues in MBD–mDNA binding.

METHODS

We performed molecular dynamics simulations for three crystal structures, namely MBD1–mDNA, MBD2–mDNA and MeCP2–mDNA, to examine the mechanism of methyl-CpG recognition by the MBD of the MBD protein family. To characterize the role of the methyl groups in MBD–mDNA binding, we compared the thermodynamics of the binding surface of MBD1–mCpG and MBD1–CpG steps. Five mutants of MBD1 binding with methylated DNA were simulated to characterize the involvement of specific side groups in recognizing mCpG steps. *Ab initio* quantum-chemistry calculations were employed to account in electronic detail for the influence of methylation on enhancing the interaction between MBD and DNA.

Molecular dynamics simulations

Atomic models. Nine molecular complexes were studied in this work: (i) wild-type MBD1–mDNA, (ii) mutant MBD1–nDNA, (iii) mutant V20A MBD1–mDNA, (iv) mutant R22A MBD1–mDNA, (v) mutant Y34A MBD1–mDNA, (vi) mutant R44A MBD1–mDNA, (vii) mutant S45A MBD1–mDNA, (viii) wild-type MBD2–mDNA and (ix) wild-type MeCP2–mDNA (see Table 1). Atomic coordinates of wild-type MBD1–mDNA, see (i), were taken from the well-defined protein residues 3–69 in the NMR structure of the MBD of human MBD1 complexed with mDNA (Protein Data Bank entry code 1IG4) (21). The mutant structure for MBD1–nDNA, see (ii), was obtained by mutating the methylated cytosines to non-methylated ones. Five mutated structures of the MBD1–mDNA complex, see (iii)–(vii), were obtained by mutating the residues of interest in the wild-type complex using the MUTATOR plug-in in VMD (38). The MUTATOR plug-in removes the structural information of a given residue and generates new coordinates for the desired mutant amino acid based on the internal coordinates provided in standard CHARMM topology files. The atomic coordinates of wild-type MBD2–mDNA, see (viii), and MeCP2–mDNA, see (ix), were taken from the NMR structure of chicken MBD2 bound to mDNA (Protein Data Bank entry code 2KY8) (23) and the X-ray structure of human MeCP2 bound to mDNA (Protein Data Bank entry code 3C2I) (28), respectively.

The topology file of DNA and protein along with the missing hydrogen atoms was generated using the psfgen plug-in of VMD (39). Each complex was placed in a water box with 0.1 mol/l KCl added. The total size of the simulated systems lies in the 50 000–60 000 atom range.

Molecular dynamics. Simulations were performed using the program NAMD 2.7 with the CHARMM27 force field for DNA (40), the CHARMM22 force field for proteins with CMAP corrections (41,42) and the TIP3P water model (43). The parameters for 5-methylcytosine are available in the CHARMM27 force field (40). Periodic boundary conditions were assumed and the particle-mesh Ewald (PME) summation method was employed for evaluating Coulomb forces. The van der Waals (vdW) energy was calculated using a smooth cutoff of 12 Å.

Table 1. List of performed simulations

Index	DNA	Protein	Mutation	Time (ns)
(i)	Methylated	MBD1	–	100
(ii)	Non-methylated	MBD1	–	30
(iii)	Methylated	MBD1	V20A	30
(iv)	Methylated	MBD1	R22A	30
(v)	Methylated	MBD1	Y34A	30
(vi)	Methylated	MBD1	R44A	30
(vii)	Methylated	MBD1	S45A	30
(viii)	Methylated	MBD2	–	30
(ix)	Methylated	MeCP2	–	30

The integration time step was 1 fs. The temperature was kept at 310 K by applying Langevin forces with a damping coefficient of $0.1\,ps^{-1}$ (44) only to the oxygen atoms of water molecules.

Each simulated system was first energy-minimized for 2000 steps and then heated to 310 K in 4 ps. After heating, each simulated system was equilibrated for 500 ps with harmonic restraints applied to all protein and DNA atoms under NPT ensemble conditions using Nosé-Andersen Langevin piston pressure control (44,45), allowing the systems to acquire a constant volume. With restraints turned off each system was then subjected to 2 ns-equilibration under NVT ensemble conditions. Finally, a 30 ns [100 ns in case of simulation (i)] production MD simulation was carried out in the NVT ensemble. Table 1 lists all the simulations carried out in the present study.

Data analysis. Analysis and snapshots of the molecular structures from MD simulations were realized with VMD (38). To evaluate the binding strength of MBD to the mDNA, we analyzed the contact area, $\sigma(t)$, between protein and DNA through MD trajectories. $\sigma(t)$ is defined as

$$\sigma(t) = \frac{S_{\mathrm{prot}}(t) + S_{\mathrm{DNA}}(t) - S_{\mathrm{prot+DNA}}(t)}{2}, \quad (1)$$

where $S_{\mathrm{prot}}(t)$, $S_{\mathrm{DNA}}(t)$ and $S_{\mathrm{prot+DNA}}(t)$ denote the time-dependent solvent accessible surface area (SASA) of protein, DNA and protein–DNA complex, respectively. The SASA values of selected atoms were calculated using the built-in SASA function in VMD (38), which uses an assigned radius (van der Waals radius) for each atom and extends the radius to a user-defined value (here we used 1.4 Å) to find the points on a sphere that are exposed to solvent. The SASA function in VMD is based on the Shrake–Rupley algorithm (46).

To characterize the thermodynamics of arginine and guanine pairing, we monitored the time evolution of hydrogen bonding between arginine and guanine. A hydrogen bond is considered formed when the distance between a hydrogen atom and an acceptor atom is <2.5 Å.

Ab initio quantum chemistry calculations

The gas-phase interaction energy of the stair motif in the MBD1–DNA complex was determined using the quantum-chemistry package GAMESS (47), employing second-order Møller–Plesset perturbation theory (MP2) (36,48–50). The electronic wave function was expanded in a sum of Gaussian functions employing the 6-311++G(*d*, *p*) Pople-type basis set of triple-ζ accuracy (51). This basis set, which includes polarization *d* and *p* functions and double-diffuse functions (51), is important for a proper MP2 estimate of the interaction energy arising in the stair motif of the MBD–DNA interface (36). Tests for accuracy of different basis sets and different quantum-chemistry methods [MP2 (52), coupled-cluster single and double excitation (CCSD) (53) and density functional theory (DFT) (54–57)] are discussed in the Supplementary Data.

Calculations were performed for the mCYT106 ∴ ARG22 ∨ GUA107 and mCYT118 ∴ ARG44 ∨ GUA119 stair motifs in the MBD1–mDNA complex. Each stair motif includes a methylated/no-methylated cytosine, a guanine and an arginine, along with hydrogen atoms terminating the side chains of these residues. The mCYT-ARG-GUA stair motif contains 51 atoms; the CYT-ARG-GUA stair motif contains 48 atoms. The geometry of the stair motifs used in the single-point energy calculations was extracted from the structures of the MBD1–DNA complex arising in the molecular dynamic simulations. Optimizations prior to the *ab initio* calculation of the MBD1–mDNA complex in water was done starting from selected conformations from the 20–30 ns MD simulation interval, and using the conjugate gradient energy minimization method with the CHARMM force field (41) in NAMD (39). Structure optimization is necessary for quantum-chemistry calculations, which should be performed on an equilibrated molecular system to reveal the energy difference in the binding energy of a MBD to methylated and non-methylated DNA, as the difference is relatively small. If non-equilibrium structures were used, the small energy difference may become masked by energy contributions from unfavorable orientation of interacting residues. The optimization was carried out for 50 000 steps. After 50 000 steps of optimization the total energy of the system did not change beyond 0.0001 kcal/mol ($\sim 2 \times 10^{-9}$ kcal/mol per atom) during one step, implying convergence of the studied structure to a local energy minimum.

In order to examine the influence of the methyl groups on stair motif energetics, the methylated cytosines in the selected snapshots from the MD simulation were substituted by non-methylated cytosine residues and the derived structures were then optimized using NAMD (39) before performing quantum-chemistry single-point energy calculations.

To characterize the MBD1–mDNA and MBD1–nDNA binding we considered the three body interaction energies in the mCYT ∴ ARG ∨ GUA and nCYT ∴ ARG ∨ GUA stair motifs which are defined as

$$E_{\mathrm{m}}^{x} = E(\mathrm{mCYT} \therefore \mathrm{ARG} \vee \mathrm{GUA})^{x} - E(\mathrm{mCYT})^{x} - E(\mathrm{ARG})^{x} - E(\mathrm{GUA})^{x}. \quad (2)$$

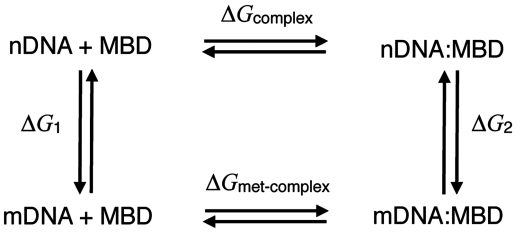

Figure 3. Thermodynamic cycle describing DNA methylation in the DNA–MBD complex. The left vertical transition was performed for free DNA (in the absence of the protein) in water; the right vertical transition was performed by mutating DNA to mDNA in the hydrated, native complex. 'DNA+MBD' represents the free DNA and the free MBD, while 'DNA:MBD' represents the DNA–MBD complex.

$$E_{\mathrm{n}}^{x} = E(\mathrm{nCYT} \therefore \mathrm{ARG} \vee \mathrm{GUA})^{x} - E(\mathrm{nCYT})^{x} - E(\mathrm{ARG})^{x} - E(\mathrm{GUA})^{x} \quad (3)$$

Here, the subscripts m and n denote the methylated and the non-methylated cases of the stair motif, respectively. The superscript $x = 1, 2$ denotes the CYT106 $\therefore$ ARG22 $\vee$ GUA107 stair motif ($x = 1$) and the CYT118 $\therefore$ ARG44 $\vee$ GUA119 stair motif ($x = 2$); E(CYT $\therefore$ ARG $\vee$ GUA) is the total energy of the stair motif, while E(CYT), E(ARG) and E(GUA) are the energies of its individual components, corresponding to the energy of cytosine, guanine and arginine, respectively.

From Equations (2) and (3) the change of the binding energy in the MBD1–DNA stair motifs due to DNA methylation can be calculated as

$$\Delta E^{x} = E_{\mathrm{m}}^{x} - E_{\mathrm{n}}^{x}. \quad (4)$$

The *ab initio* quantum chemistry calculations performed include a correction of the basis set superposition error (BSSE) using the Boys and Bernardi function counterpoise (FCP) method (58); such correction is crucial for the accurate calculation of E_{m}^{x} and E_{n}^{x} in Equations (2) and (3) (36,59).

Alchemical FEP calculations

FEP calculations (60–63) were performed to determine the free-energy change of MBD1–DNA binding upon cytosine methylation using NAMD. According to the thermodynamic cycle shown in Figure 3, the net free-energy change for the MBD–DNA $\rightarrow$ MBD–mDNA transformation ($\Delta\Delta G$) can be written

$$\Delta\Delta G = \Delta G_2 - \Delta G_1 = \Delta G_{\mathrm{met-complex}} - \Delta G_{\mathrm{complex}}. \quad (5)$$

Here, ΔG_1 was calculated by mutating non-methylated free DNA into methylated free DNA in water and ΔG_2 was calculated by mutating the nDNA–protein complex into mDNA–protein in water. We performed the transformations in both directions and used the Bennett acceptance-ratio (BAR) estimator (63,64) to obtain the maximum-likelihood value of the free energy together with the statistical error.

The FEP calculations were carried out using 25 intermediate states ($\Delta\lambda = 0.04$), which represent a total simulation time of 40 ns for each unidirectional simulation. λ is a general-extent parameter (65) introduced to transform the system between the reference, non-methylated state, i.e. the state with $\lambda = 0$, and the perturbed, methylated state, i.e. the state with $\lambda = 1$. The ParseFEP plug-in of VMD (63) was employed to determine the BAR estimator alongside with its variance by combining forward and backward transformations.

RESULTS

In the following section the interaction of methylated and non-methylated DNA with MBD is studied in detail. Particular attention is paid to the structural and energetic analyses of the CYT $\therefore$ ARG ν GUA stair motifs at the MBD–DNA interface. Combination of classical MD simulations and quantum-chemistry calculations is used to explain mDNA recognition by MBD proteins.

Stair motifs at the MBD–mDNA interface. Structures of MBD1–mDNA, MBD2–mDNA and MeCP2–mDNA complexes reveal that the conserved residues ARG22 and ARG44 form hydrogen bonds with guanines in the mCpG steps (21,23,28). The NMR structure of MBD1–mDNA shows that either ARG22 or ARG44 form a single hydrogen bond with guanine, while the X-ray structure of MeCP2–mDNA and the NMR structure of MBD2–mDNA clearly show that two hydrogen bonds are formed between each arginine and guanine. The crystal structures showing that ARG22 and ARG44 interact with mCpG steps in MBD1–mDNA, MBD2–mDNA and MeCP2–mDNA complex are provided in Supplementary Figure S1a, c and e. To investigate the role of ARG22 and ARG44 in recognizing the mCpG steps, we conducted a 100-ns molecular dynamics simulation [simulation (i), see Table 1] of the NMR structure of the human MBD1–mDNA complex (21). Interestingly, within a very short time of simulation (~1 ns), the side chains of ARG22 and ARG44 rotated, forming two hydrogen bonds with GUA107 and GUA119, respectively, as shown in Figure 4a. The length of the four hydrogen bonds formed between ARG22 and GUA107 as well as between ARG44 and GUA119 is ~1.7 Å (see Figure 4b). The fluctuation of the hydrogen bond length is small, indicating that the hydrogen bonding between arginine and guanine is stable (see Figure 4b). 30-ns MD simulations were performed for the NMR structure of chicken MBD2–mDNA [simulation (viii), see Table 1] (23) and the X-ray structure of human MeCP2–mDNA complexes [simulation (ix), see Table 1] (28). In both complexes, the four hydrogen bonds between the arginines and the guanines, observed in the crystal structures, remained stable through the entire MD simulations (see Supplementary Figures S1–S3).

Protein binding often induces structural changes in the contacted DNA (66,67). In the case of the MBD–mDNA complex, ARG22 and ARG44 insert into the major groove of the DNA, forming hydrogen bonds with the guanines and shifting them into the minor groove (see

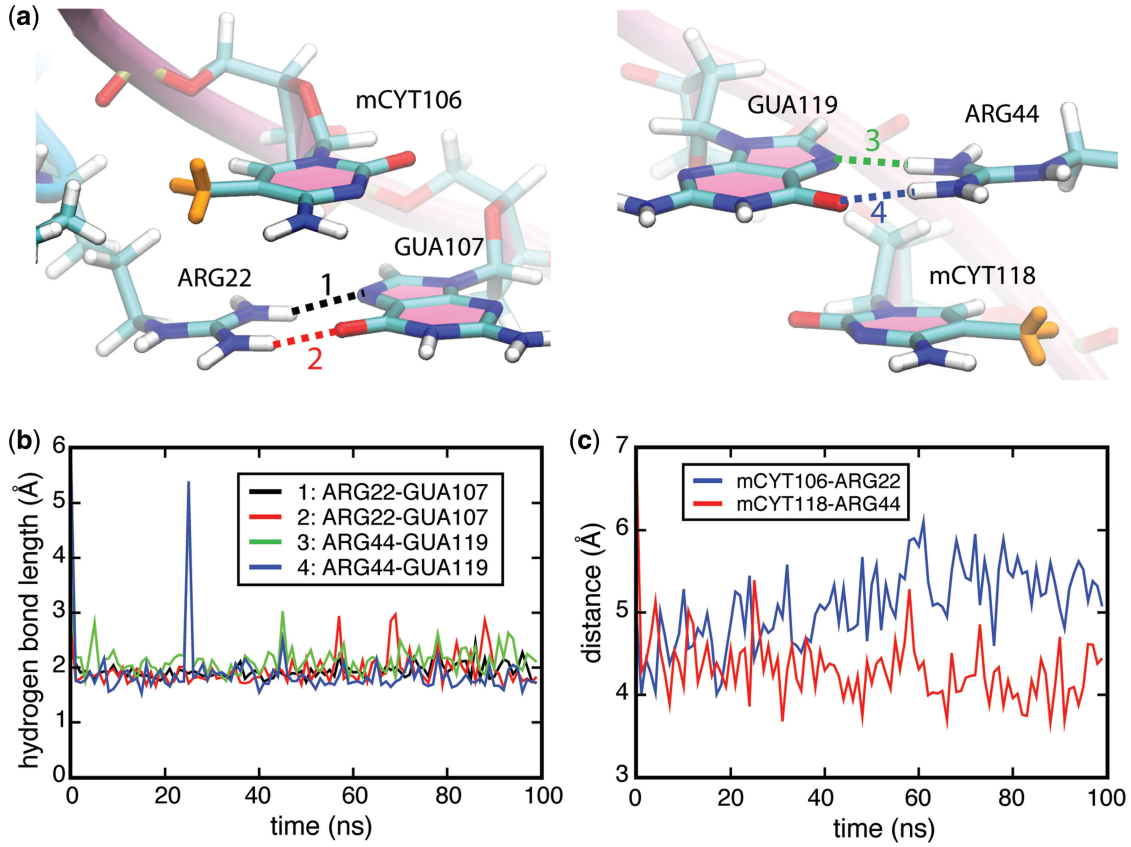

Figure 4. Stair motifs in the interface of MBD1–mDNA. (**a**) Orientation of stair motifs mCYT106 ∴ ARG22 ∨ GUA107 (left) and mCYT118 ∴ ARG44 ∨ GUA119 (right) in the MBD1–mDNA complex. (**b**) Time evolution of hydrogen bonding between the H- and N-atoms in the ARG22-GUA107 pair (black line), H- and O-atoms in the ARG22-GUA107 pair (red line), H- and N-atoms in the ARG44-GUA119 pair (green line), H- and O-atoms in the ARG44-GUA119 pair (blue line). (**c**) Time evolution of the distance between the center of mass of mCYT106 residue in the DNA and ARG22 residue in the MBD1 protein (blue line), and the distance between the center of mass of mCYT118 residue in the DNA and ARG44 residue in the MBD1 protein (red line). mCYT indicates the methylated cytosine residues.

Supplementary Figure S4). This shift decreases the stacking interaction between guanine and cytosine in mCpG steps, but allows the positively charged ARG22 and ARG44 to form cation-π interactions with cytosine bases, in addition to forming hydrogen bonds to guanines. Figure 4c shows that the distance between the center of mass of the arginine side chains and the center of mass of the methyl-cytosine base is well preserved during the 100-ns simulation. Two snapshots in Figure 4a illustrate a coupled hydrogen bonding/cation-π interaction between the arginine and the mCpG-steps in the MBD1–mDNA complex. The spatial configurations of mCYT106 ∴ ARG22 ∨ GUA107 and mCYT118 ∴ ARG44 ∨ GUA119 adopt a so-called stair motif. Similar stair motifs are formed and remain stable at the MBD2–mDNA and MeCP2–mDNA interface, as shown in Supplementary Figures S2–S3.

Methylation stabilizes stair motif. The results presented above show that arginine residues recognize mCpG steps through simultaneous cation-π and hydrogen bonding interactions. Arginine residues are expected to interact with CpG steps at the MBD–nDNA interface as well (33). However, mutational studies in which mCYT106 and mCYT118 were substituted by nCYT106 and nCYT118 (nCYT denotes non-methylated cytosine) revealed that MBD1 does not bind to non-methylated DNA (21), indicating that the 5-methyl groups on the methyl-cytosines are critical for the formation of the stair motif in MBD–mDNA binding interfaces. The influence of methylation on the stability of the stair motif was investigated through a 30-ns MD simulation [simulation (ii), see Table 1] for the MBD1–nDNA complex, where the mCYT106 and mCYT118 nucleotides were mutated into two non-methylated cytosines. The analysis of the MD simulation trajectories showed that the stair motifs in the MBD1–nDNA complex arising at mCYT106 and mCYT118 remained stable in both places for 30 ns. However, the average root-mean-square deviation (RMSD) of the distance between ARG22 and nCYT106 is 1.8 Å, while for ARG22 and mCYT106 it is 1.0 Å; the RMSD between ARG44 and nCYT118 is 1.5 Å and for ARG22 and mCYT118 it is 0.8 Å, indicating that the interactions between ARG and mCYT are more stable than those between ARG and nCYT (see Figure 5). Snapshots in Figure 5 demonstrate the spatial configurations of ARG-nCYT106, ARG-mCYT106, ARG-nCYT118 and ARG-mCYT118. The performed

analysis of the MBD1–DNA complex suggests two contributions to MBD–mDNA recognition as discussed below.

The first contribution stems from the contact area between mDNA and the MBD, which increases upon cytosine methylation. The contact area, σ, defined in Equation (1), is the surface buried at an interface between DNA and protein, which affects the binding free energy (68) and, therefore, contributes to the stability of the DNA–protein complex. In the case of the MBD–mDNA binding, the methyl groups lead to an increase of the contact area between cytosine and arginine in the stair motif, as illustrated in the snapshots in Figure 5 and Figure 6. Moreover, since several other amino acids in the MBD have direct contact with the mCpG steps, e.g. ASP32 and TYR34, replacing a hydrogen atom with a methyl group likely increases the total interfacial area between DNA and MBD. Indeed, the contact area between mCYT106 and MBD1 is ∼200 Å^2, the contact area between mCYT118 and MBD1 is ∼150 Å^2 and the total contact area between mDNA and MBD1 is ∼804 Å^2 (see Figure 6). After altering mCYT to nCYT, the values of the contact area decrease significantly: the contact area between nCYT106 and MBD1 drops to ∼170 Å^2, the contact area between nCYT118 and MBD1 decreases to ∼130 Å^2 and the total contact area between DNA and MBD1 decreases to ∼700 Å^2.

The second contribution to MBD–mDNA recognition stems from the formation of the mCYT ∴ ARG ∨ GUA stair motif that arises twice in the MBD–mDNA interface and is stabilized by cytosine methylation. The motifs exhibit a favorable contact potential energy between MBD and mDNA. In fact, the stair motif induces a cation-π interaction between the aromatic cytosine and the positively charged arginine, and two hydrogen bonds between arginine and guanine. The classical force field used in our MD simulations does not accurately account for polarizability and charge-transfer effects, especially those arising through cation-π interaction (69–72). The quantum many-body effects, i.e. when the interactions between each monomer in the stair motif are non-additive, are also not included in MD simulations (36).

For this reason we performed quantum chemistry calculations of the interaction energy in the mCYT ∴ ARG ∨ GUA and nCYT ∴ ARG ∨ GUA stair motifs in order to determine the impact of the methyl group on the stair motif stabilization. The interaction energy differences ΔE^x [see Equation (4)] between methylated and non-methylated stair motif at $x = 1$ (CYT106 ∴ ARG22 ∨ GUA107) and $x = 2$ (CYT118 ∴ ARG44 ∨ GUA119) are presented in Figure 7. The average value of ΔE^1 is -0.81 ± 0.33 kcal/mol; the fluctuations arise mainly due to the fact that, after optimizing the structure of the MBD1–DNA complex in the solvent, the structures of the stair motif in the complex are at different local minima, with slightly different potential energies. The average value of ΔE^2 is -0.80 ± 0.44 kcal/mol, i.e. very close to ΔE^1. Thus, the total change of the interaction

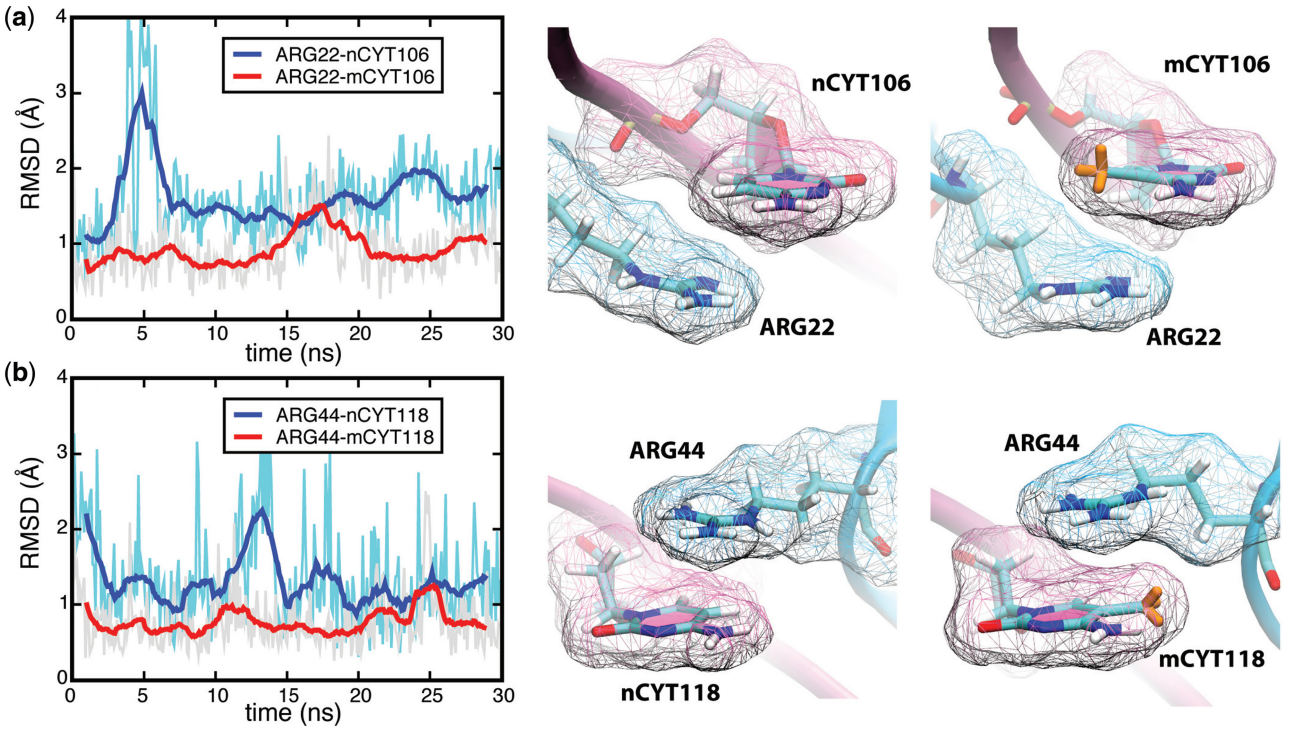

Figure 5. Stabilization of the MBD1–mDNA complex due to CYT106 and CYT118 methylation. (**a**) Comparison of the RMSD for ARG22-nCYT106 (blue line) and ARG22-mCYT106 (red line). (**b**) Comparison of the RMSD for ARG44-nCYT118 (blue line) and ARG22-mCYT118 (red line). The structure of the relevant part of the methylated and non-methylated MBD1–DNA complexes is shown on the right. The van der Waals surface of residues is shown in wireframe representation. Two movies are provided in Supplementary Data showing the dynamics of mCYT106 ∴ ARG22 ∨ GUA107 (Suplementary Movie S4) and CYT106 ∴ ARG22 ∨ GUA107 (Supplementary Movie S5).

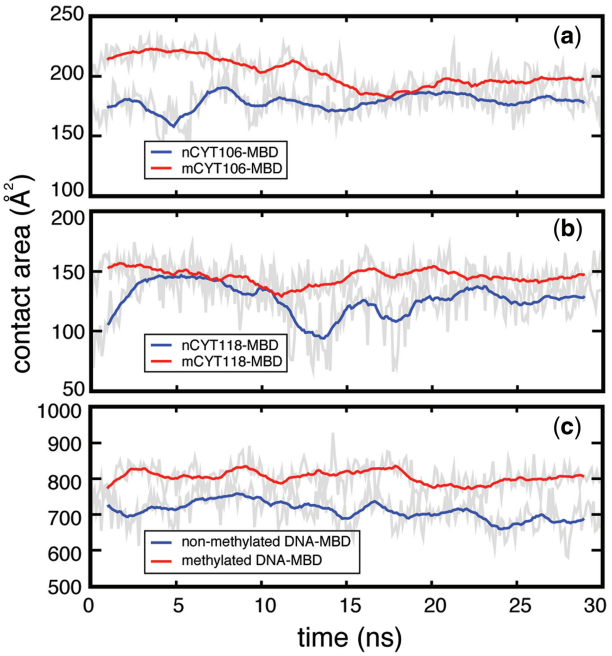

Figure 6. Effect of CYT106 and CYT118 methylation on the MBD1–DNA contact area. Time evolution of the surface contact area calculated, using Equation (1), between (**a**) CYT106 and MBD1, (**b**) CYT118 and MBD1 and (**c**) DNA and MBD1. Red and blue lines correspond to the case of methylated and the non-methylated DNA, respectively.

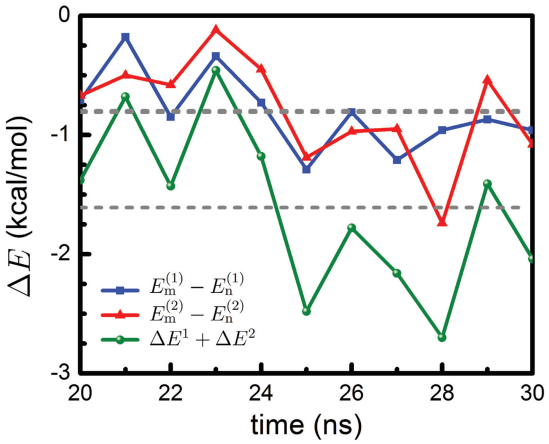

Figure 7. Influence of cytosine methylation on MBD1 binding energy. The energy difference ΔE is obtained from quantum chemistry calculations for the two stair motifs arising in the MBD1–DNA interface (see Figure 4). The total energy $\Delta E = \Delta E^1 + \Delta E^2$ [see Equation (4)] shows the three-body interaction energy difference of the mCYT ∴ ARG ∨ GUA stair motifs on the MBD1–DNA interface calculated for methylated (E_m) and non-methylated (E_n) cytosine. The blue line (squares) shows the interaction energy calculated for the mCYT106 ∴ ARG22 ∨ GUA107 stair motif, the red line (triangles) shows the interaction energy calculated for the mCYT118 ∴ ARG44 ∨ GUA119 stair motif, while the green line (dots) shows the combined energy difference.

energy in the two stair motifs is -1.61 ± 0.77 kcal/mol. The corresponding three-body interaction energies are given in Supplementary Table S1.

In order to test whether the MP2 method properly describes the interaction between CYT106 and ARG22, two alternative quantum-chemistry methods (CCSD and DFT) were used for the calculation of the energy differences between the methylated and the non-methylated states of the CYT ∴ ARG ∨ GUA stair motif. As shown in Supplementary results (see Supplementary Table S2 and Figure S7), the results obtained using the CCSD method (53) are consistent with those at the MP2 level, while the density functional theory does not describe appropriately the long-range dispersion interactions (73,74), i.e. a part of van der Waals interactions. This result is expected; the discrepancy between CCSD and DFT methods suggests that the dispersion interaction contributes strongly to stair motif stabilization. The three-body (mCYT ∴ ARG ∨ GUA) and two-body (mCYT106 ∴ ARG22) interaction energy differences upon methylation were also calculated using the CHARMM force field (see Supplementary Figures S5 and S8). Consistent with QM calculations, the results show a stabilization effect of cytosine methylation on the studied system.

Free-energy contributions of DNA methylation to MBD1–DNA complex stabilization. The protein–DNA binding affinity is affected by different interactions. In order to account for the total effect of methylation on the stability of the MBD–DNA complex, a thermodynamic analysis of the changes in binding free-energy is needed. For this purpose, FEP calculations were performed to characterize the free-energy contribution of the methyl group to the MBD–DNA complex. The free-energy changes for the MBD1–DNA complex and free DNA are shown in Supplementary Figure S6. The BAR (see 'Methods' section) free-energy difference for the free DNA was determined to be $\Delta G_1 = 54.9 \pm 0.0$ kcal/mol, with -0.12 kcal/mol hysteresis between forward and backward transition; for the MBD–DNA complex, the free-energy difference was determined to be $\Delta G_2 = 53.7 \pm 0.1$ kcal/mol, with 0.2 kcal/mol hysteresis between forward and backward transition. The hysteresis can be interpreted as an estimate of the systematic error of the FEP calculation, restricted to finite-length errors, i.e. not accounting for force-field inaccuracies or inadequate algorithms (63). Following Equation (5), one obtains $\Delta\Delta G = -1.2 \pm 0.1$ kcal/mol, hence, indicating that methylation makes a favorable contribution to the binding free energy.

Binding affinity of mutated MBD for mDNA. Previous mutagenesis studies have shown that several residues of the MBD play key roles in binding to mDNA. Ohki *et al.* (21) suggested that the recognition of mCpG steps is due to combined hydrophobic and polar contacts made by five residues of MBD, namely VAL20, ARG22, TYR34, ARG44 and SER45. The hydrophobic region of each residue forms a hydrophobic patch around the methylation sites, while the polar moieties form hydrogen bonds with the polar atoms of DNA. Mutation of these residues can significantly reduce the binding affinity for mDNA (21). Consistent with

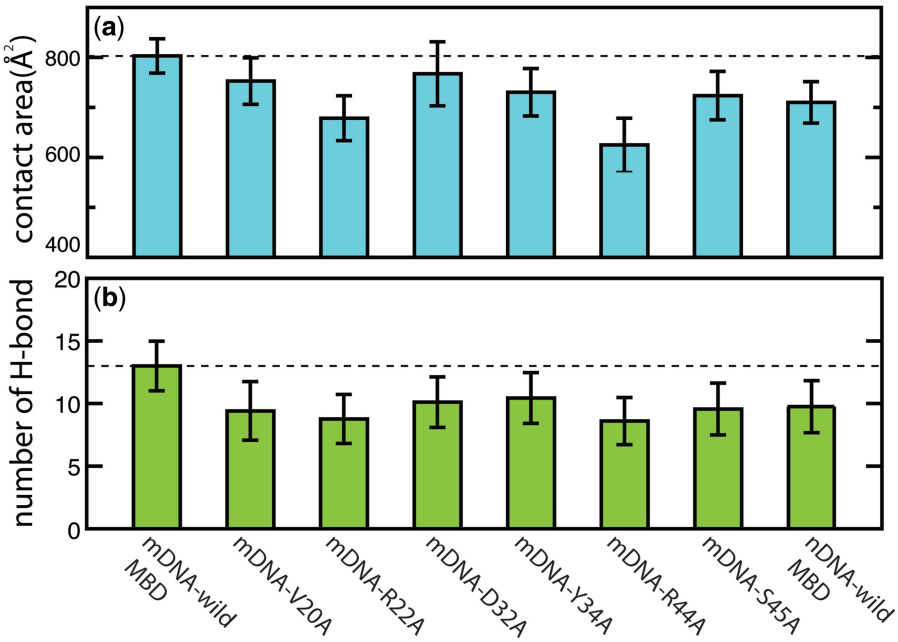

Figure 8. Mutation effects on characteristics of the MBD1–DNA interface. (**a**) Comparison of the contact area between DNA and mutants of MBD1. The dashed line shows the average contact area for wild-type MBD1–mDNA. (**b**) Comparison of the number of hydrogen bonds between DNA and mutants of MBD1. The dashed line shows the average number of interfacial hydrogen bonds for wild-type MBD1–mDNA.

mutagenesis studies, MD simulations revealed that mutants of MBD1 cause a decrease of the contact interfacial mDNA binding area, reducing the mCpG-binding affinity. Figure 8a shows that the average contact area between mDNA and wild MBD1 is 804 Å^2, while for the V20A mutant the average contact area is 752 Å^2, for the R22A mutant 678 Å^2, for the D32A mutant 767 Å^2, for the Y34A mutant 730 Å^2, for the R44A mutant 625 Å^2 and for the S45A mutant 723 Å^2. Remarkably, mutations of ARG22 and ARG44 to alanines decrease the average contact area significantly, namely by 150 Å^2, which again emphasizes the important role of ARG22 and ARG44 in mCpG recognition. We also monitored hydrogen bonding between mDNA and MBD. Figure 8b shows the average number of interfacial hydrogen bonds for wild-type MBD1 MBD (13 bonds on average), V20A mutant (9 bds.), R22A mutant (9 bds.), D32A mutant (10 bds.), Y34A mutant (10 bds.), R44A mutant (9 bds.) and S45A (10 bds.). MBD mutation is found to reduce the number of hydrogen bonds between MBD1 and mDNA. The R22A and R44A mutants are seen to have the largest impact on reducing hydrogen bonding with mDNA. The last column in the histogram in Figure 8b depicts the average number of hydrogen bonds in non-methylated DNA bound to wild-type MBD1. In the case of the MBD–DNA complex, although methyl groups are not involved in any hydrogen bonding, removing two methyl groups leads to a decrease of the MBD–DNA contact area (Figure 6), reducing the number of interfacial hydrogen bonds. We note that prior studies have also shown that mutation causes structural rearrangement at the protein–DNA interface, changing DNA-binding affinity (75,76).

DISCUSSION

In the present study, we investigated the role of the stair motif that combines arginine–guanine hydrogen bonding and arginine–cytosine cation-π stacking in the recognition of mCpG steps by MBD. Sequence alignment of human and mouse MBD family proteins shows that two arginine amino acids (ARG22 and ARG44) involved in stair motifs formation are strictly conserved, suggesting that these arginines play a crucial role in MBD–mDNA binding. Indeed, MD simulations and quantum-chemistry calculations, the latter restricted to the stair motif formed by arginine and mCpG steps at mCYT106 and mCYT118, show that methylated cytosine renders the stair motif more stable than does non-methylated cytosine. FEP calculations take into account the energy costs of methylation-induced deformations at the MBD–DNA binding interface, e.g. the displacement of mCpG step guanine to the minor groove, and the potential energy change in the stair motif, e.g. the change in cation-π interaction energy between arginine and cytosine upon methylation, quantifying the overall effect of DNA methylation on MBD–mDNA complex stability.

Our results quantified also that methylation increases the buried hydrophobic surface between MBD and DNA by ~100 Å^2. Previous theoretical and experimental studies suggested that decreasing the hydrophobic surface by 1 Å^2 reduces the interface stabilization free energy by 15 ± 1.2 cal/mol (68,77). Therefore, in the case of the MBD–mDNA complex, methylation contributes about 100×15 cal/mol = 1.5 kcal/mol to the stability of MBD–mDNA binding, which is close to the value (−1.2 ± 0.1 kcal/mol) obtained through FEP calculations.

The recognition of the 5′-pyrimidine-purine-3′ step via stair motifs through arginine residues has been observed in many protein–DNA complexes (33,36). An example is the sporulation-specific transcription factor Ndt80. Lamoureux *et al*. (32–34) reported that two 5′-TpG-3′ steps are recognized by Ndt80 via two arginines; the non-polar region of the arginines forms a hydrophobic half-pocket for the methyl group of a thymine, while the polar region contacts the guanine bases. Interestingly, the methyl group in the thymine was found to be crucial for the cation-π interaction between the thymine and the arginine, as substitution of the thymine with uracil, which lacks a methyl group, effectively results in a significant loss of Ndt80 binding affinity (32–34). Similarly, the 5-methyl group on methyl-cytosine is crucial for MBD proteins to recognize mCpG steps. However, previous studies did not address how the methyl groups in DNA lead to such a significant protein binding specificity, leaving a gap for speculation. Filling this gap, we have demonstrated in the present study that the effects of the 5-methyl group arise from a change of binding affinity due to an increased contact area and cation-π interaction.

Binding of a protein to a particular site in DNA involves guidance through protein tertiary structure, DNA local conformation and DNA sequence (78,79). It is worth noting that the binding affinities of MBD proteins for mDNA are sequence-dependent as they target different methylated promoters (17,80). For example, MeCP2 binds more efficiently to mCpG steps with adjacent A/T bases (81), while MBD1 strongly prefers an mCpG step in the context of TmCGCA and TGmCGCA *in vitro* and *in vivo* (82). A recent NMR study of MBD2 bound to mDNA revealed that MBD2 has high affinity for an mCGG sequence (23). Earlier studies suggested that the binding affinities of MBD proteins for different mDNA sequences can be attributed to specific residues within the MBD. Future investigations on specific binding properties of MBD proteins at atomic resolution are needed to explain the observed sequence specificities.

In this article we have addressed an important question in epigenetics, namely how MBD proteins are capable of recognizing methylated DNA. Although we answered some questions, others remain open, such as those regarding the mDNA interaction with other mDNA-binding proteins, e.g. Kaiso protein, the influence of the solvent on the MBD–mDNA interaction, temperature and pressure effects on mDNA binding. Epigenetics is still a long way from a thorough understanding of how methylated DNA conducts its functions in the cell and further studies should be initiated to learn more about the crucial, although subtle mDNA recognition process.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–2, Supplementary Figures 1–8, Supplementary Movies 1–5 and Supplementary Results.

ACKNOWLEDGEMENTS

The authors gladly acknowledge supercomputer time provided by the Texas Advanced Computing Center and the National Center for Supercomputing Applications via TeraGrid Resource Allocation Committee grant MCA93S028. I.S. acknowledges support as a Beckman Fellow. C.C. acknowledges the Centre National de la Recherche Scientifique for travel support through its Programme International de Coopération Scientifique.

FUNDING

Funding for open access charge: National Institutes of Health (P41-RR005969, R01 GM073655); National Science Foundation (PHY0822613).

Conflict of interest statement. None declared.

REFERENCES

1. Bird,A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
2. Razin,A. and Riggs,A.D. (1980) DNA methylation and gene function. *Science*, **210**, 604–610.
3. Rottach,A., Leonhardt,H. and Spada,F. (2009) DNA methylation-mediated epigenetic control. *J. Cell. Biochem.*, **108**, 43–51.
4. Raj,S., Bräutigam,K., Hamanishi,E.T., Wilkins,O., Thomas,B.R., Schroeder,W., Mansfield,S.D., Plant,A.L. and Campbell,M.M. (2011) Clone history shapes *Populus* drought responses. *Proc. Natl. Acad. Sci. USA*.
5. Bird,A.P. (1992) The essentials of DNA methylation. *Cell*, **70**, 5–8.
6. Jones,P.A. and Takai,D. (2001) The role of DNA methylation in mammalian epigenetics. *Science*, **293**, 1068–1070.
7. Kass,S.U., Pruss,D. and Wolffe,A.P. (1997) How does DNA methylation repress transcription? *Trends Gen.*, **13**, 444–449.
8. Klose,R.J. and Bird,A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.*, **31**, 89–97.
9. Maunakea,A.K., Nagarajan,R.P., Bilenky,M., Ballinger,T.J., D'Souza,C., Fouse,S.D., Johnson,B.E., Hong,C., Nielsen,C., Zhao,Y. *et al*. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, **466**, 253–257.
10. Keshet,I., Lieman-Hurwitz,J. and Cedar,H. (1986) DNA methylation affects the formation of active chromatin. *Cell*, **44**, 535–543.
11. Hashimshony,T., Zhang,J., Keshet,I., Bustin,M. and Cedar,H. (2003) The role of DNA methylation in setting up chromatin structure during development. *Nat. Genet.*, **34**, 187–192.
12. Lorincz,M.C., Dickerson,D.R., Schmitt,M. and Groudine,M. (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat. Struct. Biol.*, **11**, 1068–1075.
13. Lopez-Serra,L. and Esteller,M. (2008) Proteins that bind methylated DNA and human cancer: reading the wrong words. *Br. J. Cancer*, **98**, 1881–1885.
14. Choy,J.S., Wei,S., Lee,J.Y., Tan,S., Chu,S. and Lee,T.-H. (2010) DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.*, **132**, 1782–1783.
15. Rottach,A., Frauer,C., Pichler,G., Bonapace,I.M., Spada,F. and Leonhardt,H. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.*, **38**, 1796–804.
16. Hendrich,B. and Bird,A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.*, **18**, 6538–6547.
17. Fraga,M.F., Ballestar,E., Montoya,G., Taysavang,P., Wade,P.A. and Esteller,M. (2003) The affinity of different MBD proteins for

a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.*, **31**, 1765–1774.
18. Rauch,C., Trieb,M., Wibowo,F.R., Wellenzohn,B., Mayer,E. and Liedl,K.R. (2005) Towards an understanding of DNA recognition by the methyl-CpG binding domain 1. *J. Biomol. Struct. Dyn.*, **22**, 695–705.
19. Inomata,K., Ohki,I., Tochio,H., Fujiwara,K., Hiroaki,H. and Shirakawa,M. (2008) Kinetic and thermodynamic evidence for flipping of a methyl-CpG binding domain on methylated DNA. *Biochemistry*, **47**, 3266–3271.
20. Ohki,I., Shimotake,N., Fujita,N., Nakao,M. and Shirakawa,M. (1999) Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. *EMBO J.*, **18**, 6653–6661.
21. Ohki,I., Shimotake,N., Fujita,N., Jee,J.-G., Ikegami,T., Nakao,M. and Shirakawa,M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, **105**, 487–497.
22. Rupon,J.W., Wang,S.Z., Gaensler,K., Lloyd,J. and Ginder,G.D. (2006) Methyl binding domain protein 2 mediates γ-globin gene silencing in adult human βYAC transgenic mice. *Proc. Natl Acad. Sci. USA*, **103**, 6617–6622.
23. Scarsdale,J.N., Webb,H.D., Ginder,G.D. and Williams,D.C.J. (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.*, **39**, 6741–6752.
24. Brown,S.E., Suderman,M.J., Hallett,M. and Szyf,M. (2008) DNA demethylation induced by the methyl-CpG-binding domain protein MBD3. *Gene*, **420**, 99–106.
25. Screaton,R.A., Kiessling,S., Sansom,O.J., Millar,C.B., Maddison,K., Bird,A., Clarke,A.R. and Frisch,S.M. (2003) Fas-associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link between genome surveillance and apoptosis. *Proc. Natl Acad. Sci. USA*, **100**, 5211–5216.
26. Meehan,R.R., Lewis,J.D., McKay,S., Kleiner,E.L. and Bird,A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*, **58**, 499–507.
27. Wakefield,R.I.D., Smith,B.O., Nan,X., Free,A., Soteriou,A., Uhrin,D., Bird,A.P. and Barlow,P.N. (1999) The solution structure of the domain from MeCP2 that binds to methylated DNA. *J. Mol. Biol.*, **291**, 1055–1065.
28. Ho,K.L., McNae,I.W., Schmiedeberg,L., Klose,R.J., Bird,A.P. and Walkinshaw,M.D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell*, **29**, 525–531.
29. Bird,A.P. and Wolffe,A.P. (1999) Methylation-induced repression–belts, braces, and chromatin. *Cell*, **99**, 451–454.
30. Clouaire,T. and Stancheva,I. (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell. Mol. Life Sci.*, **65**, 1509–1522.
31. Luscombe,N.M., Laskowski,R.A. and Thornton,J.M. (2001) Amino acidbase interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
32. Lamoureux,J.S., Stuart,D., Tsang,R., Wu,C. and Glover,J.N.M. (2002) Structure of the sporulation-specific transcription factor Ndt80 bound to DNA. *EMBO J.*, **21**, 5721–5732.
33. Lamoureux,J.S., Maynes,J.T. and Glover,J.N.M. (2004) Recognition of 5′-YpG-3′ sequences by coupled stacking/hydrogen bonding interactions with amino acid residues. *J. Mol. Biol.*, **335**, 399–408.
34. Lamoureux,J.S. and Glover,J.N.M. (2006) Principles of protein-DNA recognition revealed in the structural analysis of Ndt80-MSE DNA complexes. *Structure*, **14**, 555–565.
35. Rooman,M., Liévin,J., Buisine,E. and Wintjens,R. (2002) Cation-π/H-bond stair motifs at protein-DNA interfaces. *J. Mol. Biol.*, **319**, 67–76.
36. Biot,C., Wintjens,R. and Rooman,M. (2004) Stair motifs at protein-DNA interfaces: nonadditivity of H-Bond, stacking, and cation-π Interactions. *J. Am. Chem. Soc.*, **126**, 6220–6221.
37. MacKerell,A.D. Jr and Nilsson,L. (2008) Molecular dynamics simulations of nucleic acid-protein complexes. *Curr. Opin. Struct. Biol.*, **18**, 194–199.
38. Humphrey,W., Dalke,A. and Schulten,K. (1996) VMD—visual molecular dynamics. *J. Mol. Graph.*, **14**, 33–38.
39. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kale,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comp. Chem.*, **26**, 1781–1802.
40. Foloppe,N. and MacKerell,A.D. Jr (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, **21**, 86–104.
41. MacKerell,A. Jr, Bashford,D., Bellott,M., Dunbrack,R.L. Jr, Evanseck,J., Field,M.J., Fischer,S., Gao,J., Guo,H., Ha,S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
42. MacKerell,A.D. Jr, Feig,M. and Brooks,C.L. III (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.*, **25**, 1400–1415.
43. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
44. Martyna,G.J., Tobias,D.J. and Klein,M.L. (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, **101**, 4177–4189.
45. Feller,S.E., Zhang,Y.H., Pastor,R.W. and Brooks,B.R. (1995) Constant pressure molecular dynamics simulation—the Langevin piston method. *J. Chem. Phys.*, **103**, 4613–4621.
46. Shrake,A. and Rupley,J.A. (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, **79**, 351–364.
47. Schmidt,M.W., Baldridge,K.K., Boatz,J.A., Elbert,S.T., Gordon,M.S., Jensen,J.H., Koseki,S., Matsunaga,N., Nguyen,K.A., Su,S. *et al.* (1993) The general atomic and molecular electronic structure system. *J. Comp. Chem.*, **14**, 1347–1363.
48. Møller,C. and Plesset,M.S. (1934) Note on an approximation treatment for many-electron systems. *Phys. Rev.*, **46**, 618–622.
49. Foresman,J.B. and Frisch,A. (1996) *Exploring Chemistry with Electronic Structure Methods*. Gaussian Inc., Pittsburgh, PA.
50. Wintjens,R., Biot,C., Rooman,M. and Liévin,J. (2003) Basis set and electron correlation effects on ab Initio calculations of cation-π/H-Bond stair motifs. *J. Phys. Chem. A*, **107**, 6249–6258.
51. Frish,M.J., Pople,J.A. and Binkley,J.S. (1984) Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.*, **80**, 3265–3269.
52. Dunning,T.H. (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, **90**, 1007–1023.
53. Piecuch,P., Kucharski,S.A., Kowalski,K. and Musial,M. (2002) Efficient computer implementation of the renormalized coupled-cluster methods: the R-CCSD[T], R-CCSD(T), CR-CCSD[T], and CR-CCSD(T) approaches. *Comput. Phys. Commun.*, **149**, 71–96.
54. Parr,R.G. and Yang,W. (1989) Density-functional theory of atoms and molecules. Oxford University Press, New York.
55. Yakubovich,A., Solov'yov,I., Solov'yov,A. and Greiner,W. (2006) Conformational changes in glycine tri- and hexapeptide. *Eur. Phys. J. D*, **39**, 23–34.
56. Solov'yov,I., Yakubovich,A., Solov'yov,A. and Greiner,W. (2006) Ab initio study of alanine polypeptide chain twisting. *Phys. Rev. E*, **73**, 021916.
57. Solov'yov,I., Yakubovich,A., Solov'yov,A. and Greiner,W. (2006) On the fragmentation of biomolecules: fragmentation of alanine dipeptide along the polypeptide chain. *J. Exp. Theor. Phys.*, **103**, 463–471.
58. Boys,S.F. and Bernardi,F. (1970) The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.*, **19**, 553–566.
59. Wintjens,R., Liévin,J., Rooman,M. and Buisine,E. (2000) Contribution of cation-π interactions to the stability of protein-DNA complexes. *J. Mol. Biol.*, **302**, 393–408.

60. Zwanzig,R.W. (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.*, **22**, 1420–1426.
61. Chipot,C. and Pearlman,D.A. (2002) Free energy calculations. The long and winding gilded road. *Mol. Sim.*, **28**, 1–12.
62. Chipot,C. and Pohorille,A. (2007) *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Springer, Berlin Heidelberg.
63. Pohorille,A., Jarzynski,C. and Chipot,C. (2010) Good practices in free-energy calculations. *J. Phys. Chem. B*, **114**, 10235–10253.
64. Bennett,C.H. (1976) Efficient estimation of free energy differences from Monte Carlo data. *J. Chem. Phys.*, **22**, 245–268.
65. Kirkwood,J. (1935) Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, **3**, 300–313.
66. Erikssona,M., Härda,T. and Nilsson,L. (1995) Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys. J.*, **68**, 402–426.
67. Prévost,C., Takahashi,M. and Lavery,R. (2009) Deforming DNA: from physics to biology. *Chem. Phys. Chem.*, **10**, 1399–1404.
68. Vallone,B., Miele,A.E., Vecchini,P., Chiancone,E. and Brunori,M. (1998) Free energy of burying hydrophobic residues in the interface between protein subunits. *Proc. Natl Acad. Sci. USA*, **95**, 6103–6107.
69. Caldwell,J.W. and Kollman,P.A. (1995) Cation-π interactions: nonadditive effects are critical in their accurate representation. *J. Am. Chem. Soc.*, **117**, 4177–4178.
70. Cubero,E., Luque,F.J. and Orozco,M. (1998) Is polarization important in cation-π interactions? *Proc. Natl Acad. Sci. USA*, **95**, 5976–5980.
71. Minoux,H. and Chipot,C. (1999) Cation-π interactions in proteins: can simple models provide an accurate description? *J. Am. Chem. Soc.*, **121**, 10366–10372.
72. Dehez,F., Angyan,J.G., Gutierrez,I.S., Luque,F.J., Schulten,K. and Chipot,C. (2007) Modeling induction phenomena in intermolecular interactions with an ab initio force field. *J. Chem. Theor. Comp.*, **3**, 1914–1926.
73. van Mourik,T. and Gdanitz,R.J. (2002) A critical note on density functional theory studies on rare-gas dimers. *J. Chem. Phys.*, **116**, 9620–9623.
74. Šponer,J., Jurečka,P. and Hobza,P. (2006) Base stacking and base pairing. In: Šponer,J. and Lankaš,F. (eds), *Computational Studies of RNA and DNA*. Springer, Netherlands, pp. 343–388.
75. Anderson,P.C. and Daggett,V. (2009) The R46Q, R131Q and R154H polymorphs of human DNA glycosylase/β-Lyase hOgg1 severely distort the active site and DNA recognition site but do not cause unfolding. *J. Am. Chem. Soc.*, **131**, 9506–9515.
76. Rutherford,K. and Daggett,V. (2010) Polymorphisms and disease: hotspots of inactivation in methyltransferases. *Trends Biochem. Sci.*, **35**, 531–538.
77. Juffer,A.H., Eisenhaber,F., Hubbard,S.J., Walther,D. and Argos,P. (1995) Comparison of atomic solvation parametric sets: applicability and limitations in protein folding and binding. *Prot. Sci.*, **4**, 2499–2509.
78. Travers,A.A. (1989) DNA conformation and protein binding. *Annu. Rev. Biochem.*, **58**, 427–452.
79. Lavery,R., Zakrzewska,K., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T. III, Dixit,S., Jayaram,B., Lankas,F., Laughton,C. *et al.* (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.
80. Ballestar,E. and Wolffe,A.P. (2001) Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur. J. Biochem.*, **268**, 1–6.
81. Klose,R.J., Sarraf,S.A., Schmiedeberg,L., McDermott,S.M., Stancheva,I. and Bird,A.P. (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell*, **19**, 667–678.
82. Clouaire,T., de las Heras,J.I., Merusi,C. and Stancheva,I. (2010) Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. *Nucleic Acids Res.*, **38**, 4620–4634.