J Mol Model (2017) 23: 152
DOI 10.1007/s00894-017-3318-8

ORIGINAL PAPER

Methylation-targeted specificity of the DNA binding proteins R.DpnI and MeCP2 studied by molecular dynamics simulations

Siba Shanak[1,2] · **Ozlem Ulucan**[1] · **Volkhard Helms**[1]

Received: 7 December 2016 / Accepted: 13 March 2017 / Published online: 3 April 2017

Abstract DNA methylation plays a major role in organismal development and the regulation of gene expression. Methylation of cytosine bases and the cellular roles of methylated cytosine in eukaryotes are well established, as well as methylation of adenine bases in bacterial genomes. Still lacking, however, is a general mechanistic understanding, in structural and thermodynamic terms, of how proteins recognize methylated DNA. Toward this aim, we present the results of molecular dynamics simulations, alchemical free energy perturbation, and MM-PBSA calculations to explain the specificity of the R.DpnI enzyme from *Streptococcus pneumonia* in binding to adenine-methylated DNA with both its catalytic and winged-helix domains. We found that adenine-methylated DNA binds more favorably to the catalytic subunit of R.DpnI (−4 kcal mol^{-1}) and to the winged-helix domain (−1.6 kcal mol^{-1}) than non-methylated DNA. In particular, N6-adenine methylation is found to enthalpically stabilize binding to R.DpnI. In contrast, C5-cytosine methylation entropically favors complexation by the MBD domain of the human MeCP2 protein with almost no contribution of the binding enthalpy.

Keywords Restriction endonuclease · DNA methylation · m6A · m5C · Sequence specificity · Binding free energy · Conformational entropy · Free energy perturbation · MM-PBSA

Electronic supplementary material The online version of this article (doi:10.1007/s00894-017-3318-8) contains supplementary material, which is available to authorized users.

✉ Volkhard Helms
volkhard.helms@bioinformatik.uni-saarland.de

[1] Center for Bioinformatics, Saarland University, PO Box/15 11 50, 66041 Saarbrücken, Germany

[2] Department of Biology and Biotechnology, Faculty of Science, Arab American University, PO Box 240, Jenin, Palestinian Territories, Israel

Introduction

DNA methylation plays a major role in a wide variety of biological processes, including the regulation of gene expression and self-recognition. In bacteria, the dominant form is the N6-adenine methylation that helps protect bacteria against invasion by foreign DNA [1]. Recently, N6-adenine methylation of DNA was also reported to be involved in the epigenetic regulation of gene expression in mouse stem cells [2]. The R.DpnI enzyme (see Fig. 1a) from *Streptococcus pneumonia* is a type IIE restriction endonuclease that consists of an N-terminal catalytic domain and a C-terminal winged helix domain (residues 183–254) [3]. R.DpnI protects Dam−, R.DpnI+ bacteria against phages that have been propagated on Dam+ hosts. Both domains of R.DpnI bind highly specifically to Dam-methylated (Gm6ATC) sites [4]. A recent X-ray structure determined by the Bochtler group [5] characterized how the two R.DpnI domains bind to methylated DNA. The authors noticed that the presence of the two methyl groups requires a structural deviation from the canonical B-DNA conformation to avoid steric conflict.

C5-cytosine methylation of DNA is one of the most crucial epigenetic modifications of eukaryotic genes, and plays an indispensable role in modulating mammalian gene expression. The human Methyl-CpG binding protein 2 (MeCP2) is crucial in gene silencing and has important neurological implications [6] (see Fig. 1b). It belongs to a highly conserved family of DNA-binding proteins that mediate gene silencing by specifically binding to methylated CpG sites in the promoter regions of genes. MeCP2 possesses a hydrophobic methyl-binding domain (MBD) that interacts directly with methyl

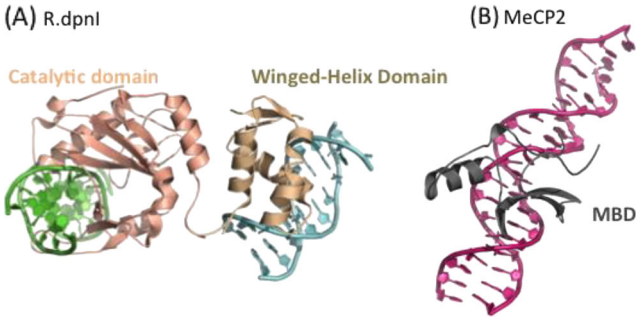

Fig. 1a,b Crystal structures of methylated DNA bound to the restriction enzyme R.DpnI from *Staphylococcus aureus* and to human MeCP2. **a** Winged-helix domain and catalytic domain of R.DpnI, each bound to DNA containing N6-methylated-adenine, are colored *cyan* and *green* (RCSB PDB code: 4KYW), respectively. **b** Methyl-binding domain of MeCP2 (*dark grey*) bound to DNA containing C5-methylated cytosine (*hot pink*) (RCSB PDB code:3C2I). Figures were generated using PYMOL

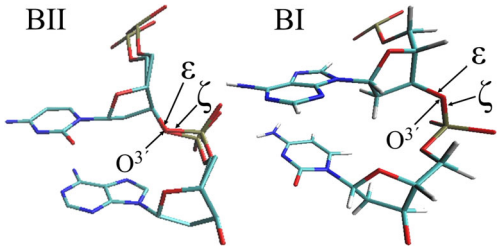

Fig. 2 BII (*left*) and BI (*right*) conformations of double-stranded DNA. The combined rotation around the two torsion angles ε and ζ generates the two conformational substates BII and BI. The BII conformation shown here was extracted from the pdb file 3GGI at the 2DCp3DT dinucleotide. The BI conformation was extracted from our simulations of MBD:mDNA (dinucleotide 7DAp8DC)

groups attached to cytosine bases of DNA, for example at the promoter III of the mouse-brain-derived neurotrophic factor (BDNF) [7]. This gene encodes the BDNF protein that seems to be involved in sustaining long-term memory [8]. A pioneering X-ray structure of the MBD domain bound to methylated DNA revealed that the methyl group of cytosine surprisingly contacts a predominantly hydrophilic surface patch on the MBD domain that includes tightly bound water molecules [9].

As mentioned, X-ray crystallography has been instrumental in elucidating the structural details of how proteins bind to methylated or non-methylated DNA sequences. When binding to DNA, proteins generally induce an increase in the width of the major DNA groove, so that the functional groups of DNA make favorable contacts to protein residues at the binding interface in a sequence- and methylation-specific context [10]. Moreover, when proteins bind to their target DNA regions, they often change the equilibrium between two alternating conformational states of DNA termed BI and BII that are characterized by different relative orientations of neighboring phosphate groups in the DNA backbone [11, 12]. As shown in Fig. 2, the two conformations differ with respect to the (epsilon-zeta) torsion angles of the DNA backbone, being negative in BI and positive in BII [13]. In the BII conformation, the bases are pushed toward the major groove of the DNA, making them more accessible to the bound protein, and in a sequence-specific manner [11]. In addition to affecting the DNA conformation, methylation of DNA may also enhance specific binding of proteins via solvent contributions or via specific interactions with protein residues [9, 14]. As it is typically not possible for X-ray crystallography to characterize proteins complexed with both methylated or non-methylated forms of their target DNA regions, there is an important need to employ molecular modeling and biomolecular simulations to unravel the mechanisms behind methylation-specific binding.

In a pioneering work, Zou and colleagues [15] conducted conventional molecular dynamics (MD) simulations and alchemical free energy perturbation calculations for the MBD:DNA system of the C5-methylcytosine binding protein MeCP2. They emphasized that the structural 'stair motif' consisting of interactions between the mCpG dinucleotide and two arginine residues at the protein-binding interface is important for methylation-specific binding. We recently characterized the binding of methylated or non-methylated BNDF fragments to MBD by plain molecular dynamics simulations [16]. In that study, we observed lower fluctuations for the methylated MBD complex as compared to unbound DNA or to the non-methylated complex. Also, binding of the methylated BDNF promoter induced a larger opening of the major groove than the non-methylated form. These observations suggested that MBD forms a tighter contact to methylated BDNF, in agreement with the experimentally detected higher efficiency of translational repression of the methylated promoter. However, in that study, we did not address the thermodynamic determinants of the MBD:BDNF promoter interaction.

This is exactly the aim of the present study. Also, we contrast the contribution of C5-cytosine- vs. N6-adenine- methylation to the specificity of protein binding. Towards this aim, we studied the R.DpnI enzyme, which binds selectively to N6-adenine-methylated DNA, and compared it to the MBD:BDNF promoter system that binds selectively to C5-cytosine-methylated DNA [9]. We present results from conventional MD simulations, alchemical free energy perturbation, and Poisson Boltzmann/Surface Area (PB/SA) calculations to characterize the conformational changes induced in the DNA strand upon binding to the protein, and to determine the free energy changes and the corresponding enthalpic and entropic contributions for both systems, either in the methylated or in the non-methylated DNA sequence context. The findings from this study provide a basis for establishing a mechanistic understanding of the structural and thermodynamic parameters that determine why and how proteins may specifically recognize the cytosine- or adenine-methylated forms of DNA.

Methods

MD simulations

The MD simulations for all systems were performed with the GROMACS 4.5.5 package [17] using the CHARMM27 force field [18] and the TIP3P water model [19]. Force field parameters for methylated adenosine were taken from [20].

Each unbound DNA or protein:DNA complex of the two R.DpnI systems (catalytic domain and winged-helix domain) was placed in a cubic water box of 6.1 or 9.9 nm box dimensions with 0.10 mol/l NaCl added. The total size of the simulated R.DpnI systems was around 23,000 atoms for the unbound DNA systems, and around 95,400 atoms for the solvated protein–DNA complexes. The simulations of MeCP2 for the unbound DNA and the bound DNA-complex were conducted using a cubic box of 9.3 nm dimensions, also with 0.10 mol/l NaCl added. The total size of these systems was about 56,200–56,300 atoms. The setup of this system is identical to [16]. Periodic boundary conditions were employed. Coulombic interactions were evaluated using a short-range cut-off of 10 Å and long-range interactions were treated by the particle-mesh Ewald (PME) summation method [21]. The non-bonded Lennard-Jones interactions were computed using a smooth cutoff of 10 Å. The integration time step was set to 1 fs. The temperature was kept at 310 K by applying leap-frog stochastic dynamics forces with a damping coefficient of 0.1 ps^{-1} [22].

At first, each simulated system was energy-minimized for 50,000 steps using the steepest descent algorithm followed by a second energy minimization for 10,000 steps using a quasi-Newtonian algorithm with the low-memory Broyden-Fletcher-Goldfarb-Shanno approach. The tolerance was set to 1.0 kJ mol^{-1} nm^{-1}. Thereafter, the system was heated to 310 K during 4 ps. Then, each system was subjected to 1 ns-equilibration in the NVT ensemble with harmonic restraints applied to all protein and DNA atoms using a force constant of 1000 kJ mol^{-1} nm^{-2}. With restraints kept, each system was further equilibrated for 500 ps in the NPT ensemble, and then for another 500 ps without restraints.

The *apo* form of R.DpnI-binding DNA was simulated in the free form starting from an ideal B-DNA conformation (methylated, nonmethylated and hemimethylated DNA). For methylated and nonmethylated DNA, we collected 500 ns simulations (two replicates with 100 ns each, plus ten shorter replicates of 30 ns each). Simulations of the hemi-methylated forms were conducted for 40 ns each. Here, we analyzed only variables (torsion angles, solvent-accessible surface area for methyl groups, radial distribution function of water around methyl groups) that converged rather fast and apparently needed no longer simulations.

Simulations of the R.DpnI:DNA complexes were based on the recent crystal structure of R. DpnI with two strands of N6-adenine-methylated DNA bound to the catalytic and the winged-helix domains [5]. In the simulations, we simulated the intact R.DpnI protein with DNA either bound to the catalytic or to the winged-helix domain. DNA in either domain was further mutated into non-methylated and hemi-methylated DNA sequences (both in the sequence context of the proximal and distal methyl groups) by replacing the respective methyl groups by hydrogen atoms. For the fully methylated and nonmethylated DNA bound to either domain, we collected 500 ns simulations (two replicates, 100 ns each plus ten replicates with 30 ns each). Simulations of the two forms of hemimethylated DNA were conducted for 40 ns each.

As a structural reference for the simulations of the MBD:DNA complex, we used the X-ray structure of the BDNF promoter bound to the human methyl-binding domain (RCSB:3C2I; [9]). The plain MD simulations for this system (two replicate simulations of 100 ns each) for free DNA in the methylated and non-methylated forms, and for bound DNA in the methylated and non-methylated forms were previously described [16]. Since we found very good agreement between the two replicate simulations of MBD:BDNF systems both in the methylated and non-methylated forms, as well as in the unbound form and in the complex (see Table 2), we saw no reason to extend the simulation time of this system to the same length as used for the R.DpnI systems.

Free energy perturbation

The free energy calculations for R.DpnI were conducted independently of the plain MD simulations. All windows were started from the initially equilibrated systems, and were conducted in a parallel fashion. Alchemical free energy perturbation (FEP) calculations using the Bennett acceptance ratio with error bars (BAR) were employed to determine the difference in binding free energy of the R.DpnI–DNA complex upon de-methylating 5-adenosine in both DNA strands [23, 24]. Both in unbound DNA and in the protein:DNA complex, the two methyl groups attached to adenosine were annihilated in two stages [25], see Fig. S1. In the first stage (corresponding to the transition from panel $1 \rightarrow 2$), the electrostatic interactions of each methyl group were switched off in a step-wise manner and its respective charge was assigned to the N6 atom. For this, the system Hamiltonian was coupled to a coupling parameter λ whereby $\lambda = 0$ corresponds to the reference state and $\lambda = 1$ to the perturbed state. No soft-core potential was used in this step since the final, uncharged atoms still exert full Lennard-Jones (LJ) interactions so that no singularities are introduced in this step (see below). In the second stage ($2 \rightarrow 3$), the atoms of the methyl group were turned into dummy atoms (by switching their epsilon and sigma LJ parameters to zero). In this stage, a soft core potential was used where soft-core alpha was set to 0.5, the soft-core power to 1.0, and soft-

core sigma to 0.3 [26]. When their LJ potentials are annihilated, atoms are converted into dummy atoms. This can introduce numerical problems during the last annihilation step(s) when the interactions of an annihilated group of atoms become so small that neighboring molecules, e.g., water molecules, can approach them very closely. At such short atom–atom distances, small displacements during an MD time step can suddenly generate an exceedingly large repulsion due to the remaining repulsive term of the LJ interaction. Exactly these problems are overcome with a soft-core potential that avoids the singularity at the origin [27]. To complete the free energy cycle, the 'dummy' hydrogen atom of the non-methylated adenosine was turned into an interacting hydrogen by switching on its Lennard-Jones interactions (stage 3; 4 → 5), and then the electrostatic interactions (stage 4; 5 → 6). Each of the four stages was decomposed into 26 intermediates states ($\Delta\lambda = 0.04$). As for the reference state at $\lambda = 0$, the simulation of each intermediate state started with a double energy minimization, followed by equilibration over 1 ns in the NVT ensemble and 500 ps equilibration in the NPT ensemble with harmonic restraints, and 500 ps without any restraints. Data was collected during another 1.5 ns for each window. This yields a total simulation time of 26 × 3.5 ns = 91 ns for each unidirectional simulation.

MM-PBSA energy calculations

In the MM-PBSA approach [28], the enthalpic contributions to the free energy of binding are calculated via:

$$H = E_{bonded} + E_{vdW} + E_{elec} + E_{PB} + E_{SA} \quad (1)$$

where bonded stands for the bonded energy terms (bond lengths, bond angles and torsion angles), vdW stands for the van der Waals interactions, and elec for the Coulombic interactions. The three terms represent the molecular mechanics (MM) terms and are computed in the gas phase. SA refers to the surface area contribution. PB stands for the solvation free energy computed here with the adaptive Poisson Boltzmann solver [29]. The net change in enthalpy upon methylation can be calculated for a single biological entity as follows:

$$\Delta\Delta\mathrm{H} = \Delta H_{met.complex} - \Delta H_{nm.complex} = H_{met.complex} - \left(H_{protein} + \mathrm{H}_{met.DNA}\right) - \left(\mathrm{H}_{nm.complex} - \left(H_{protein} + \mathrm{H}_{nm.DNA}\right)\right) \quad (2.a)$$

Here, $H_{\mathrm{met.complex}}$ belongs to the complex of methylated DNA and protein, H_{protein} to the unbound protein, and $\mathrm{H}_{\mathrm{met.DNA}}$ (or $\mathrm{H}_{\mathrm{nm.DNA}}$) to unbound methylated (or non-methylated) DNA. Since H_{protein} cancels out, this simplifies into:

$$\Delta\Delta\mathrm{H} = H_{met.complex} - \mathrm{H}_{nm.complex} - \left(\mathrm{H}_{met.DNA} + \mathrm{H}_{nmDNA}\right) \quad (2.b)$$

We note that this sum is not purely a sum of enthalpic terms, since the PB/SA terms are parameterized as solvation free energies. Enthalpy decomposition was performed using the amber MMPBSA.py tool [30]. The original set of MD simulations was performed using GROMACS (as explained in MD simulations), the trajectories of several replicates were merged, and the snapshots were superimposed on the starting structure. We then used the package AMBERTOOLS (version 2) to compute the contributions of the individual components to the free energy of binding. For consistency, the CHARMM27 force field was also used in the MM-PBSA calculations. For this, the topology files were generated with the CHAMBER package available in AMBERTOOLS, and applied to the snapshots of the GROMACS MD simulations generated with the same CHARMM27 force field [31]. Protein structure files (PSF) in CHARMM format were provided as an input to the CHAMBER package. A total of 10,000 fitted PDB snapshots per simulation type were fed as input to the CPPTRAJ utility in AMBERTOOLS, in order to generate the MDCRD trajectories required for the Amber package. Both structure file types (PDB and PSF) were processed with the PSFGEN plugin in VMD [32, 33]. Patches and parameters for the nonstandard residue 6MA were added to the input topology and parameter files.

Configurational entropy of DNA and protein

Characterizing the configurational entropy of flexible solute molecules as studied here by molecular simulations is known to require lengthy MD simulations. Thus, we performed simulations in several replicates for the apo and the holo forms of DNA started from various starting configurations, and then merged the sampled conformations to achieve better and faster convergence of the configurational entropy. For the B-DNA form of all unbound systems, as well as for the crystal forms of the DNA bound to the winged-helix domain and to the catalytic domain, ten short simulations (30 ns each) were merged with two longer simulations of 100 ns each. To check for convergence, we followed the strategy introduced by Domene and co-workers [34]. Snapshots were collected each 5 ps. For the MBD system, we used snapshots from two replicate simulations of 100 ns each [16].

The configurational entropy of the complexes, proteins and DNA in the methylated and non-methylated forms was quantified by the approach of Schlitter [35]. Entropy differences due to methylation were computed as follows:

$$\Delta\Delta S = \Delta S_{met.complex} - \Delta S_{nm.complex} = S_{met.complex} - S_{nm.complex} - \left(S_{protein} + S_{met.DNA} - \left(S_{protein} + S_{nmDNA}\right)\right) = S_{met.complex} - S_{nm.complex} - \left(S_{met.DNA} - S_{nmDNA}\right) \quad (3)$$

These contributions were computed for the individual proteins and DNA as well as for the formed complexes.

Results

In this study, we employed MD simulations to determine the structural and energetic characteristics that mediate the specific binding of the catalytic and winged-helix domains of the bacterial restriction enzyme R.DpnI and of the MBD domain of the mammalian transcription factor MeCP2 to DNA strands carrying either methylated or non-methylated adenine or cytosine bases. For the three studied systems, we found that the DNA structure was generally well maintained in all simulations, with an upper RMSD of 3 Å and an average RMSD of 1.5 Å during MD simulations of 100 ns duration.

Structural adaptation of DNA upon binding

The width of the major groove of DNA is known to have an important effect on the specificity of protein:DNA binding. Figure 3a shows the major groove width at the dinucleotide step in the binding interface for the R.DpnI simulations (both in the complexes and in the unbound DNA); Fig. 3b shows this for the MBD simulations (see also [16]). Both non-methylated DNA strains (solid blue lines in 2A and 2B) gave a clear peak around 15.5 Å. The major groove width with methylated cytosine is a bit narrower (16.0 Å, solid red, 2B) than with the methylated adenosine (17.0 Å; solid red, 2A). The protein–DNA complexes always showed an opening of the major groove. This tendency was stronger in the winged helix domain of R.DpnI (22.0 Å) than in the catalytic domain (19.0 Å). Methylated DNA bound to proteins generally gave narrower-peaked distance profiles compared to the more 'floppy' non-methylated forms. MBD-bound DNA also showed a very clear opening transition. We did not observe contacts between protein and the DNA minor groove.

Next, we checked the BI ratio of the DNA strands (see Fig. 4). In the BI conformation, the difference between the two torsion angles ε (C4′-C3′-03′-P) and ζ (C3′-03′-P-05′) is about −90°, and about +90° in the BII form [36]. Importantly, the phosphate position in the BI conformation is symmetrical with respect to the minor and the major grooves, whereas the BII conformation shifts the phosphates to the minor groove. We found that the non-methylated R.DpnI-binding sequence adopted a slightly smaller BI ratio in solution (0.88) than the MBD-binding sequence (0.92). In both cases, methylation of adenosine or cytosine hardly induced any conformational changes in solution compared to the non-methylated forms. In the complexes, the BI/BII equilibrium was shifted to slightly smaller values, indicating a higher proportion of the BII conformation with better accessible nucleic bases (see Introduction). Upon binding to MBD, the change was rather small (0.02–0.03). Upon binding to R.DpnI, bimodal

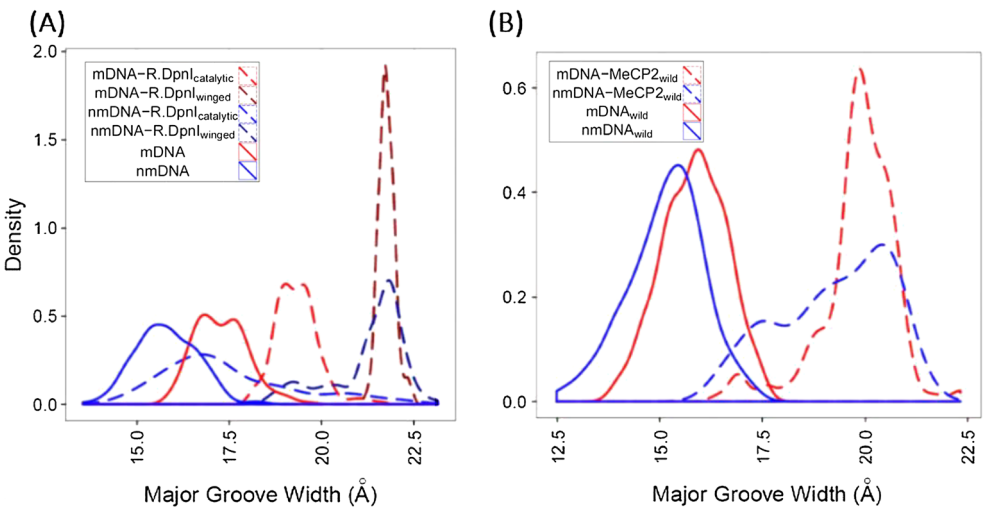

Fig. 3 Width of the major groove in an ensemble from **a** 500 ns of simulations (merged trajectories) of the R.DpnI system; methylated and non-methylated, and **b** 200 ns of simulations (merged trajectories) for the MeCP2 system. For this figure, Figs. 4 and S5, the same ensembles were used to compute the frequency distributions

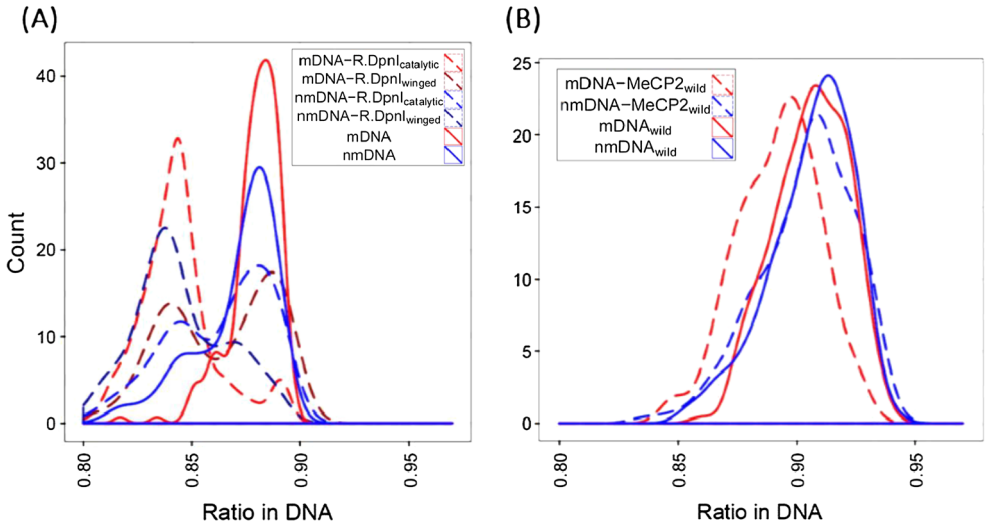

Fig. 4 BI ratio in the bound and unbound forms of DNA in **a** the R.DpnI systems and **b** the MBD:DNA system

distributions were observed with peaks near 0.88 and near 0.84. When complexed either to the catalytic or to the winged-helix domains, N6-methylation of adenine increased the occupancy of the 0.84 peak.

Investigating the DNA structure in the bound and unbound conformations (Table S1) in terms of the well-known basepair steps (rise, roll, shift, slide, tilt, and twist) revealed a certain structural strain in the DNA bound to the catalytic domain of R.DpnI and when bound to the MBD protein. These deformations were smaller in complexes with the winged-helix domain of R.DpnI.

Methyl–methyl proximity effects on methylation specific binding

DNA is fully methylated in the crystal structure of the R.DpnI enzyme so that the two N6-adenine methyl groups are tightly packed against each other. According to Fig. 5, the methyl–methyl distance is only slightly larger than the sum of their van der Waals radii (2.0 Å + 2.0 Å = 4.0 Å; [37]) in the X-ray structure of the catalytic domain of R.DpnI. This is very similar to the situation in a methylated perfect B-DNA fiber, generated for the GATC sequence with the 3DNA program and default parameters [38, 43]. When bound to the winged-helix domain, this distance is about 4.5 Å in the X-ray structure. We reasoned that interactions between the methyl groups could on one hand promote a 'shielding effect' of the hydrophobic methyl group (see below). On the other hand, close contacts could reduce the conformational flexibility of DNA in solution, and thus lower its conformational entropy [5]. To find out whether this is the case, we computationally grafted methyl groups in trans-orientation onto the N6 atoms of the adenines in the A:T dinucleotide steps in the MD snapshots (fitted to the X-ray conformation of the methylated DNA). For unbound DNA simulated in the non-methylated form, grafting showed that the methyl groups were indeed too close (3.4 Å). However, grafts applied to MD snapshots of non-methylated DNA bound to the winged-helix domain of DNA gave fluctuations around an average distance of 4.0 Å, what is an acceptable value (see above). For the non-methylated DNA bound to the catalytic domain, however, this was not the case. Here, the grafts applied to DNA showed a minimum distance of 3.0 Å and equilibrated around this value. For the MBD:meDNA complex, the distance between the two methyl groups of the C5-methylcytosines in the methylated complex (8.0 Å) was much larger than the sum of the van der Waals radii. Therefore, there appears no risk of entropic restraints in terms of the binding specificity of 5mC.

N6-methyladenosine can adopt both the cis and the trans isomers. Here, only the planar trans conformation was observed in the MD simulations for the fully and hemi-methylated Gm6ATC target sequence (Fig. S2). When complexed to R.DpnI, water molecules around the N6 adenine

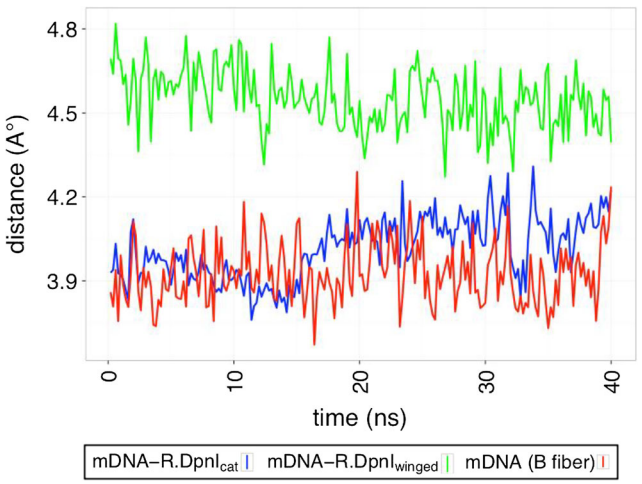

Fig. 5 Carbon–carbon distance between the two methyl groups of N6-methyladenine of DNA in the free form, and in complex with the catalytic or winged helix domains of R.DpnI during the initial 40 ns of the 100-ns-long simulations.

atoms were displaced, both in the case of methylated and non-methylated DNA (Fig. 6). Water displacement was more pronounced upon binding to the winged-helix domain than to the catalytic domain. In contrast, Fig. S3 shows that very few water molecules were displaced with respect to the C5–C5M plane connecting the C5-methylated cytosine.

Methyl groups energetically stabilize complexes of DNA with the catalytic and winged-helix domains of R.DpnI

The $\Delta\Delta G$ contribution of DNA methylation to R.DpnI binding was quantified using alchemical free energy perturbation calculations and the thermodynamic cycle shown in Fig. 7. For this, we transformed either the methyl groups on the proximal and distal strands of fully methylated DNA or the respective hydrogen atoms in the non-methylated form of DNA from a fully interacting state to a dummy state (Fig. S1). This was done in solution as well as in the complexes with the catalytic and the winged-helix domains of R.DpnI.

During the free energy calculations for the DNA:R.DpnI complex with DNA bound to either the catalytic or the winged-helix domains, the protein conformation was well maintained. During all stages and windows, the RMSD values between final and starting conformations were between 0.2 and 0.4 nm, which are values typical of MD simulations of proteins. Also, DNA conformations were well preserved throughout all simulation windows of the discharging step. However, in the last windows of the stages where the LJ potential was turned off (between $\lambda = 0.96$ and 1.00, during both perturbations of the methyl species as well as replacing hydrogens), the DNA unwound and dissociated from the complex. As the structural transition happened both in the presence and absence of protein, and, as most of the free energy change occurs for initial values of $\lambda \leq 0.8$ (see Fig.S4), the computed free energy differences should be almost unaffected. This unexpected computational result is consistent with the biochemically observed low affinity of non-methylated DNA to the catalytic domain [3]. Figure S4 shows that the cumulative changes in free energy as a function of λ were quite smooth both during the van der Waals elimination step and during the discharging step.

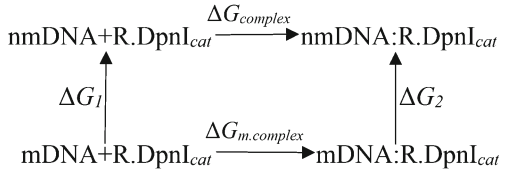

Fig. 7 Thermodynamic cycle to compute the contribution of DNA-methylation to the binding affinity of the R.dpnI protein toward DNA. The upper leg corresponds to the association of non-methylated DNA, the lower leg refers to the binding of methylated DNA. The difference of these two legs is the desired quantity. It can also be obtained as the difference of the two vertical processes 1 and 2, which are conveniently conducted as annihilation processes described in Fig. S1

Table 1 shows the free energy changes (ΔG) with standard error for the individual annihilation processes for the protein-DNA complexes and the unbound DNA strands. For unbound DNA started in the conformation extracted from the X-ray complex with the catalytic domain of R.DpnI, mutating the N6-adenine-methylated DNA to non-methylated DNA (ΔG_1) gave a favorable free energy change of ΔG_1 of −3.70 (±0.22) kcal mol^{-1} (Table 1). Mutating methylated DNA bound to the

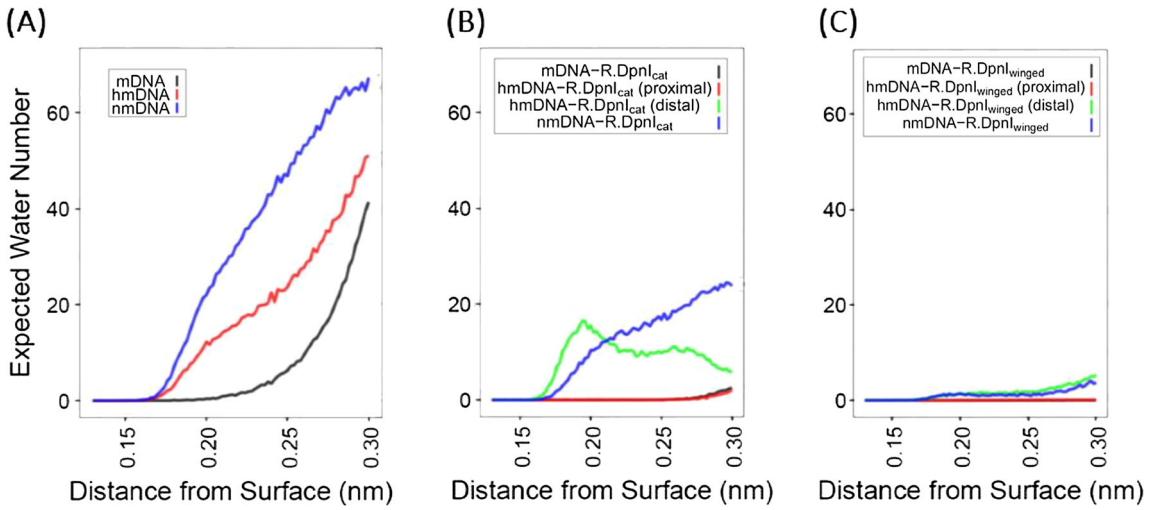

Fig. 6a–c Radial distribution of water molecules around the hypothetical surface of the two N6-amino/methylamino groups in DNA. **a**–**c** Expected value (EV) for the distribution of the number of water molecules within 0.13 to 0.3 nm distance from the surface during the 100 ns long MD simulations of methylated, hemimethylated and non-methylated DNA:R.DpnI complexes (**b**, **c**); as well as the respective unbound DNAs (**a**). The expected number of water molecules was derived by calculating the average radial distribution function of water molecules up to the indicated distance using the RDF function in GROMACS [17]. Snapshots were extracted at 10 ps intervals, and the average number of water molecules in all snapshots was calculated

Table 1 Thermodynamic cycle to compute the contribution of adenine methylation to the binding free energy of DNA to the catalytic and the winged-helix subunits of the R.DpnI protein. Values are in units of kcal mol^{-1}

	$\Delta G_{\text{discharging}}$	$\Delta G_{\text{turning LJ off}}$	ΔG_{total}
(A) Catalytic domain			
met. DNA–DNA	−3.47 (±0.11)	−0.23 (±0.18)	−3.70 (±0.22)
met. Complex–Complex	24.25 (±0.58)	−23.9 (±0.36)	0.35 (±0.68)
$\Delta\Delta$G=			−4.05 (±0.71)
(B) Winged-helix domain			
met. DNA–DNA	2.76 (±0.44)	0.09 (±0.36)	2.85 (±0.40)
met. Complex–Complex	13.66 (±0.11)	−9.49 (±0.39)	4.17 (±0.29)
$\Delta\Delta G$=			−1.32 (±0.34)

catalytic domain into its non-methylated counterpart (ΔG_2) gave a slightly unfavorable free energy difference of ΔG_2 of 0.35 (±0.68) kcal mol^{-1}. Hence, the total cycle ΔG_1–ΔG_2 adds up to a $\Delta\Delta G$ of −4.05 (±0.71) kcal mol^{-1} meaning that N6-adenine-methylated DNA binds more strongly to the catalytic domain of the R.DpnI protein by this amount than non-methylated DNA.

Table 1 also shows the free energy changes (ΔG) with standard error for the same annihilation processes in the winged-helix domain-DNA complex and the respective unbound DNA. Here, the free energy calculations were started from the final conformations of the protein–DNA complex after 100 ns of plain MD simulations. Mutating the methylated DNA in water to non-methylated DNA (ΔG_1) gave an unfavorable free energy change of $\Delta G_1 = 2.85$ (±0.40) kcal mol^{-1}. On the other hand, mutating the methylated DNA bound to the winged-helix domain of R.DpnI into the non-methylated counterpart (ΔG_2) gave an unfavorable free energy difference of $\Delta G_2 = 4.17$ (±0.29) kcal mol^{-1}. Hence, the total cycle ΔG_1–ΔG_2 adds up to $\Delta\Delta G = -1.32$ (±0.34) kcal mol^{-1}, suggesting that methylated DNA binds more strongly to the winged-helix domain of the R.DpnI protein by this amount than non-methylated DNA, given the initial conformation of the DNA and the protein.

Enthalpic contribution to the binding free energy

Using the MM-PBSA approach (one trajectory method), we characterized the enthalpic contribution of N6-adenine methylation to the binding affinity toward both domains of the R.DpnI enzyme. For comparison, we also computed the enthalpic contribution of C5-cytosine methylation to the binding to the MeCP2 protein. Using Eq. (2), the MM-PBSA calculations with an internal dielectric of four showed that adenosine methylation enthalpically favors binding to the winged-helix domain by −11.01kcal mol^{-1}, and to the catalytic domain by −9.34kcal mol^{-1}. In contrast, the calculations showed that C5-cytosine methylation enthalpically slightly destabilizes the complex with MeCP2 by 0.76kcal mol^{-1} (Table 2).

Entropic contribution to the free energy of binding

Based on the snapshots of the concatenated MD simulations (0.5 μs in total for R.DpnI and 0.2 μs for MBD, respectively) we quantified the configurational entropy of DNA using the Schlitter method implemented in GROMACS [35]. The simulations of unbound methylated and non-methylated DNA sequences were started from perfect B-DNA conformations. The computed configurational entropies are listed in Table 3 (also see Fig. S5). Figure S5 shows that the R.DpnI:DNA system needed more time than the corresponding MBD:DNA system to sample the accessible configurational space and reach convergence. A single 100 ns long simulation of the R.DpnI system appeared insufficient in this respect. Thus we added simulation snapshots from ten parallel 30 ns long simulations to improve sampling. This combined analysis appears to result in reasonable convergence so that, in particular, the entropic difference between methylated and non-methylated systems should be captured in a robust manner. When considering both protein and DNA, N6-methylation of adenine gave an unfavorable entropic contribution of −17.33 kcal mol^{-1} for binding to the catalytic subunit of R.DpnI. When separately considering DNA and protein, both gave a negative contribution each (−5.95 and −13.9, respectively). The situation is very different for binding to the winged-helix domain of R.DpnI. Here, we noticed a strong entropic decrease in the DNA, but this was completely offset by a positive contribution of the protein. Overall, methylation was predicted to give a slightly favorable entropic contribution of 2.46 kcal mol^{-1}. The case is again different for the 5C-meDNA:MBD system. Here, almost no change is found in DNA alone, whereas the protein shows a clearly lowered entropy when bound to methylated DNA. Interestingly, the entropy computed for the full protein:DNA system is higher for methylated DNA than for non-methylated DNA (9.54 kcal mol^{-1}). This reflects the important role of the relative mobility of protein and DNA that is not considered when treating the binding partners individually.

Table 2 Enthalpic contribution of methylation to the binding energies using the MM-PBSA approach. The internal dielectric constant was set to four. Values are given as mean (SE) in kcal mol^{-1}

	DNA:R.dpnI$_{cat}$	DNA:R.dpnI$_{winged}$	DNA-MBD
nmDNA			
$H_{VDWAALS}$	−134.90(0.10)	−103.08 (0.10)	−77.92 (0.11)
H_{EEL}	−748.37(0.30)	−685.60 (0.22)	−986.21 (0.51)
E_{PB}	752.39(0.25)	662.19 (0.21)	942.41 (0.48)
E_{NPOLAR}	−92.03(0.06)	−67.72 (0.05)	−61.70 (0.07)
E_{DISPER}	225.55(0.09)	167.47 (0.12)	150.59 (0.15)
ΔH_{gas}	−883.27(0.31)	−788.68 (0.24)	−1064.13 (0.57)
ΔH_{solv}	885.91(0.26)	761.93 (0.24)	1031.31 (0.53)
ΔH_{TOTAL}	2.64 (0.12)	−26.74 (0.06)	−32.82 (0.07)
mDNA			
$H_{VDWAALS}$	−137.40(0.10)	−104.72 (0.12)	−74.49 (0.08)
H_{EEL}	−760.59(0.27)	−697.71 (0.33)	−931.54 (0.40)
E_{PB}	757.76(0.24)	670.52 (0.31)	891.30 (0.36)
E_{NPOLAR}	−91.86 (0.06)	−66.75 (0.07)	−59.36 (0.05)
E_{DISPER}	225.40 (0.11)	160.91 (0.16)	141.94 (0.10)
ΔH_{gas}	−898.00 (0.28)	−802.43 (0.36)	−1006.03 (0.44)
ΔH_{solv}	891.30 (0.27)	764.68 (0.33)	973.8731 (0.40)
ΔH_{TOTAL}	−6.70 (0.09)	−37.75 (0.06)	−32.16 (0.07)
$\Delta\Delta H_{mDNA\text{-}nmDNA}$	−9.34 (0.11)	−11.01 (0.06)	0.66 (0.07)

Discussion

Previous authors have pointed out the challenge in explaining the binding selectivity of proteins to methylated DNA [5, 15]. In fact, there may exist alternative mechanisms [5] that apply to different proteins as well as DNA methylation states, involving, e.g., CH···O hydrogen-bonding interactions [14], cation–π interactions, e.g., between arginine residues and cytosine bases [15] and solvation/desolvation effects [9, 39]. In this study, we employed conventional MD simulations and free energy perturbation to unravel structural and energetic parameters that may explain the advantageous specific

Table 3 Partial entropic contribution of methylation to the binding energies (TS, kcal mol^{-1}) using the Schlitter formula at a temperature of 310 K. The entropy was calculated for protein and DNA, DNA alone, and proteins alone (only in the bound form)

	Catalytic domain of R.DPNI			Winged-helix domain of R.DPNI			MBD:DNA		
Total entropy	DNA free	DNA-Protein	Difference	DNA free	DNA-Protein	Difference	DNA free	DNA-Protein	Difference
nmDNA	237.80	1940.32	1702.52	237.80	1916.75	1678.95	369.79	1051.44	681.85
mDNA	236.01	1921.20	1685.19	236.01	1917.42	1681.41	368.05	1059.44	691.39
$T\Delta S_{MDNA\text{-}NMDNA}$			−17.33			2.46			9.54

	Catalytic domain of R.DPNI			Winged-helix domain of R.DPNI			MBD:DNA		
DNA Contribution	DNA free	DNA in Complex	Difference	DNA free	DNA in Complex	Difference	DNA free	DNA in Complex	Difference
nmDNA	237.80	244.01	6.21	237.80	222.27	−15.53	369.79	365.79	−4.00
mDNA	236.01	236.27	0.26	236.01	198.05	−37.96	368.05	364.52	−3.53
$T\Delta S_{mDNA\text{-}nmDNA}$			−5.95			−22.43			0.47

Protein contribution	Catalytic domain of R.DPNI	Winged-helix domain of R.DPNI	MBD:DNA
Protein bound to nmDNA	1718.04	1690.33	480.19
Protein bound to mDNA	1704.11	1714.19	464.78
$T\Delta S_{mDNA\text{-}nmDNA}$	−13.93	23.86	−15.41

binding of methylated DNA to the two domains of the R.DpnI protein and to the MBD domain of the MeCP2 protein over non-methylated DNA forms.

As is commonly observed upon protein:DNA association [40], both binding of methylated and non-methylated DNA to proteins was accompanied by an increased width of the major groove and a shift of the BI/BII equilibrium to smaller values.

Due to the hydrophobic nature of the methyl groups, one may suspect that desolvation of N6-methyladenosine is energetically favored over desolvation of adenosine. In a recent computational study [20], we found, however, that a single non-methylated N base has a very similar solvation free energy as a single N6-methylated adenine base. In contrast, the C5-methylated form of cytosine is strongly favored over the non-methylated form in water (reflected by a 26.7 kcal mol^{-1} difference in solvation free energy). Note that these numbers refer to single bases.

In the context of the DNA strand, steric shielding reduced the solvent accessible surface area of a methyl group attached to adenine about two-fold on average (from 65 $Å^2$ to a range of 28–38 $Å^2$, see Fig. S6). Upon protein binding, this was reduced to only about 3 $Å^2$. The process of water displacement involved more waters for the winged-helix domain, so that their release into the bulk solvent should be entropically favorable, as previously reported [41]. For comparison, the methyl group of a single C5-methylcytosine had a total solvent-exposed area of 55 $Å^2$, which was reduced to about 35–40 $Å^2$ when integrated in the DNA strand (see Fig. S7). In the complex with MBD, the SASA surface was about 10–15 $Å^2$. This suggests that shielding of the C5-methylated cytosine likely plays a smaller role for the binding specificity than for N6-methylated adenine.

According to the previous study of Zou et al. [15] for the MBD:DNA system, the C5-methylation of cytosine contributes about −1.2 kcal mol^{-1} of preferential binding free energy to the interaction between methylated DNA and the MBD domain. Similar free energy perturbation calculations presented here showed that N6-adenine-methylated DNA favored binding to the catalytic domain of the R.DpnI protein by a slightly larger amount (−4.05 ± 0.7 kcal mol^{-1}). These results are consistent with experimental findings that non-methylated DNA shows, at most, weak binding to the catalytic subunit [3]. N6-methylation of adenine also gave a favourable contribution for binding to the winged-helix domain (−1.32 ± 0.34 kcal mol^{-1}).

Computational modelling is also able to address individual contributions to the binding free energy. MM-PBSA calculations predict that N6-adenine-methylation favors binding to the R.dpnI winged-helix domain by −11.01 kcal mol^{-1}, and to the R.dpnI catalytic domain by −9.34 kcal mol^{-1}. This comparably large favorable enthalpic contribution cannot be attributed to non-bonded interactions between the two N6-adenine methyl-groups and protein residues alone. An important role is likely also played by the different conformational adaptation by DNA to the protein domains between methylated and non-methylated DNA. Also, one needs to remember that we refer to differences in binding enthalpies between the solvated and bound states, so that one always needs to consider the unbound state as well. Zou and colleagues [15] have discussed for the Mecp2 system that classical forcefields are likely not able to capture well the characteristic cation-pi interactions formed between arginine residues of the MBD domain and nucleic bases. Omission of these apparently important effects may explain why the difference in binding enthalpy computed was slightly unfavorable in our study.

According to the Schlitter formula, free DNA had lower configurational entropies in the methylated form than in the non-methylated form for all the species studied here. For comparison, one may consider the relative experimental melting points of the two species. According to Smith and colleagues, C5-cytosine-methylated DNA has higher melting temperatures than nonmethylated DNA, so that a lower entropy is expected at the same temperature [42], in agreement with our findings. We note, however, that the calculations only account for the configurational entropy of the solute molecules, whereas the experimental values reflect the entropy of the full system including solvent and ions.

An experimental study by Jen-Jacobson and colleagues characterized the enthalpic and entropic contributions to the binding of several protein:DNA complexes [41]. The authors observed an isothermal entropic–enthalpic compensation for the different systems. A similar case was observed here for DNA binding to the winged-helix domain of R.DpnI where a strong decrease in the conformational entropy of DNA was fully compensated by a corresponding gain in conformational entropy of the protein. On the other hand, N6-adenine methylation disfavored binding to the catalytic subunit of R.DpnI entropically, whereas C5-cytosine methylation had an entropically favorable effect on the interaction of the MBD:DNA complex.

Conclusions

DNA methylation of specific DNA regions is a targeting signal for particular proteins such as transcription factors. According to our findings, specific binding to N6-adenine-methylated or C5-cytosine-methylated DNA is achieved through structural adaptation in DNA and the protein on the one hand, and through the combined effects of more favorable binding enthalpies and modulation of the conformational entropy on the other hand. For the R.DpnI system, a favorable enthalpic contribution seems to play a major role in favoring binding of methylated DNA over the non-methylated DNA. In contrast, specific binding of C5-cytosine-methylated DNA

to the MBD domain of the Mecp2 protein appears to be predominately stabilized by a favorable entropic contribution due to the concerted dynamics of protein and DNA. It remains to be studied whether these characteristics are intrinsic properties of the systems investigated here or whether they transfer to other systems and are, thus, general principles of N6-adenine vs. C5-cytosine methylation.

Acknowledgments We thank Prof. Matthias Bochtler (Warsaw/Poland) for valuable discussions on mechanistic issues of the R.DpnI system and for early access to the crystallographic data. In addition, we thank Dr. Wei Gu for fruitful discussions and helpful suggestions concerning the free energy perturbation calculations.

Compliance with ethical standards

Funding sources This work is embedded in the framework of the collaborative research center SFB 1027 funded by Deutsche Forschungsgemeinschaft (DFG). S.S. thanks the German Academic Exchange Service (DAAD) for a doctoral fellowship. O.U. was supported by DFG.

References

1. Low DA, Weyand NJ, Mahan MJ (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infect Immun 69(12):7197–7204
2. Wu TP, Wang T, Seetin MG, Lai YQ, Zhu SJ, Lin KX, Liu YF, Byrum SD, Mackintosh SG, Zhong M, Tackett A, Wang GL, Hon LS, Fang G, Swenberg JA, Xiao AZ (2016) DNA methylation on N-6-adenine in mammalian embryonic stem cells. Nature 532(7599):329–333
3. Siwek W, Czapinska H, Bochtler M, Bujnicki JM, Skowronek K (2012) Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI. Nucleic Acids Res 40(15):7563–7572
4. Delacampa AG, Springhorn SS, Kale P, Lacks SA (1988) Proteins encoded by DpnI restriction gene cassette- hyperproduction and characterization of the DpnI endonuclease. J Biol Chem 263(29): 14696–14702
5. Mierzejewska K, Siwek W, Czapinska H, Skowronek K, Bujnicki J, Bochtler M (2014) Structural basis of the methylation specificity of R.DpnI. Nucleic Acids Res 42:8745–8754
6. Chahrour M, Jung SY, Shaw C, Zhou XB, Wong STC, Qin J, Zoghbi HY (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 320(5880): 1224–1229
7. Chen WG, Chang Q, Lin YX, Meissner A, West AE, Griffith EC, Jaenisch R, Greenberg ME (2003) Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science 302(5646):885–889
8. Bekinschtein P, Cammarota M, Katche C, Slipczuk L, Rossato JI, Goldin A, Lzquierdo I, Medina JH (2008) BDNF is essential to promote persistence of long-term memory storage. Proc Natl Acad Sci USA 105(7):2711–2716
9. Ho KL, McNae LW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell 29(4):525–531
10. Pabo CO, Sauer RT (1984) Protein-DNA recognition. Annu Rev Biochem 53:293–321
11. Wecker K, Bonnet MC, Meurs EF, Delepierre M (2002) The role of the phosphorus BI-BII transition in protein-DNA recognition: the NF-kappa B complex. Nucleic Acids Res 30(20):4452–4459
12. Ray BK, Dhar S, Henry C, Rich A, Ray A (2013) Epigenetic regulation by Z-DNA silencer function controls cancer-associated ADAM-12 expression in breast cancer: cross-talk between MeCP2 and NF1 transcription factor family. Cancer Res 73(2): 736–744
13. Madhumalar A, Bansal M (2005) Sequence preference for BI/II conformations in DNA: MD and crystal structure data analysis. J Biomol Struct Dyn 23(1):13–27
14. Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, Wright PE (2012) Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci USA 109(38):15229–15234
15. Zou X, Ma W, Solov'yov IA, Chipot C, Schulten K (2012) Recognition of methylated DNA through methyl-CpG binding domain proteins. Nucleic Acids Res 40(6):2747–2758
16. Schenkelberger M, Shanak S, Finkler M, Worst E, Noireaux V, Helms V, Ott A (2017) Expression regulation by a methyl-CpG binding domain in an *E. coli* based, cell-free TX-TL system. Phys Biol. doi:10.1088/1478-3975/aa5d37
17. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4(3):435–447
18. Foloppe N, MacKerell AD (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. J Comput Chem 21(2):86–104
19. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
20. Shanak S., Helms V. (2014) Hydration properties of natural and synthetic DNA sequences with methylated adenine or cytosine bases in the R.DpnI target and BDNF promoter studied by molecular dynamics simulations. J Chem Phys. p. 22D512
21. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald- an n.log(n) method for Ewald sums in large systems. J Chem Phys 98(12):10089–10092
22. Van Gunsteren WF, Berendsen HJC (1988) A leap-frog algorithm for stochastic dynamics. Mol Simul 1(3):173–185
23. Bennett CH (1976) Efficient estimation of free-energy differences from Monte-Carlo data. J Comput Phys 22(2):245–268
24. Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. J Phys Chem B 114(32):10235–10253
25. Mobley DL, Chodera JD, Dill KA (2006) On the use of orientational restraints and symmetry corrections in alchemical free energy calculations. J Chem Phys 125(8):084902
26. Hornak V, Simmerling C (2004) Development of softcore potential functions for overcoming steric barriers in molecular dynamics simulations. J Mol Graph Model 22(5):405–413
27. Beutler TC, Mark AE, Vanschaik RC, Gerber PR, Vangunsteren WF (1994) Avoiding singularities and numerical instabilities in free-energy calculations based on molecular simulations. Chem Phys Lett 222(6):529–539
28. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA (1998) Case, Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. J Am Chem Soc 120(37):9401–9409
29. Baker NA, Sept D, Holst MJ, McCammon JA (2001) The adaptive multilevel finite element solution of the Poisson-Boltzmann equation on massively parallel computers. IBM J Res Dev 45(3–4):427–438
30. Miller BR III, McGee TD Jr, Swails JM, Homeyer N, Gohlke H, Roitberg AE (2012) MMPBSA.py: an efficient program for end-

state free energy calculations. J Chem Theory Comput 8(9):3314–3321
31. Crowley MF, Williamson MJ, Walker RC (2009) CHAMBER: comprehensive support for CHARMM force fields within the AMBER software. Int J Quantum Chem 109(15):3767–3772
32. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph Model 14(1):33–38
33. Roe DR, Cheatham TE III (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J Chem Theory Comput 9(7):3084–3095
34. Furini S, Barbini P, Domene C (2013) DNA-recognition process described by MD simulations of the lactose repressor protein on a specific and a non-specific DNA sequence. Nucleic Acids Res 41(7):3963–3972
35. Schlitter J (1993) Estimation of absolute and relative entropies of macromolecules using the covariance matrix. Chem Phys Lett 215(6):617–621
36. Hartmann B, Piazzola D, Lavery R (1993) BI-BII transitions in B-DNA. Nucleic Acids Res 21(3):561–568
37. Pauling L (1992) The nature of chemical bond. J Chem Educ 69(7): 519–521
38. Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res 31(17):5108–5121
39. Liu Y, Toh H, Sasaki H, Zhang X, Cheng X (2012) An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. Genes Dev 26(21):2374–2379
40. Rohs R, Jin X, West SM, Joshi R, Honig B, Mann RS (2010) Origins of specificity in protein-DNA recognition. Annu Rev Biochem 79(79):233–269
41. Jen-Jacobson L, Engler LE, Jacobson LA (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins. Structure 8(10):1015–1023
42. Smith E, Jones ME, Drew PA (2009) Quantitation of DNA methylation by melt curve analysis. Bmc Cancer 9:123
43. Lu X-J, Olson WK (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. Nat Protoc 3(7):1213–1227