JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Molecular Dynamics Study of the Recognition of Dimethylated CpG Sites by MBD1 Protein

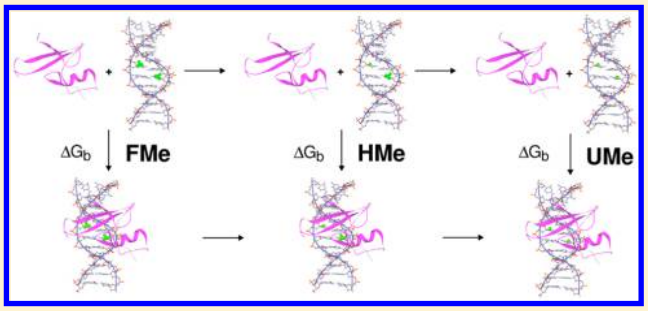
Caterina Bianchi[†] and Ronen Zangi*[,†,‡]

[†]Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018 San Sebastian, Guipúzcoa, Spain

[‡]IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Bizkaia, Spain

Supporting Information

ABSTRACT: The cell is able to regulate which genes to express via chemical marks on the DNA and on the histone proteins. In all vertebrates, the modification on the DNA is methylation at position 5 of the two cytosines present in the dinucleotide sequence CpG. The information encoded by these chemical marks on the DNA is processed by a family of protein factors containing a conserved methyl-CpG binding domain (MBD). Essential to their function, the MBD proteins are able to bind DNA containing dimethylated CpG sites, whereas binding to unmethylated sites is not observed. In this paper, we perform molecular dynamics simulations to investigate the mechanism by which the mCpG binding domain of MBD1 is able to bind specifically dimethylated CpG sites. We find that the binding affinity of MBD1 to a DNA containing dimethylated CpG site is stronger by 26.4 kJ/mol relative to binding the same DNA but with an unmethylated CpG site. The contribution of each of the methyl groups to the change in free energy is very similar and additive. Therefore, this binding affinity (to a dimethylated DNA) is halved when considered relative to binding a hemimethylated DNA, a result that is also supported by experimental observations. Despite their equal contributions, the two methyl groups are recognized differently by MBD1. In one case, demethylation induces conformational changes in which the hydrophobic patch formed by the conserved residues Val20, Arg22, and Tyr34 moves away from the (methyl)cytosine, weakening the DNA−protein interactions. This is accompanied by an intrusion of a bulk water into the binding site at the protein−DNA interface. As a consequence, there is a reduction and rearrangements of the protein−DNA hydrogen bonds including a loss of a crucial hydrogen bond between Tyr34 and the (methyl)cytosine. The methylcytosine on the opposite strand is recognized by conformational changes of the surrounding conserved hydrophobic residues, Arg44 and Ser45, in which Arg44 participate in the 5mC-Arg-G triad. More specifically, the hydrogens of the methyl group form weak hydrogen bonds with the guanidino group and backbone carbonyl of the conserved Arg44, interactions that are absent when the cytosine is unmethylated. The results presented in this paper contribute to our knowledge of the different ways the chemical mark on the DNA is recognized by the epigenetic machinery.



INTRODUCTION

Although all cells of an organism contain the same genetic information, different cells synthesize different proteins and exhibit different phenotypes tailored for executing their function. How does a cell 'know' which genes to express out of the entire genome? This is controlled by chemical marks on the DNA and on the histone proteins. In the DNA of vertebrates, this chemical mark is a covalent addition of a methyl group at position 5 of cytosine (5mC), and it occurs predominantly within CpG dinucleotide sequences.[1] Apart from regulating gene expression, the methylated cytosine marks play important roles in embryonic development, X-chromosome inactivation, and genomic imprinting.[2] The methylation patterns of the DNA pass from mother cells to daughter cells, and their faithful inheritance and maintenance is essential to the wellbeing of the organism.[3] In fact, aberrant methylation patterns of the genome, which consequently alter the genes expressed, have been demonstrated in several human cancers.[4−6]

Often DNA methylation leads to silencing of genes. Currently, two different mechanisms are known by which DNA methylation can switch-off genes.[7,8] The first is by directly preventing the transcription factors to bind their recognition sequences. The second is via modifying the high-order structure of chromatin from a lightly packed state (euchromatin) to a condensed state in which transcription is not possible (heterochromatin). This process is triggered by the binding of a family of methyl-CpG binding proteins (MBD) to the mCpG site. To date, five family members of MBD proteins have been characterized: MeCP2, MBD1, MBD2, MBD3, and MBD4. The first three, have been observed to bind mCpG

Received: October 30, 2014
Published: February 6, 2015

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644

steps and repress gene transcription from methylated gene promoters.[9] This is likely a result of a concurrent binding, via a separate transcriptional repressor domain (TRD), to a histone deacetylase whose activity causes the condensation of the structure of chromatin.[10] MBD3 is the only family member that does not bind mCpG sites, probably due to the lack of several important amino acids that are conserved in the other MBD proteins.

Among the five MBD proteins members mentioned above, MBD1 is unique in that it has CXXC motifs that are able to coordinate zinc. Five isoforms of MBD1, generated by alternative splicing, are known containing either two or three CXXC motifs, and one methyl-CpG binding domain.[11] They bind, as a monomer, to a symmetrically (di- or fully-) methylated CpG site and exhibit selectivity with respect to binding an unmethylated site. This selectivity appears to be independent of the DNA sequence surrounding the mCpG site.[12] MBD1 is also able to compete, to a lesser extent, with binding a hemimethylated step. It has been shown that full-length MBD1 efficiently represses gene expression from both unmethylated and methylated promoters.[13] Inhibition of transcription when the promoters are unmethylated requires the existence of the third CXXC motif, whereas this requirement is not needed when the promoters are methylated.

The methyl-CpG binding domain consists of 60−80 amino acid residues and displays a high sequence homology within the MBD family members.[14] Based on the NMR structure of MBD1[14] it was suggested, and later supported by mutational analysis,[15] that the five residues, Val20, Arg22, Tyr34, Arg44, and Ser45, of the mCpG binding domain are crucial for the recognition of the two methyl groups. Except for Val20 (which in MeCP2 is replaced by Lys20) these residues are highly conserved in all mCpG binding domains. Their side-chains, or segments of them, form a hydrophobic patch presumed to be responsible for the recognition. Note that based on the NMR structure of MeCP2 a similar argument has been proposed for the recognition by MeCP2.[16] In contrast, the X-ray of the same protein showed that the methyl groups contact a hydrophilic surface that includes tightly bound water molecules.[17] It was, therefore, suggested that "MeCP2 recognizes hydration of the major groove of methylated DNA rather than cytosine methylation per se".[17] In fact in MeCP2, the methyl group of the methylcytosine also interacts with a conserved arginine residue of the protein. This arginine residue concurrently forms bifurcated hydrogen bonds (via the O6 and N7 atoms) with the 3′ guanine (on the same strand) of the CpG site. The resulting motif, which is referred to as 5mC-Arg-G triad, exists also in MBD1[14] as well as in several other proteins recognizing methylated cytosine (MBD2, Kaiso, and Zfp57).[18−23]

Zou et al. studied by molecular dynamics simulations the change in the binding free energy of MBD1 to a DNA containing the mCpG site relative to a DNA containing a CpG site.[24] They obtained a preference for binding to the mCpG site; however, the magnitude of this preference was calculated to be only −5.0 kJ/mol. If both substrate DNAs are present in equal amounts, the ratio between the equilibrium populations of the bound complexes at room temperature is only 1:7. In such a case, the recognition is very weak, and it contradicts experimental observations. They also performed *ab initio* studies of only the triad motif (single point calculations from structures extracted from the classical trajectories), and the average value for the difference in the energy of the methylated triad relative to the unmethylated triad was −6.7 kJ/mol. Again, the magnitude of this energetic component to the difference in the binding affinities is too small to explain the recognition of MBD proteins to the mCpG site.

In this paper we perform molecular dynamics simulations to address the mechanism by which MBD1 recognizes dimethylated CpG sites. By using alchemical mutations we were able to reproduce substantial difference in the binding affinity of MBD1 to a dimethylated DNA relative to an unmethylated DNA. Albeit different apparent behavior, the contribution of each of the methyl groups is similar and additive, thus, MBD1 is found to bind weakly hemimethylated DNA. The molecular mechanism governing this recognition is discussed.

METHODS

System Preparation. We perform free energy molecular dynamics simulations in order to calculate the binding affinity of the methyl-binding domain (MBD) of MBD1 protein to a DNA double helix containing a CpG site. Three different methylation states were considered for this CpG site: unmethylated (UMe), hemimethylated (HMe), and fully methylated (FMe) states. The initial conformation for the simulations was taken from an NMR solution structure of MBD (75 amino acids long) of MBD1 bound to a FMe DNA double helix (PDB code: 1IG4).[15] We chose the first model, out of the 20, deposited in the Protein Data Bank. The DNA used in the experiment, as well as two other DNAs with the different methylation states considered in the simulations, are shown in Figure S1 in the Supporting Information. The CpG site is located halfway along the DNA at positions (m)C6pG7/(m)C18pG19 (the prefix (m) is used to indicate a CpG sequence that can be either methylated or unmethylated depending on whether the site is UMe, HMe, or FMe). These 12 base-paired DNAs contribute a charge of −22 *e* due to their phosphate groups. The charged residues of MBD1 are arginine and lysine, which were protonated, as well as glutamate and aspartate, which were deprotonated. In addition, the N and the C termini of the protein were taken to be protonated and deprotonated, respectively. Given these protonations states for the amino acid residues, the total charge of MBD1 was +3 *e*. Consequently, these charges of the protein and DNA were neutralized by 3 chloride anions and 22 sodium cations added at random positions in the simulation box. The dimensions of the cubic simulation box were determined by a minimum distance of 1.2 nm between the MBD1−DNA complex and each of the box edges. The systems were then solvated in water and contained a total of 8639 water molecules. The DNA was described by the parmbsc0 force-field,[25] the protein by the AMBER-99SB force-field,[26] the water molecules by the TIP3P model,[27] and the ions by the AMBER-94 force-field.[28] The partial charges of the 5-methylcytosine, which are not available in the parmbsc0 (or in AMBER-99) force-field, were taken from the work of Rauch et al.[29] These charges were obtained from an *ab initio* calculation using the Restrained ElectroStatic Potential (RESP) charge fitting procedure.[30]

Simulations Details. The molecular dynamics package GROMACS[31] version 4.5.5 was used to perform all simulations, with a time step of 0.002 ps and periodic boundary conditions applied in all three dimensions. The electrostatic forces were evaluated by the Particle-Mesh-Ewald method[32,33] with a real-space cutoff of 1.0 nm, grid spacing of 0.12 nm, and quadratic interpolation. The Lennard-Jones forces were calculated using a 1.0 nm cutoff. The entire system was maintained at 300 K by the velocity rescaling thermostat,[34] with a coupling time of 0.1

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644

ps, and at a pressure of 1.0 bar by the Berendsen barostat[35] with a compressibility of 5×10^{-5} 1/bar and a coupling time of 1.0 ps. Water bond distances and angles were constrained using the SETTLE algorithm,[36] whereas the distances of the protein and DNA covalent bonds were constrained using the LINCS algorithm.[37]

All systems were first energy minimized using the steepest descent method, followed by a 2 ns simulations in which the positions of the protein and the DNA heavy atoms were restrained by a harmonic potential with a force constant of 1000 kJ/(mol·nm^2). Then, 10 ns of unrestrained dynamics was performed to equilibrate the system.

Free Energy Calculations. The relative binding affinities were computed by the concept of a thermodynamic cycle[38] as shown in Figure 1. To this end, alchemical mutations of atom types (with soft-core potentials, $\alpha = 0.7$ and $p = 1$), bonds, angles, and dihedrals were performed to transform the FMe to HMe DNA and, subsequently, to transform the HMe to UMe DNA. Both transformations were performed for the DNA–MBD1 complex and for the DNA free in solution in the forward and backward directions. More specifically, for the FMe to HMe transformation (in the forward direction), the 5-methyl group of the methylcytosine, mC6, was mutated to a hydrogen to form an unmethylated cytosine, C6. Correspondingly, for the HMe to UMe transformation, the methylcytosine, mC18, was converted to unmethylated cytosine, C18.

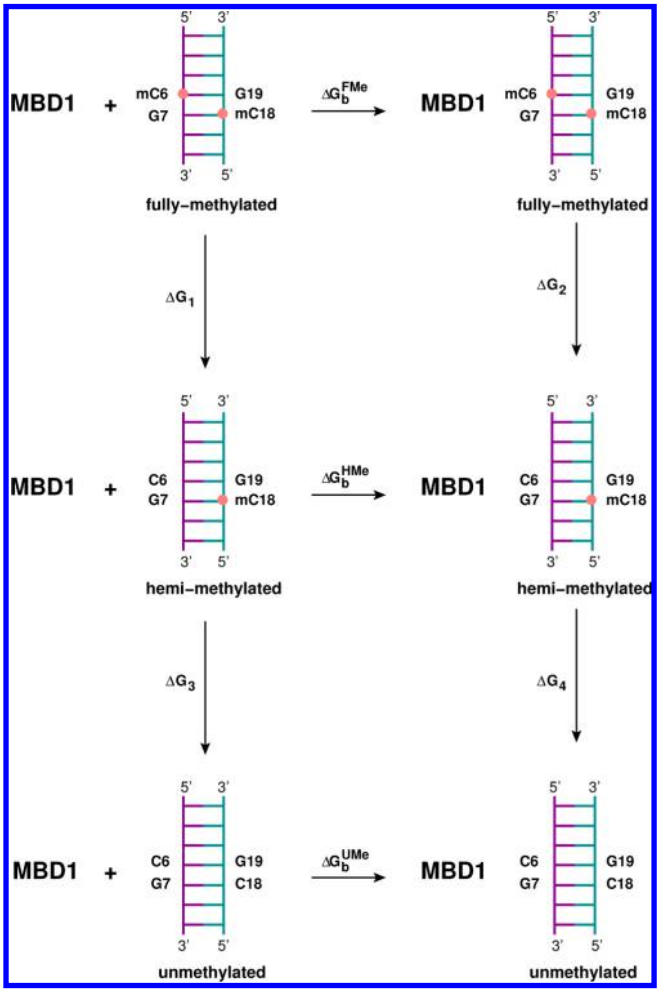


Figure 1. Thermodynamics cycle used for calculating the binding free energy between the protein MBD1 and a DNA containing a fully methylated (dimethylated) CpG site (ΔG_b^{FMe}) relative to the binding to a DNA with an unmethylated CpG site (ΔG_b^{UMe}). The alchemical transformations from fully methylated to unmethylated sites were performed in two steps wherein a hemimethylated CpG site is served as an intermediate state and its relative binding free energy (ΔG_b^{HMe}) can also be determined. The value of the free-energy change obtained from the simulations for each transformation is given in Table 1.

The free energy changes associated with these transformations were computed by the Thermodynamic Integration technique.[39] For each transformation, 13 equally spaced λ-points from $\lambda = 0$ to $\lambda = 1$ were constructed. At each λ-point the system was equilibrated for 5 ns and then data collected for 25 and 45 ns for the mutation of the DNA free in solution and complexed with MBD1, respectively (Figure S2 displays examples of the convergence properties). However, for the MBD1–DNA complex, the plot of $\partial\mathcal{H}/\partial\lambda$ as a function λ did not exhibit a smooth behavior. At locations where the curve was not smooth, we added extra λ-points (maximum eight for each mutation). In general, the starting conformation for a particular λ-point was taken after a relaxation of approximately 1 ns at the precedent λ-point. We considered the calculations to reach convergence when the values of the change in free energy obtained from the forward and backward directions are within the error of the calculations (an example of convergence for one transformation is given in Table S1).

The estimation of the errors of the free energy changes were obtained from[40]

$$\delta\Delta G = \left[\sum_{\lambda=0}^{\lambda=1} (\delta\langle\partial\mathcal{H}/\partial\lambda\rangle)^2\right]^{1/2} \quad (1)$$

where $\delta\langle\partial\mathcal{H}/\partial\lambda\rangle$ is the error in determining the average integrand at each λ-point. The value of $\delta\langle\partial\mathcal{H}/\partial\lambda\rangle$ at each λ-point was evaluated by the block averaging method[41] (see Figure S2). As a result of all the free energy transformations, each of the three methylation states of the DNA is simulated multiple times. Therefore, the average value of the hydrogen bonds were calculated from all these n multiple trajectories, and the error was estimated from the standard deviation divided by $(n-1)^{1/2}$. Correspondingly, the analyses of the distribution of the various distances considered all these multiple trajectories.

RESULTS AND DISCUSSION

Binding Affinities of MBD1 to FMe, HMe, and UMe CpG Sites. Figure 1 displays the thermodynamic cycle constructed to calculate the difference in the binding affinity of MBD1 to the FMe CpG site relative to its binding to the UMe site. The values of the free energy changes of the individual transformations are presented in Table 1. The results indicate that MBD1 prefers to bind the DNA containing FMe CpG site than to bind the same DNA but with UMe CpG. The magnitude of this recognition, $\Delta\Delta G_b^{FMe\text{-}UMe}$, is −26.4 kJ/mol. At $T = 300$ K the ratio between the equilibrium populations of both bound states (obtained from the ratio of the corresponding Boltzmann factors) is 1 to $4 \cdot 10^4$. Such a high fidelity in the recognition is to be expected from a protein involved in reading the epigenetic code.[42] As shown in Figure 1, we transformed the two methyl groups, on the two methylcytosine bases, to hydrogens one at a time. As a consequence, we obtained an intermediate state of the DNA with a HMe site (C6pG7/mC18G19). The contribution of mutating each of the methyl groups (separately) to a hydrogen

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636–644

Table 1. Free Energy Changes of the Alchemical Mutations Shown in Figure 1[a]

	forward	backward	average
ΔG_1	-456.2 ± 1.4	-455.7 ± 2.2	-455.9 ± 1.8
ΔG_2	-440.5 ± 3.2	-445.2 ± 3.0	-442.8 ± 3.1
ΔG_3	-453.3 ± 0.5	-454.3 ± 0.4	-453.8 ± 0.5
ΔG_4	-438.4 ± 2.9	-442.6 ± 3.7	-440.5 ± 3.3
$\Delta\Delta G_b^{\text{FMe-HMe}} = \Delta G_1 - \Delta G_2$		-13.1 ± 4.9	
$\Delta\Delta G_b^{\text{HMe-UMe}} = \Delta G_3 - \Delta G_4$		-13.3 ± 3.8	
$\Delta\Delta G_b^{\text{FMe-UMe}} = \Delta G_1 + \Delta G_3 - \Delta G_2 - \Delta G_4$		-26.4 ± 8.7	

[a]The relative binding free energy changes are calculated from the average of the forward and backward directions. All values are given in kJ/mol.

is almost the same and equals half of the total relative binding affinity ($\Delta\Delta G_b^{\text{FMe-HMe}} \sim \Delta\Delta G_b^{\text{HMe-UMe}} \sim -13$ kJ/mol). These results are in good agreement with *in vitro* experimental results. More specifically, using gel retardation assay, it has been reported that MBD1 (as well as MBD2 and MBD4) binds specifically DNA containing FMe CpG site(s), whereas it does not bind the same strands but with UMe CpG site(s).[43] Furthermore, by introducing a 100-fold excess of the corresponding HMe oligonucleotide, MBD1 (as well as MBD4) exhibited competitive binding to this DNA.[12] This means that MBD1 can also bind the HMe CpG site but with affinity much smaller compared with the binding to a FMe site. Note however that quantitative comparison with experimental results is not possible due to lack of experimental data.

In order to explain this specificity we calculate in Table 2 the average number of hydrogen bonds between MBD1 and the CpG site of the DNA. In the transformation from FMe to HMe DNA we modified the 5-methyl group of C6. A decrease of 0.9 hydrogen bonds is observed for this cytosine, which is partially compensated by an increase of 0.5 hydrogen bonds associated with the unmodified methylcytosine on the complementary strand. Very similar changes in the number of hydrogen bonds occur for the transformation from HMe to UMe DNA. In general, the lost hydrogen bonds of the CpG sites with the protein are replaced by hydrogen bonds with the surrounding water molecules, a process weakening the MBD1−DNA interaction. Nevertheless, in both cases, the net change in the number of MBD1−CpG hydrogen bonds is not large enough to explain the entire 13 kJ/mol reduction in the binding affinity when demethylating each of the cytosines.

Table 2. Average Number of Hydrogen Bonds between the Protein MBD1 and the CpG Site of the DNA with Different Methylation States: FMe, HMe, and UMe States[a]

	FMe	HMe	UMe
(m)C6G7 − MBD1	**7.8 ± 0.1**	**6.9 ± 0.3**	7.5 ± 0.2
(m)C6G7 − solvent	**9.4 ± 0.1**	**10.5 ± 0.4**	9.7 ± 0.2
(m)C18G19 − MBD1	3.3 ± 0.3	**3.8 ± 0.1**	**2.9 ± 0.1**
(m)C18G19 − solvent	15.7 ± 0.3	**15.2 ± 0.1**	**16.2 ± 0.2**

[a]The corresponding numbers of hydrogen bonds between the CpG site and the solvent water molecules are also displayed. Note that in the transformation from FMe to HMe DNA, mC6 is converted into C6 (mC18 is unchanged), whereas in the transformation from HMe to UMe DNA, mC18 is converted to C18 (C6 is unchanged). The values related to these transformations are bold-faced.

The Mechanism for Binding FMe DNA Stronger than HMe DNA. The conversion of the 5-methyl group on cytosine to a hydrogen is accompanied by a decrease in the excluded volume at this site. Does this decrease induce conformational changes? Do water molecules enter the space that was occupied by the methyl group? In Figure 2a we plot the radial distribution function between C5Me/H5 of (m)C6 and the oxygen atom of the water molecules. When transforming FMe to HMe DNA there is a large increase in the probability to observe a water molecule at the site of the mutation (as marked by the peak at $r < 0.4$ nm for the HMe complex). We find that the average number of waters within a radius of 0.4 nm from

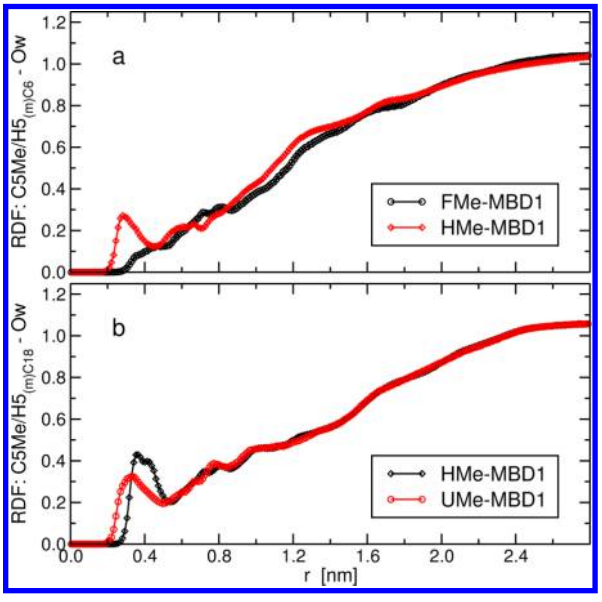


Figure 2. (a) The radial distribution function between the carbon atom of the methyl group (C5Me), or the hydrogen atom (H5), of (m)C6 and the oxygen atoms of the water molecules. These C5Me and H5 atoms define FMe and HMe DNAs, respectively, and the analyses were performed for the MBD1−DNA bound complexes. (b) Same as (a) but for C5Me/H5 of (m)C18 in HMe/UMe MBD1−DNA bound complexes.

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644

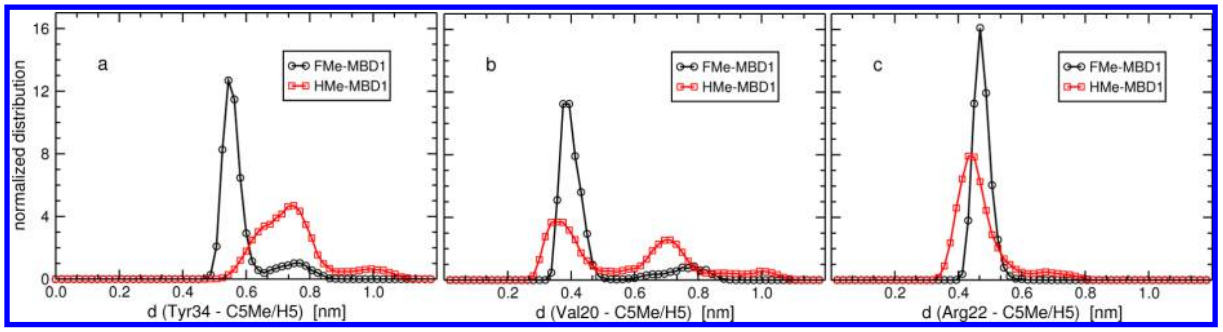


Figure 3. Normalized distribution of the distance between the atom C5Me/H5 of (m)Cyt6 (of FMe/HMe DNA) and the center of mass of the hydrophobic segment of (a) Tyr34, (b) Val20, and (c) Arg22 of MBD1 bound to FMe and HMe DNAs.

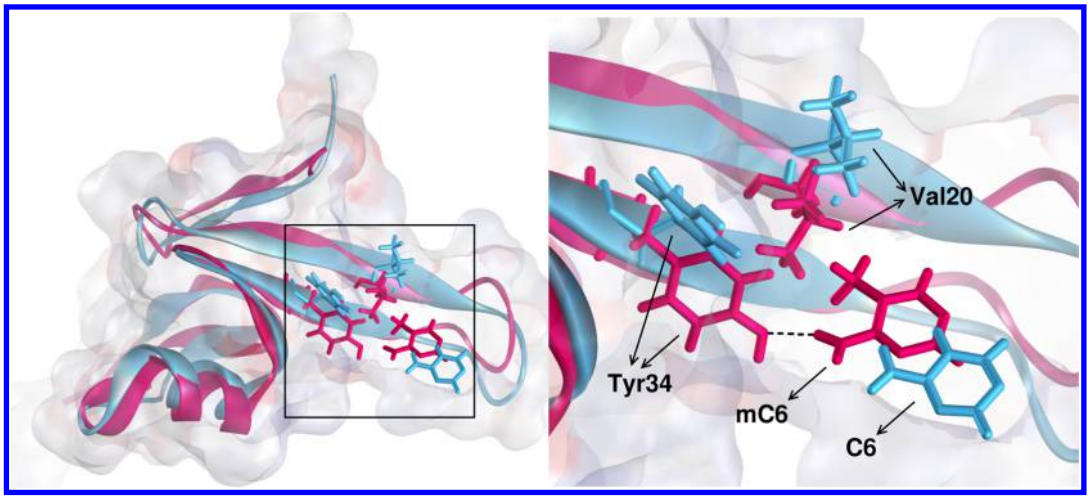


Figure 4. Superposition of the FMe (magenta) and HMe (blue) DNA−MBD1 complexes showing the conformational changes of the hydrophobic pocket around (m)C6. Tyr34 and Val20 as well as (m)C6 are represented by sticks. Magnification of the content inside the square is displayed on the right. The hydrogen bond between the NH_2 group of mC6 and the hydroxyl group of Tyr34 in the FMe DNA complex is indicated by a dashed line.

C5Me/H5 is 0.44 in the FMe complex, whereas it is 1.49 in the HMe complex. This extra water molecule is actually in contact with the hydrogen at position 5 of C6 in the HMe complex; however, analogous interaction between the methyl carbon of mC6 and surrounding waters is not evident for the FMe case. This reduces significantly the strength of the interaction between the protein and the DNA because the intruding water molecule does not form a bridged hydrogen bond between the protein and the DNA.

A methyl group also has a hydrophobic character. As mentioned in the Introduction, it is argued that the recognition of the methyl groups is due to five conserved amino acids forming a hydrophobic patch. These residues are Val20, Arg22, Tyr34, Arg44, and Ser45, where the first three and the last two are relevant for mC6 and mC18 recognition, respectively. In Figure 3 we present the distributions of the distance between the center of mass of the hydrophobic segments of the first three residues and methyl carbon or hydrogen at position 5 of (m)C6. A large increase in the distance is exhibited for the hydrophobic segment of Tyr34 when transforming FMe to HMe DNAs implying weakening MBD1−DNA hydrophobic interactions. For Val20, both the FMe and HMe complexes display bimodal distributions; however, for the HMe complex the mode with the larger distance exhibits a much higher population and again suggesting weakened protein−DNA affinity. The distributions of the Arg22 do not display a large change. The hydrophobic character of these residues is probably a necessary condition but not sufficient to keep the bound state. Electrophoretic mobility shift assay indicated that the Y34F and Y34A mutations resulted in proteins that do not bind DNA containing the FMe CpG site because the hydrogen bond between the hydroxyl of tyrosine and the NH_2 group of cytosine is also crucial for binding[15,43] (it is worth mentioning that the protein MBD3 which does not bind FMe CpG sites has a phenylalanine residue instead of tyrosine at this position).

Instantaneous conformations displaying the positioning of Tyr34 and Val20 relative to (m)C6 in FMe and HMe complexes are shown in Figure 4. This figure reveals that in addition to the increased distance between (m)C6 and Val20, the hydrogen bond between the hydroxyl of tyrosine and N4 of (methyl)cytosine is not formed in the HMe complex. Indeed, the average population of this hydrogen bond (over the all trajectories) in the FMe complex is 0.69, whereas it is only 0.11 in the HMe complex. The reduction in the number of hydrogen bonds, the intrusion of a water molecule to the interfacial cavity formed by the elimination of the methyl group at C6, and the weaker hydrophobic interactions between the protein and the DNA can explain the weaker binding of MBD1 to HMe DNA compared with that to FMe DNA.

The Mechanism for Binding HMe DNA Stronger than UMe DNA. In Figure 2b we display the radial distribution function between C5Me/H5 of (m)C18 and the oxygen atom of the water molecules. In contrast to Figure 2a, here a bulk water is present at the binding site when (m)C18 is methylated

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644

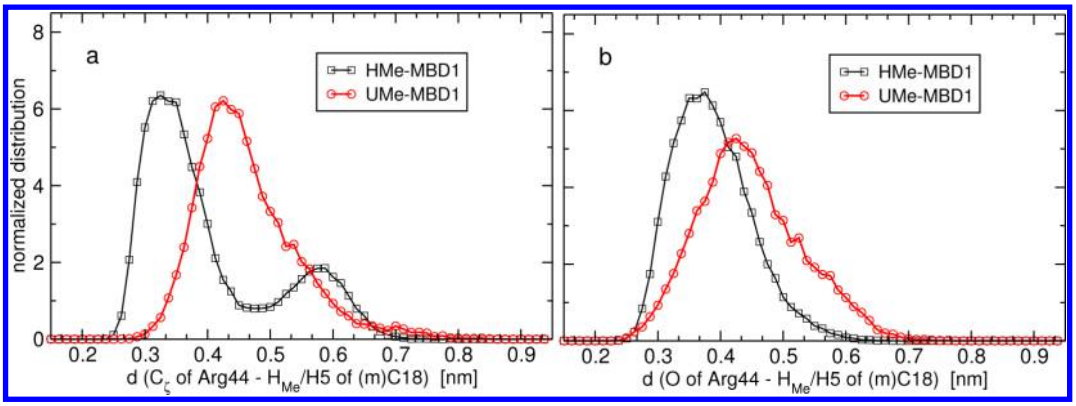


Figure 5. Normalized distribution of the distance between the closest methyl-hydrogen (HMe complex) or H5 (UMe complex) of (m)C18 and (a) the central carbon atom, C_ζ, of the guanidino group and (b) the backbone carbonyl oxygen, of Arg44.

and when it is demethylated. In Figure S3 we plot the relative positioning of the hydrophobic residues that form a hydrophobic cage around this cytosine base, (m)C18. Figure S3a indicates that Arg44 is on average at the same distance in both HMe and UMe complexes. However, in the UMe complex it displays a larger degree of flexibility. For the case of Ser45 (Figure S3b), it is evident that this amino acid moves slightly away from the (m)Cyt18 in the case of the complex with the UMe DNA. This is likely to weaken the hydrophobic interactions between (m)C18 and the protein explaining at least partially the difference in the binding affinity of MBD1 toward HMe and UMe DNAs. Note that the slight shifts of the red curves (i.e., the HMe−MBD1 complex in Figure 2a and the UMe−MBD1 complex in Figure 2b) toward smaller distances is because we considered C5Me instead of HMe in the methylated cytosine complexes. The analysis of the change in the MBD1−CpG hydrogen bonds indicates that it is rather similar in both transformations (Table 2) and can provide another energetic source for the discrimination.

Nevertheless, experimentally, it was found that also Arg44 plays a crucial role in the recognition. Mutating Arg44 to lysine or alanine abolishes the binding of the MBD1 to FMe DNA strands.[15] The mutation to lysine actually preserves the positive charge and the hydrogen bonding potential of the side-chain, thus, it might be that other characteristics of the guanidino group (of Arg44) are important, especially, the interaction of Arg44 with the 5-methyl group of mC18. Note that upon binding HMe DNA, Arg44 undergoes a large conformational change relative to its positioning when MBD1 is bound to FMe DNA. More specifically, in the HMe complex this arginine is closer to the methyl group of mC18. This rearrangement is shown as a bimodal distribution in Figure 5a for the HMe−MBD1 complex, where we plot the distance between the central atom of the guanidino group of Arg44 (C_ζ) and the closest hydrogen of the 5-methyl group or H5 of (m)C18. The larger distances with a smaller population correspond to the mode when binding DNA with a FMe site. The methyl hydrogen of the HMe complex is closer by 1.0 Å to the guanidino group than the aromatic hydrogen H5 in the UMe complex. The methyl hydrogens also interact with the backbone carbonyl oxygen of Arg44, and the corresponding change in the distance relative to that of H5 in the UMe complex is 0.7 Å (Figure 5b). Representative snapshots of these interactions are shown in Figure S4.

Are the interactions between the methyl hydrogens and the guanidino group as well as with the backbone amide carbonyl group important and do they contribute to $\Delta\Delta G_b^{\text{HMe-UMe}}$? When converting methylcytosine to cytosine, two major events occur. The first is the conversion of a methyl group to a hydrogen, and the second is changing the partial charges of the pyrimidine ring as a result of the demethylation. In previous studies of the recognition of HMe DNA strands by UHRF1 we found that the recognition can be due to either a change of the partial charges around the cytosine ring upon (de)-methylation[44] or a steric repulsion of the excluded volume of the methyl group.[45] Therefore, in order to isolate the contribution of the change in the charges of the cytosine ring from the interactions involving the methyl group, we constructed another thermodynamic cycle, that passes via an intermediate state, to calculate $\Delta\Delta G_b^{\text{HMe-UMe}}$ (see Figure S5). The results of these transformations (Table S2) indicate that the dominant contribution for the preference of MBD1 for HMe CpG sites over UMe sites arises from the step in which the methyl group at position 5 is converted to a hydrogen. Thus, the interaction between this methyl group and Arg44 is important.

As mentioned in the Introduction, the X-ray structures of several proteins recognizing methylated cytosine reveal that the methyl group of cytosine interacts with the guanidino group of an arginine of the protein that is part of the 5mC-Arg-G triad. Zou et al. calculated quantum mechanically the interaction energy of this ternary system for HMe and UMe complexes.[24] They used single point calculations of structures extracted from classical trajectories and obtained that, on average, the HMe complex does exhibit lower energy of about 3.3 kJ/mol (Table S1 in their Supplementary Information). Energy optimization of such an isolated (m)C-Arg-G triad is not possible because the minimized structure does not correspond to the structure extracted from the entire protein−DNA complex. The explanation for the stronger interaction with the methylated cytosine can be due to π-stackings that are enhanced upon methylation, dispersion interactions of the methyl group, or weak hydrogen bonding of the methyl hydrogens with electronegative atoms. The fact that our free energy calculations (performed classically) reproduced the discriminations against unmethylated cytosine suggests that the explanation is not likely to be due to a change in the stacking interactions (which require quantum description). It is known that there are interactions, albeit weak, between hydrogens attached to a carbon and an electronegative atom, A, such as oxygen and nitrogen.[46,47] This interaction has been shown to exist abundantly in proteins by analyzing their crystal structures.[48]

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644

In this case, the distance from the hydrogen to the acceptor displays a large range 2.1−3.4 Å, and the C−H···A angle can display a large distortion from linearity 78°−173°.

In order to shed light on the reason why the methylcytosine interacts stronger than cytosine in the (m)C-Arg-G triad, we calculated quantum mechanically the binding energy of a methane molecule to a guanidino group and to an amide carbonyl group. The results are shown in Table S3 indicating that the binding energy of the methane to the guanidino group and the carbonyl group are −6.6 and −3.6 kJ/mol, respectively. So when combined, the absence of these two binding energies can have a relatively large magnitude. For comparison, we also calculated the strength of the binding interaction with a water molecule which yields a value of −1.7 kJ/mol (note that quantum calculations using Symmetry Adapted Perturbation Theory yield larger values for this interaction energy[49]). It is difficult to know whether the interactions of the methane molecule with these two segments of arginine originate mainly from dispersion or from the weak CH···A hydrogen bonds. The optimized structures obtained in the calculations are shown in Figure S6. They indicate that a hydrogen of the methyl group interacts with the three nitrogens of the guanidino group and therefore points toward the central carbon atom C_ζ, a configuration very similar to that obtained from the classical simulations for both mC6−Arg22 and mC18−Arg44 interactions (see for example Figure S4). The distance to the three nitrogens range from 3.2 to 3.3 Å, and the C−H···A angle is in the range 102°−144°. In regard to the interactions with the amide carbonyl group, the closest distance between any of the methyl hydrogens and this backbone moiety is with respect to the carbonyl oxygen, 3.0 Å, with an angle of 146°. Analysis of 43 crystal structures of protein−DNA complexes found a relatively large number of CH···O contacts involving the thymine methyl group and position 5 of cytosine.[50] Whereas the magnitude of the distortion observed in this analysis is the same as in our case, the distances exhibited lower values. Taken together, these observations suggest that the interactions of a methyl group with the guanidino group within the mCyt6−Arg22 and mCyt18−Arg44 interactions might be classified as weak CH···A hydrogen bonds, however, because the distances we observed are relatively large dispersion interactions and cannot be excluded. Note that if we consider the change of these interactions upon demethylation, the (m)Cyt18−Arg44 interaction seems to contribute more to the change in the binding affinity (thus, to $\Delta\Delta G_b^{\text{HMe-UMe}}$). This is because the distance between H_{Me}/H5 and the guanidino group changed, as shown in Figure 5a, by about 1.1 Å, whereas in the (m)Cyt6−Arg22 interactions this distance changes only by about 0.4 Å. Therefore, it is not clear from the simulations to what extent these CH···A weak hydrogen bonds contribute to $\Delta\Delta G_b^{\text{FMe-HMe}}$.

CONCLUSIONS

In this study we performed molecular dynamics simulations to address the recognition of DNA containing a fully methylated CpG site by the protein MBD1. We find that the magnitude of the discrimination against the same DNA but with an unmethylated CpG site is −26.4 kJ/mol. Recognition with such an order of magnitude is in agreement with experimental results indicating that MBD1 does not bind at all unmethylated DNAs. When only one of the cytosines is methylated, thus, a DNA with a hemimethylated CpG site, the relative binding affinity of the fully methylated DNA is halved (about −13 kJ/mol), thus, the contribution of each of the methylated cytosines is more or less equal. This result is also supported experimentally by the observation of a weak binding of MBD1 to a DNA containing hemimethylated site. Nevertheless, the mechanism of the recognition of the two methyl groups is different. In the case of one of the methylcytosines, the hydrophobic pocket enclosing the methyl group undergoes large displacements upon demethylation and moves away from the cytosine, and therefore, the hydrophobic interaction between the protein and the DNA is reduced. Concurrently, a bulk water molecule intrudes into the binding site at the interface between the protein and the DNA. As a consequence, the hydrogen bond network around the CpG site rearranges, and there is a small decrease in the net number of protein−DNA hydrogen bonds, including a loss of a crucial hydrogen bond between a conserved tyrosine residue and the (methyl)-cytosine base. In the case of the methylcytosine on the opposite strand, there is a large conformational change of the surrounding conserved hydrophobic residues Arg44 and Ser45. The amino acid Arg44 forms a 5mC-Arg-G triad motif, and we find (using a thermodynamic cycle with an unphysical intermediate state) that this interaction between the methyl hydrogens and the guanidino group, as well as the backbone carbonyl, of an analogous arginine (which is also part of another 5mC-Arg-G motif) also contributes for discriminating between methylcytosine and cytosine. The nature of the interactions between the arginine groups and the methyl hydrogens are weak CH···A hydrogen bonds that are absent for unmethylated cytosine. These findings shed light on the mechanism by which the 5mC-Arg-G triad can recognize intrahelical methylated cytosines.

ASSOCIATED CONTENT

Supporting Information

A few figures and a table discussed in the manuscript. More specifically, the sequences of the DNA considered in the simulations, a scheme of the thermodynamic cycle via an intermediate state, results from *ab initio* binding energies, and few normalized distributions of interatomic distances are shown. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: r.zangi@ikerbasque.org.

Funding

This work has been funded with support from the European Commission, Marie Curie International Reintegration Grant, project number 247485 and from the Basque Government under the SAIOTEK program, project code S-PE12UN014.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Technical and human support provided by SGIker (USED SERVICES) (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged.

REFERENCES

(1) Jeltsch, A. Beyond Watson and Crick: DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *ChemBioChem* **2002**, *3*, 274−293.

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644

(2) Goto, T.; Monk, M. Regulation of X-Chromosome Inactivation in Development in Mice and Humans. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 362−378.
(3) Bird, A. DNA Methylation Patterns and Epigenetic Memory. *Genes Dev.* **2002**, *16*, 6−21.
(4) Denissenko, M. F.; Chen, J. X.; shong Tang, M.; Pfeifer, G. P. Cytosine Methylation Determines Hot Spots of DNA Damage in the Human P53 Gene. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 3893−3898.
(5) Ehrlich, M. DNA Methylation in Cancer: Too much, but also Too Little. *Oncogene* **2002**, *21*, 5400−5413.
(6) Szyf, M. Targeting DNA Methylation in Cancer. *Ageing Res. Rev.* **2003**, *2*, 299−328.
(7) Meehan, R. R.; Lewis, J. D.; McKay, S.; Kleiner, E. L.; Bird, A. P. Identification of a Mammalian Protein that Binds Specifically to DNA Containing Methylated CpGs. *Cell* **1989**, *58*, 499−507.
(8) Wade, P. A.; Gegonne, A.; Jones, P. L.; Ballestar, E.; Aubry, F.; Wolffe, A. P. Mi-2 Complex Couples DNA Methylation to Chromatin Remodelling and Histone Deacetylation. *Nat. Genet.* **1999**, *23*, 62−66.
(9) Nan, X.; Tate, P.; Li, E.; Bird, A. DNA Methylation Specifies Chromosomal Localization of MeCP2. *Mol. Cell. Biol.* **1996**, *16*, 414−421.
(10) Nan, X.; Meehan, R. R.; Bird, A. Dissection of the Methyl-CpG Binding Domain from the Chromosomal Protein MeCP2. *Nucleic Acids Res.* **1993**, *21*, 4886−4892.
(11) Fujita, N.; Takebayashi, S.-i.; Okumura, K.; Kudo, S.; Chiba, T.; Saya, H.; Nakao, M. Methylation-Mediated Transcriptional Silencing in Euchromatin by Methyl-CpG Binding Protein MBD1 Isoforms. *Mol. Cell. Biol.* **1999**, *19*, 6415−6426.
(12) Hendrich, B.; Bird, A. Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol. Cell. Biol.* **1998**, *18*, 6538−6547.
(13) Jørgensen, H. F.; Ben-Porath, I.; Bird, A. P. MBD1 Is Recruited to both Methylated and Nonmethylated CpGs via Distinct DNA Binding Domains. *Mol. Cell. Biol.* **2004**, *24*, 3387−3395.
(14) Ohki, I.; Shimotake, N.; Fujita, N.; Nakao, M.; Shirakawa, M. Solution Structure of the methyl-CpG Binding Domain of the Methylation-Dependent Transcriptional Repressor MBD1. *EMBO J.* **1999**, *18*, 6653−6661.
(15) Ohki, I.; Shimotake, N.; Fujita, N.; Jee, J.-G.; Ikegami, T.; Nakao, M.; Shirakawa, M. Solution Structure of the Methyl-CpG Binding Domain of Human MBD1 in Complex with Methylated DNA. *Cell* **2001**, *105*, 487−497.
(16) Wakefield, R. I.; Smith, B. O.; Nan, X.; Free, A.; Soteriou, A.; Uhrin, D.; Bird, A. P.; Barlow, P. N. The Solution Structure of the Domain from MeCP2 that Binds to Methylated DNA. *J. Mol. Biol.* **1999**, *291*, 1055−1065.
(17) Ho, K. L.; McNae, I. W.; Schmiedeberg, L.; Klose, R. J.; Bird, A. P.; Walkinshaw, M. D. MeCP2 Binding to DNA Depends upon Hydration at Methyl-CpG. *Mol. Cell* **2008**, *29*, 525−531.
(18) Luscombe, N. M.; Laskowski, R. A.; Thornton, J. M. Amino Acid-Base Interactions: A Three-Dimensional Analysis of Protein-DNA Interactions at an Atomic Level. *Nucleic Acids Res.* **2001**, *29*, 2860−2874.
(19) Lamoureux, J. S.; Maynes, J. T.; Glover, J. M. Recognition of 5′-YpG-3′ Sequences by Coupled Stacking/Hydrogen Bonding Interactions with Amino Acid Residues. *J. Mol. Biol.* **2004**, *335*, 399−408.
(20) Quenneville, S.; Verde, G.; Corsinotti, A.; Kapopoulou, A.; Jakobsson, J.; Offner, S.; Baglivo, I.; Pedone, P. V.; Grimaldi, G.; Riccio, A.; Trono, D. Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. *Mol. Cell* **2011**, *44*, 361−372.
(21) Scarsdale, J. N.; Webb, H. D.; Ginder, G. D.; Williams, D. C. Solution Structure and Dynamic Analysis of Chicken MBD2 Methyl Binding Domain Bound to a Target-Methylated DNA Sequence. *Nucleic Acids Res.* **2011**, *39*, 6741−6752.
(22) Buck-Koehntop, B. A.; Stanfield, R. L.; Ekiert, D. C.; Martinez-Yamout, M. A.; Dyson, H. J.; Wilson, I. A.; Wright, P. E. Molecular Basis for Recognition of Methylated and Specific DNA Sequences by the Zinc Finger Protein Kaiso. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 15229−15234.
(23) Liu, Y.; Zhang, X.; Blumenthal, R. M.; Cheng, X. A Common Mode of Recognition for Methylated CpG. *Trends Biochem. Sci.* **2013**, *38*, 177−183.
(24) Zou, X.; Ma, W.; Solov'yov, I. A.; Chipot, C.; Schulten, K. Recognition of Methylated DNA through Methyl-CpG Binding Domain Proteins. *Nucleic Acids Res.* **2012**, *40*, 2747−2758.
(25) Pérez, A.; March'an, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γConformers. *Biophys. J.* **2007**, *92*, 3817−3829.
(26) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712−725.
(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(28) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(29) Rauch, C.; Trieb, M.; Wellenzohn, B.; Loferer, M.; Voegele, A.; Wibowo, F. R.; Liedl, K. R. C5-Methylation of Cytosine in B-DNA Thermodynamically and Kinetically Stabilizes BI. *J. Am. Chem. Soc.* **2003**, *125*, 14990−14991.
(30) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.
(31) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(32) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(33) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(34) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(35) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to An External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(36) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(37) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(38) van Gunsteren, W. F.; Berendsen, H. J. C. Thermodynamic Cycle Integration by Computer Simulation as a Tool for Obtaining Free Energy Differences in Molecular Chemistry. *J. Comput.-Aided Mol. Des.* **1987**, *1*, 171−176.
(39) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(40) Guàrdia, E.; Rey, R.; Padró, J. A. Statistical Errors in Constrained Molecular Dynamics Calculations of the Mean Force Potential. *Mol. Simul.* **1992**, *9*, 201−211.
(41) Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91*, 461−466.
(42) Bird, A.; Macleod, D. Reading the DNA Methylation Signal. *Cold Spring Harbor Symp. Quant. Biol.* **2004**, *69*, 113−118.
(43) Fujita, N.; Shimotake, N.; Ohki, I.; Chiba, T.; Saya, H.; Shirakawa, M.; Nakao, M. Mechanism of Transcriptional Regulation by Methyl-CpG Binding Protein MBD1. *Mol. Cell. Biol.* **2000**, *20*, 5107−5118.

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644

(44) Bianchi, C.; Zangi, R. How to Distinguish methyl-Cytosine from Cytosine with High Fidelity. *J. Mol. Biol.* **2012**, *424*, 215−224.
(45) Bianchi, C.; Zangi, R. UHRF1 Discriminates Against Binding to Fully-methylated CpG-Sites by Steric Repulsion. *Biophys. Chem.* **2013**, *171*, 38−45.
(46) Sutor, D. J. Evidence for the Existence of C-H···O Hydrogen Bonds in Crystals. *J. Chem. Soc.* **1963**, 1105−1110.
(47) Taylor, R.; Kennard, O. Crystallographic Evidence for the Existence of C-H···O, C-H···N, and C-H···Cl Hydrogen Bonds. *J. Am. Chem. Soc.* **1982**, *104*, 5063−5070.
(48) Derewenda, Z. S.; Lee, L.; Derewenda, U. The Occurrence of C-H···O Hydrogen Bonds in Proteins. *J. Mol. Biol.* **1995**, *252*, 248−262.
(49) Akin-Ojo, O.; Szalewicz, K. Potential Energy Surface and Second Virial Coefficient of Methane-Water from ab initio Calculations. *J. Chem. Phys.* **2005**, *123*, 134311.
(50) Mandel-Gutfreund, Y.; Margalit, H.; Jernigan, R. L.; Zhurkin, V. B. A Role for CH···O Interactions in Protein-DNA Recognition. *J. Mol. Biol.* **1998**, *277*, 1129−1140.

DOI: 10.1021/ci500657d
J. Chem. Inf. Model. 2015, 55, 636−644