jmb Article

How to Distinguish Methyl-Cytosine from Cytosine with High Fidelity

Caterina Bianchi[1] and Ronen Zangi[1,2]

*1 - **Department of Organic Chemistry I,** University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018 San Sebastian, Spain*
*2 - **Ikerbasque,** Basque Foundation for Science, 48011 Bilbao, Spain*

***Correspondence to Ronen Zangi:** r.zangi@ikerbasque.org*
http://dx.doi.org/10.1016/j.jmb.2012.09.024
Edited by D. Case

Abstract

Methylation of cytosines in the DNA is central to the epigenetic code. The patterns along the DNA formed by these chemical marks instruct the cell which proteins to express and their faithful maintenance after replication are vital to the organism's life. Although Dnmt1 is the enzyme catalyzing the methylation reaction, it was found that UHRF1 (*u*biquitin-like, containing P*H*D and *R*ING *f*inger domain *1*) is the protein that actually recognizes hemi-methylated CpG sites. Nevertheless, the physical mechanism driving the strikingly robust distinction between hemi-methylated and unmethylated sites is not known. In this paper, we show that the large difference in the binding affinities of UHRF1 to these sites is possible not due to the presence of the methyl group itself but is a result of the accompanying changes in the distribution of the electrons around the cytosine ring. In particular, methylation reduces the dipole moment of cytosine and, as a consequence, unmethylated DNA in its unbound state in water is more stable than hemi-methylated DNA. Furthermore, the interaction energy of hemi-methylated DNA bound to UHRF1 with its surrounding is stronger than that of unmethylated DNA. Thus, the change in the electronic structure of cytosine upon methylation destabilizes the unbound state and stabilizes the bound state rendering discrimination with high fidelity possible.

Introduction

All cells of an organism have the same genetic information stored in their DNA sequences. However, cells belonging to different tissues exhibit very different phenotypes and perform different functions. What is the mechanism guiding different cells to express different genes? Chemical marks on the DNA (especially in the promoter regions) and on histone proteins are known to switch between active and inactive states of genes. These chemical marks have also important roles in embryonic development, X-chromosome inactivation, and genomic imprinting.[1,2] In the DNA of mammals, the epigenetic mark is a methyl group covalently attached to cytosine (at position C5) in the dinucleotide sequence CpG. The methylation patterns of the DNA pass from mother cells to daughter cells, and their faithful inheritance and maintenance is essential to the well-being of the organism.[3] This can be accomplished owing to the ability of the DNA replication machinery to distinguish hemi-methylated DNA from either unmethylated or symmetrically di-methylated DNA strands. The robustness of recognizing hemi-methylated DNA is very high as evident by the fact that patterns of methylated cytosine marks are propagated with fidelity of more than 99% and their stable inheritance for more than 80 cell generations.[4]

The enzyme responsible for maintaining the epigenetic marks on the DNA after replication is Dnmt1.[5] It catalyzes the transfer of a methyl group from *S*-adenosyl-L-methionine to the cytosine of hemi-methylated DNA.[6] It has been shown that the protein UHRF1 (*u*biquitin-like, containing P*H*D and *R*ING *f*inger domain *1*) is also essential for maintaining DNA methylation.[7,8] UHRF1 exhibits strong preferential binding to hemi-methylated DNA through its SRA (*S*ET and *R*ING-*a*ssociated)

domain, which is a methyl DNA binding domain. UHRF1 can also bind to the N-terminal domain of Dnmt1 and, therefore, is able to play a role in the correct loading of Dnmt1 to hemi-methylated CpG sites. Consequently, Dnmt1 is able to recognize and methylate the proper target cytosine on the complementary strand transforming hemi-methylated to di-methylated site (see Fig. S1). A distinct character of UHRF1 is its multiple conserved domains that are able to bind different specific sites, such as methylated histone H3 lysine 9 and histone deacetylase 1. This provides UHRF1 a unique function in regulating the epigenome because it links DNA methylation with histone marks.[9,10]

The crystal structure of the SRA domain of UHRF1 complexed with hemi-methylated DNA was obtained by four different groups.[11–14] All reported structures indicate that the protein–DNA interaction involves the flipping of the recognized methyl-cytosine out of the DNA helix, presumably to prevent the sliding of the DNA. Watson–Crick hydrogen bonds between the methyl-cytosine and its paired guanine, which are lost, are replaced by hydrogen bonds between the methyl-cytosine and the binding pocket of UHRF1, as well as by hydrogen bonds between the orphan guanine and a "finger", Asn-Lys-Arg, of the protein that intrudes into the DNA double helix. Furthermore, the stabilization due to base stacking interactions in the DNA is replaced by interactions that the flipped methyl-cytosine make with two tyrosine residues of the protein located on both sides of the base plane.

Therefore, it is the protein UHRF1 that makes the distinction between hemi-methylated DNA and either unmethylated or di-methylated DNA strands. Binding to di-methylated DNA is much less harmful than binding to unmethylated DNA because, in the former, the target cytosine on the complementary strand is already methylated, whereas in the latter, it is not. If UHRF1 will bind unmethylated DNA, the subsequent methylation of the cytosine base on the opposite strand will lead to aberrant methylation pattern that eventually can cause cancer.[4,15] In fact, *in vitro* experiments by several groups found that UHRF1 binding to unmethylated DNA is negligible.[7,11,14] Binding of fully methylated CpG site is reported by one experiment to be sevenfold weaker relative to the binding of hemi-methylated DNA,[7] whereas another experiment did not detect any complex formation at all.[11] Other experiments also reported a weaker interaction between UHRF1 and fully methylated DNA; however, the relative binding affinities were not quantified.[12,13] As indicated above, the methylated cytosine base differs from unmodified cytosine in a single methyl (or methylene if considering the aromatic hydrogen H5) group. It is intriguing, however, how such relatively small chemical alteration of the DNA (even when marked at a single CpG site) can be so accurately recognized.

Although the structure of the bound complex between hemi-methylated DNA and UHRF1 has been resolved, the molecular mechanism responsible for distinguishing between hemi-methylated and unmethylated strands is not known. In one of the studies, this discrimination is proposed to arise from a hemisphere cavity with a radius of ~2 Å in the binding pocket of UHRF1 that exactly fits the methyl group at position 5 of methyl-cytosine.[12] Since the methyl group resides deep inside the binding pocket, a water molecule that may fill-up this cavity (if unmethylated cytosine is bound) is argued to be unfavorably hydrated. However, about half of the atoms surrounding this cavity, within van der Waals distance, are hydrophilic with potential of forming hydrogen bonds with a water molecule (see Table 2 in the supplementary data of a paper reporting the X-ray structure[12]). In addition, very small adjustments in the positions of few residues of the protein, on the order of 1 Å, can readily abolish the existence of such cavity. In another study, van der Waals interactions and weak CH···O hydrogen bonds of the methyl group with surrounding atoms were proposed to explain the selectivity for hemi-methylated over unmethylated strands.[11] However, similar interactions also exist in the unbound state of the DNA. In this case, water solvent molecules are in contact with the methyl group at position 5 (in both, the flipped-in and flipped-out conformations) and can form the same types of interactions as do the surrounding residues of the protein. It is also impossible to explain this selectivity by solvent-induced (hydrophobic) interactions because, on this length scale (of one methyl group), the magnitude of the effect, even if the entire hemisphere cavity is hydrophobic, is small.[16–18]

In this paper, we perform molecular dynamics simulations to elucidate the physical mechanism by which UHRF1 is able to distinguish hemi-methylated from unmethylated DNA. We find that the strong recognition of hemi-methylated DNA by UHRF1 is, predominantly, a consequence of the change in the distribution of the electrons around the cytosine ring due to the methylation reaction.

Results and Discussion

We calculated the binding free energy of UHRF1 to hemi-methylated DNA relative to the binding to unmethylated DNA, $\Delta\Delta G_b^{\text{HMe-UMe}} = \Delta G_b^{\text{HMe}} - \Delta G_b^{\text{UMe}}$. The thermodynamic cycles constructed to perform the calculations are shown in Fig. 1, and the change in free energy associated with each transformation is given in Table 1. In the direct alchemical transformations, $\Delta\Delta G_b^{\text{HMe-UMe}} = \Delta G_1 - \Delta G_2 = -49.6$ kJ/mol. This means that the binding of UHRF1 to hemi-methylated DNA is stronger than that to unmethylated DNA and that the magnitude of the difference in

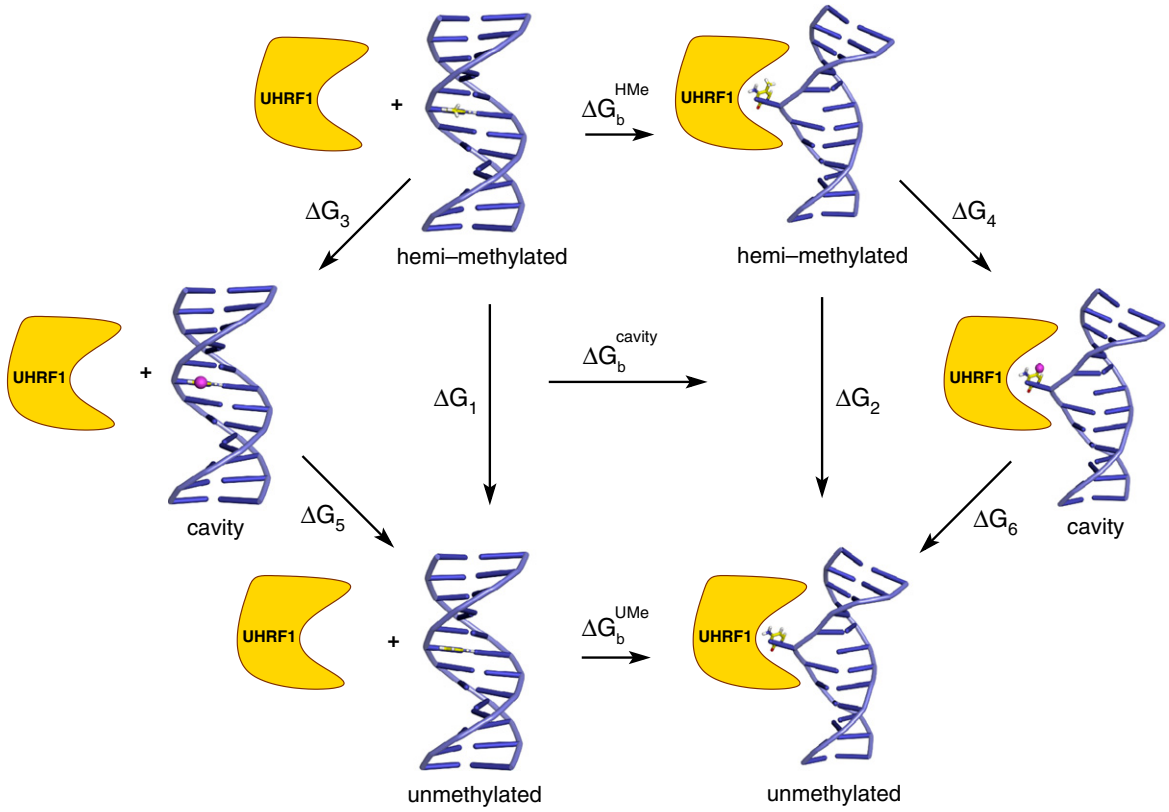


Fig. 1. Two thermodynamic cycles for calculating the difference between the binding free energy of UHRF1 to hemi-methylated DNA (ΔG_b^{HMe}) and to unmethylated DNA (ΔG_b^{UMe}). In the direct cycle, this difference, $\Delta\Delta G_b^{HMe-UMe}$, equals $\Delta G_1 - \Delta G_2$, whereas in the cycle that passes through the intermediate cavity state (see the text), it equals $\Delta G_3 - \Delta G_4 + \Delta G_5 - \Delta G_6$. The value of the free-energy change obtained in the simulations in each transformation is given in Table 1.

the binding strength (approximately, 8 orders of magnitude in the ratio of the binding constants) can readily explain the accuracy that UHRF1 discriminates between hemi-methylated and unmethylated DNA strands.

Table 1. The free-energy changes of the chemical transformations shown in Fig. 1

	Forward	Backward	Average
ΔG_1	−435.9	−436.4	−436.1
ΔG_2	−385.7	−387.2	−386.5
ΔG_3	−438.8	−439.5	−439.1
ΔG_4	−400.1	−395.9	−398.0
ΔG_5	−15.9	−15.9	−15.9
ΔG_6	−10.6	−11.2	−10.9
$\Delta\Delta G_b^{HMe-UMe} = \Delta G_1 - \Delta G_2$			−49.6
$\Delta\Delta G_b^{HMe-UMe} = \Delta G_3 - \Delta G_4 + \Delta G_5 - \Delta G_6$			−46.1

The differences in the binding free energy of UHRF1 to hemi-methylated DNA and to unmethylated DNA computed via the direct cycle and via the cycle that passes through an intermediate cavity state are also shown. All values are given in kilojoules per mole.

Given this magnitude of the difference in the binding free energies, it is reasonable to assume that one or more hydrogen bonds between the protein and the DNA are lost in the bound complex of unmethylated DNA. The majority of the hydrogen bonds between UHRF1 and the DNA are concentrated within the interaction with the CpG site. Because the difference between hemi-methylated DNA and unmethylated DNA is at position 5 of Cys6, it is likely that, if there is a loss of hydrogen bonds, it will be around this cytosine base. We calculated the average number of direct hydrogen bonds between UHRF1 and different groups of the DNA around the flipped-out methyl-cytosine base and summarized the results in Table 2†. No significant difference in the number of direct hydrogen bonds exists between the hemi-methylated and unmethylated complexes. In accordance with this, we hardly find any difference between the binding modes of the methyl-cytosine and cytosine to the SRA domain of UHRF1. Snapshots of these binding modes are displayed in Fig. 2. The numbers indicated in this figure refer to distances averaged over all trajectories. The pyrimidine rings of methyl-cytosine and cytosine establish

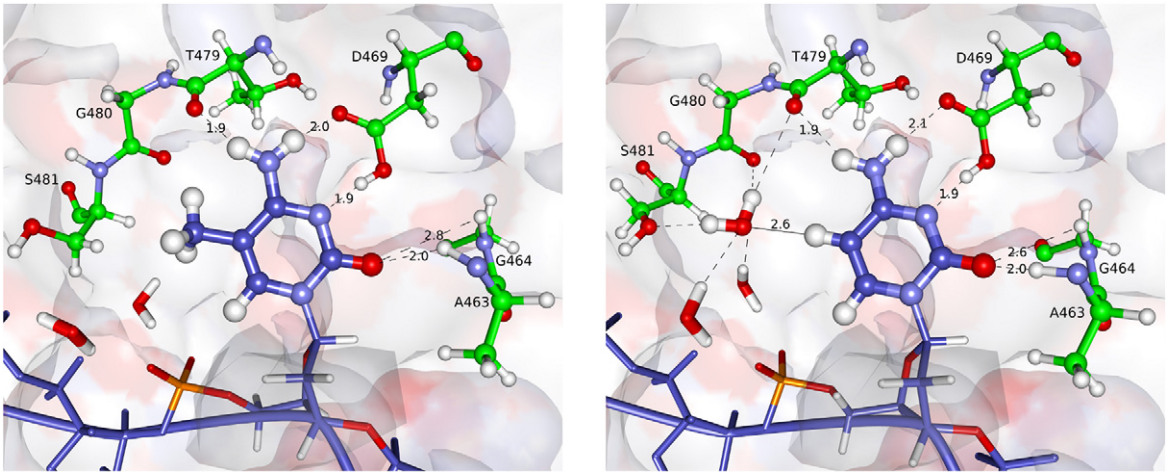
Table 2. The average number of direct and water-bridged hydrogen bonds between the SRA domain of UHRF1 and different groups of the DNA located around the recognized CpG site

	Direct hydrogen bonds		Water-bridged hydrogen bonds	
Interacting groups	Hemi-methylated	Unmethylated	Hemi-methylated	Unmethylated
Cyt6–UHRF1	4.79±0.08	4.86±0.03	<0.02	<0.02
Gua6′–UHRF1	1.92±0.01	1.88±0.03	0.26±0.06	0.03±0.01
Gua7–UHRF1	0.08±0.02	0.13±0.05	0.65±0.08	0.74±0.10
Cyt7′–UHRF1	0.86±0.07	0.91±0.03	<0.02	<0.02
Phosphate(Cyt6)–UHRF1	1.1±0.2	1.0±0.3	2.93±0.15	2.30±0.16
Phosphate(Gua6′)–UHRF1	<0.02	<0.02	0.05±0.02	0.06±0.04
Phosphate(Gua7)–UHRF1	0.88±0.04	0.87±0.05	1.05±0.08	1.16±0.09
Phosphate(Cyt7′)–UHRF1	0.29±0.08	0.38±0.08	0.10±0.03	0.15±0.04

A hydrogen bond is defined by a donor–acceptor cutoff distance of 0.35 nm and a donor–hydrogen–acceptor angle larger than 150°.

five hydrogen bonds with the binding site of UHRF1. In addition, the NKR finger of UHRF1 establishes two hydrogen bonds with the orphan guanine (via Arg491) and one hydrogen bond with the target cytosine on the opposite stand (via Asn489). All of these hydrogen bonds are stable in both complexes.

Despite the similarity in the binding modes of hemi-methylated and unmethylated complexes, one difference in the structure of both bound complexes is evident. It arises because, in the unmethylated complex, the space that was occupied by the methyl group is now filled-up by a water molecule. This is seen in the snapshots displayed in Fig. 2. In Table S1, we calculate the average number of waters within a radius of 6.0 Å from C(5-Me)/H5 that are simultaneously bound to (thus, within 4.0 Å from) the protein and DNA. The results confirm that, in the unmethylated case, there is one extra water molecule. This is also evident from the plot of the radial distribution function between C(5-Me)/H5 and the oxygen atoms of the water molecules (Fig. 3a), indicating that the probability to observe a water molecule at the space left by the methyl group is high. Thus, the possibility of discrimination driven by the existence of a "dry" state is excluded.[19] Contrary to previous suggestion,[12] this intruding water molecule is favorably hydrated, forming hydrogen bonds with two neighboring waters, as well as with three residues of UHRF1; therefore, it is not enclosed in a hydrophobic region.[20] Nevertheless, it does not bridge UHRF1 to the DNA (apart from the weak CH···O hydrogen bond with H5) as indicated by essentially the same number of bridging waters in the hemi-methylated and unmethylated complexes,



hemi–methylated unmethylated

Fig. 2. Snapshots of the interactions of UHRF1 with hemi-methylated DNA (left panel) and with unmethylated DNA (right panel) including all water molecules found within a radius of 6 Å from C(5-Me)/H5. The flipped methyl-cytosine, as well as the (intruding) water molecule that occupies the space of the methyl group in the unmethylated complex, is emphasized by thick ball-and-stick representation. The protein–DNA hydrogen bonds are indicated by broken lines, and the numbers shown correspond to the H–acceptor hydrogen bond distances (Å) averaged over all trajectories. Note that the intruding water molecule can form hydrogen bonds with two other water molecules, Ser481, Gly480, and Thr479. The distance corresponding to the first peak of the radial distribution function between H5 and oxygen of waters is also indicated and represented by a continuous line.

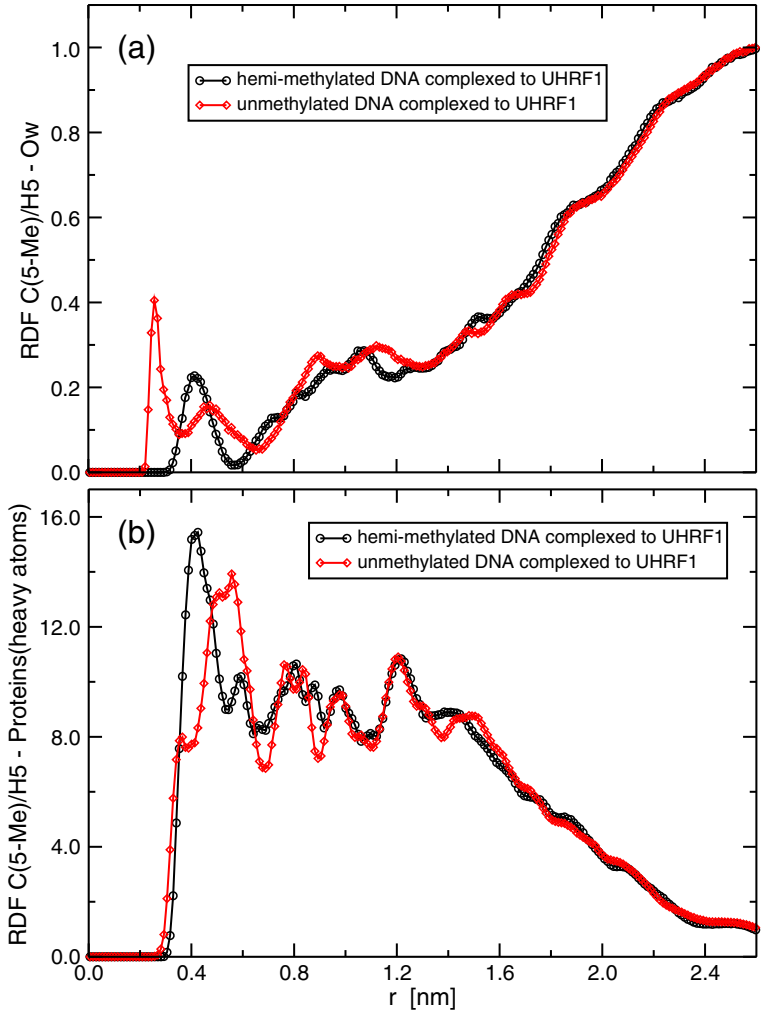


Fig. 3. The radial distribution function between C(5-Me)/H5 of the recognized cytosine and the oxygen atom of the water molecules (a), as well as between C(5-Me)/H5 and all the protein's heavy atoms (b) for UHRF1 bound to hemi-methylated and unmethylated DNA.

calculated for the region considered in Table S1. On the contrary, the presence of this water molecule appears to establish hydrogen bonds with nearby waters on the expense of the hydrogen bonds these nearby waters make with the protein. In addition, it induces larger fluctuations in the positions of these water molecules as indicated by a larger value of the local Debye–Waller factor (see Table S1). In Table 2, we present also the average number of water-bridged hydrogen bonds between different groups of the DNA and UHRF1. Significant difference is observed only in the case of the orphan guanine (Gua6′) and the phosphate group of the recognized cytosine (Cyt6). In both cases, a larger number is found in the hemi-methylated complex. Nevertheless, the combined difference is less than one hydrogen bond that corresponds to an energy with a magnitude that cannot explain the difference in the free energy of binding obtained in Table 1.

What is then the driving factor that permits the distinction between hemi-methylated and unmethylated DNA strands? To answer this question, we calculated $\Delta\Delta G_b^{\text{HMe-UMe}}$ by a two-step transformation via an intermediate (cavity) state (see Fig. 1). In this intermediate state, the partial charges of the flipped-out cytosine are the same as those of unmethylated cytosine; however, instead of H5, we constructed an atom (cavity) that has the same excluded volume as a methyl group (the charge and Lennard-Jones dispersion interactions of the cavity atom were kept as those of H5; see Fig. S2). This cavity atom prevents the entry of the intruding water molecule. Via this intermediate-state route, $\Delta\Delta G_b^{\text{HMe-UMe}} = \Delta G_3 - \Delta G_4 + \Delta G_5 - \Delta G_6 = -46.1$ kJ/mol, which is very similar to the value found via the direct transformation. The analysis of the two subcycles indicates that the large magnitude of $\Delta\Delta G_b^{\text{HMe-UMe}}$ arises predominantly from the change of the partial

charges of the pyrimidine ring due to the methylation at C5, $\Delta G_b^{HMe} - \Delta G_b^{cavity} = -41.1$ kJ/mol, whereas the removal of the excluded volume of the methyl group is rather minor, $\Delta G_b^{cavity} - \Delta G_b^{UMe} = -5.0$ kJ/mol.

When the DNA double strand is free in solution, the methyl-cytosine is inside the helix and paired to the complementary guanine. Nevertheless, it also interacts (especially the C5 and 5-methyl groups) with substantial amount of water molecules present in the first neighbor shell and beyond. When the DNA is bound to UHRF1, the methyl-cytosine is flipped-out of the helix and surrounded by the protein atoms. A major difference in the properties of bulk waters compared with those of residues inside a protein (especially at the active site) that is relevant for interactions with electrostatic charges is that the dielectric constant of the latter is much smaller relative to the former. In other words, bulk waters can easily adjust themselves to optimize the solvation of charges, whereas protein residues at the active site cannot. We calculated the potential energy between the cytosine base in unmethylated DNA free in solution and the solvent and compared it with that of the methyl-cytosine base in hemi-methylated DNA. The former interaction energy is more favorable by 33 kJ/mol compared with the latter. This implies that unmethylated DNA is more stable than hemi-methylated DNA in the unbound state. To demonstrate the effect of the different partial charges on the stability of the unbound states in water more rigorously, we constructed a simplified model to calculate the transfer free energy of methyl-cytosine from water to hexane (a solvent with a low dielectric constant) and compared this value with that of cytosine. The thermodynamic cycles for these calculations are shown in Fig. 4, and the results of the individual transformations are shown in Table 3. The value of $\Delta G_t^{mCyt} - \Delta G_t^{Cyt}$ is about −20 kJ/mol (where the corresponding potential energy difference, $\Delta E_t^{mCyt} - \Delta E_t^{Cyt}$, is −29.4 kJ), which is much larger than what is expected from hydrophobic interactions of one methylene group; the transfer free energy of a methylene group in alkane from water to a hydrophobic medium is found experimentally[21,22] and computationally[23] to be, approximately, 3.2 kJ/mol or lower. Analogous to the cycles shown in Fig. 1, also here, we calculated the change in the transfer free energy directly and via the intermediate cavity state. In fact, $\Delta G_t^{cavity} - \Delta G_t^{Cyt} = -3.7$ kJ/mol, which is very similar to the free-energy change of transferring a methylene group from water to a hydrophobic medium. Thus, most of the contribution to the difference in the transfer free energy arises from $\Delta G_t^{mCyt} - \Delta G_t^{cavity} = -16.5$ kJ/mol, which is a consequence of the change in the electrons distribution of the pyrimidine ring. Because the atoms of the hexane molecules in the model system we used are neutral, the partial charges of the methyl-cytosine base interact only with the water phase. This indicates that the transfer of methyl-cytosine from water to hexane is more favorable than that of cytosine because methyl-cytosine is less stable than cytosine in water.

What causes cytosine to be more stable than methyl-cytosine in water? We calculated the dipole moment of both molecules and found cytosine to be larger by 1.2 Debye, essentially of which the entire magnitude is projected along the axis coinciding with the C5-C(5-Me) bond (Table S2 and Fig. S3). We repeated the calculation by using charges obtained from quantum mechanical optimizations at the level of HF, MP2, and B3LYP (see Table S3). In all cases, the dipole moment of cytosine was bigger than methyl-cytosine by more than 1.2 Debye and again the magnitude of the difference was aligned almost completely along the axis of the bond between the 5-methyl and C5 (Table S2). In the supplementary data (Table S4 and Fig. S4), we describe simulations aimed at evaluating the free-energy change of creating two dipole moments in water that correspond to that of cytosine and methyl-cytosine (represented by the force field used in this work). The difference in the free-energy change, $\Delta G_{dipole}^{Cyt} - \Delta G_{dipole}^{mCyt}$, equals −32.2 kJ/mol. This means that a difference of 1.2 Debye in the dipole moment can explain the −20 kJ/mol observed for the transfer free energy from water to hexane.

In addition to the larger stability of unmethylated DNA compared with hemi-methylated DNA in the unbound state in water, the bound complex of hemi-methylated DNA with UHRF1 can be more stable than that with unmethylated DNA. Stronger interactions of methyl-cytosine with UHRF1 can be a result of the protein structure around the binding site that takes advantage of the base-altered partial charges. We calculated the interaction energy of these two bases with the protein and the solvent and found that, for hemi-methylated, the interaction was more favorable by −17.0 kJ/mol. The dominant contribution of this energy change can be attributed directly to the change in the partial charges. When we repeated the calculations for the unmethylated base, but with trajectories taken from the simulations of the hemi-methylated complex, the difference was −10.0 kJ/mol. The remaining contribution arises from a change in the protein conformation. The entry of a water molecule to the space left by the methyl group pushes some of the protein atoms away from the methyl-cytosine residue. This is displayed in Fig. 3b, where the radial distribution functions between the protein atoms and the C(5-Me)/H5 atom are plotted. The major difference is in the location of the first peak that shifts from 0.42 nm in the hemi-methylated complex to 0.54 nm in the unmethylated complex. These displacements occur probably to enlarge the space for the intruding water molecule and allow it to form its hydrogen bonds network. Most of these conformational changes occur solely due to the

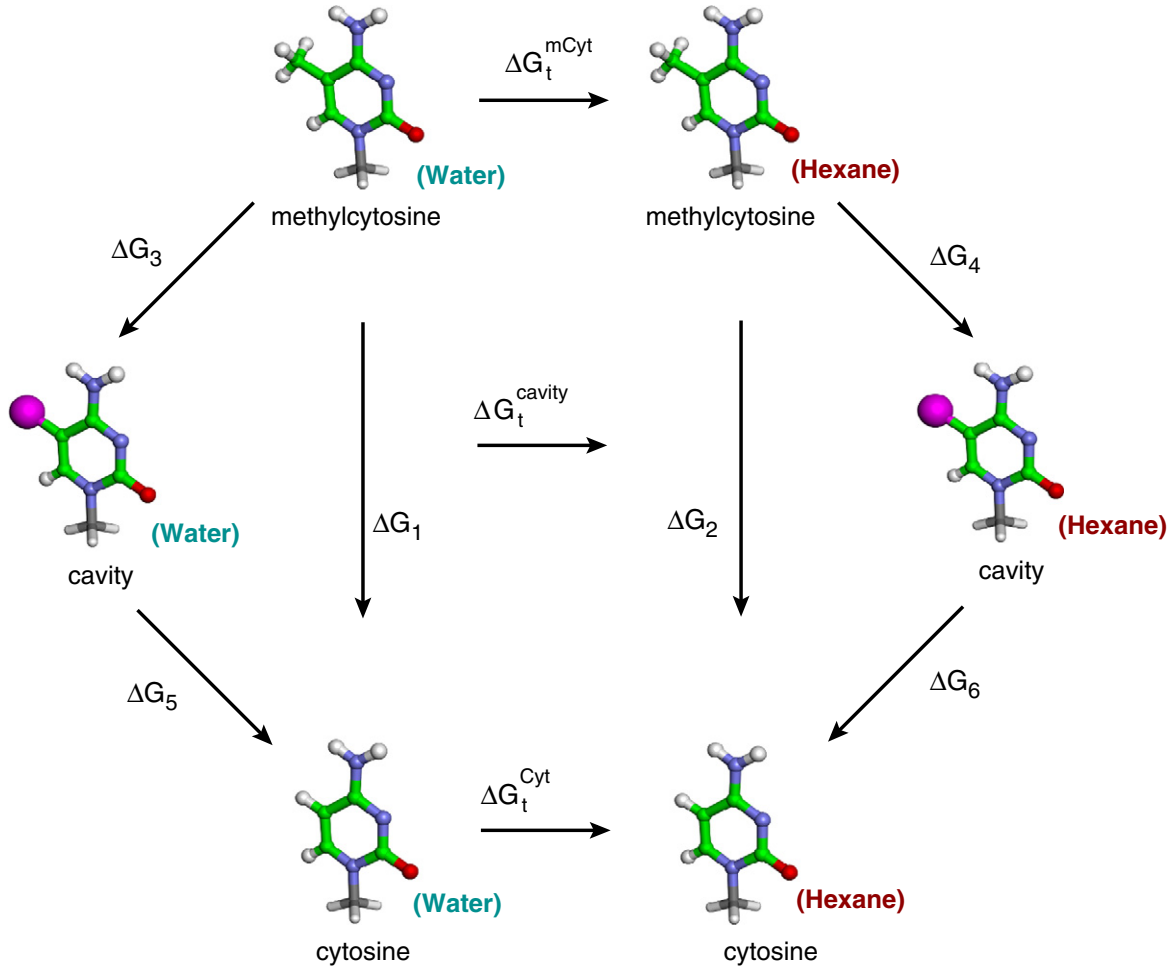


Fig. 4. The transfer free-energy change, ΔG_t, from water to hexane of cytosine, methyl-cytosine, and cytosine with the cavity atom bonded to C5. The two thermodynamic cycles shown are analogous to those constructed in Fig. 1, and the simulation result for the value of the free-energy change associated with each transformation is given in Table 3.

Table 3. The free-energy changes of the chemical transformations shown in Fig. 4

	Forward	Backward	Average
ΔG_1	−455.9	−455.9	−455.9
ΔG_2	−435.8	−435.7	−435.8
ΔG_3	−441.8	−441.5	−441.6
ΔG_4	−425.1	−425.0	−425.1
ΔG_5	−16.0	−16.0	−16.0
ΔG_6	−12.3	−12.3	−12.3
$\Delta\Delta G_t^{mCyt-Cyt}=\Delta G_1-\Delta G_2$			−20.1
$\Delta\Delta G_t^{mCyt-Cyt}=\Delta G_3-\Delta G_4+\Delta G_5-\Delta G_6$			−20.2
$\Delta\Delta E_t^{mCyt-Cyt}=\Delta E_1-\Delta E_2$			−29.4

The differences in the free-energy change of transferring cytosine and methyl-cytosine from water to hexane via the direct cycle and via the intermediate cavity state are also shown. The difference in the potential energy change of the transfer process is also indicated. All values are given in kilojoules per mole.

removal of the excluded volume of the methylene group and the entry of the intruding water molecule; however, some changes are also contingent upon the change in the partial charges of the bases. This, for example, can be seen in Fig. S5, where the distribution of the distance between the methyl-cytosine and Ser481 in the cavity intermediate complex resembles the distribution of the unmethylated more than that of the hemi-methylated complex.

Another difference between the hemi-methylated and unmethylated cases occurs in the γ-dihedral angles of the DNA backbone around (m)Cyt6. Although these dihedrals are in the *gauche* (*g*+) conformation (as in an ideal B-DNA structure) for both cases when the strands are free in solution, the C6-γ dihedral of the unmethylated strand bound to UHRF1 populates also the *trans* conformation. The *trans* population in the hemi-methylated bound

complex is very small (see Fig. S6). Note that the flipping of a base out of the double helix or the binding of a DNA to a protein regularly involves the occurrence of *trans* conformations of the DNA backbone.[24] The larger population of the *trans* conformation in the unmethylated DNA bound complex can also add an energetic penalty that will favor the binding to hemi-methylated DNA.

The results presented in this paper indicate that accurate distinction between cytosine and methyl-cytosine, which differ seemingly in only one methylene group, is possible not due to the presence of this methylene group *per se* but is a consequence of the change in the charge distribution of the aromatic ring. This change in the electron structure is a result of the methylation reaction, and it induces two effects influencing the binding free energy with UHRF1. The first is a reduction of the magnitude of the dipole moment of the base that destabilizes hemi-methylated DNA in its unbound state in water. The second is strengthening the interaction energy in the bound complex between UHRF1 and hemi-methylated DNA. The combination of these two effects produces a difference in the binding constants with magnitude that permits recognition with high fidelity not expected from such a small change.

Experimental Procedures

The initial structure for the simulations was the crystallographic structure of the SRA domain of UHRF1 (204 amino acids long) complexed with a 12-base-pair double-stranded hemi-methylated DNA (Protein Data Bank accession code: 3CLZ).[12] The hemi-methylated CpG site is located halfway along the DNA strand at position mC6pG7, and the corresponding bases on the complementary strand are C7′pG6′. From the different structures available, we chose the model with the lowest number of missing atoms. The missing atoms, eight in number belonging to three lysine residues, were then built by the software PyMOL version 1.2r1. The side chains of arginine and lysine were protonated, whereas those of glutamate and aspartate were deprotonated. Histidine was simulated in its neutral form in which the δ-position was protonated because this tautomer exhibits a slightly larger pK_a value.[25] An exception for these assignments was the case of Asp469 located at the binding pocket of UHRF1 that contributes to the stability of the flipped methyl-cytosine. In this case, the aspartate was simulated in its protonated form because the distance in the X-ray structure between one of its carboxylate oxygens and N3 of methyl-cytosine strongly suggests the presence of a proton either on the oxygen or on the nitrogen. In fact, this aspartic (or alternatively glutamic) acid is a conserved residue that is also conserved in DNA cytosine-5 methyltransferase from different organisms.[26] Quantum mechanical calculations indicate that the proton is more likely to reside on the carboxylate group,[6] and therefore, we considered Asp469 to be protonated. The N and the C termini of the protein were protonated and deprotonated, respectively. Given these protonation states for the amino acid residues, the total charge of the protein is $+7e$. In addition, a 12-base-pair double-stranded DNA contributes a charge of $-22e$ due to the phosphate groups. These charges were neutralized by 7 chlorides and 22 sodium cations added at random positions in the simulation box. All oxygen atoms of waters given in the X-ray structure (124 in total) were built into water molecules. The dimensions of the cubic simulation box were determined by a minimum distance of 0.8 nm between the DNA–protein complex and each of the box edges. The system was then solvated by additional waters and contained a total of 10,510 water molecules. Analogous construction procedure was applied for the system where the DNA is free in solution neutralized by 22 sodium atoms and solvated by 8356 water molecules.

The DNA and protein were represented by the AMBER03 force field[27,28] (note that the nucleic acid parameters are the same as those of AMBER99 force field) and the water molecules by the TIP3P model.[29] The chloride and sodium ions were also represented by the AMBER03 force field, that is, $\sigma_{Cl^-} = 0.440$ nm, $\varepsilon_{Cl^-} = 0.418$ kJ/mol, $\sigma_{Na^+} = 0.333$ nm, and $\varepsilon_{Na^+} = 0.0116$ kJ/mol. The partial charges of 5-methylcytosine, which are not available in the standard parameters of AMBER03, were taken from the work of Rauch *et al.*[30] These charges were obtained from an ab initio calculation using the RESP (*R*estrained *E*lectro*S*tatic *P*otential) charge fitting procedure.[31]

The molecular dynamics package GROMACS version 4.0.7[32] was used to perform all of the computer 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[33] (with real-space cutoff of 1.0 nm, grid spacing of 0.12 nm, and quadratic interpolation) and the Lennard-Jones forces by a cutoff of 1.0 nm. The system was maintained at a constant temperature of 300 K (with a coupling time of 0.1) by the velocity rescaling thermostat[34] and at a pressure of 1.0 bar (with a compressibility of 5×10^{-5} bar and a coupling time of 1.0 ps) by the Berendsen thermostat.[35] Water bond distances and angles were constrained using the SETTLE algorithm,[36] whereas the protein and DNA covalent bond distances were constrained using the LINCS algorithm.[37]

The system was first energy minimized using the steepest descent approach, followed by a 2-ns simulation in which the positions of the DNA and protein heavy atoms were restrained by a harmonic potential with a force constant of 1000 kJ/(mol nm^2). Then, 10 ns of unrestrained simulation was performed. The configuration emerged from this simulation was used as an input for the free-energy calculations.

The binding free energy of UHRF1 to hemi-methylated DNA relative to the binding to unmethylated DNA was computed by the concept of a thermodynamic cycle. To this end, we performed alchemical mutations of atom types (with soft-core potentials $\alpha=0.7$ and $p=1$), bonds, angles, and dihedrals to transform the hemi-methylated to unmethylated DNA and vice versa. These transformations were performed for the DNA–protein complex and for the DNA free in solution. The free-energy changes associated with these transformations were computed by the thermodynamic integration technique. For each transformation, 11 equally spaced λ-points from $\lambda=0$ to $\lambda=1$ were constructed. At each λ-point, the value of $\partial H/\partial\lambda$ was averaged over 25 ns and 35 ns, for the transformation of the DNA free in solution

and bound to UHRF1, respectively, preceded by 5 ns equilibration time. At locations where the plot of $\partial H/\partial\lambda$ as a function of λ did not exhibit a smooth behavior, we added up to 7 extra λ-points and the simulation time of some points were extended to 60 ns (see Fig. S7 for the curves of $\langle \partial H/\partial\lambda \rangle$ as a function of λ). In the calculations involving the cavity intermediate state, we constructed 13 λ-points and at each λ averaged over 35 ns and 45 ns for the transformation of the DNA free in solution and bound to the protein, respectively, after discarding the first 5 ns for equilibration. For all transformations, we calculated the free-energy change associated with the forward and backward directions. In general, the input conformation at a particular λ-point was taken after equilibration of approximately 1 ns at the preceding λ-point. We checked that, throughout the trajectories, the protein–DNA bound complexes stayed intact (see Fig. S8).

In order to obtain better statistics for the unmethylated and hemi-methylated states, we conducted six additional simulations for 60 ns for each of these states in which we saved the trajectory of the entire system every 40 ps. Thus, the thermodynamic and structural properties reported for these states are averaged over 360 ns when the information of the solvent is needed, otherwise, for 540 ns.

In the calculations of the transfer free energy of cytosine and methyl-cytosine from water to hexane, we capped both bases by a united atom methyl group at N1. Hexane and the capped methyl group at N1 were modeled by the Gromos96 (43A2) force field. The charge on the N1 methyl group was assigned to $+0.0631e$ to neutralize the charge of the entire base. Here, the interconversion between cytosine and methyl-cytosine in the forward and backward directions was performed by 13 λ-points. At each point, the system was equilibrated for 2 ns and data were collected for 8 ns. The mutations in water included 1520, whereas in hexane 521, solvent molecules. In the simulations with hexane, the value of the isothermal compressibility was set to 16.7×10^{-5} bar and the coupling time was set to 2.0 ps.

Because we calculated the electrostatic interactions using Ewald summation, the decomposition of the potential energy into contributions from different groups was performed by the direct evaluation of the interparticle distances from the trajectories (using the *-rerun* option in Gromacs).

Acknowledgements

We would like to thank Prof. Fernando P. Cossío and Prof. Ana Arrieta for simulating discussions. This work has been funded with support from the European Commission Marie Curie International Reintegration Grant (project number 247485) and from the Spanish Ministry of Science and Innovation (grant number CTQ2010-20297). We thankfully acknowledge the computer resources and technical assistance provided by the Barcelona Supercomputing Center Centro Nacional de Supercomputación. Technical and human support provided by SGIker (used services) (Universidad del País Vasco/Euskal Herriko Unibertsitatea, Ministry of Science and Innovation, Gobierno Vasco/Eusko Jaurlaritza, and European Science Foundation) is also gratefully acknowledged.

Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2012.09.024

Received 6 June 2012;
Received in revised form 14 September 2012;
Accepted 28 September 2012
Available online 4 October 2012

Keywords:
protein–DNA interactions;
epigenetics;
DNA methylation;
molecular dynamics simulations;
free-energy calculations

† The number of direct and water-bridged hydrogen bonds between the sugar groups of the CpG site on both strands and the protein is zero in most of the cases. The only exceptions are the sugar groups of Cys6 and Gua7 that make 0.35 and 0.60 water-bridged hydrogen bonds with UHRF1, respectively, with no significant difference between hemi-methylated and unmethylated complexes.

References

1. Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBioChem*, **3**, 274–293.
2. Szyf, M. (2003). Targeting DNA methylation in cancer. *Ageing Res. Rev.* **2**, 299–328.
3. Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21.
4. Goll, M. G. & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514.
5. Damelin, M. & Bestor, T. H. (2007). Biological functions of DNA methyltransferase 1 require its methyltransferase activity. *Mol. Cell. Biol.* **27**, 3891–3899.
6. Zangi, R., Arrieta, A. & Cossío, F. P. (2010). Mechanism of DNA methylation: the double role of DNA as a substrate and as a cofactor. *J. Mol. Biol.* **400**, 632–644.
7. Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S. & Jacobsen, S. E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, **317**, 1760–1764.
8. Sharif, J., Muto, M., Takebayashi, S. Ichiro, Suetake, I., Iwamatsu, A., Endo, T. A. *et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**, 908–913.
9. Hashimoto, H., Horton, J. R., Zhang, X. & Cheng, X. (2009). UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications. *Epigenetics*, **4**, 8–14.

10. Unoki, M., Brunet, J. & Mousli, M. (2009). Drug discovery targeting epigenetic codes: the great potential of UHRF1, which links DNA methylation and histone modifications, as a drug target in cancers and toxoplasmosis. *Biochem. Pharmacol.* **78**, 1279–1288.
11. Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. & Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, **455**, 818–821.
12. Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C. *et al.* (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, **455**, 822–825.
13. Hashimoto, H., Horton, J. R., Zhang, X., Bostick, M., Jacobsen, S. E. & Cheng, X. (2008). The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature*, **455**, 826–829.
14. Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M. J. & Zhou, M.-M. (2008). Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J. Biol. Chem.* **283**, 34490–34494.
15. Hervouet, E., Lalier, L., Debien, E., Cheray, M., Geairon, A., Rogniaux, H. *et al.* (2010). Disruption of Dnmt1/PCNA/Uhrf1 interactions promotes tumorigenesis from human and mice glial cells. *PLoS One*, **5**, e11333.
16. Southall, N. T., Dill, K. A. & Haymet, A. D. J. (2002). A view of the hydrophobic effect. *J. Phys. Chem. B*, **106**, 521–533.
17. Chandler, D. (2005). Interfaces and the driving force of hydrophobic assembly. *Nature*, **437**, 640–647.
18. Zangi, R. (2011). Driving force for hydrophobic interaction at different length scales. *J. Phys. Chem. B*, **115**, 2303–2311.
19. Wallqvist, A. & Berne, B. J. (1995). Computer simulation of hydrophobic hydration forces on stacked plates at short range. *J. Phys. Chem.* **99**, 2893–2899.
20. Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. (2007). Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl Acad. Sci. USA*, **104**, 808–813.
21. Wishnia, A. (1963). The hydrophobic contribution to micelle formation: the solubility of ethane, propane, butane, and pentane in sodium dodecyl sulfate solution. *J. Phys. Chem.* **67**, 2079–2082.
22. Blokzijl, W. & Engberts, J. B. F. N. (1993). Hydrophobic effects. Opinions and facts. *Angew. Chem., Int. Ed. Engl.* **32**, 1545–1579.
23. Fujimoto, K., Yoshii, N. & Okazaki, S. (2010). Molecular dynamics study of solubilization of immiscible solutes by a micelle: free energy of transfer of alkanes from water to the micelle core by thermodynamic integration method. *J. Chem. Phys.* **133**, 074511.
24. Várnai, P., Djuranovic, D., Lavery, R. & Hartmann, B. (2002). α/γ Transitions in the B-DNA backbone. *Nucleic Acids Res.* **30**, 5398–5406.
25. Tanokura, M. (1983). ^{1}H-NMR study on the tautomerism of the imidazole ring of histidine residues. 1. Microscopic p*K* values and molar ratios of tautomers in histidine-containing peptides. *Biochim. Biophys. Acta*, **742**, 576–585.
26. Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J. & Wilson, G. G. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **22**, 1–10.
27. Wang, J., Cieplak, P. & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074.
28. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W. *et al.* (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012.
29. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
30. Rauch, C., Trieb, M., Wellenzohn, B., Loferer, M., Voegele, A., Wibowo, F. R. & Liedl, K. R. (2003). C5-methylation of cytosine in B-DNA thermodynamically and kinetically stabilizes BI. *J. Am. Chem. Soc.* **125**, 14990–14991.
31. Cieplak, P., Cornell, W. D., Bayly, C. & Kollman, P. A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **16**, 1357–1377.
32. 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. Theor. Comput.* **4**, 435–447.
33. 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**, 10089–10092.
34. Bussi, G., Donadio, D. & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101.
35. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
36. Miyamoto, S. & Kollman, P. A. (1992). SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.* **13**, 952–962.
37. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472.

Supplementary Data: How to Distinguish methyl-Cytosine from Cytosine with High Fidelity

Caterina Bianchi[a], Ronen Zangi[a,b,*]

[a]*Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, San Sebastian, Spain*
[b]*IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Spain*

*Corresponding author: r.zangi@ikerbasque.org Tel: +34-943018112

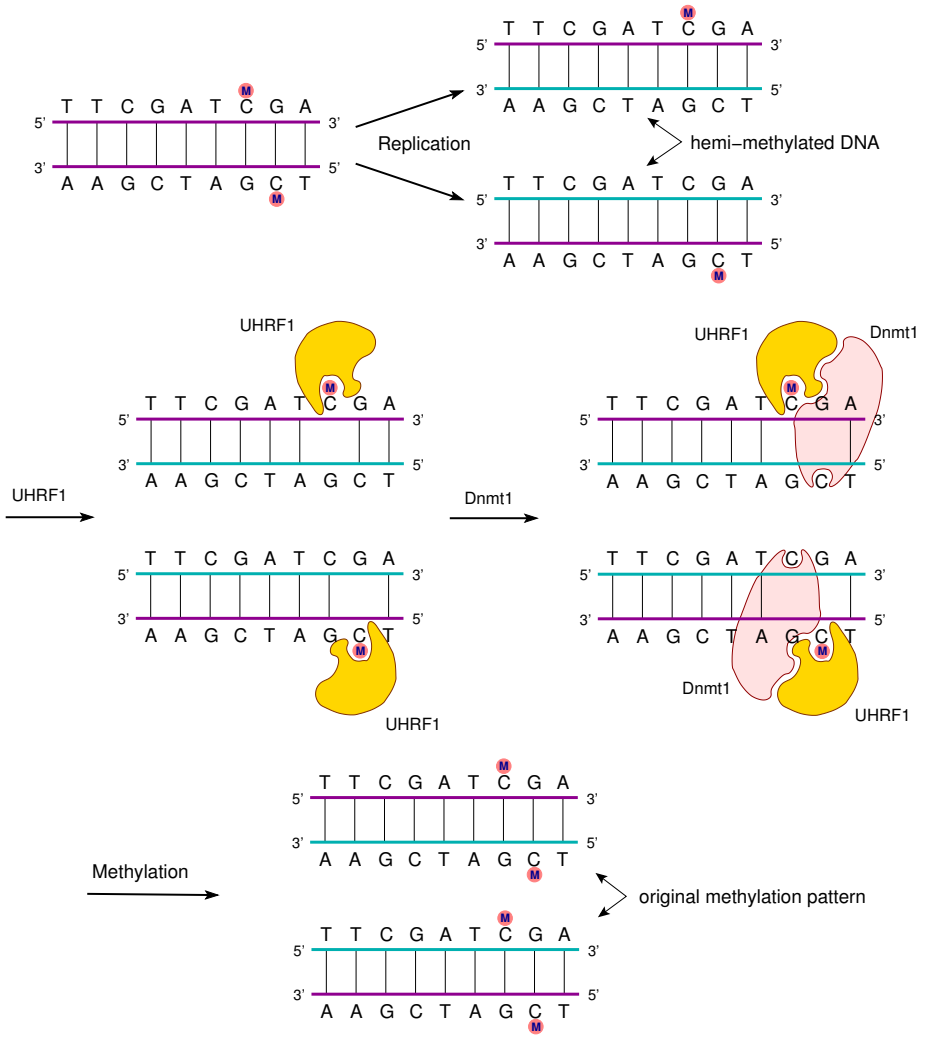


Figure S1: Schematic illustration of the role of the regulatory protein UHRF1 in maintaining the DNA methylation patterns after replication. Immediately after replication the emerging DNA strands are hemi-methylated. UHRF1 selectively binds to hemi-methylated CpG sites and concurrently can bind to Dnmt1, the enzyme catalyzing the methylation of cytosines. The binding to UHRF1 enables Dnmt1 to methylate the correct target cytosine on the opposite strand. In this way, the original methylation pattern is maintained. Thus, it is crucial that UHRF1 would not binds mistakenly to unmethylated CpG sites otherwise the methylation pattern will be altered. Note, that although in this scheme UHRF1 and Dnmt1 are shown to be bound to the DNA simultaneously, it is not clear whether this is the case. It is also possible that unbinding of UHRF1 precedes or occurs concurrently to binding of Dnmt1.

Table S1: The number of water molecules at the DNA-UHRF1 interface within a radius of 6.0 Å from C(5-Me)/H5, their Debye-Waller factor, and the number of water molecules that are simultaneously hydrogen-bonded to (thus, bridging) the DNA and UHRF1. The local Debye-Waller factor (in nm^2) for (the oxygen atom of) water i at short time was calculated by[1], $DW_i = \left\langle \left\langle (\vec{r}_i(t) - \langle \vec{r}_i \rangle_\tau)^2 \right\rangle_\tau \right\rangle$ where $\langle \ldots \rangle_\tau$ denotes average over the time period τ = 1 ns and the outer brackets, $\langle \ldots \rangle$, denotes average over the simulation trajectory (i.e., average over different time origins). Larger value of the Debye-Waller factor obtained in the unmethylated complex indicates larger magnitude of fluctuations of the water molecules around their average positions.

	Hemi-methylated	Unmethylated
# of Interfacial Waters	2.1 ± 0.4	3.2 ± 0.3
# of Bridging Waters	1.3 ± 0.1	1.5 ± 0.1
Debye-Waller factor	0.08 ± 0.02	0.14 ± 0.02

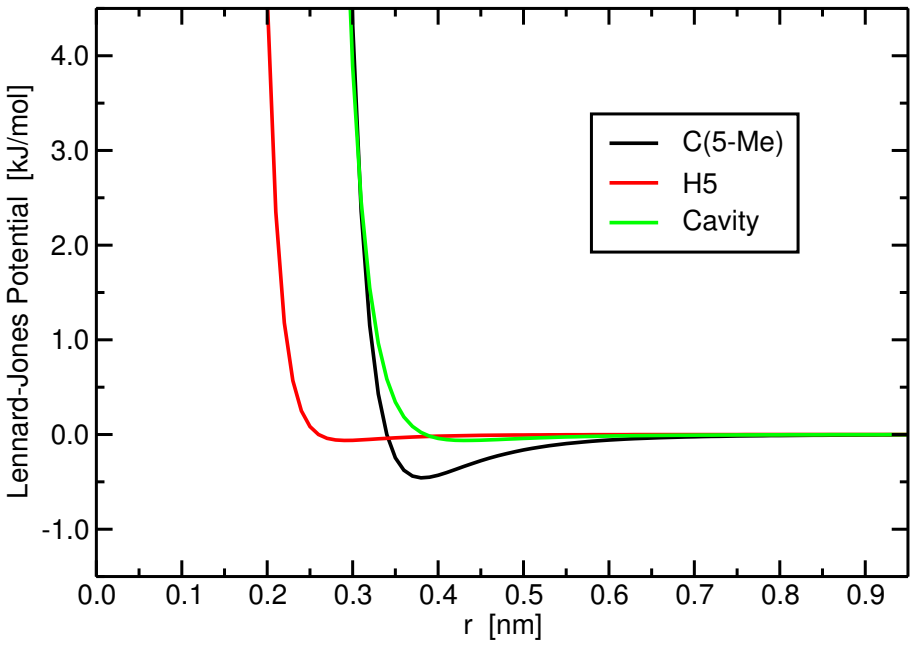


Figure S2: The Lennard-Jones potential of the hydrogen atom at position 5 in unmethylated cytosine, H5, the carbon atom of the methyl group bonded to C5 in methyl-cytosine, C(5-Me), and of the cavity atom in the cavity-intermediate state. The Lennard-Jones parameters of the cavity atom are σ=0.385 nm and ϵ=6.276 kJ/mol. The value of this ϵ is the same as that of H5 but the value of σ was chosen in such a way that the Lennard-Jones potential of this cavity atom has the same excluded volume as the 5-methyl group.

Table S2: The absolute value of the dipole moment, and its projection, μ_x (see Fig. S3), onto the axis defined by the bond between C5 and H5/C(5-Me), of cytosine and methy-cytosine. The calculations are based on the partial charges obtained by different quantum-mechanical methods (Table S3) applied on the same classical geometry as defined by the AMBER force-field. The partial charges of the hydrogens of a methyl group were incorporated into the carbon atom. All values are given in units of Debye.

	cytosine		methyl-cytosine		cytosine – methy-cytosine	
	$\|\mu\|$	μ_x	$\|\mu\|$	μ_x	$\Delta\|\mu\|$	$\Delta\mu_x$
RESP	8.25	7.95	7.03	6.74	1.22	1.21
MP2	7.61	7.47	4.88	4.82	2.73	2.65
HF	8.72	8.46	6.16	6.03	2.56	2.43
B3LYP	7.98	7.86	4.70	4.69	3.28	3.17

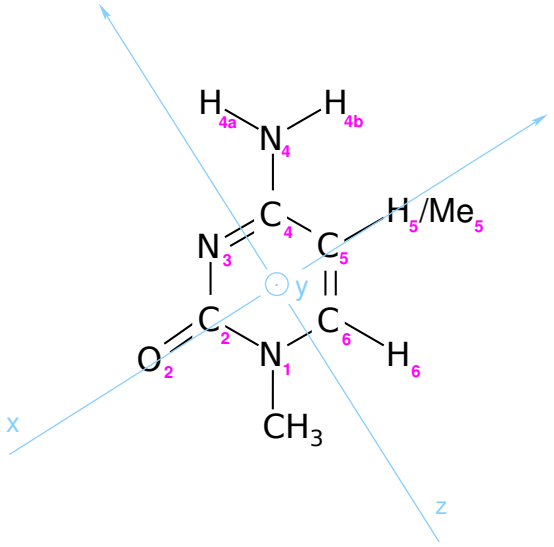


Figure S3: The definition of the different atoms in cytosine and methyl-cytosine base and the orientation of the molecule in the coordinate system used for in Tables S2 and S3.

Table S3: The partial charges of cytosine and methyl-cytosine obtained by quantum mechanical optimizations at the MP2, HF, and B3LYP levels with the 6-31++G** basis set using the Gaussian09 program[2]. The values used in this study, taken from Rauch et al.[3], were obtained by RESP charge fitting procedure in compliance with the Amber force-field derived parameters. The calculations were performed in vacuum. The starting configurations were created using the Molden processing program[4]. For the analysis, the density matrix of the current method is used. The charges of the hydrogen atoms of both methyl groups (Me5 and the capping (N1)-Me) are absorbed onto the corresponding central carbon atom. For the definition of the atom names see Fig. S3.

	cytosine				methyl-cytosine			
Atom/Group	MP2	HF	B3LYP	RESP	MP2	HF	B3LYP	RESP
N1	-0.237	-0.386	-0.154	-0.034	-0.224	-0.370	-0.138	+0.014
(N1-)CH_3	+0.275	+0.319	+0.243	+0.063	+0.260	+0.302	+0.232	+0.063
C2	+0.486	+0.702	+0.505	+0.796	+0.500	+0.742	+0.514	+0.775
O2	-0.507	-0.621	-0.513	-0.654	-0.511	-0.626	-0.511	-0.623
N3	-0.435	-0.615	-0.483	-0.775	-0.428	-0.601	-0.486	-0.717
C4	+0.103	+0.317	+0.036	+0.844	+0.112	+0.252	+0.074	+0.609
N4	-0.466	-0.564	-0.451	-0.977	-0.453	-0.567	-0.440	-0.876
H4a	+0.326	+0.355	+0.325	+0.431	+0.342	+0.374	+0.342	+0.400
H4b	+0.313	+0.345	+0.329	+0.431	+0.307	+0.338	+0.318	+0.400
C5	-0.048	-0.174	+0.080	-0.522	+0.309	+0.378	+0.591	-0.062
H5/Me5	+0.162	+0.168	+0.140	+0.186	-0.199	-0.206	-0.311	-0.002
C6	-0.102	+0.002	-0.178	-0.018	-0.158	-0.186	-0.329	-0.192
H6	+0.130	+0.152	+0.121	+0.229	+0.143	+0.170	+0.144	+0.211

Table S4: The free energy change, in kJ/mol, for creating two dipole-moments in water. The dipole moments are modeled by the system described in Fig. S4. Their magnitudes of 8.25 and 7.03 Debye correspond to the magnitude of the dipole-moments of cytosine and methy-cytosine, respectively.

	$\|\mu\| = 0 \rightarrow 8.25$ D	$\|\mu\| = 0 \rightarrow 7.03$ D	$\|\mu\| = 7.03 \rightarrow 8.25$ D
ΔG	-114.3	-82.1	-32.2

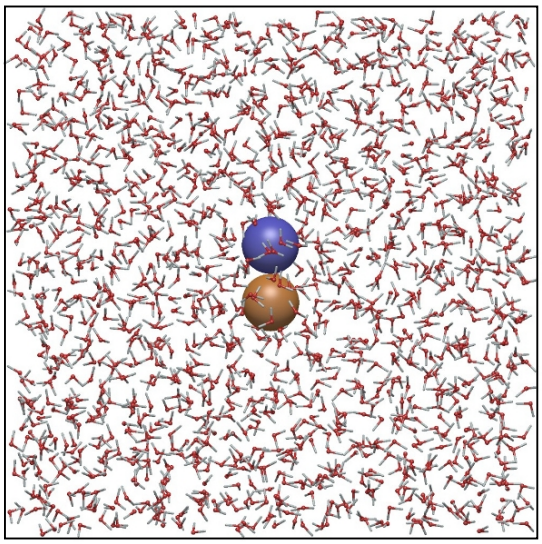


Figure S4: The model system used for the calculations of the free energy change of creating a dipole-moment in water (Table S4). Two spheres, 0.4 nm apart, were frozen in a simulation box with 1483 TIP3P water molecules. The Lennard-Jones parameters of the spheres are σ=0.4 nm and ϵ=0.4 kJ/mol. The two spheres carried charges with equal magnitudes but with opposite signs that simultaneously changed during the free energy calculations (molecular dynamics, thermodynamics integration, 11 λ-points with 15 ns simulation time at each λ-point, temperature coupling to 300K and pressure coupling to 1 atm) from 0.0 to 0.429667 e to model the creation of 8.25 D dipole, and from 0.0 to 0.366144 e to model the creation of 7.03 D dipole.

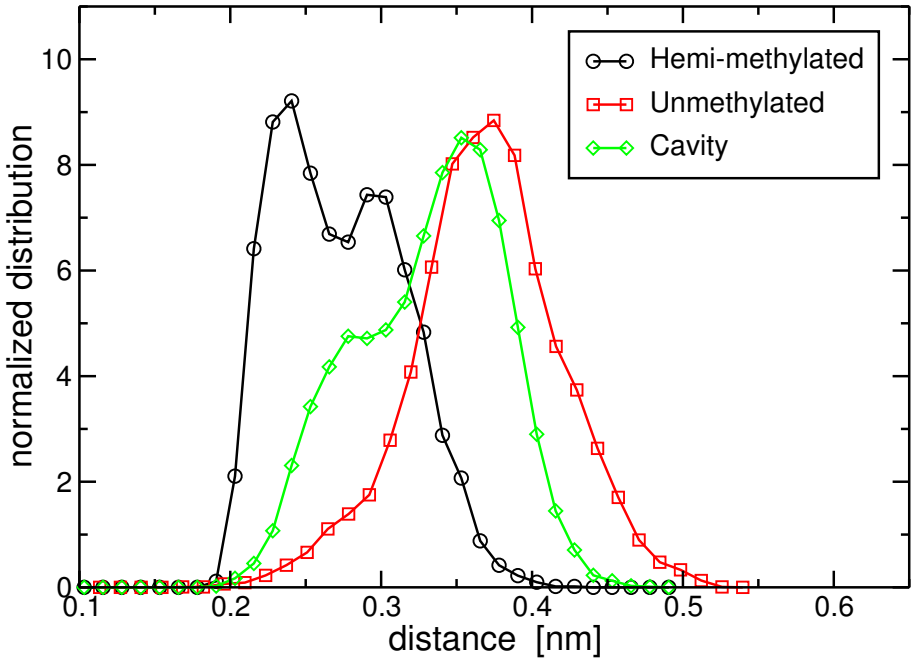


Figure S5: The normalized distribution of the closest distance between the flipped (methyl-)cytosine and Ser481 in the bound structure of UHRF1 with hemi-methylated, unmethylated, and the cavity-intermediate state DNA strands.

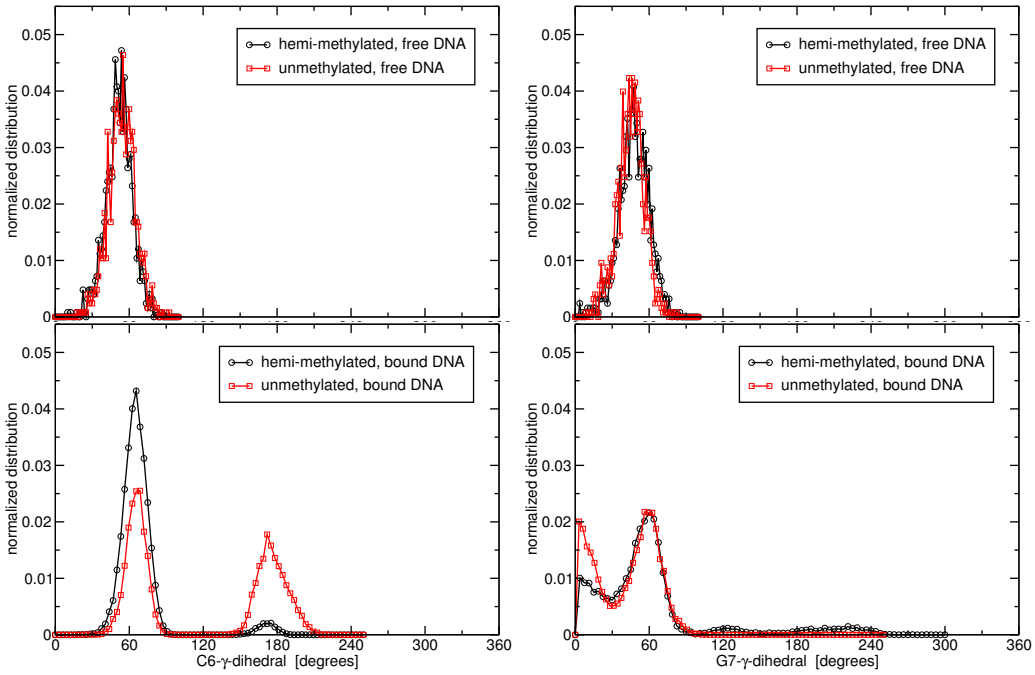


Figure S6: The γ dihedral (O5'–C5'–C4'–C3') angles of the DNA backbone around the (m)Cyt6 base, namely, those of C6 and G7. The graphs show that these dihedrals for both, hemi-methylated and unmethylated strands (averaged over the end states of the forward and backward directions) free in solution, are only in the *gauche* conformations. However, when Cyt6 flips-out of the DNA helix and binds to the binding pocket of UHRF1, the C6-γ dihedral also exhibits some degree of *trans* conformation. For the hemi-methylated strand the population of this *trans* conformation is very small.

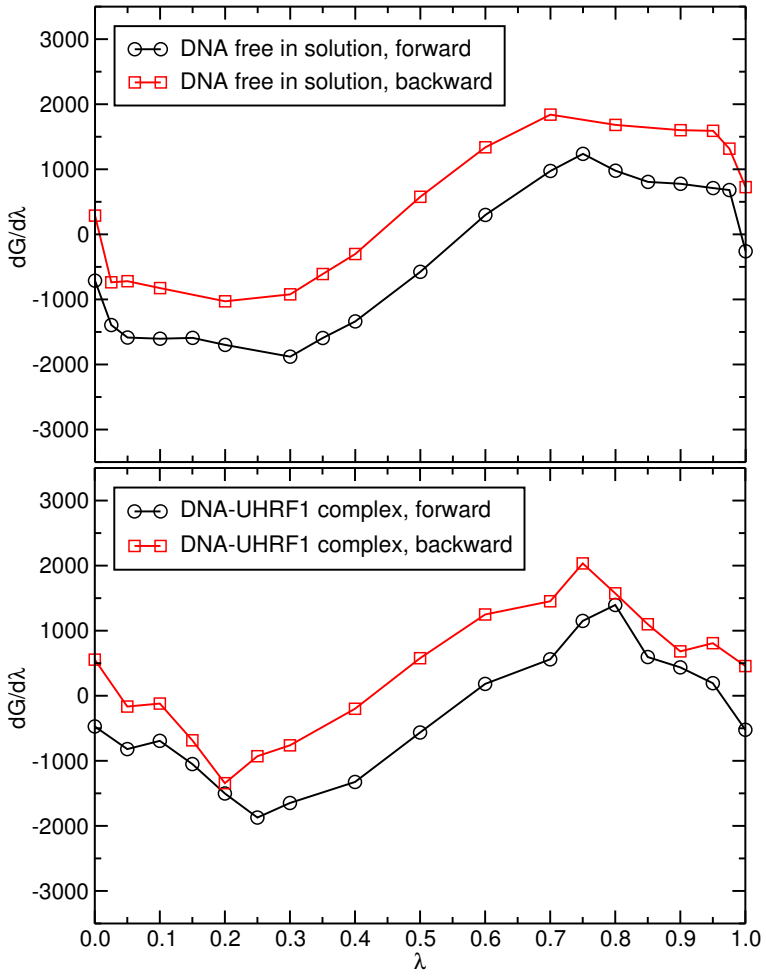


Figure S7: The value of $\langle \partial \mathcal{H} / \partial \lambda \rangle$ as a function of λ for mutating hemi-methylated to unmethylated CpG site for the unbound and bound DNA systems. Note, $\langle \partial \mathcal{H} / \partial \lambda \rangle$ does not include changes in mass (kinetic) or due to the distance constraints applied. These two contributions are cancelled out within the thermodynamic cycle and in fact in all cases their contributions to the change in binding affinity was less than 0.4 kJ/mol.

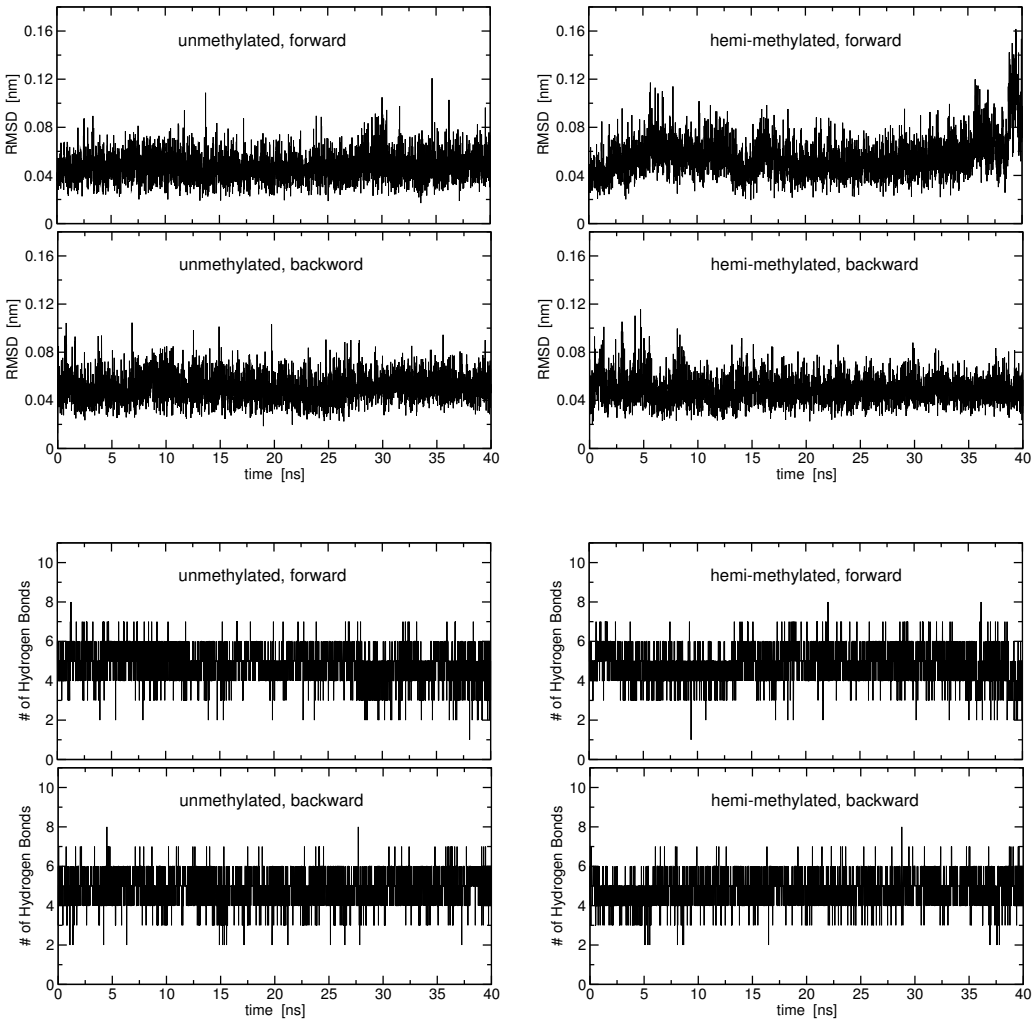


Figure S8: The root mean squared deviations of the Cyt6/mCyt6 after fitting to the four crucial residues of the binding pocket of UHRF1 of the first frame: Ala463, Gly464, Asp469, and Thr479, for unmethylated and hemi-methylated protein–DNA complexes. The trajectories are taken from the end points of the forward and backward directions of the free energy calculations. The plots of the number of hydrogen bonds between the Cyt6/mCyt6 base and UHRF1, as a function of time, for the same trajectories are shown in the lower panel. These graphs demonstrate that the protein–DNA bound complexes stay intact throughout the simulations.

References

[1] A. Widmer-Cooper, P. Harrowell, Free volume cannot explain the spatial heterogeneity of debye-waller factors in a glass-forming binary alloy, J. Non-Crystalline Solids 352 (2006) 5098–5102.

[2] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09 Revision A.02, Gaussian, Inc., Wallingford, CT (2009).

[3] C. Rauch, M. Trieb, B. Wellenzohn, M. Loferer, A. Voegele, F. R. Wibowo, K. R. Liedl, C5-Methylation of Cytosine in B-DNA Thermodynamically and Kinetically Stabilizes BI, J. Am. Chem. Soc. 125 (2003) 14990–14991.

[4] G. Schaftenaar, J. H. Noordik, Molden: a pre- and post-processing program for molecular and electronic structures, J. Comp. Aid. Mol. Design 14 (2000) 123–134.