Biophysical Chemistry 171 (2013) 38–45

Contents lists available at SciVerse ScienceDirect

Biophysical Chemistry

journal homepage: http://www.elsevier.com/locate/biophyschem

# UHRF1 discriminates against binding to fully-methylated CpG-Sites by steric repulsion

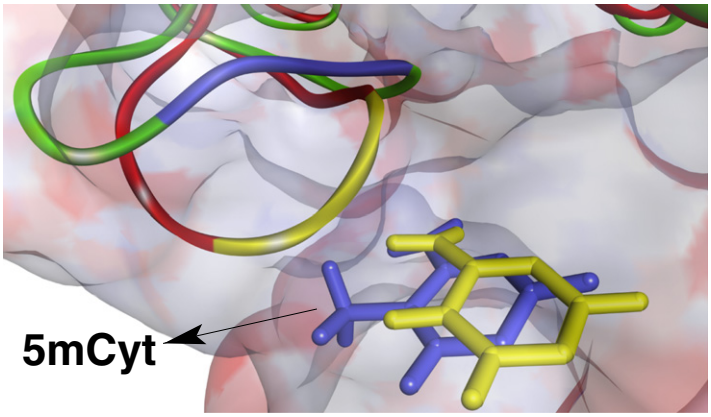
Caterina Bianchi [a], Ronen Zangi [a,b,*]

[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

## HIGHLIGHTS

- UHRF1 discriminates against binding to fully-methylated CpG sites.
- We perform MD simulations to address the mechanism of this discrimination.
- In fully-methylated complex the NKR domain is displaced away from the DNA.
- As a result, the interaction between UHRF1 and the recognized CpG site is weakened.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Article history:*
Received 11 September 2012
Received in revised form 9 October 2012
Accepted 9 October 2012
Available online 22 October 2012

*Keywords:*
Protein–DNA interaction
Epigenetics
DNA methylation
Molecular dynamics simulation
Free energy calculation

## ABSTRACT

Cytosine methylation of CpG dinucleotide sequence is an epigenetic mark on the DNA that regulates gene expression, chromatin structure, and genome stability. Although the enzyme that catalyzes the methylation reaction after replication is Dnmt1, it was found that the protein UHRF1 is essential for maintaining DNA methylation. UHRF1 exhibits preferential binding to hemi-methylated DNA relative to both unmethylated and fully-methylated DNA strands. In this paper we report results from molecular dynamics simulations aiming to elucidate the mechanism for the discrimination of UHRF1 to bind fully-methylated DNA. From alchemical mutation free energy calculations we find that the binding affinity of fully-methylated DNA to UHRF1 is weaker by 17.9 kJ/mol relative to the binding of hemi-methylated DNA. Structural analyses reveal, in agreement with the steric clash model, that a methyl group at the C5 position of the target cytosine induces a displacement of the NKR finger domain away from the DNA. As a result a net loss of, approximately, one hydrogen bond between the protein and the DNA is observed. These weakened protein–DNA interactions are located between the target cytosine and the NKR domain, as well as, between the flipped methylcytosine and the binding pocket of the SRA domain. Due to the conformational changes of the fully-methylated bound complex, water molecules intrude the protein–DNA interface and substitute the majority of the hydrogen bonds that are lost.



## 1. Introduction

The cell does not express all the genes coded in its genome. Some genes are "switched-on" whereas others are "switched-off". The current

* Corresponding author at: Department of Organic Chemistry I, University of the Basque Country UPV/EHU, Avenida de Tolosa 72, 20018, San Sebastian, Spain. Tel.: +34 943018112.
*E-mail address:* r.zangi@ikerbasque.org (R. Zangi).

0301-4622/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bpc.2012.10.002

knowledge of why certain genes are transcribed whereas others are not is attributed, among other factors, to chemical modifications of the DNA and core-histone proteins. The histone proteins are unique in the large number and types of their post-translational modifications. However, in the DNA of mammals the modifications occur only at the cytosine base, especially at the promoter regions. In particular, a methyl group substitutes a hydrogen atom at position 5 of the cytosine ring in the dinucleotide sequence CpG. This chemical mark has also an important roles in embryonic development, X-chromosome inactivation, and genomic imprinting [1–3]. The pattern of the methylated cytosines along the DNA passes from mother cells to daughter cells and their faithful inheritance and maintenance is essential to the wellbeing of the organism [4–7]. This can be accomplished owing to the ability of the DNA replication machinery to distinguish hemi-methylated DNA from either unmethylated or symmetrically di-(fully)-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 [5].

The enzyme that catalyzes the methylation of cytosines after replication is Dnmt1 [8,9]. The reaction involves a transfer of a methyl group from the cofactor S-adenosyl-L-methionine to the cytosine of hemi-methylated DNA [10]. Nevertheless, the recognition of hemi-methylated DNA is mediated by the protein UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1) and was found to be essential for maintaining DNA methylation [11,12]. UHRF1 contains a SET and RING-associated (SRA) domain, which is a methyl DNA binding domain, and as a consequence exhibits strong binding to hemi-methylated DNA. 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 fully-methylated CpG site. Essential for its function, UHRF1 is able to discriminate efficiently hemi-methylated from either unmethylated or fully-methylated CpG sites.

Interests in the structure and functions of UHRF1 have grown significantly in the last few years due to a distinct character of this protein. It contains 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 [13,14]. Due to these multiple functions of UHRF1 in gene regulation, it is not surprising that this protein is found to be over-expressed in many forms of cancer [15–17].

Among the X-ray structures solved recently for UHRF1 are its complexes with hemi-methylated DNA [18–21]. More specifically, the SRA domain of UHRF1 from different species bound to a model double stranded DNA were crystallized. A very distinct feature of the UHRF1–DNA interaction is the flipping of the recognized methylcytosine out of the DNA helix. The flipped methylcytosine forms several hydrogen bonds with the active site of the protein which prevents its sliding. Furthermore, base stacking interactions in the fipped-in state of the methylcytosine are replaced by stacking interactions this base makes with two tyrosine residues of the protein located on both sides of its plane. The binding pocket of the SRA domain specifically recognizes methylcytosine. It is too small to fit purines, and the discrimination against thymine is due to Asp469 and Thr479 (amino acids labeling is adopted from the human UHRF1 protein [18]) that accept two hydrogen bonds from N4 of the methylcytosine. When flipping out of the helix, the Watson-Crick hydrogen bonds between the methylcytosine and its paired guanine on the complementary strand are lost. In order to compensate for this energy loss and to strengthen the protein–DNA interactions, a finger, Asn–Lys–Arg (NKR), of the protein intrudes into the DNA double-helix and forms, via Arg491, three hydrogen bonds with the orphan guanine (G6′). The NKR finger, via Asn489, also forms hydrogen bonds with the following G7–C7′ base-pair.

The binding of unmethylated DNA to the SRA domain of UHRF1 is found to be negligible by several groups [11,19,21]. However, the binding affinity of fully-methylated is reported by one *in-vitro* experiment to be seven-fold weaker relative to the binding of hemi-methylated DNA [11], whereas another *in-vitro* experiment did not detect any complex formation at all with fully-methylated strand [19]. Other experiments also reported a weaker interaction between UHRF1 and fully-methylated DNA, however, the relative binding affinities were not quantified [18,20]. What is the mechanism by which UHRF1 recognizes hemi-methylated DNA? In a previous study we found that the discrimination against unmethylated DNA is a result of the change in the electron distribution around the cytosine ring upon methylation [22]. This has the effect of destabilizing the unbound state of methylcytosine in aqueous solution as well as stabilizing its bound state to UHRF1. Based on the X-ray structures, the discrimination against binding fully-methylated DNA is suggested to be due to a steric clash [18–20]. More specifically, it is proposed [18] that Asn489 and Arg491 "work together to effectively mask the C5 position of the non-methylated cytosine (Cyt7′)′′. The C5 atom of this unmethylated cytosine is within van der Waals distance of the backbone carbonyl-oxygen (which possibly interact through C O ⋯ H—C weak hydrogen bond) and amide side chain nitrogen of Asn489. At the same time this amide side chain of Asp489 is hydrogen bonded to the guanidino group of Arg491 as well as with the backbone phosphate group of Cyt7′.

Depending on the flexibility of the protein's residue at the active site, the volume taken by a methyl group relative to a hydrogen atom can cause some structural rearrangements that may or may not lead to loss of hydrogen bonds. What are then the conformational changes the protein undergo upon binding to fully-methylated DNA? Does it involve an overall loss of hydrogen bonds? If yes, between which protein–DNA groups of atoms? Given the conflicting reports in the literature about the relative binding affinity of UHRF1 to fully-methylated DNA, we address these questions by free energy molecular dynamics simulations. We find that by growing a methyl group at position 5 of Cyt7′ the NKR finger domain is substantially displaced away from the DNA. As a consequence, the interaction of the NKR domain with Cyt7′ is disrupted. In addition, this conformational change also induces a disruption of the interaction between the binding pocket of UHRF1 and the flipped-out methylcytosine (Cyt6). In total, we observe a net loss of, approximately, one hydrogen bond between UHRF1 and the fully-methylated DNA strand. The hydrogen bonds that are lost are substituted by hydrogen bonds with water molecules that penetrate the DNA–protein interface.

## 2. Methods

We perform molecular dynamics simulations in order to calculate the binding affinity of UHRF1 to a DNA strand containing a fully-methylated CpG site, and to elucidate possible conformational rearrangements, relative to a DNA containing a hemi-methylated CpG site. The initial structure taken for the simulations was the crystallographic structure of the SRA domain of UHRF1 (204 amino-acids long) complexed with a 12-mer hemi-methylated DNA (PDB accession code: 3CLZ) [18]. The sequence of one of the hemi-methylated strands is 5′-GGGCC-mC-GCAGGG-3′ (mC denotes 5-methylcytosine) which is base-paired to the complementary strand, 5′-CCCTGCGGGCCC-3′. Thus, the hemi-methylated CpG site is located halfway along the DNA double helix at positions mC6pG7 and the corresponding bases on the complementary strand are C7′pG6′. From the different X-ray structures deposited 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. An exception for these assignments was the case of Asp469 located at the active site of the SRA domain that contributes

to the binding affinity of the flipped methylcytosine to UHRF1. 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 methylcytosine indicates 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 [23]. Quantum mechanical calculations indicate that the proton is more likely to reside on the carboxylate group [10] 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 $+7\,e$. In addition, a 12 base-pair double-stranded DNA contributes a charge of $-22\,e$ 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 resulting in a total of 10510 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. In these simulations the nucleotide mC6 was in its flipped-in state because this is, by far, the more stable conformation when it is unbound to the protein. The initial structure of the DNA was an ideal B-DNA double helix conformation built using the PREDICTOR software [24].

The DNA, protein, and counterions were represented by the AMBER03 force-field [25,26] (note that for the nucleic acids the parameters are the same as those of AMBER99 force-field) and the water molecules by the TIP3P model [27]. 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. [28]. These charges were obtained from an ab-initio calculation using the Restrained ElectroStatic Potential (RESP) charge fitting procedure [29].

The molecular dynamics package GROMACS version 4.0.7 [30] 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 [31] (with real-space cut-off 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 by the velocity rescaling thermostat [32] with a coupling time of 0.1 ps, and at a pressure of 1.0 bar by the Berendsen thermostat [33] with a compressibility of $5\times10^{-5}$ 1/bar and a coupling time of 1.0 ps. Water bond distances and angles were constrained using the SETTLE algorithm [34] whereas the protein and DNA covalent bond distances were constrained using the LINCS algorithm [35]. 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/($mol\cdot$ nm$^2$). Then, a 10 ns of unrestrained simulation was performed. The configuration emerged from these simulations is used as an input for the free energy calculations.

The binding free energy of UHRF1 to hemi-methylated DNA relative to the binding to fully-methylated DNA was computed by the concept of a thermodynamic cycle [36] (see Fig. 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 hemi-methylated to fully-methylated DNA and vice versa [37]. More specifically, in the forward direction H5 of C7′ and three dummy atoms covalently bonded to it were mutated to a carbon atom and three hydrogen atoms, thus, transforming cytosine to methylcytosine. Whereas, the backward transformation describes the mutation of methylcytosine (mC7′) to cytosine (C7′). These transformations were performed for the DNA-UHRF1 complex and for the DNA free in solution. The free energy changes associated with these transformations were computed by the Thermodynamic Integration technique [38]. For each transformation, eleven equally spaced λ -points from $\lambda = 0$ to $\lambda = 1$ were constructed. At each λ -point the system was

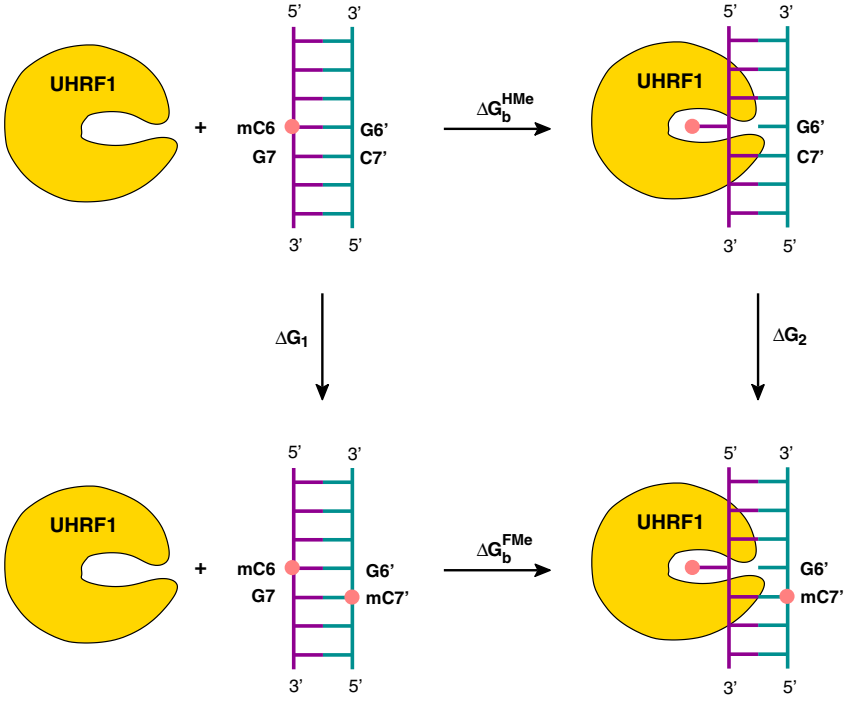



**Fig. 1.** The thermodynamic cycle constructed to calculate the binding free energy of UHRF1 to fully-methylated DNA (lower process, $\Delta G_b^{FMe}$) relative to the binding to hemi-methylated DNA (upper process, $\Delta G_b^{HMe}$). This relative binding affinity, $\Delta\Delta G_b^{FMe\text{-}HMe}$, equals $\Delta G_2 - \Delta G_1$ and the results from the simulations of the individual mutations are shown in Table 1.

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 UHRF1, respectively. However, for the DNA-UHRF1 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 two for each mutation) and the simulation time of some the λ -points were extended to 100 ns. For all transformations we calculated the free energy change in the forward and backward directions. For the DNA-UHRF1 complex, the hemi-methylated → fully-methylated transformation was performed first, and then, the fully-methylated DNA was mutated back to the hemi-methylated state. In general, the starting conformation for a particular λ -point was taken from the simulation of the preceding λ after a relaxation time in the range of 2–3 ns. At $\lambda = 0$ and $\lambda = 1$ for the forward and backward directions, thus for hemi-methylated and fully-methylated states, the system was simulated for 100 ns in order to obtain better statistics. The estimation of the errors of the free energy changes were obtained by integrating the error estimate at each λ -point. The error at each λ is calculated by dividing the standard deviation of $\partial\mathcal{H}/\partial\lambda$ by the square root of the number of independent data-points. The latter was estimated from the total time of the simulation for each λ divided by the autocorrelation time of $\partial\mathcal{H}/\partial\lambda$ [39].

All structural analyses for the hemi-methylated and fully-methylated states were averaged over the forward and backward transformations, thus, for 60 and 200 ns for the unbound and bound DNA, respectively. A hydrogen bond is defined by a donor–acceptor cutoff distance of 0.35 nm and a donor–hydrogen–acceptor angle larger than 150°. The errors in the number of hydrogen bonds were estimated using the block averaging method [40].

## 3. Results and discussion

The results of the mutation free energy changes defined in Fig. 1 are shown in Table 1. Given these values, the binding free energy change of UHRF1 to fully-methylated DNA, relative to hemi-methylated DNA, is $\Delta G_b^{FMe} - \Delta G_b^{HMe} = +17.9 \pm 6.4$ kJ/mol. Thus, at room temperature, fully-methylated CpG sites bind weaker to UHRF1 by, approximately, three orders of magnitudes. As discussed in the introduction, the experimental data are not settled yet and our results seem to be somewhere in between the factor of seven [11] and the absence of binding [19] reported in the literature. However, an error in the force-field on the order of 5 kJ/mol can change the number we obtained by one order of magnitude. Furthermore, it is assumed that if fully-methylated DNA binds to UHRF1 its binding mode can be reached from the binding mode of the hemi-methylated bound complex within the simulation time. Along the trajectory, the binding of fully-methylated DNA to UHRF1 results in small rearrangements of some of the protein residues (see below) compared with the UHRF1-hemi-methylated bound structure. It can be that the simulation time at each λ-point along the mutation in the bound complex was not long enough to observe a complete relaxation, in which case, the magnitude of the relative binding affinity will likely to be reduced. This incomplete relaxation might be somewhat implied by the relatively large difference of 5.5 kJ/mol between the forward and backward directions when the transformation took place inside the protein, whereas this difference for the DNA free in solution was very small, 0.6 kJ/mol.

**Table 1**
The free energy changes of the alchemical mutations shown in Fig. 1. The change in the binding free energy of UHRF1 to fully-methylated (FMe) DNA relative to hemi-methylated (HMe) DNA is also indicated. All values are given in kJ/mol.

| | Forward | Backward | Average |
|---|---|---|---|
| $\Delta G_1$ | $+448.4 \pm 5.1$ | $-447.8 \pm 5.3$ | $+448.1 \pm 3.7$ |
| $\Delta G_2$ | $+463.3 \pm 3.7$ | $-468.8 \pm 3.9$ | $+466.0 \pm 2.7$ |
| $\Delta\Delta G_b^{FMe-HMe} = \Delta G_2 - \Delta G_1$ | | | $+17.9 \pm 6.4$ |

In Table 2 we calculate the direct and water bridged hydrogen bonds between the protein and all the bases, phosphates, and sugars of the CpG dinucleotide on both strands. The difference between the two DNA strands is the presence of a methyl group instead of a hydrogen at position 5 of Cyt7′ (i.e., the cytosine that is not flipped out). Not surprisingly, the largest difference we observe is between the number of direct hydrogen bonds this cytosine nucleotide is forming with the SRA domain of UHRF1. In hemi-methylated DNA it forms on average 0.88 hydrogen bonds, whereas in fully-methylated DNA this value is only 0.27. A more detailed analysis of the groups participating in the Cyt7′-UHRF1 interaction, reveals that a major contribution for this excess (of 0.61 hydrogen bonds) arises almost entirely from the hydrogen bond between the carbonyl-oxygen of Asn489 and N4 amide hydrogens of Cyt7′ (an average difference of 0.56 hydrogen bonds). This is shown in Fig. 2a–c displaying this hydrogen–Acceptor distance for both trajectories and its total distribution. The excess of 0.25 in the number of direct hydrogen bonds in hemi-methylated DNA associated with the phosphate of Cyt7′ is also due to the interaction with Asn489, via its side-chain $-NH_2$ group. Furthermore, an excess of 0.38 hydrogen bonds is observed between Gua7 (via its O6 atom) and the amide backbone of Arg491.

As suggested based on the X-ray structures, the NKR finger domain is positioned in a location that would cause a steric clash if a fully-methylated CpG site is bound to UHRF1. Indeed, the results shown in Table 2 indicate that a significant portion of the hydrogen bonds that are lost in the fully-methylated complex involves interactions with the NKR finger domain. In Fig. 3 we superimpose an instantaneous conformation of the bound complex of hemi-methylated DNA on that of fully-methylated DNA. The conformation of the protein and DNA is not very different for the two complexes except for the NKR finger domain. Because of the 5-methyl group of Cyt7′ the NKR domain is pushed away from the DNA and the interaction with Cyt7′ seems to be weaker. A more quantitative description of the excluded volume effect of the C5 methyl group of Cyt7′ in fully-methylated DNA, is plotted in Fig. 4a. The figure exhibits the distribution of the distance between the C5-methyl atom and the closest atom of the NKR domain, i.e. the carbonyl oxygen of Asn489. In comparison with the corresponding distance to H5 in hemi-methylated DNA, a clear shift of more than 1 Å, toward larger distances is observed in the fully-methylated complex. Thus, the rearrangement of Asn489 to prevent the steric clash with the methyl group results in the loss of 0.86 hydrogen bonds with the Cyt7′ nucleotide (taking into account the base and the phosphate). Due to its weaker interactions with the DNA, Asn489 has a less defined structure. This is

**Table 2**
The average number of direct and water-bridged hydrogen bonds between the SRA domain of UHRF1 and different groups around the bound hemi-methylated and fully-methylated CpG sites. The total number of hydrogen bonds between the entire DNA and the SRA domain of UHRF1 is also indicated.

| | Direct HB | | Water-Bridged HB | |
|---|---|---|---|---|
| Interacting groups | FMe | HMe | FMe | HMe |
| mCyt6–UHRF1 | $4.14 \pm 0.1$ | $4.76 \pm 0.05$ | $0.16 \pm 0.03$ | <0.02 |
| Gua6′–UHRF1 | $1.89 \pm 0.02$ | $1.90 \pm 0.01$ | $0.23 \pm 0.06$ | $0.15 \pm 0.03$ |
| Gua7–UHRF1 | $0.06 \pm 0.01$ | $0.44 \pm 0.06$ | $0.59 \pm 0.06$ | $0.40 \pm 0.05$ |
| (m)Cyt7′–UHRF1 | $0.27 \pm 0.08$ | $0.88 \pm 0.05$ | $0.03 \pm 0.01$ | < 0.02 |
| Phosphate(mCyt6)–UHRF1 | $1.14 \pm 0.06$ | $1.10 \pm 0.07$ | $2.0 \pm 0.4$ | $1.80 \pm 0.09$ |
| Phosphate(Gua6′)–UHRF1 | <0.02 | <0.02 | <0.02 | <0.02 |
| Phosphate(Gua7)–UHRF1 | $1.02 \pm 0.04$ | $0.91 \pm 0.08$ | $1.0 \pm 0.1$ | $1.1 \pm 0.2$ |
| Phosphate((m)Cyt7′)–UHRF1 | <0.02 | $0.25 \pm 0.02$ | $0.05 \pm 0.03$ | $0.10 \pm 0.03$ |
| Sugar(mCyt6)–UHRF1 | <0.02 | <0.02 | $0.19 \pm 0.04$ | $0.20 \pm 0.06$ |
| Sugar(Gua6′)–UHRF1 | <0.02 | <0.02 | <0.02 | <0.02 |
| Sugar(Gua7)–UHRF1 | <0.02 | <0.02 | $0.63 \pm 0.07$ | $0.46 \pm 0.06$ |
| Sugar((m)Cyt7′)–UHRF1 | <0.02 | <0.02 | <0.02 | <0.02 |
| DNA–UHRF1 | $18.1 \pm 0.7$ | $19.4 \pm 0.6$ | $5.2 \pm 0.6$ | $5.1 \pm 0.5$ |

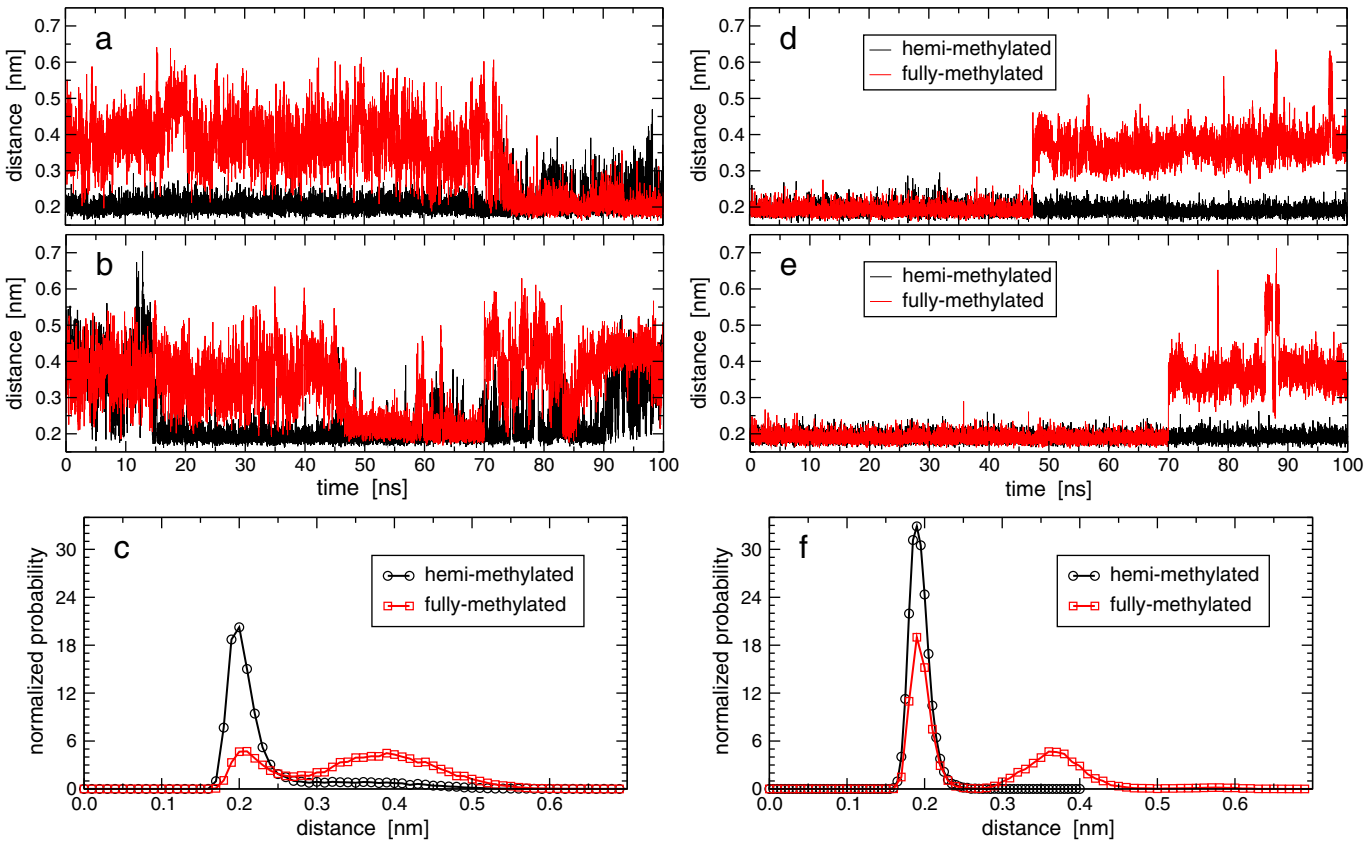



**Fig. 2.** The distance between the carbonyl-oxygen of Asn489 and the closest N4 hydrogen of (m)Cyt7′ for the (a) forward and (b) backward transformations, for hemi-methylated and fully-methylated DNA strands bound to UHRF1. (c) The normalized distribution of this distance. The corresponding plots for the distance between the N3 atom of mCyt6 and $H_\delta$ (the hydrogen of the protonated carboxy-oxygen) of Asp469 are plotted in (d), (e), and (f). Note that the backward trajectories (at $\lambda = 0$) do not start from the point the forward trajectories (at $\lambda = 1$) end because the starting conformation at a given $\lambda$ -point was taken after only 2–3 ns simulation time of the preceding $\lambda$ -point (see Methods section for more details).

demonstrated in Fig. 4b displaying larger root mean squared deviations of the Asn489 atoms, with respect to the protein excluding the NKR finger domain, in fully-methylated DNA.

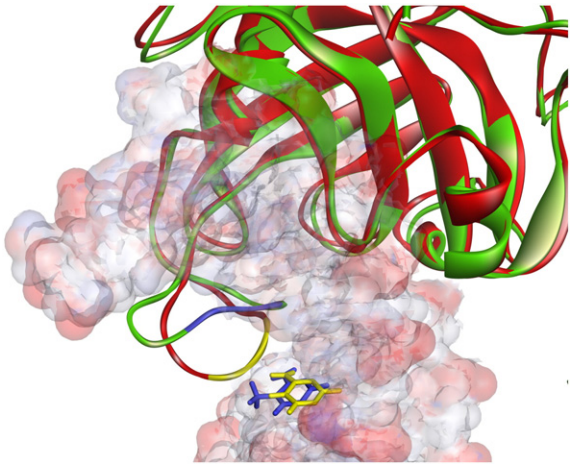


**Fig. 3.** Superposition of instantaneous configurations of UHRF1 complexed with hemi-methylated (red) and fully-methylated (green) DNA strands. In both cases, the DNA double helix is displayed only by its van der Waals surface except of (m)Cyt7′ that is shown also as colored sticks. Note that the main difference in the structure between the SRA domain in both complexes is the conformation and positioning of the NKR finger domain. The backbone of the NKR finger domain, as well as, (m)Cyt7′ are colored by yellow and blue for hemi-methylated and fully-methylated complexes, respectively.

Instantaneous configurations of the interaction of Cyt7′ and Gua7 with the NKR finger domain in both complexes are shown in Fig. 5. The relative motion of Asn489 with respect to Cyt7′, and as a consequence the loss of the hydrogen bond between its carbonyl oxygen and N4 amide group of Cyt7′, is clearly shown in the fully-methylated complex. The numbers shown in the plots correspond to the average distances over the entire trajectories. In order to satisfy the loss of the hydrogen bond that broke with the N4 of Cyt7′, water molecules intrude into the interface between Asn489 and the DNA. We calculated the number of water molecules within a radius of 4 Å from the C(5-Me)/H5 atom and found values of 1.2 and 3.1 in hemi-methylated and fully-methylated complexes, respectively. These waters make an average of 0.4 hydrogen bonds with the N4 amide group in fully-methylated DNA whereas this number is essentially zero for the hemi-methylated complex. The compensation of the lost hydrogen bond of the carbonyl oxygen of Asn489 is a little bit more complex. This carbonyl oxygen does not make a hydrogen bond with these intruding waters but it does with the guanidino group of Arg491. This interaction comes on the expense of the hydrogen bond this guanidino group makes with $O_\gamma$ of Asn489 in hemi-methylated DNA. The hydrogen bond with $O_\gamma$ of Asn489 is then replaced by a hydrogen bond this residue makes with N4 of Cyt5. Cyt5 is not part of the CpG dinucleotide sequence that is recognized in hemi-methylated DNA. These rearrangements of hydrogen bonds within the NKR finger domain are illustrated in Fig. 6 which exhibits the distances between the atoms that are involved in the formation and breakage of these hydrogen bonds. Two other important hydrogen bonds that are lost due to the rearrangement of the NKR domain in fully-methylated DNA are between the O6 of Gua7 and the amide backbone of Arg491, as well as, between the

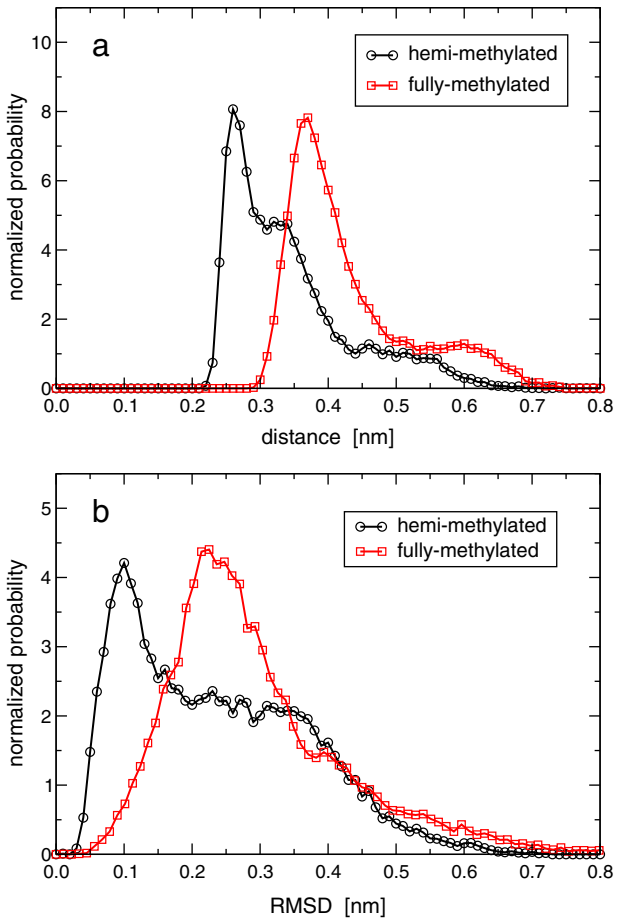



**Fig. 4.** (a) The normalized distributions of (a) the distance between C(5-Me)/H5 of (m) Cyt7′ and the carbonyl-oxygen of Asn489, and (b) the root-mean squared deviations of the Asn489 residue atoms after fitting the protein, excluding the NKR domain, to its initial conformation in both trajectories, for the hemi-methylated and fully-methylated DNA–UHRF1 complexes.

phosphate of (m)Cty7′ and the side-chain amide of Asn489. These hydrogen bonds are replaced by hydrogen bonds the intruding waters make with these groups except for the side-chain amide of Asn489 for which only partial substitution with these water molecules is observed.

In Table 2 it is indicated that not only the interactions with the NKR finger domain of UHRF1 are disrupted by the C(5-Me) group of Cyt7′ but also the interactions with the active site of the SRA domain to which the flipped mCyt6 binds to. In this case, an average of 0.62 hydrogen bonds are lost. The major contribution of this loss (0.41 hydrogen bonds) is between the N3 of mCyt6 and $H_\delta$ (the hydrogen of the protonated carboxy-oxygen) of Asp469. In Fig. 2d–f we plot the distance between these atoms for hemi-methylated and fully-methylated complexes from both the forward and backward transformations. This hydrogen bond is very stable in the hemi-methylated complex throughout the entire trajectories. However, in the fully-methylated complex, although also very stable at the initial segment of the trajectories, a sudden (after 47 ns and 70 ns for the forward and backward mutations, respectively) loss of this hydrogen bond is observed without any sign of recovery until the end of these trajectories. Thus, it is very likely that if these trajectories were extended for longer times, the loss of this hydrogen bond would have had a larger contribution to the different averages collected.

Snapshots of the interactions between mCyt6 and the binding pocket of the SRA domain of UHRF1 for hemi-methylated and fully-methylated DNA strands are shown in Fig. 7. In general, the interaction of the flipped-out methylcytosine and the binding pocket of the SRA domain has to be strong to compensate for the lost Watson-Crick base pairing and support the extrahelical conformation. In the complex with hemi-methylated DNA, this is done by the formation of five hydrogen bonds. Despite the partial loss of some of these hydrogen bonds, the positioning of the methylcytosine in the binding pocket for fully-methylated complex was similar to that of the hemi-methylated complex. However, also here, water molecules entered the protein–DNA interface and compensated for the lost hydrogen bonds between the N3 of mCyt6 and $H_\delta$ of Asp469. We computed the average number of water molecules within a radius of 4.0 Å from the protonated $O_\delta$ of Asp469 to be 0.1 and 1.0 for hemi-methylated and fully-methylated complexes, respectively. This allows the protonated $O_\delta$ of Asp469 to act as a donor for a hydrogen bond with this water molecule. On average, 0.92 hydrogen bonds are formed between the intruding waters and $O_\delta$ of Asp469 when we consider the trajectories from the onset of the loss of the Asp469—mCyt6 hydrogen bond. The replacement of the N3 of mCyt6 hydrogen bond is only partial; a corresponding calculation yielded a value of 0.31 hydrogen bonds. The configuration we choose for the fully-methylated DNA complex shown in Fig. 7 corresponds to a case in which the intruding water molecule is engaged with hydrogen bonding with $H_\delta$ of Asp469 and N3 of mCyt6. Other contributions, albeit smaller, to the weaker interaction between mCyt6 and the binding pocket of the SRA domain of UHRF1 arises also from the hydrogen bonds between N4 of mCyt6 and Asp469/Thr479 (a difference of 0.12 hydrogen bonds), as well as between the carbonyl group of mCyt6 and Ala463/Gly465 (a difference of 0.11 hydrogen bonds).

Note that no substantial difference is observed in the number of protein–DNA hydrogen bonds bridged by water molecules between all groups analyzed (Table 2). For some groups belonging to the

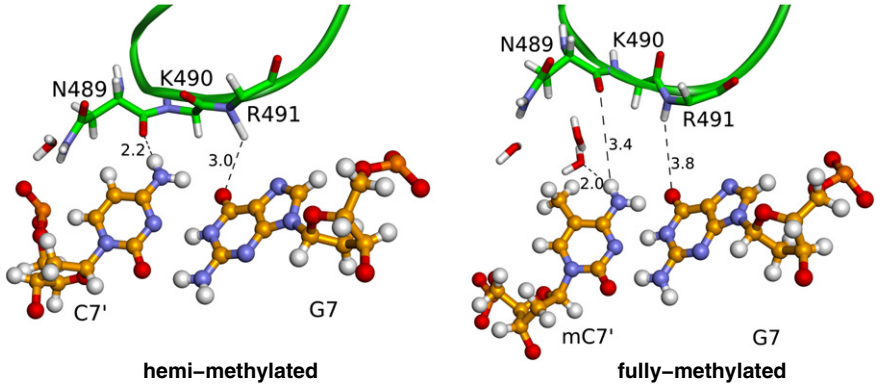



**Fig. 5.** Snapshots of the interaction between the NKR finger domain of UHRF1 and hemi-methylated DNA (left), as well as, with fully-methylated DNA (right). All water molecules within a radius of 4.0 Å from the C(5-Me)/H5 atom are shown. The hydrogen bonds (for hemi-methylated) and the corresponding distances (for fully-methylated) between the residues Asn489 and Arg491 of UHRF1 and the DNA are denoted by dashed lines. The numbers shown correspond to the H–acceptor distances (in Å) averaged over all trajectories. Note that the waters intruding the protein–DNA interface in the fully-methylated complex substitute the hydrogen bond that N4 of mCyt7′ lost due to the 5-methyl-group on this cytosine base.

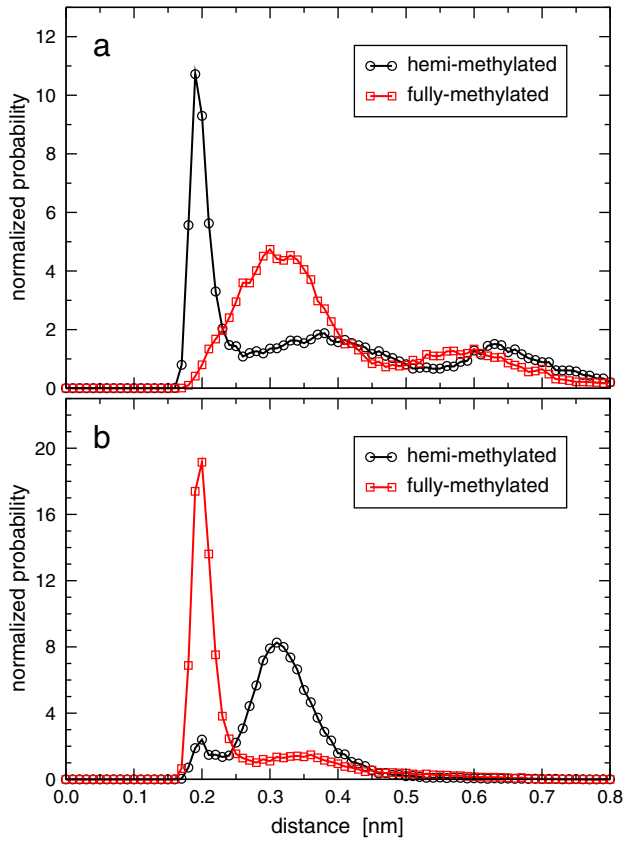



**Fig. 6.** The rearrangements of the hydrogen bonds within the NKR finger domain. The normalized distribution between the closest hydrogen of the guanidino group of Arg491 and (a) the $O_\gamma$, as well as (b) the carbonyl-oxygen, of Asn489 in hemi-methylated and fully-methylated DNA complexed to UHRF1.

CpG site, slightly larger numbers are observed in the fully-methylated complex. This can be attributed to the larger number of water molecules present between the protein and the DNA at these locations. Nevertheless, the total number is essentially the same. The sugar groups of the CpG site do not form direct interactions with the protein. When considering the total number of direct (and water-bridged) hydrogen bonds between the entire protein and the entire DNA, there is a loss of approximately one hydrogen bond in fully-methylated DNA. A loss of one hydrogen bond corresponds, approximately, to the relative change in free energy of + 17.9 kJ/mol we obtained in Table 1. However, larger loss is observed for the interaction between the protein and the CpG dinucleotide. As mentioned above, a compensation for this loss is established between Asn489 and Cyt5 which is not part of the CpG site. It might be that because the interaction with the CpG site, including the flipped methylcytosine, is weakened whereas the interaction with other residues are slightly strengthened, the binding to fully-methylated will induce further conformational changes in which no base-flipping occur and the binding is not specific to the CpG site. This might explain the contrasting results obtained experimentally on the strength of binding between UHRF1 and (different) fully-methylated DNA strands which actually have different DNA sequences.

## 4. Conclusions

To perform its function, UHRF1 preferentially binds (via its SRA domain) DNA strands with a hemi-methylated, over fully-methylated, CpG site. Because the difference between these two sites is the presence a methyl group instead of a hydrogen atom, this methyl group must be responsible for the recognition observed. Aiming to address this discrimination against binding to fully-methylated strands we performed molecular dynamics simulations with free energy calculations. The starting conformation for the simulations is the X-ray structure of the SRA domain of UHRF1 bound to hemi-methylated DNA. We find that the excluded volume of a methyl group on the target cytosine to be methylated (in fully-methylated DNA) induces a conformational change in the NKR finger domain and displaces this domain away from the DNA. As a result, weakened protein–DNA interactions are observed at two different locations involving the recognized CpG site. The first location is directly related to these rearrangements and is associated with the loss of hydrogen bonds between the NKR finger domain and the second G–C base-pair of the CpG site. The second location is at the binding pocket of the SRA domain to which the flipped-out methylcytosine is bound to. In this case, a hydrogen bond between N3 of the methylcytosine and $H_\delta$ (the hydrogen of the protonated carboxy-oxygen) of Asp469 is lost. These lost protein–DNA hydrogen bonds are replaced by hydrogen bonds with water molecules that intrude the protein–DNA interface in the fully-methylated complex due to the NKR domain rearrangements. Furthermore, a rearrangement of hydrogen bonds within the NKR domain is also observed. In total, a net loss of, approximately, one hydrogen bond is observed in the fully methylated complex and, correspondingly,

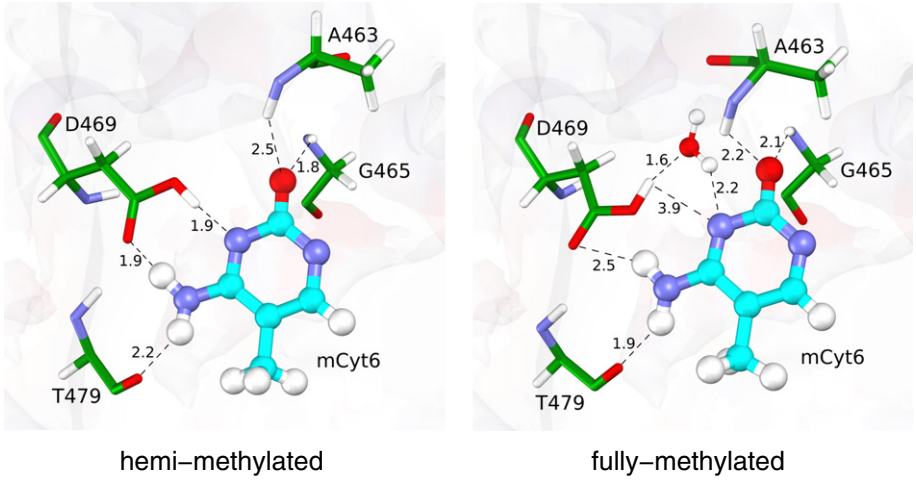



**Fig. 7.** Snapshots of the interaction between the flipped methylcytosine and the binding pocket of the SRA domain of UHRF1 for hemi-methylated and fully-methylated complexes. Note that in the latter complex, the hydrogen bond between $H_\delta$ (the protonated carboxy-oxygen) of Asp489 and N3 of mCyt6 is lost and substituted by a hydrogen bond with an intruding water molecule. In this figure, the numbers indicated correspond to hydrogen–acceptor instantaneous distances.

the results from the free energy calculations indicate that the binding affinity to UHRF1 is weaker by $17.9 \pm 6.4$ kJ/mol. We conjecture that because the flipped-out methylcytosine is not entirely bound to the protein in the fully-methylated complex, it might be that this flipped-out state is not stable enough. In this case, base flipping will not take place and the binding of UHRF1 to fully-methylated DNA is not specific to CpG sites. The magnitude of such nonspecific binding is likely then to dependent on the sequence of the bound DNA segment.

## Acknowledgments

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, MICINN, grant number CTQ2010-20297. Technical and human support provided by SGIker (USED SERVICES) (UPV/EHU, MICINN, GV/EJ, ESF) is gratefully acknowledged. The authors also thankfully acknowledge the computer resources and technical assistance provided by the Barcelona Supercomputing Center — Centro Nacional de Supercomputación.

## References

[1] T.H. Bestor, The DNA methyltransferases of mammals, Human Molecular Genetics 9 (2000) 2395–2402.
[2] A. Jeltsch, Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases, Chembiochem 3 (2002) 274–293.
[3] M. Szyf, Targeting DNA methylation in cancer, Ageing Research Reviews 2 (2003) 299–328.
[4] A. Bird, DNA methylation patterns and epigenetic memory, Genes & Development 16 (2002) 6–21.
[5] M.G. Goll, T.H. Bestor, Eukaryotic cytosine methyltransferases, Annual Review of Biochemistry 74 (2005) 481–514.
[6] G. Egger, S. Jeong, S.G. Escobar, C.C. Cortez, T.W.H. Li, Y. Saito, C.B. Yoo, P.A. Jones, G. Liang, Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival, Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 14080–14085.
[7] E. Hervouet, L. Lalier, E. Debien, M. Cheray, A. Geairon, H. Rogniaux, D. Loussouarn, S.A. Martin, F.M. Vallette, P.-F. Cartron, Disruption of dnmt1/pcna/uhrf1 interactions promotes tumorigenesis from human and mice glial cells, PloS One 5 (2010) e11333.
[8] M. Damelin, T.H. Bestor, Biological functions of DNA methyltransferase 1 require its methyltransferase activity, Molecular and Cellular Biology 27 (2007) 3891–3899.
[9] F. Spada, A. Haemmer, D. Kuch, U. Rothbauer, L. Schermelleh, E. Kremmer, T. Carell, G. Längst, H. Leonhardt, DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells, The Journal of Cell Biology 176 (2007) 565–571.
[10] R. Zangi, A. Arrieta, F.P. Cossío, Mechanism of DNA methylation: the double role of DNA as a substrate and as a cofactor, Journal of Molecular Biology 400 (2010) 632–644.
[11] M. Bostick, J.K. Kim, P.-O. Estève, A. Clark, S. Pradhan, S.E. Jacobsen, UHRF1 plays a role in maintaining DNA methylation in mammalian cells, Science 317 (2007) 1760–1764.
[12] J. Sharif, M. Muto, S. ichiro Takebayashi, I. Suetake, A. Iwamatsu, T.A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, H. Koseki, The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA, Nature 450 (2007) 908–913.
[13] H. Hashimoto, J.R. Horton, X. Zhang, X. Cheng, UHRF1, a modular multi-domain protein, regulates replication-coupled crosstalk between DNA methylation and histone modifications, Epigenetics 4 (2009) 8–14.
[14] M. Unoki, J. Brunet, M. Mousli, 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, Biochemical Pharmacology 78 (2009) 1279–1288.
[15] M. Unoki, T. Nishidate, Y. Nakamura, ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain, Oncogene 23 (2004) 7601–7610.
[16] Y. Jenkins, V. Markovtsov, W. Lang, P. Sharma, D. Pearsall, J. Warner, C. Franci, B. Huang, J. Huang, G.C. Yam, J.P. Vistan, E. Pali, J. Vialard, M. Janicot, J.B. Lorens, D.G. Payan, Y. Hitoshi, Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth, Molecular Biology of the Cell 16 (2005) 5621–5629.
[17] G.J. Schaaf, J.M. Ruijter, F. van Ruissen, D.A. Zwijnenburg, R. Waaijer, L.J. Valentijn, J. Benit-Deekman, A.H.C. van Kampen, F. Baas, M. Kool, Full transcriptome analysis of rhabdomyosarcoma, normal, and fetal skeletal muscle: statistical comparison of multiple SAGE libraries, The FASEB Journal 19 (2005) 404–406.
[18] G.V. Avvakumov, J.R. Walker, S. Xue, Y. Li, S. Duan, C. Bronner, C.H. Arrowsmith, S. Dhe-Paganon, Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1, Nature 455 (2008) 822–825.
[19] K. Arita, M. Ariyoshi, H. Tochio, Y. Nakamura, M. Shirakawa, Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism, Nature 455 (2008) 818–821.
[20] H. Hashimoto, J.R. Horton, X. Zhang, M. Bostick, S.E. Jacobsen, X. Cheng, The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix, Nature 455 (2008) 826–829.
[21] C. Qian, S. Li, J. Jakoncic, L. Zeng, M.J. Walsh, M.-M. Zhou, Structure and hemimethylated CpG binding of the SRA domain from human UHRF1, The Journal of Biological Chemistry 283 (2008) 34490–34494.
[22] C. Bianchi, R. Zangi, How to Distinguish methyl-Cytosine from Cytosine with High Fidelity, Journal of Molecular Biology 424 (3–4) (December 7, 2012) 215–224, http://dx.doi.org/10.1016/j.jmb.2012.09.024.
[23] S. Kumar, X. Cheng, S. Klimasauskas, S. Mi, J. Posfai, R.J. Roberts, G.G. Wilson, The DNA (cytosine-5) methyltransferases, Nucleic Acids Research 22 (1994) 1–10.
[24] J. Farwer, M.J. Packer, C.A. Hunter, PREDICTOR: a web-based tool for the prediction of atomic structure from sequence for double helical DNA with up to 150 base pairs, In Silico Biology 7 (2007) 595–600.
[25] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? Journal of Computational Chemistry 21 (2000) 1049–1074.
[26] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, Journal of Computational Chemistry 24 (2003) 1999–2012.
[27] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, The Journal of Chemical Physics 79 (1983) 926–935.
[28] 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, Journal of the American Chemical Society 125 (2003) 14990–14991.
[29] P. Cieplak, W.D. Cornell, C. Bayly, P.A. Kollman, Application of the multimolecule and multiconformational RESP methodology to biopolymers: charge derivation for DNA, RNA, and proteins, Journal of Computational Chemistry 16 (1995) 1357–1377.
[30] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, Journal of Chemical Theory and Computation 4 (2008) 435–447.
[31] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems, The Journal of Chemical Physics 98 (1993) 10089–10092.
[32] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, The Journal of Chemical Physics 126 (2007) 014101.
[33] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, The Journal of Chemical Physics 81 (1984) 3684–3690.
[34] S. Miyamoto, P.A. Kollman, SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water models, Journal of Computational Chemistry 13 (1992) 952–962.
[35] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, Journal of Computational Chemistry 18 (1997) 1463–1472.
[36] W.F. van Gunsteren, H.J.C. Berendsen, Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry, Journal of Computer-Aided Molecular Design 1 (1987) 171–176.
[37] A. Villa, R. Zangi, G. Pieffet, A.E. Mark, Sampling and convergence in free energy calculations of protein–ligand interactions: the binding of triphenoxypyridine derivatives to factor Xa and trypsin, Journal of Computer-Aided Molecular Design 17 (2003) 673–686.
[38] J.G. Kirkwood, Statistical Mechanics of Fluid Mixtures, The Journal of Chemical Physics 3 (1935) 300–313.
[39] S.A. Hassan, Intermolecular potentials of mean force of amino acid side chain interactions in aqueous medium, The Journal of Physical Chemistry. B 108 (2004) 19501–19509.
[40] H. Flyvbjerg, H.G. Petersen, Error estimates on averages of correlated data, The Journal of Chemical Physics 91 (1989) 461–466.