THE IMPACT OF THE HYDROXYMETHYL-CYTOSINE EPIGENETIC SIGNATURE ON DNA STRUCTURE AND FUNCTION

Federica Battistini1, Pablo D. Dans1,2,3, Montserrat Terrazas1, Chiara L. Castellazzi1, Guillem Portella1,4, Mireia Labrador1, Núria Villegas1, Isabelle Brun-Heath1, Carlos González5 and Modesto Orozco1,6*

1Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology. Baldiri Reixac 10–12, 08028 Barcelona, Spain.  
2Department of Biological Sciences, CENUR North Riverside, University of the Republic (UdelaR), Gral. Rivera 1350, 50000 Salto, Uruguay. 
3Functional Genetics Lab., Institute Pasteur of Montevideo, Mataojo 2020, 11400 Montevideo, Uruguay. 
4Chemistry Department, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK. 
5Instituto Química Física Rocasolano. Consejo Superior de Investigaciones Científicas (CSIC), Serrano 119, 28006 Madrid, Spain. 
6Department of Biochemistry and Molecular Biology. University of Barcelona, 08028 Barcelona, Spain.

* Correspondence to: Prof. Modesto Orozco: modesto.orozco@irbbarcelona.org
ABSTRACT

We present a comprehensive, experimental and theoretical study of the impact of 5-hydroxymethylation of DNA cytosine. Using molecular dynamics, biophysical experiments and NMR spectroscopy, we found that Ten-Eleven translocation (TET) dioxygenases generate an epigenetic variant with structural and physical properties not too different to those of 5-methylcytosine. Experiments and simulations demonstrate that 5-methylcytosine (mC) and 5-hydroxymethyl-cytosine (hmC) generally lead to more rigid duplexes with poorer circularization efficiencies and lower ability to form nucleosomes. In particular, we can rule out the hypothesis that hydroxymethylation reverts to unmodified cytosine physical properties, as hmC is even more rigid than mC. Thus, we do not expect dramatic changes in the chromatin structure induced by differences in physical properties between d(mCpG) and d(hmCpG). On the contrary, our simulations suggest that methylated-DNA binding domains (MBD), associated with repression activities, are very sensitive to the substitution d(mCpG) → d(hmCpG), while MBD3 which has a dual activation/repression activity is not sensitive to the d(mCpG) → d(hmCpG) change. Overall, while changes in gene activity due to cytosine methylation are the result of the combination of stiffness-related chromatin reorganization and MBD binding, those associated to 5-hydroxylation of methylcytosine could be explained by a change in the balance of repression/activation pathways related to differential MBD binding.
INTRODUCTION

Genomic expression is a finely controlled process regulated by a myriad of external and internal effectors which ensure that genetic information is expressed whenever and wherever required in the interest of the entire organism (1). External effectors are proteins or RNAs, which can regulate directly DNA transcription activation or repression. Internal effectors are chromatin chemical modifications (epigenetic changes) (2) that can affect DNA expression by recruiting activator or repressor proteins; or by controlling the accessibility of the transcriptional machinery to the target DNA sequences.

In developed organisms, such as mammals, DNA epigenetic changes are highly prevalent and most CpG steps (except for those in CpG islands) are methylated at the 5'-position of the pyrimidine ring (3, 4). The presence of 5-methylcytosine (mC) has been traditionally related to gene inactivation, (5–8) but analysis of methylation changes in Leukaemia strongly suggests that the real connection between methylation and gene expression is probably more complex and differs in each individual gene (9). It is known that methylated DNA (met-DNA) recruits proteins containing methylated-DNA binding domains (MBD;10) triggering protein-specific responses. Furthermore, it has been described that DNA methylation modulates DNA accessibility, (11) mainly by making the DNA stiffer and less prone to wrap around nucleosomes (12, 13), leading to nucleosome shifts and accordingly to changes in the pattern of accessible and hidden sequences.

The portfolio of DNA epigenetic modifications in mammals has been recently expanded with 5-hydroxymethylcytosine (hmC) (14, 15). This modified nucleobase was initially considered just an intermediate step in the de-methylation pathway (5-Methylâ†’ hydroxymethylâ†’ formylâ†’ carboxylâ†’ CO₂) (16, 17), but different studies on the expression of the protein catalyzing mCâ†’hmC conversion (Tet: Ten-Eleven translocation family of dioxygenases)(18–22) and genome wide analysis have shown a crucial role of hmC in the transcription (23–25), especially in processes related to tissue differentiation (26–28). Interestingly, for still unknown reasons, hmC is prevalent in specific brain cell types, as well as in embryonic stem cells (ESCs), with consequential reduction during differentiation (29–33). When present, hmC signals are not randomly distributed, but appear located near transcriptional starting sites and promoters, as well as in certain CpG islands (24, 34–36), suggesting a specific role of this epigenetic modification in gene activity.

The presence of hmC seems to correlate with a reversion of mC effects (37), but the molecular mechanisms connecting the presence of the hmC with gene activity are unclear. Two main possibilities emerge: i) a change in the DNA physical properties reverting to the unmethylated situation that leads to nucleosome repositioning; ii) a direct protein-mediated mechanism. Using a variety of experimental and theoretical methods, we explore here both possibilities. We found that physical properties of hmC are significantly different to those of unmodified cytosine, but not dramatically different to those of mC, suggesting that the reversion of methylation effects in gene expression by TET action cannot be explained simply by change in physical properties of DNA involving nucleosome remodelling. No specific hmC-binding proteins have been described, to our knowledge, but the analysis of known MBDs (16, 38) show a surprisingly wide range of substrate specificity. Our calculations suggest MBD with dual activation/repression activity shows nearly equal affinity for mC and hmC, while those with exclusively inactivating role have a large preference for the methylated form. This can be the ultimate reason of the differential effect of mC and hmC in gene expression.
MATERIALS AND METHODS

UV-monitored thermal-denaturation studies

Modified and unmodified 8-nucleotide DNAs (CGAC*GTCG) were synthetized (details in Supporting Methods, Supplementary Table S1) and their absorbance versus temperature curves of each duplex was measured at 5 µM, 20 µM and 66 µM strand concentration in phosphate buffer (25 mM) containing NaCl (100 mM). Experiments were performed in Teflon-stoppered quartz cells of 1 cm and 1 mm length path on a JASCO V-650 spectrophotometer equipped with thermoprogrammer. The samples were heated to 90 °C, allowed to cool slowly to 25 °C, and then warmed during the denaturation experiments at a rate of 0.5 °C min\(^{-1}\) to 90 °C. The absorbance was monitored at 260 nm. The data were analyzed by the denaturation-curve processing program, MeltWin v.3.0. Melting temperatures (T\(_m\)) were determined by computer fit of the first derivative of absorbance with respect to 1/T. These values were designed to give uniformly separated data points on ln (C\(_t\)/4) scale. ∆H, ∆S, and ∆G values were estimated from dependence of melting temperatures on DNA concentrations (39). The reciprocal values of average melting temperatures were plotted against ln (C\(_t\)/4) and fitted to linear relationships, 1/T\(_m\) = (R/∆H) ln(CT/4) + (∆S/∆H).

NMR

Samples of the duplexes with a central CpG step: d(CGTC*GACG)·d(CGTC*GACG) and d(CGAC*GTCG)·d(CGAC*GTCG) with and without epigenetic modifications (synthesis details in Supplementary Methods) were dissolved in either D\(_2\)O or 9:1 H\(_2\)O/D\(_2\)O in 100 mM NaCl and 10 mM sodium phosphate, pH=7 buffer. All NMR spectra were acquired in Bruker Avance spectrometers operating at 600 and 800 MHz, equipped with cryoprobes and processed with the TOPSPIN software. DQF-COSY, TOCSY and NOESY experiments were recorded in D\(_2\)O at 25°C. In the experiments in D\(_2\)O, presaturation was used to suppress the residual H\(_2\)O signal. The NOESY spectra in D\(_2\)O were acquired with a mixing time of 250 ms. TOCSY spectra were recorded with standard MLEV17 spinlock sequence, and a mixing time of 80 ms. NOESY spectra in H\(_2\)O were acquired with 150 ms mixing time at 5°C to reduce exchange with the water. In 2D experiments in H\(_2\)O, water suppression was achieved by including a WATERGATE module prior to acquisition. The spectral analysis program SPARKY (D. K. TD Goddard and D. G. Kneller, SPARKY v3, University of California, San Francisco) was used for semiautomatic assignment of the NOE cross-peaks and quantitative evaluation of the NOE intensities. Additional details of the NMR analysis are given in Supplementary Material.

Refinement of the experimental NMR structures

Ensembles of structures using atomistic force-fields based on the distance constraints obtained experimentally were calculated using classical and usual annealing procedure similar to that used to refine most NMR structures (40). Accordingly, ideal fiber B-DNA and A-DNA structures are thermalized (298 K) and equilibrated for 100 ps each (using the same options described previously), applying harmonic restraints of 100 kcal/mol·Å\(^2\) on the DNA. Then, a 500 ps MD simulation is performed where the global restraints are replaced by the specific NMR distance constraints obtained experimentally. To obtain the final ensemble, 50 structures were chosen and minimized in vacuo, keeping the NMR constraints.

HydroxyMethylCytosine Parameterization

Geometrical parameters for hmC were derived starting from parmbsc1 (41) cytosine and serine from ff14SB for hydroxymethyl group. Final hmC geometry was obtained by energy minimization (42). Charges were derived from RESP fit at HF/6-31G* level starting from an optimized hmC capped at the C1’ using the procedure described in reference (43). Library with
parameters for hmC are available upon request.

Molecular dynamics simulations

We performed molecular dynamics (MD) simulations for 33 12-mer DNA duplexes of sequence CGCGXC*GYCCCG, being C*= C, mC or hmC, with X and Y selected to represent all the 10 possible unique tetranucleotide environments containing a central CG step, were each simulated for 0.5 μs. Additionally the poli(CpG) sequence C*GC*GC*GC*GC*GC*G being C*= C, mC or hmC were simulated. Starting structures were taken from Arnott canonical B-DNA (44), and oligomers were built using the Nucleic Acid Builder (45), epigenetic modification (mC and hmC) were afterwards added to the duplex. The oligomers were simulated using periodic boundary conditions with a truncated octahedral box and an explicit solvent of TIP3P water molecules (46), with a minimum thickness of 10 Å around the solute. The DNA net charge was neutralized with K+ cations and K+Cl− ion pairs substituted water molecules to reach a concentration of ∼0.15 M. Each oligomer was simulated using the AMBER 18 (47) program suite and considering the 3 epigenetic modifications led to a total of 33 simulations and 16.5 μs of unrestrained trajectories. The simulation was performed using our well-established multistep protocol (48–50) which involves energy minimizations of the solvent, slow thermalization and a final re-equilibration for 10 ns, prior to the 0.5 μs production runs. All simulations were carried out in the isothermal-isobaric ensemble (T = 298 K, P = 1 atm) using the parmbsc1 force field (41) and Dang parameters for ions (51, 52). Long-range electrostatic effects were treated using the Particle Mesh Ewald method (53) using a cutoff of 10 Å. The bonds involving hydrogen were restrained using SHAKE (54) and The temperature and the pressure, with a coupling constant of 5 ps were controlled using Berendsen algorithm (55). Trajectories were stored in our BigNAsim database (56), following recent recommendations for FAIR trajectory storage (57).

Analysis of the trajectories

All the trajectories were processed with the cpptraj module of the AmberTools 18 package (47). The ability of DNA to recognize sodium was analysed using our classical molecular interaction potential (cMIP (58)). The electrostatic interaction term was determined by solving the linear Poisson–Boltzmann equation, while the van der Waals contribution was determined using standard AMBER Lennard–Jones parameters. The ionic strength and the reaction-field dielectric constant were set to 0.15 and 78.4 M, respectively, while the dielectric constant for DNA was set to 8 (59). The molecular plots were generated using either VMD 1.9 (60) or the UCSF Chimera package version 1.8.1 (61).

Analysis of DNA physical properties and Deformation Energy calculation

DNA helical parameters and backbone torsion angles were measured with the Curves+ and Canal programs (62) along the MD simulations and for X-ray crystal structures. From the ensemble of MD simulations, the covariance matrix describing the deformability of the helical parameters for each dinucleotide step was computed and inverted to generate the 6×6 stiffness matrix, resulting in the force constants for helical deformation at the base pair step level (63, 64). Average and force constants were calculated for the last 200ns of each MD simulation. The position of each cation in curvilinear cylindrical coordinates for each snapshot of the simulations with respect to the instantaneous helical axis was determined using Canion, a utility program that reads the ion counts (65). Deformation energy for the wild type, mCpG and hmCpG DNA sequences was calculated using our mesoscopic energy model (13), with parameters fitted (see above) for each epigenetic modification in each tetramer environment (see above).

Circularization Assays

Phosphorylation and radioactive labeling. Every single stranded oligonucleotide (1 nmol) was
5'-end-phosphorylated with 40U of T4 Polynucleotide kinase (New England Biolabs) by incubation at 37°C in 1x T4 DNA Ligase Reaction Buffer. **Annealing and ligation.** The complementary strands were denatured at 90°C and subsequently annealed by gradually decreasing the temperature to room conditions. Double stranded oligonucleotides were then multimerized with 400U of T4 DNA ligase (New England Biolabs) by an overnight incubation at room temperature. The reaction was stopped by inactivating the ligase at 65°C for 10 minutes. The DNA was ethanol precipitated and dissolved in 10 µl of DNAse free water. **Two-Dimensional Gel Electrophoresis.** The ligation/digestion products were loaded on a 5% polyacrylamide native gel (19:1 acrylamide:bisacrylamide) and electrophoresed at room temperature at 4 V/cm in TBE buffer. After separating the different DNA species on the first dimension, the lanes of interest were excised from the gel and loaded on a two dimensional gel (8% polyacrylamide) to separate circular and linear DNA. The separation of the two families of DNA was facilitated by the presence of chloroquine phosphate (250 µg/ml) that distorts the linear molecules of DNA in a different way from the circular ones (3). The gels were stained with Sybr Safe (Invitrogen) and visualized on the ImageQuant system (GE Healthcare) under UV light (see Results). **In vitro nucleosome reconstitution** Free energies for histone binding in **in vitro** nucleosome reconstitution were measured for the wild type 601 sequence and the epigenetically mutate versions, mCpG601 and hmCpG601 (details in Supplementary Methods). The 3 forms of the high affinity 601.2 DNA sequence (sequences in Supplementary Information), with normal C-, mC-, hmC-DNA were chemically synthetized starting from the oligos 65-, 66-, 80- and 81-nucleotide long DNAs on the 0.2 µmol scale on an Applied Biosystems 394 synthesizer (see Supplementary Methods for details, Supplementary Table S1). We introduced 2 CpG base pair steps 5 or 11 base pairs apart: in positions that face the protein core through the minor and major groove or, alternatively, only the major grooves. Such positions explore sites in the nucleosomal DNA of negative and positive opening of the base pairs along their long axis (13). Bands intensities corresponding to nucleosome-bound or free DNA were measured by densitometry using the PhosphorImager system (GE Healthcare) (Supplementary Figure S1). Bands intensities quantitation allowed to calculate the affinity for each sequence to form nucleosome as the ratio of counts in nucleosome bound DNA to counts in free DNA. Experimental normalised affinity was then computed dividing the relative affinity for methylated or hydroxymethylated sequences by the control (unmodified cytosine). Each experiment, for each sequence, was repeated four times. **Thermodynamic Integration** We used a thermodynamic cycle to compute the reversible work associated to the alchemical transformation between two DNA sequences (between mC and hmC), both in the bound and in the unbound state for a list of all MDB protein-DNA complexes (Supplementary Methods). The details on how these calculations were carried out are explained in a previous work in which we used the same method to calculate the variation in free energy adding methylation in a nucleosome complex (13). In this case, to achieve the alchemical transformation between mC and hmC, our molecular dynamics simulations convert the hydroxyl group in position 5 of the cytosine ring into a hydrogen atom. In each complex for each mutation, the variation of λ is discretized in 21 windows, i.e. dλ=0.05, and the final ΔG is computed via numerical integration. We used soft-core potentials implemented in GROMACS to avoid singularities in the Lennard-Jones and Coulomb potentials, with α=0.3, σ=0.25 and a soft-core power of 1. The initial structures for each window were obtained from a 10 ns simulation in which λ was continuously varied from 0 to 1. From this simulation we extracted frames corresponding to a given λ value, and we relaxed the structures by minimizing the energy of the system in those configurations. We simulated each window at fixed λ value for 1 ns, and we discarded the first 100 ps of simulation. For each window we collected 9 estimates for ⟨(δH)/δλ⟩_λ by using 9 blocks of 100 ps, which were then integrated.
through the entire mutation pathway to obtain mutation free energies (with associated statistical errors).

RESULTS AND DISCUSSION

The effect of 5-hydroxymethylcytosine on the stability of DNA. We first analysed the changes in stability due to methylation and hydroxymethylation of the model sequence d(CGAC*GTCG):d(CGAC*GTCG). UV-Melting experiments demonstrated (see Methods and Supplementary Table S2) a slight increase in the melting temperature of hmC with respect to unmodified DNA (2-3 degrees), and a slight decrease (around 1 degree) from mC-DNA. Processing of melting profiles confirms the small impact of C→mC→hmC substitutions in duplex stability, and the predictable enthalpy/entropy compensation, with the hydrophobicity term favouring (as expected, (66)) the stability of mC (mC > hmC > C, see Supplementary Table S3).

The effect of 5-hydroxymethylcytosine on dsDNA structure. NMR studies. We studied the impact of hmC at d(CpG) steps on the conformation of DNA by NMR experiments using two model sequences with a central CpG step: d(CGTC*GACG):d(CGTC*GACG) (central base pair very flexible, (67)) and d(CGCGAC*GTCGG):d(CGCGAC*GTCGG) (central base pair very stiff, (67)). NMR data was collected for C*=C, hmC and C*=C, mC and hmC, respectively; allowing us to obtain, for the first time direct experimental information on the structure of DNA related to epigenetic changes C→mC→hmC placed on both strands of the duplex (i.e., d(C*pG):d(C*pG)) (see Methods and Supplementary information). Sequential assignments of exchangeable and non-exchangeable proton found small, but significant changes in chemical shifts of the groups involved in H-bonding at the substitution site (chemical shifts at C* amino move from 8.21 ppm (C*=C) to 8.41 ppm (C*=mC) and 8.54 ppm (C*=hmC)) (data in Supplementary Tables S4-S9 and Figures S2-S5). The NMR signal of –OH in the hmC duplex could not be detected, most probably due to rapid exchange with the solvent, suggesting that OH-N6 hydrogen bond is not highly populated. Angular and distance constraints derived from J-couplings and NOE cross-peaks, respectively, allowed us to determine the solution structures (see Methods and Supplementary information) for the mC and hmC containing duplexes in the two distinct tetrameric environments. In all cases structures agree very well with a general B-like duplex (Figure 1). The changes induced by the presence of mC or hmC are small, in general less important than those induced by tetramer-dependency (TC*GA vs AC*GT) and a detailed analysis at the central base pair step (d(CpG); see Table 1) demonstrates that at the very local level the structural impact of epigenetic changes in DNA are modest.
Figure 1. Average NMR structure of the sequences d(CGTC*GACG)·d(CGTC*GACG) and d(CGCGAC*GTCGCG)·d(CGCGAC*GTCGCG) (C* being C, mC or hmC respectively) for the central octamer (A) and the central tetramer (B).

**AC*GT**

| Base pair parameter | NMR hmCpG       | NMR mCpG       | NMR CpG        |
|---------------------|-----------------|----------------|----------------|
| Shift (Å)           | -0.13 ± 0.79    | -0.16 ± 0.23   | 0.11 ± 0.25    |
| Slide (Å)           | -0.33 ± 0.20    | -0.33 ± 0.21   | -0.09 ± 0.18   |
| Rise (Å)            | 3.25 ± 0.21     | 3.35 ± 0.01    | 3.20 ± 0.17    |
| Tilt (degrees)      | 1.0 ± 2.0       | -0.35 ± 0.73   | 2.8 ± 2.1      |
| Roll (degrees)      | 11.0 ± 3.5      | 11.85 ± 4.1    | 8.7 ± 3.6      |
| Twist (degrees)     | 34.9 ± 3.5      | 37.9 ± 2.8     | 37.2 ± 1.9     |

**TC*GA**

| Base pair parameter | NMR hmCpG       | NMR CpG        |
|---------------------|-----------------|----------------|
| Shift (Å)           | 0.37 ± 0.73     | 0.37 ± 0.28    |
| Slide (Å)           | -0.40 ± 0.32    | 0.79 ± 0.25    |
| Rise (Å)            | 3.08 ± 0.21     | 3.02 ± 0.22    |
| Tilt (degrees)      | 1.4 ± 3.4       | 2.8 ± 2.65     |
| Roll (degrees)      | 10.5 ± 4.3      | 9.5 ± 3.0      |
| Twist (degrees)     | 33.1 ± 2.9      | 28.0 ± 2.6     |

Table 1. Average parameters (in Å and Degrees) with standard deviation for the central step in the NMR structure of the sequences d(CGTC*GACG)·d(CGTC*GACG) and d(CGCGAC*GTCGCG)·d(CGCGAC*GTCGCG) (C* being C, mC or hmC).

The effect of 5-hydroxymethylcytosine on the physical properties of DNA. Unbiased MD ensembles of d(CGTC*GACG)·d(CGTC*GACG) and d(CGAC*GTCG)·d(CGAC*GTCG) agrees very well with the MD-NMR-biased ones (RMSd deviation around 0.7-1.1 Å from the
average NMR structure for the central tetramers; see detailed analysis at the C*pG step in Supplementary Figure S6), supporting the quality of parmbsc1 simulations (41). This allows us to analyse the impact of C→mC→hmC substitutions in a much wider range of sequence environments and in the absence of the artificial restraints used in NMR refinements. Such restraints which are nearly-harmonic bias the intrinsic deformability of the DNA hindering the possibility to sample for example, the bimodality at the central CpG step, a well-known polymorphism found both in MD simulation and in high-resolution experimental structures (67–70). With this aim we collected 0.5 µs trajectories (see Methods) of 10 dodecamers containing all the unique tetramer environments of the d(C*pG) step (for C*=C, mC or hmC). Additional trajectories were collected for d(C*pG)₆·d(C*pG)₆, a poly(CpG) dodecamer mimicking CpG islands, repetitive polymers with unique biological importance. In summary, we collected 16.5 µs of unbiased trajectories of dodecamers containing C*pG tetramers (0.5 µs x 33 structures), which confirm that the introduction of epigenetic variants of cytosine leads to only mild structural alterations in the duplex, and the differences between mC- and hmC-DNAs being very small (Supplementary Table S10). Modifications induced by epigenetic changes are typically smaller than tetramer-induced changes, but there is a clear general tendency for a reduction in twist and an increase in roll when C is substituted by mC or hmC (Figure 2).

![Figure 2](image_url)

**Figure 2.** Average value for the base pair parameters roll (left panel) and twist (right) for C*G step in the different tetramer environment. Being C* normal cytosine (blue dots), methylated cytosine (green dots) and hydroxymethylated cytosine (red dots).

In the case of twist, the epigenetic changes induce (in general) a displacement of the bimodal CpG distribution towards low twist values (Figure 3), while in the case of roll, it results in the movement of the normal distribution towards higher values, probably related to the steric hindrance of the cytosine 5-substituent (Figure 3). Note that it is difficult, due to the refinement procedures, to detect changes in twist bimodality in NMR structures, but the high twist to low twist transition detected by MD in flexible sequences has been already detected in a high-resolution crystal structure (PDB ID 4GLH, 4HLI and 4GLC) (see Supplementary Figure S7), supporting the validity of our theoretical findings. Finally, and very interestingly, we found that some tetramers show quite unique distributions (for example the high twist of AC*GT) and that
polymer effects are not negligible as shown in the d(CGCG)-d(CGCG) tetramer behaves differently when embedded in the reference dodecamer or in a poly(CpG).

**Figure 3.** Roll (left panel) and twist (right panel) distributions for each C*pG step, in a different tetramer environment, being C* normal (A), methylated (B) and hydroxymethylated (C). Dashed line for C*pG step in a poly(CpG) sequence.

Stiffness analysis (see Methods) (63, 64) illustrates the impact of epigenetic changes on the elastic deformability of DNA. In general, cytosine methylation increases the stiffness of CpG steps,
mainly due to the restriction of roll flexibility. Such a stiffness increase is even more evident for hmC (see Table 2 and Supplementary Table S1), due to the narrowing of the roll distribution (see Figure 3), probably related to the formation of transient hydrogen bond interactions involving the hydroxyl group of hmC (see Supplementary Figure S8 as examples). It is worth noting, again, that these general trends are not universal and significant deviations are evident for certain tetramers (for example ACGT and TCGA). Again poly(CpG) has a rather differential behaviour, as for this repetitive sequence methylation increases rather than decrease flexibility, which combined with results above suggest that epigenetic changes can affect physical properties of CpG islands in a quite different way than in the rest of the DNA. This unique behaviour may explain claims by other authors of a general increase in DNA flexibility induced by cytosine methylation (71).

| A | $K_{hn-tw}^t$ | $K_{roll-roll}^t$ | $K_{tilt-tilt}^t$ |
|---|---|---|---|
| CpG | hmCpG | CpG | hmCpG | CpG | hmCpG | CpG | hmCpG |
| ACGA | 0.025 | 0.024 | 0.025 | 0.016 | 0.018 | 0.025 | 0.033 | 0.031 |
| ACGC | 0.023 | 0.024 | 0.026 | 0.014 | 0.015 | 0.023 | 0.034 | 0.031 |
| ACGG | 0.025 | 0.024 | 0.026 | 0.015 | 0.016 | 0.027 | 0.033 | 0.031 |
| ACGT | 0.029 | 0.034 | 0.027 | 0.015 | 0.021 | 0.016 | 0.036 | 0.038 |
| CCGA | 0.027 | 0.027 | 0.028 | 0.019 | 0.022 | 0.028 | 0.033 | 0.032 |
| CCGC | 0.025 | 0.026 | 0.026 | 0.018 | 0.021 | 0.027 | 0.033 | 0.032 |
| CCGG | 0.027 | 0.027 | 0.026 | 0.018 | 0.022 | 0.028 | 0.031 | 0.031 |
| GCCA | 0.027 | 0.028 | 0.029 | 0.018 | 0.022 | 0.029 | 0.034 | 0.033 |
| GCCG | 0.024 | 0.026 | 0.024 | 0.016 | 0.018 | 0.025 | 0.034 | 0.032 |
| TCGA | 0.025 | 0.027 | 0.025 | 0.020 | 0.025 | 0.024 | 0.032 | 0.033 |
| Poly(CG) | 0.022 | 0.024 | 0.027 | 0.018 | 0.015 | 0.018 | 0.033 | 0.030 |

| B | $K_{tot}^{rot}$ | | hmCpG |
|---|---|---|---|
| CpG | mCpG | hmCpG | CpG | mCpG | hmCpG |
| ACGA | 0.024 | 0.024 | 0.028 |
| ACGC | 0.022 | 0.022 | 0.029 |
| ACGG | 0.023 | 0.023 | 0.030 |
| ACGT | 0.025 | 0.030 | 0.025 |
| CCGA | 0.026 | 0.027 | 0.031 |
| CCGC | 0.025 | 0.026 | 0.031 |
| CCGG | 0.025 | 0.026 | 0.031 |
| GCCA | 0.025 | 0.027 | 0.032 |
| GCCG | 0.024 | 0.025 | 0.029 |
| TCGA | 0.025 | 0.028 | 0.027 |
| average | 0.024 | 0.026 | 0.029 |
| Poly(CG) | 0.024 | 0.022 | 0.024 |

**Table 2.** A) Diagonal stiffness constants for rotational movements in kcal/mol deg$^2$ for the central C*pG step (C*=C, mC and hmC) in the different tetrameric environments. B) Cubed root of the product of the diagonal stiffness constants for rotational movements in kcal/mol deg$^2$.

To validate our claim that epigenetic variants tend to increase the DNA stiffness (in the majority of the sequence contexts), we carried out DNA circularization experiments using the double stranded 21mers (d(GAAAAACCGGGC*GAAAAACCGG)·d(TCCCGTTTTTC*GCCCCCGTTTT)) with a central C*pG dinucleotide (C*=C, mC, hmC, respectively) as described in Perez et al (12). Under favorable ligation conditions (see Methods),
the multimers form circles that are as short as allowed by the flexibility of the DNA. Results shown in Figure 4 clearly demonstrate that, as suggested by our MD simulations (Table 2), epigenetic variants increase (in most sequence context) the stiffness of DNA, following the order: hmC>mC>C. In summary, the reversion of methylation signature by hydroxymethylation cannot be explained as a recovery of the properties of the unmethylated DNA.

![Figure 4](image)

**Figure 4.** (A-C) 2D polyacrylamide native gels showing different migrations of linear and circular DNA species oligomers of 21 bp, respectively for (A) Cytosine (C), (B) Methylcytosine (mC) and (C) Hydroxymethylcytosine (hmC) containing fragments. Linear DNA molecules are positioned on the lower diagonal, and circular DNA molecules are positioned on the upper diagonal. (D) The circularization efficiency expressed as the ratio between Circular and Linear molecules of the same size (base pairs, bp).

The reduction of flexibility introduced by epigenetic variants of cytosine is expected to increase the energetic cost of wrapping the DNA around the nucleosome core (13), leading to a decrease in nucleosome affinity. To confirm this hypothesis, we compute deformation energies (see Methods) associated to nucleosome wrapping for DNA containing unmodified CpG, mCpG and hmCpG steps. To include phasing issues into consideration we explored two scenarios: i) 4 epigenetic changes (2 d(CpG) steps modified) “in-phase”, i.e., each modification separated by 11 bases (the periodicity of DNA in a nucleosome (72)) and ii) 4 epigenetic changes in “anti-phase”, i.e. each modification separated by 5 bps. Results in Figure 5A demonstrate that epigenetic variations increase the energy cost required for nucleosome wrapping, both when introduced in “in-phase” and “anti-phase” (Supplementary Figure S9), suggesting that change in flexibility rather than in the intrinsic curvature (roll and tilt) are the main responsible of the epigenetic-induced difference in nucleosome affinity. To validate these theoretical findings, we performed nucleosome reconstitution experiments (see Methods, Supplementary Methods) (73) using both “in-phase” and “anti-phase” arrangements (the same sequences as before). Results (Figure 5B) confirm that the C⇒mC⇒hmC substitution leads to a reduction in the efficiency of nucleosome reconstitution, the difference between methylated and hydroxymethylated DNAs being small. Interestingly, experimental changes in nucleosome reconstitution induced by epigenetic alterations “in-phase” are not dramatically different than those found in “anti-phase”, where the
effect of changes in intrinsic curvature should cancel, supporting the claims that changes in flexibility are the main responsible for epigenetic-related changes in nucleosome affinity.

Figure 5. A) Deformation energy variation cost required for nucleosome wrapping, difference between the energy cost of the methylated (mC) and hydroxymethylated (hmC) DNA respect to the control (unmodified cytosine). At the top a schematic view of the positioning of the modifications (red dots), respectively 5 and 11 base pairs apart on the DNA wrapped around the histones (in yellow). B) Experimental affinity ratio, with standard deviation, between the methylated (mC) and hydroxymethylated (hmC) DNA respect to the control (unmodified cytosine) calculated using nucleosome reconstitution.

Altogether, theoretical and experimental results strongly suggest that C→mC→hmC substitution at d(CpG) steps lead to small changes in structure and to a small, and environment specific increase in the stiffness (hmC>mC>C), which is reflected in a reduction of nucleosome affinity. Very interestingly, the behavior of DNA containing hmC is much closer to that of met-DNA than that of unmodified DNA, which means that the “reversion” to the unmodified DNA behavior found when methylated DNA is processed by TET-proteins to yield hmC-DNA (22) cannot be explained by changes in the intrinsic physical properties of DNA and the associated recovering of the nucleosome distribution.

The impact of epigenetic modifications on the recognition properties of DNA. The preferred recognition pattern of DNA changes with epigenetic modifications and this can modulate DNA-protein contacts. Thus, DNA containing CpG steps exhibits a preferred region of interaction with cations along the minor groove (Figure 6) (see Methods) (65). Cytosine methylation widens slightly the minor groove, displacing the preferred interaction region to the backbone and hydroxymethylation makes the major groove the preferred interaction region (see examples in Supplementary Figure S10), probably as a consequence of the presence of a polar group. The polarity of the CH₂OH group is also visible in the water distributions (see examples in Supplementary Figure S11), where the integration of the hydroxyl into highly ordered strings of waters leads in general to a very dense solvation atmosphere along the major groove.
Figure 6. Examples of classical molecular interaction potentials using sodium as a probe (58). For the sake of comparison, all the averaged structure with the 3 epigenetic modifications of the cytosine were aligned and the same isosurface -3.5 kcal/mol was computed.

Results above suggest epigenetic modifications alter DNA interaction pattern, which might impact the binding of methylated DNA binding domains (MBDs). We focus here on the four MBD proteins related to the regulation of gene expression: i) MeCP2, a protein with direct (via repression transcription domain; TRD) and indirect (via recruitment of histone deacetylases) silencing activity in mature nerve cell (3, 38, 74), ii) MBD1 which represses directly (via TRD) or indirectly (via recruitment of histone methylases) gene expression (38), iii) MBD2 which is also coupled to repression (directly through TRD, or indirectly through histone methylation and deacetylation), and which is known to induce NuRD-dependent nucleosome remodeling (38, 75, 76) and finally iv) MBD3 which lacks TRD and which has been related to a myriad of secondary interactions with other partners such as NuRD, DNMT and TET leading to either gene expression or activation (30, 38, 76).

We built models of the four MBD-DNA complexes from the available experimental data (details in Methods and Supplementary Methods). The complexes were solvated and used as starting conformations for MD simulations, coupled to thermodynamic integration (MD/TI), which combined with standard thermodynamic cycles (Figure 7) allowed us to determine the impact in binding of methyl-cytosine to hydroxymethyl-cytosine change in protein binding (see Methods). Quite interestingly, despite the very high sequence similarity the four studied MBD have a different discriminant hmC vs mC power. Based on our simulations, MBD1 has the highest specificity for mC compared to hmC, MBD2 and MeCP2 show a small preference for the methylated DNA, MBD3 seems unable to distinguish between mC and hmC.
Figure 7. Diagram of the thermodynamic cycle used to extract the free energy variation (DDG (kcal/mol)) in nucleosome-DNA stability due to methylation of CpG steps. The calculations were performed for each different MBD protein-DNA complex, interchanging hmC and mC and computing the associated reversible work by MD/TI. Positive $\Delta G$ values mean that the complex binding to met-DNA is more favorable than to the hydroxymethylated one.

| MBD     | $\Delta G_{bind1} - \Delta G_{bind2} = \Delta G_1 - \Delta G_2 = \Delta \Delta G$ (kcal/mol) |
|---------|------------------------------------------------------------------------------------------|
| MDB1    | 1.19 ± 0.21                                                                               |
| MBD2    | 0.69 ± 0.17                                                                               |
| MeCP2   | 0.58 ± 0.25                                                                               |
| MBD3    | -0.16 ± 0.10                                                                              |

Analysis of the binding sites (Figure 8) allowed us to determine the origin of the differential specificity. The four proteins have a large sequence similarity and identical fold. They recognize d(CpG) steps (i.e. d(CpG)·d(CpG)) by means of interactions along the major groove (N7 and O6) of the two guanines. Such interactions are made by two Arg in three of the proteins (MeCP2, MBD1 and MBD2), while for MBD3 recognition is made by one Arg and one His, suggesting a lower specificity (Figure 8A). The electrostatic environment of the protein pocket where the hydroxyl group is placed is also different, in the MBD1, the protein with highest affinity for mC, the methyl group is placed in a pocket defined by residues such as VAL and PHE defining a quite hydrophobic cavity. When the methyl is hydroxylated, the OH-group is embedded in a unfavourable apolar environment (see white surface in Figure 8B left panel). On the contrary, in
MDB3 the hydroxyl group are placed in a hydrophilic/polar cavities, interacting with ASP and a SER (see red and green surface in Figure 8B). These theoretical results are supported by a myriad of experimental evidences, for example, MBD1 is known to be the protein with the largest specificity for methylated DNA, while MBD3 is known to be the most promiscuous of all MBD (38). There is also good agreement with in vitro binding estimates by Hashimoto et al. (77), which suggests the same range of mC/hmC specificity than that found here.
Figure 8. A) Detail of the Arg-recognition model for each MBD-DNA complex (see Methods for detail), arginines interacting with the guanines paired to the epigenetically modified cytosines. B) Protein surface closest to the modified cytosine (hmC), showing the interaction cavity hydroxyl.
CONCLUSIONS

The presence of hmC in both strands of a d(CpG) step leads to minor structural changes in the DNA duplex, but to non-negligible changes in physical properties, which are reflected in a general increase in the stiffness of DNA that becomes evident in MD simulations and is validated by circularization experiments. This increase in stiffness implies a reduced ability to wrap around nucleosomes, as suggested by MD simulations and demonstrated by gel reconstitution experiments. Very interestingly, both simulations and experiments suggest that hydroxymethylated DNA is even stiffer than methylated DNA, precluding that physical properties could justify the known reversion to unmethylated behavior which is coupled to the d(mCpG)→d(hmCpG) transformation catalyzed by TET. In fact, simulation and experiments suggest a small impact of d(mCpG)→d(hmCpG) transformation on chromatin conformation.

The presence of the hydroxymethyl in the major groove of changes the recognition properties of DNA and this is expected to modify binding of many effector properties, modulating then in a complex way DNA activation and repression pathways. In this paper we have centered our attention in MBD, the major elements for recognition of methylation signals. Accurate MD/TI calculations strongly suggest that d(mCpG)→d(hmCpG) transformation leads to reduction of the affinity of binding for all MBD coupled to repressive action, while binding to MBD3, the only methyl-binding domain with dual (activation/repression) activities is not affected. Simulations suggest then that a significant part of the biological effect of d(mCpG)→d(hmCpG) transformation can be related to an unbalance in the recognition of DNA by MBD proteins.

ACKNOWLEDGEMENTS

We thank Irene Gomez-Pinto of the Instituto Quimica Fisica Rocabolano. Consejo Superior de Investigaciones Cientificas (CSIC) in Madrid for her support in the NMR experiments. NMR experiments were performed in the “Manuel Rico” NMR laboratory (LMR), a node of the Spanish Large-Scale National Facility (ICTS R-LRB).

FUNDING

Spanish Ministry of Science [BFU2014-61670-EXP, BFU2017-89707-P]; Catalan SGR, the Instituto Nacional de Bioinformática; European Research Council (ERC SimDNA), European Union’s Horizon 2020 Research and Innovation Program [676556]; Biomolecular and Bioinformatics Resources Platform co-funded by the Fondo Europeo de Desarrollo Regional (FEDER) [ISICIII PT 13/0001/0030 to M.O.]; MINECO Severo Ochoa Award of Excellence (Government of Spain) (awarded to IRB Barcelona); PEDECIBA (Programa de Desarrollo de las Ciencias Básicas) (to P.D.D.);
SNI (Sistema Nacional de Investigadores, Agencia Nacional de Investigación e Innovación, Uruguay) (to P.D.D.); ICREA (Institució Catalana de Recerca i Estudis Avançats) (to M.O.). the Biomolecular and Bioinformatics Resources Platform (ISCIII PT 13/000/0030 co-funded by the Fondo Europeo de Desarrollo Regional [FEDER]) [grants Elixir-Excelerate: 676559; BioExcel2:823830 and MuG: 676566].

Conflict of interest statement. None declared.

REFERENCES

1. Garland Science - Book: Molecular Biology of the Cell + 5.
2. Jaenisch,R. and Bird,A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat. Genet., 33 Suppl, 245–54.
3. Ng,H.H. and Bird,A. (1999) DNA methylation and chromatin modification. Curr. Opin. Genet. Dev., 9, 158–63.
4. Ehrlich,M., Gama-Sosa,M.A., Huang,L.H., Midgett,R.M., Kuo,K.C., McCune,R.A. and Gehrke,C. (1982) Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res., 10, 2709–21.
5. Kass,S.U., Landsberger,N. and Wolfe,A.P. (1997) DNA methylation directs a time-dependent repression of transcription initiation. Curr. Biol., 7, 157–165.
6. Hashimshony,T., Zhang,J., Keshet,I., Bustin,M. and Cedar,H. (2003) The role of DNA methylation in setting up chromatin structure during development. Nat. Genet., 34, 187–92.
7. Miranda,T.B. and Jones,P.A. (2007) DNA methylation: the nuts and bolts of repression. J. Cell. Physiol., 213, 384–90.
8. Suzuki,M.M. and Bird,A. (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat. Rev. Genet., 9, 465–76.
9. Kulis,M., Heath,S., Bibikova,M., Queirós,A.C., Navarro,A., Clot,G., Martinez-Trillos,A., Castellano,G., Brun-Heath,I., Pinyol,M., et al. (2012) Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet., 44, 1236–1242.
10. Klose,R.J. and Bird,A.P. (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci., 31, 89–97.
11. Jones,P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet., 13, 484–92.
12. Pérez,A., Castellazzi,C.L., Battistini,F., Collinet,K., Flores,O., Deniz,O., Ruiz,M.L., Torrents,D., Eritja,R., Soler-López,M., et al. (2012) Impact of Methylation on the Physical Properties of DNA. Biophys. J., 102, 2140–2148.
13. Portella,G., Battistini,F., Orozco,M., Pérez,A. and Soler-López,M. (2013) Understanding the Connection between Epigenetic DNA Methylation and Nucleosome Positioning from Computer Simulations. PLoS Comput. Biol., 9, e1003354.
14. Branco,M.R., Ficz,G. and Reik,W. (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat. Rev. Genet., 13, 7–13.
15. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L., et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner
TET1. Science, 324, 930–5.
16. Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013) Dynamic readers for 5-(hydroxymethyl)cytosine and its oxidized derivatives. Cell, 152, 1146–59.
17. Kellinger, M.W., Song, C.-X., Chong, J., Lu, X.-Y., He, C. and Wang, D. (2012) 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat. Struct. Mol. Biol., 19, 831–3.
18. He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science, 333, 1303–7.
19. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science, 333, 1300–3.
20. Guo, J.U., Su, Y., Zhong, C., Ming, G. and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell, 145, 423–34.
21. Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeier, C. and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew. Chem. Int. Ed. Engl., 50, 7008–12.
22. Wu, H. and Zhang, Y. (2014) Tet1 and 5-hydroxymethylation. Cell Cycle, 10, 2428–2436.
23. Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W. and Walter, J. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat. Commun., 2, 241.
24. Yu, M., Hon, G.C., Szulwach, K.E., Song, C.-X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell, 149, 1368–80.
25. Song, C.-X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.-H., Zhang, W., Jian, X., et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol., 29, 68–72.
26. Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmiiller, V., Michalakis, S., Müller, M., Biel, M. and Carell, T. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew. Chem. Int. Ed. Engl., 49, 5375–7.
27. Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W. and Balasubramanian, S. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science, 336, 934–7.
28. Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science, 324, 929–30.
29. Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011) Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol. Cell, 42, 451–64.
30. Yildirim, O., Li, R., Hung, J.-H., Chen, P.B., Dong, X., Ee, L.-S., Weng, Z., Rando, O.J. and Fazzio, T.G. (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell, 147, 1498–510.
31. Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011) Genome-
wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature*, 473, 394–7.

32. Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviuciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al. (2012) 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.*, 19, 1037–43.

33. Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., et al. (2011) Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.*, 7, e1002154.

34. Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S. and Reik, W. (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*, 473, 398–402.

35. Wu, H., D’Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E. and Zhang, Y. (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.*, 25, 679–84.

36. Shen, L. and Zhang, Y. (2013) 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr. Opin. Cell Biol.*, 25, 289–96.

37. Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., et al. (2012) Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of Melanoma. *Cell*, 150, 1135–1146.

38. Du, Q., Luu, P.L., Stirzaker, C. and Clark, S.J. (2015) Methyl-CpG-binding domain proteins: Readers of the epigenome. *Epigenomics*, 7, 1051–1073.

39. Schroeder, S.J. and Turner, D.H. (2009) Optimal melting measurements of nucleic acid thermodynamics. *Methods Enzymol.*, 468, 371–387.

40. Dans, P.D., Ivani, I., Hospital, A., Portella, G., González, C. and Orozco, M. (2017) How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.*, 44, gkw1355.

41. Ivani, I., Dans, P.D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrión, P., Goñi, R., Balaceanu, A., et al. (2015) Parmbsc1: a refined force field for DNA simulations. *Nat. Methods*, 13, 55–8.

42. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratch, D.J. (2016) Gaussian 16. *Gaussian, Inc.*, Wallingford CT.

43. Cieplak, P., Cornell, W.D., Bayly, C. and 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.

44. Arnott, S. and Hukins, D.W. (1972) Optimised parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.*, 47, 1504–9.

45. Macke, T.J. and Case, D.A. (1997) Modeling Unusual Nucleic Acid Structures. In pp. 379–393.

46. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79, 926–935.

47. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A.D., R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz,
48. Pérez, A., Luque, F.J. and Orozco, M. (2007) Dynamics of B-DNA on the Microsecond Time Scale. J. Am. Chem. Soc., 129, 14739–14745.
49. Dans, P.D., Danilāne, L., Ivani, I., Dršata, T., Lankaš, F., Hospital, A., Walther, J., Pujagut, R.I., Battistini, F., Gelpí, J.L., et al. (2016) Long-timescale dynamics of the Drew–Dickerson dodecamer. Nucleic Acids Res., 44, 4052–4066.
50. Dans, P.D., Walther, J. and Gómez, H. (2016) Multiscale simulation of DNA. Curr. Opin. Struct. Biol., 37, 29–45.
51. Dang, L.X. (1995) Mechanism and Thermodynamics of Ion Selectivity in Aqueous Solutions of 18-Crown-6 Ether: A Molecular Dynamics Study. J. Am. Chem. Soc., 117, 6954–6960.
52. Dang, L.X. and Kollman, P.A. (1990) Free Energy of Association of the 18-Crown-6:K+ Complex in Water: A Molecular Dynamics Simulation. J. Am. Chem. Soc., 112, 5716–5720.
53. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. J. Chem. Phys., 98, 10089–10092.
54. Ryckaert, J.-P., Cicotti, G. and Berendsen, H.J.C. (1977) Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. J. Comput. Phys., 23, 321–341.
55. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. J. Chem. Phys., 81, 3684–3690.
56. Hospital, A., Andrio, P., Cugnasco, C., Codo, L., Becerra, Y., Dans, P.D., Battistini, F., Torres, J., Goni, R., Orozco, M., et al. (2016) BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. Nucleic Acids Res., 44, D272–D278.
57. Hospital, A., Battistini, F., Soliva, R., Gelpí, J.L. and Orozco, M. (2020) Surviving the deluge of biosimulation data. WIREs Comput. Mol. Sci., 10.
58. Gelpí, J.L., Kalko, S.G., Barril, X., Cirera, J., de La Cruz, X., Luque, F.J. and Orozco, M. (2001) Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. Proteins, 45, 428–37.
59. Cuervo, A., Dans, P.D., Carrascosa, J.L., Orozco, M., Gomila, G. and Fumagalli, L. (2014) Direct measurement of the dielectric polarization properties of DNA. Proc. Natl. Acad. Sci. U. S. A., 111, E3624-30.
60. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: Visual molecular dynamics. J. Mol. Graph., 14, 33–38.
61. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera – A visualization system for exploratory research and analysis. J. Comput. Chem., 25, 1605–1612.
62. Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res., 37, 5917–29.
63. Lankas, F., Sponer, J., Langowski, J. and Cheatham, T.E. (2003) DNA basepair step deformability inferred from molecular dynamics simulations. Biophys. J., 85,
64. Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. (1998) DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 11163–8.

65. Lavery, R., Maddocks, J.H., Pasi, M. and Zakrzewska, K. (2014) Analyzing ion distributions around DNA. *Nucleic Acids Res.*, **42**, 8138–8149.

66. Thalhammer, A., Hansen, A.S., El-Sagheer, A.H., Brown, T. and Schofield, C.J. (2011) Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. *Chem. Commun. (Camb)*, **47**, 5325–7.

67. Dans, P.D., Faustino, I., Battistini, F., Zakrzewska, K., Lavery, R. and Orozco, M. (2015) Unraveling the sequence-dependent polymorphic behavior of d (CpG) steps in B-DNA. *Nucleic Acids Res.*, **42**, 11304–11320.

68. Kielkopf, C.L., Ding, S., Kuhn, P. and Rees, D.C. (2000) Conformational flexibility of B-DNA at 0.74 Å resolution: d(CCAGTACTGG). *J. Mol. Biol.*, **296**, 787–801.

69. Maehigashi, T., Hsiao, C., Kruger Woods, K., Moulaei, T., Hud, N. V. and Williams, L.D. (2012) B-DNA structure is intrinsically polymorphic: Even at the level of base pair positions. *Nucleic Acids Res.*, **40**, 3714–3722.

70. Dans, P.D., Pérez, A., Faustino, I., Lavery, R. and Orozco, M. (2012) Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res.*, **40**, 10668–78.

71. Liebl, K. and Zacharias, M. (2019) How methyl-sugar interactions determine DNA structure and flexibility. *Nucleic Acids Res.*, **47**, 1132–1140.

72. Chen, K., Meng, Q., Ma, L., Liu, Q., Tang, P., Chiu, C., Hu, S. and Yu, J. (2008) A novel DNA sequence periodicity decodes nucleosome positioning. *Nucleic Acids Res.*, **36**, 6228–6236.

73. Lowary, P.T. and Widom, J. (1997) Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 1183–1188.

74. Zhou, Z., Hong, E.J., Cohen, S., Zhao, W. ning, Ho, H. yi H., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., et al. (2006) Brain-Specific Phosphorylation of MeCP2 Regulates Activity-Dependent Bdnf Transcription, Dendritic Growth, and Spine Maturation. *Neuron*, **52**, 255–269.

75. Cramer, J.M., Scarsdale, J.N., Walalvarkar, N.M., Buchwald, W.A., Ginder, G.D. and Williams, D.C. (2014) Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. *J. Biol. Chem.*, **289**, 1294–1302.

76. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A. and Bird, A. (2001) Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.*, **15**, 710–723.

77. Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X. and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.*, **40**, 4841–4849.