RESEARCH ARTICLE

The Impact of the HydroxyMethylCytosine epigenetic signature on DNA structure and function

Federica Battistini[1], Pablo D. Dans[1,2,3], Montserrat Terrazas[1], Chiara L. Castellazzi[1], Guillem Portella[1,4], Mireia Labrador[1], Núria Villegas[1], Isabelle Brun-Heath[1], Carlos González[5], Modesto Orozco[1,6]*

1 Institute for Research in Biomedicine (IRB Barcelona). The Barcelona Institute of Science and Technology, Barcelona, Spain, 2 Department of Biological Sciences, CENUR Litoral Norte, Universidad de la República (UdelaR), Salto, Uruguay, 3 Functional Genomics Lab., Institut Pasteur of Montevideo, Montevideo, Uruguay, 4 Chemistry Department, University of Cambridge, Cambridge, United Kingdom, 5 Instituto Química Física Rocasolano, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain, 6 Department of Biochemistry and Molecular Biology, University of Barcelona, Barcelona, Spain

* 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 similar to those of 5-methylcytosine. Experiments and simulations demonstrate that 5-methylcytosine (mC) and 5-hydroxymethylcytosine (hmC) generally lead to stiffer DNA than normal cytosine, 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). Conversely, our simulations suggest that methylated-DNA binding domains (MBDs), associated with repression activities, are 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 gene activity changes 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.

Author summary

In Eukaryotic cells, DNA epigenetic modifications play an important role in gene expression and regulation, and protein recognition. In this work we investigate the physical implications of cytosine 5-hydroxymethylation on DNA, its structural and flexibility

differences with methylated and unmodified cytosine using molecular dynamics, biophysical experiments and NMR spectroscopy.

In particular the effect of hydroxyl group on free energy of nucleosome and Methyl binding Protein (MBD) binding, comparing in silico and experimental data to shed light on the effect of the reduced flexibility and the direct protein-DNA recognition.

1. 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 another covalently modified cytosine, the 5-hydroxymethylcytosine (hmC) [14,15]. This modified nucleobase was initially considered just an intermediate step in the de-methylation pathway (5-Methyl➔ hydroxymethyl➔ formyl➔ carboxyl➔ CO_2) [16,17], but different studies on the expression of the protein catalyzing mC➔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].

In the past years approaches to detect and differentiate mC and hmC in vitro and in vivo along the genome at base-resolution have been developed [29,30].

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 [31–35]. When present, hmC signals are not randomly distributed, but appear located near transcriptional starting sites and promoters, as well as in certain CpG islands [24,36–38], suggesting a specific role of this epigenetic modification in gene activity. The presence of hmC seems to correlate with a reversion of mC effects [39], 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. However, there is still a controversy on the physical/conformational effect of mC and hmC on DNA flexibility and its ability to wrap around histones to form nucleosomes [12,40–45]. Using a variety of experimental and

theoretical methods, here we explored this issue to have a complete view on all the sequence-dependent effects within the nearest-neighbor framework. We found that changes in flexibility induced by cytosine modifications strongly depend on the sequence of the tetrameric environments, considering CpG as central step. In general, we discovered that physical properties of hmC are significantly different to those of unmodified cytosine, but similar to those of mC. This suggests that the reversion of methylation effects in gene expression by TET action cannot be explained simply by changes 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 [16,46] 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.

2. Results and discussion

2.1 The effect of 5-hydroxymethylcitosine 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 Materials and Methods and Table A in S1 File) 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. This different becomes significative at higher concentrations while still inside the experimental error at lower concentrations, suggesting a mild effect in stability. 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, [47]) the stability of mC (mC >hmC>C, see Table B in S1 File). 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, [47]) the stability of mC (mC >hmC>C, see Table C in S1 File).

2.2 The effect of 5-hydroxymethylcitosine 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, [48]) and d(CGCGAC*GTCGCG)·d(CGCGAC*GTCGCG) (central base pair very stiff, [48]). 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 Materials and Methods and S1 File). 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 Tables C-H and Figs A-D in S1 File). 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 Materials and Methods and S1 File) for the mC and hmC containing duplexes in the two distinct tetrameric environments. In all cases structures agree very well with a general B-
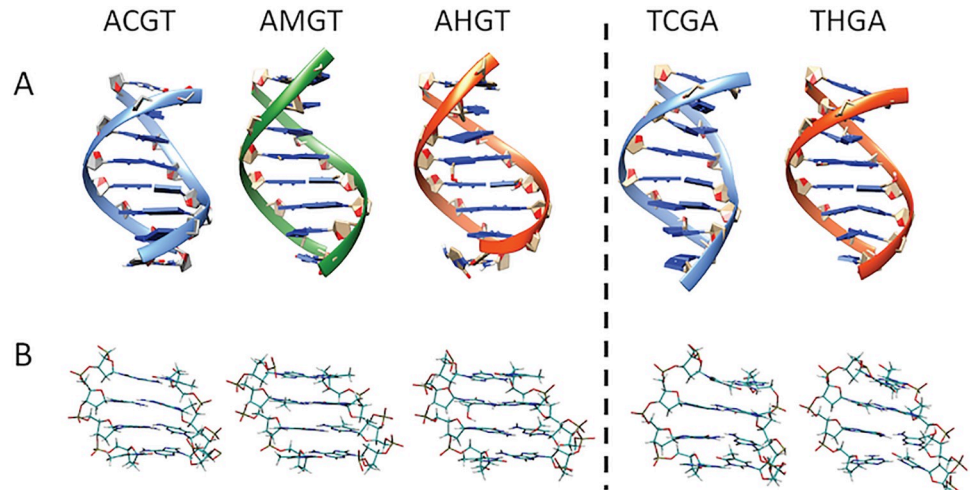
Fig 1. Average NMR structure of the sequences d(CGTC*GACG)·d(CGTC*GACG) and d(CGCGAC*GTCGCG)·d(CGCGAC*GTCGCG) (C* being C, mC or hmC respectively) for the central octamer (A) and the central tetramer (B).

https://doi.org/10.1371/journal.pcbi.1009547.g001

like duplex (Fig 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.

2.3 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

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).

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

https://doi.org/10.1371/journal.pcbi.1009547.t001

C*pG step in Fig E in S1 File), supporting the quality of parmbsc1 simulations [49]. 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 restrains 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 [48,50–52]. With this aim we collected 0.5 μs trajectories (see Materials and 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 (Table I in S1 File). 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 (Fig 2).

In the case of twist, the epigenetic changes induce (in general) a displacement of the bimodal CpG distribution towards low twist values (Fig 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 (Fig 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
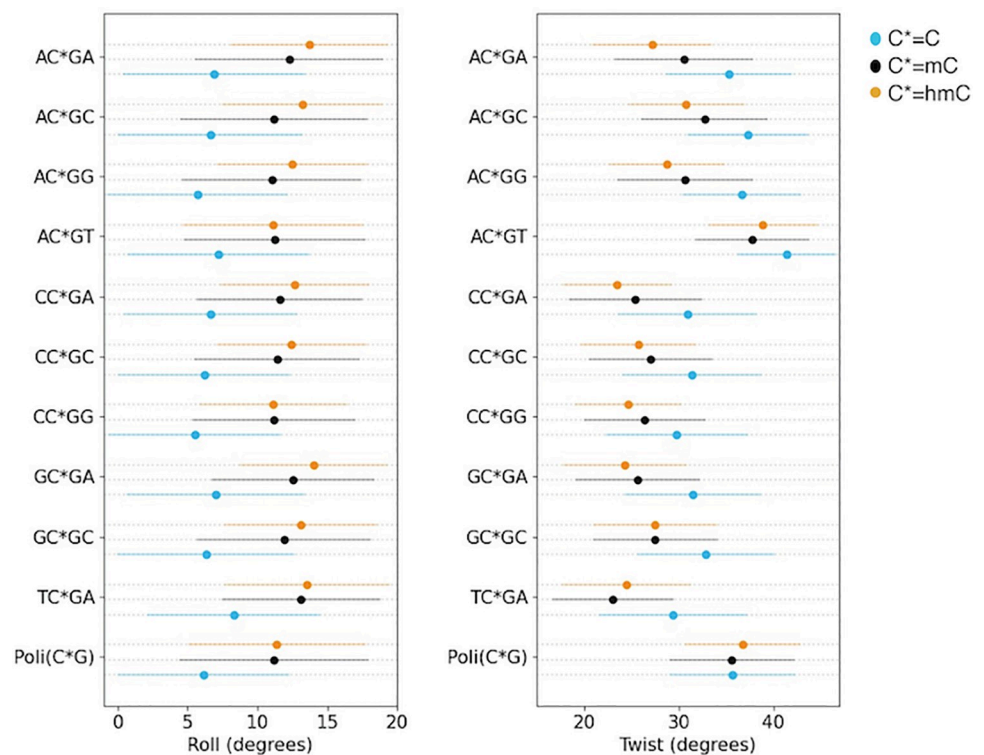
Fig 2. Average value with relative standard deviation for the base pair parameters roll (left panel) and twist (right) for C*G step in the different tetramer environment. Being C* normal cytosine (blue dots), methylated cytosine (green dots) and hydroxymethylated cytosine (red dots).

https://doi.org/10.1371/journal.pcbi.1009547.g002
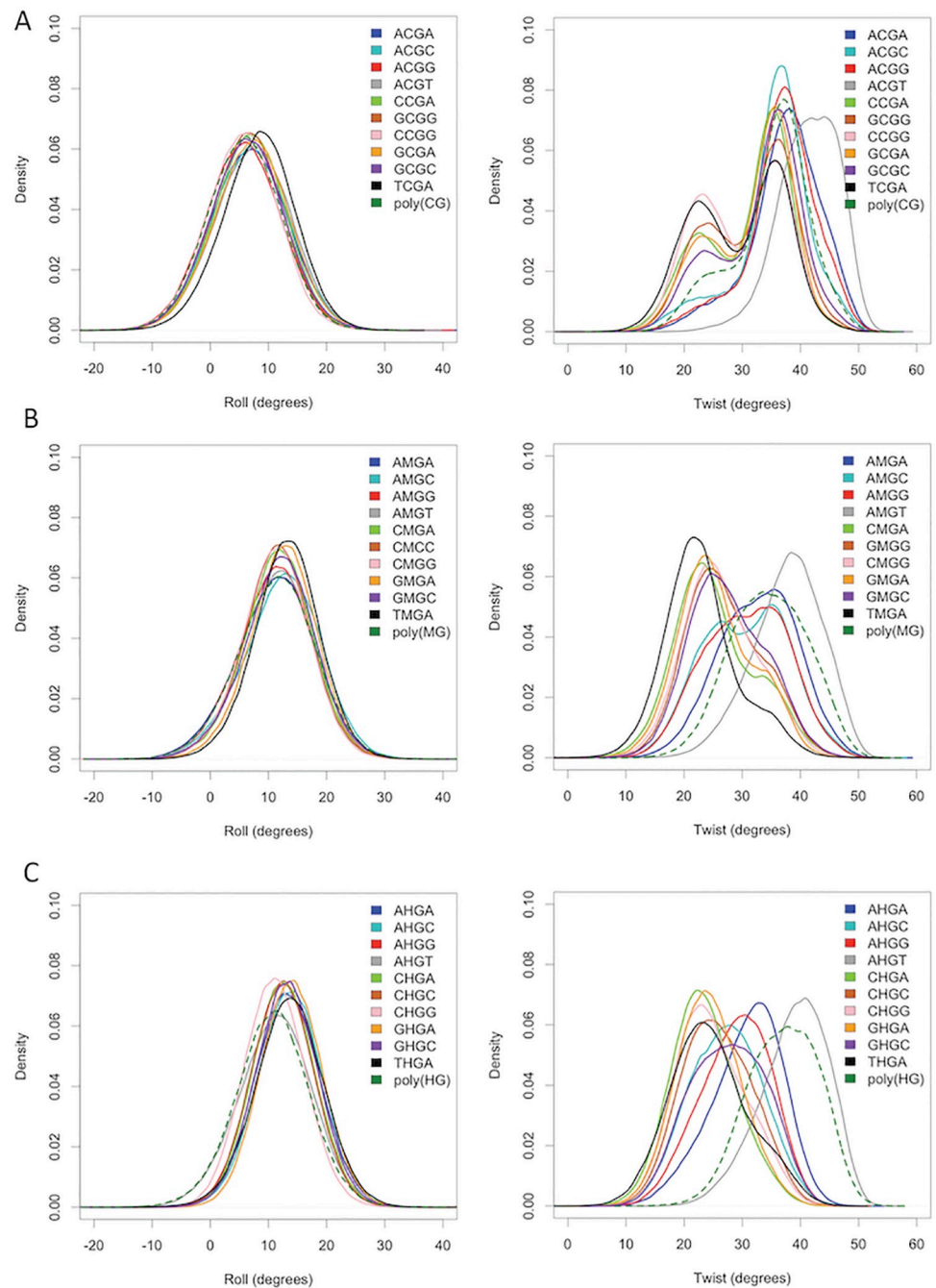
Fig 3. Roll and Twist distributions. Roll (left panel) and twist (right panel) distributions for each C*pG step, in a different tetramer environment, being C* normal (A), methylated (B) and hydroxymethylated (C). Dashed line for C*pG step in a poly(CpG) sequence.

https://doi.org/10.1371/journal.pcbi.1009547.g003

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 Fig F in S1 File), 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).

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. Full dataset of the stiffness matrices in S2 File.

A

	K^b_{tw-tw}			$K^b_{roll-roll}$			$K^b_{tilt-tilt}$		
	CpG	mCpG	hmCpG	CpG	mCpG	hmCpG	CpG	mCpG	hmCpG
ACGA	0.025	0.024	0.025	0.016	0.018	0.025	0.033	0.031	0.037
ACGC	0.023	0.024	0.026	0.014	0.015	0.023	0.034	0.031	0.039
ACGG	0.025	0.024	0.026	0.015	0.016	0.027	0.033	0.031	0.039
ACGT	0.029	0.034	0.027	0.015	0.021	0.016	0.036	0.038	0.035
CCGA	0.027	0.027	0.028	0.019	0.022	0.028	0.033	0.032	0.038
CCGC	0.025	0.026	0.026	0.018	0.021	0.027	0.033	0.032	0.041
CCGG	0.027	0.027	0.026	0.018	0.022	0.028	0.031	0.031	0.039
GCGA	0.027	0.028	0.029	0.018	0.022	0.029	0.034	0.033	0.039
GCGC	0.024	0.026	0.024	0.016	0.018	0.025	0.034	0.032	0.041
TCGA	0.025	0.027	0.025	0.020	0.025	0.024	0.032	0.033	0.032
average	0.026	0.027	0.026	0.017	0.020	0.025	0.033	0.032	0.038
Poly(CG)	0.022	0.024	0.027	0.018	0.015	0.018	0.033	0.030	0.030

B

	K^{tot}_{rot}		
	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
GCGA	0.025	0.027	0.032
GCGC	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

https://doi.org/10.1371/journal.pcbi.1009547.t002

Stiffness analysis (see Materials and Methods) [53,54] 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 diagonal terms of the stiffness matrices in Table 2 and Table J in S1 File, and data in S3 File for the entire stiffness matrices), due the narrowing of the roll distribution (see Fig 3), probably related to the formation of transient hydrogen bond interactions involving the hydroxyl group of hmC (see Fig G in S1 File as examples). The formation of those hydrogen bonds covers only a very small part of the trajectory, consequently they cannot justify completely the stiffness due to the hydroxyl group but combined with the loss of conformational sub-states these results show a narrower sampling of the conformations and restrictions for some movements due to the hydroxyl group interactions. 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 (in agreement with previous findings by Zacharias and coworkers [44]), 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.

We also investigated the stiffness at the base pair level and calculated the 6x6 stiffness matrices for the intra-base pair movements (translational: Shear, Stretch, Stagger, and rotational: Buckle, Propeller, Opening) (data in S3 File). We found that the translational movements have in general higher force constants for normal cytosine (C>hmC>mC), while the rotational parameters are stiffer when hmC is present and the unmodified cytosine is the most flexible. Also in this case the sequence-dependent variability looks more important that the effect of the epigenetic change.

We calculated the correlations between the neighboring base pair (j and j+1), starting from the central C*G step. As previously reported [55] for unmodified CpG steps we found correlations of shift-shift; twist-twist and tilt-tilt base pair parameters between neighboring steps. However, this correlation going from the base pair C*G to the neighboring step decreases (see Fig H in S1 File) and does not go further than the j+1. The epigenetic modifications clearly diminish the neighboring correlations (Fig 3), as they simplify the conformational landscape by reducing the diversity of possible sub-states of several degrees of freedom.

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(GAAAAAACGGGC*GAAAAACGG)· d(TCCCGTTTTTC*GC CCGTTTTT)) with a central C*pG dinucleotide (C* = C, mC, hmC, respectively) as described in Perez *et al* [12]. Under favorable ligation conditions (see Materials and Methods), the multimers form circles that are as short as allowed by the flexibility of the DNA. Results shown in Fig 4 (and Fig I in S1 File) 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.
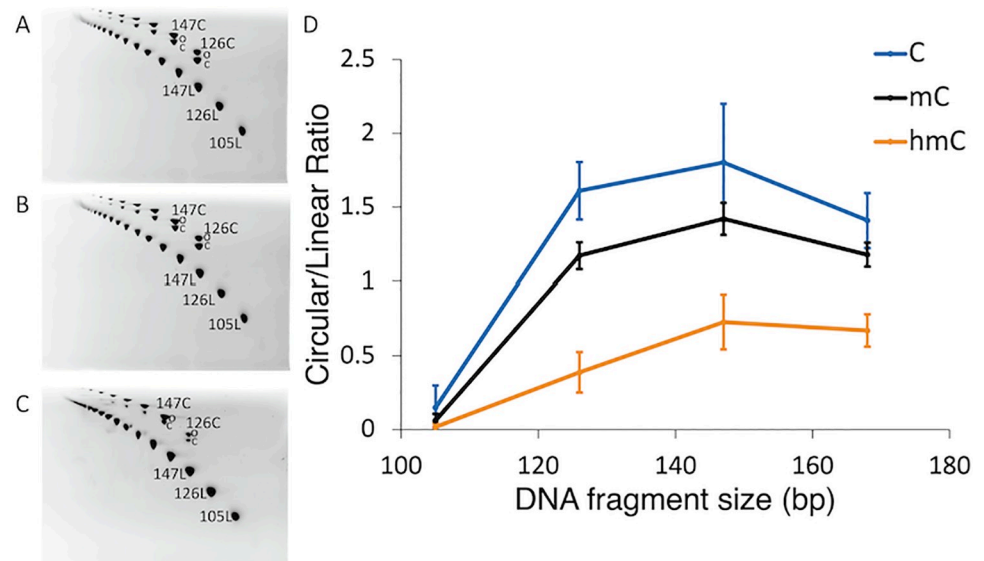
Fig 4. (A-C) 2D polyacrylamide native gels showing different migrations of linear and circular DNA species oligomers of 21 bp, respectively for (A) Cytosine (C), (B) Methylcytosine (mC) and (C) Hydroxymethylcytosine (hmC) containing fragments. Linear DNA molecules are positioned on the lower diagonal, 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).

https://doi.org/10.1371/journal.pcbi.1009547.g004

Our experimental and theoretical results can be compared with previous data from the literature. Thus, the impact of methylation reducing DNA flexibility found here agrees well with a myriad of experiments from our own group and others [12,40,41,43–45], but disagrees with the results in a recent study by Zaichuk and Marko [42] that point in the opposite direction, probably due to a different experimental set-up where no full control of the placement and level of methylation exists. The impact of hydroxymethylation is less clear, and our results disagree with previous claims by Ngo et al. [41] derived from circularization experiments using short oligomers. We can speculate that the differences might arise from the use in those experiments [41] of a short oligomer that cannot circularize by harmonic deformations (condition in which stiffness analysis can be applied), and possible phasing problems which can be prevalent when experiments are done using only a single duplex length.

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 Materials and 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 [56]) and ii) 4 epigenetic changes in "anti-phase", i.e. each modification separated by 5 bps. Results in Fig 5A demonstrate that epigenetic variations increase the energy cost required for nucleosome wrapping, both when introduced in "in-phase" and "anti-phase" (Fig J in S1 File), 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 Materials and Methods, Supporting Methods in S1 File) [57] using both "in-phase" and "anti-phase" arrangements (the same sequences as before). The results obtained (Fig 5B), characterized by high standard deviation but statistically significant (pvalue of 0.006), confirm that the C➔mC➔hmC substitution leads to a reduction in the efficiency of nucleosome reconstitution, the difference
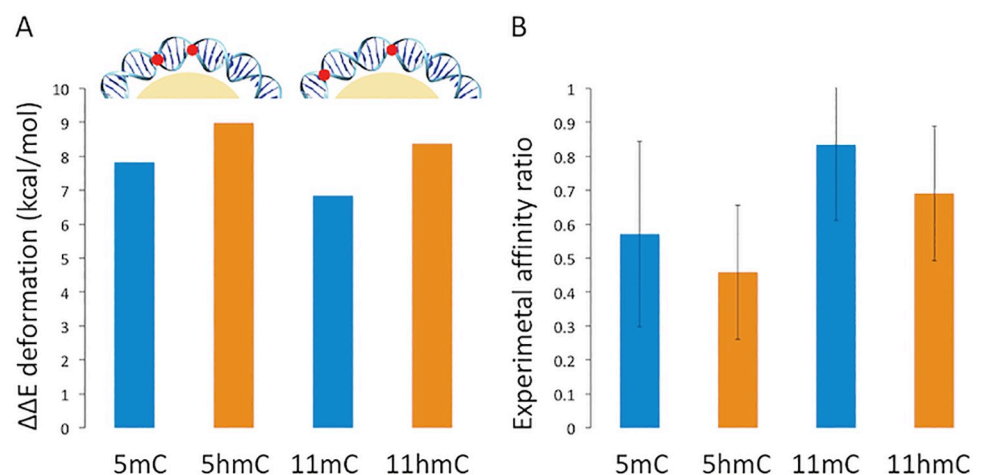
Fig 5. A) Deformation energy variation cost required for nucleosome wrapping, difference between the energy cost of the methylated (mC) and hydroxymethylated (hmC) DNA respect to the control (unmodified cytosine). At the top a schematic view of the positioning of the modifications (red dots), respectively 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.

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. To our knowledge, this is the first time that hmC impact on nucleosome binding has been analyzed, but our findings on methylation fully agree with previous experimental data [12,40,41,43,44] including recent one with cellular models [45], which inspires confidence on the reliability of our hmC predictions.

Altogether, theoretical and experimental results strongly suggest that C➔mC➔hmC substitution at d(CpG) steps leads 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.

2.4 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 (Fig 6) (see Materials and Methods) [58]. 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 Fig K in S1 File), probably because of the presence of a polar

ACGT   AMGT   AHGT
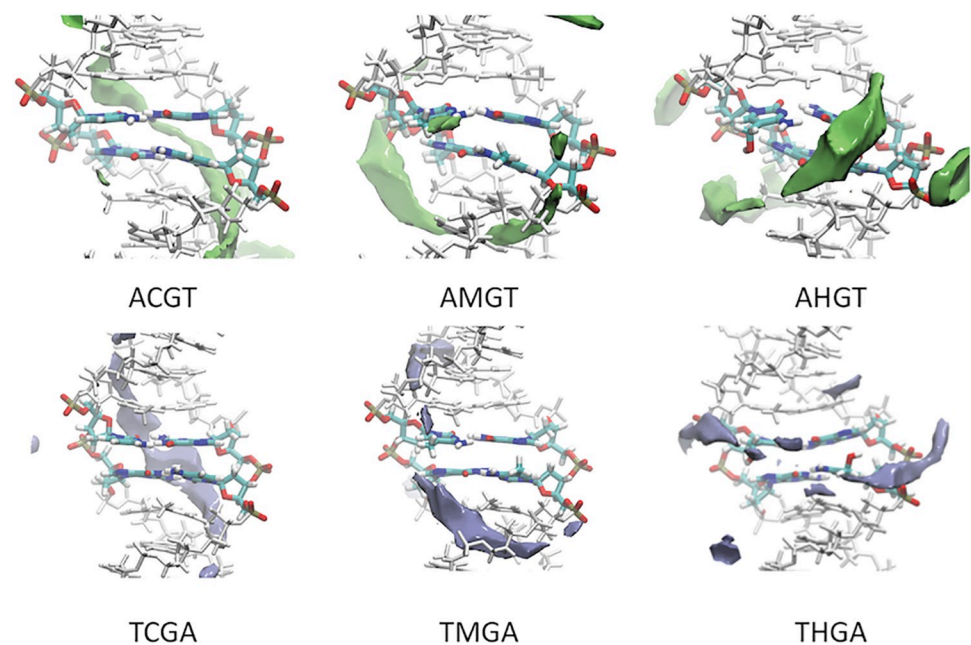
TCGA   TMGA   THGA

Fig 6. Examples of classical molecular interaction potentials using sodium as a probe [82]. 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.

https://doi.org/10.1371/journal.pcbi.1009547.g006

group. The polarity of the CH_2OH group is also visible in the water distributions (see examples in Fig L in S1 File), 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.

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 [3,46,59], ii) MBD1 which represses directly (via TRD) or indirectly (via recruitment of histone methylases) gene expression [46], 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 [46,60,61]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.[32,46,61].

We built models of the four MBD-DNA complexes from the available experimental data (details in Materials and Methods and Supporting Methods S1 File), we used as starting structures well-solved complexes and with a clear definition of the interaction site between the modified cytosine and the protein, without the need to sample the positioning [62]. The complexes were solvated and used as starting conformations for MD simulations, coupled to thermodynamic integration (MD/TI), which combined with standard thermodynamic cycles (Fig 7) allowed us to determine the impact in binding of methyl-cytosine to hydroxymethyl-cytosine change in protein binding (see Materials and 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.

Analysis of the binding sites (Fig 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 (Fig 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 Fig 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 Fig 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 [46]. There is also good agreement with in vitro binding estimates by Hashimoto et al. [63], which suggests the same range of mC/hmC specificity than that found here.

In summary, MD/TI simulations strongly suggest that mC➜hmC introduces a general decrease in binding affinity in MBD proteins with repressive activities, reversing then the inactivating signals associated with cytosine methylation. On the contrary, MBD3, which shows that largest sequence variability at the recognition site has no preference between the two epigenetic variants. This suggests that MBD3-associated signaling is not affected, which might yield to specific activation or deactivation of genes. The final consequences of the presence of d(hmCpG) steps should be then not only a general activation of gene activity (reversing

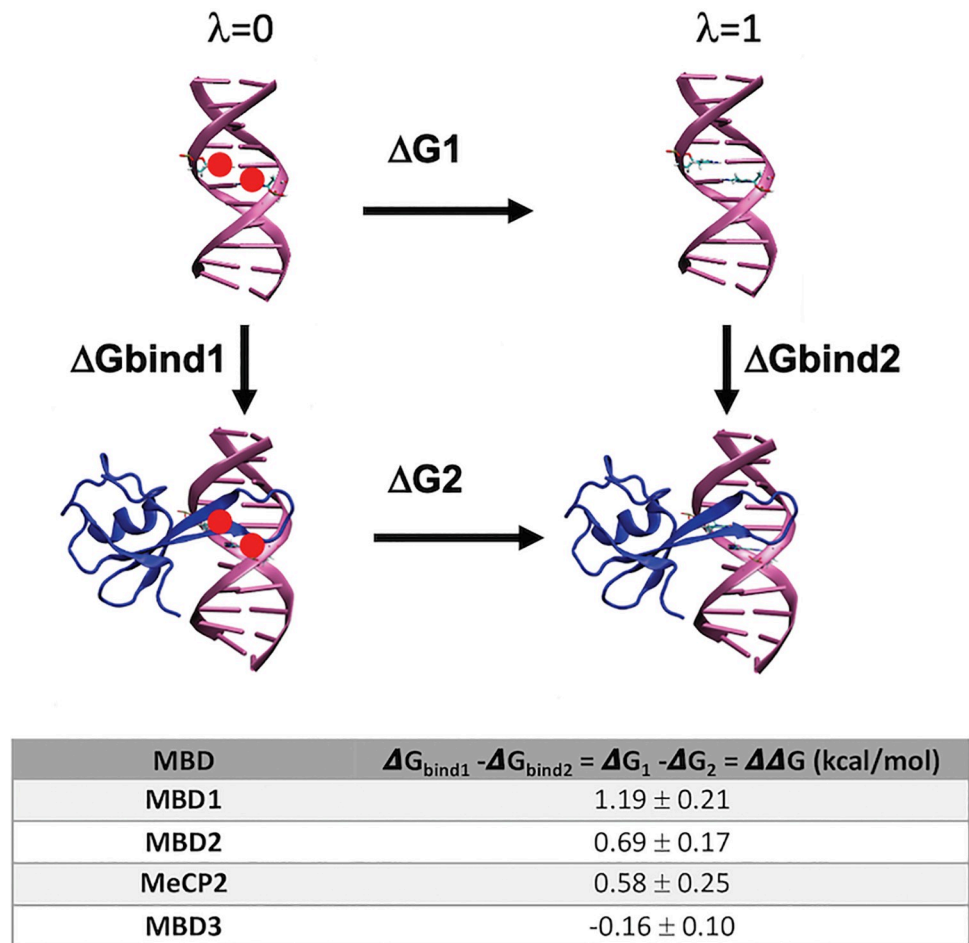| MBD	$\Delta G_{bind1} - \Delta G_{bind2} = \Delta G_1 - \Delta G_2 = \Delta\Delta G$ (kcal/mol)
MBD1	1.19 ± 0.21
MBD2	0.69 ± 0.17
MeCP2	0.58 ± 0.25
MBD3	-0.16 ± 0.10

Fig 7. Diagram of the thermodynamic cycle used to extract the free energy variation ($\Delta\Delta G$ (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 $\Delta\Delta G$ values mean that the complex binding to met-DNA is more favorable than to the hydroxymethylated one. As a control of the quality of our theoretical estimates, free energy calculations were done also for the transition from mC to C. The results suggest a strong destabilization of MDB1 binding (2.2±0.4 kcal/mol), a moderate destabilization of MBD2 and MeCP2 binding (0.8±0.6 and 0.5±0.4) and no appreciable effect on MBD3 binding (0 ±0.6), values which agree well with in vitro experimental estimates [63].

methylation signal), but a qualitative gain in the complex and promiscuous MBD3 associated signaling.

Experiments and calculations suggest that 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, our results suggest a small impact of d(mCpG)➜d(hmCpG) transformation on chromatin conformation.
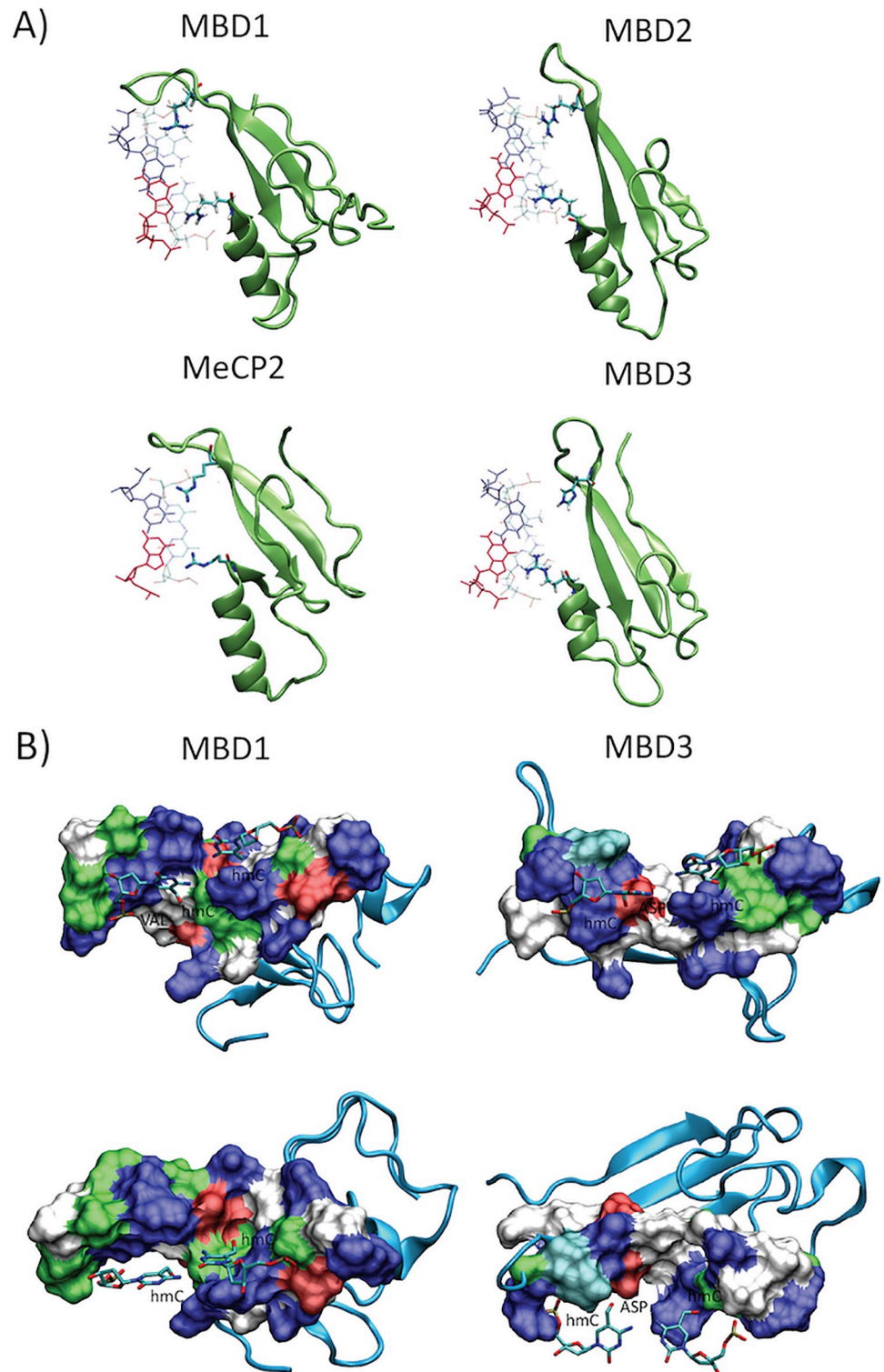
Fig 8. A) Detail of the Arg-recognition model for each MBD-DNA complex (see Methods for detail), arginines interacting with the guanines paired to the epigenetically modified cytosines. B) Protein surface closest to the modified cytosine (hmC), showing the interaction cavity hydroxyl group for the two extreme examples MDB1 (left panel) and MDB3 (right panel).

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.

3 Materials and methods

3.1 UV-monitored thermal-denaturation studies

Modified and unmodified 8-nucleotide DNAs (CGAC*GTCG) were synthetized (details in Supporting Methods and Table K in S1 File) 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 1mm 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. ΔH, ΔS, and ΔG values were estimated from dependence of melting temperatures on DNA concentrations [64]. The reciprocal values of average melting temperatures were plotted against ln (C_t/4) and fitted to linear relationships, $1/T_m = (R/\Delta H) \ln(CT/4) + (\Delta S/\Delta H)$.

3.2 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 Supporting Methods in S1 File) were dissolved in either D_2O or 9:1 H_2O/D_2O 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_2O at 25˚C. In the experiments in D_2O, presaturation was used to suppress the residual H_2O signal. The NOESY spectra in D_2O 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_2O were acquired with 150 ms mixing time at 5˚C to reduce exchange with the water. In 2D experiments in H_2O, 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 NOESY cross-peaks and quantitative evaluation of the NOE intensities. Additional details of the NMR analysis are given in S1 File. Coordinates and NMR restraints have been deposited in the Protein Data Bank, PDB IDs 7OGV, 7OHE, 7OHJ and 7OHM.

3.3 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 [65]. 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.

3.4 HydroxyMethylCytosine parameterization

Geometrical parameters for hmC were derived starting from parmbsc1 [49] cytosine and serine from ff14SB for hydroxymethyl group. Final hmC geometry was obtained by energy minimization [66]. 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 [67]. Library with parameters for hmC are available upon request.

3.5 Molecular dynamics simulations

We performed molecular dynamics (MD) simulations for 33 12-mer DNA duplexes of sequence CGCGXC*GYCGCG, 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 [68], and oligomers were built using the Nucleic Acid Builder [69], 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 [70], 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 [71] 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 [72–74] 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 [49] and Dang parameters for ions [75,76]. Long-range electrostatic effects were treated using the Particle Mesh Ewald method [77] using a cutoff of 10 Å. The bonds involving hydrogen were restrained using SHAKE [78] and The temperature and the pressure, with a coupling constant of 5 ps were controlled using Berendsen algorithm [79]. Trajectories, data, analyses and parameters for methylated and hydroxymethylated cytosines are available in our BigNAsim database [80], following recent recommendations for FAIR trajectory storage [81].

3.6 Analysis of the trajectories

All the trajectories were processed with the cpptraj module of the AmberTools 18 package [71]. The ability of DNA to recognize sodium was analysed using our classical molecular interaction potential (cMIP [82]). 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 [83]. The molecular plots were generated using either VMD 1.9 [84] or the UCSF Chimera package version 1.8.1 [85].

3.7 Analysis of DNA physical properties and Deformation Energy calculation

DNA helical parameters and backbone torsion angles were measured with the Curves+ and Canal programs [86] 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 [53,54]. To be noted that our method is a simplified model, where we considered a reduce stiffness matrix and that our deformation energy model does not consider explicitly the nearest neighbor correlations but only the tetrameric effect on the central CpG step. 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 [58]. 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).

3.8 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).

3.9 *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 (methods details in S1 File). The 3 forms of the high affinity 601.2 DNA sequence (sequences in S1 File), 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 Supporting Methods in S1 File for details and Table K in S1 File). 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) (Fig M in S1 File). 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.

3.10 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 (Supporting Methods in S1 File). 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. $d\lambda = 0.05$, 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 $\alpha = 0.3$, $\sigma = 0.25$ 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 estimates for $\langle(\delta H)/\delta\lambda\rangle_\lambda$ by using blocks of 20 ns, which were then integrated through the entire mutation pathway to obtain mutation free energies (with associated statistical errors). All the sequences used in this work, for computational and/or experimental analyses are listed in Supporting Methods in S1 File.

Supporting information

S1 File. This file contains Supporting Methods and Tables A-K and Figs A-M. Table A. Melting Temperatures (Tm in ˚C) for the oligonucleotide CGAC*GTCG, where C* stands for cytosine (C), methylated-cytosine (mC) and hydroxymethylated-cytosine (hmC) respectively, calculated by UV experiments at different oligo concentrations. **Table B.** Variation in thermodynamic parameters, enthalpy (ΔH in Kcal/mol), entropy (ΔS in cal/mol*K) and free energy ΔG (Kcal/mol), at 25˚C calculated using Van't Hoff equation (see Methods) from UV data for the oligomer in Table K with cytosine, methylcytosine and hydroxymethycytosine respectively. **Table C.** Assignment of the proton resonances of the hMC duplex (CGAhCGTCG)2, where hC = 5-hydroxymethylcytosine. Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, T = 5˚C, pH 7. **Table D.** Assignment of the proton resonances of the 5MC duplex (CGAmCGTCG)2, where mC = 5-methylcytosine. Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, T = 5˚C, pH 7. **Table E.** Assignment of the proton resonances of the control duplex (CGACGTCG)2. Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, T = 5˚C, pH 7. **Table F.** Assignment of the proton resonances of the hMC duplex d

(CGCGAhCGTCGCG)2. Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, T = 5˚C, pH 7. **Table G.** Assignment of the proton resonances of the control duplex d (CGCGACGTCGCG)2. Buffer conditions: 25 mM sodium phosphate, 100 mM NaCl, T = 5˚C, pH 7. **Table H.** NMR restraints and structural calculation statistics. **Table I.** Average parameters (in Å and Degrees) averaged over the last 200 ns for the central step (d(C*pG)·d(C*pG)) in the different tetrameric environments between the different forms of cytosine, HydroxyMethylC, MethylC, Cytosine. **Table J.** Diagonal stiffness constants for translational movements in kcal/mol ang2 for the central C*pG step (C* = C, mC and hmC) in the different tetrameric environments. **Table K.** Mass spectrometry analysis of synthesized oligonucleotides. **Fig A.** Regions of NOESY spectra (150 ms mixing time) of hMC (CGA*CGTCG)2 duplexes, *C = hMC (top), 5MC duplex (middle) and C (control) duplex (bottom). H1'-base assignment pathways are indicated. Buffer conditions: 100 mM NaCl, 25 mM sodium phosphate, T = 5˚C, pH 7. **Fig B.** Imino region of the NOESY spectra in H2O (τm = 150 ms) of (CGA*CGTCG)2 duplexes, *C = hMC (right), 5MC (middle), and C (control) duplex (left). Buffer conditions: 100 mM NaCl, 25mM sodium phosphate, T = 5˚C, pH 7. **Fig C.** 1H-31P correlation spectra for hMC (right), 5MC (middle), and control duplex (left). Buffer conditions were 100 mM NaCl, 25mM sodium phosphate, T = 5˚C, pH 7. **Fig D.** Chemical shifts differences for non-exchangeable protons between hMC (top) and 5MC (bottom) with respect to the control duplex. **Fig E.** Overlap of the central tetramer of the average NMR structures (light blue) with the average structure from MD (green), for the tetramers ACGT, AMGT, AHGT, TCGA and THGA. RMDs calculated between the structures for heavy atoms 0.95, 0.69, 0.98, 1.1, 0.76 Å respectively. **Fig F.** Twist profile for the hemi-hydroxymethylated sequence in the X-ray crystal structures 4GLH (red line), 4HLI (green line) and 4GLC (blue line). HG/CG steps are characterised by low-twist state. **Fig G.** Hydrogen bonds detected along the MD simulations among hydroxy group of hmC (OH, hmC position 6) and the flanking bases (guanine 7, adenine 5). We detected the formation of hydrogen bonds between the hydrogen of the hydroxyl group and the neighbouring guanine (G7); in particular between the hydrogen of the hydroxyl group and the nitrogen 7 (panel A) OHN7-G7O6-G7OHAHGT A(N7)—hmC(OH) N7-A5OHABCDH42-hmCOH or/and the oxygen 6 (panel B) of the guanine. We also detected a HB between the hydroxyl hydrogen and the oxygen of the backbone (panel C). It is worth noticing that these hydrogen bonds are formed maximum the 0.039 fraction of the simulation time and they are not stable. Nevertheless, the interaction between hmC and G7 indicates that the base pair can assume a more distorted conformation, that reflects into an opening of the minor groove, high roll. In the AC*GT tetramer (panel D) the hydroxyl group interacts also with the adenine (A6), and not only with the following guanine. This behaviour translates into the stiffer and higher twist of the tetramers XCGY where X is a purine and Y a pyrimidine base. **Fig H.** Correlations between the neighboring base pair (j-> j+1), starting from the central C*G (j = 6), averaged over all the dodecamers simulated. We found correlation, as previously reported [1], between twist-twist, tilt-tilt, shift-shift (left to right panels) base pair parameters. The correlation going from the central base pair C*G to the neighboring steps decreases and does not go further than the j+1 (step = 7). **Fig I.** Replicas of the 2D polyacrylamide native gels showing different migrations of linear and circular DNA species oligomers of 21 bp, respectively for Cytosine (C), Methylcytosine (mC) and Hydroxymethylcytosine (hmC) containing fragments. Linear DNA molecules are positioned on the lower diagonal, and circular DNA molecules are positioned on the upper diagonal (see Fig 4). **Fig J.** Periodicity of the variation in deformation energy given by the positioning of the epigenetic modifications respect to the histones in the nucleosome. In the top panel, in blue (light blue mC and dark blue hmC) the variation in energy when the modifications are 5 base pairs apart. IIn the bottom panel, energy variations (light red mC and dark red hmC) when the modifications are 11

base pairs apart. In both plots a schematic image of the DNA modifications (red dots) when they are positioned 5bps or 11 bps apart, in the minor or major grooves facing the histones (in yellow). **Fig K.** Cation K+ concentration (molarity) along the minor and major grooves averaged over the last 200 ns of the trajectories. K+ molarity distribution as a function of the angular dependence (in degrees) for the tetramers AC*GT and TC*GA where C* = C,mC,hmC (in blue, green and red respectively). **Fig L**. Occupancy maps of water molecules in the major and minor groove for the unmodified CpG, mCpG and hmCpG respectively in the two tetramer AC*GG and TC*GA. **Fig M.** In vitro nucleosome core particle reconstitution. Results of the four replicas of the gel mobility shift assays of nucleosomes reconstituted in vitro with a 147-bp 601 (with normal cytosines, methylated and hydroxylated respectively). The upper bands (Nucleosome) correspond to histone core-bound DNA, and lower bands correspond to unbound DNA (free DNA). Mk: DNA ladder for size band estimation.
(PDF)

S2 File. This file contains all the force constants for each base parameter and each possible trimer with C*G base pair (C* = hmC, mC or C respectively).
(XLSX)

S3 File. This file contains all the force constants for each base pair parameter and each possible tetramer with central C*G base pair (C* = hmC, mC or C respectively).
(TXT)

Acknowledgments

Author Contributions

Conceptualization: Federica Battistini, Pablo D. Dans, Modesto Orozco.

Data curation: Federica Battistini, Pablo D. Dans, Guillem Portella, Mireia Labrador, Isabelle Brun-Heath, Carlos González, Modesto Orozco.

Formal analysis: Pablo D. Dans, Montserrat Terrazas, Chiara L. Castellazzi, Guillem Portella, Mireia Labrador, Isabelle Brun-Heath, Carlos González, Modesto Orozco.

Funding acquisition: Modesto Orozco.

Investigation: Federica Battistini, Modesto Orozco.

Methodology: Federica Battistini, Pablo D. Dans, Montserrat Terrazas, Guillem Portella, Mireia Labrador, Núria Villegas, Isabelle Brun-Heath, Carlos González.

Supervision: Modesto Orozco.

Validation: Montserrat Terrazas, Chiara L. Castellazzi, Carlos González.

Writing – original draft: Federica Battistini, Isabelle Brun-Heath, Carlos González, Modesto Orozco.

Writing – review & editing: Federica Battistini, Modesto Orozco.

References

1. Garland Science—Book: Molecular Biology of the Cell + 5. [cited 26 Jan 2015]. Available: http://www.garlandscience.com/product/isbn/9780815341055

2. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003; 33 Suppl: 245–54. https://doi.org/10.1038/ng1089 PMID: 12610534

3. Ng HH, Bird A. DNA methylation and chromatin modification. Curr Opin Genet Dev. 1999; 9: 158–63. Available: http://www.ncbi.nlm.nih.gov/pubmed/10322130 https://doi.org/10.1016/s0959-437x(99)80024-0 PMID: 10322130

4. Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. 1982; 10: 2709–21. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=320645&tool=pmcentrez&rendertype=abstract https://doi.org/10.1093/nar/10.8.2709 PMID: 7079182

5. Kass SU, Landsberger N, Wolffe AP. DNA methylation directs a time-dependent repression of transcription initiation. Curr Biol. 1997; 7: 157–165. https://doi.org/10.1016/s0960-9822(97)70086-1 PMID: 9395433

6. Hashimshony T, Zhang J, Keshet I, Bustin M, Cedar H. The role of DNA methylation in setting up chromatin structure during development. Nat Genet. 2003; 34: 187–92. https://doi.org/10.1038/ng1158 PMID: 12740577

7. Miranda TB, Jones PA. DNA methylation: the nuts and bolts of repression. J Cell Physiol. 2007; 213: 384–90. https://doi.org/10.1002/jcp.21224 PMID: 17708532

8. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008; 9: 465–76. https://doi.org/10.1038/nrg2341 PMID: 18463664

9. Kulis M, Heath S, Bibikova M, Queirós AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet. 2012; 44: 1236–1242. https://doi.org/10.1038/ng.2443 PMID: 23064414

10. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci. 2006; 31: 89–97. https://doi.org/10.1016/j.tibs.2005.12.008 PMID: 16403636

11. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13: 484–92. Available: http://www.nature.com/nrg/journal/v13/n7/full/nrg3230.html?WT.ec_id=NRG-201207 https://doi.org/10.1038/nrg3230 PMID: 22641018

12. Pérez A, Castellazzi CL, Battistini F, Collinet K, Flores O, Deniz O, et al. Impact of Methylation on the Physical Properties of DNA. Biophys J. 2012; 102: 2140–2148. Available: https://doi.org/10.1016/j.bpj.2012.03.056 PMID: 22824278

13. Portella G, Battistini F, Orozco M. Understanding the connection between epigenetic DNA methylation and nucleosome positioning from computer simulations. PLoS Comput Biol. 2013; 9: e1003354. https://doi.org/10.1371/journal.pcbi.1003354 PMID: 24278005

14. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet. 2012; 13: 7–13. https://doi.org/10.1038/nrg3080 PMID: 22083101

15. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; 324: 930–5. https://doi.org/10.1126/science.1170116 PMID: 19372391

16. Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell. 2013; 152: 1146–59. https://doi.org/10.1016/j.cell.2013.02.004 PMID: 23434322

17. Kellinger MW, Song C-X, Chong J, Lu X-Y, He C, Wang D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol. 2012; 19: 831–3. https://doi.org/10.1038/nsmb.2346 PMID: 22820989

18. He Y-F, Li B-Z, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011; 333: 1303–7. https://doi.org/10.1126/science.1210944 PMID: 21817016

19. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011; 333: 1300–3. https://doi.org/10.1126/science.1210597 PMID: 21778364

20. Guo JU, Su Y, Zhong C, Ming G, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell. 2011; 145: 423–34. https://doi.org/10.1016/j.cell.2011.03.022 PMID: 21496894

21. Pfaffeneder T, Hackner B, Truss M, Münzel M, Müller M, Deiml CA, et al. The discovery of 5-formylcyto-sine in embryonic stem cell DNA. Angew Chem Int Ed Engl. 2011; 50: 7008–12. https://doi.org/10.1002/anie.201103899 PMID: 21721093

22. Wu H, Zhang Y. Tet1 and 5-hydroxymethylation. Cell Cycle. 2014; 10: 2428–2436. https://doi.org/10.4161/cc.10.15.16930 PMID: 21750410

23. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, et al. 5-Hydroxy-methylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011; 2: 241. https://doi.org/10.1038/ncomms1240 PMID: 21407207

24. Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxy-methylcytosine in the mammalian genome. Cell. 2012; 149: 1368–80. https://doi.org/10.1016/j.cell.2012.04.027 PMID: 22608086

25. Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol. 2011; 29: 68–72. https://doi.org/10.1038/nbt.1732 PMID: 21151123

26. Münzel M, Globisch D, Brückl T, Wagner M, Welzmiller V, Michalakis S, et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl. 2010; 49: 5375–7. https://doi.org/10.1002/anie.201002033 PMID: 20583021

27. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methyl-cytosine and 5-hydroxymethylcytosine at single-base resolution. Science. 2012; 336: 934–7. https://doi.org/10.1126/science.1220671 PMID: 22539555

28. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009; 324: 929–30. https://doi.org/10.1126/science.1169786 PMID: 19372393

29. Wang Y, Zhang X, Wu F, Chen Z, Zhou X. Bisulfite-free, single base-resolution analysis of 5-hydroxy-methylcytosine in genomic DNA by chemical-mediated mismatch. Chem Sci. 2019; 10: 447–452. https://doi.org/10.1039/c8sc04272a PMID: 30746092

30. Zeng T, Liu L, Li T, Li Y, Gao J, Zhao Y, et al. Detection of 5-methylcytosine and 5-hydroxymethylcyto-sine in DNA via host-guest interactions inside α-hemolysin nanopores. Chem Sci. 2015; 6: 5628–5634. https://doi.org/10.1039/c5sc01436k PMID: 28757950

31. Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol Cell. 2011; 42: 451–64. https://doi.org/10.1016/j.molcel.2011.04.005 PMID: 21514197

32. Yildirim O, Li R, Hung J-H, Chen PB, Dong X, Ee L-S, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell. 2011; 147: 1498–510. https://doi.org/10.1016/j.cell.2011.11.054 PMID: 22196727

33. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, et al. Genome-wide mapping of 5-hydro-xymethylcytosine in embryonic stem cells. Nature. 2011; 473: 394–7. https://doi.org/10.1038/nature10102 PMID: 21552279

34. Khare T, Pai S, Koncevicius K, Pal M, Kriukiene E, Liutkeviciute Z, et al. 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. Nat Struct Mol Biol. 2012; 19: 1037–43. https://doi.org/10.1038/nsmb.2372 PMID: 22961382

35. Szulwach KE, Li X, Li Y, Song C-X, Han JW, Kim S, et al. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS Genet. 2011; 7: e1002154. https://doi.org/10.1371/journal.pgen.1002154 PMID: 21731508

36. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011; 473: 398–402. https://doi.org/10.1038/nature10008 PMID: 21460836

37. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, et al. Genome-wide analysis of 5-hydroxymethylcy-tosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. 2011; 25: 679–84. https://doi.org/10.1101/gad.2036011 PMID: 21460036

38. Shen L, Zhang Y. 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. Curr Opin Cell Biol. 2013; 25: 289–96. https://doi.org/10.1016/j.ceb.2013.02.017 PMID: 23498661

39. Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, et al. Loss of 5-hydroxymethylcytosine is an epige-netic hallmark of Melanoma. Cell. 2012; 150: 1135–1146. https://doi.org/10.1016/j.cell.2012.07.033 PMID: 22980977

40. Jimenez-Useche I, Shim D, Yu J, Yuan C. Unmethylated and methylated CpG dinucleotides distinc-tively regulate the physical properties of DNA. Biopolymers. 2014; 101: 517–524. https://doi.org/10.1002/bip.22411 PMID: 24122444

41. Ngo TTM, Yoo J, Dai Q, Zhang Q, He C, Aksimentiev A, et al. Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. Nat Commun 2016 71. 2016;7: 1–9. https://doi.org/10.1038/ncomms10813 PMID: 26905257

42. Zaichuk T, Marko J. Single-molecule micromanipulation studies of methylated DNA. Biophys J. 2021; 120: 2148–2155. https://doi.org/10.1016/j.bpj.2021.03.039 PMID: 33838135

43. Mendonca A, Chang EH, Liu W, Yuan C. Hydroxymethylation of DNA influences nucleosomal conformation and stability in vitro. Biochim Biophys Acta. 2014; 1839: 1323–9. https://doi.org/10.1016/j.bbagrm.2014.09.014 PMID: 25263161

44. Liebl K, Zacharias M. How methyl-sugar interactions determine DNA structure and flexibility. Nucleic Acids Res. 2019; 47: 1132–1140. https://doi.org/10.1093/nar/gky1237 PMID: 30541032

45. Buitrago D, Labrador M, Arcon JP, Lema R, Flores O, Esteve-Codina A, et al. Impact of DNA methylation on 3D genome structure. 2020 [cited 10 Mar 2021]. https://doi.org/10.21203/rs.3.rs-36311/v1

46. Du Q, Luu PL, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: Readers of the epigenome. Epigenomics. Future Medicine Ltd.; 2015. pp. 1051–1073. https://doi.org/10.2217/epi.15.39 PMID: 25927341

47. Thalhammer A, Hansen AS, El-Sagheer AH, Brown T, Schofield CJ. Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. Chem Commun (Camb). 2011; 47: 5325–7. https://doi.org/10.1039/c0cc05671e PMID: 21451870

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

49. Ivani I, Dans PD, Noy A, Pérez A, Faustino I, Hospital A, et al. Parmbsc1: a refined force field for DNA simulations. Nat Methods. 2015; 13: 55–8. https://doi.org/10.1038/nmeth.3658 PMID: 26569599

50. Kielkopf CL, Ding S, Kuhn P, Rees DC. Conformational flexibility of B-DNA at 0.74 Å resolution: d (CCAGTACTGG)2. J Mol Biol. 2000; 296: 787–801. https://doi.org/10.1006/jmbi.1999.3478 PMID: 10677281

51. Maehigashi T, Hsiao C, Kruger Woods K, Moulaei T, Hud N V., Williams LD. B-DNA structure is intrinsically polymorphic: Even at the level of base pair positions. Nucleic Acids Res. 2012; 40: 3714–3722. https://doi.org/10.1093/nar/gkr1168 PMID: 22180536

52. Dans PD, Pérez A, Faustino I, Lavery R, Orozco M. Exploring polymorphisms in B-DNA helical conformations. Nucleic Acids Res. 2012; 40: 10668–78. https://doi.org/10.1093/nar/gks884 PMID: 23012264

53. Lankas F, Sponer J, Langowski J, Cheatham TE. DNA basepair step deformability inferred from molecular dynamics simulations. Biophys J. 2003; 85: 2872–83. https://doi.org/10.1016/S0006-3495(03)74710-9 PMID: 14581192

54. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. Proc Natl Acad Sci U S A. 1998; 95: 11163–8. https://doi.org/10.1073/pnas.95.19.11163 PMID: 9736707

55. Dans PD, Balaceanu A, Pasi M, Patelli AS, Petkevičiūtė D, Walther J, et al. The static and dynamic structural heterogeneities of B-DNA: extending Calladine-Dickerson rules. Nucleic Acids Res. 2019; 47: 11090–11102. https://doi.org/10.1093/nar/gkz905 PMID: 31624840

56. Chen K, Meng Q, Ma L, Liu Q, Tang P, Chiu C, et al. A novel DNA sequence periodicity decodes nucleosome positioning. Nucleic Acids Res. 2008; 36: 6228–6236. https://doi.org/10.1093/nar/gkn626 PMID: 18829715

57. Lowary PT, Widom J. Nucleosome packaging and nucleosome positioning of genomic DNA. Proc Natl Acad Sci U S A. 1997; 94: 1183–1188. https://doi.org/10.1073/pnas.94.4.1183 PMID: 9037027

58. Lavery R, Maddocks JH, Pasi M, Zakrzewska K. Analyzing ion distributions around DNA. Nucleic Acids Res. 2014; 42: 8138–8149. https://doi.org/10.1093/nar/gku504 PMID: 24906882

59. Zhou Z, Hong EJ, Cohen S, Zhao W ning, Ho H yi H, Schmidt L, et al. Brain-Specific Phosphorylation of MeCP2 Regulates Activity-Dependent Bdnf Transcription, Dendritic Growth, and Spine Maturation. Neuron. 2006; 52: 255–269. https://doi.org/10.1016/j.neuron.2006.09.037 PMID: 17046689

60. Cramer JM, Scarsdale JN, Walavalkar NM, Buchwald WA, Ginder GD, Williams DC. Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. J Biol Chem. 2014; 289: 1294–1302. https://doi.org/10.1074/jbc.M113.512236 PMID: 24307175

61. Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 2001; 15: 710–723. https://doi.org/10.1101/gad.194101 PMID: 11274056

62. Khabiri M, Freddolino PL. Deficiencies in Molecular Dynamics Simulation-Based Prediction of Protein–DNA Binding Free Energy Landscapes. J Phys Chem B. 2017; 121: 5151–5161. https://doi.org/10.1021/acs.jpcb.6b12450 PMID: 28471184

63. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. 2012; 40: 4841–4849. https://doi.org/10.1093/nar/gks155 PMID: 22362737

64. Schroeder SJ, Turner DH. Optical melting measurements of nucleic acid thermodynamics. Methods Enzymol. 2009; 468: 371–387. https://doi.org/10.1016/S0076-6879(09)68017-4 PMID: 20946778

65. Dans PD, Ivani I, Hospital A, Portella G, González C, Orozco M. How accurate are accurate force-fields for B-DNA? Nucleic Acids Res. 2017; 44: gkw1355. https://doi.org/10.1093/nar/gkw1355 PMID: 28088759

66. 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 DJ. Gaussian 16. Gaussian, Inc, Wallingford CT. 2016.

67. Cieplak P, Cornell WD, Bayly C, Kollman PA. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. J Comput Chem. 1995; 16: 1357–1377. https://doi.org/10.1002/jcc.540161106

68. Arnott S, Hukins DW. Optimised parameters for A-DNA and B-DNA. Biochem Biophys Res Commun. 1972; 47: 1504–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/5040245 https://doi.org/10.1016/0006-291x(72)90243-4 PMID: 5040245

69. Macke TJ, Case DA. Modeling Unusual Nucleic Acid Structures. 1997. pp. 379–393. https://doi.org/10.1021/bk-1998-0682.ch024

70. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79: 926–935. https://doi.org/10.1063/1.445869

71. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro TAD, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer YH, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo DJ, Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan R, Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao DMY and PAK. AMBER 2018. Univ California, San Fr. 2018.

72. Pérez A, Luque FJ, Orozco M. Dynamics of B-DNA on the Microsecond Time Scale. J Am Chem Soc. 2007; 129: 14739–14745. https://doi.org/10.1021/ja0753546 PMID: 17985896

73. Dans PD, Danilãne L, Ivani I, Dršata T, Lankaš F, Hospital A, et al. Long-timescale dynamics of the Drew–Dickerson dodecamer. Nucleic Acids Res. 2016; 44: 4052–4066. https://doi.org/10.1093/nar/gkw264 PMID: 27084952

74. Dans PD, Walther J, Gómez H. Multiscale simulation of DNA. Curr Opin Struct Biol. 2016; 37: 29–45. https://doi.org/10.1016/j.sbi.2015.11.011 PMID: 26708341

75. Dang LX. Mechanism and Thermodynamics of Ion Selectivity in Aqueous Solutions of 18-Crown-6 Ether: A Molecular Dynamics Study. J Am Chem Soc. 1995; 117: 6954–6960. https://doi.org/10.1021/ja00131a018

76. Dang LX, Kollman PA. Free Energy of Association of the 18-Crown-6:K+ Complex in Water: A Molecular Dynamics Simulation. J Am Chem Soc. 1990; 112: 5716–5720. https://doi.org/10.1021/ja00171a006

77. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. J Chem Phys. 1993; 98: 10089–10092. https://doi.org/10.1063/1.464397

78. Ryckaert J-P, Ciccotti+ G, Berendsen HJC. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. J Comput Phys. 1977; 23: 321–341. Available: http://physics.ujep.cz/~mlisal/md/shake.pdf

79. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys. 1984; 81: 3684–3690. https://doi.org/10.1063/1.448118

80. Hospital A, Andrio P, Cugnasco C, Codo L, Becerra Y, Dans PD, et al. BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. Nucleic Acids Res. 2016; 44: D272–D278. https://doi.org/10.1093/nar/gkv1301 PMID: 26612862

81. Hospital A, Battistini F, Soliva R, Gelpí JL, Orozco M. Surviving the deluge of biosimulation data. WIREs Comput Mol Sci. 2020; 10. https://doi.org/10.1002/wcms.1449

82. Gelpí JL, Kalko SG, Barril X, Cirera J, de La Cruz X, Luque FJ, et al. Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. Proteins. 2001; 45: 428–37. Available: http://www.ncbi.nlm.nih.gov/pubmed/11746690 https://doi.org/10.1002/prot.1159 PMID: 11746690

83. Cuervo A, Dans PD, Carrascosa JL, Orozco M, Gomila G, Fumagalli L. Direct measurement of the dielectric polarization properties of DNA. Proc Natl Acad Sci U S A. 2014; 111: E3624–30. https://doi.org/10.1073/pnas.1405702111 PMID: 25136104

84. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph. 1996; 14: 33–38. https://doi.org/10.1016/0263-7855(96)00018-5 PMID: 8744570

85. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem. 2004; 25: 1605–1612. https://doi.org/10.1002/jcc.20084 PMID: 15264254

86. Lavery R, Moakher M, Maddocks JH, Petkeviciute D, Zakrzewska K. Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res. 2009; 37: 5917–29. https://doi.org/10.1093/nar/gkp608 PMID: 19625494