Impact of Methylation on the Physical Properties of DNA

Alberto Pérez,[†‡△] Chiara Lara Castellazzi,[‡△] Federica Battistini,[‡] Kathryn Collinet,[‡] Oscar Flores,[‡] Ozgen Deniz,[‡] Maria Luz Ruiz,[‡] David Torrents,[†¶] Ramon Eritja,[§] Montserrat Soler-López,[‡] and Modesto Orozco[†‡‖*]

[†]IRB-BSC Joint Research Program on Computational Biology, Barcelona Supercomputing Center, Barcelona, Spain; [‡]IRB-BSC Joint Research Program on Computational Biology and [§]Chemistry and Molecular Pharmacology Program, Institute for Research in Biomedicine, Barcelona, Spain; [¶]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain; and [‖]Departamento de Bioquímica, Facultat de Biología, Barcelona, Spain

ABSTRACT There is increasing evidence for the presence of an alternative code imprinted in the genome that might contribute to gene expression regulation through an indirect reading mechanism. In mammals, components of this coarse-grained regulatory mechanism include chromatin structure and epigenetic signatures, where d(CpG) nucleotide steps are key players. We report a comprehensive experimental and theoretical study of d(CpG) steps that provides a detailed description of their physical characteristics and the impact of cytosine methylation on these properties. We observed that methylation changes the physical properties of d(CpG) steps, having a dramatic effect on enriched CpG segments, such as CpG islands. We demonstrate that methylation reduces the affinity of DNA to assemble into nucleosomes, and can affect nucleosome positioning around transcription start sites. Overall, our results suggest a mechanism by which the basic physical properties of the DNA fiber can explain parts of the cellular epigenetic regulatory mechanisms.

INTRODUCTION

Determining the mechanisms that regulate gene expression in complex organisms is the next frontier of genomics research (1). In the traditional paradigm, specific proteins regulate gene expression through the recognition of certain sequence signals (by means of specific hydrogen-bond interactions) upstream of the transcription start sites (TSSs) (2). Nevertheless, there is increasing evidence about the presence of an alternative code that may contribute to a rough regulation of gene expression through an indirect reading mechanism, probably signaled by chromatin structure and epigenetic marks (3,4). This mechanism is unlikely (even in a synergistic manner) to achieve the fine-tuning and specificity of the direct protein-DNA readout. Conversely, it probably plays a pivotal role in basal gene expression, which requires less regulation and for which the extreme cost of developing a highly specific protein regulation infrastructure seems unjustified. Key players in this regulatory mechanism may be d(CpG) steps, which despite being largely underrepresented in the genome of complex organisms are enriched in nearly 60% of human promoters, where they often define ultrarich CpG segments, the so-called CpG islands (5,6). Even if CpG islands do not contain specific transcription binding motifs, they clearly favor the downstream binding of the transcription machinery (7), particularly for those genes that are usually active. The molecular basis of the d(CpG) effect on gene regulation remains unclear, although it has been suggested to be related to the definition of DNA fiber properties (8).

One of the most intriguing features of the d(CpG) step is its ability to undergo nonmutagenic chemical modifications such as cytosine methylation (9). In mammalian genomes, DNA methyltransferases (DNMTs) can transfer a methyl group from S-adenosylmethionine to cytosine at CpG dinucleotides (10). The bulk of the methylation takes place during DNA replication in the S-phase of the cell cycle (11), and is the most abundant form of post-replicative DNA modification observed in eukaryotic organisms (12). During this process, the cytosine is flipped 180° out of the DNA backbone into an active-site pocket of the enzyme (13) where methylation of cytosine takes place.

Intriguingly, methylation of cytosines seems to be an erroneous decision of evolution because it dramatically increases the chances of C→T mutation, but this seeming disadvantage is compensated for by the gain in regulatory possibilities offered by methylation. Indeed, highly methylated DNA is typically associated with inactive genes, whereas methylation depletion is observed for active genes (14,15). Furthermore, most cytosines in CpG steps, except those in CpG islands, are methylated in vertebrate somatic cells (16,17). The first step of methylation occurs early in mammalian development as a result of de novo DNMTs (Dnmt3a and Dnmt3b) (18) that methylate CpG steps in both DNA strands. The methylation profile is conserved by maintenance DNA methyltransferase (Dnmt1) throughout cell divisions. During replication, daughter strands are nonmethylated, resulting in hemimethylated DNA. Dnmt1 recognizes hemimethylated CpG steps and methylates the daughter strand (19). Recent studies have demonstrated that changes in methylation patterns along CpG islands and CpG shores (methylation

Submitted December 27, 2011, and accepted for publication March 22, 2012.

△Alberto Pérez and Chiara Lara Castellazzi contributed equally to this work.

*Correspondence: modesto@mmb.pcb.ub.es

Editor: Michael Levitt.

0006-3495/12/05/2140/9 $2.00

doi: 10.1016/j.bpj.2012.03.056

hotspots on the outskirts of CpG islands (20)) correlate with tissue differentiation and cancer, proving the role of methylation in cellular reprogramming (20,21,23,24). The regulatory function of methylated cytosines (hereafter referred to as ^{Me}C) was traditionally explained by their interaction with methyl-CpG binding domain proteins (MBDs) (18), but considering the prevalence of cytosine methylation, MBDs alone cannot entirely account for the role of d(CpG) methylation in the cell. An increased level of complexity is provided by the almost nonexistent sequence specificity of the DNMTs, precluding the mechanism underlying the methylation reaction (26).

Here we present a comprehensive theoretical analysis of the d(CpG) properties along with an experimental validation of key theoretical findings. We show that d(CpG) steps display unique physical properties, especially in the context of CpG islands, that severely change upon methylation. Our calculations suggested, and experiments confirmed, that DNA segments containing ^{Me}C are very stiff and hard to bend, and display a lower tendency to circularize or form nucleosomes by wrapping around histones. The latter effect has striking consequences for the organization of nucleosome arrays near TSSs, which in turn modifies the accessibility of regulatory proteins, leading to alterations in the pattern of DNA expression. Overall, without diminishing the role of specific regulatory proteins, basic descriptors of DNA physical properties can help us rationalize several seemingly disconnected pieces of the puzzle of DNA regulation.

METHODS

Molecular-dynamics simulations

We performed molecular-dynamics (MD) simulations on an array of different oligomers containing CpG steps, in both their methylated and nonmethylated forms (see Table S1 in the Supporting Material). We also included in our analysis trajectories of unmethylated CpG step data from the Ascona B-DNA Consortium (ABC) (27) database in all tetramer environments to enrich the dynamics database. All simulations were carried out in duplicates for 100 ns (after equilibration) using explicit solvent, the parmbsc0 refinement of the Amber force field (28), and state-of-the-art simulation conditions (Supporting Material).

Mesoscopic model of DNA flexibility

We derived a flexibility model from different MD equilibrium trajectories using a harmonic model (29–31). Accordingly, we projected the MD trajectories onto a helical reference system to obtain equilibrium values and derive the covariance matrix, which we then inverted to recover the stiffness matrices for each basepair step, from which a mesoscopic estimate of the energy associated to a given deformation can be easily computed (29,32) as

$$E = 0.5\sum_{i=1}^{6}\sum_{j=1}^{6} f_{ij}\Delta X_i \Delta X_j,$$

where ΔX_i is the perturbation from equilibrium geometry (Fig. 1), and f_{ij} are elements of Θ, where Θ is the 6×6 stiffness matrix expressing the stiffness of a given step to deformation in roll, tilt, twist, slide, shift, and rise (see Supporting Material, Fig. 2, and Pérez et al. (33)). Alternatively, global deformation parameters can be derived using a similar approach, but considering global instead of local geometric descriptors (see Fig. S1, Supporting Material, and Lankas et al. (34)). Note that elastic parameters derived from protein-DNA crystal complexes (29,31) or simulation data are in good agreement (33), supporting their use to describe DNA flexibility. Here, we favored the use of MD-derived values for consistency with the newly derived parameters describing ^{Me}CpG steps (which cannot be derived from analysis of crystal structures).

In this work, we used a mesoscopic method to estimate deformation energy related to nucleosome formation and circularization assays. This implies that indirect readout mechanisms prevail over the direct readout for the description of sequence preferences in nucleosome binding. A second implication is that these deformations follow a harmonic behavior. Both of these assumptions represent simplifications, and thus the validity of the method is not always guaranteed (32,35,36).

Circularization assays and modeling

We carried out DNA circularization experiments to validate our theoretical estimations about the impact of methylated cytosines on DNA physical properties. For this purpose, we designed a short polymerizable oligonucleotide (d(GAAAAAACGGGCGAAAAACGG)·d(TCCCGTTTTTCGCCCGTTTTT)) based on a reported sequence favoring the formation of minicircles (37), with the incorporation of a central CpG dinucleotide subject to be methylated and 5′-sticky ends to enable the formation of multimers. Thus, under favorable ligation conditions, the multimers form circles that are as short as allowed by the geometry and flexibility of the DNA (Fig. 3). As a negative control, we selected a previously reported nonbendable oligonucleotide (d(GCAAATATTGAAAAC)·d(GCGTTTTCAATATTT; see Supporting Material for details). The ligation products were analyzed by atomic force microscopy (AFM) and by two-dimensional gel electrophoresis (Fig. 3, Fig. S2, Fig. S3, and Supporting Material).

Circularization efficiency was determined based on the *J*-factor, which defines the ratio of circular and linear DNA species for a given sequence length (37). Experimental *J*-factors were extracted from the linear and circle DNA signal intensities as detected on two-dimensional gels (see Supporting Material and Fig. S3). Although the absolute *J*-factors depend on the experimental setup, the ratio of *J*-factors of methylated versus nonmethylated oligos provide a reliable measure of the impact of methylation on DNA circularization. Accordingly, experimental *J*-factors can validate whether theoretical suggestions regarding physical changes induced by methylation are correct. We derived theoretical *J*-factors from Monte Carlo simulations (Supporting Material) using mesoscopic descriptors derived from MD simulations (see above).

Mesoscopic model of nucleosome deformation energy

We theoretically determined the ability of a 147-mer DNA sequence to wrap around a nucleosome using harmonic deformation energy as described above considering mesoscopic descriptors. To reduce the noise, we determined the deformation vector (ΔX above) using the target geometry obtained by Fourier-averaging all available crystal structures of nucleosome particles. As noted above, our mesoscopic model is useful in so far as histone-bound DNA deforms harmonically and the indirect readout has an important contribution in directing nucleosome formation. Regarding the first point, the rotational degrees of motion in nucleosomes (twist, roll, and tilt) clearly fall within the normal fluctuations of DNA (38,39), and only the translational parameter slide shows slightly more positive values than expected. Clearly, evolution has optimized nucleosome positioning sequences to have flexible steps such as d(CpA) and d(TpA) in crucial positions to accommodate the deformations required for nucleosome binding (see results of SELEX experiments in Thåström et al.

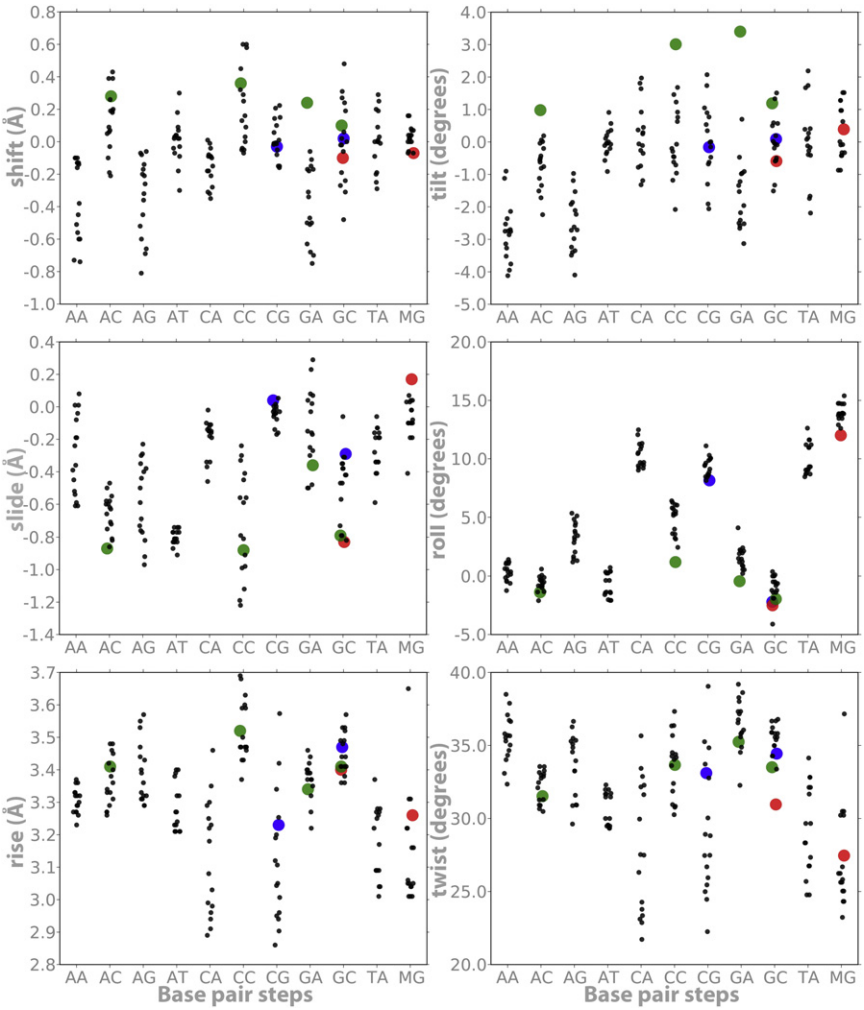

FIGURE 1 Average helical parameters (translations in angstroms and rotations in degrees) derived from MD simulations of the usual 10 dinucleotides plus d(^{Me}CpG; referred to as MG in the figure). The black dots correspond to control simulations (those performed for this work and those obtained from the ABC database (27)) for each central basepair step representing the different tetramer environments; notice the slight displacement from the vertical to avoid stacking of data. Blue dots stand for d(CpG) and d(GpC) steps when embedded in a poly d(CpG) track. Green dots refer to neighboring methylated steps (Xp^{Me}C): d(Ap^{Me}C), d(Tp^{Me}C)=d(GpA), d(Cp^{Me}C), and d(Gp^{Me}C). Finally, red dots stand for d(^{Me}CpG) and d(Gp^{Me}C) in the context of a poly d(CpG) track.

(40)). Note also that by using a single conformation to model all nucleosomes, Tolstorukov et al. (38) and Balasubramanian et al. (39) found a high degree of correlation between the predictions arising from harmonic deformation energies and nucleosome positioning sequences, a result that was consistent with nucleosome location experiments performed by our group (41). Thus, without ignoring its limitations, we believe the model can be useful to rationalize nucleosome positioning.

In vitro nucleosome reconstitution

To assess the effect of methylated DNA on nucleosome assembly, we selected a nucleosome positioning sequence (DNA construct 601.2 in Anderson and Widom (42)) to reconstitute nucleosomes in vitro after incubation with histones, before and after extensive DNA methylation (see Supporting Material). Methylated states were verified by DNA sequencing. The reconstituted nucleosomes were subsequently analyzed by gel shift assays (see Fig. 4).

RESULTS

Physical properties of CpG steps and CpG islands

MD simulations performed here in conjunction with those retrieved from the ABC database (27) revealed that sequence is crucial for defining the DNA equilibrium geometries (Fig. 1). In general, Pyr-Pur steps show a lower rise and twist, as well as higher roll values, than the rest of the dinucleotide steps. Additionally, they display an unusually large dispersion in certain key equilibrium helical parameters (e.g., twist), arising from different tetramer environments. Focusing on the different tetramer environments, we can see that the d(CpG) steps are peculiar in presenting bimodal distributions of some parameters (e.g., twist; see Fig. S4), which confirms previous ABC findings (27). Such bimodality is not an artifact that arises from incomplete sampling, because it is also present in multi-microsecond trajectories (43), and suggests that the d(CpG) step is specially flexible, as confirmed by a stiffness analysis (Fig. 2). It is worth noting that the large deformability in twist and roll, combined with large roll values, suggests that protein-induced curvature may be favored in DNA with CpG steps. Fig. S5 shows that CpG, despite the neighboring basepair, is more curved than most basepair steps (only TA and CA are comparable) and is directed preferentially toward the major groove.

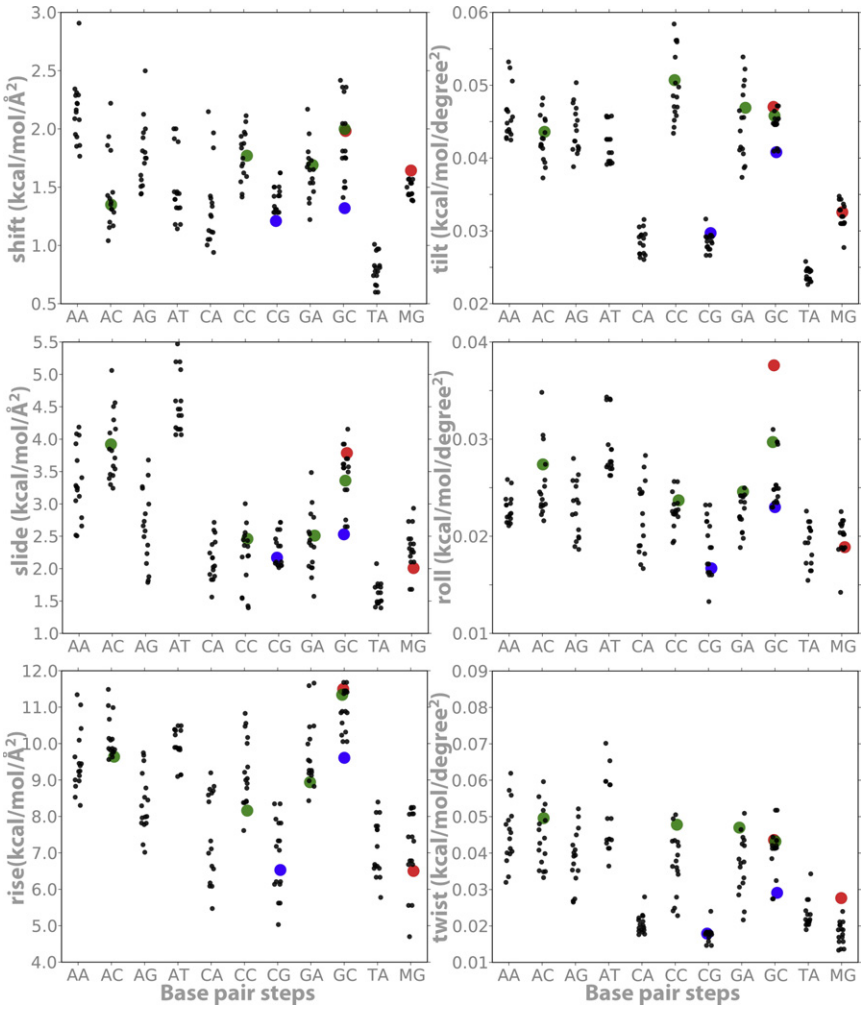

FIGURE 2 Average stiffness helical parameters (translations in kcal/mol Å^2 and rotations in kcal/mol deg^2) derived from MD simulations of the usual 10 dinucleotides plus d(^{Me}CpG; referred to as MG in the figure). Same notation as in Fig. 1.

One might expect the unique physical properties of d(CpG) steps to be amplified in long d(CpG)$_n$ tracks (i.e., CpG islands). Surprisingly, this is not the case, and the global geometrical properties of long d(CpG) segments are different from those expected by extrapolating the individual characteristics of the d(CpG) steps (Fig. S1). Thus, the high flexibility of the d(CpG) steps suggests that poly d(CpG) should be extremely flexible. Conversely, the d(CpG)$_9$ segment studied here is hardly distinguishable from other 18-mer duplexes in terms of global unwinding and isotropic bending. This is hardly surprising when one considers the differences between an individual d(CpG) step and a poly d(CpG). The former has the properties of an individual d(CpG) step, whereas the d(CpG)$_9$ segment has properties due to the alternation of d(CpG) and d(GpC) steps. Thus, the lower flexibility of d(GpC) steps (Fig. S1) makes the whole sequence overall stiffer than a simple extrapolation of d(CpG) properties. Furthermore, whereas the large roll of individual d(CpG) steps would imply a strong curvature of the entire oligonucleotide, the d(CpG)$_9$ curvature is actually very moderate due to the low roll values of d(GpC) steps. In conclusion, the properties of long d(CpG) segments are distinct from the extrapolation of properties of isolated d(CpG) steps, warning against the use of oversimplified rules of DNA flexibility.

Effect of CpG methylation

Early structural experiments suggested that cytosine methylation (Fig. S5) might induce helical transitions from B- to Z-DNA (44). However, a secondary structure analysis of CpG methylated oligonucleotides by circular dichroism spectroscopy (Supporting Material and Fig. S6) revealed that the transition only occurred at nonphysiological salt concentrations (from 1 to 2 M NaCl). This evidence is in agreement with our MD results and previously reported Fourier transform infrared (FTIR) spectroscopy data (45), demonstrating that in vivo DNA remains in the B-form upon methylation, and accordingly, the transition to the Z-form is not the underlying determinant for the physiological role of CpG methylation.

The results of the MD simulations suggest that when methylated, d(CpG) steps increase their average roll value and reduce their twist (Fig. 1), leading to an increase in local

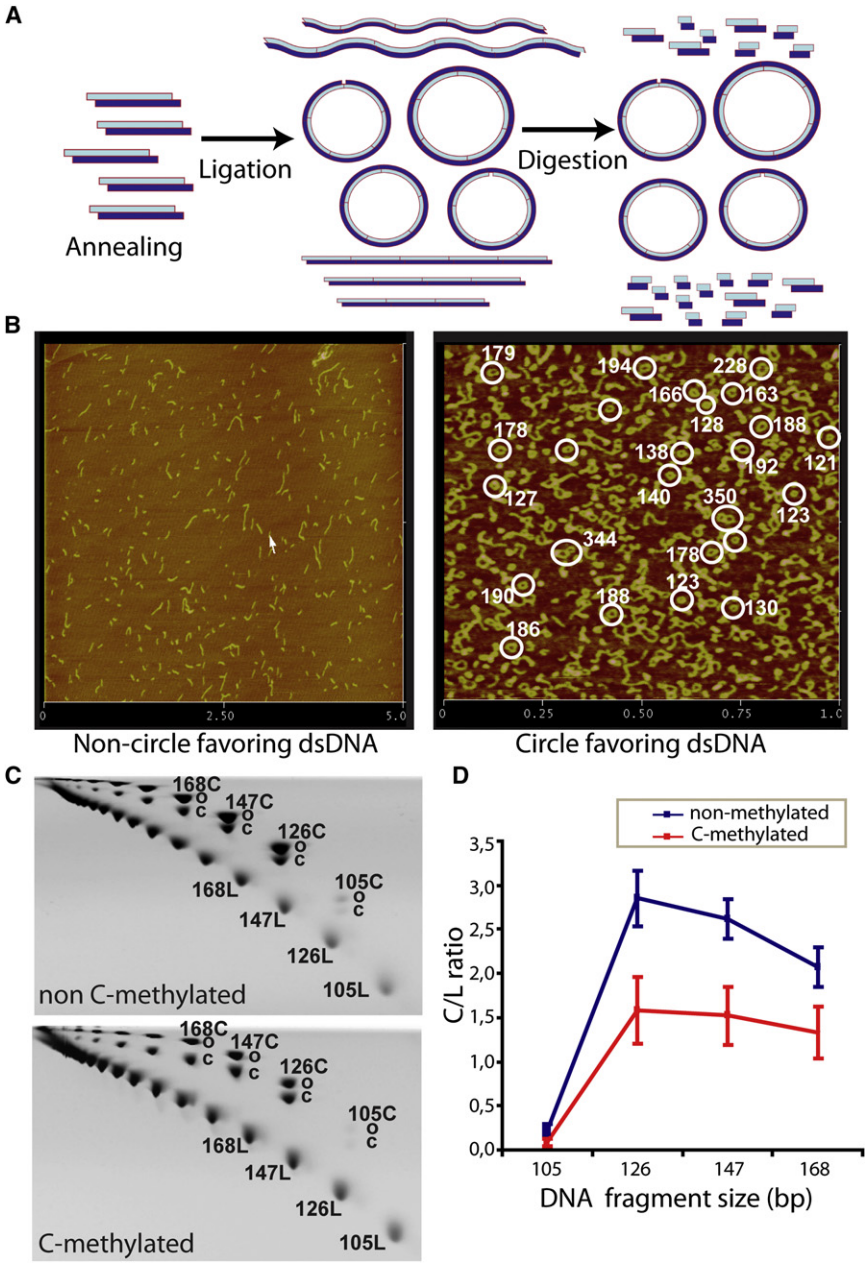

FIGURE 3 Overview of circularization assays. (*A*) Schematic diagram of the underlying principle of circularization assays. A DNA oligonucleotide is first annealed to form duplexes and subsequently is multimerized-circularized by a ligation reaction. Under favorable ligation conditions, DNA forms circles as short as allowed by the geometry and flexibility of the DNA. Only circularized DNA is resistant to an exonuclease digestion; linear multimers will be degraded. (*B*) AFM images of ligation products for 15-bp nonfavoring (*left*) and 21-bp favoring (*right*) circularization oligonucleotides; circle size estimations are highlighted in white. (*C*) 2D polyacrylamide native gels showing different migrations of linear (L) and circular (C) DNA species (which can be either covalently closed (cC) or nicked open (oC)) for nonmethylated and ^{Me}C oligomers of 21 bp, respectively. Linear DNA molecules are positioned on the lower diagonal, and circular DNA molecules are positioned on the upper diagonal. (*D*) The circularization efficiency is expressed as the ratio between C and L molecules of the same size.

curvature. Furthermore, methylation makes d(CpG) steps stiffer, especially in terms of roll and tilt deformations: the ^{Me}CpG step has larger tilt and roll force constants on average than the CpG step (MG and CG, respectively, in Fig. 2). Methylation also alters the geometric properties of the basepair step previous to the d(CpG) site, here denoted as d(XpC), where X = A, C, G, or T (Figs. 1 and 2). In canonical DNA, d(XpC) steps tend to compensate for the geometry and relative stiffness of d(CpG) in twist, tilt, and roll. However, upon methylation, we observed an increase of force constants for rotational parameters in both the ^{Me}CpG and d(Xp^{Me}C) steps (*green dots* in Fig. 2). Hence, the additive effect of methylation leads to significant alterations in the global physical properties of DNA, especially for CpG islands (Figs. 1 and 2, and Fig. S1).

The higher stiffness of the d(^{Me}CpG) steps should lead to a decrease in the DNA circularization efficiency, which must be especially visible for the smallest circles. Indeed, Monte Carlo calculations using the MD-derived stiffness parameters (see Materials and Methods, and Supporting Material) suggested that circularization of 126- to 189-bp-long methylated oligos (one ^{Me}CpG every 21 bp) is more difficult than circularization of unmethylated ones (with a relative *J*-factor of 0.05–0.2). This result was confirmed by circularization experiments on the same sequence with a relative *J*-factor of ~0.5 (see Fig. 3, Materials and Methods, and Supporting Material).

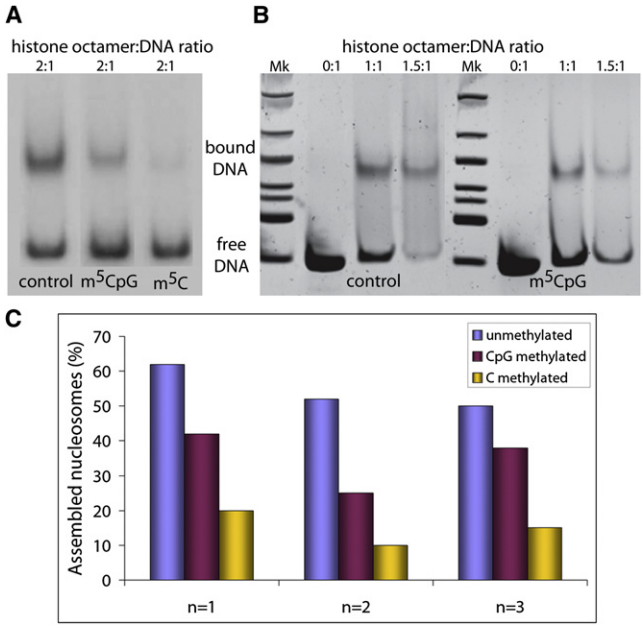

FIGURE 4 In vitro nucleosome core particle reconstitution. Results of gel mobility shift assays of nucleosomes reconstituted in vitro with a 147-bp 601.2 DNA fragment containing either C or ^{Me}C at different histone octamer concentrations are shown. The upper bands correspond to histone core-bound DNA, and lower bands correspond to unbound (or free) DNA. Mk: DNA ladder for size band estimation. (*A*) Radiolabeled DNA bands. (*B*) DNA bands stained with SyBr Safe (Invitrogen) and visualized by ultraviolet light. (*C*) Histograms displaying the percentage of in vitro assembled nucleosomes using the same sequence in different methylation conditions coming from triplicate experiments.

Potential impact on chromatin structure

The sequence-dependent physical properties of DNA play a crucial role in determining nucleosome positioning (46–52) and thus are instrumental in genome regulation (53,54). An analysis of nucleosome distribution in *Saccharomyces cerevisiae* (55,56) revealed that d(CpG)s are enriched in nucleosome-bound regions but depleted in internucleosomal segments, supporting the idea that d(CpG) steps are easily fitted into the nucleosome structure. Protein-DNA interactions are governed by direct and indirect readouts. The former arises from protein-DNA direct interactions, whereas the latter is based on the ability of a DNA sequence to deform into a conformation that makes the interaction happen. Without underestimating the importance of direct readout mechanisms in nucleosome binding, we note that indirect readout models seem to capture well the global positioning profile of nucleosomes (41). Thus, changes in the physical properties of the DNA fiber related to methylation should have a direct impact on nucleosome affinity and positioning. Our models prompted us to hypothesize that in the absence of external factors (e.g., MBD proteins or chromatin remodelers), the increased stiffness due to d(CpG) methylation leads to a higher deformation energy required to wrap DNA around a nucleosome (Fig. 5). We tested this hypothesis by conducting in vitro nucleosome reconstitution experiments (see Materials and Methods, and Supporting Material) with normal and methylated DNAs. The results confirm that the d(^{Me}CpG) DNA has a lower ability to form nucleosomes than the nonmethylated sequence (Fig. 4).

Interestingly, nucleosome formation was further decreased when all cytosines in the DNA were methylated, which confirms that a reduced flexibility is mainly responsible for the lower affinity of methylated DNA for the histones. Other nonmammalian organisms that have alternative cytosine methylation patterns besides the methylation of CpG steps could use this strategy for gene regulation. Tillo and Hughes (57) established that increasing C+G contents correlate well with higher nucleosome formation. Therefore, it is not surprising that a mechanism in which more cytosines are methylated would further rigidify the sequence, further changing the nucleosome positioning preferences (57).

DISCUSSION

d(CpG) steps are statistically underrepresented in the genome, but they appear concentrated in regulatory regions,

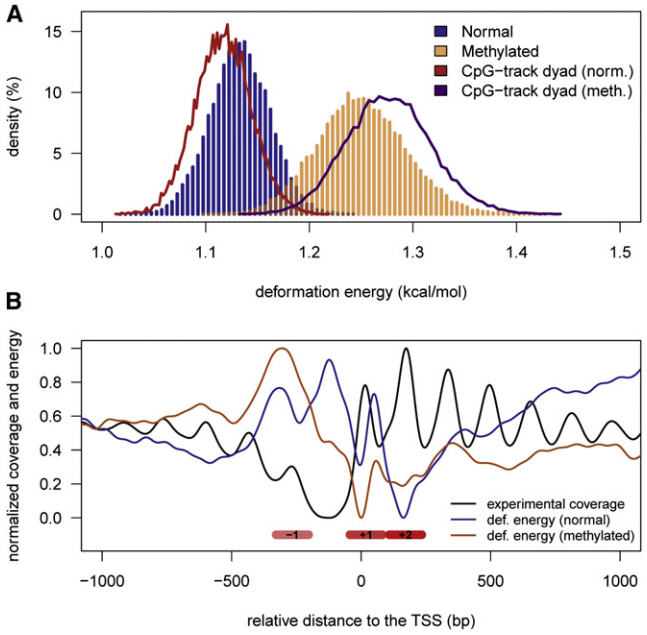

FIGURE 5 Impact of methylation on nucleosome positioning. (*A*) The distribution of predicted energies (per base step) for 147-bp-long random DNAs in normal (*blue histogram*) and methylated (*orange histogram*) forms, respectively. The curves correspond to the predicted energy (per base step) when random oligos contain a poly d(CpG) track at the dyad, in normal (*red*) and methylated (*magenta*) forms. (*B*) The nucleosome distribution surrounding the TSSs of yeast genes determined from MNase digestion experiments (*black line*), compared with the predicted distortion energy to wrap a nucleosome in those sites when genomic DNA is normal (*blue line*) or methylated (*orange line*). All values were normalized to facilitate the interpretation of the plots. Nucleosome positions −1, +1, and +2 are indicated by red boxes for clarity.

which suggests that d(CpG) steps provide the suitable physical properties that enable the DNA to efficiently interact with regulatory proteins (58,59) and help define the correct nucleosome positioning. Indeed, our MD results suggest that isolated d(CpG) steps are curved (see Fig. S5) and particularly flexible, making d(CpG) steps appropriate for those regions in which DNA needs to be locally distorted to facilitate protein binding, in agreement with early NMR measures (60). This hypothesis is supported by an analysis of whole-genome nucleosome positioning in yeast, which revealed a d(CpG) step depletion in internucleosomal segments (41,55,56).

Mesoscopic calculations (Fig. 5 *A*) indicate that a $d(CpG)_5$ segment located at the dyad axis may favor nucleosome formation, and hence very long d(CpG) tracks (CpG islands) are very likely to assemble into nucleosomes. Nucleosomes can easily move along the d(CpG) track, presumably leading to a nucleosome-depleted region at the external borders of the island, and thereby imprinting a distinct nucleosome array organization that would define the accessibility to regulatory regions downstream of the CpG islands, where many promoters are located. This also accords with the fact that high C+G content correlates with nucleosome positioning (57)

Methylation increases the curvature of d(CpG) steps (Fig. S5), although this local geometric effect tends to be compensated for by neighboring steps. In fact, all tested 18-mer methylated oligos (except CpG islands) were less curved and less flexible than their unmethylated counterparts. On the other hand, as previously suggested (61,62), methylation increases d(CpG) stiffness, and this effect propagates to neighboring steps, leading to a global increase in the rigidity of DNA. Our results suggest that this effect alone explains (within the indirect readout model) the limited ability of methylated DNA to interact with certain proteins, such as transcription factors (63–65). Our in silico simulations and in vitro nucleosome reconstitution experiments showed that methylation reduces DNA affinity for nucleosomes, probably due to the increased rigidity of the DNA fiber. Our results are in full agreement with previous findings (66,67), with recent data about the anticorrelation between nucleosome formation and methylation (68), and with physical intuition that suggests that a more flexible fiber should wrap more easily than a rigid one. Additionally, the rigidifying effect of ^{Me}CpG is also observed in recent fluorescence resonance energy transfer (FRET)-derived data (69,70) showing that methylating nucleosome-bound DNA results in nucleosome compaction and rigidity. Taken together, these results indicate that DNA methylation may regulate nucleosome dynamics by increasing the rigidity of DNA either before or during nucleosome assembly.

Our combined theoretical and experimental results demonstrate that methylation decreases nucleosome formation. Nucleosome depletion is usually considered as a signal of gene activity (71,72). At the same time, gene activity correlates with low levels of methylation (73). We postulate that the presence of almost 10^5 ^{Me}CpG steps present in the genome could significantly modify the nucleosome positioning landscape. This hypothesis would explain how the gene expression pattern can change while the number of nucleosomes (but not positions) is kept constant in either methylated or canonical genomes. Hence, we analyzed the nucleosome organization around TSSs on the unmethylated yeast genome and subsequently compared the in silico effects of methylation on nucleosome positioning (Fig. 5 *B*). TSSs are typically characterized by a nucleosome-depleted region and well-positioned −1, +1, and +2 nucleosomes (55,74). As expected, these positions are clearly marked in the energy profiles: regions with high deformation energy signal nucleosome-depleted areas and vice versa (Fig. 5). These profiles support recent claims about particular nucleosome positioning sites (signaled by large in vitro propensities for nucleosome assembly) anchoring the formation of nucleosomal arrays in vivo (75,76). Methylation of d(CpG) steps modifies the deformation energy profile associated with nucleosome wrapping around the TSS, which ultimately may be reflected by a change in the nucleosome array (Fig. 5 *B*; note that this does not necessarily make the nucleosomes more diffuse) and, accordingly, in gene expression (77). In particular, it seems that upon methylation, the nucleosome-free region is less defined and the nucleosome −1 is moved downstream. Furthermore, our calculations suggest that when CpG islands are methylated, nucleosomes are concentrated at the CpG island edges, leading to a completely different configuration of the nucleosome array around the TSS and consequently to a change in gene activity. Hence, we can partially rationalize the striking effect of methylated CpG islands on the activity of several genes, particularly those involved in cancer (78), by considering the highly unfavorable impact that methylation has on the ability of poly CpG tracks to wrap around nucleosomes (Fig. 5 *A*). Further work is required to shed more light on this interesting hypothesis.

Taken together, our studies show how an apparently minor covalent change such as methylation can alter the physical properties of DNA, and how such a change can modify the ability of DNA to organize the chromatin fiber, which may be reflected by significant alterations in gene regulation, even in the absence of specific MBDs.

CONCLUSIONS

In summary, simple physical properties of DNA (described from calculations based on first principles) can provide a rationale for the seemingly chaotic diversity of gene regulatory signals in developed organisms, particularly epigenetic signatures such as cytosine methylation. Our results support the hypothesis that physical properties define a basal regulatory code that is superposed onto more elaborated

mechanisms involving the action of specific proteins when fine-tuning of gene function is required. Furthermore, our findings suggest that simply by varying the physical properties of some distant regions to a particular gene while keeping the specific protein-binding boxes unaltered, we may be able to modulate that gene's biological functionality. This raises interesting possibilities in the emerging field of synthetic biology. From the results of this study, it follows that methylated DNA is not as likely to form nucleosomes. However, the complete picture is even more complex when one considers that DNMTs have a greater preference to target nucleosome-bound DNA, slightly enriching it (1%) in ^{Me}CpG steps (79).

SUPPORTING MATERIAL

Supplementary materials and methods, seven figures, a table, and references are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00397-9.

We thank Anna Aviñó for help with the experiments. The authors declare that they have no competing interests and acknowledge no conflict of interest.

This work was supported by the Spanish Ministry of Science and Innovation (BIO2009-10964 and Consolider E-Science, INB-Genoma España, and COMBIOMED RETICS), the European Research Council, and the Fundación Marcelino Botín. A.P. received support from a Juan de la Cierva postdoctoral fellowship.

REFERENCES

1. Collins, F. S., E. D. Green, …, M. S. Guyer, US National Human Genome Research Institute. 2003. A vision for the future of genomics research. *Nature.* 422:835–847.
2. Hélène, C. 1981. Recognition of base sequences by regulatory proteins in procaryotes and eucaryotes. *Biosci. Rep.* 1:477–483.
3. Schueler, M. G., and B. A. Sullivan. 2006. Structural and functional dynamics of human centromeric chromatin. *Annu. Rev. Genomics Hum. Genet.* 7:301–313.
4. Butcher, L. M., and S. Beck. 2008. Future impact of integrated high-throughput methylome analyses on human health and disease. *J. Genet. Genomics.* 35:391–401.
5. Bird, A. P. 1986. CpG-rich islands and the function of DNA methylation. *Nature.* 321:209–213.
6. McClelland, M., and R. Ivarie. 1982. Asymmetrical distribution of CpG in an 'average' mammalian gene. *Nucleic Acids Res.* 10:7865–7877.
7. Elango, N., and S. V. Yi. 2011. Functional relevance of CpG island length for regulation of gene expression. *Genetics.* 187:1077–1083.
8. Kundu, T. K., and M. R. Rao. 1999. CpG islands in chromatin organization and gene expression. *J. Biochem.* 125:217–222.
9. Doerfler, W. 1983. DNA methylation and gene activity. *Annu. Rev. Biochem.* 52:93–124.
10. Goll, M. G., and T. H. Bestor. 2005. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74:481–514.
11. Leonhardt, H., A. W. Page, …, T. H. Bestor. 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell.* 71:865–873.
12. Clark, T. A., I. A. Murray, …, J. Korlach. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res.* 40:e29.
13. Klimasauskas, S., S. Kumar, …, X. Cheng. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell.* 76:357–369.
14. Chandler, L. A., and P. A. Jones. 1988. Hypomethylation of DNA in the regulation of gene expression. *Dev. Biol.* 5(N Y 1985):335–349.
15. Cedar, H., and Y. Bergman. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* 10: 295–304.
16. Southern, E. M. 1984. DNA sequences and chromosome structure. *J. Cell Sci. Suppl.* 1:31–41.
17. Suzuki, M. M., and A. Bird. 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9:465–476.
18. Hermann, A., H. Gowher, and A. Jeltsch. 2004. Biochemistry and biology of mammalian DNA methyltransferases. *Cell. Mol. Life Sci.* 61:2571–2587.
19. Hermann, A., R. Goyal, and A. Jeltsch. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.* 279: 48350–48359.
20. Irizarry, R. A., C. Ladd-Acosta, …, A. P. Feinberg. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41: 178–186.
21. Doi, A., I. H. Park, …, A. P. Feinberg. 2009. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* 41:1350–1353.
22. Reference deleted in proof.
23. Lister, R., M. Pelizzola, …, J. R. Ecker. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* 462:315–322.
24. Bhutani, N., J. J. Brady, …, H. M. Blau. 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature.* 463:1042–1047.
25. Reference deleted in proof.
26. Choi, S. H., K. Heo, …, A. S. Yang. 2011. Identification of preferential target sites for human DNA methyltransferases. *Nucleic Acids Res.* 39:104–118.
27. Lavery, R., K. Zakrzewska, …, J. Sponer. 2010. A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.* 38:299–313.
28. Pérez, A., I. Marchán, …, M. Orozco. 2007. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.* 92:3817–3829.
29. Olson, W. K., A. A. Gorin, …, V. B. Zhurkin. 1998. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA.* 95:11163–11168.
30. Lankas, F., J. Sponer, …, T. E. Cheatham, 3rd. 2003. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.* 85:2872–2883.
31. Morozov, A. V., K. Fortney, …, E. D. Siggia. 2009. Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res.* 37:4707–4722.
32. Orozco, M., A. Noy, and A. Pérez. 2008. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr. Opin. Struct. Biol.* 18:185–193.
33. Pérez, A., F. Lankas, …, M. Orozco. 2008. Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res.* 36:2379–2394.
34. Lankas, F., J. Sponer, …, J. Langowski. 2000. Sequence-dependent elastic properties of DNA. *J. Mol. Biol.* 299:695–709.

35. Paillard, G., and R. Lavery. 2004. Analyzing protein-DNA recognition mechanisms. *Structure.* 12:113–122.

36. Pérez, A., F. J. Luque, and M. Orozco. 2012. Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.* 45:196–205.

37. Podtelezhnikov, A. A., C. Mao, …, A. Vologodskii. 2000. Multimerization-cyclization of DNA fragments as a method of conformational analysis. *Biophys. J.* 79:2692–2704.

38. Tolstorukov, M. Y., A. V. Colasanti, …, V. B. Zhurkin. 2007. A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning. *J. Mol. Biol.* 371:725–738.

39. Balasubramanian, S., F. Xu, and W. K. Olson. 2009. DNA sequence-directed organization of chromatin: structure-based computational analysis of nucleosome-binding sequences. *Biophys. J.* 96:2245–2260.

40. Thåström, A., P. T. Lowary, …, J. Widom. 1999. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* 288:213–229.

41. Deniz, O., O. Flores, …, M. Orozco. 2011. Physical properties of naked DNA influence nucleosome positioning and correlate with transcription start and termination sites in yeast. *BMC Genomics.* 12:489.

42. Anderson, J. D., and J. Widom. 2000. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *J. Mol. Biol.* 296:979–987.

43. Pérez, A., F. J. Luque, and M. Orozco. 2007. Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.* 129:14739–14745.

44. Behe, M., and G. Felsenfeld. 1981. Effects of methylation on a synthetic polynucleotide: the B—Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc. Natl. Acad. Sci. USA.* 78:1619–1623.

45. Banyay, M., and A. Gräslund. 2002. Structural effects of cytosine methylation on DNA sugar pucker studied by FTIR. *J. Mol. Biol.* 324: 667–676.

46. Schones, D. E., K. Cui, …, K. Zhao. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell.* 132:887–898.

47. Kaplan, N., I. K. Moore, …, E. Segal. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature.* 458:362–366.

48. Chung, H. R., and M. Vingron. 2009. Sequence-dependent nucleosome positioning. *J. Mol. Biol.* 386:1411–1422.

49. Cui, F., and V. B. Zhurkin. 2010. Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro. *J. Biomol. Struct. Dyn.* 27:821–841.

50. Travers, A., E. Hiriart, …, E. Di Mauro. 2010. The DNA sequence-dependence of nucleosome positioning in vivo and in vitro. *J. Biomol. Struct. Dyn.* 27:713–724.

51. Trifonov, E. N. 2010. Nucleosome positioning by sequence, state of the art and apparent finale. *J. Biomol. Struct. Dyn.* 27:741–746.

52. Olson, W. K., and V. B. Zhurkin. 2011. Working the kinks out of nucleosomal DNA. *Curr. Opin. Struct. Biol.* 21:348–357.

53. Field, Y., N. Kaplan, …, E. Segal. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLOS Comput. Biol.* 4:e1000216.

54. Bai, L., and A. V. Morozov. 2010. Gene regulation by nucleosome positioning. *Trends Genet.* 26:476–483.

55. Segal, E., Y. Fondufe-Mittendorf, …, J. Widom. 2006. A genomic code for nucleosome positioning. *Nature.* 442:772–778.

56. Lee, W., D. Tillo, …, C. Nislow. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* 39:1235–1244.

57. Tillo, D., and T. R. Hughes. 2009. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics.* 10:442.

58. Goñi, J. R., C. Fenollosa, …, M. Orozco. 2008. DNAlive: a tool for the physical analysis of DNA at the genomic scale. *Bioinformatics.* 24:1731–1732.

59. Goñi, J. R., A. Pérez, …, M. Orozco. 2007. Determining promoter location based on DNA structure first-principles calculations. *Genome Biol.* 8:R263.

60. Bertrand, H., T. Ha-Duong, …, B. Hartmann. 1998. Flexibility of the B-DNA backbone: effects of local and neighbouring sequences on pyrimidine-purine steps. *Nucleic Acids Res.* 26:1261–1267.

61. Nathan, D., and D. M. Crothers. 2002. Bending and flexibility of methylated and unmethylated EcoRI DNA. *J. Mol. Biol.* 316:7–17.

62. Mirsaidov, U., W. Timp, …, G. Timp. 2009. Nanoelectromechanics of methylated DNA in a synthetic nanopore. *Biophys. J.* 96:L32–L34.

63. Bell, A. C., and G. Felsenfeld. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature.* 405:482–485.

64. Clark, S. J., J. Harrison, and P. L. Molloy. 1997. Sp1 binding is inhibited by (m)Cp(m)CpG methylation. *Gene.* 195:67–71.

65. Hark, A. T., C. J. Schoenherr, …, S. M. Tilghman. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature.* 405:486–489.

66. Davey, C. S., S. Pennings, …, J. Allan. 2004. A determining influence for CpG dinucleotides on nucleosome positioning in vitro. *Nucleic Acids Res.* 32:4322–4331.

67. Pennings, S., J. Allan, and C. S. Davey. 2005. DNA methylation, nucleosome formation and positioning. *Brief. Funct. Genomics Proteomics.* 3:351–361.

68. Felle, M., H. Hoffmeister, …, G. Längst. 2011. Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Res.* 39:6956–6969.

69. Choy, J. S., S. Wei, …, T. H. Lee. 2010. DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.* 132:1782–1783.

70. Lee, J. Y., and T. H. Lee. 2012. Effects of DNA methylation on the structure of nucleosomes. *J. Am. Chem. Soc.* 134:173–175.

71. Lee, C. K., Y. Shibata, …, J. D. Lieb. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* 36: 900–905.

72. Field, Y., Y. Fondufe-Mittendorf, …, E. Segal. 2009. Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nat. Genet.* 41:438–445.

73. Choi, J. K., J. B. Bae, …, Y. J. Kim. 2009. Nucleosome deposition and DNA methylation at coding region boundaries. *Genome Biol.* 10:R89.

74. Feng, J., X. Dai, …, C. He. 2010. New insights into two distinct nucleosome distributions: comparison of cross-platform positioning datasets in the yeast genome. *BMC Genomics.* 11:33.

75. Kaplan, N., I. Moore, …, E. Segal. 2010. Nucleosome sequence preferences influence in vivo nucleosome organization. *Nat. Struct. Mol. Biol.* 17:918–920, author reply 920–912.

76. Valouev, A., J. Ichikawa, …, S. M. Johnson. 2008. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res.* 18:1051–1063.

77. Wang, X., G. O. Bryant, …, M. Ptashne. 2011. An effect of DNA sequence on nucleosome occupancy and removal. *Nat. Struct. Mol. Biol.* 18:507–509.

78. Esteller, M. 2006. Epigenetics provides a new generation of oncogenes and tumour-suppressor genes. *Br. J. Cancer.* 94:179–183.

79. Chodavarapu, R. K., S. Feng, …, M. Pellegrini. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature.* 466:388–392.