Supplementary Materials and Methods

Oligonucleotides

The oligonucleotides were synthesized and gel-purified by Sigma-Aldrich. As a positive control for multimerization-cyclization, we designed a 21bp sequence with 5'sticky ends derived from (1), which we called *PosSeq*: sense strand 5'-d(GAAAAAACGGGCGAAAAACGG) and antisense strand 5'-d(TCCCGTTTTTCGCCCGTTTTT) (here the underlined part of the sequence is complementary in both strands, with the overhanging bases at each end being complementary with the overhanging bases in the other strand, thus allowing for oligomerization). Conversely, as a negative control, we selected a oligo that was previously reported to have poor circularization properties (2), *NegSeq:* sense 5'-d(GCAAATATTGAAAAC) and antisense 5'-d(GCGTTTTCAATATTT).

The circularization favoring sequence was subsequently methylated at different positions, in a central cytosine on a CpG dinucleotide, *PosSeq_1met2*: sense 5'-d(GAAAAAACGGG**C**GAAAAACGG) and antisense 5'-d(TCCCGTTTTT**C**GCCCGTTTTT) (the C in bold-face corresponds to the methyl-cytosines). We also selected a set of molecules containing 2, 3 and 4 sequence repetitions of the 21bp *PosSeq* oligo (named *oligo 42*, *oligo 63* and *oligo 84*, respectively) as references for the size assignment.

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 37ºC in 1x T4 DNA Ligase Reaction Buffer. In case of radioactive phosphorylation, one picomole of DNA was mixed with 2.2 μmol of γ–32ATP (6 mCi/mL), 3U of T4 polynucleotide kinase and incubated for 1 hour at 37°C. The reaction mixture was purified using *MicroSpin G-25* columns (GE Healthcare) and resuspended in T4 DNA Ligase Reaction buffer.

Annealing and ligation

The complementary strands were denatured at 90 ºC 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 65ºC for 10 minutes. The DNA was ethanol precipitated and dissolved in 10 μl of DNAse free water.

Digestion

The ligation samples were then digested with 20U of *Exonuclease III* (New England Biolabs) for 30 minutes at 37ºC. The reaction was stopped by inactivating the nuclease at 65ºC for 20 minutes.

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 second dimension gel (8% polyacrylamide) to separate circular and linear molecules of 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 (Fig. S2).

Atomic Force Microscopy (AFM)

The ligation products were ethanol precipitated, resuspended in AFM Buffer (40mM HEPES, 10mM $MgCl_2$) and analyzed with AFM in tapping mode. AFM enabled us to have a rough estimate of the size of DNA molecules (see Fig. 2B) and confirmed the formation of minicircles in the experiments.

Linear and Circular DNA Size Determination

To determine the size of linear and circular DNA species more accurately, we performed individual circularization assays with *PosSeq*, *42*, *63* and *84* oligos, separately. Every sample was resolved in both 1D and 2D polyacrylamide gels (Fig. S3). By comparing the bands appearing or disappearing in the first and second dimension gels of the four samples, we were able to determine the exact size of every band. The actual size of the smallest circular molecules we obtained was determined to be 105 bp, with the following molecules increasing by 21 bp each.

Circular dichroism (CD) experiments

In order to rule out the hypothesis that massive cytosine methylation might induce a local change towards a Z-form of the DNA that could ultimately explain the functional role of CpG islands (4), we analyzed the secondary structure of the $d(CpG)_7$ oligonucleotide (with none or all the CpG steps methylated) by CD spectroscopy (Fig. S7). The DNA was exposed to an increasing ionic strength to force the B to Z transition. The transition B→Z was clearly favored in methylated DNA compared with the normal one, but in any case it occurred far beyond physiological conditions (Fig. S7), demonstrating that the transition to the Z-form is not the underlying principle for the physiological role of CpG methylation.

***In Vitro* Nucleosome Reconstitution**

Histone octamer purification

Histone octamers were purified from the nuclei of mammalian cells by a similar procedure described in detail elsewhere (5).

Accumulation of DNA substrate

The 146bp "601.2 sequence" (6, 7) was PCR amplified and cloned into the *PCR Script Amp+* vector (Stratagene), and subsequently used as a template for further amplification by PCR using *AccuPrime* Taq Polymerase (Invitrogen). Following PCR amplification, the oligo was agarose gel isolated and column purified (GE Healthcare). The DNA was ethanol precipitated, resuspended in water and quantified using *Nanodrop* spectrophotometer (Thermo Scientific)

In case of radioactive labeling, one picomole of DNA was mixed with 2.2 μmol of γ–32ATP (6 mCi/mL), 3U of T4 polynucleotide kinase and incubated for 1 hour at 37°C. The reaction mixture was purified using *MicroSpin G-25* columns (GE Healthcare) and resuspended in T4 DNA Ligase Reaction buffer.

CpG methylation of 601.2 sequence

The 601.2 sequence was methylated at all 12 CpG dinucleotides using the CpG methyltransferase MSssI (New England Biolabs). Briefly, 20μg of the 601 oligo was treated with 60U MSssI in the presence of 160 μM SAM (S-Adenosyl Methyionine) at 37ºC for 30 min. The reaction was replenished with another 60U MSssI and 160 μM SAM and incubated further for an hour at 37ºC. The reaction was stopped by incubation for 20 minutes at 65ºC. The CpG methylated oligo was purified from the enzyme using DNA purification columns (GE Healthcare). The localized methylation was verified by the sequencing of bisulfite treated DNA sample, which confirmed that all the cytosines in CpG dinucleotides were indeed methylated.

Full methylation of 601.2 sequence

Full cytosine methylation of the 601.2 sequence was achieved by replacing CTP with methyl-CTP (Fermentas) in the nucleotide mix during PCR amplification, using *Pfx* Taq Polymerase (Invitrogen). Subsequent agarose gel isolation, ethanol precipitation, and quantification were performed exactly as for the unmethylated 601 oligo.

Reconstitution reaction

Hydroxyapatite-purified histones were stored in aliquots in 2M NaCl/phosphate buffer pH 6.7 at -80ºC. Histones were thawed on ice and centrifuged 10,000 x g for 20 min at 4ºC to remove aggregates (improperly folded histones). Histones were subsequently quantified by measuring their absorption at 230nm.

The reconstitution reactions were prepared following a similar procedure as described in (8). All DNA and histones to be used in the reactions were freshly quantified immediately prior to use. Briefly, 750 ng of the respective forms of the 601.2 oligo (unmethylated control, CpG methylated, and full cytosine methylated)

were brought to 2M NaCl by adding an equal volume of 4M NaCl. The DNA was further mixed with histones at histone:DNA ratios ranging from 1.0-3.0 (w/w). The final volume of the reaction was brought to 35 µl using 2M NaCl / 50mM Tris pH 8.0 /1mM EDTA.

Each reconstitution reaction was mixed and transferred to a dialysis chamber (membrane 3,500 MWCO, Pierce). An initial volume of 100 ml of 2M NaCl /50mM Tris pH 8.0 / 1mM EDTA was diluted to 0.1M NaCl with continual addition of 50mM Tris pH 8.0 / 1mM EDTA to a final volume of 2 L using a peristaltic pump set at a flow rate of 40-60ml / hr at 4ºC. The dialyzed reaction was transferred to a microtube and stored at 4ºC.

Gel mobility shift assays

Nucleosome reconstitution was analyzed on 5% native polyacrylamide gels, which were pre-electrophoresed for 1hour at 100V at 4ºC in TBE. A 30% sucrose solution was added to the reconstitution reactions as a loading buffer immediately prior to loading the gel. The gels were run at 40 V for 6 hours at 4ºC and stained for 20 minutes in *SyBr Safe* (Invitrogen) diluted in 1x TBE. The intensities of the bands corresponding to nucleosome-bound or free DNA were subsequently estimated using the *ImageQuant* system (GE Healthcare) under UV light (see Fig. 4B).

In case of radioactive DNA, the bands were size-cut and the number of radiolabelled DNA molecules was then measured using a scintillation counter. In addition, the band intensities were also measured by densitometry using the *PhosphorImager* system (GE Healthcare) (see Fig. 4A).

Molecular Dynamics Simulations

MethylCytosine Parameters

Bonded parameters for ^{Me}C were based on the CHARMM27 parameters for ^{Me}C(9). Charges were derived from RESP fit at 6-31G** level starting from an optimized mC capped at the C1’ using the procedure described in reference (10). Library with parameters for d^{Me}C are available upon request.

Molecular Dynamics

DNA oligos were neutralized with Na+ counterions and hydrated with a octahedral box extending at least 10 Å in every direction from the DNA. The system was then optimized, thermalized (T = 298 K), and pre-equilibrated using a multistep protocol(11) in which we doubled the equilibration windows. The solvated systems were then allowed to evolve without restrains for an additional 100 ps before starting production runs.

Mesoscopic flexibility descriptors

Local parameters

MD simulations were projected into helical reference space to explore the sampled values of roll, twist, tilt, rise, shift and slide (12) for each of the ten unique base steps (plus methylated variants) in all unique tetramer environments. Equilibrium and stiffness parameters were obtained by following Lankas's procedure (13). Accordingly, MD trajectories were projected into a helical reference system to obtain equilibrium values and to derive the covariance matrix, which was then inverted following Einstein's equation to recover stiffness matrices, from which a mesoscopic estimate of the energy associated to a given deformation can be easily computed (see equation 1):

$$E = 0.5\sum_{i=1}^{6}\sum_{j=1}^{6} f_{ij}\Delta X_i \Delta X_j;\ \text{with the elements } f_{ij} \text{ corresponding to:}$$

$$\Theta = k_B T \subset^{-1} = \begin{pmatrix} k_w & k_{wr} & k_{wt} & k_{ws} & k_{wl} & k_{wf} \\ k_{wr} & k_r & k_{rt} & k_{rs} & k_{rl} & k_{rf} \\ k_{wt} & k_{rt} & k_t & k_{ts} & k_{tl} & k_{tf} \\ k_{ws} & k_{rs} & k_{ts} & k_s & k_{sl} & k_{sf} \\ k_{wl} & k_{rl} & k_{tl} & k_{sl} & k_l & k_{lf} \\ k_{wf} & k_{rf} & k_{tf} & k_{sf} & k_{lf} & k_f \end{pmatrix} \quad (1)$$

where k_b is the Boltzman constant, T is the absolute temperature, E is the energy associated to the deformation ΔX, and k stands for the different stiffness constants defining the 36 elements of the stiffness matrix (Θ) (twist (w), roll (r), tilt (t), rise

(s), slide (l) and shift (f)) at the dinucleotide level (in different tetramer environments) obtained by inversion of the MD-associated covariance matrix (⊂).

Global parameters

Global descriptors of equilibrium geometry and deformability were obtained for the different 18-mer oligonucleotides considering the 8mer central portion and Lavery's definitions (12) combined with Lankas's procedure (14). Descriptors considered here describe the ability of short tracts of DNA to bend in different directions (bending rigidity) and to undertwist/overtwist (twisting rigidity). They are calculated by considering fluctuations on this measures calculated at different oligomer lengths (up to a maximum of the full 18mer) following procedure described in (14), results shown in Fig. S1 are shown as the average for the twisting and bending rigidity of all possible 8mer combinations in each of the depicted 18mers.

Monte Carlo calculations of circularization efficiency

Theoretical J factors are calculated on the basis of the mesoscopic model described in methods (average and force constants for shift, slide, rise, tilt, roll and twist) and following a Monte Carlo procedure as described in the work of Olson and coworkers (15). Accordingly, J factors were calculated for PosSeq using different repeats of this sequence in the interval where circularization is successful experimentally (selected range of 126 to 210 bp). Since closed conformations are unlike to be sampled during the MC, efficiency was improved by: i.) Implementing Olson's gaussian sampling and ii.) using half-chain sampling enhancement. Both of them described in reference (15), labelling of an accepted conformation as closed was carried out again as described in reference (15). The J index is calculated as:

$$J = \frac{M_{closed}}{QM} \qquad (2)$$

where M_{closed} is the number of events in closed conformation, M is the number of conformations not closed and Q is defined in equation 11 of (15) as a normalization factor for the conditions imposed for determining a closed conformation. Note again

that since we are interested in relative rather than absolute J-factors arbitrary selections on when a conformation is “closed” do not affect our results.

Supplementary Figures

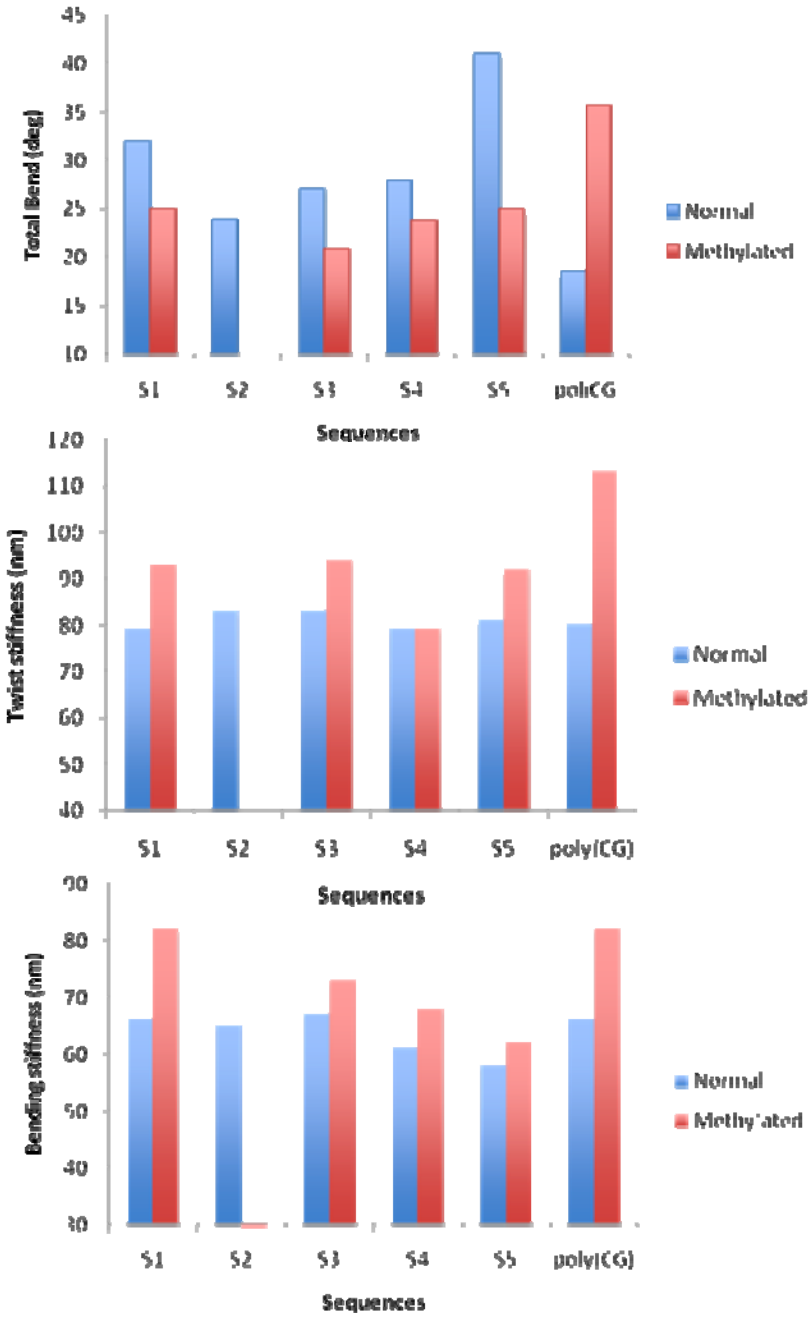

Figure S1. Description of the overall bend observed in 18mers in presence and absence of methylation (global bend as defined in (12)). Global twisting and bending rigidities are reported following work in (14). Results presented here show the rigidification of the sequence due to the presence of methylated d(CpG) steps. Note that sequence S2 (see Table S1 for definitions) has no d(CpG) steps.

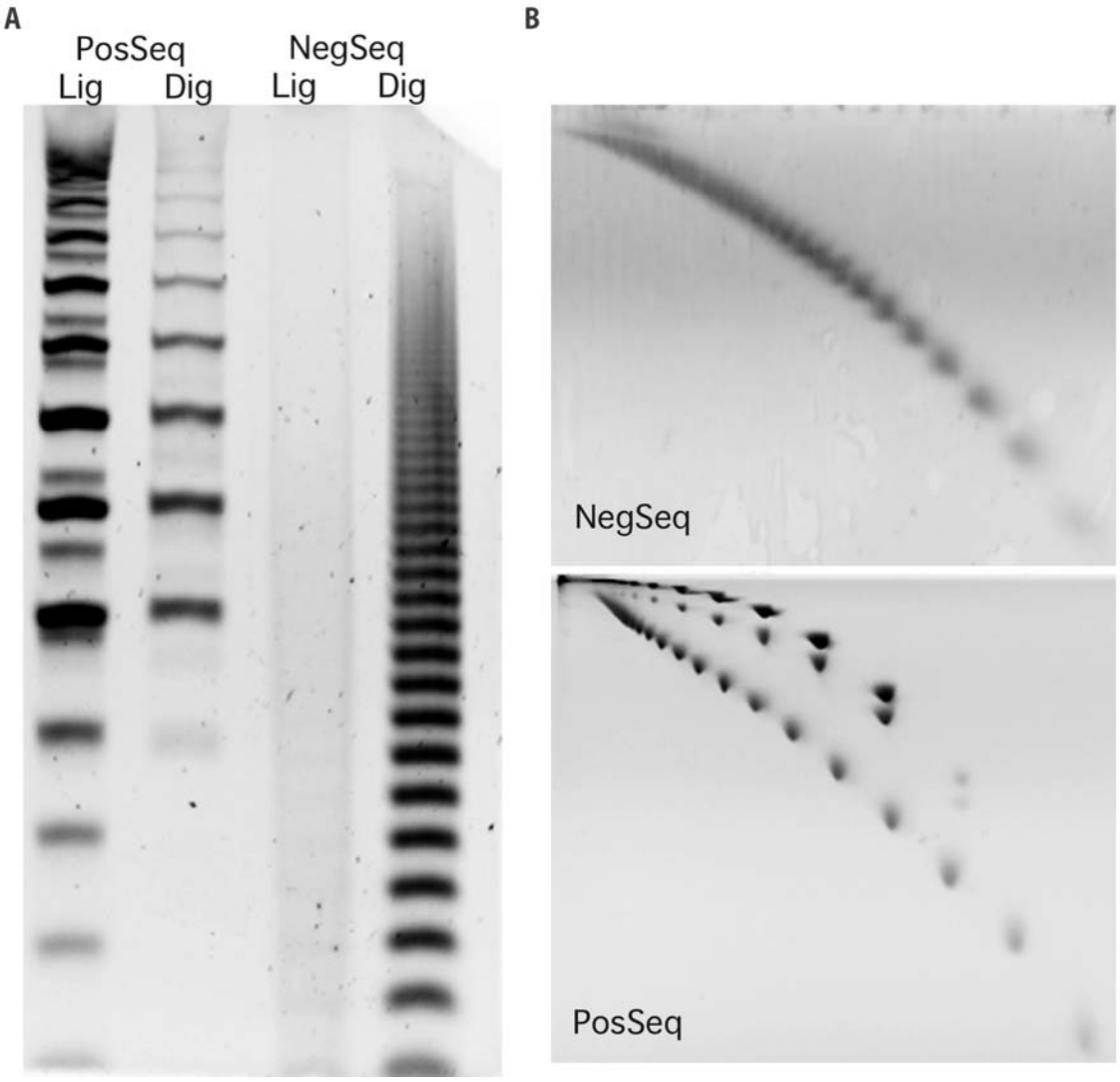

Figure S2. Circularization assays. (A) 5% polyacrylamide 1D-gel showing ligation (Lig) and digestion (Dig) products for the 21bp favoring (*PosSeq*) and 15bp non-favoring (*NegSeq*) cyclization oligonucleotides, respectively. The observed ladder patterns correspond to the different size multimers of the duplexes (both linear and circular). After digestion, only Exonuclease III-resistant DNA circles were observed. Thus, for the negative control sample, which is not expected to form circles, all the ligation products were digested. (B) 8% polyacrylamide 2D-native polyacrylamide gels showing the different migration of two species of DNA (linear and circular) from the ligation products. In the case of the positive control, linear DNA molecules are positioned on the lower diagonal, whereas closed and nicked circular DNA molecules in the upper diagonal. Conversely, for the negative control, only linear molecules are forming.

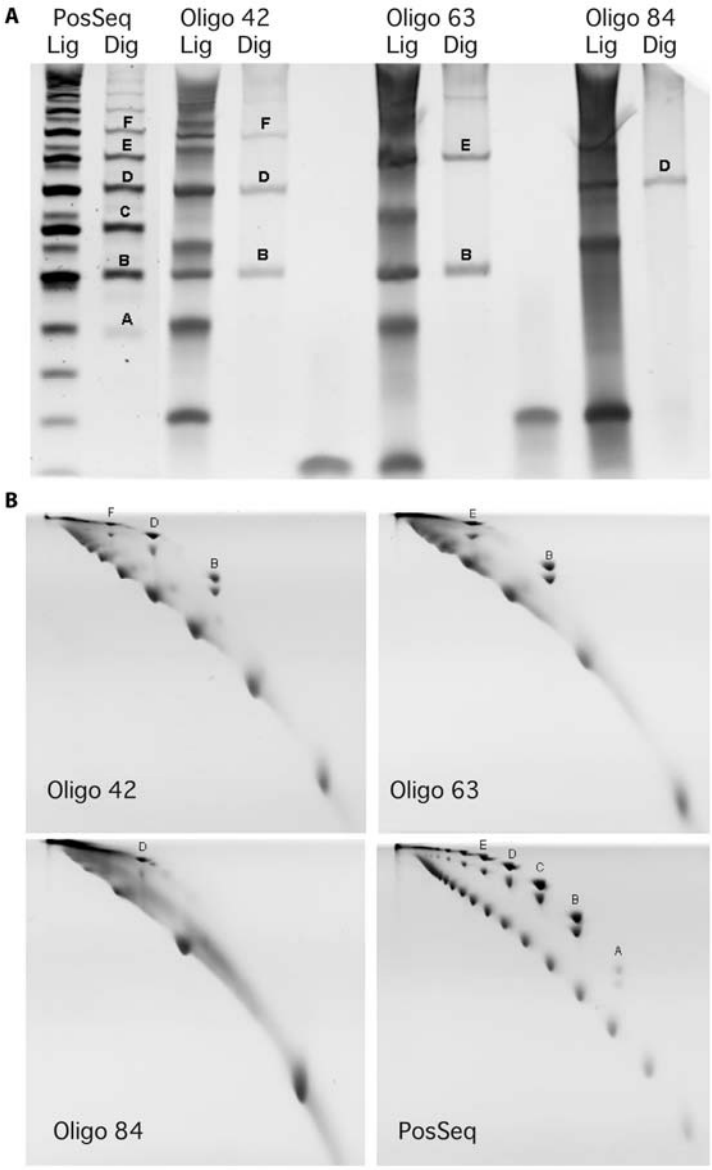

Figure S3. Circular and linear DNA size determination. (A) 1D gel showing ligation (Lig) and digestion (Dig) products for different size oligonucleotides. Lanes 1 and 2 contain *PosSeq* oligo ligation and digestion products, respectively; lanes 3, 6, 9: ligation products of oligo *42*, oligo *63* and oligo *84*, respectively; lanes 4, 7, 10: digestion products of the same oligos, respectively (corresponding to circular DNA); lanes 5, 8: duplexes of oligo *63* and oligo *84*. (B) 2D gels showing ligation products of oligo *42*, oligo *63*, oligo *84* and *PosSeq*, respectively. By comparing those bands appearing or disappearing on the different gels, we were able to calculate the exact size of each band.

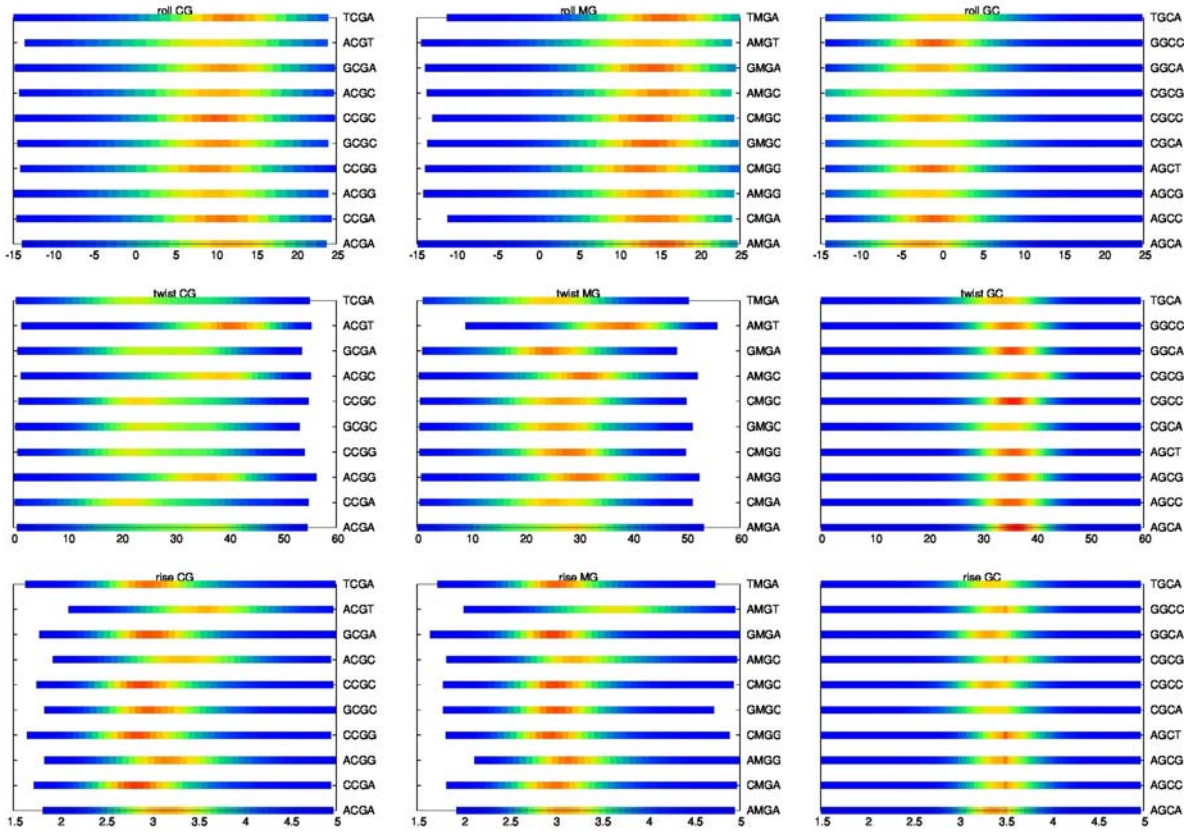

Figure S4. Histograms of roll(degree), twist(degree) and rise(Å) helical parameters for d(CpG), d(^{Me}CpG) and d(GpC). Each base pair step is described in all possible tetrad environments (denoted on the right hand axis 5'-ABCD-3'). Colours range from red (highly populated value) to blue (low populated).

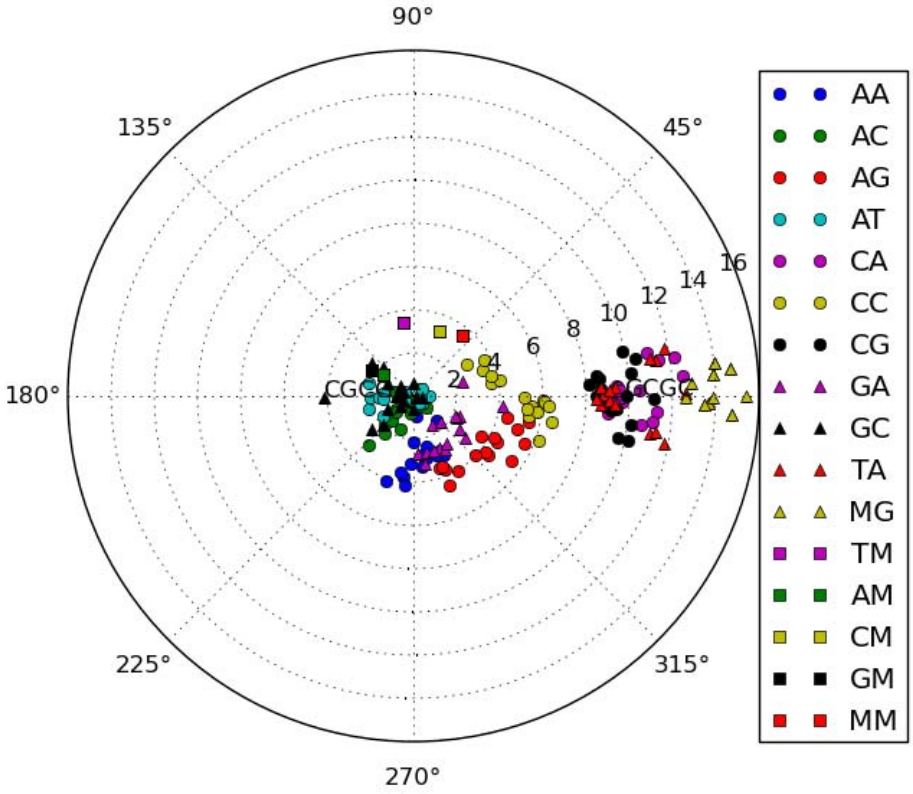

Figure S5. Bending direction and magnitude for the 10 base pair steps. Angles of ~0° denote bending towards the major groove and ~180° indicate bending towards the minor groove. For each base pair step, different data points represent the different nearest neighbour environments. A label has been added to d(GC) in d(CGCG) environment and d(CG) in d(GCGC) environment to note the counteracting effect leading to low overall curvature in poli(CpG) oligomers.

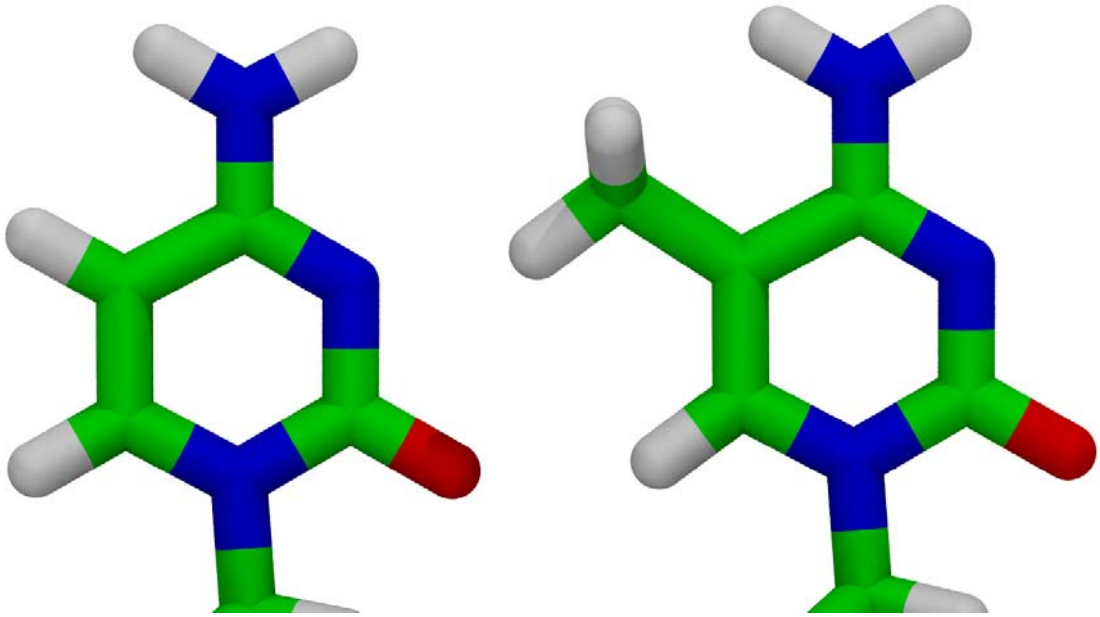

Figure S6. Chemical representation of Cytosine (left) and 5-methylCytosine (right). Color coding is: green for carbon atoms, red for oxygens, blue for nitrogens and white for hydrogens.

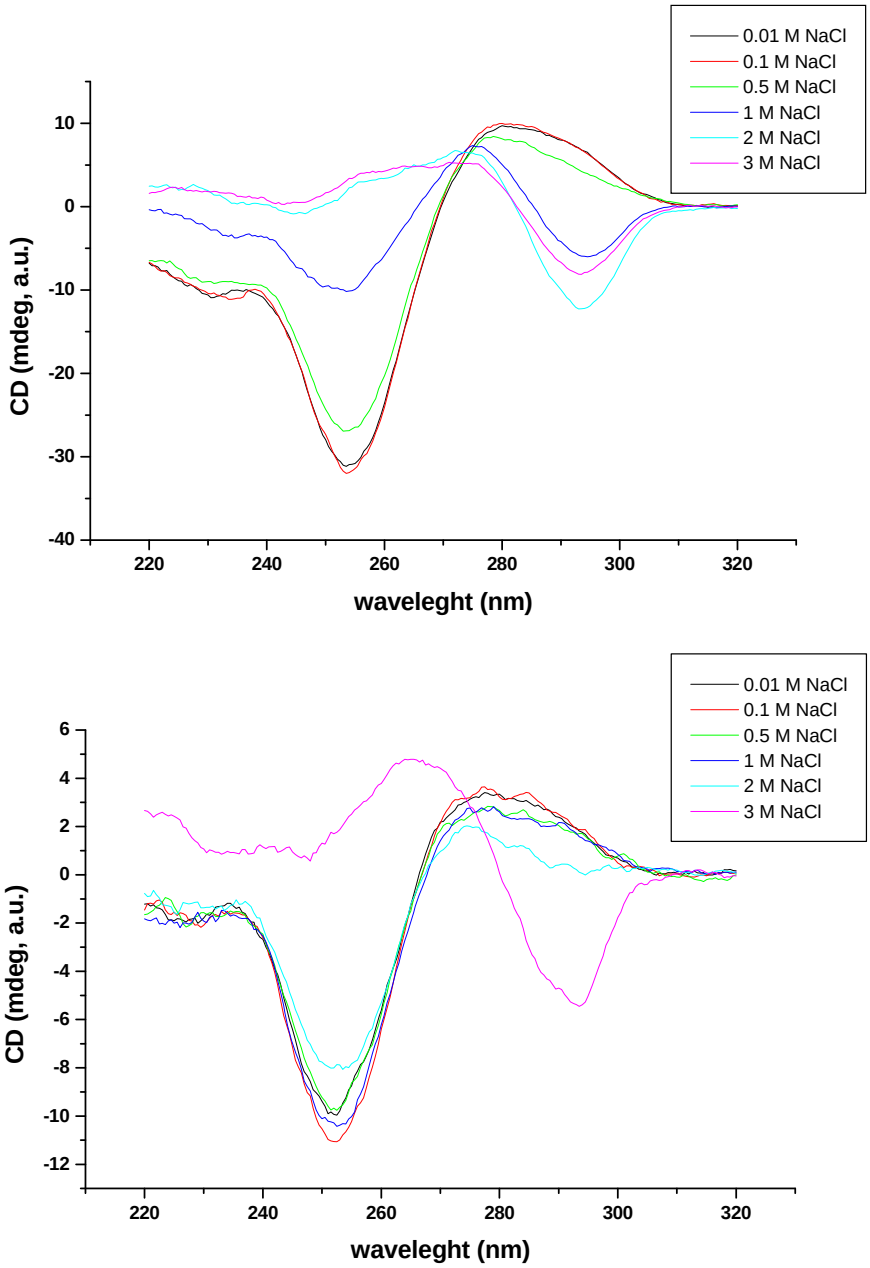

Figure S7. Circular Dichroism spectra monitoring the B to Z DNA transition for sequence d(^{Me}CpG)$_7$ (TOP) and d(CpG)$_7$ (BOTTOM): Curves are shown for different Ionic strength. The transition denoted by a shift in the peaks can be observed for concentrations much higher than those observed in physiological conditions, thus ruling out the change in conformation (B to Z) as the driving principle affecting the physical properties changes in d(CpG)n.

Supplementary Tables

Sequence name	Sequence identifier	Length	Sequence (5'→3')	Simulated time
Poli(MG)	P1/PM1	18mer	GMGMGMGMGMGMGMGMGM	100ns
Seq1*	S1	18mer	GCCTATAAA**MG**CCTATAA	100ns
Seq2*	S2	18mer	CTAGGTGGATGACTCATT	100ns
Seq3*	S3	18mer	CA**MG**GAAC**MG**GTTC**MG**TG	100ns
Seq4*	S4	18mer	GG**MGMG**CACCA**MGMGMG**G	100ns
Seq5	S5	18mer	GCCTATAAG**MGMG**TATAA	100ns
14MER-M1	M1	14mer	CGGA**MG**AC**MG**CGCG	100ns
14MER-M2	M2	14mer	CGGC**MG**AA**MG**CGCG	100ns
14MER-M3	M3	14mer	CGGA**MG**GG**MG**AGCG	100ns
14MER-M4	M4	14mer	CGGC**MG**GA**MG**TGCG	100ns
14MER-M5	M5	14mer	CGGG**MG**CT**MG**AGCG	100ns
ABC-dataset**	ABC	18mer	Multiple (39 different oligomers)	≥50ns

Table S1. Different sequences simulated for this study. The character M represents a methylated cytosine, all the corresponding non-methylated sequences were studied for the same length of time. The ABC dataset contains the most extensive set of trajectories and thus the most converged view on non-methylated sequences. Thus we considered its use for control of convergence in our dataset.
*These sequences correspond to the methylated counterparts of those published in reference 25 of the main text.
**These sequences correspond to the dataset generated in reference 20 of the main text.

Supplementary references

1. Podtelezhnikov, A. A., C. Mao, N. C. Seeman, and A. Vologodskii. 2000. Multimerization-cyclization of DNA fragments as a method of conformational analysis. Biophys J 79:2692-2704.
2. Payet, D., A. Hillisch, N. Lowe, S. Diekmann, and A. Travers. 1999. The recognition of distorted DNA structures by HMG-D: a footprinting and molecular modelling study. J Mol Biol 294:79-91.
3. Ulanovsky, L., M. Bodner, E. N. Trifonov, and M. Choder. 1986. Curved DNA: design, synthesis, and circularization. Proc Natl Acad Sci U S A 83:862-866.
4. Zacharias, W., A. Jaworski, and R. D. Wells. 1990. Cytosine methylation enhances Z-DNA formation in vivo. J Bacteriol 172:3278-3283.
5. Côté J., Utley R.T., and W. J.L.. 1995. Basic Analysis of transcription factor binding to nucleosomes. Methods Mol.Genet 6:108-129.
6. Lowary, P. T., and J. Widom. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol 276:19-42.
7. 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.
8. Steger, D. J., T. Owen-Hughes, S. John, and J. L. Workman. 1997. Analysis of transcription factor-mediated remodeling of nucleosomal arrays in a purified system. Methods 12:276-285.
9. MacKerell, A. D., Jr., N. Banavali, and N. Foloppe. 2000. Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56:257-265.
10. Cieplak, P., W. D. Cornell, C. Bayly, and P. A. Kollman. 1995. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. Journal of Computational Chemistry 16.
11. Shields, G. C., C. A. Laughton, and M. Orozco. 1997. Molecular dynamics simulations of the d(T center dot A center dot T) triple helix. Journal of the American Chemical Society 119:7463-7469.
12. Lavery, R., M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska. 2009. Conformational analysis of nucleic acids revisited: Curves+. Nucleic Acids Res 37:5917-5929.
13. Lankas, F., J. Sponer, J. Langowski, and T. E. Cheatham, 3rd. 2003. DNA basepair step deformability inferred from molecular dynamics simulations. Biophys J 85:2872-2883.
14. Lankas, F., J. Sponer, P. Hobza, and J. Langowski. 2000. Sequence-dependent elastic properties of DNA. J Mol Biol 299:695-709.
15. Czapla, L., D. Swigon, and W. K. Olson. 2006. Sequence-dependent effects in the cyclization of short DNA. J Chem Theory Comput 2:685-695.