Structural Dynamics of DNA Depending on Methylation Pattern

Takeru Kameda,[1,2] Miho M. Suzuki,[3] Akinori Awazu,[4] and Yuichi Togashi[4,5,*]

[1] *Graduate School of Science, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan*
[2] *RIKEN Center for Biosystems Dynamics Research (BDR), Wako, Saitama, Japan*
[3] *Graduate School of Medicine, Nagoya University, Nagoya, Aichi, Japan*
[4] *Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan*
[5] *RIKEN Center for Biosystems Dynamics Research (BDR), Higashi-Hiroshima, Hiroshima, Japan*

(Dated: November 4, 2020)

Abstract

DNA methylation is associated with a number of biological phenomena, and plays crucial roles in epigenetic regulation of eukaryotic gene expression. It is also suggested that DNA methylation alters the mechanical properties of DNA molecules, which is likely to affect epigenetic regulation. However, it has not been systematically investigated how methylation changes the structural and dynamic features of DNA. In this research, to elucidate the effects of methylation on DNA mechanics, a fully atomic molecular dynamics simulation of double-stranded DNA with several methylation patterns was performed. Through the analysis of the relative positioning of the nucleotides (base-step variables), characteristic changes in terms of local flexibility were observed, which further affected the overall DNA geometry and stiffness. These findings may serve as a basis for a discussion on methylation-dependent DNA dynamics in physiological conditions.

I. INTRODUCTION

Methylation is a universal biochemical modification of DNA [1–3]. A methyl group ($-CH_3$) binds to the cytosine base in a CpG dinucleotide, referred to as mCpG [1]. DNA methylation is a widespread and crucial epigenetic modification that influences gene expression [2, 4]. Vertebrate genomes are globally methylated at CpG dinucleotides; only CpG islands (CGI) are unmethylated [5]. CGIs are short unique DNA sequences that are over 200 bp (typically 500–1000 bp) in length and have high G+C content, as well as high CpG density [6]. The promoter regions of 60–70 % of all vertebrate genes overlap with CGIs [7, 8]. CGIs eliminate nucleosomes and attract proteins that create a transcriptionally permissive chromatin state [6, 9]. Whereas, when DNA methylation occurs at CpG sites, the CGI-promoter strongly silences transcription by inhibiting transcriptional factor binding or recruitment of mCpG binding proteins and co-repressor complexes [10]. It has been proposed that methylated CGI functions to stably maintain a silenced state of gene transcription (called the "long-term locking model") [4].

DNA methylation affects the mechanical properties of DNA. For example, DNA methylation reduces the flexibility of DNA [11], inhibits or facilitates DNA strand separation [12], attracts homologous DNA segments [13], and changes the dynamics of nucleosomes [14]. DNA methylation is comparable to the sequence-dependent mechanics of DNA [15, 16], which affects basal processes of molecular recognition [17]. The extremely opposite roles of CGIs in transcriptional regulation are likely to be defined by the mechanical properties of hyper- and hypo-methylated DNA. However, how DNA methylation changes the structural and dynamic features of DNA molecules remains poorly understood. Although effects of methylation in general on DNA dynamics have been recently studied using molecular dynamics (MD) simulations [18–22], there still remain microscopic properties (e.g. positional effects and their ranges) to be elucidated; thus, systematic analysis of different patterns of methylation is desired.

In this research, the relationship between methylation patterns and DNA dynamics was investigated using all-atom MD simulations. The structures of DNA with several methylation patterns were sampled in equilibrium. Through the analysis of fluctuations in relative positioning of nucleotides, characteristic changes of base-step parameters (BSP) in terms of local flexibility were identified, which further affected overall DNA dynamics.

II. MATERIALS AND METHODS

A. Simulation Procedure

All-atom MD simulations were employed for double-stranded DNA dynamics in solution. DNA structures were constructed using X3DNA [23, 24] with a helical parameter set obtained by *in vitro* experiments and X-ray crystal structure analysis [15, 25, 26]. The atomic coordinates of double-stranded DNA were generated according to a given DNA sequence (see [23, 24] for details). In this study, for simplicity, repeats of CpG dinucleotides were used, and regular patterns of methylation were introduced, as shown in Tab. I. These DNA models were soaked in a 60 Å × 60 Å × 200 Å water box, neutralized by K^+, and added 150 mM KCl. TIP3P water model was

* togashi@hiroshima-u.ac.jp

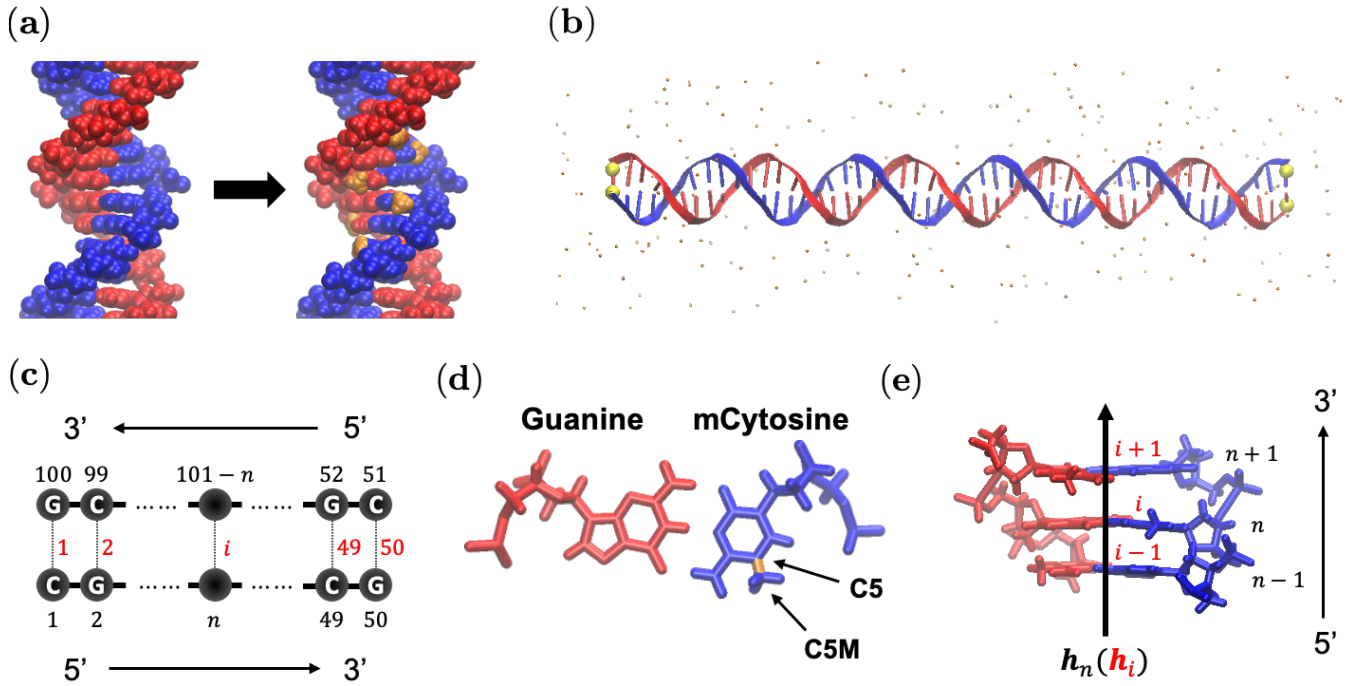

FIG. 1. Overview of the Model. (a) Structure of DNA with poly-CpG (left) and poly-mCpG (right). The methyl groups are shown in orange. (b) Solvated DNA model. Small particles represent ions. C1' atoms located at both ends (shown in yellow) are restrained in the simulation (see Sec. II A). (c) Nucleotide indices n (shown in black) and base-pair indices i (shown in red) of the double-stranded DNA. (d) Locations of C5 and C5M atoms in 5-methylated cytosine. (e) Schematic representation of helical axis $\boldsymbol{h}_n$ ($\boldsymbol{h}_i$).

employed. VMD [27] was used to infer missing atom coordinates, solvate the model, and visualize the structure throughout the study.

All simulations were performed using NAMD (version 2.13 multi-core with CUDA) [28, 29]. The CHARMM36 force-field (July 2017 update) was used and a 5MC2 nucleotide patch was applied to introduce 5-methylated cytosine (mC) (Fig. 1 (a)). A periodic boundary condition with Particle-Mesh Ewald electrostatics was employed and a cutoff of 12 Å (with switching from 10 Å) was used for non-bonded interactions. Temperature and pressure were set at 300 K and 1 atm, respectively; a Langevin thermostat (damping coefficient: 5/ps) and Langevin-piston barostat were adopted. After energy minimization (10,000 steps), the system was equilibrated for 10 ns, and then simulated for 50 ns (time-step: 2 fs). The conformation was sampled over the last 50 ns at 10 ps intervals (i.e. 5,000 snapshots), and each DNA model was simulated 10 times (i.e. 50,000 snapshots in total).

Through the equilibration and production run, harmonic restraints were applied to C1' atoms of nucleotides at both ends of the DNA segment (Fig. 1 (b); spring constant: 1.0 pN/Å, centered at the position after the energy minimization), to prevent rotation of DNA. It was confirmed that the restraints did not excessively stretch the DNA and allowed bending (Fig. 2 (c)).

B. Overall Geometry of Double-Stranded DNA

Overall structural variation of the DNA model was characterized by the ratios of the square root of the three principal components of the distribution of atomic positions (except hydrogen), $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, and $\sqrt{\lambda_3}$ ($\sqrt{\lambda_1} > \sqrt{\lambda_2} \geq \sqrt{\lambda_3} > 0$). Here, λ_1 to λ_3 were obtained as eigenvalues of the covariance matrix I:

$$I := \begin{pmatrix} \langle(\Delta x_k)^2\rangle_k & \langle\Delta x_k \Delta y_k\rangle_k & \langle\Delta x_k \Delta z_k\rangle_k \\ \langle\Delta y_k \Delta x_k\rangle_k & \langle(\Delta y_k)^2\rangle_k & \langle\Delta y_k \Delta z_k\rangle_k \\ \langle\Delta z_k \Delta x_k\rangle_k & \langle\Delta z_k \Delta y_k\rangle_k & \langle(\Delta z_k)^2\rangle_k \end{pmatrix} \quad (1)$$

where $(\Delta x_k, \Delta y_k, \Delta z_k) = (x_k - \langle x_m\rangle_m, y_k - \langle y_m\rangle_m, z_k - \langle z_m\rangle_m)$; (x_k, y_k, z_k) is the position of k-th atom, $(\langle x_m\rangle_m, \langle y_m\rangle_m, \langle z_m\rangle_m)$ is the center position of a given DNA molecule, and $\langle ...\rangle_m$ indicates the average for all ms. Note that $(\lambda_1, \lambda_2, \lambda_3)$ were defined for each snapshot τ, and thus 5,000 data-points were obtained in each simulation trajectory (see Sec. II A).

In this case, λ_1 and λ_2 corresponded to the extension and distortion of double-stranded DNA, respectively.

C. Persistence Length of Double-Stranded DNA

The persistence length of polymers statistically characterize their stiffness [31], and the case of double-stranded

TABLE I. Target Sequences. For each model, methylated CpG (mCpG) content ratio, methylated cytosine (mC) sites, and repeat units are listed. All sequences are composed of only CpG dinucleotides. mC sites are indexed as shown in Fig. 1 (c). Note that the models, except SEQ 0.00 and SEQ 1.00, include partial (fractional) sequences at the ends (see Tab. S1 [30]).

Model	mCpG Content	mC Site (Nucleotide Indices)	Repeat Unit
SEQ 0.00	0% (0)	-	CG
SEQ 0.25	25% (1/4)	1 9 17 25 33 41 49 51 59 67 75 83 91 99	CGCGmCGCG
SEQ 0.33	33% (1/3)	1 7 13 19 25 31 37 43 49 51 57 63 69 75 81 87 93 99	CGmCGCG
SEQ 0.50	50% (1/2)	1 5 9 13 17 21 25 29 33 37 41 45 49 51 55 59 63 67 71 75 79 83 87 91 95 99	CGmCG
SEQ 0.67	67% (2/3)	3 5 9 11 15 17 21 23 27 29 33 35 39 41 45 47 53 55 59 61 65 67 71 73 77 79 83 85 89 91 95 97	CGmCGmCG
SEQ 0.75	75% (3/4)	3 5 7 11 13 15 19 21 23 27 29 31 35 37 39 43 45 47 53 55 57 61 63 65 69 71 73 77 79 81 85 87 89 93 95 97	CGmCGmCGmCG
SEQ 1.00	100% (1)	1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99	mCG

DNA has been studied well [32–34]. In this research, the persistence length of DNA was defined as follows. First, A_d, the inner-product of helical axes separated by d nm ($d \in [0.3, 10.0]$ at intervals Δd ($= 0.01$)) was defined as

$$A_d := \left\{ \boldsymbol{h}_i \cdot \boldsymbol{h}_j \mid d - \frac{1}{2}\Delta d \leq C_{ij} < d + \frac{1}{2}\Delta d \right\}$$
$$C_{ij} := \sum_{l=i}^{j-1} |\boldsymbol{c}_{l+1} - \boldsymbol{c}_l| \tag{2}$$
$$(11 \leq i < j \leq 40)$$

where $\boldsymbol{c}_i$ is the center of base pair i, and $\boldsymbol{h}_i$ is the helical axis of base-pair i (Fig. 1 (e)). These vectors were determined using X3DNA [23, 24]. The length of $\boldsymbol{h}_i$ was equal to 1. To eliminate the effects of the boundaries and restraints (see Sec. II A), 10 base pairs from each end were eliminated (the base-pair indices (i, j) satisfy $11 \leq i < j \leq 40$). Then, the dataset $\{(d, \langle H \rangle_{H \in A_d})\}$ ($\langle ... \rangle_H$, which indicates the average for all Hs, sampled over all 5,000 snapshots for each trial) was fitted by function $f(d) := \exp(-d/L_p)$, where L_p is the persistence length to be estimated.

D. Local Geometry and Flexibility

Profiles of base-step geometry were obtained using X3DNA [23, 24], from which the local flexibility was estimated. First, the displacement in terms of mode represented by each base-step parameter (BSP) from the (meta)stable conformation was examined. The score $X_i^\bullet(\tau)$ for BSP $\bullet$ of base-step i (base pair indices from i to $i+1$ shown in Fig. 1 (c)) at the τ-th frame, and its displacement $\hat{X}_i^\bullet(\tau) := X_i^\bullet(\tau) - X_i^\bullet(0)$ were calculated, where $X_i^\bullet(0)$ is the score $X_i^\bullet$ just after the energy minimization ($\tau = 0$) (see Sec. II A). Then, the flexibility score $\sigma_i^\bullet$ for BSP $\bullet$ of base-step i was defined as the standard deviation (S.D.) of $\hat{X}_i^\bullet(\tau)$, either in each simulation trial or over 10 trials for the same model.

Additionally, to quantify the positional effects of methylation (i.e. effects depending on the relative position to the mCpG sites) on the structural variation, the flexibility scores $\sigma_i^\bullet$ were aligned and averaged in terms of the repeat unit (Tab. I), to obtain $\overline{\sigma_l^\bullet}$ defined in Eq. (3). The repeat units and partial (fractional) sequences at the ends for each model are shown in Tab. S1 [30]. To eliminate boundary effects, 10 base-pairs were omitted from each end, and only complete repeat units within the central 30 base-pairs were analyzed (for base-step parameters, the base-step between the last base-pair to the next one was also included). For example, for SEQ 0.33, five iterations of CGmCGCG in the center (and the next C for base-step parameters) were used. The averaged scores $\overline{\sigma_l^\bullet}$ at base-step l in the repeat unit were calculated as

$$\overline{\sigma_l^\bullet} := \langle \sigma_i^\bullet \rangle_{i \sim l} \tag{3}$$

where $\langle \cdots \rangle_{i \sim l}$ represents the average for all base-step is that correspond to position l in the repeat unit (Tab. I). $\overline{\sigma_l^\bullet}$ were calculated for each trial, and then their mean and S.D. over 10 trials were evaluated, for which Welch's t-test was applied.

Furthermore, relationships between $\hat{X}_i^\bullet(\tau)$ were evaluated. Specifically, (i) the same BSP $\bullet$ between two base-steps, and (ii) two BSPs $\bullet$ of the same base-step, were compared. For case (i), correlation coefficient of BSP $\bullet$ between base-steps i and j was calculated as

$$\frac{\langle (\hat{X}_i^\bullet(\tau) - \langle \hat{X}_i^\bullet(\tau) \rangle_\tau)(\hat{X}_j^\bullet(\tau) - \langle \hat{X}_j^\bullet(\tau) \rangle_\tau) \rangle_\tau}{\sigma_i^\bullet \sigma_j^\bullet} \tag{4}$$

where $\langle \cdots \rangle_\tau$ indicates the average for all τs over 10 trials (50,000 samples in total, for which $\sigma_i^\bullet$ and $\sigma_j^\bullet$ were also calculated) for each model. For case (ii), correlation coefficient between two individual BSPs $\bullet_1$ and $\bullet_2$ (of base-step i) was calculated as

$$\frac{\langle (\hat{X}_i^{\bullet_1}(\tau) - \langle \hat{X}_i^{\bullet_1}(\tau) \rangle_\tau)(\hat{X}_i^{\bullet_2}(\tau) - \langle \hat{X}_i^{\bullet_2}(\tau) \rangle_\tau) \rangle_\tau}{\sigma_i^{\bullet_1} \sigma_i^{\bullet_2}} \tag{5}$$

where $\langle \cdots \rangle_\tau$ is the same as in Eq. (4). Then, it was averaged over the central 30 base-steps ($11 \leq i \leq 40$).

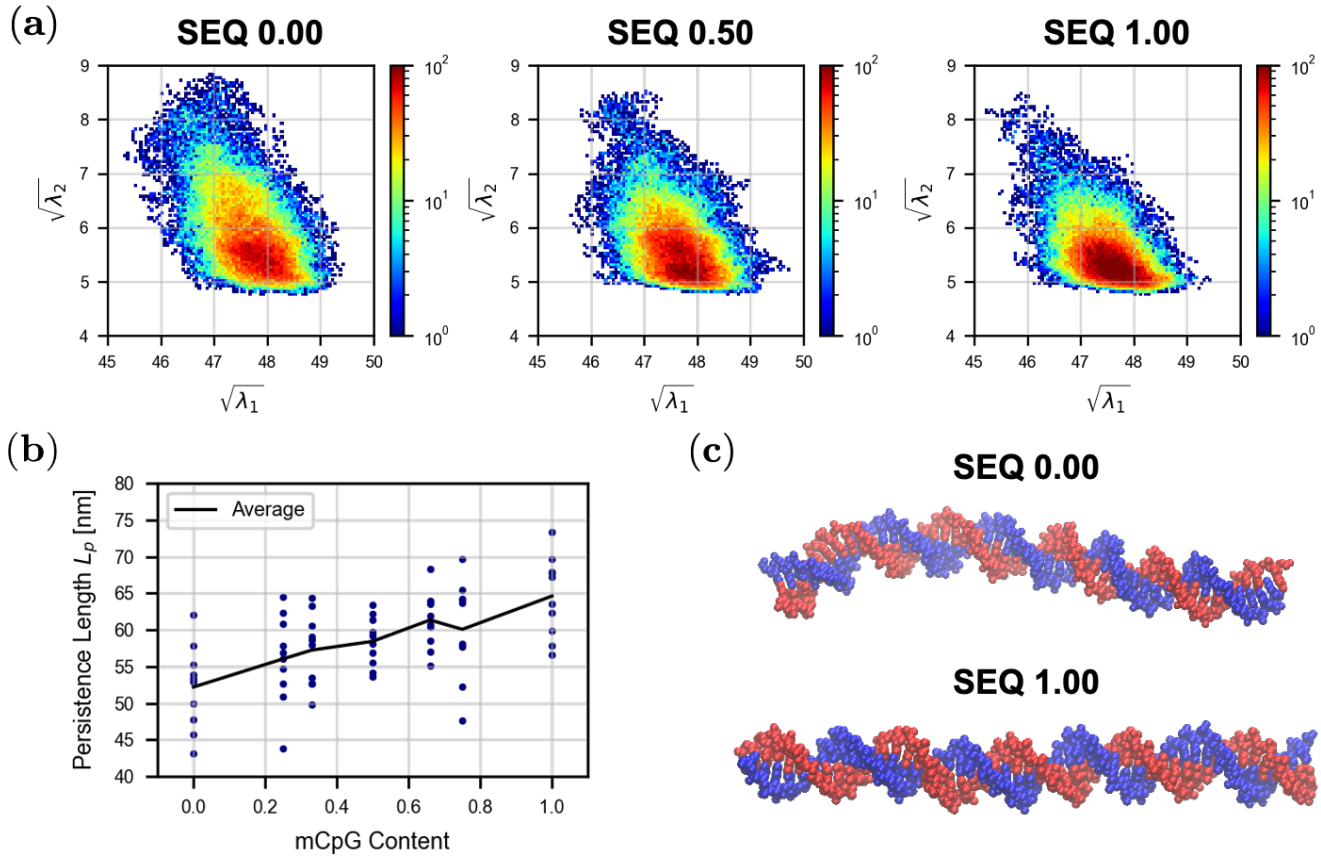

FIG. 2. Profiles of Overall Geometry and the Persistence Length of DNA. (a) Overall geometry. The distribution of ($\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$) summed over 10 trials is shown. (b) Persistence length L_p. The average within each trial and over 10 trials are shown. The horizontal axis corresponds to the mCpG content (Tab. I). p-values (by Welch's t-test) between mCpG content 0.0 and 0.5, 0.5 and 1.0, and 0.0 and 1.0 are 9×10^{-3}, 8×10^{-3}, and 1×10^{-4}, respectively. (c) Snapshots of the structure of SEQ 0.00 and SEQ 1.00 models at 40 ns.

E. Dynamics of Methyl Groups

The relative orientation of each methyl group was evaluated by the angle between the direction of the methyl group and the plane formed by the base-pair. The angle θ_n was defined as follows:

$$\begin{aligned} \theta_n &:= \theta_n^m - \theta_n^p \\ \theta_n^m &:= \cos^{-1}(\boldsymbol{h}_n \cdot \boldsymbol{d}_n^m) \\ \theta_n^p &:= \cos^{-1}(\boldsymbol{h}_n \cdot \boldsymbol{d}_n^p) \end{aligned} \tag{6}$$

where $\boldsymbol{h}_n$ is the vector of the helical axis at the base-pair where nucleotide n belongs (Fig. 1 (e)), obtained using X3DNA [23, 24], $\boldsymbol{d}_n^m$ is the direction vector of the methyl group (from C5 to C5M shown in Fig. 1 (d)), and $\boldsymbol{d}_n^p$ is the direction vector from the C1' atom to the C1' atom in the complementary base. The lengths of these three vectors were equal to 1. If $\theta_n > 0$ ($\theta_n < 0$), the methyl group tilted toward the next (prior) base.

Additionally, physical contact (interaction) frequency between the methyl group and its neighbor nucleotides was evaluated. A methyl group and a nucleotide were regarded in contact if their shortest atomic distance was less than 3.0 Å.

III. RESULTS

A. Profiles of Double-Stranded DNA Stiffness

The overall geometry of DNA was represented by the extension ($\sqrt{\lambda_1}$) and distortion ($\sqrt{\lambda_2}$) of its atomic coordinates. The distribution of $\sqrt{\lambda_1}$ and $\sqrt{\lambda_2}$ for each model is shown in Figs. 2 (a) and S1 [30]. Mostly, ($\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$) was in the range of $47 \leq \sqrt{\lambda_1} \leq 49$ and $5 \leq \sqrt{\lambda_2} \leq 6$, corresponding to a conformation similar to the shape of stretched DNA (e.g. Fig. 1 (b)).

The distribution outside of this range was different for these models. The probability within this range was higher for SEQ 1.00 than that for SEQ 0.00. $\sqrt{\lambda_2}$ of SEQ 1.00 tended to be smaller than that of SEQ 0.00. This implied that high mCpG content resulted in a stretched, or stiffer DNA shape. The other models (Tab. I) showed intermediate stiffness between SEQ 0.00 and SEQ 1.00

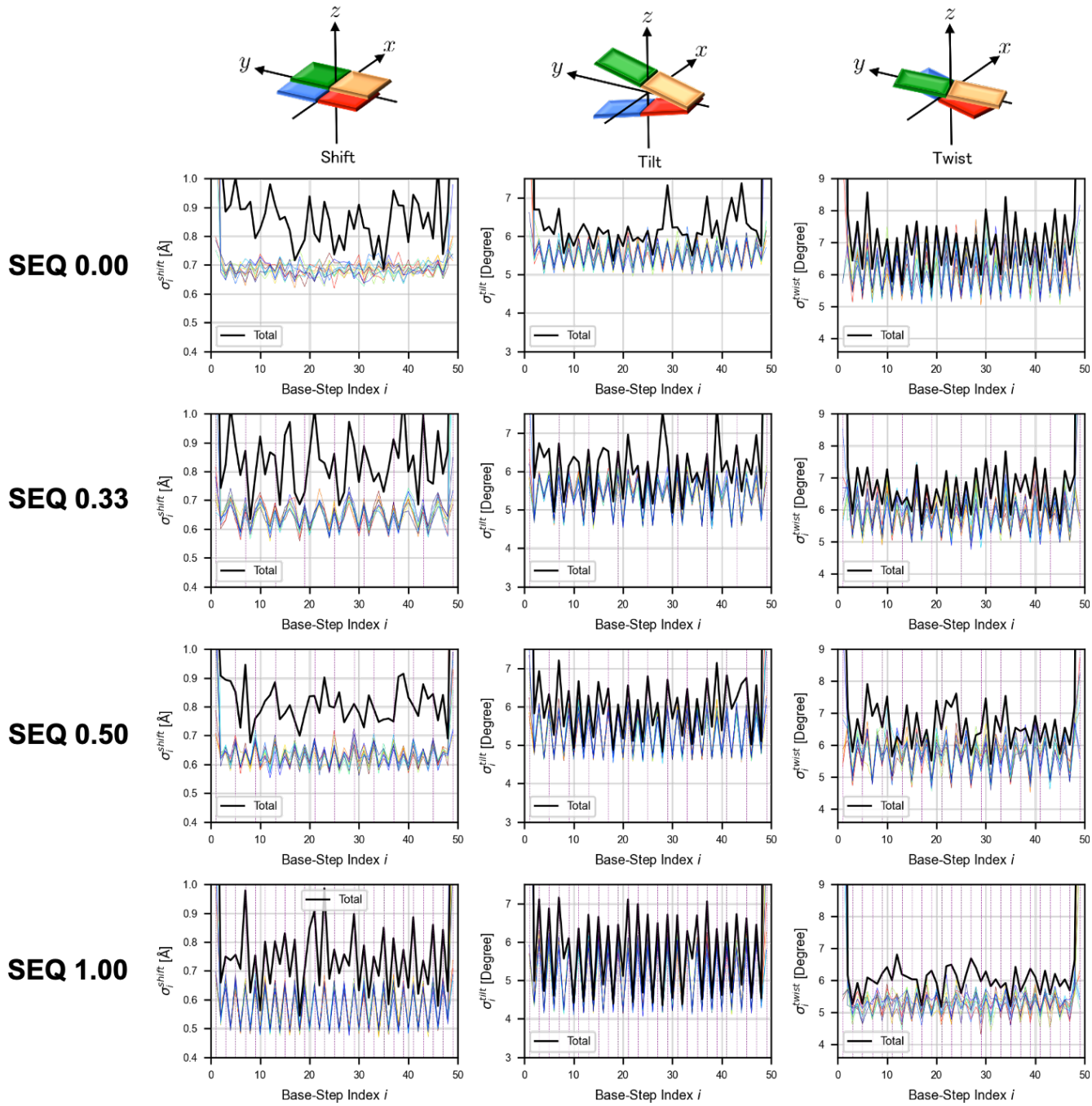

FIG. 3. Profiles of σ_i^{shift}, σ_i^{tilt}, and σ_i^{twist}. Colors show different simulation trials and the black line shows the overall flexibility (structural variation) in 10 trials. In the horizontal axis, odd and even numbers correspond to base steps of C→G and G→C, respectively (see Fig. 1 (c)). Purple lines show mC→G (C→G in the case of mC) base steps.

(Figs. 2 (a) and S1 [30]). The variation of $\sqrt{\lambda_2}$ was suppressed as the mCpG content increased. These results showed that higher methylation content stiffened DNA in equilibrium conditions.

The persistence length L_p (see Sec. II C) for each model is shown in Fig. 2 (b), which increases approximately linearly with mCpG content. Though estimated L_p values showed large variation, this trend was significant (e.g., $p < 1 \times 10^{-4}$ between SEQ 0.00 and SEQ 1.00). Note that the typical persistence length of DNA is approximately 50 nm ($\sim$ 150 base-pairs), with which the estimated values are consistent.

The difference in DNA stiffness can also be observed in the snapshots, as shown in Fig. 2 (c).

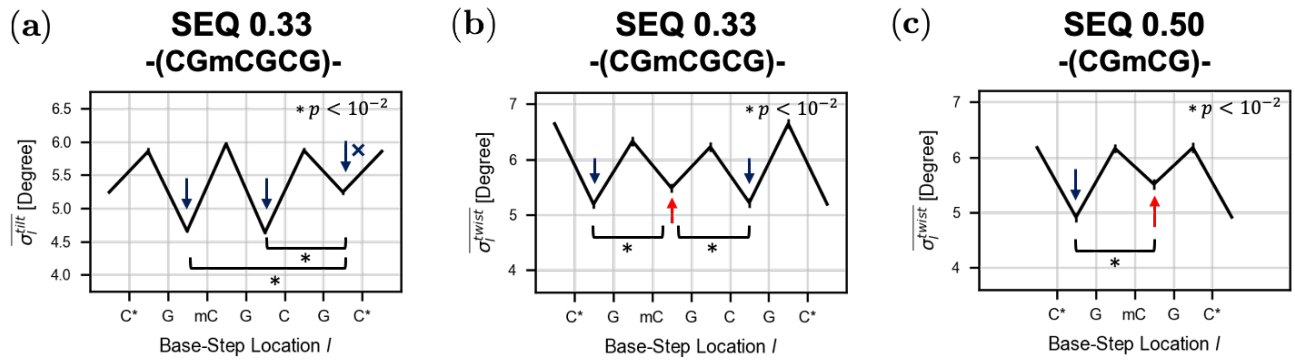

FIG. 4. Profiles of $\overline{\sigma_l^{\bullet}}$ for the Repeat Unit. σ_i^{tilt} and σ_i^{twist} (shown in Fig. 3) are aligned and averaged in terms of the repeat unit (Tab. I). Error bars show mean ± S.D. over 10 trials; see Sec. II D for the method. As these profiles are for base-step parameters, data points are shown between two consecutive bases. Nucleotides at both ends (shown with * in the horizontal axis) are identical. (a) $\overline{\sigma_l^{tilt}}$ of SEQ 0.33 (CGmCGCG, averaged over five iterations); (b) $\overline{\sigma_l^{twist}}$ of SEQ 0.33 (CGmCGCG, averaged over five iterations); (c) $\overline{\sigma_l^{twist}}$ of SEQ 0.50 (CGmCG, averaged over seven iterations). In (a), reduction of S.D. at G→C is significantly larger at both neighbors of mCpG (* $p = 3 \times 10^{-16}$ each). In (b) and (c), S.D. at mC→G is significantly higher than unmethylated C→G (* $p = 1 \times 10^{-7}$ (b; left), 1×10^{-6} (b; right), and 6×10^{-11} (c)).

B. Local Base-Step Flexibility

The overall DNA geometry analyzed above is a result of sequential base-pair stacking [35, 36]. To investigate the effects of methylation in detail, local profiles of flexibility at each base-step were evaluated. Through the analysis using X3DNA (see Sec. II D), profiles of BSPs were obtained (Fig. S2 [30]). Profiles of $\sigma_i^{\bullet}$ (Fig. S2 [30]) changed with increase of mCpG content; among them, shift, tilt, and twist (shown in Fig. 3) were the most significantly affected.

In the profiles of σ_i^{shift} and σ_i^{tilt} shown in Fig. 3, colored curves (i.e. 10 individual simulation trials) show alternating patterns, getting steeper as mCpG content increases. More specifically, with higher mCpG content, σ_i^{shift} and σ_i^{tilt} at base-steps G→C became smaller, while that at C→G was almost unchanged. The overall σ_i^{tilt} for 10 trials clearly showed alternating patterns similar to individual trials as mCpG content increased, while σ_i^{shift} did not show such clear patterns. Taking into account the results of partially methylated cases (e.g. SEQ 0.50), methylated DNA made the tilt-mode stiffer at base-steps G→C, and restricted their deformation in determined ranges (Fig. 3).

Furthermore, mCpG suppressed tilt and shift dynamics only at the base-steps adjacent to the mCpG. To clearly show the positional effect, the profiles of $\sigma_i^{\bullet}$ in terms of the repeat unit (i.e. relative position with respect to the mCpG) have been aligned and averaged (see definition of $\overline{\sigma_l^{\bullet}}$ in Sec. II D), as shown in Figs. 4 and S3 [30]. In the profile of SEQ 0.33 in Fig. 4 (a), $\overline{\sigma_l^{tilt}}$ decreases at both neighbors of mCpG (Tab. I), while almost unchanged at G→C between non-methylated CpGs (indicated by $\downarrow^{\times}$ in Fig. 4).

In contrast, for σ_i^{twist}, alternating patterns became less significant as mCpG content increased (Fig. 3). In SEQ 0.00, σ_i^{twist} at base-steps G→C was larger than that at C→G, both in individual trials and the overall profile. In SEQ 1.00, the difference between G→C and C→G was smaller, and almost disappeared in the overall σ_i^{twist} profile. This result suggested that DNA methylation irregularly prevented twist dynamics. In general, DNA methylation made twisting at base-steps G→C stiffer; however, another considerable trend existed. In SEQ 0.50, $\overline{\sigma_l^{twist}}$ was actually higher at mCpG sites than at non-methylated CpG sites (Fig. 4 (c); indicated by red arrow); this trend also appeared in SEQ 0.33, although it was slight (Fig. 4 (b)). This result suggested that, in terms of twisting, methylation generally stiffened DNA; however, C→G in the mCpG itself showed relatively large fluctuations.

Besides the positional effects on each BSP, correlations between BSPs were surveyed (see Sec. II D), to detect long-range interactions or characteristic motion. However, notable results were not found (Figs. S4 and S5 [30]). Correlation of the same BSP between two base-steps (Fig. S4) was high only for the (near-)diagonal components (i.e. nearby base-steps). Some correlation between BSPs within the same base-step (Fig. S5) was observed, reflecting the structure of DNA (e.g. between roll and twist), which was however independent of methylation.

C. Methyl Groups Distribution

The orientation of methyl group θ_n (see Sec. II E) was evaluated (Figs. 5 (a) and S6 [30]). In every case, the θ_n distribution was both positive and negative, and the aver-

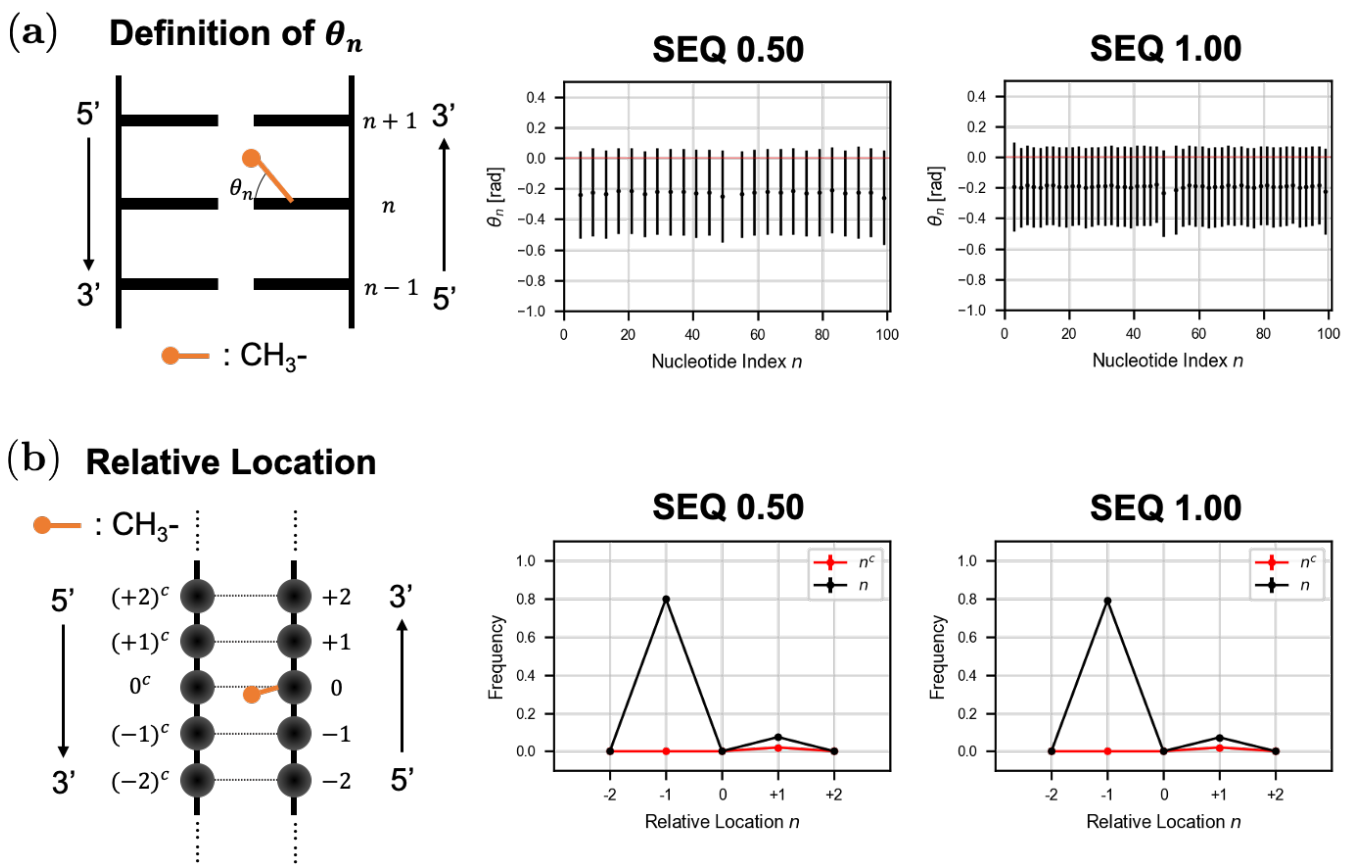

FIG. 5. Orientation of Methyl Groups and Their Interactions with DNA. (a) The definition and observed distribution of θ_n. The average ± the standard deviation (S.D.) of θ_n for each methyl group is shown. (b) The definition of nucleotide indices relative to the methyl group, and estimated contact frequencies. The average ± S.D. of contact frequency for all methyl groups is shown.

age was approximately -0.2 [radian]. This suggested that methyl groups slightly tilt toward the 5'-end, but could fluctuate. No specific interactions (e.g. electrostatic interactions) were identified between the methyl group and other parts of the DNA molecule; instead, physical contacts between methyl groups and their 5'-side neighbor nucleotides were frequently observed (Figs. 5 (b) and S7 [30]). Hence, it was suggested that weak van der Waals interactions and simple excluded volume effects played a role in changing local base-step flexibility.

IV. DISCUSSION AND CONCLUSIONS

In this study, how different patterns of methylation affected DNA dynamics, both overall geometry (Fig. 2) and local flexibility at each base-step in terms of shift, tilt, or twist (Fig. 3) were investigated. Methylation generally stiffens DNA, which is consistent with previous studies. Specifically, the shifting and tilting dynamics of base-steps adjacent to the mCpG was clearly restricted. Nevertheless, in terms of twisting, C→G base-steps in mCpG were certainly more flexible than other sites. This result might originate from the orientation of the methyl groups and their interactions with neighbors (Fig. 6). The twisting flexibility of C→G, maintained at the mCpG itself, may enhance accessibility of mCpG binding proteins [37, 38], although further investigation is required.

Recently, mechanical response of hypermethylated CGI DNA to stretching was assessed by nanometry using optical tweezers [39]; stiffening of DNA was observed, which however disappeared when DNA was overstretched. In our analysis, it was suggested that the base-step flexibility (Fig. 3) was suppressed by dynamic interactions between DNA and methyl groups (Fig. 6). These interactions are expected only in B-DNA conformation and likely to disappear in the overstretched DNA; hence the suggested mechanism (Fig. 6) is consistent with the experimental result [39]. Although our study suggested the importance of methyl group dynamics (Fig. 5), interactions between DNA and water molecules have not been discussed. Another preceding research suggested that hydration around methylation sites affects local base-step

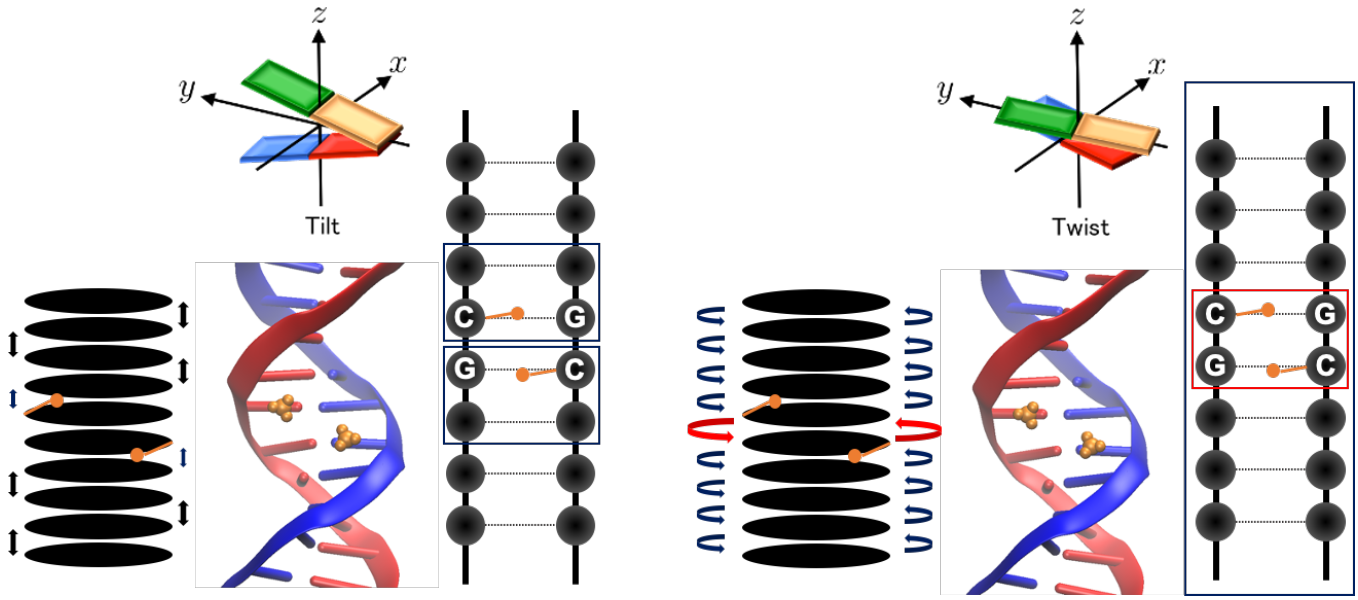

FIG. 6. Schematics of Directed Effects Induced by DNA Methylation. Suggested effects on tilt and twist dynamics. Navy (or red) squares represent ranges where site-specific DNA methylation (i.e. existence of mCpG) makes the tilt or twist mode stiffer (or relatively flexible).

flexibility [22]. Both physical and chemical effects of methylation might be involved in the mechanism.

Previous findings suggest that DNA methylation stabilizes nucleosome positioning [40]. X-ray crystal structures of the nucleosome have been obtained using special sequences of DNA showing high affinity to core histones; in these structures, DNA is not uniformly curved around histones, but bent at several sites [41]. These results also support the hypothesis that site-specific DNA methylation may change local base-step flexibility, which may induce a similar situation and enhance nucleosome positioning.

In this analysis, DNA methylation globally stiffens DNA (Fig. 2), which is consistent with previous experiments [11]. These findings suggest that effects of site-specific methylation on DNA dynamics should be evaluated both at the methylated site and also as global correlated dynamics (e.g. normal modes), as we previously proposed [42, 43]. In recent years, coarse-grained (reduced degree of freedom) physical models of DNA have been proposed, which also include sequence-dependent mechanical properties (e.g. [42–45]). To reproduce the dynamics of longer DNA in physiological conditions, methylation-dependent properties should be also incorporated into these models; this will be performed in future studies.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Zenodo repository DOI: 10.5281/zenodo.3992685.

AUTHOR CONTRIBUTIONS

TK, MS, AA, and YT conceived and designed the research and wrote the paper. TK, AA, and YT analyzed the data. TK and YT performed the simulations.

ACKNOWLEDGMENTS

The authors are grateful to I. Nikaido, S. Shinkai, and R. Erban for fruitful discussions. This work was supported by RIKEN Junior Research Associate Program, “TSUBAME Encouragement Program for Young/Female Users” of Global Scientific Information and the Computing Center at Tokyo Institute of Technology, JSPS KAKENHI Grant Numbers JP16H01408, JP18H04720, and 18KK0388, and by JSPS and NRF under the Japan-Korea Basic Scientific Cooperation Program. Part of the numerical calculations were carried out on the TSUBAME 3.0 supercomputer at Tokyo Institute of Technology, and the RIKEN supercomputer HOKUSAI.

[1] D. N. Cooper, Hum. Genet. **64**, 315 (1983).
[2] A. Bird, Cell **70**, 5 (1992).
[3] Z. D. Smith and A. Meissner, Nat. Rev. Genet. **14**, 204 (2013).

[4] P. A. Jones, Nat. Rev. Genet. **13**, 484 (2012).
[5] M. M. Suzuki and A. Bird, Nat. Rev. Genet. **9**, 465 (2008).
[6] A. M. Deaton and A. Bird, Genes Dev. **25**, 1010 (2011).
[7] H. K. Long, D. Sims, A. Heger, N. P. Blackledge, C. Kutter, M. L. Wright, F. Grützner, D. T. Odom, R. Patient, C. P. Ponting, *et al.*, eLife **2**, e00348 (2013).
[8] F. Antequera, Cell. Mol. Life. Sci. **60**, 1647 (2003).
[9] R. Fenouil, P. Cauchy, F. Koch, N. Descostes, J. Z. Cabeza, C. Innocenti, P. Ferrier, S. Spicuglia, M. Gut, I. Gut, *et al.*, Genome Res. **22**, 2399 (2012).
[10] R. J. Klose and A. P. Bird, Trends Biochem. Sci. **31**, 89 (2006).
[11] T. T. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, and T. Ha, Nat. Commun. **7**, 10813 (2016).
[12] P. M. Severin, X. Zou, H. E. Gaub, and K. Schulten, Nucleic Acids Res. **39**, 8740 (2011).
[13] J. Yoo, H. Kim, A. Aksimentiev, and T. Ha, Nat. Commun. **7**, 11045 (2016).
[14] I. Jimenez-Useche, J. Ke, Y. Tian, D. Shim, S. C. Howell, X. Qiu, and C. Yuan, Sci. Rep. **3**, 2121 (2013).
[15] W. K. Olson, A. A. Gorin, X.-J. Lu, L. M. Hock, and V. B. Zhurkin, Proc. Natl. Acad. Sci. USA **95**, 11163 (1998).
[16] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, Nat. Struct. Biol. **6**, 346 (1999).
[17] R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig, Nature **461**, 1248 (2009).
[18] A. T. P. Carvalho, L. Gouveia, C. R. Kanna, S. K. T. S. Wärmländer, J. A. Platts, and S. C. L. Kamerlin, Epigenetics **9**, 1604 (2014).
[19] A. Pérez, C. L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, M. L. Ruiz, D. Torrents, R. Eritja, M. Soler-López, *et al.*, Biophys. J. **102**, 2140 (2012).
[20] C. Hognon, V. Besancenot, A. Gruez, S. Grandemange, and A. Monari, J. Phys. Chem. B **123**, 7365 (2019).
[21] J. Hörberg and A. Reymer, Chem. Commun. **54**, 11885 (2018).
[22] X. Teng and W. Hwang, Biophys. J. **114**, 1791 (2018).
[23] X.-J. Lu and W. K. Olson, Nucleic Acids Res. **31**, 5108 (2003).
[24] X.-J. Lu and W. K. Olson, Nat. Protoc. **3**, 1213 (2008).
[25] W. K. Olson, A. V. Colasanti, Y. Li, W. Ge, G. Zheng, and V. B. Zhurkin, in *Computational Studies of RNA and DNA* (Springer, 2006) pp. 235–257.
[26] A. V. Morozov, K. Fortney, D. A. Gaykalova, V. M. Studitsky, J. Widom, and E. D. Siggia, Nucleic Acids Res. **37**, 4707 (2009).
[27] W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 33 (1996).
[28] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).
[29] J. C. Phillips, D. J. Hardy, J. D. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, *et al.*, J. Chem. Phys. **153**, 044130 (2020).
[30] See Supplemental Material for additional data.
[31] P. J. Flory and M. Volkenstein, Biopolymers **8**, 699 (1969).
[32] J. Bednar, P. Furrer, V. Katritch, A. Stasiak, J. Dubochet, and A. Stasiak, J. Mol. Biol. **254**, 579 (1995).
[33] P. Gross, N. Laurens, L. B. Oddershede, U. Bockelmann, E. J. Peterman, and G. J. Wuite, Nat. Phys. **7**, 731 (2011).
[34] G. S. Manning, Biophys. J. **91**, 3607 (2006).
[35] C. Calladine, J. Mol. Biol. **161**, 343 (1982).
[36] C. A. Hunter, J. Mol. Biol. **230**, 1025 (1993).
[37] I. Ohki, N. Shimotake, N. Fujita, J.-G. Jee, T. Ikegami, M. Nakao, and M. Shirakawa, Cell **105**, 487 (2001).
[38] X. Zou, W. Ma, I. A. Solov'Yov, C. Chipot, and K. Schulten, Nucleic Acids Res. **40**, 2747 (2012).
[39] C. I. Pongor, P. Bianco, G. Ferenczy, R. Kellermayer, and M. Kellermayer, Biophys. J. **112**, 512 (2017).
[40] R. K. Chodavarapu, S. Feng, Y. V. Bernatavichute, P.-Y. Chen, H. Stroud, Y. Yu, J. A. Hetzel, F. Kuo, J. Kim, S. J. Cokus, *et al.*, Nature **466**, 388 (2010).
[41] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond, J. Mol. Biol. **319**, 1097 (2002).
[42] S. Isami, N. Sakamoto, H. Nishimori, and A. Awazu, PLoS One **10**, e0143760 (2015).
[43] T. Kameda, S. Isami, Y. Togashi, H. Nishimori, N. Sakamoto, and A. Awazu, Front. Physiol. **8**, 103 (2017).
[44] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. Doye, and A. A. Louis, J. Chem. Phys. **137**, 135101 (2012).
[45] G. S. Freeman, D. M. Hinckley, J. P. Lequieu, J. K. Whitmer, and J. J. de Pablo, J. Chem. Phys. **141**, 165103 (2014).