PLOS ONE

RESEARCH ARTICLE

Simple Elastic Network Models for Exhaustive Analysis of Long Double-Stranded DNA Dynamics with Sequence Geometry Dependence

Shuhei Isami[1], Naoaki Sakamoto[1,2], Hiraku Nishimori[1,2], Akinori Awazu[1,2]*

1 Department of Mathematical and Life Sciences, Hiroshima University, Kagami-yama 1-3-1, Higashi-Hiroshima 739-8526, Japan, **2** Research Center for Mathematics on Chromatin Live Dynamics, Hiroshima University, Kagami-yama 1-3-1, Higashi-Hiroshima 739-8526, Japan

* awa@hiroshima-u.ac.jp

Citation: Isami S, Sakamoto N, Nishimori H, Awazu A (2015) Simple Elastic Network Models for Exhaustive Analysis of Long Double-Stranded DNA Dynamics with Sequence Geometry Dependence. PLoS ONE 10(12): e0143760. doi:10.1371/journal.pone.0143760

Editor: Rafael J. Najmanovich, Universite de Sherbrooke, CANADA

Received: July 23, 2015

Accepted: November 9, 2015

Published: December 1, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Platform Project for Supporting in Drug Discovery and Life Science Research (Platform for Dynamic Approaches to Living System) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and Japan Agency for Medical Research and Development (AMED). Grant-in-Aid for Scientific Research on Innovative Areas "Spying minority in biological phenomena" (No. 3306) (24115515) and "Initiative for High-Dimensional Data-Driven Science

Abstract

Simple elastic network models of DNA were developed to reveal the structure-dynamics relationships for several nucleotide sequences. First, we propose a simple all-atom elastic network model of DNA that can explain the profiles of temperature factors for several crystal structures of DNA. Second, we propose a coarse-grained elastic network model of DNA, where each nucleotide is described only by one node. This model could effectively reproduce the detailed dynamics obtained with the all-atom elastic network model according to the sequence-dependent geometry. Through normal-mode analysis for the coarse-grained elastic network model, we exhaustively analyzed the dynamic features of a large number of long DNA sequences, approximately ∼150 bp in length. These analyses revealed positive correlations between the nucleosome-forming abilities and the inter-strand fluctuation strength of double-stranded DNA for several DNA sequences.

1 Introduction

Elastic network models of proteins, including all-atom models [1, 2] and coarse-grained models [3–9] represent some of the simplest and most powerful types of theoretical models that can accurately reveal structure-dynamics relationships and the mechanisms underlying a protein's functional activities [10–14]. Such models have also been widely employed to accurately reproduce the temperature factors on the crystal structure of a protein via normal-mode analysis. These models can thus demonstrate the large and slow deformations of proteins that are essential for protein functions, but remain difficult to demonstrate via all-atom molecular dynamics simulations.

Along with proteins, DNA is the most important biomolecule for the activities of all living organisms. Recent developments in molecular biology have revealed that DNA contains several functional regions. Therefore DNA is no longer considered to have the sole function in the

through Deepening of Sparse Modeling" (No. 4503) (26120525) of MEXT of Japan (A. A) Grants-in-Aid for Scientific Research (C) (No. 25430169) of MEXT of Japan (N. S).

Competing Interests: The authors have declared that no competing interests exist.

storage of genetic information, but is now known to be actively involved in gene regulation [15–18], insulator activity [19–24], and in the construction of chromosomal architectures [25–28] such as heterochromatins and topologically associated domains through nucleosome formation and protein bindings [29–33]. The functional behavior of each strand of DNA is determined not only by chemical aspects of the nucleotides and base pairs but also by its physical characteristics such as the structure and dynamics of the nucleotide sequences in each functional region. However, comprehensive understanding of the physical properties of nucleotide sequences lags far behind the knowledge accumulated of their biochemical aspects [15, 16, 26].

Since the last century, much progress has been made in revealing the physical aspects of DNA using all-atom normal-mode analysis [34, 35] and molecular dynamics simulations [36–42]. Several coarse-grained models of DNA (and RNA) have also been proposed. Some of these describe the detailed shape of each nucleotide using three or more particles [43–49], whereas others describe each base pair by simply one or two particles [50–54]. Molecular dynamics simulations and normal-mode analysis of these models have identified the flexibilities, nucleosome-forming abilities, and zip-unzip transitions of the double helices of some specific DNA sequences from tens to a few hundred base pairs in length. Although these methods have proven to be very powerful for the analysis of the physical aspects of DNA, molecular dynamics simulations are not suitable for conducting exhaustive analyses of several sequences simultaneously owing to the high computational costs of such extensive simulations; for example, analyses of whole genomic and whole possible sequences. Moreover, the normal-mode analysis of all-atomic models of long DNA sequences is also computationally heavy.

Alternatively, data-driven methods have been proposed for determining the mechanical properties of DNA with respect to the helical parameters and local flexibilities of base pairs from X-ray crystal structure analysis and all-atom molecular dynamics [55–60]. These methods also seem to be powerful and can be applied to the analysis of the mechanical properties of several DNA sequences simultaneously. However, the methods thus far proposed have focused on static and local mechanical properties. Therefore, they are not particularly useful for the study of the functional contribution of the dynamic and correlated motions of DNA.

The objective of this study was to construct a simple coarse-grained elastic network model of DNA to allow for exhaustive analysis of the dynamic correlated motions of several long DNA sequences. For this purpose, we first constructed a simple all-atom elastic network model of short DNA sequences based on the method introduced by Tirion [1] for the modeling of protein dynamics. We confirmed that this model could accurately reproduce the temperature factors of some DNA fragments from data obtained with X-ray crystal structure analysis.

Second, we developed a simple coarse-grained elastic network model of long DNA sequences, where each nucleotide is described by only one node. We confirmed that this simplified model has lower computational costs but can nonetheless reproduce the nucleotide sequence-dependent dynamics revealed by the all-atom elastic network model.

Finally, through the normal-mode analysis of this coarse-grained model, we conducted an exhaustive analysis of the general features of the sequence-dependent structure-dynamics relationships among several DNA sequences. We specifically focused on the dynamic properties of a large number of long DNA sequences, approximately $\sim$150 bp in length (with respect to the length of the nucleosome-forming regions), for the genomes of some model organisms as well as for random sequences with several A, T, C, G ratios. Through these analyses, we found that the dynamic aspects of DNA are highly influenced by their sequences, and found positive correlations between the nucleosome-forming abilities and inter-strand fluctuations of double-stranded DNA.

In particular, we focus on the geometry dependencies of the dynamics of several DNA sequences, since a recent study demonstrated that the geometry of DNA sequences has a

dominant contribution to their mechanical features [47]. In the following arguments, for simplicity, we use "sequence-dependent" to mean "dynamics that depend on sequence geometry".

2 Models and Methods

2.1 Basic structures of Elastic Network Models of DNA

In order to construct elastic network models of DNA, the basic DNA structure first needs to be determined. In the following arguments, we construct two types of models: i) a simple all-atom elastic network model (AAENM) to reproduce the characteristics of the crystal structures of DNA, and ii) a simple coarse-grain elastic network model (CGENM) to reproduce the characteristics of the AAENM. We obtained the basic DNA structures in the following two ways for the respective purposes of constructing each model.

For construction of the AAENM, we employed the atom coordinate sets of several naked DNA crystal structures included in the Protein Data Bank (PDB) as the given basic structures of the model. We used 9 DNA crystal structures containing only DNA and water molecules under different conditions of crystallization, where all temperature factors are given as positive values (Table 1). The suitability of the model was evaluated by comparisons of the temperature factors obtained between the model and those obtained from X-ray crystal structure analysis.

The objective of our constructed arguments was to unveil the sequence-dependent dynamic features of several long DNA sequences simultaneously. In recent crystal structure analysis, only shorter DNA sequences (i.e., less than 12 bp in length) were studied. On the other hand, several types of helical parameter sets, base pair parameters, and base step parameter sets have been proposed through experiments or molecular dynamics simulations [36, 61–66](Table 2, S1 Table). Here, we used the application X3DNA [67] to obtain the coordinates of each atom for any sequence from these helical parameters, and used these parameters to construct the AAENMs and CGENMs of longer DNA sequences (for detailed methods of the generation of the atom coordinates, see Lu and Olson [67]). In the following arguments, we compare the characteristics of the AAENM and CGENM constructed by X3DNA from three different helical parameter sets: i) a parameter set obtained by an *in vitro* experiment and X-ray crystal structure analysis [48, 61–63] (Table 2), ii) a parameter set obtained only by the X-ray crystal structure analysis, and iii) a parameter set obtained by all-atom molecular dynamics simulations [58, 65, 66] (S1 Table).

Table 1. Information of the analyzed DNA segments. PDB ID, sequences of X-ray crystal structure analysis of DNA fragments, parameter sets of the AAENM for each DNA structure, and correlation coefficients of TF_i and $MATF_i$ between the AAENM and X-ray crystal structure.

PDB ID	Sequence	T [K]	C_a [kJ/(Å²mol)]	B_B	B_w	ρ (all atom)	ρ (MATF)
1BNA	5'-CGCGAATTCGCG-3'	290	2.91	0.018	0.002	0.6775	0.8463
7BNA	5'-CGCGAATTCGCG-3'	290	13.30	0.010	0.002	0.6349	0.7038
9BNA	5'-CGCGAATTCGCG-3'	290	5.90	0.007	0.002	0.6764	0.7475
1D91	5'-GGGGTCCC-3'	277	0.89	0.632	0.252	0.5750	0.6925
1DC0	5'-CATGGGCCCATG-3'	277	1.18	0.380	1.192	0.5910	0.7038
122D	5'-CCAGGCCTGG-3'	277	2.85	0.002	0.050	0.8219	0.8744
123D	5'-CCAGGCCTGG-3'	277	6.87	0.003	0.007	0.7804	0.8230
181D	5'-CACGCG-3'	296	0.94	1.522	0.332	0.4217	0.5788
330D	5'-ACCGCCGGCGCC-3'	277	4.53	0.013	0.002	0.5418	0.5568

doi:10.1371/journal.pone.0143760.t001

Table 2. Helical parameter sets (i). Helical parameter sets (i) obtained by *in vitro* experiments and X-ray crystal structure analysis [48, 61–63].

Base-step parameters						
	Shift [Å]	Slide [Å]	Rise [Å]	Tilt [°]	Roll [°]	Twist [°]
AA	–0.05	–0.21	3.27	–1.84	0.76	35.31
AT	0.00	–0.56	3.39	0.00	–1.39	31.21
AC	0.21	–0.54	3.39	–0.64	–1.39	31.52
AG	0.12	–0.27	3.38	–1.48	3.15	33.05
TA	0.00	0.03	3.34	0.00	5.25	36.20
TT	0.05	–0.21	3.27	1.84	0.76	35.31
TC	0.27	–0.03	3.35	1.52	3.87	34.80
TG	0.16	0.18	3.38	0.05	5.95	35.02
CA	–0.27	–0.03	3.35	–1.52	3.87	34.80
CT	–0.12	–0.27	3.38	1.48	3.15	33.05
CC	0.02	–0.47	3.28	0.40	3.86	33.17
CG	0.00	0.57	3.49	0.00	4.29	35.30
GA	–0.27	–0.03	3.35	–1.52	3.87	34.80
GT	–0.21	–0.54	3.39	0.64	0.91	31.52
GC	0.00	–0.07	3.38	0.00	0.67	34.38
GG	–0.02	–0.47	3.28	–0.40	3.86	33.17
Base pair parameters						
	Shear [Å]	Stretch [Å]	Stagger [Å]	Buckle [°]	Propeller [°]	Opening [°]
A-T	0.07	–0.19	0.07	1.80	–15.00	1.50
T-A	–0.07	–0.19	0.07	–1.80	–15.00	1.50
C-G	0.16	–0.17	0.15	–4.90	–8.70	–0.60
G-C	–0.16	–0.17	0.15	4.90	–8.70	–0.60

doi:10.1371/journal.pone.0143760.t002

2.2 All-atom Elastic Network Model of DNA

A simple all-atom elastic network model of double-stranded DNA was constructed based on the model proposed by Tirion [1]. In this model, we regard all of the atoms of given DNA sequences as the nodes comprising the elastic network. For the analysis of the crystal structures of DNA involving water molecules, we also regard the oxygen atoms of water around DNA as the nodes of the elastic network. We define the mass and position of atom i as m_i and $\mathbf{r}_i$ ($\mathbf{r}_i = (x_i, y_i, z_i)$), respectively. The potential V of all atoms is given as

$$V = \sum_{i,j} \frac{C_a}{2} (|\mathbf{r}_i - \mathbf{r}_j| - |\mathbf{r}_i^0 - \mathbf{r}_j^0|)^2 \theta(R_i + R_j + R_c - |\mathbf{r}_i^0 - \mathbf{r}_j^0|) + \sum_{boundary} \frac{B_i C_a}{2} (\mathbf{r}_i - \mathbf{r}_i^0)^2. \quad (1)$$

Here, $\mathbf{r}_i^0$ is the position of atom i of the basic DNA structure, as described above.

The first term indicates the interaction potential among atoms that are spatially closed in the basic DNA structure (Fig 1(a)). Here, R_i refers to the Van der Waals radius of atom i, R_c is an arbitrary cut-off parameter that models the decay of interactions with distance, and θ indicates the Heaviside function, where $\theta(z) = 1$ ($\theta(z) = 0$) for $z \geq 0$ ($z < 0$). We assume $R_c = 2.0\text{Å}$, which is empirically considered to be an appropriate range for biomolecules, at least for proteins [1]. The results of the arguments presented in this paper were qualitatively unchanged for the appropriate range of R_c. In the crystal structures of DNA, the motions of water molecules are also restricted by the crystal packing. Thus, for all atoms, we assume that spatially closed

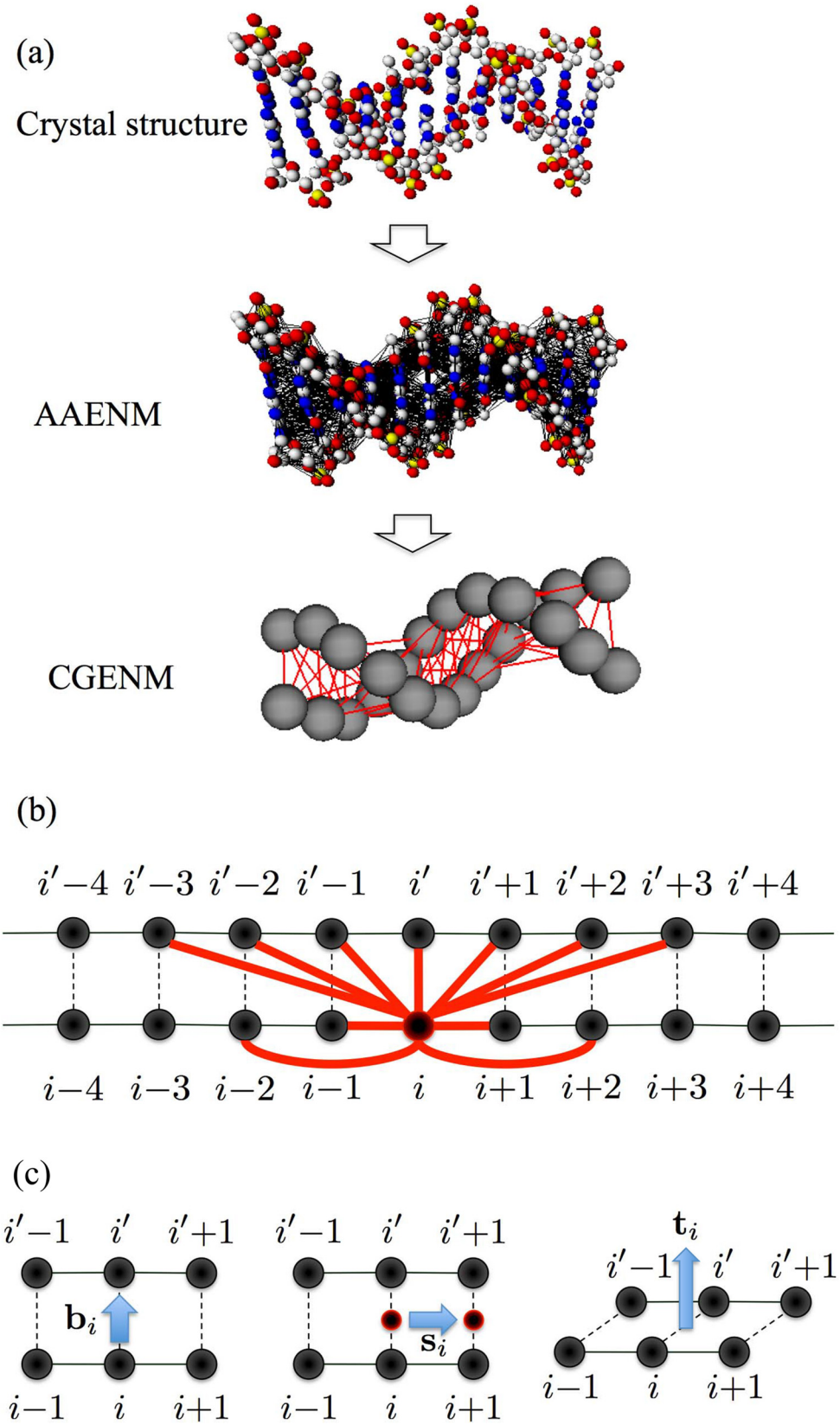

Fig 1. Illustrations of (a) all-atom elastic network models (AAENMs) from the crystal structures, and the coarse-grained elastic network models (CGENMs) from the AAENMs; (b) detailed interactions among the nucleotides (nodes) of the CGENM where nucleotide *i* interacts with all nucleotides connected by the 11 bold curves; and (c) $\mathbf{b}_i$, $\mathbf{s}_i$, and $\mathbf{t}_i$.

doi:10.1371/journal.pone.0143760.g001

pairs of atoms are connected by linear springs with their respective natural lengths. The elastic coefficient C_a ($[kJ/(\mathring{A}^2 mol)]$) is a phenomenological constant, which is assumed to be the same for all interacting pairs.

The second term indicates the boundary effects of each DNA and water molecule in each DNA crystal structure. This term plays a crucial role for the analysis of the fluctuations of the crystal structure of DNA, such as temperature factors, since the fluctuations of the nucleotides at the edges of DNA and water molecules are restricted due to the following facts.

In the crystal of DNA, the motion of edges at upper and lower streams of each DNA segment (left and right edges in Fig 1(a)) is influenced by atoms of other adjacent segments in the long-axis direction of DNA segments. Moreover, the motions of water molecules around each DNA segment are influenced by atoms of other adjacent water molecules or DNA segments in the direction perpendicular to the longitudinal axis of the DNA segment. Thus, we need to consider the second term of Eq (1) to implement such effects, where B_j indicates the strength of such effects for atoms in the edge nucleotides of each DNA and water molecule, respectively.

Remarkably, as shown in the Results section, the distributions of the temperature factor of atoms exhibited various patterns from the same sequence of crystallized DNA (S1 and S2 Figs) owing to the dependency on the conditions of crystallization. Therefore, in order to compare the characteristics of the present DNA model to those of the crystal structure of DNA, appropriate values of B_j need to be assigned to the atom j that belongs to the edge base pairs and water molecules. For simplicity, we assign $B_j = B_B$ and B_w to the atoms j at the edge base pairs and the water molecules, respectively, and $B_j = 0$ otherwise.

It is noted that the internal structure of the crystal of DNA is spatially anisotropic. Thus, it is reasonable to assume that the interactions among local parts of the crystal of DNA exhibit different strengths in different directions. Accordingly, in general, the strengths of the restrictions of atoms belonging to the edge of DNA differ from those of water molecules. Thus, we assume that B_B and B_w are different values (Table 1).

For simplicity, the mass of each atom is assumed as a constant value ($m_i = m = 10^{-3}/N_A[kg]$, $N_A = 6.02214129 \times 10^{23}[/mol]$ is Avogadro's number). However, we confirmed that the results were almost identical when using the precise masses of the atoms.

2.3 Coarse-grained Elastic Network Model of DNA

A simple coarse-grained elastic network model of double-stranded DNA, where each nucleotide is described as one node, was constructed as follows (Fig 1(a)). We define the coordinate of the C1' carbon of nucleotide i, $\mathbf{x}_i$ ($\mathbf{x}_i = (x_i, y_i, z_i)$), as the position of nucleotide i, and regard the motion of the C1' carbon as that of the nucleotide. Here, we assume that the mass of the C1' carbon obeys $m_i = 10^{-3}/N_A[kg]$. The potential V of all nucleotides is given as

$$V = \sum_{i,j} \frac{C_g}{2} (|\mathbf{x}_i - \mathbf{x}_j| - |\mathbf{x}_i^0 - \mathbf{x}_j^0|)^2. \quad (2)$$

Here, $\mathbf{x}_i^0$ is the position of nucleotide i of the basic structure of DNA, as defined above. We assume that nucleotides i and i' belong to the same base pair. For nucleotide i, the sum is restricted to the pair of nucleotides in the same base pair ($j = i'$), in the neighboring base pairs ($j = i + 1, i' + 1, i - 1, i' - 1$), in the next neighboring bases pairs ($j = i + 2, i' + 2, i - 2, i' - 2$), and in the next-next neighboring bases in another strand ($j = i' + 3, i' - 3$) (Fig 1(b)). The elastic coefficient C_g ($[kJ/(\mathring{A}^2 mol)]$) is a phenomenological constant that is assumed to be the same for all interacting pairs. This model is considered as a simplified version of previously proposed one-site-per-nucleotide models [50, 53].

2.4 Normal-mode Analysis

An overview of the theory of normal-mode analysis is provided in several recent studies [1–7, 10–13]. Thus, we here briefly show the results of this analysis. For this analysis, we define $\mathbf{q}(t)$ ($\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, ...\mathbf{q}_N)$, $\mathbf{q}_i = (x_i, y_i, z_i)$) as a $3N$-dimensional position vector, and $\mathbf{q}^0$ as the position vector of the basic structure. Here, $\mathbf{q} = (\mathbf{r}_1, \mathbf{r}_2, ...\mathbf{r}_N)$ for the AAENM and $\mathbf{q} = (\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N)$ for the CGENM. The motions of small deviations of $\mathbf{q}(t)$ from $\mathbf{q}^0$, $\delta\mathbf{q}(t) = \mathbf{q} - \mathbf{q}^0$ obey

$$\delta\mathbf{q}(t) = \sum_{\omega_k \neq 0} A_k \mathbf{v}_k e^{i\omega_k t} \tag{3}$$

where $-(\omega_k)^2$ and $\mathbf{v}^k = (\mathbf{v}_1^k, \mathbf{v}_2^k, ..., \mathbf{v}_N^k)$ ($\mathbf{v}_i^k = (v_{x_i}^k, v_{y_i}^k, v_{z_i}^k)$) are the k-th largest eigenvalue and its eigenvector of the $3N \times 3N$ Hessian matrix $\mathbf{H}$ as

$$H_{ij} = -\left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_{\mathbf{q}=\mathbf{q}^0}. \tag{4}$$

We assume that the system is in thermodynamic equilibrium with temperature T. Thus, the amplitude A_k is given as

$$(A_k)^2 = \frac{k_B T}{(\omega_k)^2 m} \tag{5}$$

with Boltzmann constant $k_B = 1.3806488 \times 10^{-23} [m^2\ kg/s^2\ K]$. Using this solution, the mean square fluctuation of the i-th atom in the AAENM ($\delta\mathbf{q}_i = \delta\mathbf{r}_i$) is obtained as

$$AF_i = <|\delta\mathbf{r}_i|^2> = \sum_{\omega_k \neq 0} \frac{k_B T}{(\omega_k)^2 m} |\mathbf{v}_i^k|^2 \tag{6}$$

with Boltzmann constant, and the temperature factor is displayed as $TF_i = \frac{8}{3}\pi^2 AF_i$. Here, $<\ldots>$ represents the temporal average.

For the CGENM, we define the mean square fluctuation of the n-th nucleotide ($\delta\mathbf{q}_n = \delta\mathbf{x}_n$) as $CF_n = <|\delta\mathbf{x}_n|^2>$. To consider the motion of the n-th nucleotide in the AAENM, we define the average nucleotide motions as $\delta\mathbf{R}_n = <\delta\mathbf{r}_j>_{n-th\ nucleotide}$. Using this vector, we define the mean square fluctuation of the n-th nucleotide as $NF_i = <|\delta\mathbf{R}_n|^2>$. For the motif *mo* of the n-th nucleotide (*mo* = sugar, base, and phosphoric acid), the average temperature factor of the motif ($MATF_{(mo\ in\ n-th\ nucleotide)}$) is defined as the average of TF_j belonging to each motif of each nucleotide, $< TF_j >_{(mo\ in\ n-th\ nucleotide)} = \frac{8}{3}\pi^2 < AF_j >_{(mo\ in\ n-th\ nucleotide)}$.

2.5 Treatment of the Temperature Factor in X-ray crystal Structure Analysis

In order to evaluate the validity of the AAENM, we measured the correlation coefficient between the profile of the temperature factor obtained from PDB data (via X-ray crystal structure analysis) and that obtained from the AAENM based on this crystal structure. It is noted that the temperature factor profiles for some of the PDB data often include unnaturally large or small values. Thus, the correlation coefficients were estimated using data excluding such outliers. In the present evaluations, the value g_i was considered as an outlier if $|g_i - \mu| > s\sigma$, where μ

and σ are the average and standard deviation of $\{g_i\}$, respectively, and $s = 2.5$ is used based on the standard arguments of statistics.

2.6 Evaluations of Anisotropic Fluctuations of DNA

We also focused on the relationships between the fluctuations of each nucleotide in the AAENMs and CGENMs in the directions parallel to the base pair axis, parallel to the helix axis, and vertical to both the base pair and helix axes.

Here, we name the nucleotides in one and the other strand constructing the i-th base pair as the i-th and i'-th nucleotide. We define the position vectors of the C1' atoms belonging to the i-th and i'-th nucleotides as $\mathbf{c}_i$ and $\mathbf{c}_{i'}$, and consider

$$\mathbf{b}_i = \frac{\mathbf{c}_i^0 - \mathbf{c}_{i'}^0}{|\mathbf{c}_i^0 - \mathbf{c}_{i'}^0|}, \tag{7}$$

$$\mathbf{s}_i = \frac{(\mathbf{c}_{i+1}^0 + \mathbf{c}_{i'+1}^0) - (\mathbf{c}_i^0 + \mathbf{c}_{i'}^0)}{|(\mathbf{c}_{i+1}^0 + \mathbf{c}_{i'+1}^0) - (\mathbf{c}_i^0 + \mathbf{c}_{i'}^0)|}, \tag{8}$$

and

$$\mathbf{t}_i = \frac{\mathbf{b}_i \times \mathbf{s}_i}{|\mathbf{b}_i \times \mathbf{s}_i|} \tag{9}$$

(Fig 1(c)). It is noted that $\mathbf{b}_i$ and $\mathbf{s}_i$ are not orthogonal in general; however, we confirmed that the angles of these vectors were always sufficiently close to $\pi/2$ rad for each i.

Using these vectors, the mean square fluctuations of the i-th and i'-th nucleotides of the AAENM in the directions parallel to the base pair axis are defined as $NF_i^b = <|\delta\mathbf{R}_i\mathbf{b}_i|^2>$ and $NF_{i'}^b = <|\delta\mathbf{R}_{i'}\mathbf{b}_i|^2>$, those parallel to the helix axis are defined as $NF_i^s = <|\delta\mathbf{R}_i\mathbf{s}_i|^2>$ and $NF_{i'}^s = <|\delta\mathbf{R}_{i'}\mathbf{s}_i|^2>$, and those in the torsional direction are defined as $NF_i^t = <|\delta\mathbf{R}_i\mathbf{t}_i|^2>$ and $NF_{i'}^t = <|\delta\mathbf{R}_{i'}\mathbf{t}_i|^2>$, respectively. The mean square fluctuations of the i-th and i'-th nucleotides of the CGENM in these same directions are obtained by $CF_i^b = <|\delta\mathbf{x}_i\mathbf{b}_i|^2>$ and $CF_{i'}^b = <|\delta\mathbf{x}_{i'}\mathbf{b}_i|^2>$, $CF_i^s = <|\delta\mathbf{x}_i\mathbf{s}_i|^2>$ and $CF_{i'}^s = <|\delta\mathbf{x}_{i'}\mathbf{s}_i|^2>$, and $CF_i^t = <|\delta\mathbf{x}_i\mathbf{t}_i|^2>$ and $CF_{i'}^t = <|\delta\mathbf{x}_{i'}\mathbf{t}_i|^2>$, respectively. Moreover, we consider the mean square fluctuation of the relative base position of each base pair of the CGENM in the three directions listed above given by $DF_i^b = <|(\delta\mathbf{x}_i - \delta\mathbf{x}_{i'})\mathbf{b}_i|^2>$, $DF_i^s = <|(\delta\mathbf{x}_i - \delta\mathbf{x}_{i'})\mathbf{s}_i|^2>$, and $DF_i^t = <|(\delta\mathbf{x}_i - \delta\mathbf{x}_{i'})\mathbf{t}_i|^2>$, respectively.

2.7 Evaluations of the Overall Geometry of DNA

The overall geometry of each modeled DNA molecule is characterized by the ratios among the square root of the three principal components of the populations of atoms, $\sqrt{\lambda_1}$, $\sqrt{\lambda_2}$, and $\sqrt{\lambda_3}$ ($\lambda_1 > \lambda_2 > \lambda_3 > 0$). Here, λ_1, λ_2, and λ_3 are obtained as eigenvalues of the covariant matrix

$$I = \begin{pmatrix} <(\Delta x_i)^2>_i & <\Delta x_i \Delta y_i>_i & <\Delta x_i \Delta z_i>_i \\ <\Delta y_i \Delta x_i>_i & <(\Delta y_i)^2>_i & <\Delta y_i \Delta z_i>_i \\ <\Delta z_i \Delta x_i>_i & <\Delta z_i \Delta y_i>_i & <(\Delta z_i)^2>_i \end{pmatrix}, \tag{10}$$

where $(\Delta x_i, \Delta y_i, \Delta z_i) = (x_i - x_{CM}, y_i - y_{CM}, z_i - z_{CM})$, (x_i, y_i, z_i) is the position of i-th atom, $(x_{CM},$

y_{CM}, z_{CM}) is the position of the center of mass of a given DNA molecule, and $<\ldots>_i$ indicates the average for all *i*s. We evaluate the overall geometry of a given DNA molecule using the linearity $\sigma_1 = \sqrt{\lambda_1}/\sqrt{\lambda_2}$ and the line symmetry with respect to the λ_1-axis $\sigma_2 = \sqrt{\lambda_3}/\sqrt{\lambda_2}$. Here, σ_1 is large when the DNA is straightened, and σ_2 is large (small) when the DNA forms a three (two)-dimensional curve with wide (flat) envelope.

2.8 Evaluations of Correlations Among the Results of AAENM, CGENM, and Experiments

We employ Pearson's correlation coefficients, ρ, to evaluate the correlations among the profiles of temperature factors and several anisotropic fluctuations of atoms obtained by AAENM, CGENM, and experiments.

3 Results and Discussion

3.1 Comparisons of Fluctuations between the AAENM and the Crystal Structure of DNA

The fluctuations of atoms of the AAENMs of several short DNA sequences were measured with normal-mode analysis. Here, the basic structures of the DNAs are given by the crystal structures of the naked DNAs with the following PDB IDs: 1BNA, 7BNA, 9BNA, 1D91, 1DC0, 122D, 123D, 181D, and 330D (Table 1 [68–75]). To confirm the validities of the AAENMs, the correlations between the distribution profiles of the temperature factor of atoms (TF_i) and the average temperature factor of the motifs ($MATF_i$) in the crystal structures and those in the corresponding AAENMs were measured. In the following arguments, we employ the optimal values of C_a, B_B, and B_w for each crystal structure (Table 1), which were manually found to maximize ρ of TF_i between the results of the crystal structure analysis and those of the AAENM.

By choosing the appropriate conditions for the atoms of the two edge base pairs and water molecules (B_B and B_W) for each PDB ID, TF_i of each AAENM exhibited a similar profile to that of the crystal structure with a significant correlation coefficient ρ (Fig 2(a), Table 1, and S1 Fig). Therefore, the AAENM seemed to reproduce the overall structure of the temperature factor profile of each crystal structure well, although the detailed profiles among atoms showed some deviations.

Furthermore, we focused on the average temperature factor for the motifs, $MATF_i$, as recently discussed [35]. We found that the $MATF_i$ obtained with the AAENMs could accurately reproduce those of the corresponding crystal structures with high correlation coefficients ρ, where $\rho > \sim 0.7$ was obtained for most cases (Fig 2(b), Table 1, and S2 Fig). These results demonstrate that the AAENM is a suitable model for describing the sequence-dependent fluctuations of the nucleotide motifs of several double-stranded DNA sequences, despite the simplicity of model construction and its easy implementation. Moreover, these results show that the sequence-dependent forms of DNA have a dominant contribution to their overall flexibilities and fluctuations.

Nevertheless, the present AAENM could not accurately reproduce $MATF_i$ well for some crystal structures. These deviations are considered to have arisen from the following primary assumption: we only considered the effects of the restrictions by the packing of DNAs in crystal form for two edge base pairs and water, whereas such effects have an influence on all atoms. Therefore, obtaining and incorporating more detailed knowledge of the restrictions of bulk sequences caused by crystal packing should help to achieve a more accurate reproduction of the molecular fluctuations for all cases.

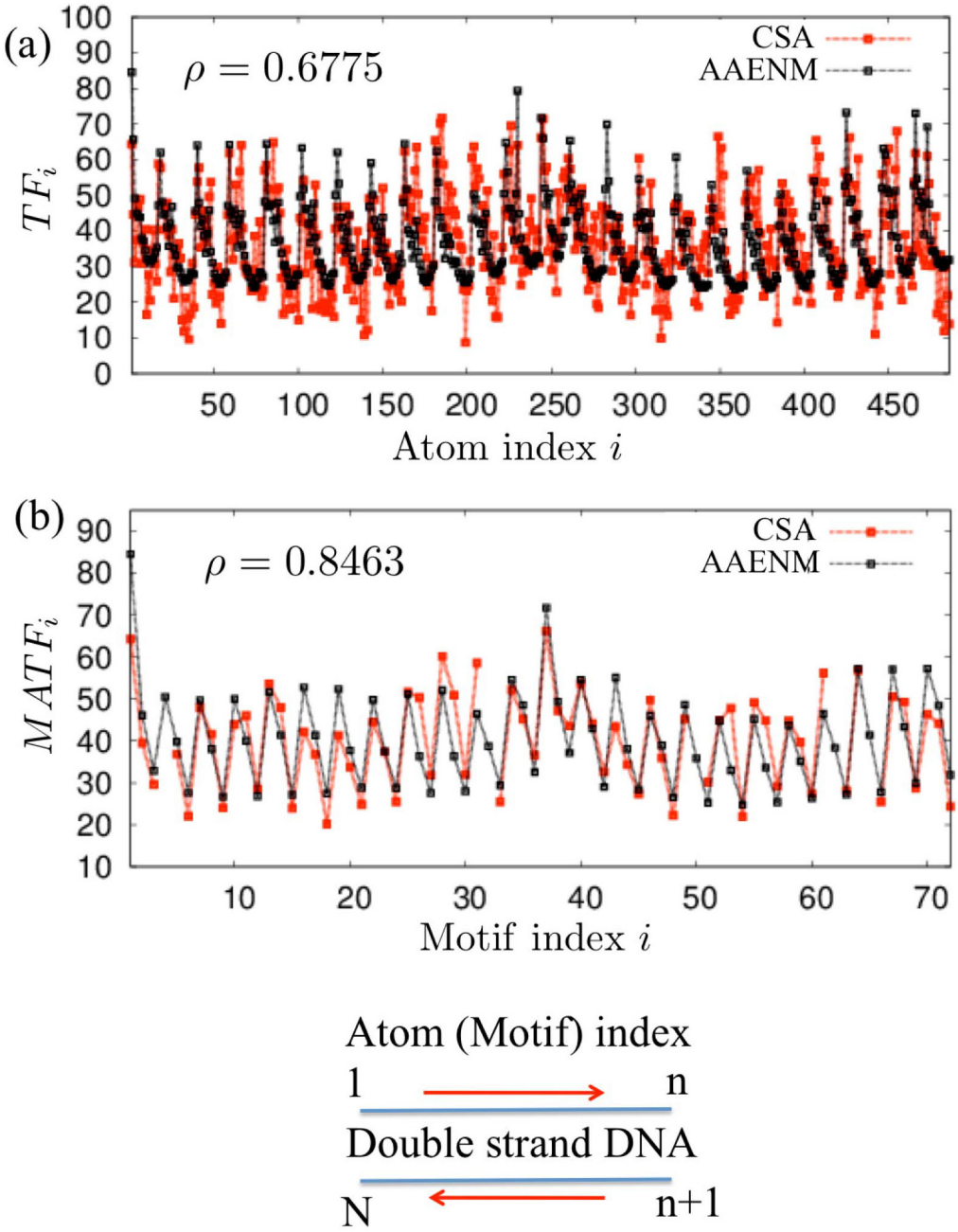

Fig 2. Temperature factors obtained with the AAENMs and crystal structure analysis. (a) Temperature factor of each atom (TF_i) and (b) average temperature factor of the motifs ($MATF_i$) obtained by the AAENMs (black curve) and crystal structure analysis (CSA, gray (red) curve) of typical double-stranded DNA (PDB ID: 1BNA). Here, $C_a = 2.91[kJ/(\AA^3 mol)]$, $B_B = 0.018$, and $B_W = 0.02$. Atom and motif indices in (a) and (b) are given in the same order as shown for (c). ρ indicates the Pearson correlation coefficient of the profiles between the two curves.

doi:10.1371/journal.pone.0143760.g002

3.2 Comparison between the AAENMs and CGENMs of DNA

The main objective of this study was to unveil the sequence-dependent dynamic correlated motions of several long DNA sequences. We next describe these dynamics of DNA sequences with longer lengths than considered in the previous subsection. For this purpose, we also constructed a coarse-grained model, which is often useful for focusing on the slow and large-scale dynamics of molecules that essentially influence their function. Thus, to propose a coarse-grained model of long double-stranded DNA, we evaluated whether the CGENM proposed provides an appropriate coarse-grained model of the present AAENM.

We performed normal-mode analysis of the AAENM and corresponding CGENM for 500 randomly chosen 50-bp sequences, and compared the mean square fluctuations of the i-th nucleotide (NF_i and CF_i) in the directions parallel to the base pair axis (NF_i^b and CF_i^b), parallel to the helix axis (NF_i^s and CF_i^s), and in the torsional direction (NF_i^t and CF_i^t). Here, $C_a = 1.29kJ/(\text{Å}^2 mol)$ is employed, as in the previous study [1], and $C_g = 7.7kJ/(\text{Å}^2 mol)$ is assumed, which was manually found to provide the best fit of fluctuation profiles between AAENM and CGENM. Here, the overall fluctuation profiles of CGENM are independent of the value of C_g since C_g influences only on the absolute values of fluctuations. Independent of the sequences and employed helical parameters, we found that the fluctuations of each nucleotide in several directions were very similar between the AAENMs and CGENMs when these models are constructed with the same helical parameters, with average correlation coefficients $\rho > 0.98$ (Fig 3, Table 3, S2 Table, S3 and S4 Figs).

It is noted that the present CGENM contains only one node per nucleotide, whereas the AAENM contains $19 \sim 22$ nodes (atoms) per nucleotide. This fact demonstrates that the computational costs of the CGENMs are much lower than those of the AAENMs, although the accuracies of the obtained statistical aspects are almost identical between these two models. Thus, this CGENM could be used for exhaustive analysis and comparisons of the dynamic features of several sequence-dependent DNAs related to protein binding affinities, functions of transcription regulation sequences, and nucleosome positioning [16–26, 46, 55–60]. In the next subsection, we provide an example of such an analysis to determine the relationships between the nucleosome-forming abilities of several double-stranded DNA sequences and their inter-strand dynamic features.

3.3 Exhaustive Analysis of the Sequence-dependent Behaviors of 150-bp DNAs with the CGENM

Nucleosome positioning is important not only for compacting DNA but also for appropriate gene regulation. Several recent studies have been performed for genome-wide nucleosome mapping and the identifications and predictions of nucleosome-forming and -inhibiting sequences for some model organisms [57, 59, 76–80].

As an example of the applications of the CGENM to an exhaustive analysis of the sequence-dependent behavior of DNA, we compared the dynamic features of DNA sequences of $\sim$150 bp that were predicted as nucleosome-forming or nucleosome-inhibiting sequences in the genome of budding yeast *Saccharomyces cerevisiae* (5000 forming sequences and 5000 inhibiting sequences of 150 bp) [57], nematode *Caenorhabditis elegans* (2567 forming sequences and 2608 inhibiting sequences of 147 bp), *Drosophila melanogaster* (2900 forming sequences and 2850 inhibiting sequences of 147 bp), and *Homo sapiens* (2273 forming sequences and 2300 inhibiting sequences of 147 bp) [59]. The histograms of the average relative fluctuations of DNAs for the three directions $< CF_i >_i$, $< CF_i^b >_i$, $< CF_i^s >_i$, $< CF_i^t >_i$, $< DF_i >_i$, $< DF_i^b >_i$, $< DF_i^s >_i$, and $< DF_i^t >_i$ ($<\ldots>_i$ indicates the average for all is.) for the nucleosome-forming

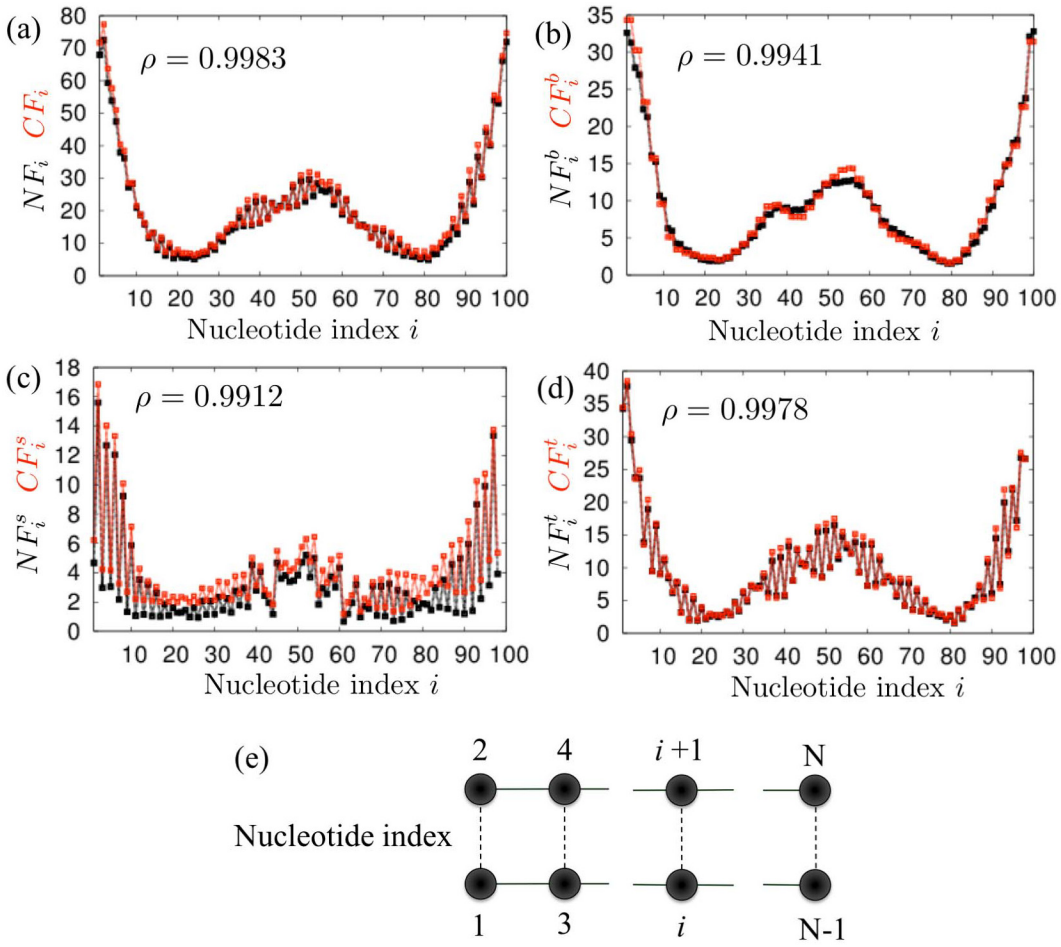

Fig 3. Comparisons between the AAENM and CGENM. Comparisons of the fluctuations between each nucleotide in the AAENM (black curves) and CGENM (gray (Red) curves) for a typical 50-bp random DNA sequence (5'—GAGGCTAAAGTCTATTTAGACCGGAGTTGACGTGGAAGCCCGGCTAGTCT—3'). (a) NF_i and CF_i, (b) NF_i^b and CF_i^b, (c) NF_i^s and CF_i^s, and (d) NF_i^t and CF_i^t. Helical parameter set (i) (Table 2) was used for both models. The nucleotide indices in (a) to (d) are given in the same order shown for (e). ρ indicates the Pearson correlation coefficient of the profiles between the two curves.

doi:10.1371/journal.pone.0143760.g003

Table 3. Comparisons between the AAENM and CGENM. Average and standard deviation of the correlation coefficients of 500 random samples of 50-bp sequences between NF_i and CF_i, NF_i^b and CF_i^b, NF_i^s and CF_i^s, and NF_i^t and CF_i^t. Helical parameter set (i) (Table 2) was used in all cases.

	Ave. correlation	STD
$NF_i - CF_i$	0.9987737779	0.0005488169
$NF_i^b - CF_i^b$	0.9966993676	0.0020356817
$NF_i^s - CF_i^s$	0.9899659572	0.0029525582
$NF_i^t - CF_i^t$	0.9977074526	0.0007460563

doi:10.1371/journal.pone.0143760.t003

sequences and the nucleosome-inhibiting sequences are shown in Fig 4 and S5 Fig ∼ S7 Fig. Here, we employed the helical parameter set (i) (Table 2) that was used in the coarse-grained molecular dynamics simulations by Freeman et al., which exhibited consistent results to some experiments [48, 61–63].

The histograms for budding yeast showed that that nucleosome-forming sequences tend to exhibit larger fluctuations in several directions compared to the inhibiting sequences. In particular, the histogram of $< DF_i^b >_i$ for the nucleosome-forming sequences showed a clear shift in the direction toward larger values compared to that for the nucleosome-inhibiting sequences (Fig 4). For the other organisms, most of the histograms showed few differences between the nucleosome- forming and -inhibiting sequences. However, similar to the case of yeast, the distribution of $< DF_i^b >_i$ for the nucleosome-forming sequences shifted largely in the direction of larger values compared to that for the nucleosome-inhibiting sequences in these organisms (S5 Fig ∼ S7 Fig). These results indicate that the nucleosome-forming ability is highly correlated to the fluctuations of the inter-strand distances of DNAs, in which sequences with larger fluctuations tend to form the nucleosome.

Similar to the previous arguments, we compared the dynamical features among the random 150-bp DNA sequences varying in average GC-contents; the histograms, averages, and standard deviations of $< CF_i >_i$, $< DF_i >_i$, $< DF_i^b >_i$, $< DF_i^s >_i$, and $< DF_i^t >_i$ were measured from 10,000 random sequences for each GC content (Fig 5 and S8 Fig). In this case, $< CF_i >_i$ exhibited a minimum at a GC content of ∼0.2, which indicates that the AT-rich sequences tend to be more rigid than the GC-rich sequences. However, the fluctuations of sequences consisting of only A or T were as large as those of the GC-rich sequences. Moreover, the GC content dependencies of $< DF_i >_i$, $< DF_i^b >_i$, $< DF_i^s >_i$, and $< DF_i^t >_i$ showed different characteristics for GC contents larger or smaller than 0.6 ∼ 0.7. In particular, the results for $< DF_i^b >_i$ were similar for cases with a large GC content (0.7 ∼ 1) but monotonically decreased with a GC content with little variance. A recent experimental study showed that the probability of nucleosome formation tends to increase with increases in the GC content ratio [81]. Thus, the present results indicate that sequences with larger fluctuations of inter-strand distances tend to form the nucleosome, which is consistent with the results described above from the analysis of the genomes of the four model organisms.

Finally, we focus on the relationships between the overall geometries and fluctuations of the considered DNA sequences. The overall geometries of several DNA sequences, the nucleosome-forming and -inhibiting DNA sequences for four model organisms and random DNA sequences with different GC contents, were evaluated using scatter plots of σ_1 (linearity) and σ_2 (line symmetry) (Fig 6 and S9 Fig). It is noted that σ_1 and σ_2 are highly correlated. The average, standard deviation, and distributions of σ_1 and σ_2 exhibited slight but not significant deviations between the nucleosome-forming and -inhibiting sequences.

For the random sequences, the average value of σ_1 exhibited similar variations to $< CF_i >_i$ with an increase in GC content. In particular, both values decreased with increases in GC content for GC contents ≤0.2, whereas they increased with increases in GC content for GC contents ≥0.3 (Figs 5(c) and 6(c)). In fact, σ_1 and $< CF_i >$ were highly correlated for random DNA sequences, regardless of their CG content. The Pearson correlation coefficient for the 110,000 sequences analyzed above with GC contents = 0.0 ∼ 1.0 was 0.9433. It is noted that σ_1 showed large dispersion for each GC content, and there were significant overlaps among σ_1 distributions with different GC contents (Fig 6(b) and 6(c)). This fact indicates that different DNA sequences can often show similar geometries, and such sequences also tend to show similar overall fluctuations. On the other hand, the fluctuations of inter-strand distances $< DF_i^b >$ that may correlate to the nucleosome-forming ability did not correlate significantly to either σ_1

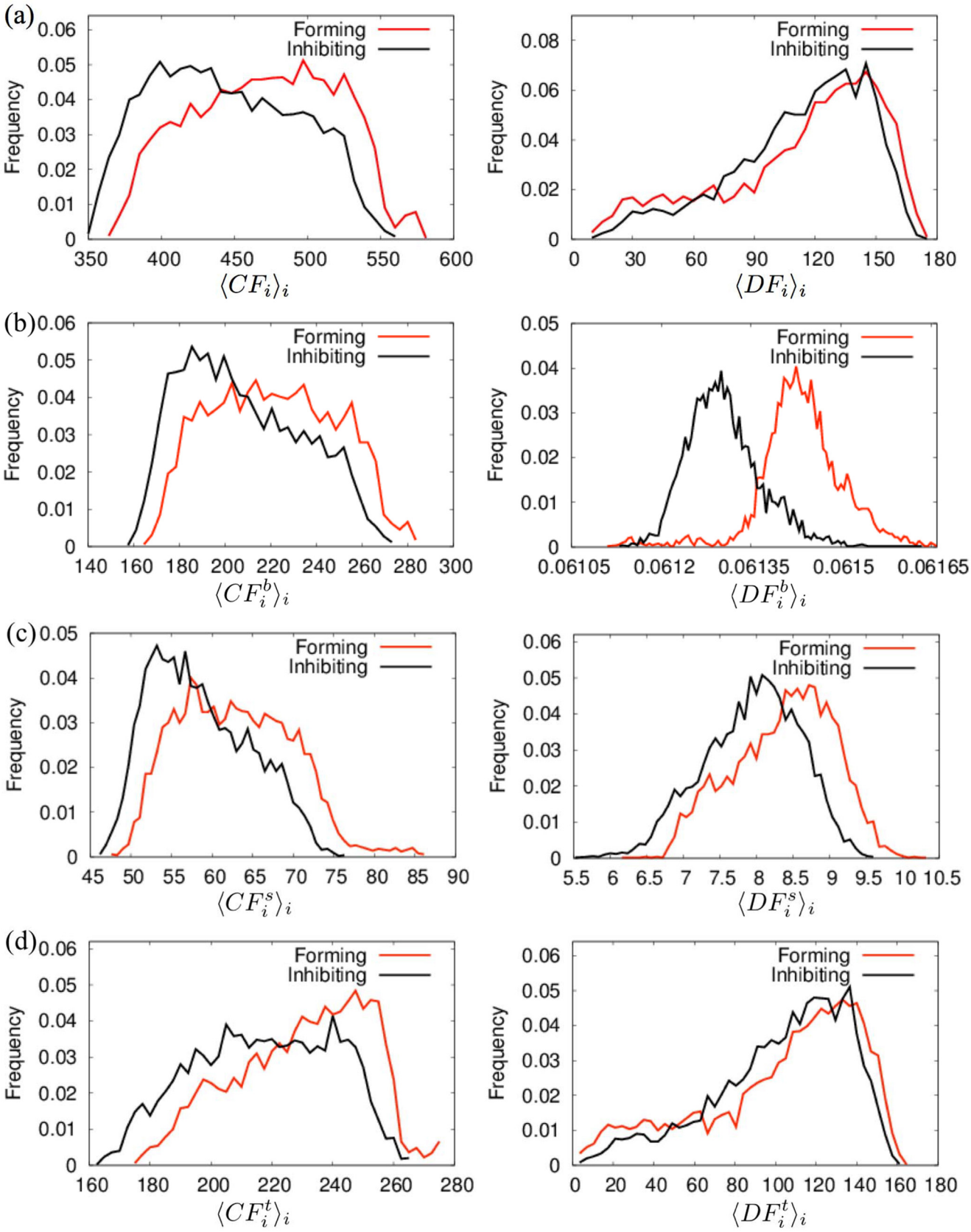

Fig 4. Fluctuations of the CGENM of long DNA sequences. Histograms of the average fluctuations, (a) $< CF_i >_i$ and $< DF_i >_i$, (b) $< CF_i^b >_i$ and $< DF_i^b >_i$, (c) $< CF_i^s >_i$ and $< DF_i^s >_i$, and (d) $< CF_i^t >_i$, and $< DF_i^t >_i$, for nucleosome-forming and -inhibiting sequences of budding yeast *Saccharomyces cerevisiae* (150 bp). Helical parameter set (i) (Table 2) was used.

doi:10.1371/journal.pone.0143760.g004

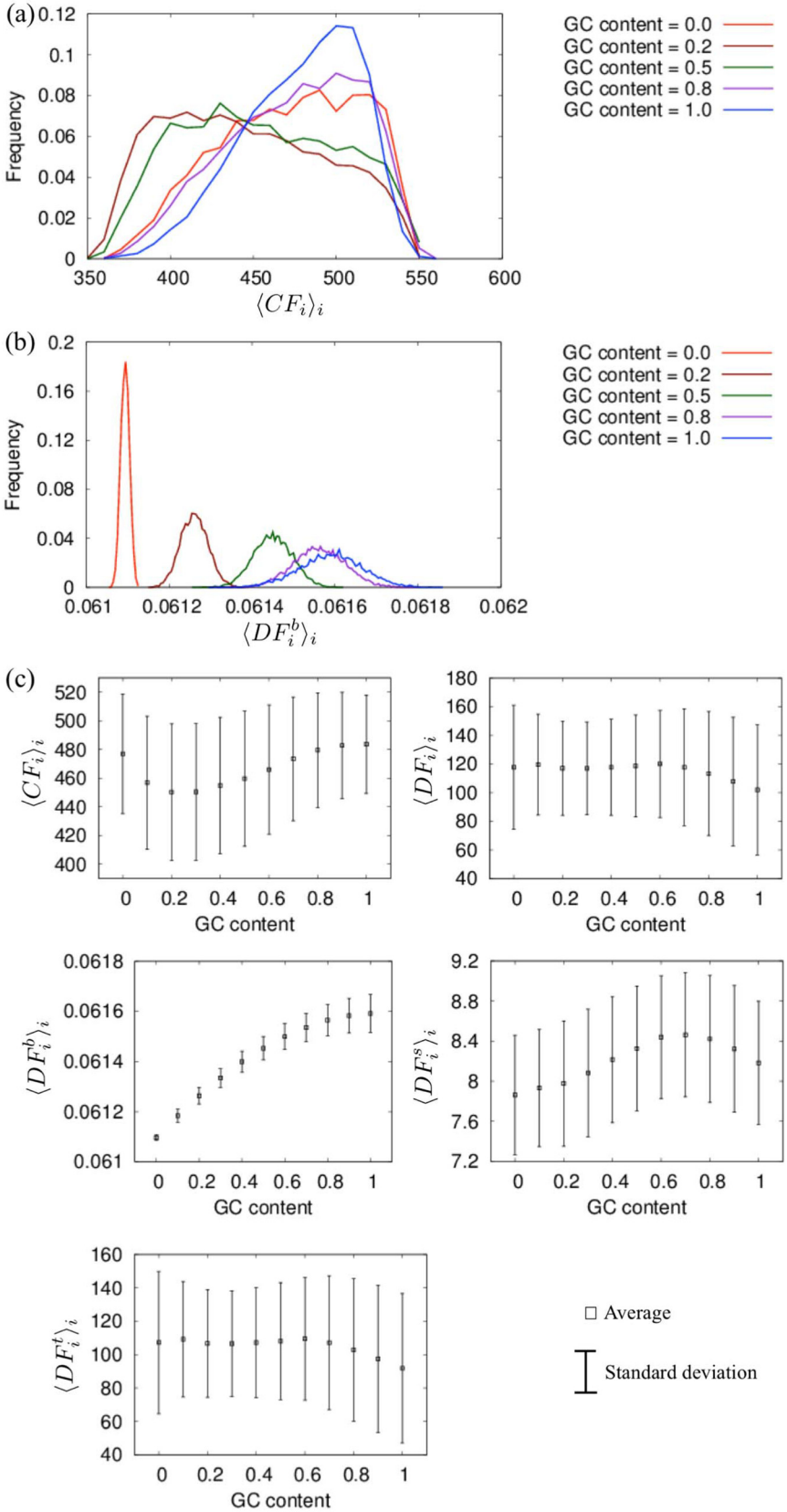

Fig 5. Fluctuations of the CGENM of long DNA sequences. Histograms of (a) $< CF_i >_i$ and (b) $< DF_i^b >_i$, and (c) averages and standard deviations of $< CF_i >_i$, $< DF_i >_i$, $< DF_i^b >_i$, $< DF_i^s >_i$, and $< DF_i^t >_i$ for 10,000 samples of random 150-bp sequences with an average GC content = 0, 0.1, 0.2, ⋯, and 1. Helical parameter set (i) (Table 2) was used.

doi:10.1371/journal.pone.0143760.g005

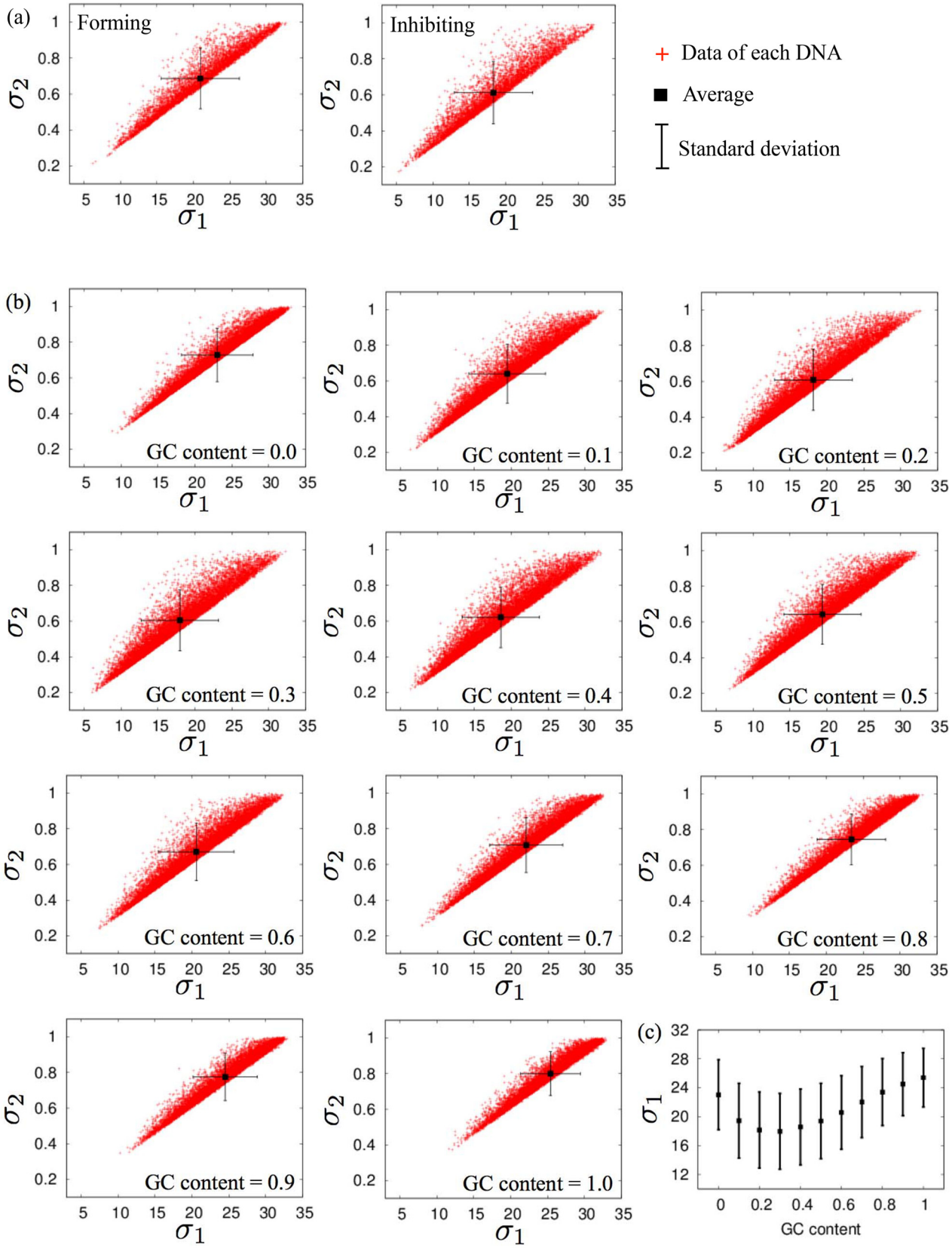

Fig 6. Overall geometries of long DNA sequences. Scatter plots of σ_1 and σ_2 for (a) nucleosome-forming and -inhibiting sequences of budding yeast *Saccharomyces cerevisiae* (150 bp), and for (b) random sequences with several GC contents (150 bp). (c) Averages and standard deviations of σ_1 for 10000 samples of random 150-bp sequences with an average GC content = 0, 0.1, 0.2, ⋯, and 1. Helical parameter set (i) (Table 2) was used.

doi:10.1371/journal.pone.0143760.g006

or σ_2, with Pearson correlation coefficients of 0.2250 and 0.1967. This fact indicates that the nucleosome-forming ability of DNA sequences are not only determined by the overall DNA geometries but also by their dynamic properties.

4 Summary and Conclusion

In this study, simple elastic network models of double-stranded DNAs were developed in order to perform an exhaustive analysis of several sequence-dependent dynamical features. First, we constructed a simple all-atom elastic network model that could reproduce the fluctuations of the motifs of each nucleotide (sugar, phosphoric acid, and bases) of several crystal structures of short DNA sequences. Second, we proposed a simple coarse-grained elastic network model that could reproduce the dynamic features of the long DNA sequences obtained by the all-atom elastic network model. Through exhaustive analysis of the dynamic features of several DNA sequences with normal-mode analysis of the presented coarse-grained elastic network model, we found that the dynamic aspects of DNA are highly influenced by the properties of nucleotide sequence such as GC content. We also found that the nucleosome-forming abilities of double-stranded DNA exhibited positive correlations with their sequence-dependent inter-strand fluctuations.

In the present study, we demonstrated the sequence-dependent dynamic features for several ∼150-bp DNA sequences to evaluate the relationships between the nucleosome-forming abilities and DNA dynamics. Of course, DNA sequences longer than ∼150 bp can also be analyzed using the presented coarse-grained model. Moreover, coarse-grained molecular dynamics simulations can be performed to consider the large deformations of DNA, such as formation of a super helix and nucleosome that are the basic structures of higher-order chromosome architectures, using the presented elastic network models with the excluded effect of the volume of each atom or each nucleotide. We are currently conducting these molecular dynamics simulations, and the results will be reported in the future. We did not consider the effects of solvents such as temperature and salt concentrations in the present elastic network models. Therefore, we are also planning to attempt modifications of the models so that several solvent conditions can be incorporated in future work.

Supporting Information

S1 Table. Helical parameter sets. (a) Helical parameter sets (ii) obtained by X-ray crystal structure analysis, [61, 62], and (b) helical parameter sets (iii) obtained by all-atom molecular dynamics simulations. [58, 65, 66].
(EPS)

S2 Table. Comparisons between CGENM and AAENM. Average and standard deviation of the correlation coefficients of 500 samples of random 50-bp sequences between NF_i and CF_i, NF_i^b and CF_i^b, NF_i^s, and CF_i^s, and NF_i^t and CF_i^t. Helical parameter sets (ii) and (iii) (S1 Table) were used.
(EPS)

S1 Fig. Temperature factor of each atom. The temperature factor of each atom obtained by the AAENMs (black curve) and X-ray crystal structure analysis (gray (red) curve) of typical double-stranded DNAs obtained from PDB ID (a) 7BNA, (b) 9BNA, (c) 1D91, (d) 1DC0, (e) 122D, (f) 123D, (g) 181D, and (h) 330D. Parameters C_a, B_B, and B_W are given in Table 1.
(EPS)

S2 Fig. Average temperature factor of motifs. Average temperature factor of motifs obtained by the AAENMs (black curve) and X-ray crystal structure analysis (gray (red) curve) of typical double-stranded DNAs obtained from PDB ID (a) 7BNA, (b) 9BNA, (c) 1D91, (d) 1DC0, (e) 122D, (f) 123D, (g) 181D, and (h) 330D. Parameters C_a, B_B, and B_W are given in Table 1.
(EPS)

S3 Fig. Comparisons between AAENM and CGENM. Comparisons of the fluctuations between each nucleotide in the AAENM (black curves) and CGENM (gray (red) curves) for a typical random 50-bp DNA sequence (5'—AGTGGTAAGGCATGGTTCTC-GAATCTCGGTTTATTTACACTGCTGCTCCA—3'). (a) NF_i and CF_i, (b) NF_i^b and CF_i^b, (c) NF_i^s and CF_i^s, and (d) NF_i^t and CF_i^t using helical parameter set (ii) S1 Table.
(EPS)

S4 Fig. Comparisons between AAENM and CGENM. Comparisons of the fluctuations between each nucleotide in the AAENM (black curves) and CGENM (gray (red) curves) for a typical random 50-bp DNA sequence (5'—ATATGCTGTA-GAGCGTCCCGTCCGCGCGTTGTGGTTTTTTCGGTGCTCTA—3'). (a) NF_i and CF_i, (b) NF_i^b and CF_i^b, (c) NF_i^s and CF_i^s, and (d) NF_i^t and CF_i^t using helical parameter set (iii) S1 Table.
(EPS)

S5 Fig. Histograms of the average fluctuations in *Caenorhabditis elegans*. Histograms of the average fluctuations of (a) $<CF_i>_i$ and $<DF_i>_i$, (b) $< CF_i^b >_i$ and $< DF_i^b >_i$, (c) $< CF_i^s >_i$ and $< DF_i^s >_i$, and (d) $< CF_i^t >_i$ and $< DF_i^t >_i$ for nucleosome-forming and nucleosome-inhibiting sequences of the nematode *Caenorhabditis elegans* (147 bp). Helical parameter set (i) (Table 2) was used.
(EPS)

S6 Fig. Histograms of the average fluctuations in *Drosophila melanogaster*. Histograms of the average fluctuations of (a) $< CF_i >_i$ and $< DF_i >_i$, (b) $< CF_i^b >_i$ and $< DF_i^b >_i$, (c) $< CF_i^s >_i$ and $< DF_i^s >_i$, and (d) $< CF_i^t >_i$ and $< DF_i^t >_i$ for nucleosome-forming and nucleosome-inhibiting sequences of *Drosophila melanogaster* (147 bp). Helical parameter set (i) (Table 2) was used.
(EPS)

S7 Fig. Histograms of the average fluctuations in *Homo sapiens*. Histograms of the average fluctuations of (a) $< CF_i >_i$ and $< DF_i >_i$, (b) $< CF_i^b >_i$ and $< DF_i^b >_i$, (c) $< CF_i^s >_i$ and $< DF_i^s >_i$, and (d) $< CF_i^t >_i$ and $< DF_i^t >_i$ for nucleosome-forming and nucleosome-inhibiting sequences of *Homo sapiens* (147 bp). Helical parameter set (i) (Table 2) was used.
(EPS)

S8 Fig. Histograms of the average fluctuations with different GC contents. Histograms of (a) $< DF_i >_i$, (b) $< DF_i^s >_i$, and (c) $< DF_i^t >_i$ for 10,000 samples of random 150-bp sequences with different average GC contents. Helical parameter set (i) (Table 2) was used.
(EPS)

S9 Fig. Overall geometries of long DNA sequences in model organisms. Scatter plots of σ_1 and σ_2 for nucleosome-forming and -inhibiting sequences of (a) *Caenorhabditis elegans* (147 bp), (b) *Drosophila melanogaster* (147 bp), and (c) *Homo sapiens* (147 bp). Helical parameter set (i) (Table 2) was used.
(EPS)

Acknowledgments

The author is grateful to S. Tate and Y. Murayama for fruitful discussions.

Author Contributions

Conceived and designed the experiments: AA NS. Performed the experiments: SI AA. Analyzed the data: SI AA. Wrote the paper: AA NS HN.

References

1. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77: 1905–1908. doi: 10.1103/PhysRevLett.77.1905 PMID: 10063201
2. Hinsen K, Petrescu AJ, Dellerue S, Bellisent-Funel MC, Kneller GR (2000) Harmonicity in slow protein dynamics. Chem Phys 261: 25–37. doi: 10.1016/S0301-0104(00)00222-6
3. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, et al. (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80: 505–515. doi: 10.1016/S0006-3495(01)76033-X PMID: 11159421
4. Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15: 586–592. doi: 10.1016/j.sbi.2005.08.007 PMID: 16143512
5. Bahar I, Lezon TR, Yang LW, Eyal E (2010) Global dynamics of proteins: bridging between structure and function. Annu Rev Biophys 39: 23–42. doi: 10.1146/annurev.biophys.093008.131258 PMID: 20192781
6. Bahar I, Lezon TR, Bakan A, Shrivastava IH (2010) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev 110: 1463–1497. doi: 10.1021/cr900095e PMID: 19785456
7. Dykeman EC, Sankey OF (2010) Normal mode analysis and applications in biological physics. J Phys Condens Matter 22: 423202. doi: 10.1088/0953-8984/22/42/423202 PMID: 21403307
8. Flechsig H, Mikhailov AS (2010) Tracing entire operation cycles of molecular motor hepatitis C virus helicase in structurally resolved dynamical simulations. Proc Natl Acad Sci (USA) 107: 20875–20880. doi: 10.1073/pnas.1014631107
9. Atilgan C, Okan OB, Atilgan AR (2012) Network-based models as tools hinting at nonevident protein functionality. Annu Rev Biophys 41: 205–25. doi: 10.1146/annurev-biophys-050511-102305 PMID: 22404685
10. Van Wynsberghe AW, Cui Q (2006) Interpreting correlated motions using normal mode analysis. Structure 14: 1647–1653. doi: 10.1016/j.str.2006.09.003 PMID: 17098190
11. Tama F, Brooks CL (2006) Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu Rev Biophys Biomol Struct 35: 115–33. doi: 10.1146/annurev.biophys.35.040405.102010 PMID: 16689630
12. Moritsugu K, Smith JC (2007) Coarse-grained biomolecular simulation with REACH: realistic extension algorithm via covariance Hessian. Biophys J 93: 3460–3469. doi: 10.1529/biophysj.107.111898 PMID: 17693469
13. Yang L, Song G, Jernigan RL (2009) Protein elastic network models and the ranges of cooperativity. Proc Natl Acad Sci (USA) 106: 12347–12352. doi: 10.1073/pnas.0902159106
14. Dehouck Y, Mikhailov AS (2013) Effective harmonic potentials: insights into the internal cooperativity and sequence-specificity of protein dynamics. PLoS Comp Bio 9: e1003209 doi: 10.1371/journal.pcbi.1003209
15. Albert B, et al. (2008) Molecular Biology of the Cell 5th edition. Garland Science, Taylor & Francis Group, LLC.
16. Latchman DS (2010) Gene Control. Garland Science, Taylor & Francis Group, LLC.
17. Fukue Y, Sumida N, Nishikawa J, Ohyama T (2004) Core promoter elements of eukaryotic genes have a highly distinctive mechanical property. Nucleic Acids Res 32: 5834–5840. doi: 10.1093/nar/gkh905 PMID: 15520466
18. Ohyama T (2005) DNA Conformation and Transcription. New York, NY, USA: Springer, Landes Bioscience.
19. Valenzuela L, Kamakaka RT (2006) Chromatin insulators. Annu Rev Genet 40: 107–138. doi: 10.1146/annurev.genet.39.073003.113546 PMID: 16953792

20. Bushey AM, Dorman ER, Corces VG (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. Mol Cell 32: 1–9. doi: 10.1016/j.molcel.2008.08.017 PMID: 18851828
21. Geyer PK, Clark I (2002) Protecting against promiscuity: the regulatory role of insulators. Cell Mol Life Sci 59: 2112–2127. doi: 10.1007/s000180200011 PMID: 12568337
22. Gaszner M, Felsenfeld G (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet 7: 703–713. doi: 10.1038/nrg1925 PMID: 16909129
23. Raab JR, Kamakaka RT (2010) Insulators and promoters: closer than we think. Nat Rev Genet 11: 439–446. doi: 10.1038/nrg2765 PMID: 20442713
24. Takagi H, Inai Y, Watanabe S, Tatemoto S, Yajima M, Akasaka K, et al. (2012) Nucleosome exclusion from the interspecies-conserved central AT-rich region of the Ars insulator. J Biochem 151: 75–87. doi: 10.1093/jb/mvr118 PMID: 21930654
25. Turner BM (2001) Chromatin and Gene Regulation. Blackwell Science Ltd.
26. Dekker J, Rippe K, Dekker M, and Kleckner N (2002) Capturing chromosome conformation. Science 295: 1306–1311. doi: 10.1126/science.1067799 PMID: 11847345
27. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. Mol Cell 10: 1453–1465. doi: 10.1016/S1097-2765(02)00781-5 PMID: 12504019
28. Allis CDet al. (2007) Epigenetics. Cold Spring Harbor Laboratory Press.
29. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485: 376–380. doi: 10.1038/nature11082 PMID: 22495300
30. Apostolou E, Ferrari F, Walsh RM, Bar-Nur O, Stadtfeld M, Cheloufi S, et al. (2013) Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell 13: 1–14.
31. Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. Mol Cell 49: 773–782. doi: 10.1016/j.molcel.2013.02.011 PMID: 23473598
32. Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Gen 14: 390–403. doi: 10.1038/nrg3454
33. Bickmore WA, van Steensel B (2013) Genome architecture: domain organization of interphase chromosomes. Cell 152: 1270–1284. doi: 10.1016/j.cell.2013.02.001 PMID: 23498936
34. Lin D, Matsumoto A, Go N (1997) Normal mode analysis of a double-stranded DNA dodecamer d (CGCGAATTCGCG). J Chem Phys 107: 3684–3690. doi: 10.1063/1.474724
35. Matsumoto A, Olson WK (2002) Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. Biophys J 83: 22–41. doi: 10.1016/S0006-3495(02)75147-3 PMID: 12080098
36. Lankas F, Sponer J, Langowski J, Cheatham TE 3rd (2003) DNA basepair step deformability inferred from molecular dynamics simulations. Biophys J 85: 2872–2883. doi: 10.1016/S0006-3495(03)74710-9 PMID: 14581192
37. Fujii S, Kono H, Takenaka S, Go N, Sarai A (2007) Sequence-dependent DNA deformability studied using molecular dynamics simulations. Nucl Acids Res 35: 6063–6074. doi: 10.1093/nar/gkm627 PMID: 17766249
38. Orozco M, Noy A, Pe̋rez A (2008) Recent advances in the study of nucleic acid flexibility by molecular dynamics. Curr Opin Struct Biol 18: 185–193. doi: 10.1016/j.sbi.2008.01.005 PMID: 18304803
39. Bomble YJ, Case DA (2008) Multiscale modeling of nucleic acids: Insights into DNA flexibility. Biopolymers 89: 722–731. doi: 10.1002/bip.21000 PMID: 18412139
40. Beveridge DL, Cheatham TE 3rd, Mezei M (2012) The ABCs of molecular dynamics simulations on B-DNA, circa 2012. J Biosci 37: 379–397. doi: 10.1007/s12038-012-9222-6 PMID: 22750978
41. Cheatham TE 3rd, Case DA (2013) Twenty-five years of nucleic acid simulations. Biopolymers 99: 969–977. doi: 10.1002/bip.22331 PMID: 23784813
42. Pasi M, Maddocks JH, Beveridge D, Bishop TC, Case DA, Cheathm T 3rd, et al. (2014) ABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. Nucl Acids Res 42: 12272–12283. doi: 10.1093/nar/gku855 PMID: 25260586
43. Knotts IV TA, Rathore N, Schwartz DC, de Pablo JJ (2007) A coarse grain model for DNA. J Chem Phys 126: 084901. doi: 10.1063/1.2431804
44. Morriss-Andrews A, Rottler J, Plotkin SS (2010). A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. J Chem Phys 132: 035105. doi: 10.1063/1.3269994 PMID: 20095755

45. Dans PD, Zida A, Machado MR, Pantano S (2010) A coarse grained model for atomic-detailed DNA simulations with explicit electrostatics. J Chem Theor Comp 6: 1711–1725. doi: 10.1021/ct900653p

46. Freeman GS, Hinckley DM, de Pablo JJ (2011) A coarse-grain three-site-per-nucleotide model for DNA with explicit ions. J Chem Phys 135: 165104. doi: 10.1063/1.3652956 PMID: 22047269

47. Freeman GS, Lequieu JP, Hinckley DM, Whitmer JK, de Pablo JJ (2014) DNA shape dominates sequence affinity in nucleosome formation. Phys Rev Lett 113: 168101. doi: 10.1103/PhysRevLett.113.168101 PMID: 25361282

48. Freeman GS, Hinckley DM, Lequieu JP, Whitmer JK, Pablo JJ (2014) Coarse-grained modeling of DNA curvature. J Chem Phys 141: 165103. doi: 10.1063/1.4897649 PMID: 25362344

49. Pinamonti G, Bottaro S, Micheletti C, Bussi G (2015) Elastic network models for RNA: a comparative assessment with molecular dynamics and SHAPE experiments. Nucleic acids res. gkv708.

50. Savelyev A, Papoian GA (2010) Chemically accurate coarse graining of double-stranded DNA. Proc Nat Acad Sci 107: 20340–20345. doi: 10.1073/pnas.1001163107 PMID: 21059937

51. Setny P, Zacharias M (2013) Elastic network models of nucleic acids flexibility. J Chem Theor Comp 9: 5460–5470. doi: 10.1021/ct400814n

52. Gonzalez O, Petkeviciute D, Maddocks JHJ (2013) A sequence-dependent rigid-base model of DNA. J Chem Phys 138: 055102. doi: 10.1063/1.4789411 PMID: 23406150

53. Naome A, Laaksonen A, Vercauteren DP (2014) A solvent-mediated coarse-grained model of DNA derived with the systematic Newton inversion method. J Chem Theor Comp 10: 3541–3549. doi: 10.1021/ct500222s

54. Korolev N, Luo D, Lyubartsev AP, Nordensklold L (2014) A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. Polymers 6: 1655–1675. doi: 10.3390/polym6061655

55. Goñi JR, Fenollosa C, Perez A, Torrents D, Orozco M (2008) DNAlive: a tool for the physical analysis of DNA at the genomic scale. Bioinformatics 24: 1731–1732. doi: 10.1093/bioinformatics/btn259

56. Stolz RC, Bishop TC (2010) ICM Web: the interactive chromatin modeling web server. Nucleic Acids Res 38: W254–W261. doi: 10.1093/nar/gkq496 PMID: 20542915

57. Chen W, Lin H, Feng PM, Ding C, Zuo YC, Chou KC (2012) iNuc-PhysChem: A sequence-based predictor for identifying nucleosomes via physicochemical properties. PLos One 7: e47843. doi: 10.1371/journal.pone.0047843 PMID: 23144709

58. Hospital A, Faustino I, Collepardo-Guevara R, Gonzalez C, Gelpi JL, Orozco M (2013) NAFlex: a web server for the study of nucleic acid flexibility. Nucleic Acids Res 41: W47–W55. doi: 10.1093/nar/gkt378 PMID: 23685436

59. Guo SH, Deng EZ, Xu LQ, Ding H, Lin H, Chen W (2014) iNuc-PseKNC: a sequence-based predictor for predicting nucleosome positioning in genomes with pseudo k-tuple nucleotide composition. Bioinformatics 30: 1522–1529. doi: 10.1093/bioinformatics/btu083 PMID: 24504871

60. Petkeviciūtė D, Pasi M, Gonzalez O, Maddocks JH (2014) cgDNA: a software package for the prediction of sequence-dependent coarse-grain free energies of B-form DNA. Nucleic Acids Res 42: e153. doi: 10.1093/nar/gku825

61. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB (1998) DNA sequence-dependent deformability deduced from protein DNA crystal complexes. Proc Nat Acad Sci 95: 11163–11168. doi: 10.1073/pnas.95.19.11163 PMID: 9736707

62. Olson WK, Colasanti AV, Li Y, Ge W, Zheng G, Zhurkin VB (2006) DNA simulation benchmarks as revealed by X-ray structures. In: Computational studies of RNA and DNA. Springer Netherlands: 235–257.

63. Morozov AV, Fortney K, Gaykalova DA, Studitsky VM, Widom J, Siggia ED (2009) Using DNA mechanics to predict in vitro nucleosome positions and formation energies. Nucleic Acids Res 37: 4707–4722. doi: 10.1093/nar/gkp475 PMID: 19509309

64. Perez A, Lankas F, Luque FJ, Orozco M (2008) Towards a molecular dynamics consensus view of B-DNA flexibility. Nucleic Acids Res 36: 2379–2394. doi: 10.1093/nar/gkn082 PMID: 18299282

65. Lavery R, Zakrzewska K, Beveridge D, Bishop TC, Case DA, Cheatham T, et al. (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. doi: 10.1093/nar/gkp834 PMID: 19850719

66. Dans PD, Perez A, Faustino I, Lavery R, Orozco M (2012). Exploring polymorphisms in B-DNA helical conformations. Nucleic Acids Res 40: 10668–10678. doi: 10.1093/nar/gks884 PMID: 23012264

67. Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three dimensional nucleic acid structures. Nucleic Acids Res 31:5108–5121. doi: 10.1093/nar/gkg680 PMID: 12930962

68. Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, et al. (1981) Structure of a B-DNA dodecamer: conformation and dynamics. Proc Natl Acad Sci USA 78: 2179–2183. doi: 10.1073/pnas.78.4.2179 PMID: 6941276

69. Holbrook SR, Dickerson RE, Kim SH (1985). Anisotropic thermal-parameter refinement of the DNA dodecamer CGCGAATTCGCG by the segmented rigid-body method. Acta Crystallogr B 41:255–262. doi: 10.1107/S0108768185002087

70. Westhof E (1987). Re-refinement of the B-dodecamer d(CGCGAATTCGCG) with a comparative analysis of the solvent in it and in the Z-hexamer d(5BrCG5BrCG5BrCG). J Biomol Struct Dyn 5:581–600. doi: 10.1080/07391102.1987.10506414 PMID: 3271485

71. Kneale G,Brown T, Kennard O, Rabinovich D (1985). G.T base-pairs in a DNA helix: the crystal structure of d(G-G-G-G-T-C-C-C). J Mol Biol 186: 805–814. doi: 10.1016/0022-2836(85)90398-5 PMID: 4093986

72. Ng HL, Kopka ML, Dickerson RE (2000). The structure of a stable intermediate in the A;-> B DNA helix transition. Proc Natl Acad Sci USA 97:2035–2039. doi: 10.1073/pnas.040571197 PMID: 10688897

73. Hahn M, Heinemann U (1993). DNA helix structure and refinement algorithm: comparison of models for d(CCAGGCm5CTGG) derived from NUCLSQ, TNT and X-PLOR. Acta Crystallogr D 49: 468–477. doi: 10.1107/S0907444993004858 PMID: 15299506

74. Sadasivan C, Gautham N (1995). Sequence-dependent microheterogeneity of Z-DNA: the crystal and molecular structures of d(CACGCG).d(CGCGTG) and d(CGCACG).d(CGTGCG). J Mol Biol 248: 918–930. doi: 10.1006/jmbi.1995.9894 PMID: 7760333

75. Timsit Y, Vilbois E, Moras D (1991). Base-pairing shift in the major groove of (CA)n tracts by B-DNA crystal structures. Nature 354: 167–170. doi: 10.1038/354167a0 PMID: 1944598

76. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom AC, Field Y, Moore IK, et al. (2006). A genomic code for nucleosome positioning. Nature 442: 772–778. doi: 10.1038/nature04979 PMID: 16862119

77. Lee W, Tillo D, Bray N, Morse RH, Davis RW, et al. (2007) A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet 39: 1235–1244. doi: 10.1038/ng2117 PMID: 17873876

78. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, et al. (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell 132: 887–898. doi: 10.1016/j.cell.2008.02.022 PMID: 18329373

79. Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, et al. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res 18: 1073–1083. doi: 10.1101/gr.078261.108 PMID: 18550805

80. Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, et al. (2008) Nucleosome organization in the Drosophila genome. Nature 453: 358–362. doi: 10.1038/nature06929 PMID: 18408708

81. Tillo D, Hughes TR (2009). G+ C content dominates intrinsic nucleosome occupancy. BMC Bioinform 1: 442. doi: 10.1186/1471-2105-10-442