Supporting Information

Confinement-Driven Organization of a Histone-Complexed DNA Molecule in a Dense Array of Nanoposts

Heesun Joo,[†] Jun Soo Kim[*,†]

[†]Department of Chemistry and Nanoscience, Ewha Womans University, Seoul 03760, Republic of Korea

*Corresponding author: jkim@ewha.ac.kr

Contents

I. Simulation Models

1. A Histone-Complexed DNA (hcDNA) Molecule

A DNA molecule was modeled as a semiflexible polymer chain composed of 1,450 monomers. The diameter of each monomer, denoted σ, was set to 2 nm (the same as the diameter of B-DNA) and used as a unit of length in this work. Since B-DNA molecules have a rise of 0.34 nm per base pair, each monomer unit of 2-nm size represented 6 base pairs (bp) of a DNA molecule and the total DNA length amounted to 8.7 kilobase pairs (kbp) or 3 μm. The bond between neighboring monomers of a DNA molecule was modeled by a combination of a finitely extensible nonlinear elastic (FENE) potential energy

$$U_{fene}(r) = -\frac{1}{2}k_{fene}R_{fene}^2 \ln\left[1-\left(\frac{r}{R_{fene}}\right)^2\right]$$

and the repulsive part of the Lennard-Jones (LJ) potential energy

$$U_{rLJ}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & 0 < r < r_{cut} \\ 0 & elsewhere \end{cases},$$

where $k_{fene} = 30k_BT/\sigma^2$, $R_{fene} = 1.5\sigma$, $\epsilon = k_BT$, and $r_{cut} = 2^{1/6}\sigma$. Here, k_B is the Boltzmann constant, T is the temperature, and $\sigma = 2$ nm (as defined above). The use of these parameters ensures the topological constraint on the DNA molecule, preventing the bonds from crossing each other.[1] The angle formed by two consecutive bond vectors was restrained by harmonic potential energy

$$U_{angle} = \frac{1}{2}k_{angle}(\theta - \theta_0)^2,$$

where $k_{angle} = 25\, k_BT/\text{rad}^2$ and $\theta_0 = 0$, resulting in the persistence length of 24.7σ (= 49.4 nm) from the bond correlation function as explained below in the discussion of the persistence length of a hcDNA molecule. The non-bonded interaction between DNA monomers was modeled by the repulsive part of the LJ potential energy, U_{rLJ}, described above.

A histone-complexed DNA (hcDNA) molecule is a complex formed by a DNA molecule and histone protein complexes. In this work, it was modeled by a

complex of the DNA model and spherical particles with a diameter of 3.5σ (= 7 nm), to mimic the beads-on-a-string conformation of a nucleosome array as found in the eukaryotic cell nucleus. 50 spherical particles were equally distributed over the 1,450 DNA monomers. Each spherical particle was wrapped by 24 DNA monomers by the harmonic restraining potential energy

$$U_{wrap} = \frac{1}{2} k_{wrap} \left(r - r_{wrap}\right)^2,$$

where $k_{wrap} = 1{,}000\, k_B T/\sigma^2$ and $r_{wrap} = 2.25\sigma$. Neighboring complexes of a spherical particle and 24 DNA monomers were separated by 5 DNA monomers (corresponding to 30 bps of linker DNA), as shown in Figure 1 (a) of the main text. The presence of linker histones was not considered in this work. In addition, electrostatic interactions were not considered by assuming the electroneutrality of nanoposts. The non-bonded interactions between a DNA monomer and a spherical particle and between spherical particles were modeled by the repulsive part of the LJ potential energy, as follows.

$$U_{nb}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r - r_0} \right)^{12} - \left(\frac{\sigma}{r - r_0} \right)^{6} \right] + \epsilon & r_0 < r < r_0 + r_{cut} \\ 0 & elsewhere \end{cases},$$

where $r_0 = \frac{D_1 + D_2}{2} - \sigma$ and $r_{cut} = 2^{1/6}\sigma$. Here, D_1 and D_2 are the diameters of DNA monomers (1σ = 2 nm) or histone-mimicking spherical particles (3.5σ = 7 nm), depending on the non-bonded interactions considered, that is, between a DNA monomer and a spherical particle or between spherical particles.

The persistence length, l_p, of a hcDNA molecule was estimated to be 14.5σ (= 29 nm), as shown in Figure S1. The persistence length of polymer chains is typically obtained by calculating the bond correlation function[2] of $\langle \hat{b}_i \cdot \hat{b}_j \rangle = \exp\left(-|i-j|\bar{b}/l_p\right)$, where $\hat{b}_i$ and $\hat{b}_j$ are unit bond vectors and $\bar{b}$ is the average bond length. Initially, the persistence length was obtained for a chain constructed by connecting consecutive spherical particles bound to a DNA molecule (shown as a chain of black arrows in Figure S1 (a)). However, due to the zigzag conformation of this "sphere-connected chain" (as seen in a nucleosome array), the persistence length obtained showed a rapidly decaying oscillation of the correlation function. Instead, we smoothed the "sphere-

connected chain" by constructing an alternative chain that connects the points, each of which is defined at an average location of two consecutive spherical particles. This smoothed chain, termed a "2-particle average" chain, is depicted as a collection of red spheres in Figure S1 (a), and its bond correlation function is presented in Figure S1 (b). In addition, we also calculated the persistence length based on "3-particle average", "4-particle average", and "5-particle average" chain conformations that were constructed by connecting the average locations of 3, 4, and 5 consecutive spherical particles, respectively. The bond correlation functions obtained for these chain conformations are also presented in Figure S1 (b). It was found that the "4-particle average" chain provided the best fit to the exponential function, and we concluded that the persistence length of a hcDNA molecule is 14.5σ (= 29 nm).

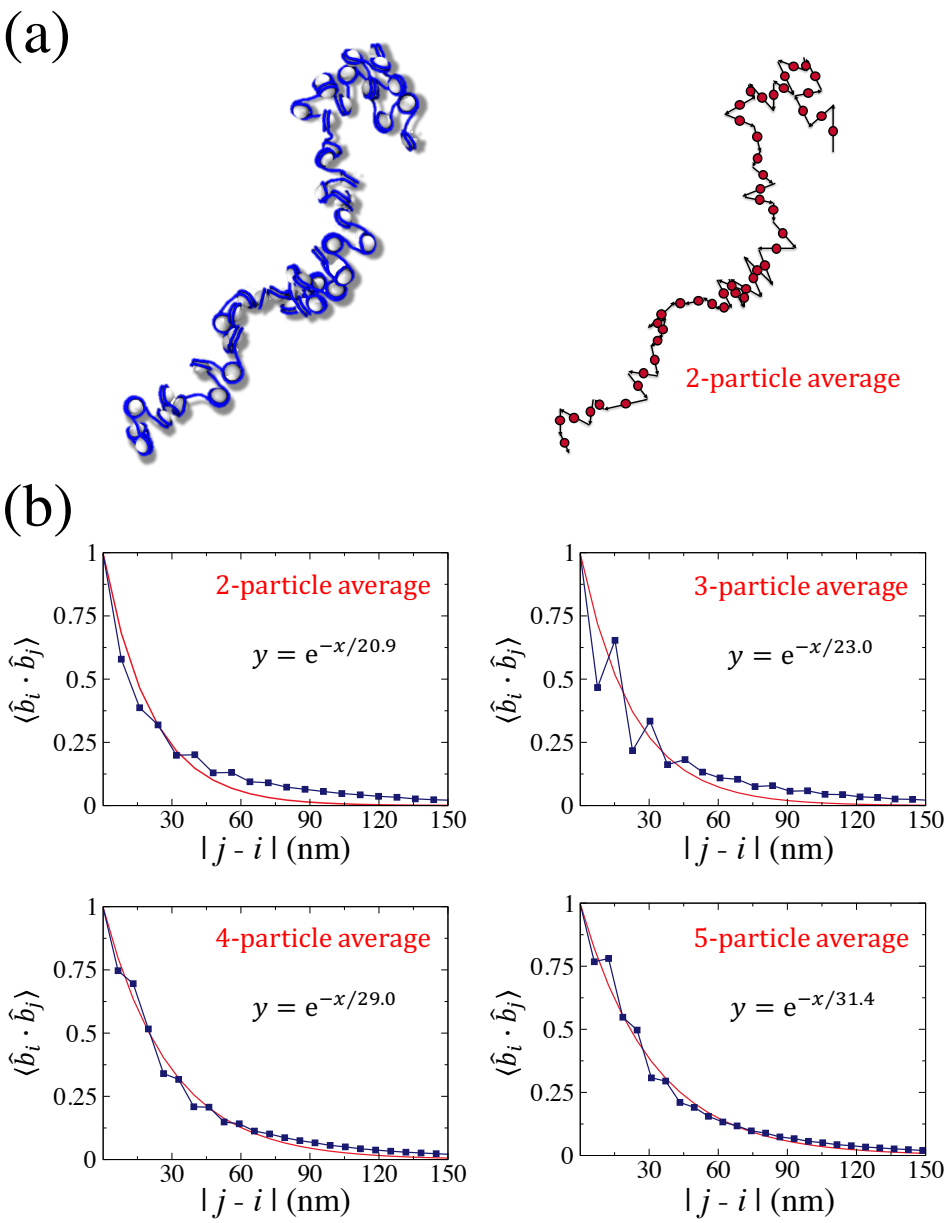

Figure S1. Calculation of the persistence length of a hcDNA molecule in bulk. (a) A snapshot of one of the hcDNA conformations in bulk and its "sphere-connected chain" representation (in black lines) that was constructed by connecting the histone-mimicking spherical particles. The average locations of two consecutive spherical particles are depicted as red, filled circles, whose connection was termed a "2-particle average" chain as described in the text above. (b) Bond correlation functions obtained for 2-, 3-, 4-, and 5-particle average chains. The "4-particle average" shows the best fit to the exponential function with a persistence length of 29 nm.

2. An Array of Nanoposts

The main purpose of this work was to understand the confinement effect of various nanopost arrays on the conformation of a hcDNA molecule. Therefore, we varied the dimensions of the nanopost arrays for each simulation. The dimensions of a nanopost array are defined by the set of thickness (D_p) and separation (S_p) of nanoposts. The parameters D_p and S_p are described in Figure 1 (b) of the main text. Alternatively, the dimensions of a nanopost array can be described by the size of the inter-post spaces surrounded by four nanoposts (D_c) and the width of the passage between two neighboring nanoposts (W_p), as also shown in Figure 1 (b) of the main text. D_c and W_p can be calculated from D_p and S_p using $D_c = \sqrt{2}S_p - D_p$ and $W_p = S_p - D_p$. In the main text, the simulation results in Figure 3 are interpreted in terms of D_p and W_p for convenience. All the values of S_p, D_p, D_c, and W_p are given in units of nanometers in this work.

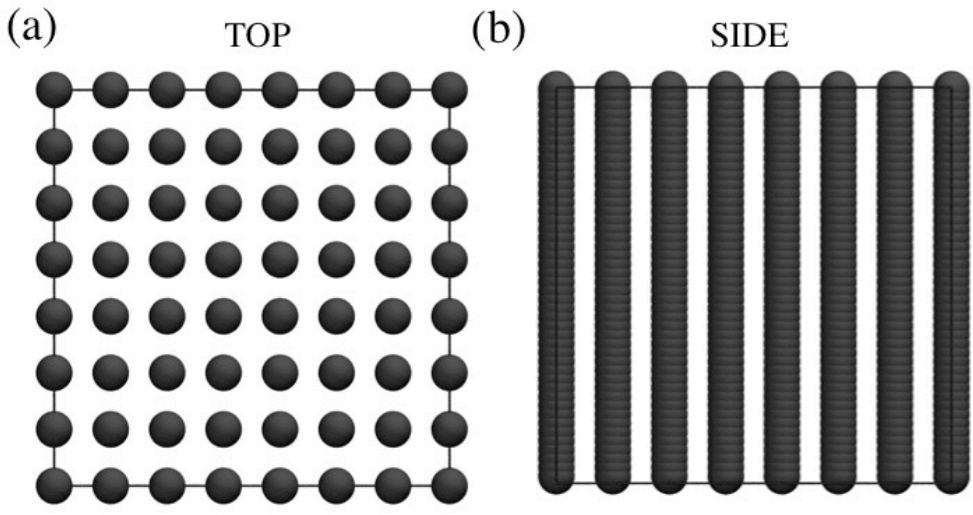

Figure S2. A model of an array of nanoposts with (S_p, D_p) = (64, 40) in units of nanometers. Each nanopost was constructed by closely stacking spherical particles with a diameter of D_p. The side view in (b) shows the stacking of spheres, which is close enough to mimic a smooth surface of a cylindrical nanopost. The square box represents the boundary of the simulation system used with the periodic boundary condition.

Nanoposts were modeled as cylindrical posts constructed by stacking large spherical particles with a diameter of D_p. These spheres were stacked so closely

(with a center-to-center distance of 10 nm) compared to the value of D_p (ranging from 40 nm to 80 nm) that each pile of spheres mimicked a smooth cylindrical post, as shown in Figure S2.

The spheres comprising nanoposts were frozen at each position without any interaction between them. However, they interact with DNA monomers and histone-mimicking spherical particles by the repulsive part of the shifted LJ potential as shown below.

$$U_{post}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r-r_0}\right)^{12} - \left(\frac{\sigma}{r-r_0}\right)^{6}\right] + \epsilon & r_0 < r < r_0 + r_{cut} \\ 0 & elsewhere \end{cases},$$

where $r_0 = \frac{D_p + D_0}{2} - \sigma$ and $r_{cut} = 2^{1/6}\sigma$. Here, D_p is the diameter of the nanopost spheres (ranging between 20σ and 40σ), and D_0 is either the diameter of DNA monomers (1σ = 2 nm) or histone-mimicking spherical particles (3.5σ = 7 nm). The use of the shifted LJ potential with the given parameters was intended to model the same extent of stiff repulsions on nanopost surfaces with that on each DNA monomer and histone-mimicking spherical particles.

II. Simulation Methods

Choice of N_{MD} and N_{MC} in Hybrid Monte Carlo (HMC) Simulation

The advantage of the HMC method[3] is that momenta and coordinates for all of the particles in a hcDNA molecule can be updated collectively by the MD simulations that may accelerate sampling of the conformations. In this work, the MD simulations were performed in the constant N, V, and E condition using GROMACS v. 4.6.5.[4] In the MD sampling, a very large time step dt for numerical integration is allowed even though the total energy may not be conserved, as long as the integrator is time reversible and area preserving to satisfy the detailed balance. However, if dt is too large, the acceptance ratio becomes very small due to larger discretization errors, *i.e.*, large values of δH. For a given value of $dt = 0.005\,\tau_{MD}$, the acceptance ratio of trial conformations decreases as N_{MD} increases. Therefore, we tested several values of N_{MD} while keeping the same value of dt. For a test simulation, we calculated the probability, $P(n)$, that a hcDNA molecule is distributed over n inter-post spaces for a nanopost array with (S_p, D_p) = (68, 40) because large conformational changes were expected in this nanopost array from the preliminary simulations. The comparison of $P(n)$ with different N_{MD} is presented in Figure S3. It was shown that $P(n)$'s with $N_{MD} \geq$ 30,000 were converged, and therefore we used N_{MD} = 40,000 for all other simulations. It is notable that the total number of the MD steps in the HMC simulations was kept fixed at $N_{tot} = 8\times10^8$. Therefore, N_{MC} = 20,000 when N_{MD} = 40,000.

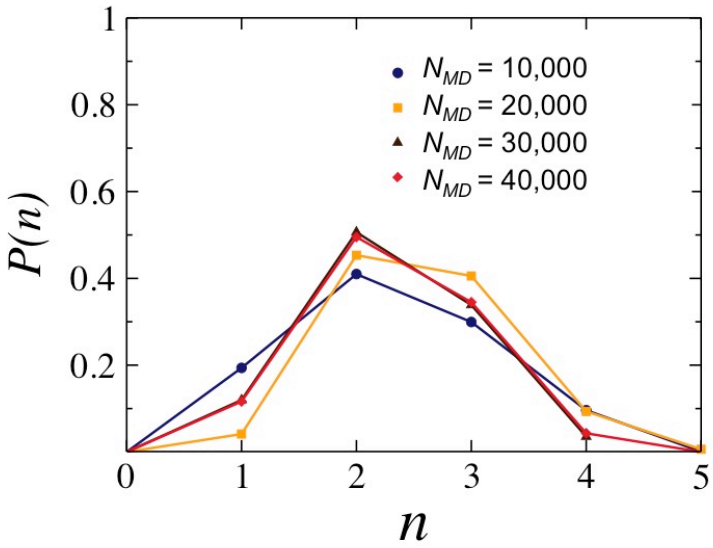

Figure S3. Probability, $P(n)$, that a hcDNA molecule is distributed over n inter-post spaces, calculated for a hcDNA molecule in a nanopost array with $(S_p, D_p) = (68, 40)$ using different values of N_{MD}.

III. Derivation of Equation (1) of the Main Text

The partition function Z of a hcDNA molecule in a dense array of nanoposts can be written as follows.

$$Z = \sum_{\{\alpha\}} e^{-\beta E_\alpha} \tag{S1}$$

where $\{\alpha\}$ refers to all possible hcDNA conformations in the nanopost array, β is $1/k_B$T, and E_α is an energy of hcDNA molecule in a conformation α. All possible conformations can be divided into those partitioned among different numbers of inter-post spaces, that is, $\{\alpha\}=\{\alpha_1, \alpha_2, \alpha_3, \cdots\}$, where $\{\alpha_n\}$ is all conformations partitioned among n inter-post spaces. Then, the partition function can be rewritten as

$$Z = \sum_{\{\alpha_1\}} e^{-\beta E_{\alpha_1}} + \sum_{\{\alpha_2\}} e^{-\beta E_{\alpha_2}} + \sum_{\{\alpha_3\}} e^{-\beta E_{\alpha_3}} + \cdots \tag{S2}$$

We assume that the energies of all conformations partitioned among the same number (n) of inter-post spaces are equivalent, that is, $E_{\alpha_n} = \langle E \rangle_n$, where $\langle \cdots \rangle_n$ is the ensemble average calculated for all hcDNA conformations spread over n inter-post spaces. This also means that all hcDNA conformations with the same n are equally probable. Each term in Equation (S2) can be rewritten as

$$\sum_{\{\alpha_n\}} e^{-\beta E_{\alpha_n}} = W_n \cdot e^{-\beta \langle E \rangle_n} \tag{S3}$$

where W_n is the total number of the conformations partitioned among n inter-post spaces.

W_n can be expressed as a product of the number of conformations partitioning over a specific set of connected n inter-post spaces W_n' and the realization of all possible sets of connected n inter-post spaces Ω_n. Then the partition function can be written as follows.

$$Z = W_1 \cdot e^{-\beta\langle E\rangle_1} + W_2 \cdot e^{-\beta\langle E\rangle_2} + \cdots \quad (S4)$$
$$= \Omega_1 \cdot W_1' \cdot e^{-\beta\langle E\rangle_1} + \Omega_1 \cdot W_1' \cdot e^{-\beta\langle E\rangle_2} + \cdots$$

W_n' can be expressed in terms of their entropy, $W_n' = e^{lnW_n'} = e^{S_n/k_B}$, where S_n is the conformational entropy of a hcDNA molecule when partitioned over a specific set of connected *n* inter-post spaces.

$$\mathrm{Z} = \Omega_1 \cdot e^{-\beta(\langle E\rangle_1 - TS_1)} + \Omega_1 \cdot W_1' \cdot e^{-\beta(\langle E\rangle_2 - TS_2)} + \cdots \quad (S5)$$
$$= \Omega_1 \cdot e^{-\beta\langle F\rangle_1} + \Omega_1 \cdot W_1' \cdot e^{-\beta\langle F\rangle_2} + \cdots$$

where $\langle F\rangle_n = \langle E\rangle_n - TS_n$. Therefore, the probability of a hcDNA molecule partitioned among a given combination of *n* inter-post spaces is expressed as follows.

$$P_n = \frac{\Omega_n \cdot e^{-\beta\langle F\rangle_n}}{Z} \quad (S6)$$

This ends the derivation of Equation (1) in the main text.

The values of Ω_n were obtained numerically in our previous work,[5] as below.

n	1	2	3	4
Ω_n/Ω_1	1	4	24	144

Table S1. The number of possible ways Ω_n to select different sets of connected *n* inter-post spaces.

IV. Additional Data

1. Radius of gyration of a hcDNA molecule in nanopost arrays with W_p = 24 nm.

When localized to a single inter-post space with a large $P(1)$, a hcDNA molecule was elongated along the nanoposts. On the other hand, when distributed over several inter-post spaces with a very small $P(1)$, a hcDNA molecule was overall spread perpendicular to the nanoposts. These cases were shown quantitatively by calculating the component of the radius of gyration of a hcDNA molecule along the nanoposts. Since the nanoposts were aligned in the z-direction, we calculated $\langle R_{g,z}^2 \rangle$ of a hcDNA molecule in nanopost arrays with W_p = 24 nm. In addition, the mean squared radius of gyration $\langle R_g^2 \rangle$ as well as the ratio of $\frac{\langle R_{g,z}^2 \rangle}{\langle R_g^2 \rangle}$ were also calculated. These results are presented in Figure S4 (a).

The results in Figure S4 (a) revealed that the value of $\langle R_{g,z}^2 \rangle$ increased from a nanopost array with D_p = 40 nm through D_p = 56 nm, then decreased for D_p = 64 nm, and slightly recovered with D_p = 72 nm. The value of $\langle R_{g,z}^2 \rangle$ is obtained by $\langle R_{g,z}^2 \rangle = P(1)R_{g,z}^2(1) + P(2)R_{g,z}^2(2) + P(3)R_{g,z}^2(3) + \cdots = \sum_n P(n)R_{g,z}^2(n)$, where n is the number of occupied inter-post spaces. First, changing from D_p = 40 nm through D_p = 56 nm, the value of $P(1)$ increases while $P(2)$ and $P(3)$ decrease as shown in Figure 3 of the main text. Because $R_{g,z}^2(1)$ is larger than $R_{g,z}^2(2)$ or $R_{g,z}^2(3)$ (shown in Figure S4 (b)) and its contribution becomes more important, $\langle R_{g,z}^2 \rangle$ increases with the change from D_p = 40 nm through D_p = 56 nm. However, changing from D_p = 56 nm through 72 nm, $P(1)$ is already larger than $P(2)$ or $P(3)$ and thus, $\langle R_{g,z}^2 \rangle$ is determined more importantly by the value of $R_{g,z}^2(1)$. Because $R_{g,z}^2(1)$ decreases from D_p = 56 nm through 72 nm (due to wider inter-post spaces), $\langle R_{g,z}^2 \rangle$ does not increase any more. It decreases slightly from D_p = 56 nm to 64 nm and increases again from D_p = 64 nm to 72 nm depending on the balance between the values of $P(1)R_{g,z}^2(1)$ and $P(2)R_{g,z}^2(2)$.

For W_p = 24 nm

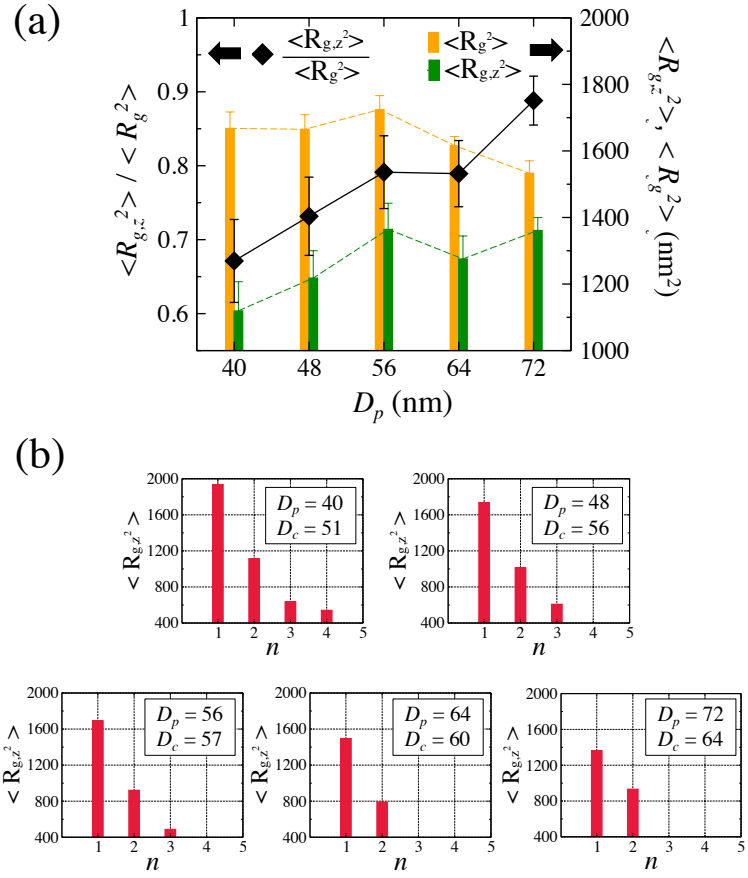

Figure S4. (a) The elongation of a hcDNA molecule along nanoposts in various nanopost arrays with W_p = 24 nm. The scale on the left axis is for $\frac{\langle R_{g,z}^2 \rangle}{\langle R_g^2 \rangle}$ with data in black diamonds, and the scale on the right axis is for $\langle R_{g,z}^2 \rangle$ and $\langle R_g^2 \rangle$ with green and yellow bars, respectively. The ratio of $\frac{\langle R_{g,z}^2 \rangle}{\langle R_g^2 \rangle}$ is presented alone in Figure 4 of the main text. (b) Values of $\langle R_{g,z}^2 \rangle$ for each conformation distributed in n inter-post spaces in different nanopost arrays with W_p = 24 nm.

In Figure S4 (a), the mean squared radius of gyration $\langle R_g^2 \rangle$ was maintained for nanopost arrays with D_p = 40 nm through D_p = 56 nm. The value of $\langle R_g^2 \rangle$ is obtained by $\langle R_g^2 \rangle = P(1)R_g^2(1) + P(2)R_g^2(2) + P(3)R_g^2(3) + \cdots = \sum_n P(n)R_g^2(n)$, where $R_g^2(n) = R_{g,x}^2(n) + R_{g,y}^2(n) + R_{g,z}^2(n)$. In a nanopost arrays with D_p = 40 nm and 48 nm, a hcDNA molecule was spread most probably over two inter-post spaces (n = 2) and $\langle R_g^2 \rangle$ is determined most significantly from $R_g^2(2)$. In a nanopost array with D_p = 56 nm, a hcDNA molecule was more localized compared to those with D_p = 40 nm and 48 nm and therefore the value of $R_g^2(1)$

contribute more than $R_g^2(2)$ to determine the value of $\langle R_g^2 \rangle$. Although $R_{g,x}^2(1)$ and $R_{g,y}^2(1)$ are less than $R_{g,x}^2(2)$ and $R_{g,y}^2(2)$, the larger value of $R_{g,z}^2(1)$ than $R_{g,z}^2(2)$ compensate to maintain the value of $\langle R_g^2 \rangle$. However, in nanopost arrays with D_p = 64 nm and 72 nm, a hcDNA molecule was highly probably localized in a single inter-post space (n = 1), and since the value of $R_{g,z}^2(1)$ was less than in a nanopost array with D_p = 56 nm the values of $\langle R_g^2 \rangle$ are less than in a nanopost array with D_p = 56 nm.

Combining $\langle R_g^2 \rangle$ and $\langle R_{g,z}^2 \rangle$, the ratio of $\frac{\langle R_{g,z}^2 \rangle}{\langle R_g^2 \rangle}$ increased overall from D_p = 40 nm through D_p = 72 nm, which indicates that a hcDNA molecule was elongated along the nanoposts when localized in a single inter-post space. The data of $\frac{\langle R_{g,z}^2 \rangle}{\langle R_g^2 \rangle}$ are presented alone in Figure 4 of the main text.

2. Values of D_c and D_c/W_p Corresponding to the Data Presented in Figure 3 of the Main Text

In Figure 3 of the main text, the probability $P(1)$ that a hcDNA molecule is localized in a single inter-post space was given in terms of D_p and W_p. Since the competitive effects of confinement in the inter-post spaces and in the passages are revealed more clearly by the parameters D_c and W_p or D_c/W_p, we specified the values of D_c and D_c/W_p in Figure S5 for each probability given in Figure 3 of the main text.

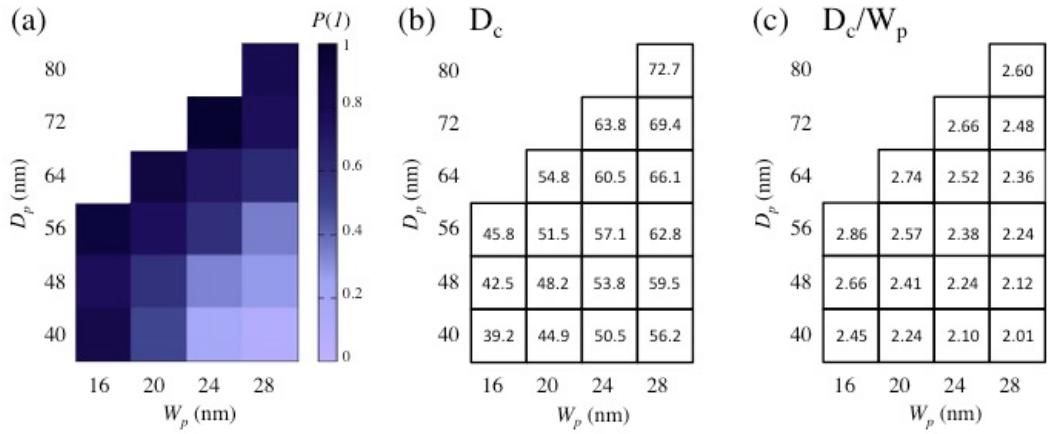

Figure S5. (a) The same data as given in Figure 3 (a) of the main text. (b) and (c) Values of D_c and D_c/W_p corresponding to each data point in (a).

V. References

(1) Kremer, K.; Grest, G. S. Dynamics of Entangled Linear Polymer Melts: A Molecular-Dynamics Simulation. *J. Chem. Phys.* **1990**, 92, 5057-5086

(2) Rubinstein, M.; Colby, R. H. *Polymer Physics*, Oxford University Press, New York, **2003**

(3) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed., Computational Science Series, Academic Press, San Diego, **2002**

(4) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435-447

(5) Joo, H.; Kim, J. S. Confinement and Partitioning of a Single Polymer Chain in a Dense Array of Nanoposts. *Soft Matter* **2015**, *11*, 8262