Modeling Spatial Correlation of DNA Deformation: DNA Allostery in Protein Binding

Citation: Xu, Xinliang, Hao Ge, Chan Gu, Yi Qin Gao, Siyuan S. Wang, Beng Joo Reginald Thio, James T. Hynes, X. Sunney Xie, and Jianshu Cao. "Modeling Spatial Correlation of DNA Deformation: DNA Allostery in Protein Binding." The Journal of Physical Chemistry B 117, no. 42 (October 24, 2013): 13378–13387.

As Published: http://dx.doi.org/10.1021/jp4047243

Publisher: American Chemical Society

Persistent URL: http://hdl.handle.net/1721.1/88163

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Modeling Spatial Correlation of DNA Deformation: DNA Allostery in Protein Binding

Xinliang Xu[1, 2, §], Hao Ge[3, 4, §], Chan Gu[3, 5], Yi Qin Gao[3, 5], Siyuan S. Wang[6], Beng Joo Reginald Thio[2], James T. Hynes[7, 8, *], X. Sunney Xie[3, 6, *], Jianshu Cao[1, 9, *]

[1]Department of Chemistry, MIT, Cambridge, MA 02139, USA

[2]Pillar of Engineering Product Development, Singapore University of Technology and Design, Singapore, 138682

[3]Biodynamic Optical Imaging Center (BIOPIC), Peking University, Beijing 100871, China

[4]Beijing International Center for Mathematical Research (BICMR), Peking University, Beijing 100871, China

[5]Institute of Theoretical and Computational Chemistry, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

[6]Department of Chemistry & Chemical Biology, Harvard University, Cambridge, MA 02138, USA

[7]Department of Chemistry & Biochemistry, University of Colorado, Boulder, CO 80309, USA

[8]Department of Chemistry, UMR ENS-CNRS-UPMC-8640, Ecole Normale Superieure, 75005 Paris, France

[9]Singapore-MIT Alliance for Research and Technology (SMART), Singapore, 138602

[§]These authors contributed equally to this work

Corresponding Authors

[*]Email: James.Hynes@colorado.edu. Phone: +1 303 492 6926 (J.T.H.).

[*]Email: xie@chemistry.harvard.edu. Phone: +1 617 496 9925 (X.S.X.).

[*]Email: jianshu@mit.edu. Phone: +1 617 253 1563 (J.C.).

Abstract

We report a study of DNA deformations using a coarse-grained mechanical model and quantitatively interpret the allosteric effects in protein-DNA binding affinity. A recent single molecule study (Kim et al. (2013) *Science*, **339**, 816) showed that when a DNA molecule is deformed by specific binding of a protein, the binding affinity of a second protein separated from the first protein is altered. Experimental observations together with molecular dynamics simulations suggested that the origin of the DNA allostery is related to the observed deformation of DNA's structure, in particular the major groove width. In order to unveil and quantify the underlying mechanism for the observed major groove deformation behavior related to the DNA allostery, here we provide a simple but effective analytical model where DNA deformations upon protein binding are analyzed and spatial correlations of local deformations along the DNA are examined. The deformation of the DNA base orientations, which directly affect the major groove width, is found in both an analytical derivation and coarse-grained Monte Carlo simulations. This deformation oscillates with a period of 10 base pairs with an amplitude decaying exponentially from the binding site with a decay length $l_D \sim 10$ base pairs, as a result of the balance between two competing terms in DNA base stacking energy. This length scale is in agreement with that reported from the single molecule experiment. Our model can be reduced to the worm-like chain form at length scales larger than l_P but is able to explain DNA's mechanical properties on shorter length scales, in particular the DNA allostery of protein-DNA interactions.

Keywords

Protein-DNA interactions, mechanical deformation, network model, base orientations.

I. Introduction

Protein-DNA interactions play a vital role in many important biological functions, such as chromosomal DNA packaging[1, 2], repair of damaged DNA sites[3, 4], target location[5, 6] and unwinding of DNA[7]. Many studies have explored the local deviations from the canonical helical structure of DNA[8] as the consequence of protein-DNA binding interactions[9, 10]. Nonetheless, understanding of protein-DNA interactions at the microscopic level is still incomplete, in part because the relevant interactions span a wide range of length scales. In particular, previous theoretical descriptions of DNA typically work well on either very small length scales with atomic resolution or very large length scales, at least comparable to the persistent length. This leaves an important lacuna for intermediate length scales. In this connection, our understanding of protein-DNA interactions has recently been advanced by single molecule measurements by Kim et al.[11] of the binding affinities of specific binding of protein to DNA under the influence of the binding of another protein to the same DNA at a distance of intermediate length scales, which presents the challenge to create a theoretical model to bridge the mesoscopic thermodynamic or mechanical properties observed and the underlying molecular mechanism. In the following, we expand on these issues.

At one end of the length scale spectrum, with local details incorporated at the atomic level, molecular dynamic (MD) simulations based on force fields such as CHARMM[12], and AMBER[13] have been proven very successful in studying many different phenomena of DNA including DNA allostery[11], especially with the aid of other numerical techniques such as umbrella sampling[14] and replica exchange[15]. However, the complexity of the DNA molecule with its atomic level details together with the lack of a sufficiently realistic continuous field model in describing the solvent makes these simulations computationally expensive. These

studies are in general limited by their computational requirements to length scales of the order of 10 base pairs (bps) and time scales of the order of microseconds.

At the other end of the length scale spectrum, a widely used theoretical model—the worm-like chain (WLC) model[16], proposes to treat DNA as a semi-flexible polymer chain that behaves like an elastic rod[17]. In this continuous description of DNA, all the local details of the DNA molecule are coarse-grained into a quadratic bending potential that can be characterized by one single parameter, the bending persistence length l_P. By fitting to experimental results that measure extensions of DNA molecules subject to external forces, the model shows a very good agreement between theory and experiment with $l_P \sim 150bps \sim 50nm$ for double-strand DNA under physiological conditions[18] as well as in a flow field[19]. Detailed variations of this model have been proposed over the years by introducing a small number of additional independent parameters[20], such as the twisting persistence length l_t. Since they have only a few parameters, models of this type prove to be very efficient and accurate in treating long DNA molecules at length scales larger than $10^3 bps$. But the coarse graining of all local details also deprives these models of any ability to describe DNA at molecular length scales smaller than the persistence lengths.

For a number of problems of biological significance, the length scale of interest falls in the gap between the atomistic description and the continuous description. These problems call for the creation of a model at the intermediate level, which incorporates the correct amount of local details while at the same time provides the computational efficiency for relatively long chains of DNA. An excellent example is a recent experimental single molecule study by Kim et al.[11], which has motivated the present study. In this experiment, a single DNA molecule of medium size (contour length $100bps \sim 200bps$) is deformed by specific binding of a protein, and

the rate constant of the dissociation of a second protein from the same DNA chain was measured as a function of the separation L between the two binding sites. The experimental results were analyzed with the assumption that the measured dissociation rate constant k is related to the free-energy difference between the binding of the protein and DNA through $\Delta F = -k_B T ln(K_D)$, where the dissociation constant K_D is the dissociation rate k divided by the bimolecular association constant. With this assumption, the experimental results showed that the binding free-energy difference of the second protein oscillates with a period of $10bps$ (the helical pitch of the double helical structure of B form DNA) while the envelop envelope of the amplitude decays very quickly and becomes virtually zero at separations larger than $40bps$. Additional experiments were conducted with the DNA deformation caused by attachment to a hairpin loop instead of the specific binding of the first protein. A similar oscillation of the dissociation rate was observed, indicating that this observed free-energy landscape is related to the underlying correlations between deformed structures along the DNA chain under study rather than to direct protein-protein interactions. The observed allostery was interpreted in terms of the modulation of the major groove width of the DNA induced by the binding of a protein[11]. But, given the observed length scales involved, a quantitative description of the observed correlation requires a mesoscopic model with base pair resolution that applies to a DNA chain of contour length on the order of $100\ bps$.

Following several pioneering works[21-23] in the development of models of intermediate length scale, here we propose a mechanical model of DNA to interpret the observed allosteric phenomenon. As one component of this model, the stacking potential between neighboring bases is modeled by a variant of the Gay-Berne potential[24, 25] between ellipsoids, while the sugar-phosphate backbone as well as the hydrogen bonding between bases within a base pair is

modeled as springs. We find that interhelical distance changes caused by either protein binding or the attached hairpin loop **(**as used in the experimental study[11]**)** induce deformation in the DNA base orientations. Analysis of our model shows that the deformation of the major groove width, which is related to DNA base orientation, exhibits an oscillatory change with an exponentially decaying amplitude. The length scale for the decay is derived analytically and confirmed by our coarse-grained Monte Carlo simulation. These results are in good agreement with the experimental observations of Ref. 11.

The outline of the remainder of this contribution is as follows. In Sec. 2, the description of the model is given and an analytic theory is developed, which produces the key decay and oscillation lengths results (some portions of the analysis are given in an Appendix). The Monte Carlo simulation procedures are described in Sec. 3. Our analytical theory results are successfully compared with both experiment and the Monte Carlo simulations in Sec. 4. Section 5 offers concluding remarks and discussion, including some directions for future efforts.

II. Model description

Here we present and analytically develop a mechanical model to study DNA deformations at zero temperature. We show in Sec. 5 that the mechanism underlying the behavior of the major groove deformations is an intrinsic feature of the DNA system and that our study is applicable to the DNA deformations at room temperature. In this coarse-grained representation of a DNA molecule which incorporates an intrinsic twist at every base pair step, the double helical structure of an ideal B-type DNA helps us define a right-handed coordinate system with the *z* axis in the longitudinal direction (Figure 1). As illustrated in Figure 2, in our model each phosphate-sugar-base unit of DNA is modeled by a sphere representing the

phosphate-sugar group attached to a thin plate (representing the base) with thickness *c*, depth of the short side *b* and length of the long side a. These units are connected into two strands, color-coded as blue and red. The two strands are connected together—forming a double helical structure, by springs representing the hydrogen bonds between each base pair. The orientation for each DNA unit is defined by the unit vector $\hat{u}$ normal to the corresponding thin plate and by definition $\hat{u} = \hat{z}$ for all units of an ideal B-type double helical structure (Figure 3A). According to previous studies[23], the stacking interactions between neighboring bases within each strand with orientation $\hat{u}_n^s$ and $\hat{u}_{n+1}^s$, where $s = blue$ for the blue strand and $s = red$ for the red strand can be well modeled by a variant of the Gay-Berne potential as a product of three terms:

$$U(\hat{u}_n^s, \hat{u}_{n+1}^s, r) = U_r * \eta(\hat{u}_n^s, \hat{u}_{n+1}^s) * \chi(\hat{u}_n^s, \hat{u}_{n+1}^s). \quad (1)$$

The first term, in a form of a simple Lennard-Jones potential, controls the distance dependence of the interaction; while the last two terms relate the interaction to the orientation $\hat{u}_n^s$ and the relative orientation $\hat{u}_{n+1}^s \cdot \hat{u}_n^s$.

As suggested by the experimental studies of Ref. 11, here we assume that one base pair with index $n = 0$ is pulled apart along its long side. This deformation causes an interhelical distance change that involves backbone chemical bonds, stacking interactions and hydrogen bonds. Since the stiffness of the backbone bonds as well as the distance dependent part of the stacking interactions (U_r in eq. 1) is much higher than for other kinds of energies, these two kinds of bonds can be regarded as almost rigid. This approximation exerts a strong geometric constraint such that the distorted interhelical distance at the base pair $n = 0$ will relax along the DNA chain back to equilibrium length in a few base pair steps, by the induction of an alteration of orientations for neighboring bases, from $\hat{u} = \hat{z}$ at equilibrium to an altered orientation

$\hat{u}(\theta,\varphi) = \sin\theta * \cos\varphi\,\hat{x} + \sin\theta * \sin\varphi\,\hat{y} + \cos\theta\,\hat{z}$ (Figure 3A and 3B). The induced alteration of orientations itself relaxes slowly back to $\hat{u} = \hat{z}$ along the DNA chain. Due to the symmetry of the system, the orientations of the two bases in a base pair $\hat{u}(\theta_{red},\varphi_{red})$ and $\hat{u}(\theta_{blue},\varphi_{blue})$ satisfy the conditions $\theta_{red} = \theta_{blue}$ and $\varphi_{red} = \pi + \varphi_{blue}$. Depending on the alignment between the alteration of orientation and the long side of the base plate, such induced alteration of orientation can be manifest as a combination of a buckling deformation and a propeller twist deformation (Figure 3C). Since the stacking energy prefers adjacent bases on the same strand to have the same orientations, the induced alteration of orientations decays very slowly, as noted above. For illustration purposes we show in Figure 4 a case where it is a constant within one helical pitch of DNA. This Figure shows that as a result of the intrinsic twist, the relative alignment between the alteration of orientation and the long side of the base plate changes periodically, yielding periodic structure changes from buckling backward to propeller twist outward to buckling forward to propeller twist inward within each helical pitch.

In order to quantitatively describe the deformation relaxation along the DNA chain, we propose here a simplified two-dimensional model that yields analytical results. In this simplified model illustrated in Figure 5, centers of identical solid rectangles (side length $a \gg c$)each representing one DNA base are connected into two strands (color coded as blue and red) extending to infinity on both sides. By means of the pairing of each rectangle on one strand to its corresponding rectangle on the other strand with springs of stiffness k_{HY} and equilibrium length $2l_{HY}$, the two parallel strands are connected together and form a two dimensional network. Here we denote the direction parallel to each strand as the *z* axis and the direction perpendicular as the *x* axis, with the two strands at $x_{blue} = l_{HY}$ and $x_{red} = -l_{HY}$ respectively. The orientation of each rectangle can be characterized by the angle θ between its main axis perpendicular to side *a*

and the *z* axis. For an ideal B-type DNA molecule $\theta = 0$ for all bases. In order to study the relaxation of an interhelical distance deformation, one pair of rectangles (denoted as the 0th pair in sequence) are pulled slightly apart in the *x* direction as their centers are now located at $x_{blue\_0} = l_{HY} + d_0$ and $x_{red\_0} = -l_{HY} - d_0$, respectively. As a result of this deformation, all rectangles relocate (to $x_{blue\_n} = l_{HY} + d_n$ and $x_{red\_n} = -l_{HY} - d_n$) and reorient ($\theta_n$ for the n^{th} base in the blue strand and $-\theta_n$ for the n^{th} base in the red strand) so that on each rectangle such that the force balance and the torque balance are restored. If we assume that all rectangles in one strand (e.g., the blue strand) are properly relocated so that the distance-dependent contribution U_r in eq. 1 stays fixed, we can simplify the interaction defined in that equation as:

$$U(\theta_n, \theta_{n+1}) = const * (1 - \omega_1 \theta_n^2) * (1 - \omega_1 \theta_{n+1}^2) * (1 - \omega_2 \Delta\theta_n^2), \quad (2)$$

where θ_n is the orientation of the n^{th} base in the blue strand, $\Delta\theta_n = \theta_{n+1} - \theta_n$ and the coefficients ω_1 and ω_2 can be obtained from eq. 1. Due to the symmetry of the system, the orientation of the n^{th} base in the other strand (in this case the red strand) is $-\theta_n$. Now for the n^{th} rectangle away from the deformed boundary, the torque balance requires that

$$\omega_2 * (\theta_{n+1} - \theta_n) - \omega_1 * \theta_n - \omega_2 * (\theta_n - \theta_{n-1}) + \tau_n = 0, \quad (3)$$

where τ_n is the torque on the base exerted by the hydrogen bonds within the n^{th} base pair.

Solution of eq. 3 is not straightforward since the torque τ_n is coupled with the orientation deformation θ_n. For a simpler problem of interest, in which we have torque $\tau'_n = \tau * \delta_{i,n}$, where τ is a constant and $\delta_{i,n}$ is the Kronecker delta function (a constant torque at the i^{th} base and 0 torque at any other bases), eq. 3 can be reduced to a simpler form for $j > 0$

$$\theta_{i+j-1} - \left(2 + \frac{\omega_1}{\omega_2}\right) * \theta_{i+j} + \theta_{i+j+1} = 0. \quad (4)$$

Equation 4 should hold for all $j > 0$, which means that the ratio $\theta_{i+j}/\theta_{i+j-1} = \alpha$ is independent of j and is parameterized by ω_1 and ω_2 through the quadratic equation $1 - \left(2 + \frac{\omega_1}{\omega_2}\right) * \alpha + \alpha^2 = 0$. There are two solutions to this equation satisfying $\alpha_1 * \alpha_2 = 1$, corresponding to one decaying mode $|\alpha_1| < 1$ and one growing mode $|\alpha_2| > 1$. It is implied in this derivation that the deformation is induced by the external torque at the i^{th} base and decays towards the boundary at infinity where $\theta_\infty = 0$, so that the constant ratio $\theta_{i+j}/\theta_{i+j-1} = \alpha$ is uniquely determined as α_1. The amplitude of the deformation characterized by θ_n is then determined to decay exponentially along the chain as $\theta_{i+j} = \theta_i * \alpha_1^j = \theta_i * e^{-j/l_D} \sim \tau * e^{-j/l_D}$, where the deformation correlation length scale $l_D = -(ln\alpha_1)^{-1}$. In the limiting case where $\omega_1 \ll \omega_2$, this can be reduced to a simple form $l_D = \sqrt{\omega_2/\omega_1}$.

An analytical approximation to the complete solution to the full eq. 3 as opposed to the simplified eq. 4 can be found in the Appendix. To summarize the result, for the n^{th} base away from the deformed boundary we find

$$\theta_n \sim \left(1 - e^{-n/l_1}\right) * e^{-n/l_D}, \tag{5}$$

where l_1 shows the relaxation length scale of interhelical distance changes and is estimated to be on the order of one base pair step.

The last two terms in eq. 1 have been studied previously[23], providing some information on the ratio ω_2/ω_1. An evaluation of these two terms following this early formulation shows that $\eta(\theta_n = \varepsilon, \theta_{n+1} = \varepsilon + \Delta) \sim 1 - \varepsilon^2 - \frac{X^2}{2c^2}\Delta^2$ and $\chi(\theta_n = \varepsilon, \theta_{n+1} = \varepsilon + \Delta) \sim 1 - 2\varepsilon^2 - \frac{X^2}{c^2}\Delta^2$ for small ε and Δ, where $X = a$ for orientation changes parallel to the long side of the plate and

$X = b$ for orientation changes parallel to the short side of the plate. Comparing this result to eq. 2 we see that $\omega_2/\omega_1 = X^2/c^2$.

Our modeling of the DNA base as a rectangular thin plate with long side length *a*, short side length *b* and thickness *c* is of course a phenomenological approximation and the appropriate values for these parameters must yield the minimum center-to-center distance for perfect stacking. Previous study[23] shows that one good choice is that $a = 9\dot{A}$, $b = 4\dot{A}$ and $c = 0.7\dot{A}$. From this we obtain an expectation of the ratio $\frac{\omega_2}{\omega_1} = \frac{X^2}{c^2} \sim (a^2 + b^2)/2c^2 \sim 100 \gg 1$. This supports the simple approximation for l_D obtained at the end of the discussion of the solution of eq. 4 and gives a decay length scale $l_D = \sqrt{\omega_2/\omega_1} \sim 10\ (bps)$.

In our development above, we have dealt with the simplified two-dimensional case. In a more realistic three-dimensional DNA model the unit vector representing the orientation is characterized by both θ and φ, where θ characterizes the overall amplitude of the change of orientation from equilibrium where $\hat{u} = \hat{z}$ and φ characterizes the relative direction of the change of orientation. As illustrated by our own Monte Carlo simulation results shown later in Sec. 4, the change in φ at each base pair step is small and as an approximation we can assume that in the real DNA system the change in φ is negligible. Under this approximation our results on $\{\theta_n\}$ for the simplified two-dimensional model can be extended to the orientations of bases $\{\hat{u}_n(\theta_n', \varphi_n')\}$ in a realistic three-dimensional DNA model which incorporates the intrinsic twist, in a fashion that $\theta_n' = \theta_n$ and $\varphi_n' = const.$ If we assume that the backbone phosphate group relocates according to the edge of the base plate in the longitudinal direction by attachment, we have the major groove width of the DNA molecule defined as the distance between the phosphate group in the n^{th} blue unit and the phosphate group in the $(n+7)^{\text{th}}$ red unit

$$W_n = \left|\vec{P}_{n\_Blue} - \vec{P}_{(n+7)\_Red}\right| = 6h + a\left(\theta_n \cos\frac{n\pi}{5} - \theta_{n+7}\cos\frac{(n+7)\pi}{5}\right) + O(\theta_{ind}^2), \quad (6)$$

where $h = 3.4\dot{A}$ is the base step of an ideal B-type DNA, and θ_{ind} is the overall induced amplitude defined through $\theta_n = \theta_{ind} * \left(1 - e^{-n/l_1}\right) * e^{-n/l_D}$ (see eq. 5) which is assumed to be small so that all higher order terms can be neglected.

III. Monte Carlo simulation

To test if the analytical approach of Sec.3 is reasonable, we carried out a simple coarse-grained Monte Carlo simulation on a DNA molecule with $N = 100$ base pairs. We simplified the system by keeping only base stacking, hydrogen bonding between bases within each base pair and backbone bonding interactions. The base stacking interaction has been limited to the interaction between neighboring bases within the same strand**;** it is decoupled into a distance-dependent part and an orientation-dependent part as $U(\hat{u}_n, \hat{u}_{n+1}, r) = U_r * U_\Theta(\hat{u}_n, \hat{u}_{n+1})$, where the distance *r* between two neighboring bases is obtained from $r \equiv \min\left(\left|\vec{r}_i - \vec{r}_j\right|\right), \forall(i,j)$ with $i \in plate\ n$ and $j \in plate\ n+1$. All the distance-dependent interactions included in our simulation are modeled as elastic springs around their corresponding equilibrium distances. That is, we use an elastic spring of stiffness K_r for the distance-dependent part U_r, an elastic spring with stiffness k_{HY} for hydrogen bonding, and an elastic spring with stiffness K_P for backbone bonding (see Table I for the parameters used in the simulation).The orientation-dependent part of the stacking is modeled as $U_\Theta\left(\hat{u}_n(\theta_n, \varphi_n), \hat{u}_{n+1}(\theta_{n+1}, \varphi_{n+1})\right) = U_0 * [1 - \omega_1\theta_n^2 - \omega_1\theta_{n+1}^2 - 2\omega_2(1 - \hat{u}_n \cdot \hat{u}_{n+1})]$ with amplitude U_0, which reduces to the two dimensional case eq. (2) the two dimensional case when $\varphi_n = \varphi_{n+1}$.

To start each simulation run, all the bases are placed at the corresponding positions of an ideal B-type DNA except for one base pair which is pulled apart in the long side direction by $1\dot{A}$. The orientation of each base $\hat{u}_n(\theta_n, \varphi_n)$ is initiated with $\theta_n$being a random number between 0 to 0.001 and φ_n being a random number between 0 to 2π, except for the one base pair which is pulled apart where the orientations of the two are kept fixed at $\theta = 0$ and $\varphi = 0$ throughout the simulation run. As described in previous studies[23], each base taken as a thin plate has six degrees of freedom. Three of them are translational—Rise, Shift, Slide, and the other three are rotational—Tilt, Twist, Roll. Due to the symmetry of the system in our problem, to study the deformation relaxation of our interest we assume that only one base in a base pair is free to move and that the other will move symmetrically. In each trial move of our simulation, we fixed the Twist degree of freedom and made random displacements in the other five degrees of freedom for each base pair. The moves are accepted or rejected according to the Metropolis scheme[26]. Since we are only interested in the deformation relaxation of DNA as a result of its mechanical properties, we have chosen to downplay the role of thermal excitations and conducted the simulation with the very low temperature $T' = 10^{-4}T$, where T denotes room temperature =293K.

IV. Results

In this section, we compare our analytic predictions with both experiment and our Monte Carlo simulations.

Our analytical predictions of the base orientation change are compared with the results obtained in the simulations in Figure 6. For the parameter θ, the amplitude of the change in orientation, our analytical prediction (eq. 5) agrees very well with the results obtained in our

Monte Carlo simulations. For the base orientation parameter φ, results from the simulations show that the changes at each base step are fairly small (on the order of $0.1\ radian \sim 5^{\circ}$) as compared to the intrinsic twist which is $\pi/5 = 36^{\circ}$ at each base step. This slow variance in φ supports the approximation used in our analytical analysis in Sec. 3, where φ is treated as a constant. This can be understood as a result that the change in φ raises a large amount of energy but does not explicitly help the relaxation of the deformation.

Most proteins primarily interact with the DNA major grooves. Therefore distortion of the major groove would have the largest influence on protein binding affinity. Our theoretical results are compared with recent experimental results of Ref. 11, which demonstrated the correlation and anticorrelation between bindings of two proteins on two specific sites of DNA with a separation of L. Figure 7 shows our results from simulations for the positions of the phosphate groups. The major groove width of the DNA can be obtained either from these locations or analytically from eq. 6. In figure 8 our theoretical results concerning the major groove width are shown in comparison with the experimentally observed 2nd protein binding free-energy $\Delta G(L)$ as a function of separation L in the form of $\Delta\Delta G(L) = \Delta G(L) - \Delta G(\infty)$. The comparison shows a quite good agreement between the experiment and theory for $L > 5bps$; the quantitative discrepancy at small separation regime for $L < 5bps$ is still poorly understood and requires more detailed studies.

V. Conclusion and Discussion

Our coarse-grained mechanical model proves to be generally useful for studying DNA deformation at an intermediate length scale and leads to theoretical predictions that are in good agreement with recent experimental results[11] and Monte Carlo simulations. The new decay

lengthscale l_D, first demonstrated in the recent single molecule experiment in Ref. 11, is proposed here as a result of the balance between two competing terms in DNA base stacking energy. Since this competition is a generic feature of the DNA system, it is of considerable interest to see whether the same general exponential decaying behavior is at work for deformations other than interhelical distance changes, such as bending, supercoiling deformation.

The results demonstrated within have been obtained from DNA either at zero temperature (analytical analysis) or at very low temperature (Monte Carlo simulations). Here we argue that these results also apply at room termperature, and so are relevant for the experiments of Ref. 11. At room temperature the DNA molecule undergoes thermal excitations resulting from its interactions with the surrounding solvent (typically water) molecules. The time scale over which these interactions occur is denoted as T_1, typically comparatively small ($1ps \sim 1ns$). Over this time scale, the thermal excitations can be considered as an instantaneous thermal "kick"—an external force (or torque) at each base pair. On the other hand, typical experimental observations happen at time scale T_0 around $1ms \sim 1sec$, at which the DNA has undergone many thermal "kicks". Since these interactions are uncorrelated in nature, the effects observed in experiments are the statistical averages of many instantaneous thermal "kicks" over T_0. In a simple approach, here we model each of these uncorrelated thermal "kicks" as an external force (or torque) at each base pair site, of amplitude f_0 pointing in a random direction, where the statistical time average of these "kicks" over a time scale of T_0 has a square amplitude proportional to the thermal energy, $\langle f_0 * f_0 \rangle_{T_0} \sim ckT$, where c is the suitable proportionality factor. In order to study the thermally driven deformation of DNA, it involves no loss of generality to keep the DNA chain at zero temperature except for one base pair with index $n = 0$, since the molecule is treated as a linear system in our mechanical model. The forces of thermal origin mentioned above are not

fundamentally different in terms of deforming DNA from other external forces treated in our current study.

Therefore, in the simplest case we can consider only one mode of the thermal "kick" which acts as an external torque of amplitude f_0 pointing in a random direction in the *xy* plane. In the spirit of our earlier analytical analysis in Sec. 2, at any instant t the DNA molecule can be described by its two-dimensional projection with normal direction of the two dimensional plane (characterized by $\varphi(t)$) determined by the external torque $\vec{\tau}(t)$ and the z axis. According to our simplified two-dimensional model, such an external torque induces a change of orientations of bases $\{\hat{u}_n(\theta_n''(t))\}$. We have already shown that the behavior of $\{\theta_n''(t)\}$ is governed by eq. 4, which yields a result of $\theta_n''(t) = \theta_0''(t) * e^{-n/l_D}$ with amplitude $\theta_0''(t) \sim f_0$. Since the thermal "kicks" are totally uncorrelated, $\varphi(t)$ is random. On the time scale T_0, the statistical averages show that the deformation in base orientation $\overrightarrow{\delta u_n}(t) = \hat{u}_n(t) - \hat{z} = \sin\theta_n''(t) * \cos\varphi(t)\,\hat{x} + \sin\theta_n''(t) * \sin\varphi(t)\,\hat{y}$ satisfies $\langle\overrightarrow{\delta u_n}(t)\rangle_{T_0} = 0$ as a result of the randomness. However—and this is the key point—the correlation $\langle\overrightarrow{\delta u_n}(t) * \overrightarrow{\delta u_0}(t)\rangle_{T_0} / \langle\overrightarrow{\delta u_0}(t) * \overrightarrow{\delta u_0}(t)\rangle_{T_0} = e^{-n/l_D}$ remains just the same as the result obtained in Sec. 2 for our model developed for the zero temperature system. This important result can be generalized as $\langle\overrightarrow{\delta u_j}(t) * \overrightarrow{\delta u_i}(t)\rangle_{T_0} / \langle\overrightarrow{\delta u_i}(t) * \overrightarrow{\delta u_i}(t)\rangle_{T_0} = e^{-|i-j|/l_D}$ for the more realistic case where all of the DNA base pair sites are thermally excited. As a direct result of this correlation, the major groove widths at different locations exhibit a similar correlation as $\langle W_j(t) * W_i(t)\rangle_{T_0} / \langle W_i(t) * W_i(t)\rangle_{T_0} = e^{-|i-j|/l_D}$. The above analysis indicates that the mechanism unveiled by our model—the correlation between local deformations of DNA structures at different locations—is general and is an intrinsic feature of the DNA system.

Conventional models based on the elastic rod treatment of DNA (e.g. the worm-like chain model) describe the DNA molecule in terms of its centerline and cross sections. These models provide reliable descriptions of the DNA molecule at length scales larger than the persistence length $l_P \sim 150bps$, where the amplitude of the bending angle Θ_i between two consecutive segments (labeled with index i and $i+1$, respectively) of DNA of length L_0 is accurately predicted as $\langle \Theta_i \rangle = e^{-L_0/l_P}$. However, since they lack local details, these continuous models fail to provide a good description at length scales smaller than that persistence length. This failure is caused by the breakdown of one key assumption that the cross sections (as a point in the worm-like chain[16] and as a circle in other models[27]) are rigid and are "stacked" along the centerline, which requires that all bending angles are independent as $\langle \Theta_i \Theta_j \rangle = \delta_{i,j}$. Our results show that local deformations are correlated at short length scale $l_D \sim 10bps$ and the failure of these continuous descriptions at short length scales can be avoided by incorporating modifications that follow naturally from the model presented in this paper. The conclusion $l_D \sim 10bps \ll l_P$ from the present model is consistent with these elastic rod descriptions since the molecular details included in our model can be renormalized into the fitting parameter l_P at length scales larger than l_P. This new description, which incorporates local details into traditional continuous models, is expected to be of considerable importance in studying DNA structures at length scales comparable to the persistence length and should help us understand many mechanical properties of DNA such as the enhanced flexibility at short length scales and DNA repair mechanism inside cells.

Strictly speaking, the analytical results obtained in this study only apply to an infinitely large system consisting of identical units. Extension of the study to finite system with sequence-dependent properties can be made by bundling all the linear torque balance equations on all

bases in an equivalent matrix representation. In this representation, a so-called resistance matrix can be given with neighboring interaction coefficients ω_1 and ω_2 being the matrix elements. The final structure of the system upon deformation can be expressed in terms of the eigenvalues and the eigenvectors of this resistance matrix. When all units are identical the matrix is a Toeplitz matrix, that is, elements are constant along diagonals. For a finite DNA chain of N base pairs, the convergence of the eigenvalues and eigenvectors of the N by N Toeplitz matrix to the $N \to \infty$ analytical limit has been studied[28]. The close agreement between results from our analytical analysis with an infinitely large system by eq. 5 and our simulation studies for $N = 100$ shows consistence with the mathematical study in Ref. 28; the DNA chain length satisfies $N{\sim}100bps \gg l_D$ so that $N{\sim}\infty$ serves as a good approximation.

Of course, in reality these DNA units are in general different. The variations of the DNA molecule at the base pair level, including mismatches[29, 30] (broken hydrogen bonds and poor stacking forces) and sequence-dependent features [31, 32] (hydrogen bond strength and stacking force vary for different sequences), actually have important biological implications and accordingly are of great interest. The rugged free-energy landscape associated with the sequence dependent interactions between DNA and the binding protein has been probed[33] and its important role on many processes of great biological importance, e.g. the sliding kinetics of the binding protein along DNA, has been discussed[34]. Qualitatively, we know that GC stacking interactions are more stable than AT stacking interactions, that is $\left|U_{0\_GC}\right| > \left|U_{0\_AT}\right|$. This leads to a smaller overall amplitude of the induced alteration of orientation for GC-rich DNA segments than for AT-rich segments, in qualitative agreement with experimental observations[11]. However, a highly desired quantitative study is left for the future, although we do note here that for small variations this can be realized by perturbation of the resistance matrix M'around the Toeplitz

matrix M as $(M+\varepsilon X)^{-1} = M^{-1} - \varepsilon M^{-1} X M^{-1} + \mathcal{O}(\varepsilon^2)$. The sequence dependence and other issues will be subjects of further studies.

In conclusion, we have proposed a mechanical model and analytic analysis to explain the recent experimentally observed DNA allostery phenomenon. We attributed the observed DNA allostery to major groove distortions, which result from the deformation of DNA base orientations. Since the DNA base orientation is much more flexible than the backbone or the interhelical distance, the local deformation of the interhelical distance transfers to the distortion of the base orientation very rapidly, which can propagate to a long range at a length scale about $10\ bps$. The major groove length oscillates because of the intrinsic double helix structure of DNA. Local deformations, major groove width in particular as shown in recent experimental study, induced by the first protein bound in turn affects the binding of a second protein and vice versa, which is the underlying mechanism for DNA allostery.

Appendix. Approximate solution to eq. 3.

In order to solve the full eq. 3, we assume that the system is linear. When one base pair is pulled apart, changes of orientations for neighboring base pairs are induced. Along the DNA chain we see that spatially the interhelical distance change deformation transforms into an orientation change deformation. Under the linear system assumption, we assume that the external torque on the n^{th} base $\tau_n \equiv \rho * d_n$. Equation 3 then becomes

$$\omega_2 * (\theta_{n+1} - \theta_n) - \omega_1 * \theta_n - \omega_2 * (\theta_n - \theta_{n-1}) + \rho * d_n = 0. \quad \text{(A1)}$$

Without the external torques, we have seen that the solution to equation

$$\omega_2 * (\theta_{n+1} - \theta_n) - \omega_1 * \theta_n - \omega_2 * (\theta_n - \theta_{n-1}) = 0 \quad \text{(A2)}$$

satisfies $\theta_{n+1} = \theta_1 * e^{-n/l_D}$. As an extension of this result to a system with linear coupling between the interhelical distance change and the orientation change, we assume that there exists a linear combination $q_n = \theta_n + \gamma d_n$ that obeys

$$q_{n+1} = q_1 * e^{-n/l_D}, \quad \text{(A3)}$$

where γ is constant showing the coupling between the two deformations just mentioned.

Equations A1 and A3 can be solved together numerically, with any specified constant γ. Based on the fact that in our case the decaying length scale l_D is about ten times larger than the length scale l_1 over which the interhelical distance change transforms into an orientation change, an analytical solution can be achieved with an additional approximation. This approximation considers that the decaying length scale l_D is much larger than the lengthscale l_1 so that the decaying regime and the transformation regime can be regarded as decoupled. That is, in the transformation regime, the decaying terms can be regarded as negligible so that we have:

$$\begin{cases} \omega_2 * (\theta_{n+1} - \theta_n) - \omega_2 * (\theta_n - \theta_{n-1}) + \rho * d_n = 0 \\ \theta_n + \gamma d_n = \gamma d_0 \end{cases}. \quad \text{(A4)}$$

Equation A4 can be solved analytically with $d_n = d_0 e^{-n/l_1}$ and $\theta_n = \gamma d_0 \left(1 - e^{-n/l_1}\right)$, where $l_1 = -(ln\epsilon)^{-1}$ and ϵ satisfies:

$$\gamma\omega_2 - (\rho + 2\gamma\omega_2) * \epsilon + \gamma\omega_2 * \epsilon^2 = 0. \quad \text{(A5)}$$

Outside the transformation regime we can assume that the external torque is negligible so that $\theta_{n+N} = \theta_N * e^{-n/l_D}$, where $N > l_1$. So overall an analytical approximation of the solution to equation (3) can be written as:

$$\theta_n = \gamma d_0 \left(1 - e^{-n/l_1}\right) e^{-n/l_D}. \quad \text{(A6)}$$

Table I

Parameters used for ideal B-type DNA:

Base step in z direction	Base step intrinsic twist	Radius of the double helix
$h = 3.4\dot{A}$	$\omega_0 = \pi/5$	$R_0 = 9\dot{A}$

Other parameters used in Monte Carlo simulation:

Backbone strength	Base stacking distance part	Hydrogen bond strength	Base stacking orientation part I	Base stacking orientation part II
$K_P = 50k_BT/\dot{A}^2$	$K_r = 50k_BT/\dot{A}^2$	$K_{HY} = 3k_BT/\dot{A}^2$	$U_0\omega_1 = -2k_BT$	$U_0\omega_2 = -200k_BT$

Acknowledgements

X.L.X. would like to thank J. Wu, L. Lai, C. Chern and J. Moix for helpful discussions. X.L.X. and J.C. acknowledge the financial assistance of Singapore-MIT Alliance for Research and Technology (SMART), National Science Foundation (NSF CHE-112825), Department of Defense (DOD ARO W911NF-09-0480), and a research fellowship by Singapore University of Technology and Design (to X.L.X.). H.G. is supported by the Foundation for the Author of National Excellent Doctorial Dissertation of China (No. 201119). The research work of X.S.X. is supported by NIH Director's Pioneer Award. The research work by B.J.R.T. is supported by Singapore University of Technology and Design Start-Up Research Grant (SRG EPD 2012 022). The research work by J.T.H. is supported by research grant NSF CHE-1112564.

Additional information

The authors declare no competing financial interests. Correspondence and requests for numerical results should be addressed to J.T.H., X.S.X. and J.C.

References

1. Richmond, T.J.; Davey, C.A. The Structure of DNA in the Nucleosome Core. *Nature* **2003**, 423, 145—150.
2. Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J.D. *Molecular Biology of the Cell*; Garland Publishing: New York, 1994.
3. Lukas, J.; Bartek, J. DNA Repair: New Tales of an Old Tail. *Nature* **2009**, 458, 581—583.
4. Misteli, T.; Soutoglou, E. The Emerging Role of Nuclear Architecture in DNA Repair and Genome Maintenance. *Nat. Rev. Mol. Cell Biol.* **2009**, 10, 243—254.
5. Zhou, H.X. Rapid Search for Specific Sites on DNA through Conformational Switch of Nonspecifically Bound Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, 108, 8651—8656.
6. Gorman, J.; Greene, E.C. Visualizing One-Dimensional Diffusion of Proteins along DNA. *Nat. Struct. Mol. Biol.* **2008**, 15, 768—774.
7. Boule, J.B.; Vega, L.R.; Zakian, V.A. The Yeast Pif1p Helicase Removes Telomerase from Telomeric DNA. *Nature* **2005**, 438, 57—61.
8. Watson, J.D.; Crick, F.H.C. A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, 171, 737—738.
9. Rohs, R.; Jin, X.S.; West, S.M.; Joshi, R.; Honig, B.; Mann, R.S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, 79, 233—269.
10. Olson, W.K.; Zhurkin, V.B. Modeling DNA Deformations. *Curr. Opin. Struct. Biol.* **2000**, 10, 286—297.
11. Kim, S.; Brostromer, E.; Xing, D.; Jin, J.; Chong, S.; Ge, H.; Wang, S.; Gu, C.; Yang, L.; Gao, Y.; Su, X.; Sun, Y.; Xie, X.S. Probing Allostery Through DNA. *Science* **2013**, 339, 816—819.
12. Mackerell, A.D.; Wiorkiewiczkuczera, J.; Karplus, M. An All-Atom Empirical Energy Function for the Simulation of Nucleic Acids. *J. Am. Chem. Soc.* **1995**, 117, 11946—11975.
13. Cheatham, T.E.; Cieplak, P.; Kollman, P.A. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, 16, 845—862.
14. Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J.T. On the Molecular Mechanism of Drug Intercalation into DNA: A Simulation Study of the Intercalation Pathway, Free Energy, and DNA Structural Changes. *J. Am. Chem. Soc.* **2008**, 130, 9747—9755.
15. Kannan, S.; Zacharias, M. Simulation of DNA Double-Strand Dissociation and Formation during Replica-Exchange Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2009**, 11, 10589—10595.
16. Kratky, O.; Porod, G. Rontgenuntersuchung geloster fadenmolekule. *Recueil des Travaux Chimiques des Pays-Bas* **1949**, 68, 1106—1122.
17. Landau, L.D.; Lifschitz, E.M. *Theory of Elasticity*; Pergamon Press: New York, 1986.

18. Bustamante, C.; Marko, J.F.; Siggia, E.D.; Smith, S. Entropic Elasticity of λ-Phage DNA. *Science* **1994**, 265, 1599—1600.
19. Yang, S.; Witkoskie, J.; Cao, J. First-Principle Path Integral Study of DNA under Hydrodynamic Flows. *Chem. Phys. Lett.* **2003**, 377, 399—405.
20. Moroz, J.D.; Nelson, P. Torsional Directed Walks, Entropic Elasticity, and DNA Twist Stiffness. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, 94, 14418—14422.
21. Tepper, H.L.; Voth, G.A. A Coarse-Grained Model for Double-Helix Molecules in Solution: Spontaneous Helix Formation and Equilibrium Properties. *J. Chem. Phys.* **2005**, 122, 124906.
22. Knotts, T.A., IV; Rathore, N.; Schwartz, D.C.; de Pablo, J.J. A Coarse Grain Model for DNA. *J. Chem. Phys.* **2007**, 126, 084901.
23. Mergell, B.; Ejtehadi, MR.; Everaers, R. Modeling DNA Structure, Elasticity, and Deformations at the Base-Pair Level. *Phys. Rev. E* **2003**, 68, 021911.
24. Everaers, R.; Ejtehadi, M.R. Interaction Potentials for Soft and Hard Ellipsoids. *Phys. Rev. E* **2003**, 67, 041710.
25. Gay, J.G.; Berne, B.J. Modification of the Overlap Potential to Mimic a Linear Site-Site Potential. *J. Chem. Phys.* **1981**, 74, 3316—3319.
26. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.N.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, 21, 1087—1092.
27. Balaeff, A.; Mahadevan, L.; Schulten, K. Modeling DNA Loops Using the Theory of Elasticity. *Phys. Rev. E* **2006**, 73, 031919.
28. Dai, H.; Geary, Z.; Kadanoff, L.P. Asymptotics of Eigenvalues and Eigenvectors of Toeplitz Matrices. *J. stat. Mech.: Theo. and Exp.* **2009**, P05012.
29. Jiricny, J. The Multifaceted Mismatch-Repair System. *Nat. Rev. Mol. Cell Biol.* **2006**, 7, 335—346.
30. Kunkel, T.A.; Erie, D.A. DNA Mismatch Repair. *Annu. Rev. Biochem.*, **2005**, 74, 681—710.
31. Olson, W.K.; Gorin, A.A.; Lu, X.J.; Hock, L.M.; Zhurkin, V.B. DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 11163—11168.
32. Nelson, P. Sequence-Disorder Effects on DNA Entropic Elasticity. *Phys. Rev. Lett.* **1998**, 80, 5810—5812.
33. Maiti, P.K.; Bagchi, B. Structure and Dynamics of DNA-Dendrimer Complexation: Role of Counterions, Water, and Base Pair Sequence. *Nano Lett.* **2006**, 6, 2478—2485.
34. Leith, J.S.; Tafvizi, A.; Huang, F.; Uspal, W.E.; Doyle, P.S.; Fersht, A.R.; Mirny, L.A.; van Oijen, A.M. Sequence-Dependent Sliding Kinetics of p53. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 109, 16552—16557.

Figure captions

Figure 1. Coordinate system. The coordinate system used is defined as illustrated: the longitudinal direction of the double helical structure is defined as *z*. In the plane perpendicular to *z*, an arbitrary direction is selected as *x*. Then *y* is defined through the right hand rule.

Figure 2. Our coarse-grained model of DNA.DNA is modeled as two strands (color-coded red and blue) of identical units. Each unit of DNA is modeled as a sphere representing the sugar-phosphate group attached to a thin plate representing a base, where the long sides of the plates are represented by solid lines with length *a*, short sides of the plates are represented by dotted lines with length *b*, and the thickness of the plates is represented by dashed lines with length *c*. (**A**) Projection of our three dimensional model in the *xz* plane. (**B**) Projection of our three dimensional model in the *xy* plane.

Figure 3. DNA unit orientations ($\hat{\boldsymbol{u}}(\boldsymbol{\theta_{red}}, \boldsymbol{\varphi_{red}})$ for units in the red strand and $\hat{\boldsymbol{u}}(\boldsymbol{\theta_{blue}}, \boldsymbol{\varphi_{blue}})$ for units in the blue strand). The orientation of each unit of DNA is defined as the unit vector normal to the corresponding base plate. (**A**) By definition, the orientations for the all units of an ideal B-type DNA are in the *z* direction, that is $\hat{u} = \hat{z}$. (**B**) The orientation of each unit can change as the DNA molecule is deformed from the ideal double helical structure. The change in orientation can be characterized by two parameters θ and φas shown. (**C**) In case that $\theta_{red} = \theta_{blue} = \theta_{const}$ and $\varphi_{red} = \pi + \varphi_{blue} = \varphi_{const}$ for two units within one base pair, the deformation can manifest in the form of a buckling deformation or in the form of a propeller twist deformation, depending on the angle between the long sides of the plates and φ.

Figure 4. Alteration of orientations. As the base pair with index $n = 0$ is pulled apart, it induces orientation changes in neighboring base pairs. For the case where the change of orientation is a constant over one DNA helical pitch, we see periodic structure changes from buckling backward ($n = 1$) to propeller twist outward ($n = 3$ or $n = 4$) to buckling forward ($n = 6$) to propeller twist inward ($n = 8$ or $n = 9$) as a result of the intrinsic twist of DNA.

Figure 5.A simplified two dimensional model. Identical solid rectangles each representing one DNA base are connected into two strands (one colored blue and the other colored red). By pairing one rectangle in the blue strand to its corresponding rectangle in the red strand we form a two dimensional network resembling a DNA molecule. The behavior of the orientation change

for each DNA base, as defined by the angle between the *z* axis and the corresponding plate main axis perpendicular to side *a*, can be studied by examining the torque balance of the network.

Figure 6. Comparison between results from analytical analysis and simulations. (**A**) Comparison for the orientation parameter θ between analytical theory (eq. 5) as given by solid line and Monte Carlo simulation as given by solid squares. The solid line is obtained by setting the parameters in eq. 5 to the values $l_1 = 1bps$ and $l_D = 9.5bps$. (**B**) Results from the simulations show small variations at each base step for the orientation parameter φ.

Figure 7. Displacements of the Phosphate group as a result of the orientation changes of DNA bases. (**A**) The positions of the phosphate groups according to Monte Carlo our simulations, where for phosphate groups at positions px, py and pz, we have $\cos\theta = px/\sqrt{px^2 + py^2}$ and $\sin\theta = py/\sqrt{px^2 + py^2}$. (**B**) Another version of the positions of the phosphate groups, where θ follows the double helix instead of being confined between 0 to 2π. In both figures, H is the length of the helical pitch of an ideal B-type DNA and the amplitudes of all displacements are multiplied by a factor of 15 for illustration purposes.

Figure 8. Comparison between results from analytical analysis, simulations and experimental observations. The experimental relative binding free-energy of the 2nd protein as a function of the separation between the two protein binding sites on DNA from Ref. 11 are shown as solid red circles with error bars. Our theoretical results of the major groove width changes of the DNA are also shown, with the results from analytical analysis shown by black solid line and results from simulations shown by solid blue squares. Both the black solid line and the solid blue squares are scaled to match the experimentally observed amplitude around $L = 10bps$.

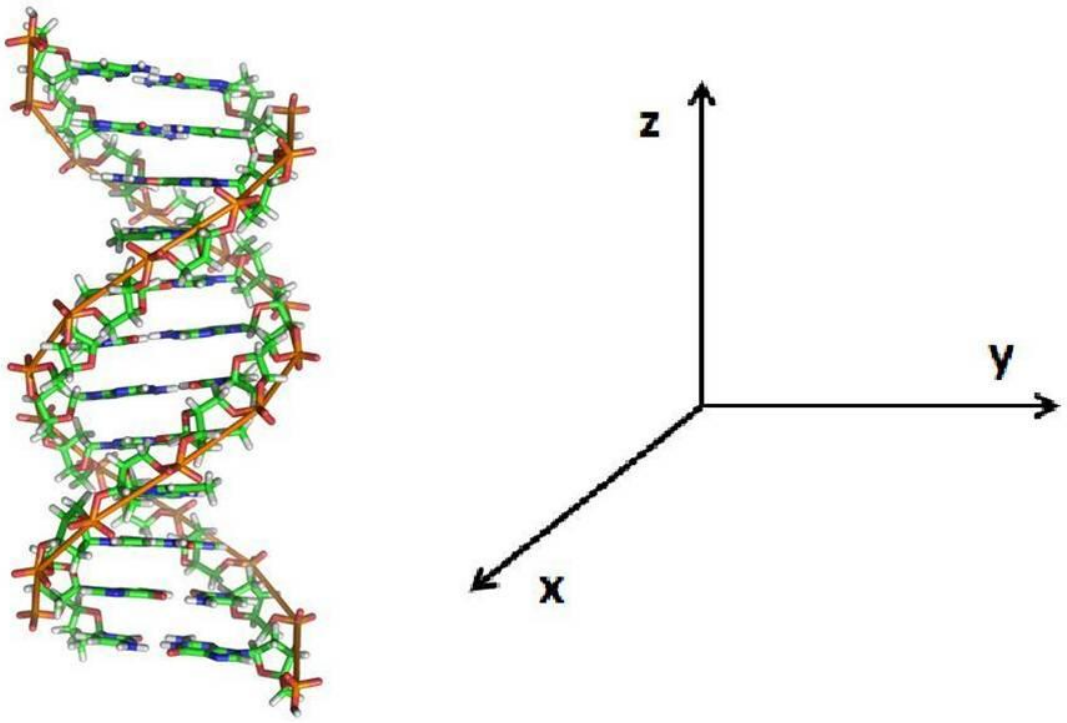


Figure 1

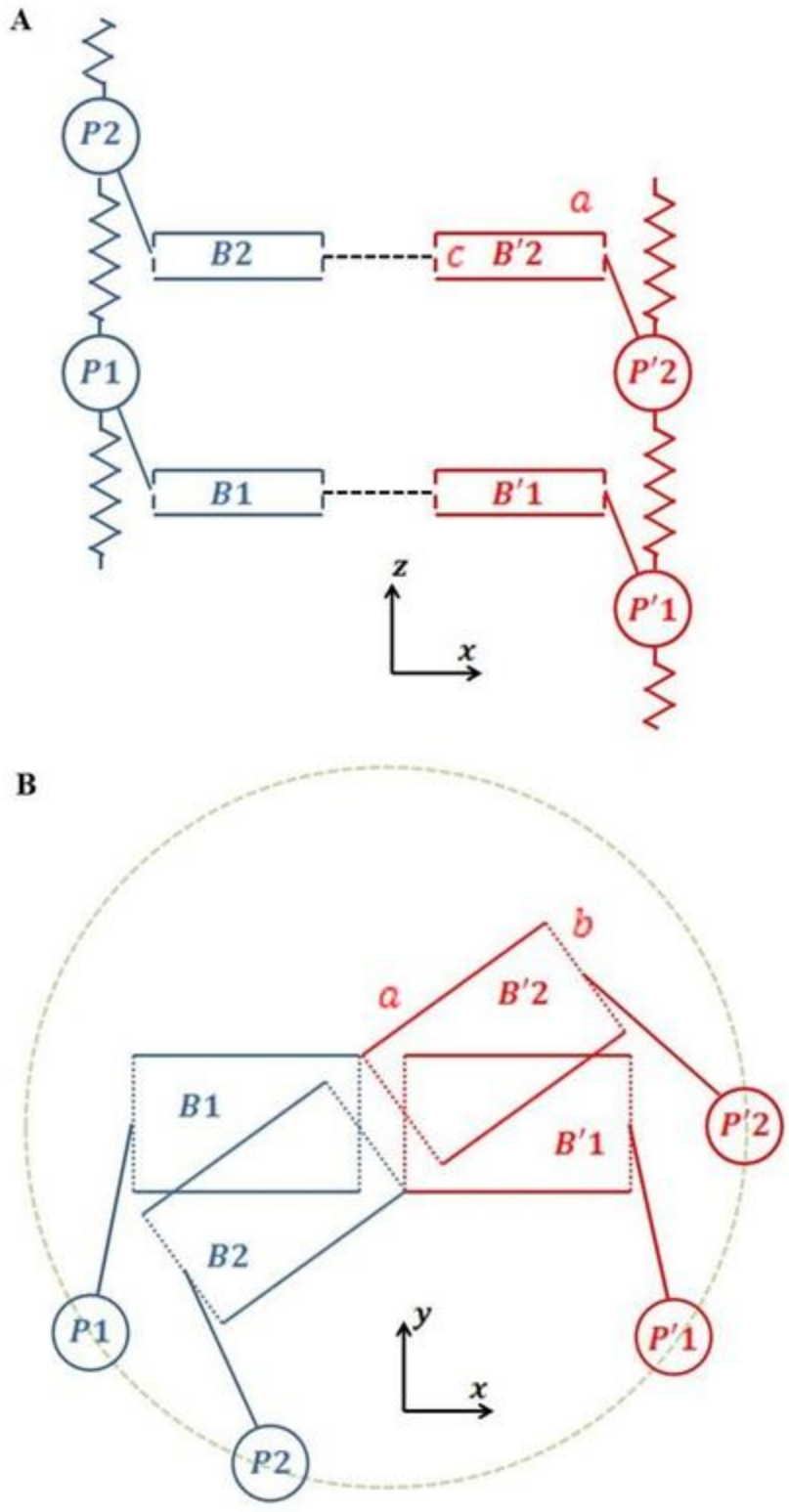


Figure 2

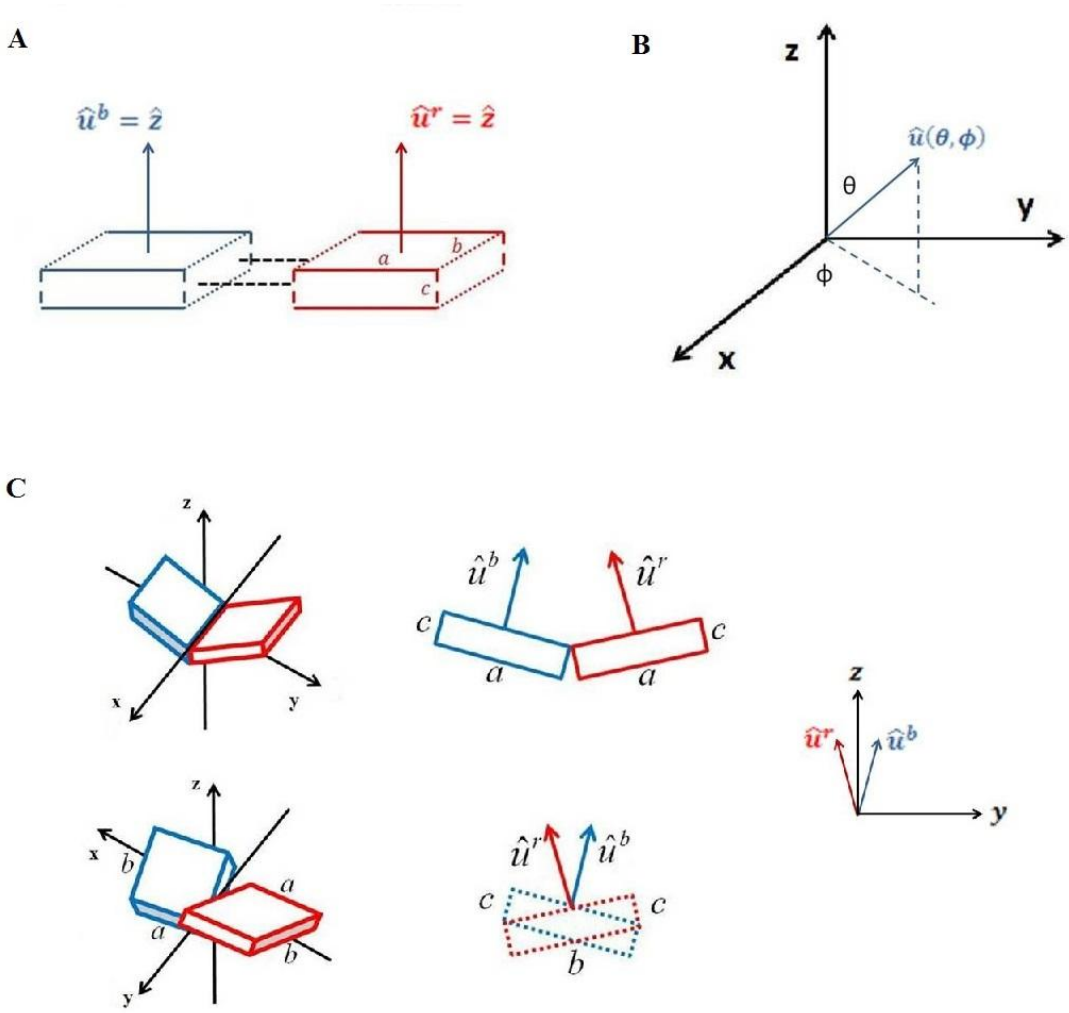

A
$\hat{u}^b = \hat{z}$
$\hat{u}^r = \hat{z}$
a
b
c
B
z
$\hat{u}(\theta, \phi)$
θ
y
ϕ
x
C
z
x
y
$\hat{u}^b$
$\hat{u}^r$
c
a
a
c
z
$\hat{u}^r$
$\hat{u}^b$
y
z
x
b
a
a
b
y
$\hat{u}^r$
$\hat{u}^b$
c
c
b

Figure 3

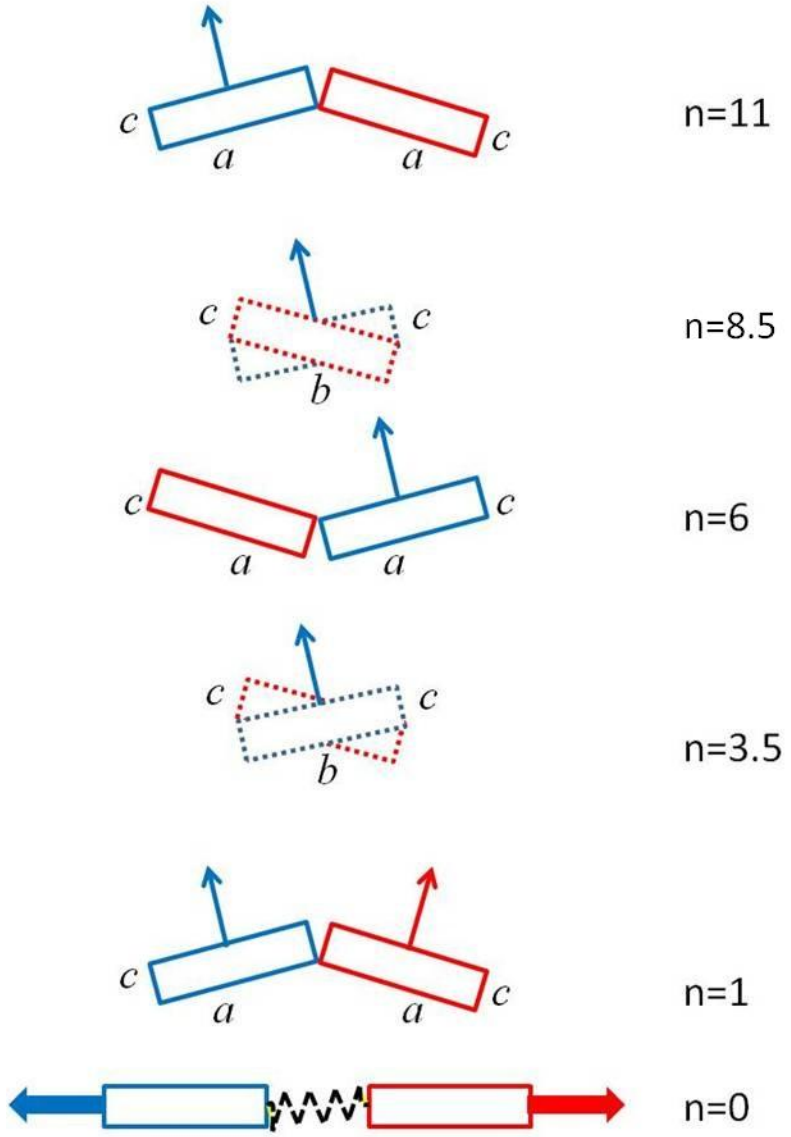


Figure 4

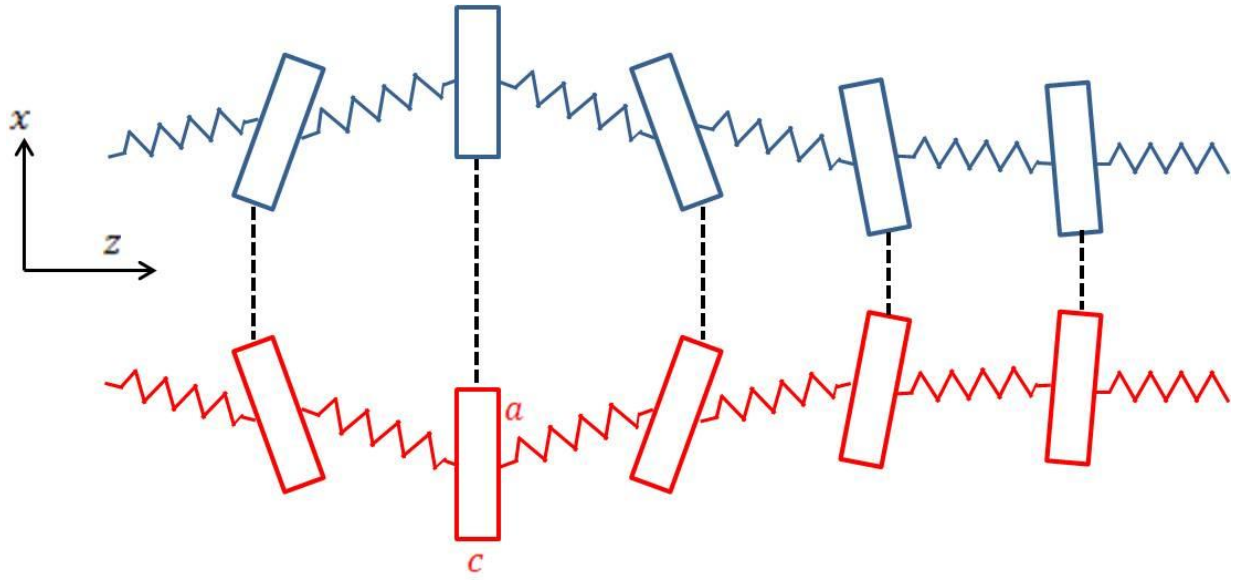


Figure 5

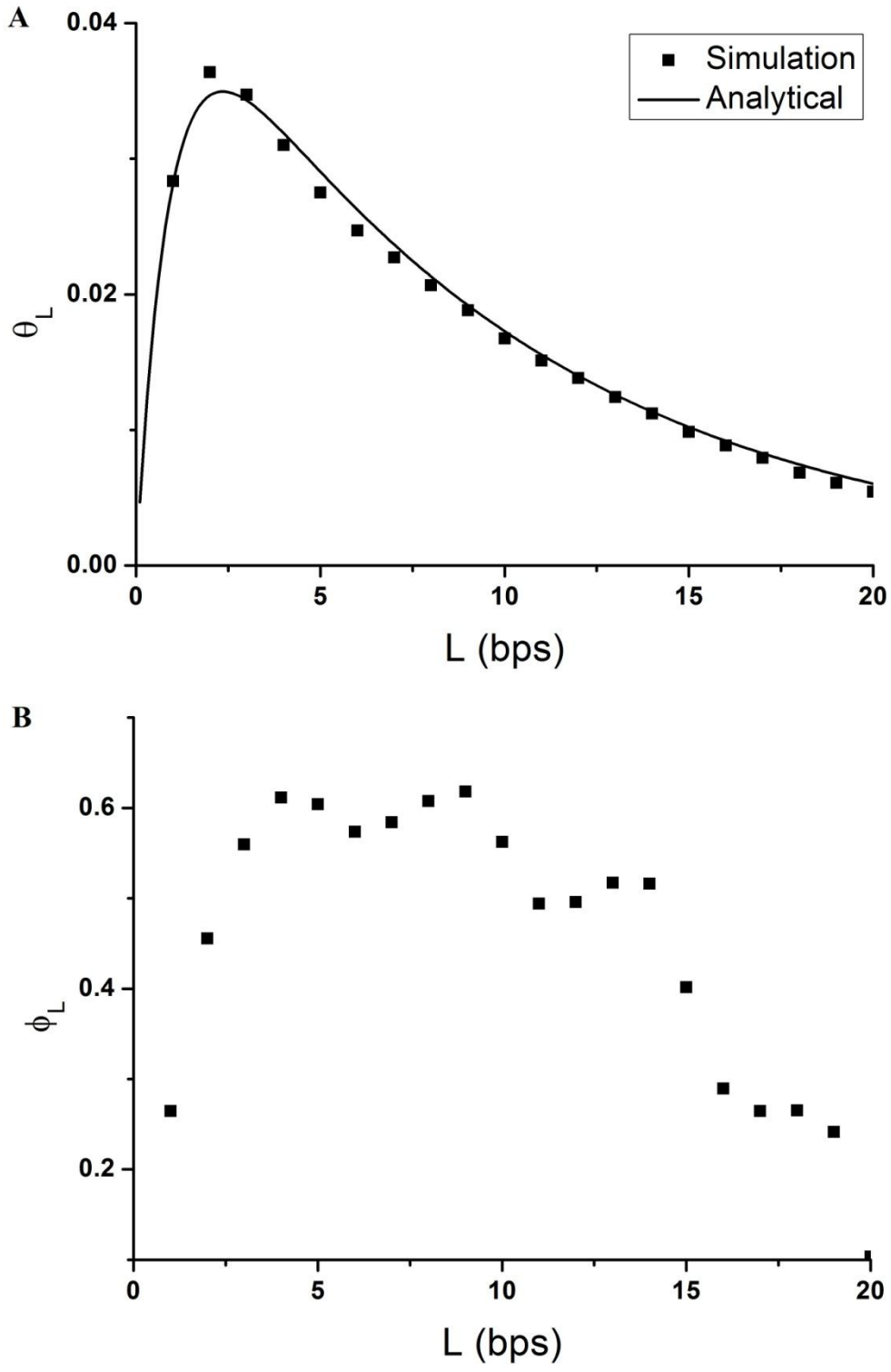
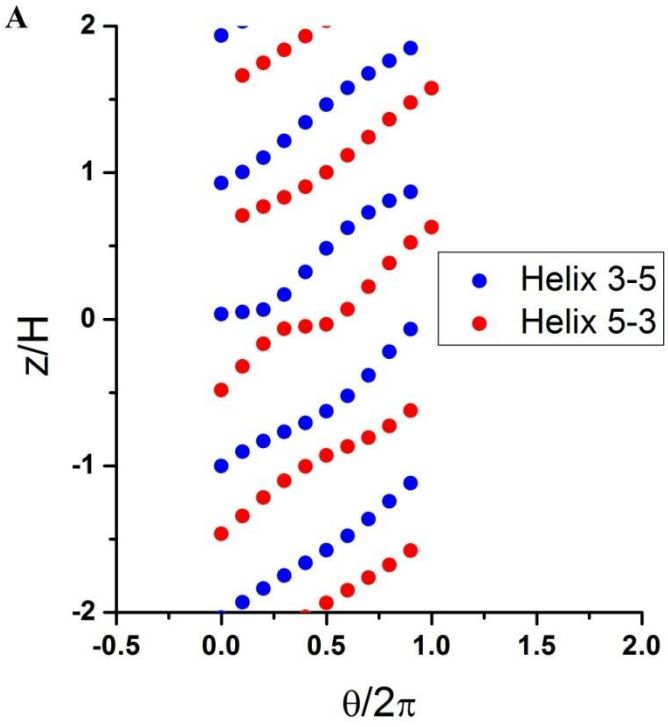

A
0.04
0.02
0.00
θL
Simulation
Analytical
0
5
10
15
20
L (bps)
B
0.6
0.4
0.2
ϕL
0
5
10
15
20
L (bps)

Figure 6


A
2
1
0
-1
-2
z/H
-0.5
0.0
0.5
1.0
1.5
2.0
θ/2π
Helix 3-5
Helix 5-3

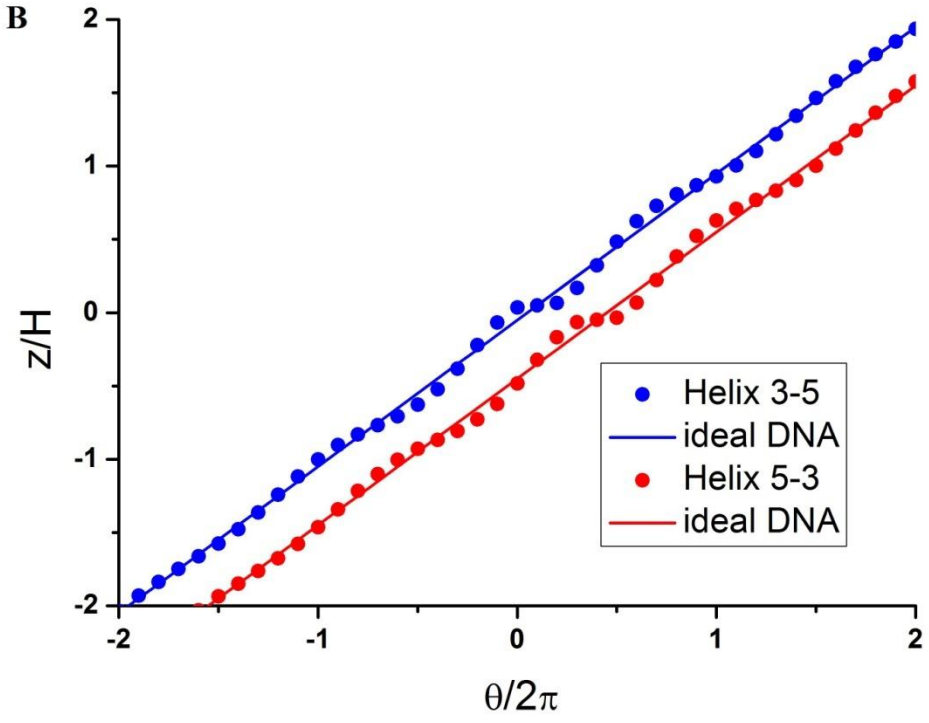

B
2
1
0
-1
-2
z/H
-2
-1
0
1
2
θ/2π
Helix 3-5
ideal DNA
Helix 5-3
ideal DNA

Figure 7

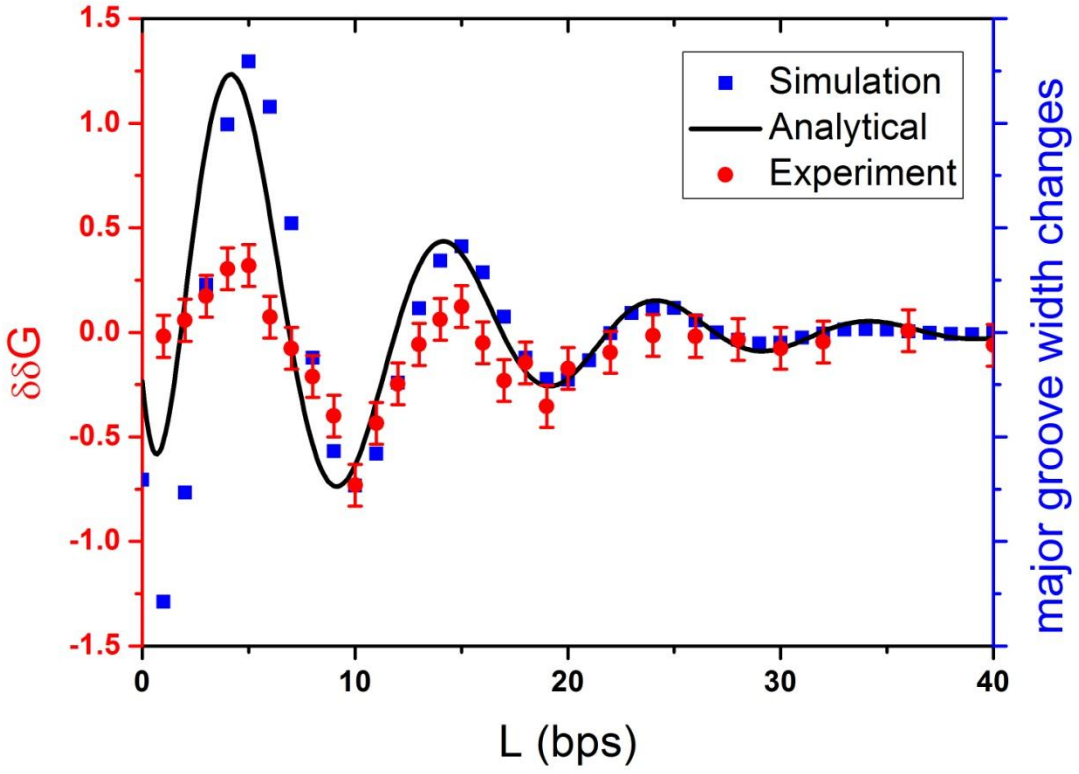

Simulation
Analytical
Experiment
δδG
major groove width changes
L (bps)
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
0
10
20
30
40

Figure 8