Torsional Deformation of Double Helix in Interaction and Aggregation of DNA

A. G. Cherstvy,[†,‡] A. A. Kornyshev,[§] and S. Leikin*[,||]

Institut für Festkörperforschung (IFF), Forschungszentrum Jülich, D-52425 Jülich, Germany, Department of Chemistry, Faculty of Physical Sciences, Imperial College London SW7 2AY, U.K., and National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland 20892

Received: December 30, 2003; In Final Form: March 10, 2004

We incorporate sequence-dependent twisting between adjacent base pairs and torsional elasticity of double helix into the theory of DNA−DNA interaction. The results show that pairing and counterion-induced-aggregation of nonhomologous DNA are accompanied by considerable torsional deformation. The deformation tunes negatively charged phosphate strands and positively charged grooves on opposing molecules to stay "in register", substantially reducing nonideality of the helical structure of DNA. Its cost, however, makes interaction between nonhomologous DNA less energetically favorable. In particular, interaction between double helical DNA may result in sequence homology recognition and selective pairing of homologous fragments containing more than 100−200 base pairs. We also find a weak, but potentially measurable, increase in the expected counterion concentration required for aggregation of nonhomologous DNA and slightly higher solubility of such DNA above the critical concentration.

I. Introduction

Electrostatic interactions between highly charged DNA molecules are believed to play an important role in packaging of genetic material inside cells and viruses and in many other fundamentally important biological processes.[1] Extensive in vitro studies of DNA−DNA interactions,[2] DNA condensation,[3] and structure of DNA aggregates[4−6] were performed, and a variety of models were proposed.[1−3,6] Most of the early models were based on the description of DNA as a simple linear polyelectrolyte (a thin, uniformly charged line) or as a homogeneously charged cylinder.[7−10] While such models captured some features of intermolecular forces, they could not explain many reported experimental observations, e.g., counterion specificity of DNA condensation,[11,12] torsional deformation of double helix upon formation of aggregates,[13] transitions between different forms of DNA,[4,14] etc. The latter observations suggested that intermolecular interactions should be intimately related to finer details of molecular structure, particularly to the helical nature of DNA surface charge pattern.

In order to account for these "details", a theoretical formalism for interaction between rods with arbitrary patterns of surface charges was developed.[15] Applied to molecules with helical surface charge patterns, this theory revealed the following. Counterion specificity of DNA condensation might be caused by preferential adsorption of condensing counterions in the major groove.[16] B-to-A transition in dense aggregates might be promoted by a gain in the electrostatic interaction energy upon the change in the relative width of the minor and major grooves.[17] DNA overwinding from 10.5 base pairs (bp) per helical turn in solution to 10.0 bp/turn in aggregates[13,18,19] might be explained by energetically favorable axial alignment of phosphates.[20] Nontrivial cholesteric pitch behavior[21,22] upon compression of DNA aggregates, subsequent transition from the cholesteric to hexagonal (hexatic) phase,[23,24] and multiple quasicrystalline phases of even more densely packed aggregates[4] might result from alignment of charged phosphate strands against grooves on opposing molecules.[25−29]

Later, this model was extended by incorporation of a realistic sequence-dependent twist between adjacent base pairs and corresponding distortions in the helical pattern of surface charges.[30] It was found that double helices with homologous sequences might recognize each other through electrostatic interactions due to the effect of sequence-dependent distortions on intermolecular interaction. This electrostatic recognition might contribute, e.g., to pairing of intact DNA double helices observed prior to genetic recombination.[31,32]

It was also proposed that close juxtaposition of nonhomologous DNA sequences might cause torsional deformation of double helices. The basic equation describing this effect for two opposing molecules was suggested, but no analysis of a possible role of torsional deformation was performed.[30] Such analysis is the focus of the present study. Here, we evaluate the extent of torsional deformation and clarify the role of such deformation in sequence homology recognition and counterion-induced aggregation of DNA.

The results of this work are conceptually clear although some of the equations appear quite elaborate and their derivation involves a number of nonobvious steps. Thus, to simplify the task for the readers interested more in the physics of the results and their implications rather than in the mathematical details, we describe the theory in the extensive Appendix (which includes the discussion of the model, derivations, and the resulting expressions for the torsional deformation and interaction free energy). In the main text, we present the basic concepts and only few simple equations essential for understanding these concepts (section II). In section III, we proceed directly to

* Corresponding author. E-mail: leikins@mail.nih.gov.
[†] Institut für Festkörperforshung (IFF).
[‡] On leave from Institute of Physics, National Academy of Sciences of Belarus, 220072 Minsk, Belarus.
[§] Imperial College London.
[||] National Institutes of Health.

10.1021/jp0380475 CCC: $27.50
Published on Web 04/27/2004

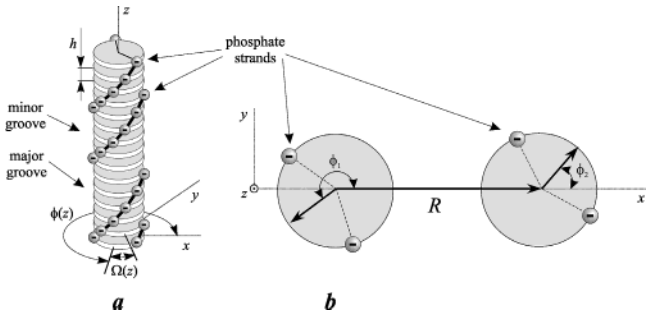

Figure 1. (a) Schematic representation of B-DNA and (b) cross-section of two parallel B-DNA molecules separated by distance *R*. The disks in part a represent DNA base pairs stacked with the axial step *h* and twisted by the angle $\Omega(z)$ with respect to each other. For calculation of electrostatic interaction energy, water-impermeable, low-dielectric-constant cores of DNA are modeled as dielectric cylinders shown by large shaded circles in part b. Charged phosphate strands, shown as chains of small spheres in part a, are modeled as negatively charged helical lines at the surfaces of the dielectric cores. *Azimuthal orientation* $\phi_i(z)$ of each molecule is defined as the angle between the *x*-axis and the vector pointing from the center of the molecule to the middle of the smaller arc (minor groove) between the strands.

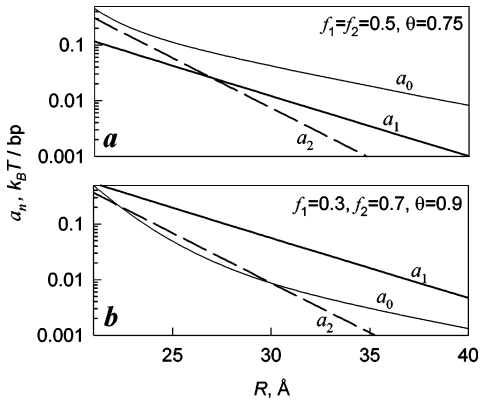

Figure 2. Zeroth (a_0), first (a_1), and second (a_2) helical harmonics of electrostatic interactions between two B-DNA molecules surrounded by electrolyte solution (eq 5). The following parameters were used for the calculations: DNA core radius $r = 9.5$ Å, effective width of surface charged groups $w = 5$ Å (see eq A7), electrolyte screening length $\kappa_D^{-1} = 7$ Å, water dielectric constant $\epsilon = 80$, minor groove half-width $\tilde{\phi}_s = 0.4\,\pi$. The distance dependence and relative contributions to the energy of different helical harmonics in hexagonal aggregates are qualitatively similar.

displaying the results and describing their physical meaning. We conclude by discussing the most important predictions and comparing them with available experimental data (section IV).

II. Basic Concepts

II.1. Structure and Torsional Deformation of Double Helical DNA. We describe the helical conformation of DNA in terms of the azimuthal orientation of the middle of its minor grove at the axial position *z* (Figure 1). Since DNA consists of base pairs stacked with the axial step *h*, in the absence of torsional deformation, this orientation is given by the following recursive relationship[30]

$$\phi(z + h) = \phi(z) + \Omega(z) \tag{1}$$

where $\Omega(z)$ is the preferred (intrinsic) twist angle between adjacent base pairs.

In a geometrically ideal helix, the twist angle between adjacent base pairs is constant along the molecular axis ($\Omega(z) = \text{const}$) and $\phi(z) = \phi(0) + \Omega z/h$. However, in nature, DNA is not an ideal helix. Instead, it has 10 distinct combinations of two adjacent base pairs, all of which have different preferred values of Ω.[33−35] The axial pattern of this intrinsic twist angle $\Omega(z)$ is a unique, sequence-dependent "fingerprint" of DNA structure. The deviation from the average twist angle $\langle\Omega\rangle$

$$\omega(z) = \Omega(z) - \langle\Omega\rangle \tag{2}$$

is relatively small ($\sqrt{\langle\omega^2\rangle} \equiv \Delta\Omega \approx$ 4-6°, $\langle\Omega\rangle \approx 34°$, $\langle\ \rangle$ denotes an ensemble average over all possible $\Omega(z)$). Nevertheless, this nonideality has important implications for interactions between DNA, as we will see below.

Because $\Delta\Omega/\Omega \ll 1$, we can replace the discrete recursive relationship (eq 1) by a more convenient continuous description

$$h\ \mathrm{d}\phi(z)/\mathrm{d}z = \Omega(z) \tag{3}$$

which defines the azimuthal orientation of DNA in the absence of torsional deformation. Torsional deformation results in deviation of $\phi(z)$ from this relationship. Its energetic cost can then be described by[30]

$$E^{\mathrm{tors}} = \frac{C}{2}\int_0^L \mathrm{d}z\left(\frac{\mathrm{d}\phi(z)}{\mathrm{d}z} - \frac{\Omega(z)}{h}\right)^2 \tag{4}$$

where *L* is the length and *C* ($\approx 3 \times 10^{-19}$ erg cm)[36] is the torsional elasticity modulus of DNA.[37]

II.2. Electrostatic Interactions between DNA Molecules. Basic expressions for the energy of electrostatic interaction between DNA double helices were derived in several previous studies.[15−17,25,30] It was shown that DNA can be approximated by a dielectric cylinder with two negatively charged helical lines on its surface, which represent charged phosphate strands (Figure 1). In biologically relevant DNA pairs and in most commonly studied DNA aggregates, surface-to-surface separations between the molecules are smaller or of the order of the distances between the charged strands on the DNA surface. As a result, the energy of electrostatic interaction between two opposing molecules ($i = 1, 2$) is a functional of the difference in their local azimuthal orientations, $\delta\phi(z) = \phi_1(z) - \phi_2(z)$, at the axial position *z* (Figure 1).[15]

The general functional form of the dependence of the electrostatic interaction energy on $\delta\phi(z)$ is given by[30]

$$E = a_0 L - a_1\int_0^L \cos[\delta\phi(z)]\ \mathrm{d}z + a_2\int_0^L \cos[2\delta\phi(z)]\ \mathrm{d}z \tag{5}$$

where the first, second, and third terms describe the zero, first, and second helical harmonics of charge−charge interactions correspondingly (for B-DNA higher harmonics can be neglected[15]) and the first term also contains a contribution of image-charge forces. A more detailed description of this energy and general expressions for a_n are presented in the Appendix. These coefficients depend on the interaxial distance *R* between the molecules, on the fraction θ of bare DNA charge neutralized by adsorbed counterions, on the fractions *f* of adsorbed counterions located in the minor (f_1) and major (f_2) grooves, and on the Debye screening length κ_D^{-1} in surrounding solution. Figure 2 illustrates the dependencies of a_n on *R* under conditions unfavorable ($\theta = 0.75$, $f_1 = f_2 = 0.5$, $\kappa_D^{-1} = 7$ Å) and favorable ($\theta = 0.9$, $f_1 = 0.3$, $f_2 = 0.7$, $\kappa_D^{-1} = 7$ Å) for DNA aggregation. Note that the functional dependence of the electrostatic energy

on $\delta\phi(z)$ has the same form of eq 5 both for two molecules surrounded by electrolyte solution and for aggregates. However, because pair electrostatic interactions in aggregates are not additive, the coefficients a_n are described by different formulas in these two cases (see Appendix).

Because of counterion condensation, at least ~75% of bare DNA charge is expected to be neutralized ($\theta \geq 0.75$) even at low ionic strength.[9] As a result, the $\delta\phi$-dependent terms give important contributions to the energy of electrostatic interaction between DNA at all relevant distances (Figure 2). Indeed, ideal helices can establish an energetically favorable alignment, $\delta\phi(z) = \delta\phi^0 = \text{const}$, minimizing unfavorable phosphate−phosphate interactions and maximizing favorable phosphate−counterion interactions on opposing surfaces. As follows from eq 5 and Figure 2a, $\delta\phi^0$ depends on R, and such alignment can substantially reduce intermolecular repulsion.[15,16]

The dependence of the interaction energy on $\delta\phi$ is further enhanced by preferential adsorption of counterions in the grooves (as illustrated in Figure 2 by the increase in a_1 at $\theta = 0.9$, $f_1 = 0.3$, $f_2 = 0.7$). Accumulation of a sufficiently large positive charge in the grooves allows the alignment of negatively charged phosphate strands opposite to positively charged grooves resulting in intermolecular attraction and DNA aggregation. It was argued that this mechanism might explain, e.g., the observed DNA condensation by spermine, spermidine, and other biologically important DNA counterions.[15,16]

II.3. Effect of Structure on Electrostatic Interactions. Ideal alignment of two opposing DNA molecules in the most energetically favorable conformation requires constant $\delta\phi$, $\delta\phi(z) = \delta\phi^0$, along the whole juxtaposition length. For nonideal helices with nonhomologous sequences this is impossible without torsional deformation.[30] In this case, only an average optimal alignment, $\langle\delta\phi\rangle = \delta\phi^0$, can be established. (Hereafter, $\langle\ \rangle$ denotes averaging over all possible realizations of $\Omega(z)$ and over the juxtaposition length L.[38]) The deviation from this alignment, $\langle[\delta\phi(z) - \langle\delta\phi\rangle]^2\rangle$, plays an important role in DNA pairing and aggregation.

Indeed, for two ideal helices or for two homologous molecules, $\langle[\delta\phi(z) - \langle\delta\phi\rangle]^2\rangle = 0$. Such molecules can remain perfectly aligned over any juxtaposition length. For two nonhomologous molecules, straightforward integration of eq 3 shows that in the absence of torsional deformation

$$\langle[\delta\phi(z) - \langle\delta\phi\rangle]^2\rangle = \frac{L}{2\lambda_c} \tag{6}$$

Here we assumed that the range of pair correlations in ω does not exceed several base pairs and introduced the helical coherence length of DNA, λ_c, as[39]

$$\lambda_c = \frac{h^2}{\int_0^L \langle\omega(z)\omega(z')\rangle_\Omega \, dz'} \tag{7}$$

where $\langle\ \rangle_\Omega$ denotes that the averaging should be performed only over possible realizations of $\Omega(z)$, but not over the juxtaposition length. For molecules with completely random base pair sequences, $\lambda_c = h/\Delta\Omega^2 \approx 300-700$ Å.[30] Even though nonideality of DNA helix is small, it results in accumulation of large alignment errors over large juxtaposition lengths.

Thus, intact, nonhomologous DNA cannot retain the energetically favorable strand-groove alignment for juxtaposition lengths exceeding λ_c.[30] Their alignment can be restored by torsional deformation, but at the cost of the corresponding elastic energy (eq 4). Minimization of the sum of the energies shows that the electrostatic cost of misalignment is higher than the elastic cost of at least some deformation. Therefore, pairing or aggregation of nonhomologous DNA should cause torsional deformation, and the cost of the deformation should contribute to the overall interaction energy. The corresponding calculations are fairly elaborate because of the nonlinear dependence of the electrostatic energy on $\delta\phi(z)$. Their mathematical details are reported in the Appendix, and their results are presented in the next section.

III. Results

III.1. Torsional Deformation, Alignment, and Interaction between Two DNA Molecules in Juxtaposition. In summary, the results of minimization of the sum of torsional and electrostatic energies with respect to torsional deformation are as follows. The average optimal alignment $\langle\delta\phi(z)\rangle = \delta\phi^0$ is given by eqs A37 and A39 and is plotted in Figure 3a,b. The mean square deviation from this alignment is shown in Figure 3c,d and is described by (see Appendix, section A3.2)

$$\langle[\delta\phi(z) - \delta\phi^0]^2\rangle = \frac{\lambda}{2\lambda_c} G\left(\frac{2L}{\lambda}\right) \tag{8}$$

where $G(x) = [1 - (1 - e^{-x})/x] \leq 1$ and λ is determined by eqs A37 and A39. The total interaction energy per base pair, F_{bp}, is given by eqs A35−A39 and is plotted in Figure 3e,f. The values of $\delta\phi^0$, λ, and F_{bp} depend on the juxtaposition length, electrolyte concentration, and counterion adsorption pattern. They are plotted in Figure 3 at $L \gg \lambda$, physiological electrolyte concentration ($\kappa_D^{-1} \approx 7$ Å), and two different counterion patterns.

Probably the most important consequence of torsional deformation is the appearance of the torsional adaptation length λ, which limits accumulation of the alignment error. At $L \ll \lambda$, the torsional deformation is minimal and $\langle(\delta\phi - \delta\phi^0)^2\rangle \approx L/2\lambda_c$ as determined by eq 6. At $L \gg \lambda$, the torsional deformation becomes significant, and it prevents unlimited accumulation of the alignment error. Instead of unlimited growth at large L described by eq 6, the deviation from the optimal alignment levels off at $\langle(\delta\phi - \delta\phi\ ^0)^2\rangle \approx \lambda/2\lambda_c$ and becomes independent of L.

Torsional adaptation of nonhomologous molecules improves their alignment, but $\delta\phi^0$ still differs from that expected for ideal helices. Figure 3a,b shows that the optimal alignment of ideal and of nonhomologous DNA is $\delta\phi^0 = 0$ at large interaxial distances R and $\delta\phi^0 \neq 0$ at small R. In the case of ideal helices, the transition from zero to nonzero $\delta\phi^0$ upon decreasing R is of the second-order.[15] In the case of two nonhomologous DNA, it occurs at smaller R and stepwise as a first-order transition. (Apparently, the difference in the preferred twist angle, $\delta\omega(z)$, plays the role of an "external field" causing the change in the transition order.)

The torsional adaptation length and, therefore, the mean square deviation from the optimal alignment tend to decrease with increasing strength of intermolecular interaction. Therefore, they depend on the interaxial distance and tend to decrease at smaller R (Figure 3c,d). A small rise near the transition point is caused by increased susceptibility with respect to fluctuations in $\delta\phi$ (as expected near a phase transition). At small interaxial separations, torsional deformation becomes so strong and torsional adaptation length becomes so small that the alignment of nonhomologous DNA becomes almost as good as between ideal helices.

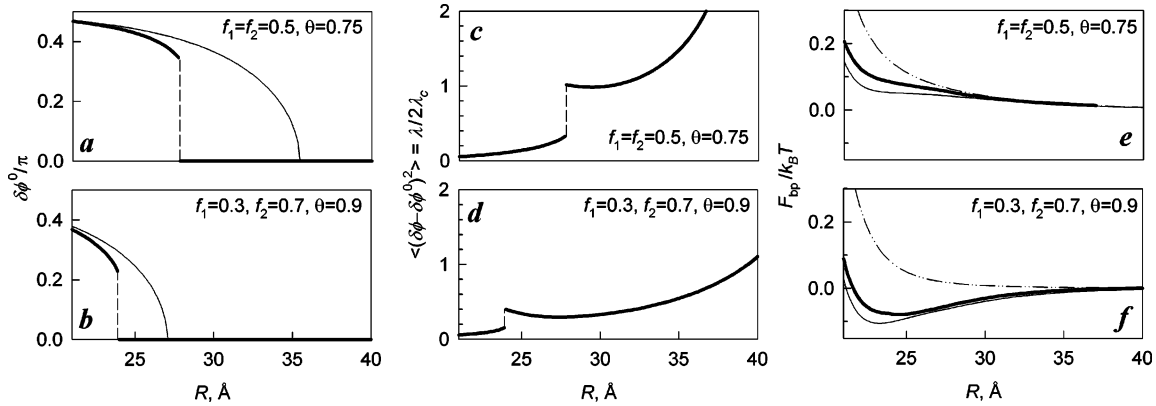

Figure 3. (a, b) Average mutual alignment, (c, d) alignment error, and (e, f) interaction energy per base pair between parallel, long ($L \gg \lambda$, $G(2L/\lambda) \approx 1$) DNA under (a, c, e) unfavorable and (b, d, f) favorable conditions for aggregation. Bold lines show the most energetically favorable conformation and interaction energy for DNA with unrelated sequences. Dashed lines mark first-order transitions. Thin solid lines show alignment and interaction energy for ideal helices. Dash−dotted lines in parts e and f show the interaction energy between nonhomologous, rigid ($C \rightarrow \infty$) helices. The following parameters were used: $C \approx 3 \times 10^{-19}$ erg cm,[36] $h = 3.4$ Å, and $\Delta\Omega = 6°$ ($\lambda_c = 310$ Å). Other parameters were the same as in Figure 2.

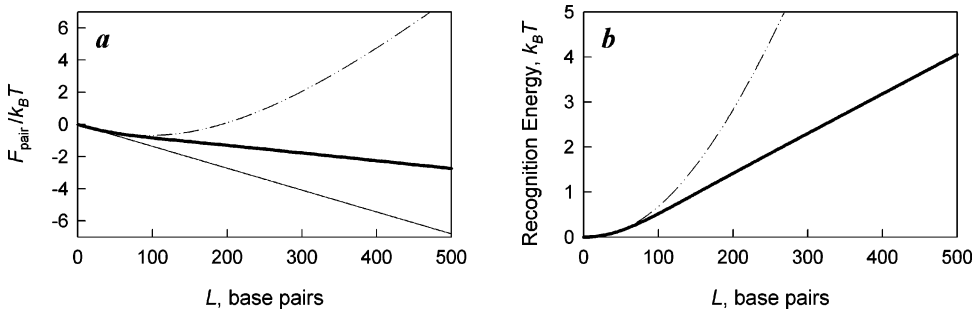

Figure 4. (a) Pair interaction energy (F_{pair}) for nonhomologous torsionally rigid helices (dash−dot lines), nonhomologous DNA (bold line), and ideal helices (F_{ideal}, thin, solid line). The interaction energy between two homologous DNA molecules at optimal alignment is the same as between two ideal helices. (b) Recognition energy ($F_{pair} - F_{ideal}$) for torsionally rigid helices (dash−dot lines) and DNA (bold line). The energies were calculated from eqs A35−A39 with the following parameters: $f_1 = 0.3$, $f_2 = 0.7$, $\theta = 0.8$, $R = 30$ Å, $C = \infty$ for torsionally rigid helices and $C = 3 \times 10^{-19}$ erg cm[36] for DNA. All other parameters were the same as in Figures 2 and 3.

Figure 3e,f shows that torsional adaptation results in a qualitative change in intermolecular interaction compared to the behavior expected in the absence of the deformation ($C \rightarrow \infty$). Most importantly, it prevents the loss of counterion-induced intermolecular attraction at conditions favorable for counterion induced aggregation (Figure 3f). Overall, because of the relatively low cost of torsional deformation, nonideality of the helical structure of DNA has only a minor effect on the interaction energy, instead of dramatic changes expected for torsionally rigid helices.

III.2. Electrostatic Sequence Homology Recognition in Pairing of DNA Duplexes. Although torsional deformation prevents the loss of counterion induced attraction, the residual twist and the cost of the deformation weaken the attraction between nonhomologous DNA compared to ideal helices (Figure 4a). The higher the extent of sequence homology is, the smaller the torsional deformation required to retain the alignment and the lower the cost of this deformation. Indeed, the preferred twist angles $\Omega(z)$ of two DNA with identical sequences are the same and, therefore, the molecules can remain perfectly aligned without any torsional deformation. (The energy of interaction between DNA with identical sequences at optimal mutual alignment is the same as between two ideal helices.)

The difference in the interaction energy between two nonhomologous and between two homologous DNA fragments of the same length is the sequence homology recognition energy.[30] It is plotted versus the juxtaposition length in Figure 4b. Torsional deformation reduces the recognition energy compared to infinitely rigid helices, but this energy still exceeds k_BT for sufficiently large fragments. In particular, it becomes sufficient for selective pairing of homologous DNA when more than 100−200 base pairs come into close juxtaposition.

III.3. DNA Interactions in Hexagonal Aggregates and Counterion-Induced Condensation. In many respects the results for hexagonal aggregates of DNA are similar to the results for two DNA samples in juxtaposition. Intermolecular interaction causes substantial torsional deformation. The resulting torsional adaptation allows better molecular alignment and prevents the loss of intermolecular attraction at conditions favorable for counterion-induced DNA aggregation. The alignment undergoes a first-order transition upon decreasing interaxial distance. But, there are several important differences: (i) At large distances mutual alignment of DNA in aggregates is the same as in pairs, $\delta\phi^0(R) = 0$. At small R, however, the same optimal value of nonzero $\delta\phi^0$ cannot be realized for all nearest neighbor pairs in a hexagonal aggregate because this would be incompatible with the aggregate symmetry. Thus, the alignment becomes more complex, as illustrated in Figure 5a,b.

(ii) The mean square deviation of $\delta\phi$ for every pair of nearest neighbor molecules in a hexagonal aggregate from $\delta\phi^0$ is still described by eq 8. However, each molecule in an aggregate is involved in six pair interactions with its six nearest neighbors. As a result, the total electrostatic interaction energy becomes even more dominant compared to the elastic deformation energy, the torsional deformation becomes stronger, the torsional adaptation length decreases, and $\langle(\delta\phi\text{-}\delta\phi^0)^2\rangle$ becomes smaller (Figure 5c).

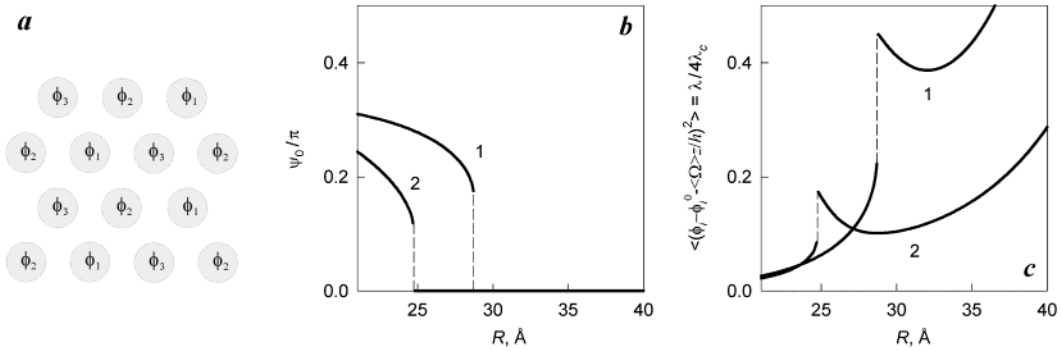

Figure 5. (a) Packing, (b) mutual alignment, $\psi_0 = |\phi_1 - \phi_2| = |\phi_1 - \phi_3| = 0.5|\phi_2 - \phi_3|$, and (c) deviation of DNA from an ideal helix in hexagonal aggregates calculated from eqs A43−A47 at $f_1 = f_2 = 0.5$, $\theta = 0.75$ (curves 1 in parts b and c) and $f_1 = 0.3$, $f_2 = 0.7$, $\theta = 0.9$ (curves 2 in parts b and c). All other parameters used for the calculation were the same as in Figures 2 and 3.

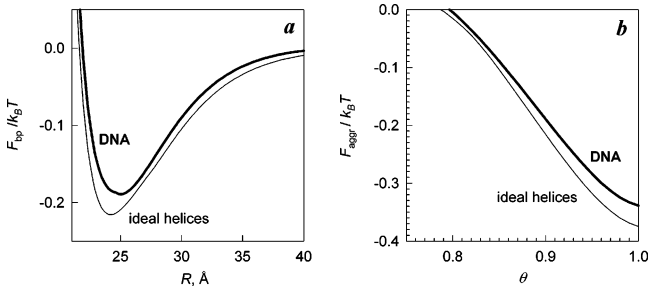

Figure 6. (a) Dependence of the interaction energy per base pair (F_{bp}) on interaxial spacing in hexagonal aggregate for ideal helices (thin line) and nonhomologous DNA (bold line) at $f_1 = 0.3$, $f_2 = 0.7$, $\theta = 0.9$. (b) Dependence of the aggregation energy (F_{bp} at the energy minimum) on the fraction of DNA charge neutralized by bound counterions at $f_1 = 0.3$, $f_2 = 0.7$. All other parameters were the same as in Figures 2 and 3.

(iii) In DNA pairs, $\delta\phi = \phi_1(z) - \phi_2(z)$ approaches the alignment of ideal helices, but $\Sigma\phi = \phi_1(z) + \phi_2(z)$ still follows the preferential twist profiles and might be far from ideal ($h\, d\Sigma\phi/dz = \Omega_1(z) + \Omega_2(z)$). In hexagonal aggregates, simultaneous optimization of $\delta\phi$ and $\Sigma\phi$ for all nearest neighbor pairs of molecules is impossible. As a result, torsional adaptation limits the deviation of each molecule i from a geometrically ideal helix $\phi_i^0 + \langle\Omega\rangle z/h$ (see Appendix, section A3.2)

$$\left\langle \left[\phi_i(z) - \phi_i^{\,0} - \frac{\langle\Omega\rangle z}{h} \right]^2 \right\rangle = \frac{\lambda}{4\lambda_c} G\left(\frac{2L}{\lambda}\right) \qquad (9)$$

as illustrated in Figure 5c.

Similar to interaction between two molecules (Figure 3e,f), neutralization of a sufficiently high fraction of DNA charge by counterions and preferential adsorption of these counterions in the major groove lead to intermolecular attraction in aggregates. Under such conditions, formation of hexagonal DNA aggregates becomes energetically favorable. In particular, Figure 6a shows the interaction energy versus separation for molecules with nonhomologous sequences upon 90% neutralization of DNA charge by counterions adsorbed in the major (70%) and minor (30%) grooves and a similar curve for ideal helices. Figure 6b compares the dependence of the energy gain upon aggregation on the extent of charge neutralization in both cases. Although higher extent of charge neutralization is required for aggregation of nonhomologous DNA, this difference is not significant. For typical B-DNA parameters, the energy gain upon aggregation is only weakly affected by deviation of DNA structure from an ideal helix.[40]

IV. Discussion

Overall, the following lessons can be drawn from the present study.

Intermolecular interaction causes torsional deformation of DNA which reduces sequence-dependent variation of the twist angle between adjacent base pairs. Our calculations show that this deformation enables long-range alignment of negatively charged strands and positively charged grooves on opposing nonhomologous DNA molecules and makes electrostatic interaction between them more favorable. Its extent and role in intermolecular interaction grow with increasing length of DNA and increasing number of molecules involved. In hexagonal aggregates formed by long, nonhomologous DNA, this deformation becomes so strong that it almost completely eliminates the variation of the twist angle between adjacent base pairs and produces nearly ideal helical conformation of the molecules (Figure 5c).

These results provide a natural explanation of several important observations. For instance, already early X-ray studies of long, nonhomologous DNA from natural sources found its conformation in aggregates to be close to ideal helices with constant twist angle between base pairs.[4,14] Only later studies of DNA structure in solution and analyses of structures of crystals formed by identical synthetic DNA oligonucleotides revealed that the base pair twist angles vary substantially dependent on the sequence.[33−35,41]

It was also found that nonhomologous DNA fragments have different average twist angle in solution (∼10.5 bp/turn)[13,18] and in aggregates (10.0 bp/turn),[13,19] but the origin of the torsional deformation was not understood at the time.[13] We believe that such change is caused by a combination of (i) the gain in the electrostatic interaction energy upon removal of sequence-dependent twist variation and (ii) an extra gain in the electrostatic interaction energy upon axial alignment of phosphates at integral number of base pairs per turn.[20]

Nonideal helical structure of DNA backbone has only a minor effect on counterion induced DNA condensation. DNA forms hexagonal aggregates when it is mixed with water at such high concentration that it has to be densely packed. However, it also forms hexagonal aggregates upon spontaneous precipitation from low concentration solutions by some divalent transition metal ions (e.g., Mn^{2+} and Cd^{2+}) and biologically active polyions (e.g., spermine, spermidine, protamine, etc.).[3,42] It was argued that these ions condense DNA by binding in grooves and inducing electrostatically favorable zipperlike alignment of positively charged grooves and negatively charged phosphate strands on opposing molecules.[16] This "electrostatic zipper" mechanism explains counterion specificity of DNA condensation, and it

correctly predicts the spacing between DNA in aggregates and the value of the aggregation energy[12,43] as well as different propensity of different forms of DNA toward aggregation.[17,20,44] However, this model was proposed on the basis of calculations for idealized double helices[15,16] while DNA is not an ideal helix. Long-range zipperlike alignment of strands and grooves of opposing nonideal helices would be impossible without torsional deformation. Thus, it remained unclear whether the energetic cost of necessary deformation would have a substantial effect on the ability of counterions to condense DNA or even completely preclude this condensation mechanism.

In the present study, we calculated the extent and energetic cost of the torsional deformation upon aggregation. Even though we used the torsional elasticity modulus at the upper end of experimentally measured values,[36,45] we found that the cost of necessary torsional deformation in hexagonal aggregates is small compared to the gain from favorable electrostatic interactions. The nonideality of DNA structure does not preclude the electrostatic zipper mechanism and has only a minor effect on its predicted features, e.g., the condensation energy (Figure 6) and the fraction of charge neutralization required for the process to proceed (Figure 6b).

Softening of DNA torsional rigidity by counterion binding or elevated temperature may promote DNA condensation. Although the cost of torsional deformation upon DNA condensation into aggregates is small compared to the total aggregation energy at most favorable aggregation conditions, it is still larger than k_BT per DNA persistence length ($\sim$150 bp).[40] Thus, it might become important at marginal aggregation conditions causing DNA to aggregate more readily when this cost is reduced. For instance, torsional rigidity of DNA is known to decrease at elevated temperature,[45] and it should vanish completely upon approach to the denaturation temperature. Binding of Mn^{2+} reduces denaturation temperature of DNA,[46,47] and therefore, it should lead to even stronger decrease in the torsional rigidity of DNA with temperature. This effect may enhance the ability of Mn^{2+} to condense DNA at elevated temperature (for more detailed discussion of this phenomenon and its other possible mechanisms see, e.g., ref 48 and references therein).

Torsional deformation weakens but does not eliminate electrostatic sequence homology recognition between double helical DNA. Since torsional deformation reduces the sequence-dependent variation of the twist angle between adjacent base pairs, it reduces the energetic advantage of interaction between molecules with identical sequences (Figure 4a) and, therefore, weakens sequence homology recognition between double helical DNA (Figure 4b). Still, as suggested by earlier estimates for torsionally rigid molecules,[30] the recognition energy does exceed the thermal energy (k_BT) for DNA fragments longer than 100−200 bp (Figure 3b). This calculation confirms our previous conclusion[30] that the electrostatic recognition might contribute to selective pairing of homologous double helical DNA fragments in cells prior to homologous recombination.

Of course, processes in vivo are much more complex than interaction between two DNA molecules in an electrolyte solution modeled in the present study. They occur in a very dense soup of other macromolecules and are likely to involve some proteins. Nevertheless, in cellulo experiments have given indications of transient pairing of homologous fragments of double helical DNA which precedes genetic recombination.[31,32] Such pairing was assumed to be caused by some weak direct interactions between DNA whose nature remained unclear.[31,32] Despite all its simplifications, our theory might offer a natural explanation for these interactions. It is also tempting to speculate that the 100−200 bp sequence homology requirement for recombination[49−51] is related to 100−200 bp requirement for efficient electrostatic recognition.

Acknowledgment. The authors are grateful to A. M. Berezhkovskij, V. A. Parsegian, D. C. Rau, J. M. Schurr, and V. Zhurkin for many stimulating discussions. A.A.K. acknowledges the support of this work by Royal Society (Wolfson Research Merit Award) and Deutsche Forschungsgemeinschaft, Grant KO 1391/4-1. A.A.K. and A.G.C. acknowledge the financial support of their visits to Bethesda by the National Institute of Child Health and Human Development, NIH. The work is a part of EPSRC funded research program (GR/S31068/01).

Appendix: Theory

A1. Model and Approximations. *A1.1. Electrostatic Interaction between Two DNA Molecules in Juxtaposition*. The energy of electrostatic interaction between two long ($L \gg 2\pi h/\langle\Omega\rangle$), parallel DNA molecules ($i = 1, 2$) surrounded by electrolyte solution can be approximated by[15,30]

$$E_{pair} \approx E^{cyl}_{pair}(\kappa_D, R) + E^{hel}_{1,2}(\kappa_D, R, \{\delta\phi(z)\}) \qquad (A1)$$

where κ_D is the reciprocal Debye screening length in the electrolyte solution; R is the interaxial distance between the molecules; $\delta\phi(z) = \phi_1(z) - \phi_2(z)$; $E^{cyl}_{pair}(\kappa, R)$ is the energy of interaction between two homogeneously charged, dielectric cylinders with the same average surface charge density as DNA; $E^{hel}_{i,j}(\kappa, R, \{\delta\phi(z)\})$ is the contribution to the energy of interaction between molecules i and j associated with helical (inhomogeneous) distribution of their surface charges. Here and thereafter, we use curly parentheses around $\delta\phi(z)$ to indicate that the corresponding energy depends on $\delta\phi(z)$ as a functional (see eq A3).

The expressions for the cylindrical, $E^{cyl}_{pair}(\kappa,R)$, and helical, $E^{hel}_{i,j}(\kappa, R, \{\delta\phi(z)\})$, contributions to the interaction energy were derived in refs 15 and 30. In particular

$$E^{cyl}_{pair}(\kappa, R) = \frac{8\pi^2\sigma^2(1-\theta)^2L}{\epsilon\kappa^2[K_1(\kappa r)]^2}\left[K_0(\kappa R) + \sum_{m=-\infty}^{\infty}\frac{(K_m(\kappa R))^2 I'_m(\kappa r)}{-K'_m(\kappa r)}\right] \qquad (A2)$$

where r is the radius of DNA, σ is the surface charge density of phosphates; θ is the fraction of phosphate charge σ neutralized by bound counterions, ϵ is the dielectric constant of water, and $I_n(x)$, $K_n(x)$, $I'_n(x)$, and $K'_n(x)$ are the modified Bessel functions and their derivatives, respectively. Note that the term with the sum adds the contribution of the image-charge repulsion associated with low-dielectric-constant molecular cores to the well-known formula[52] for interaction between two homogeneously charged cylinders.[15,53]

The helical contribution to the interaction energy is given by[15,16,30]

$$E^{hel}_{i,j}(\kappa, R, \{\delta\phi(z)\}) = \alpha(\kappa, R)L - a_1(\kappa, R)\int_0^L \cos[\delta\phi(z)]\,dz + a_2(\kappa, R)\int_0^L \cos[2\delta\phi(z)]\,dz \qquad (A3)$$

where

$$\alpha(\kappa, R) = \frac{16\pi^2\sigma^2}{\epsilon}\sum_{n=1}^{2}\frac{[f(n,\theta)]^2}{\kappa_n{}^2[K_n'(\kappa_n r)]^2}\sum_{m=-\infty}^{\infty}\frac{(K_{n-m}(\kappa_n R))^2 I_m'(\kappa_n r)}{-K_m'(\kappa_n r)} \quad \text{(A4)}$$

is the sum of the first two ($n = 1, 2$) helical harmonics of image-charge repulsion between surface charges on one molecule and the dielectric core of the opposing molecule

$$a_{n=1,2}(\kappa, R) = \frac{16\pi^2\sigma^2}{\epsilon}\frac{[f(n,\theta)]^2}{\kappa_n{}^2}\frac{K_0(\kappa_n R)}{[K_n'(\kappa_n r)]^2} \quad \text{(A5)}$$

are the first ($n = 1$) and second ($n = 2$) helical harmonics of direct charge−charge interactions on opposing molecules, and

$$\kappa_n = \sqrt{\kappa^2 + n^2\left(\frac{2\pi}{H}\right)^2} \quad \text{(A6)}$$

Note that the contribution of higher helical harmonics can be neglected and that only several terms with small $|m|$ contribute significantly into the sum over m in eq A4.[15]

The interaction energy depends on the pattern of fixed phosphates and on the distribution of bound counterions through the coefficients

$$f(n,\theta) = e^{-(\pi^2 n^2 w^2/2H^2)}[f_1\theta + f_2(-1)^n\theta - (1 - f_3\theta)\cos(n\tilde{\phi}_s)] \quad \text{(A7)}$$

where $\tilde{\phi}_s$ ($\approx 0.4\pi$ for B-DNA) is the azimuthal half-width of the narrow arch separating the phosphates (Figure 1), f_i are the fractions of counterions bound in the minor groove (f_1), in the major groove (f_2), and on the phosphate strands (f_3), $f_1 + f_2 + f_3 = 1$; and w is the width of charge density distribution across each phosphate strand or groove.[54]

Derivation of eqs A1−A7 involves several assumptions and approximations described in detail in refs 15−17, 25. Briefly, the most important assumptions and approximations are the following: (a) The length of the molecules is much larger than the helical pitch (for B-DNA this means that $L \gg 10h$). (b) Local deviation of molecular conformation from an ideal helix is small (see section II.1). (c) Molecular cores (shaded cylinder of the radius r in Figure 1) exclude electrolyte solution and have much lower dielectric constant than water. (d) Counterions providing nonlinear electrostatic screening can be formally treated as adsorbed (for B-DNA this means that its average surface charge density is $\sigma(1 - \theta)$ where $\theta \geq 0.7$). (e) The electrostatic potential outside the layer of adsorbed counterions does not exceed k_BT/q_e, where k_B is the Boltzmann constant, T is the temperature, and q_e is the elementary charge, and its variation can be described within the linear Debye−Huckel theory.

A1.2. Electrostatic Interaction between DNA in Hexagonal Aggregates. At small interaxial spacings, the average electrostatic potential inside the bulk of a multimolecular DNA aggregate might exceed k_BT/q_e necessitating a different approach to calculation of the electrostatic energy. Such an approach was recently proposed in ref 48. The main ideas and the result of this calculation are as follows.

Average Electrostatic Potential. The first simplifying factor is that the variation of the potential in the interstitial space between the molecules generally remains smaller than k_BT/q_e at all interaxial spacings. Therefore, the equations for the potential can be linearized near the average value of the potential in the aggregate, but this average potential has to be calculated within a nonlinear theory. The second simplifying factor is that the average potential is determined by the average surface charge density of DNA, $\sigma(1 - \theta)$, rather than by the helical nature of the charge pattern. Therefore, it can be approximated by the potential created by homogeneously charged cylinders with the surface charge density $\sigma(1 - \theta)$ and calculated within the cylindrical cell approximation as previously described.[48]

Briefly, we calculate the average potential Ψ_s at the surface of the Wigner−Seitz cell surrounding each molecule. We approximate it by the potential at the outer surface of the cylindrical cell with the same volume. The radius of this cylindrical cell is, therefore, equal to

$$R_s = R\sqrt{\frac{\sqrt{3}}{2\pi}} \quad \text{(A8)}$$

where R is the interaxial distance between DNA in the aggregate. To calculate Ψ_s, we assume that the potential is created by a cylinder with the radius r and homogeneous surface charge density $\sigma(1 - \theta)$ located in the center of the cell and that the electric field at the outer surface of the cell is zero. The variation of the potential within the cell is smaller than k_BT/q_e and is described by the linearized Poisson−Boltzmann equation, but with renormalized reciprocal screening length

$$\kappa_\Psi = \kappa_D\sqrt{\cosh(q_e\Psi_s/k_BT)} \quad \text{(A9)}$$

The solution for the potential, which satisfies appropriate boundary conditions at the surface of the charged cylinder and at the surface of the cell, yields[48]

$$\tanh\left(\frac{q_e\Psi_s}{k_BT}\right) = \frac{2\xi(1-\theta)[I_0(\kappa_\Psi R_s)K_1(\kappa_\Psi R_s) + I_1(\kappa_\Psi R_s)K_0(\kappa_\Psi R_s)]}{\kappa_\Psi r[I_1(\kappa_\Psi R_s)K_1(\kappa_\Psi r) - I_1(\kappa_\Psi r)K_1(\kappa_\Psi R_s)]} \quad \text{(A10)}$$

where $\xi \approx 4$ is the ratio of the Bjerrum length (~7 Å in water) and the axial length per elementary charge on "naked" DNA (~1.7 Å). This procedure predicts virtually the same Ψ_s and the same radial distribution of the potential as the full numerical solution of the nonlinear Poisson−Boltzmann equation.[48]

Electrostatic Interaction Energy. To calculate the electrostatic interaction energy, we utilize that the contribution of the helical nature of DNA surface charge pattern to the variation of the interstitial electrostatic potential inside the aggregate is also smaller than k_BT/q_e. Then, as argued in ref 48, the electrostatic energy of the aggregate can be approximated as

$$E_{aggr} \approx E_{aggr}^{cyl}(\kappa_\Psi, R) + \frac{1}{2}\sum_{i,j=1}^{N}{}^*E_{i,j}^{hel}(\kappa_\Psi, R, \{\delta\phi_{i,j}(z)\}) \quad \text{(A11)}$$

Here

$$E_{aggr}^{cyl} \approx 2\sqrt{3}LNn_0k_BT\int_R^\infty\left[\cosh\left(\frac{q_e\Psi_s(R')}{k_BT}\right) - 1\right]R'dR' \quad \text{(A12)}$$

is the energy of interaction between homogeneously charged cylinders whereas $E_{i,j}^{hel}(\kappa_\Psi, R, \{\delta\phi_{i,j}(z)\})$ are helical contributions to pair interaction energies between molecules i and j defined by eqs A3−A7, $\delta\phi_{i,j}(z) = \phi_i(z) - \phi_j(z)$. Σ^* denotes summation which includes only nearest neighbor pairs of molecules i and j ($i \neq j$), N is the number of molecules in the

aggregate, and n_0 is the molar concentration of 1:1 electrolyte in the surrounding bulk solution.

A1.3. Free Energy Functionals. Thus, as we pointed out above, both the elastic, torsional energy and the electrostatic interaction energy of DNA depend on conformation and alignment of each molecule i described by $\phi_i(z)$. Therefore, pairing of DNA and formation of aggregates must be accompanied by torsional deformation which minimizes the sum of torsional and electrostatic energies. The corresponding free energy functionals are

$$F_{\text{pair}} = E^{\text{cyl}}_{\text{pair}}(\kappa_D, R) + E^{\text{hel}}_{1,2}(\kappa_D, R, \{\delta\phi(z)\}) + \sum_{i=1}^{2} E_i^{\text{tors}}(\{\Omega_i(z)\}, \{\phi_i(z)\}) \quad \text{(A13)}$$

for DNA pairs and

$$F_{\text{aggr}} = E^{\text{cyl}}_{\text{aggr}}(\kappa_\Psi, R) + \frac{1}{2}\sum_{i,j=1}^{N} {}^{*}E^{\text{hel}}_{i,j}(\kappa_\Psi, R, \{\delta\phi_{i,j}(z)\}) + \sum_{i=1}^{N} E_i^{\text{tors}}(\{\Omega_i(z)\}, \{\phi_i(z)\}) \quad \text{(A14)}$$

for aggregates. These equations reduce the problem of finding the torsional deformation of DNA to minimization of F_{pair} or F_{aggr} with respect to possible realizations of $\phi_i(z)$.

A2. Pair Interaction between "Soft" Helices. Consider interaction between two long, parallel molecules. The free energy functional defined by eqs A1−A7 can be rewritten in the following form

$$F_{\text{pair}} \approx F^{\text{cyl}}_{\text{pair}}(R) + \int_0^L dz[\alpha(\kappa, R) - a_1(\kappa, R)\cos(\delta\phi(z)) + a_2(\kappa, R)\cos(2\delta\phi(z))] + \frac{C}{4}\int_0^L dz\left[\left(\frac{d(\delta\phi(z))}{dz} - \frac{\delta\Omega(z)}{h}\right)^2 + \left(\frac{d(\phi_1(z)+\phi_2(z))}{dz} - \frac{\Omega_1(z)+\Omega_2(z)}{h}\right)^2\right] \quad \text{(A15)}$$

where $\alpha(\kappa, R)$ and $a_{1,2}(\kappa, R)$ are defined by eqs A4 and A5, $\delta\phi(z) = \phi_1(z) - \phi_2(z)$, and $\delta\ \Omega(z) = \Omega_1(z) - \Omega_2(z)$. Minimization of this functional yields the following equations for $\delta\phi(z)$ and $\phi_1(z) + \phi_2(z)$:

$$\frac{d^2(\delta\phi)}{dz^2} - \frac{2a_1(R)}{C}\sin(\delta\phi)\left[1 - \frac{4a_2(R)}{a_1(R)}\cos(\delta\phi)\right] = \frac{1}{h}\frac{d(\delta\Omega)}{dz} \quad \text{(A16)}$$

and

$$\phi_1(z) + \phi_2(z) = \phi_1(0) + \phi_2(0) + \frac{1}{h}\int_0^z (\Omega_1(z') + \Omega_2(z'))\, dz' \quad \text{(A17)}$$

It follows from eqs A15 and A17 that the optimal alignment of $\phi_1(z) + \phi_2(z)$ has no energetic cost. Therefore, we can focus only on the energetic cost associated with $\delta\phi(z)$.

Note that eq A16 is reminiscent of the time-independent sine-Gordon equation in the external field $h^{-1}[d(\delta\Omega)/dz]$. At $\delta\Omega = 0$, this equation has a set of exact analytical solutions, including, e.g., $\delta\phi = 0$, $\delta\phi = \arccos(a_1/4a_2)$, and kink solitons. At large interaxial distances ($a_1 > 4a_2$), $\delta\phi = 0$ has the lowest energy. At smaller distances ($a_1 < 4a_2$), $\delta\phi = \arccos(a_1/4a_2)$ becomes more energetically favorable.[15] Kink solitons have higher free energy and describe excitations in the molecular pair.

The problem becomes dramatically more complex at nonzero $\delta\Omega$ when eq A16 has no general solutions.[55] In the simplest case of two "soft" ($|\delta\phi(z) - \delta\phi(z')| \ll 1$) molecules, eq A16 can be linearized, and its lowest energy solution reads

$$\delta\phi(z) = \delta\phi^0 + \frac{1}{2h}\int_0^L \delta\Omega(z')e^{-(|z-z'|/\lambda_t)}\frac{|z-z'|}{z-z'}dz' \quad \text{(A18)}$$

where $\delta\phi^0 = 0$, $\lambda_t = \sqrt{C/[2a_1(R) - 8a_2(R)]}$ at $a_1 > 4a_2$ and $\delta\phi^0 = \arccos[a_1(R)/4a_2(R)]$, $\lambda_t = \sqrt{C/(8a_2(R) - a_1^2(R)/[2a_2(R)])}$ at $a_1 < 4a_2$. However, calculation of $\langle(\delta\phi - \delta\phi^0)^2\rangle$ described below shows that this approximation can be used only at small R, when the interaction is substantially stronger than torsional rigidity and $\langle(\delta\phi - \delta\phi^0)^2\rangle$ becomes small (Figure 3b).

A3. Methods and Calculations. To obtain a more accurate approximation for optimal $\phi_i(z)$ within a wider range of parameters, we use a different approach similar to variational methods in quantum mechanics and which is described below. This variational approximation accounts for the important nonlinearity of the dependence of the interaction energy on $\delta\phi$. Our estimates and Monte Carlo simulations[56] indicate that it works well at all relevant interaxial distances R and that it correctly captures the most important features of the torsional deformation and intermolecular interaction. This approximation does not describe soliton-like solutions, but they appear to be important only under special circumstances[56] and their discussion is beyond the scope of this paper.

A3.1. Variational Method. The essence of the variational method is that we look for $\phi_i(z)$ within a set of trial functions $\tilde{\phi}_i(z)$ dependent on a set of variational parameters. We minimize the nonlinear free energy functional (eq A13 or eq. A14) with respect to these variational parameters and approximate $\phi_i(z)$ by optimal $\tilde{\phi}_i(z)$.

Specifically, for two molecules (i,j=1,2; $i \neq j$) we select the trial function set based on generalized eqs A17 and A18

$$\tilde{\phi}_i(z) \approx \phi_i^0 + \frac{\langle\Omega\rangle z}{h} + \frac{1}{2h}\int_0^z (\omega_i(z') + \omega_j(z'))\, dz' + \frac{1}{4h}\int_0^L (\omega_i(z') - \omega_j(z'))e^{-(|z-z'|/\lambda)}\frac{|z-z'|}{z-z'}\, dz' \quad \text{(A19)}$$

where ϕ_i^0 and λ are used as the variational parameters.[57] For hexagonal aggregates we use

$$\tilde{\phi}_i(z) \approx \phi_i^0 + \frac{\langle\Omega\rangle z}{h} + \frac{1}{2h}\int_0^L \omega_i(z')e^{-(|z-z'|/\lambda)}\frac{|z-z'|}{z-z'}dz' \quad \text{(A20)}$$

where the variational parameters are also ϕ_i^0 and λ. The latter is the simplest trial function set compatible with the presence of six nearest neighbors. It gives the same $\tilde{\phi}_i(z) - \tilde{\phi}_j(z)$ as eq A19 so that all electrostatic pair interaction energies in the hexagonal aggregate are optimized with the same accuracy as for two molecules.

Because eq A19 includes eq A18 as one of the trial functions, the variational method should always give a better (or at least equally good) approximation for $\phi_i(z)$. In particular, eq A18 is limited to small twist angle variations ($|\phi_i(z) - \phi_j(z)| \ll 1$) when the expressions for the free energy can be linearized. In this case, the variational method gives exactly the same result. However, this method remains reasonably accurate also when the free energy cannot be linearized, and our estimates indicate that it works well up to $|\phi_i(z) - \phi_j(z)| \sim 1$.

A3.2. Ensemble Averaging and Mean-Field Approximation. To further simplify the problem, we use free energy functionals

which are ensemble-averaged over possible realizations of $\Omega(z)$; i.e., we assume that

$$\int_0^L \cos(n\delta\tilde{\phi}_{i,j}(z))\,\mathrm{d}z \approx \int_0^L \langle\cos(n\delta\tilde{\phi}_{i,j}(z))\rangle_\Omega\,\mathrm{d}z \quad \text{(A21)}$$

and

$$\int_0^L\left(\frac{\mathrm{d}\tilde{\phi}_i(z)}{\mathrm{d}z}-\frac{\Omega_i(z)}{h}\right)^2\mathrm{d}z \approx \int_0^L\left\langle\left(\frac{\mathrm{d}\tilde{\phi}_i(z)}{\mathrm{d}z}-\frac{\Omega_i(z)}{h}\right)^2\right\rangle_\Omega\mathrm{d}z \quad \text{(A22)}$$

where $\langle\ \rangle_\Omega$ denotes ensemble averaging over realizations of Ω at given z.

At infinite juxtaposition length ($L \to \infty$), integration over the juxtaposition length in eqs A13 and A14 is equivalent to ensemble averaging so that this is a rigorous procedure which does not introduce any additional approximations. However, for finite size fragments this is an analogue of a "mean-field" approximation.

Before we proceed to calculation of the free energies, let us determine the mean square deviations of $\tilde{\phi}_i(z)$ from ideal helical conformation, $\phi_i^0 + \langle\Omega\rangle z/h$. From eqs A19 and A20 we find that both for pairs of nonhomologous DNA and for hexagonal aggregates

$$\langle(\delta\tilde{\phi}_{i,j}(z)-\delta\phi_{i,j}{}^0)^2\rangle_\Omega = \frac{1}{2h^2}\int_0^L \mathrm{d}z'\int_0^L \mathrm{d}z''\,\langle\omega_i(z')\omega_i(z'')\rangle_\Omega \mathrm{e}^{-(|z-z'|+|z-z''|)/\lambda}\frac{|z-z'|}{z-z'}\frac{|z-z''|}{z-z''} \quad \text{(A23)}$$

where we used that $\langle\omega_i(z')\omega_i(z'')\rangle_\Omega = \langle\omega_j(z')\omega_j(z'')\rangle_\Omega$ and $\langle\omega_i(z')\omega_j(z'')\rangle_\Omega = 0$ at $i \neq j$). Assuming that λ is much larger than the range of $\omega_i(z')\omega_i(z'')$ correlation (which is not expected to extend beyond several base pairs), we arrive at

$$\langle(\delta\tilde{\phi}_{i,j}(z)-\delta\phi_{i,j}{}^0)^2\rangle_\Omega = \left(\frac{\lambda}{2\lambda_c}\right)\left(1-\mathrm{e}^{-L/\lambda}\cosh\left[\frac{L-2z}{\lambda}\right]\right) \quad \text{(A24)}$$

where the helical coherence length λ_c is defined by eq 7. Similarly, we find that in hexagonal aggregates

$$\left\langle\left(\tilde{\phi}_i(z)-\phi_i{}^0-\frac{\langle\Omega\rangle z}{h}\right)^2\right\rangle_\Omega = \left(\frac{\lambda}{4\lambda_c}\right)\left(1-\mathrm{e}^{-L/\lambda}\cosh\left[\frac{L-2z}{\lambda}\right]\right) \quad \text{(A25)}$$

Note that these mean square deviations explicitly depend on the position z along the juxtaposition at L comparable to or smaller than λ. This dependence is a "finite size effect" in the alignment of such helical fragments. Therefore, upon averaging of functions or functionals of $\phi(z)$ for finite size DNA fragments, it is important to distinguish an ensemble average over possible realizations of $\Omega(z)$, an average over the juxtaposition length, and a combined average over realizations of $\Omega(z)$ and over the juxtaposition length.

Averaging of eqs A24 and A25 over the juxtaposition length L yields eqs 8 and 9 correspondingly.

A3.3. Electrostatic Energy. To calculate $\langle\cos(n\delta\tilde{\phi}_{i,j}(z))\rangle_\Omega$ in eq A21, we take into account that $\tilde{\phi}_i(z)$ is described by Gaussian probability distribution regardless of the probability distribution for different values of $\omega_i(z)$, as long as λ is larger than the range of $\langle\omega_i(z)\omega_i(z')\rangle$ correlation. This conclusion is based on the application of the central limit theorem for sufficiently long molecules or sufficiently large ensembles.[58] Using the Gaussian probability distribution for $\delta\tilde{\phi}_{i,j}(z)$, we find

$$\langle\cos(n\delta\tilde{\phi}_{i,j}(z))\rangle_\Omega = \cos(n\delta\phi_{i,j}{}^0)\exp\left(-\frac{n^2}{2}\langle(\delta\tilde{\phi}_{i,j}(z)-\delta\phi_{i,j}{}^0)^2\rangle_\Omega\right) \quad \text{(A26)}$$

After substitution of eq A24 into eq A26 and then into eq A21, we arrive at

$$\int_0^L \cos(n\delta\tilde{\phi}_{i,j}(z))\,\mathrm{d}z \approx L\exp\left(-\frac{n^2\lambda}{4\lambda_c}\right)\cos(n\delta\phi_{i,j}{}^0)S\left(\frac{n^2\lambda}{4\lambda_c},\frac{L}{\lambda}\right) \quad \text{(A27)}$$

where

$$S(x,y) = \int_0^1 \exp[x\,\mathrm{e}^{-y}\cosh(ty)]\,\mathrm{d}t \quad \text{(A28)}$$

Within the range of applicability of the variational approximation (up to $x \sim 1$), eq A28 can be approximated by a simple interpolation formula

$$S(x,y) \approx \exp\left(x\left[\frac{1-\exp(-2y)}{2y}\right]\right) \quad \text{(A29)}$$

which works with better than 4% accuracy at all y (this formula becomes exact at all y when $x \to 0$ and at all x when $y \to 0$ or $y \to \infty$). Therefore

$$\int_0^L \cos(n\delta\tilde{\phi}_{i,j}(z))\,\mathrm{d}z \approx L\exp\left(-\frac{n^2\lambda}{4\lambda_c}G\left(\frac{2L}{\lambda}\right)\right)\cos(n\delta\phi_{i,j}{}^0) \quad \text{(A30)}$$

where

$$G(x) = 1-\frac{1-\mathrm{e}^{-x}}{x} \quad \text{(A31)}$$

and

$$\frac{E_{i,j}^{\mathrm{hel}}(\kappa,R,\{\delta\tilde{\phi}(z)\})}{L} = \alpha(\kappa,R) - a_1(\kappa,R)\mathrm{e}^{-(\lambda/4\lambda_c)G(2L/\lambda)}\cos(\delta\phi_{i,j}{}^0) + a_2(\kappa,R)\mathrm{e}^{-(\lambda/\lambda_c)G(2L/\lambda)}\cos(2\delta\phi_{i,j}{}^0) \quad \text{(A32)}$$

As follows from its derivation, eq A32 is valid for both trial function sets defined by eq A19 and eq A20 as long as $L \gg 10h$. Since the pitch of DNA double helix is $10h$ and eqs A1−A7 were derived for molecules much longer than the pitch, this derivation does not introduce any additional restrictions on possible values of the parameters.

A3.4. Torsional Energy. After similar calculation of the integrals associated with the torsional energy, we find that for two molecules ($\tilde{\phi}_i(z)$ defined by eq A19)

$$\frac{E_1^{\mathrm{tors}}+E_2^{\mathrm{tors}}}{L} \approx \frac{C}{8\lambda\lambda_c}G\left(\frac{2L}{\lambda}\right) \quad \text{(A33)}$$

while for hexagonal aggregates ($\tilde{\phi}_i(z)$ defined by eq A20)

$$\frac{E_i^{\mathrm{tors}}}{L} \approx \frac{C}{8\lambda\lambda_c}G\left(\frac{2L}{\lambda}\right) \quad \text{(A34)}$$

Here the two trial function sets give the results which differ by a factor of 2 because eq A19 optimizes both $\tilde{\phi}_i(z) - \tilde{\phi}_j(z)$ and $\tilde{\phi}_i(z) + \tilde{\phi}_j(z)$ while eq A20 optimizes only $\tilde{\phi}_i(z) - \tilde{\phi}_j(z)$. (Simultaneous optimization of both $\tilde{\phi}_i(z) - \tilde{\phi}_j(z)$ and

$\tilde{\phi}_i(z) + \tilde{\phi}_j(z)$ for all nearest neighbor pairs in a hexagonal aggregate is impossible.)

A3.5. Minimization of Free Energy Functionals for Two DNA Molecules in Juxtaposition. After substitution of eqs A32 and A33 into eq A13, we find that the free energy functional for two nonhomologous DNA molecules is given by

$$\frac{F_{\text{pair}}}{L} \approx \frac{E_{\text{pair}}^{\text{cyl}}(R)}{L} + \alpha(R) - A_1(\lambda,R)\cos(\delta\phi^0) + A_2(\lambda,R)\cos(2\delta\phi^0) + \frac{C}{8\lambda\lambda_c}G\left(\frac{2L}{\lambda}\right) \quad \text{(A35)}$$

where

$$A_1(\lambda, R) = a_1(R)e^{-(\lambda/4\lambda_c)G(2L/\lambda)},\ A_2(\lambda, R) = a_2(R)e^{-(\lambda/\lambda_c)G(2L/\lambda)} \quad \text{(A36)}$$

and

$$\delta\phi^0 = \phi_1{}^0 - \phi_2{}^0$$

Minimization of this functional with respect to $\delta\phi^0$ and λ yields the following solutions. At $A_1(\lambda,R) > 4A_2(\lambda,R)$, the optimal $\delta\phi^0$ and λ are given by

$$\delta\phi^0(R) = 0 \text{ and } \lambda = \sqrt{\frac{C\tilde{G}(L/\lambda)}{2[A_1(\lambda, R) - 4A_2(\lambda, R)]}} \quad \text{(A37)}$$

where

$$\tilde{G}(x) = \frac{x\tanh(x)}{x - \tanh(x)} \quad \text{(A38)}$$

At $A_1(\lambda, R) < 4A_2(\lambda, R)$, we find

$$\delta\phi^0(R) = \pm\arccos\left[\frac{A_1(\lambda, R)}{4A_2(\lambda, R)}\right] \text{ and } \lambda = \sqrt{\frac{2CA_2(\lambda, R)\tilde{G}(L/\lambda)}{16A_2{}^2(\lambda, R) - A_1{}^2(\lambda, R)}} \quad \text{(A39)}$$

Here $\lambda(R, L)$ is defined through the corresponding transcendental equations, which can be easily solved numerically.

The first solution exists at $R \geq R_1$, where R_1 is the root of $a_1(R)/a_2(R) = 4e^{-3L/2\lambda_c}$. The second solution exists at $R \leq R_2$, where $R_2 > R_1$ and the rather cumbersome set of equations for R_2 can be obtained by substitution of eq A39 into $\delta^2F_{\text{pair}}/\delta\lambda^2 = 0$ (F_{pair} loses the minimum defined by eq A39 at the point where $\delta F_{\text{pair}}/\delta\lambda = 0$ and $\delta^2F_{\text{pair}}/\delta\lambda^2 = 0$). The two solutions coexist at $R_1 < R < R_2$, when both $A_1(\lambda, R) > 4A_2(\lambda, R)$ and $A_1(\lambda, R) < 4A_2(\lambda, R)$ are possible depending on the value of λ. From numerical solution of eqs A37 and A39 within this range, we find that the first solution is more energetically favorable at larger R while the second solution is more energetically favorable at smaller R. The switch from the first to the second solution occurs as a first-order transition at the point where the energies of the two solutions become equal, as illustrated in Figure 3.

A3.6. Minimization of Free Energy Functionals for Hexagonal Aggregates. Because a hexagonal aggregate consists of equilateral triangles formed by nearest neighbor molecules, the pattern of molecular orientations illustrated in Figure 5a gives the lowest interaction free energy. Indeed, in this pattern all triangles are identical. Thus, the azimuthal orientations $\phi_1{}^0$, $\phi_2{}^0$, and $\phi_3{}^0$ minimizing the free energy of any given triangle minimize the free energy of the whole ensemble.

For this alignment, the aggregation free energy per molecule, $F_{\text{DNA}}(R) = \mu_{\text{DNA}}(R) - \mu_{\text{DNA}}(\infty)$, where μ_{DNA} is the chemical potential of DNA, is given by substitution of eqs A32 and A34 into eq A14

$$\frac{F_{\text{DNA}}}{L} \approx \frac{F_{\text{aggr}}^{\text{cyl}}(R)}{L} + 3a_0^{\text{hel}}(R) + \frac{C}{8\lambda\lambda_c}G\left(\frac{2L}{\lambda}\right) - A_1(\lambda, R)\{\cos(\phi_1{}^0 - \phi_2{}^0) + \cos(\phi_2{}^0 - \phi_3{}^0) + \cos(\phi_1{}^0 - \phi_3{}^0)\} + A_2(\lambda, R)\{\cos[2(\phi_1{}^0 - \phi_2{}^0)] + \cos[2(\phi_2{}^0 - \phi_3{}^0)] + \cos[2(\phi_1{}^0 - \phi_3{}^0)]\} \quad \text{(A40)}$$

Similar to the case of two molecules, minimization of this energy with respect to $\phi_1{}^0$, $\phi_2{}^0$, $\phi_3{}^0$, and λ yields two different solutions at $A_1(\lambda, R) > 4A_2(\lambda, R)$ and $A_1(\lambda, R) < 4A_2(\lambda, R)$. One solution is more energetically favorable at larger and the other at smaller R. Also, as in the case of two molecules, the switch between these two solutions upon changing R occurs as a first-order transition.

Specifically, from minimization with respect to $\phi_1{}^0$, $\phi_2{}^0$, and $\phi_3{}^0$ we find that at $A_1(\lambda, R) > 4A_2(\lambda, R)$

$$\phi_1{}^0 = \phi_2{}^0 = \phi_3{}^0 \quad \text{(A41)}$$

and at $A_1(\lambda, R) < 4A_2(\lambda, R)$

$$\phi_2{}^0 - \phi_1{}^0 = \phi_1{}^0 - \phi_3{}^0 = \pm\psi_0(R) \quad \text{(A42)}$$

where

$$\cos(\psi_0) = \frac{1}{4}\left(1 + \sqrt{1 + \frac{2A_1(\lambda, R)}{A_2(\lambda, R)}}\right) \quad \text{(A43)}$$

and $\phi_1{}^0$ can have any value.[59]

After substitution of eq A41 into eq A40, we find that at $A_1(\lambda, R) > 4A_2(\lambda, R)$

$$\frac{F_{\text{DNA}}}{L} \approx \frac{F^{\text{cyl}}}{L} + 3a_0^{\text{hel}}(R) - 3A_1(\lambda, R) + 3A_2(\lambda, R) + \frac{C}{8\lambda\lambda_c}G\left(\frac{2L}{\lambda}\right) \quad \text{(A44)}$$

and optimal λ is the root of the following equation

$$\lambda = \sqrt{\frac{C\tilde{G}(L/\lambda)}{6[A_1(\lambda, R) - 4A_2(\lambda, R)]}} \quad \text{(A45)}$$

Similarly, at $A_1(\lambda, R) < 4A_2(\lambda, R)$

$$\frac{F_{\text{DNA}}}{L} \approx \frac{F^{\text{cyl}}}{L} + \frac{A_1(\lambda, R)}{4} - \frac{A_2(\lambda, R)}{4}\left\{5 + \left[1 + \frac{2A_1(\lambda, R)}{A_2(\lambda, R)}\right]^{3/2}\right\} - \frac{A_1{}^2(\lambda, R)}{8A_2(\lambda, R)} + \frac{C}{8\lambda\lambda_c}G\left(\frac{2L}{\lambda}\right) \quad \text{(A46)}$$

and λ is the root of

$$\lambda = \sqrt{C\tilde{G}(L/\lambda)/\left\{2A_2(\lambda, R)\left[5 + \left(1 - \frac{A_1(\lambda, R)}{4A_2(\lambda, R)}\right)\left(1 + \frac{2A_1(\lambda, R)}{A_2(\lambda, R)}\right)^{1/2}\right] - \frac{A_1(\lambda, R)}{2}\left(1 + \frac{A_1(\lambda, R)}{A_2(\lambda, R)}\right)\right\}} \quad \text{(A47)}$$

References and Notes

(1) Gelbart, W. M.; Bruinsma, R. F.; Pincus, P. A.; Parsegian, V. A. *Phys. Today* **2000**, *53*, 38.

(2) Strey, H. H.; Podgornik, R.; Rau, D. C.; Parsegian, V. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 309.

(3) Bloomfield, V. A. *Curr. Opin. Struct. Biol.* **1996**, *6*, 334.

(4) Langridge, R.; Wilson, H. R.; Hooper, C. W.; Wilkins, M. H. F.; Hamilton, L. D. *J. Mol. Biol.* **1960**, *2*, 19.

(5) Livolant, F. *Physica A* **1991**, *176*, 117.

(6) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic acids: structures, properties, and functions*; University Science Books: Sausalito, CA, 2000.

(7) Oosawa, F. *Polyelectrolytes*; Marcel Dekker: New York, 1971.

(8) Stigter, D. *Biopolymers* **1977**, *16*, 1435.

(9) Manning, G. S. *Q. Rev. Biophys.* **1978**, *11*, 179.

(10) Frank-Kamenetskii, M. D.; Anshelevich, V. V.; Lukashin, A. V. *Sov. Phys. Usp.* **1987**, *30*, 317.

(11) Knoll, D. A.; Fried, M. G.; Bloomfield, V. A. Heat-induced DNA aggregation in the presence of divalent metal salts. In *DNA and its drug complexes*; Sarma, M. H., Sarma, R. H., Eds.; Adenine Press: Schenectady, NY, 1988.

(12) Rau, D. C.; Parsegian, V. A. *Biophys. J.* **1992**, *61*, 260.

(13) Rhodes, D.; Klug, A. *Nature* **1980**, *286*, 573.

(14) Saenger, W. *Principles of nucleic acid structure*; Springer-Verlag: New York, 1984.

(15) Kornyshev, A. A.; Leikin, S. *J. Chem. Phys.* **1997**, *107*, 3656; **1998**, *108*, 7035 (Erratum).

(16) Kornyshev, A. A.; Leikin, S. *Phys. Rev. Lett.* **1999**, *82*, 4138.

(17) Kornyshev, A. A.; Leikin, S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13579.

(18) Wang, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 200.

(19) Zimmerman, S. B.; Pheiffer, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 2703.

(20) Kornyshev, A. A.; Leikin, S. *Biophys. J.* **1998**, *75*, 2513.

(21) Van Winkle, D. H.; Davidson, M. W.; Chen, W. X.; Rill, R. L. *Macromolecules* **1990**, *23*, 4140.

(22) Leonard, M.; Hong, H.; Easwar, N.; Strey, H. H. *Polymer* **2001**, *42*, 5823.

(23) Podgornik, R.; Strey, H. H.; Gawrisch, K.; Rau, D. C.; Rupprecht, A.; Parsegian, V. A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4261.

(24) Strey, H. H.; Wang, J.; Podgornik, R.; Rupprecht, A.; Yu, L.; Parsegian, V. A.; Sirota, E. B. *Phys. Rev. Lett.* **2000**, *84*, 3105.

(25) Kornyshev, A. A.; Leikin, S. *Phys. Rev. E* **2000**, *62*, 2576.

(26) Lorman, V.; Podgornik, R.; Zeks, B. *Phys. Rev. Lett.* **2001**, *8721*, 218101.

(27) Kornyshev, A. A.; Leikin, S.; Malinin, S. V. *Eur. Phys. J. E* **2002**, *7*, 83.

(28) Harreis, H. M.; Kornyshev, A. A.; Likos, C. N.; Lowen, H.; Sutmann, G. *Phys. Rev. Lett.* **2002**, *89*, 018303.

(29) Harreis, H. M.; Likos, C. N.; Lowen, H. *Biophys. J.* **2003**, *84*, 3607.

(30) Kornyshev, A. A.; Leikin, S. *Phys. Rev. Lett.* **2001**, *86*, 3666.

(31) Weiner, B. M.; Kleckner, N. *Cell* **1994**, *77*, 977.

(32) Burgess, S. M.; Kleckner, N.; Weiner, B. M. *Genes Dev.* **1999**, *13*, 1627.

(33) Kabsch, W.; Sander, C.; Trifonov, E. N. *Nucl. Acid Res.* **1982**, *10*, 1097.

(34) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. *J. Mol. Biol.* **1995**, *247*, 34.

(35) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163.

(36) Crothers, D. M.; Drak, J.; Kahn, J. D.; Levene, S. D. DNA bending, flexibility and helical repeat by cyclization kinetics. In *Methods in enzymology*; Lilley, D. M. J., Dahlberg, J. E., Eds.; Academic Press: San Diego, CA, 1992; Vol. 212 B; p 3.

(37) In principle, strong electrostatic interaction at close separations may not only result in torsional deformation but also lead to changes in stacking distance and roll and tilt angles between base pairs. Intermolecular interaction may also cause more global conformational changes such as B−A transition, which occurs in dense DNA aggregates at low humidity. In the present study, however, we focus on recognition and formation of sufficiently well hydrated DNA aggregates where only the twist deformation appears to be important.

(38) Upon averaging of functions or functionals of $\phi(z)$, it is important to distinguish an ensemble average over possible realizations of $\Omega(z)$, an average over the juxtaposition length, and a combined average over realizations of $\Omega(z)$ and over the juxtaposition length (for more details see section A3.2).

(39) The helical coherence of DNA can be evaluated from the mean square deviation of $\phi(z) - \phi(z')$ from the value expected for an ideal helix, $\langle\Omega\rangle(z - z')/h$, i.e., $\langle[\phi(z) - \phi(z') - \langle\Omega\rangle(z - z')/h]^2\rangle = |z - z'|/\lambda_c$.

(40) For 100−200 bp fragments, the expected difference in the aggregation energy of homologous and nonhomologous DNA is $\sim 1-2\ k_BT$.

(41) Dickerson, R. E. *Methods Enzymol.* **1992**, *211*, 67.

(42) Pelta, J., Jr.; Durand, D.; Doucet, J.; Livolant, F. *Biophys. J.* **1996**, *71*, 48.

(43) Rau, D. C.; Parsegian, V. A. *Biophys. J.* **1992**, *61*, 246.

(44) Sitko, J. C.; Mateescu, E. M.; Hansma, H. G. *Biophys. J.* **2003**, *84*, 419.

(45) Schurr, J. M.; Delrow, J. J.; Fujimoto, B. S.; Benight, A. S. *Biopolymers* **1997**, *44*, 283.

(46) Eichhorn, G. L.; Shin, Y. A. *J. Am. Chem. Soc.* **1968**, *90*, 7323.

(47) Duguid, J. G.; Bloomfield, V. A.; Benevides, J. M.; Thomas, G. J., Jr. *Biophys. J.* **1995**, *69*, 2623.

(48) Cherstvy, A. G.; Kornyshev, A. A.; Leikin, S. *J. Phys. Chem. B* **2002**, *106*, 13362.

(49) Singer, B. S.; Gold, L.; Gauss, P.; Doherty, D. H. *Cell* **1982**, *31*, 25.

(50) Rubnitz, J.; Subramani, S. *Mol. Cell Biol.* **1984**, *4*, 2253.

(51) Watt, V. M.; Ingles, C. J.; Urdea, M. S.; Rutter, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4768.

(52) Brenner, S. L.; Parsegian, V. A. *Biophys. J.* **1974**, *14*, 327.

(53) The value of the dielectric constant of the molecular cores ϵ_c does not enter the expressions for the image charge repulsion because it is much smaller than the dielectric constant of water ϵ and the strength of the repulsion between a charge and an interface depends on ϵ_c as $(\epsilon - \epsilon_c)/(\epsilon + \epsilon_c) \sim 1$.

(54) Here we account for charge "smearing" caused by finite width and thermal motions of charged groups (in contrast to the approximation of infinitely thin charged lines for strands and grooves used in our previous studies). For simplicity, we assume that the smearing can be described by a Gaussian function with the width w independent of charge group identity and location. Such smearing suppresses the contribution of the interaction mode with $n = 2$, as compared to the case of $w = 0$.

(55) To the best of our knowledge, exact solutions of the time-independent sine-Gordon equation with an inhomogeneous righ-hand-side were obtained only for few pointlike defects in the context of soliton pinning (see, e.g.: Reisinger, H.; Schwalb, F. *Z. Phys. B* **1983**, *52*, 151). The derivative of $\delta\Omega$ in the right-hand-side further increases the complexity of the problem by making it nonlocal.

(56) Kornyshev, A. A.; Wynveen, A. *Phys. Rev. E* **2004**, in press.

(57) In principle, an even more accurate approximation would be based on using an independent variational parameter as a factor in front of the second integral. However, the resulting expanded set of trial functions gives exactly the same result for infinitely long helices and only minor corrections for finite-length helical fragments.

(58) At $\lambda \ll L$ and infinite L, the Gaussian probability distribution for $\phi_i(z)$ follows from the central limit theorem. The accuracy of the description of $\phi_i(z)$ by Gaussian statistics at finite λ, $L \gg h$, might depend on the type and range of $\omega_i(z)\omega_i(z')$ correlations. However, numerical simulations based on experimental values of $\omega_i(z)$ showed that the Gaussian approximation works extremely well at all relevant values of the parameters, as long as the ensemble of different realizations of $\omega_i(z)$ is sufficiently large (either due to large molecular length or due to large number of molecules with different, unrelated $\omega_i(z)$).

(59) Note that mutual alignment of each pair of nearest neighbours deviates from the optimal one for pair interaction within the region of nonzero ϕ_i^0 (cf. Figure 3a,b and Figure 5b). This frustration of pairwise alignment is caused by multimolecular correlations. It leads to a rather shallow energy minimum at the optimal alignment, and therefore, it might result in substantial thermal torsional fluctuations. Such fluctuations reduce biaxial correlations and might cause complete loss of the long-range order in azimuthal orientations. Although this effect might be important for understanding the loss of cholesteric packing of molecules at small interaxial distances, its detailed analysis is beyond the scope of the present work.