DNA Double Helices Recognize Mutual Sequence Homology in a Protein Free Environment

Geoff S. Baldwin,*,† Nicholas J. Brooks,‡ Rebecca E. Robson,†,‡ Aaron Wynveen,‡ Arach Goldar,‡,§ Sergey Leikin,*,‖ John M. Seddon,*,‡ and Alexei A. Kornyshev*,‡

Division of Molecular Biosciences, Imperial College London, SW7 2AZ London, U.K., Department of Chemistry, Imperial College London, SW7 2AZ London, U.K., and Section on Physical Biochemistry, National Institute of Child Health and Human Development, National Institutes of Health, DHHS, Bethesda, Maryland 20892

Received: November 27, 2007

The structure and biological function of the DNA double helix are based on interactions recognizing sequence complementarity between two single strands of DNA. A single DNA strand can also recognize the double helix sequence by binding in its groove and forming a triplex. We now find that sequence recognition occurs between intact DNA duplexes without any single-stranded elements as well. We have imaged a mixture of two fluorescently tagged, double helical DNA molecules that have identical nucleotide composition and length (50% GC; 294 base pairs) but different sequences. In electrolytic solution at minor osmotic stress, these DNAs form discrete liquid-crystalline aggregates (spherulites). We have observed spontaneous segregation of the two kinds of DNA within each spherulite, which reveals that nucleotide sequence recognition occurs between double helices separated by water in the absence of proteins, consistent with our earlier theoretical hypothesis. We thus report experimental evidence and discuss possible mechanisms for the recognition of homologous DNAs from a distance.

Introduction

Organization, compaction, and processing of genetic material involve direct interactions between DNA double helices at small distances.[1] These interactions have many surprising features even in simple electrolytic solutions without proteins.[2−10] Invariably, they are assumed to be independent of the base pair sequence because the nucleotides are buried inside the double helix and shielded by the charged sugar−phosphate backbone.[7,8] However, a recent theory[10] challenged this concept, predicting that the sequence dependence of the backbone structure[11−13] might affect the DNA−DNA interactions and even result in sequence homology recognition without unzipping the double helix.[14,15]

The hypothesis of homology recognition between intact double helices resulting from the sequence-dependent structure of the double helix remains controversial. For instance, a recent observation of the apparent self-assembly of DNA fragments by gel retardation and atomic force microscopy (AFM) was interpreted in terms of transient cross-hybridization between single-stranded "bubbles" and flipped-out bases.[16] To test the hypothesis of structure-mediated homology recognition, we imaged mixtures of fluorescently tagged DNAs with identical nucleotide composition and length but different sequences. To reduce the probability of cross-hybridization and to eliminate potential contributions from interactions other than between DNA fragments, we investigated interactions between double-stranded DNA with blunt ends within cholesteric spherulites under weak osmotic stress. Here, we report spontaneous segregation of DNAs, revealing recognition between the double helices separated by more than a nanometer of water.

Materials and Methods

DNAs were amplified by polymerase chain reaction (PCR) with Pfu polymerase (prepared in our laboratory) from ΦX174 template DNA using 5′-ACTTTGCGACCTTTCGC-3′ and 5′-AAAACGGCAGAAGCCTG-3′ primers for $^{469}_{176}$DNA and 5′-AACAATCCGTACGTTTCC-3′ and 5′-TAAATTCAGCGC-CTTCC-3′ primers for $^{699}_{406}$DNA (MWG-Biotech). 5′-Amine-modified DNAs were amplified using forward primers of the same sequence but with a 5′-C6 amine. The PCR products were purified using a PCR cleanup kit (Sigma) and verified by UV absorption spectra and gel electrophoresis (Supporting Information Figure S1). Unlabeled DNAs were used without further purification. 5′-Amine-modified PCR products were ethanol precipitated, resuspended (1 mg/mL) in 75 mM tetraborate buffer, pH 8.5 (Sigma), and labeled overnight at room temperature by adding 120 μg of Alexa Fluor (555 or 647, Invitrogen) in 7 μL of anhydrous dimethylformide (Sigma) to 50 μL of DNA solution. The labeled DNAs were loaded on a 4.6 mm TSKgel DNA-NPR ion-exchange column (Tosoh) in 20 mM tris(hydroxymethyl)aminomethane (Tris), pH 9, and eluted with a 0−1 M NaCl gradient in the same buffer. DNA and fluorophore absorbance was monitored from 240 to 700 nm. The

* To whom correspondence should be addressed. E-mail: g.baldwin@imperial.ac.uk (G.S.B.); leikins@mail.nih.gov (S.L.); j.seddon@imperial.ac.uk (J.M.S.); a.kornyshev@imperial.ac.uk (A.A.K.).
† Division of Molecular Biosciences, Imperial College London.
‡ Department of Chemistry, Imperial College London.
§ Current address: Service de Biologie Integrative et de Genetique Moleculaire, CEA/Saclay, 91191 Gif-sur-Yvette Cedex, France.
‖ Section on Physical Biochemistry, NICHD/NIH.

Letters

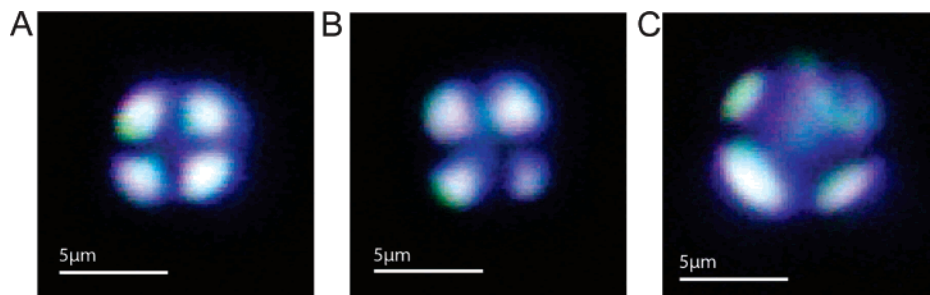*J. Phys. Chem. B, Vol. 112, No. 4, 2008* **1061**

Figure 1. Typical polarizing microscopy images of DNA spherulites show uniform optical textures for $^{469}_{176}$DNA (A) or $^{699}_{406}$DNA (B) but disrupted patterns in 1:1 mixtures of $^{469}_{176}$DNA with $^{699}_{406}$DNA (C). The spherulites were formed from 50 μg/mL DNA in 0.5 M NaCl, 10 mM Tris, pH 7.5, 1 mM EDTA, and 5% PEG and imaged after 2 week equilibration in sealed microscopy slides at room temperature (20 °C).
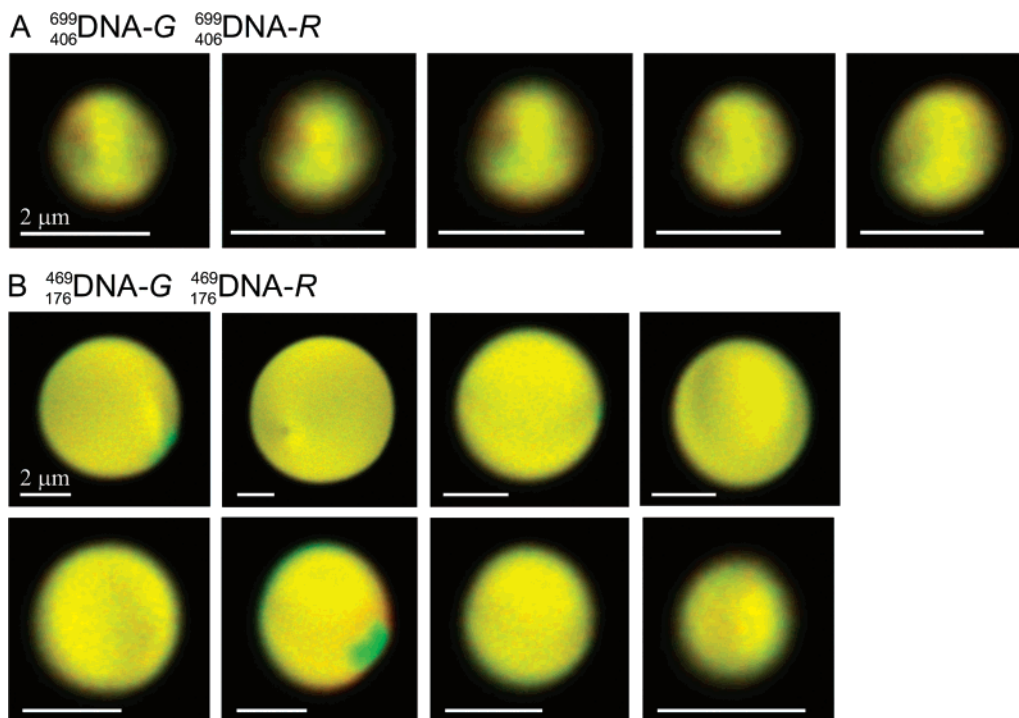

Figure 2. Confocal imaging of single DNA sequences in spherulites demonstrates that fluorescent dyes do not affect the localization of DNA within the spherulite. Typical confocal microscopy sections of spherulites condensed as described in Figure 1 from equal mixtures of (A) $^{699}_{406}$DNA-G with $^{699}_{406}$DNA-R and (B) $^{469}_{176}$DNA-G with $^{469}_{176}$DNA-R. The two color channels were normalized for intensity and corrected for chromatic aberration.

eluted sample was dialyzed, concentrated, ethanol precipitated, redissolved in 10 mM Tris, pH 7.5, and 1 mM ethylenediaminetetraacetic acid (EDTA), and retested by UV absorbance and gel electrophoresis.

3D *z*-stacks and 2D confocal slices were collected on a Leica SP2 or Zeiss LSM 510 microscope with oversampling (60 × 60 nm pixel size) for subsequent noise reduction. Alexa Fluor 555 fluorescence was excited with a 514 nm (SP2) or 543 nm (LSM510) laser and detected from 535 to 630 nm (SP2) or from 560 to 615 nm (LSM510); Alexa Fluor 647 fluorescence was excited using a 633 nm laser (SP2 and LSM510) and detected from 650 to 745 nm (SP2) or above 650 nm (LSM510). The detector gain, scanning speed, and laser intensities were optimized for best image quality and minimal photobleaching.

The image of each spherulite was processed and analyzed independently. (i) The intensities of the two color channels were normalized to give the same integral value for each spherulite. (ii) The channels were then realigned according to the "center of mass" for each color, to correct for chromatic aberrations sometimes visible as a shift between the two channels. Such chromatic shifts can be caused, for example, by nonparallel

alignment of the cover glass with respect to the objective. For consistency, this correction was applied to each spherulite image, although it may result in underestimating the extent of color separation in samples with significant DNA segregation. (iii) To reduce the detector and digital noise, the color intensity at each pixel was replaced by the average intensity of a 5 × 5 pixel square centered on that pixel. The small size of this square (300 × 300 nm) allowed for the noise reduction without affecting the optical resolution of the image. The color separation fraction within the 2D confocal midsection of each spherulite was then calculated as

$$f = \frac{\sum_{i,j} |IG_{ij} - IR_{ij}|}{\sum_{i,j} IG_{ij} + IR_{ij}} \qquad (1)$$

Here, the sum is taken over all pixels i,j within a spherulite section and IG and IR are the normalized intensities of the green and red channel at each pixel, respectively. As follows from
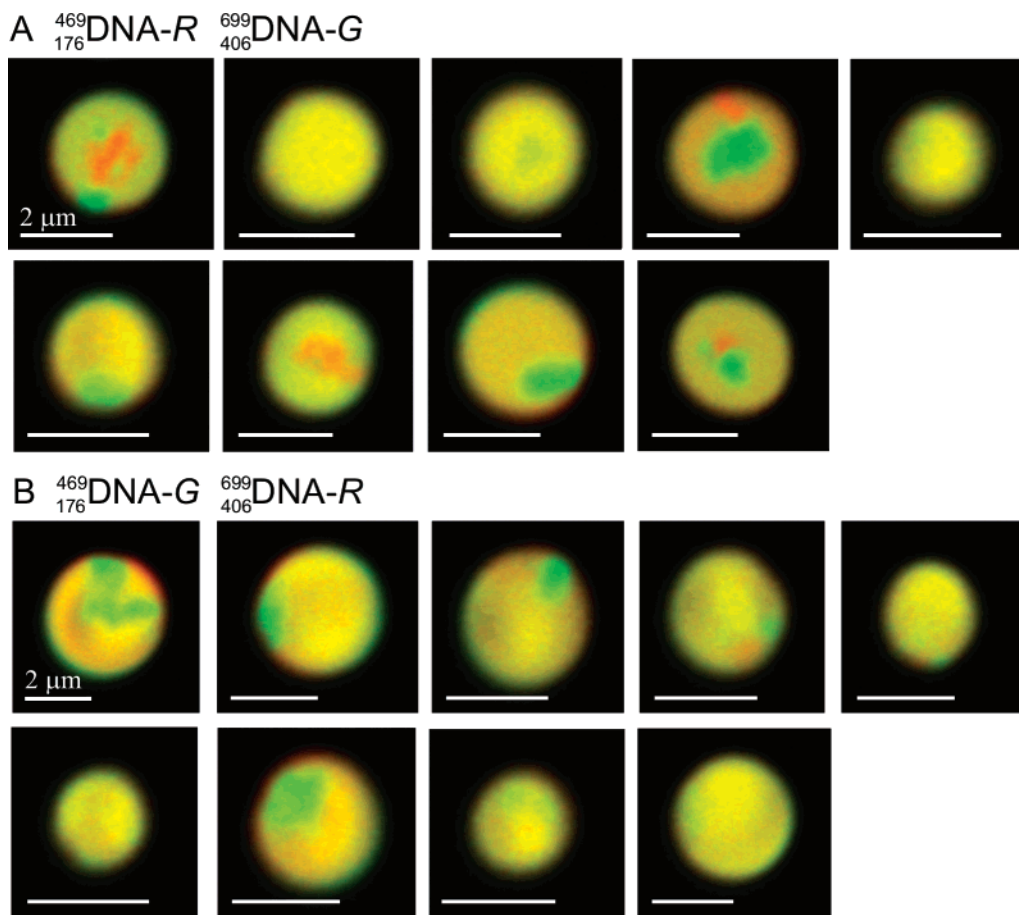
Figure 3. Confocal cross sections of spherulites containing equal mixtures of (A) $^{469}_{176}$DNA-R with $^{699}_{406}$DNA-G and (B) $^{469}_{176}$DNA-G with $^{699}_{406}$DNA-R demonstrate spontaneous segregation of DNAs with identical sequences within the spherulite. The samples were from the same DNA preparations, and microscopy slides were assembled, equilibrated, and imaged in parallel with the samples in Figure 2; imaged data were treated in the same way.

eq 1, $f = 0$ corresponds to complete colocalization of the fluorescent dyes in each pixel while $f = 1$ corresponds to complete separation of the dyes.

Results and Discussion

Two 294 base pair (bp) DNA sequences with similar GC content were selected from the circular bacteriophage ϕX174 genome and amplified by PCR to produce blunt-ended $^{469}_{176}$DNA and $^{699}_{406}$DNA, where the sub- and superscript label the first and last nucleotide of each fragment within the ϕX174 sequence (Supporting Information Figure S1). 5′-Amine-modified-DNAs were labeled with Alexa Fluor 555, yielding green fluorescent $^{469}_{176}$DNA-G and $^{699}_{406}$DNA-G, or with Alexa Fluor 647, yielding far-red fluorescent $^{469}_{176}$DNA-R and $^{699}_{406}$DNA-R. Different mixtures of 50 μg/mL labeled and unlabeled fragments in 0.5 M NaCl, 10 mM Tris, and 1 mM EDTA, pH 7.5, were condensed by polyethylene glycol (PEG 6000) into aggregates, from which PEG is excluded.[17] The condensation started at 5 wt % PEG and was essentially complete at 15 wt % or higher PEG concentrations (Supporting Information Figure S2). To reduce possible kinetic traps preventing DNA segregation, further experiments were performed with samples mixed at 5 wt % PEG, corresponding to the condensation onset. Freshly prepared condensed DNA samples were mounted on microscope slides, sealed, and equilibrated for 2 weeks.

Examination between crossed polarizers revealed ~1−10 μm spherical aggregates (spherulites) with the Maltese cross optical textures characteristic of the liquid-crystalline, cholesteric phase

(Figure 1A,B). In cholesteric spherulites, DNA fragments are preferentially oriented along a particular direction that rotates helically through the sample with a well-defined pitch, usually in the micron range.[2,18] The surfaces of nearly parallel nearest neighbor DNAs are separated by 1−3 nm of water.[19,20] The spherulites from DNA molecules with the same sequence exhibited uniform Maltese cross textures (Figure 1A,B). The spherulites from $^{469}_{176}$DNA and $^{699}_{406}$DNA mixtures exhibited disrupted textures (Figure 1C), suggesting some effect of the sequence on intermolecular interactions.

The distribution of DNA within the spherulites was determined by confocal imaging of discrete spherulite layers. To minimize interactions between the fluorescent dyes on adjacent DNA, approximately 1 in 25 DNA molecules (4 mol %) in each mixture were labeled. All permutations of single DNA molecules and dyes were first examined to establish that the labels did not interfere with spherulite formation (Supporting Information Figure S3). Each DNA sequence was then examined with 2 mol % labeled with one dye and 2 mol % labeled with the other dye. Binary mixtures of $^{469}_{176}$DNA-G with $^{469}_{176}$DNA-R and $^{699}_{406}$DNA-G with $^{699}_{406}$DNA-R showed almost perfect overlay of the signals from each fluorescence channel (Figure 2). To quantify the extent of color separation within the spherulite cross sections, the fluorescence channels were aligned to remove chromatic aberrations. The average color separation fraction of $f = 0.04 \pm 0.01$ (Figure 4) calculated from eq 1 was consistent with that expected upon complete colocalization of the same fragments labeled by different dyes.[21]
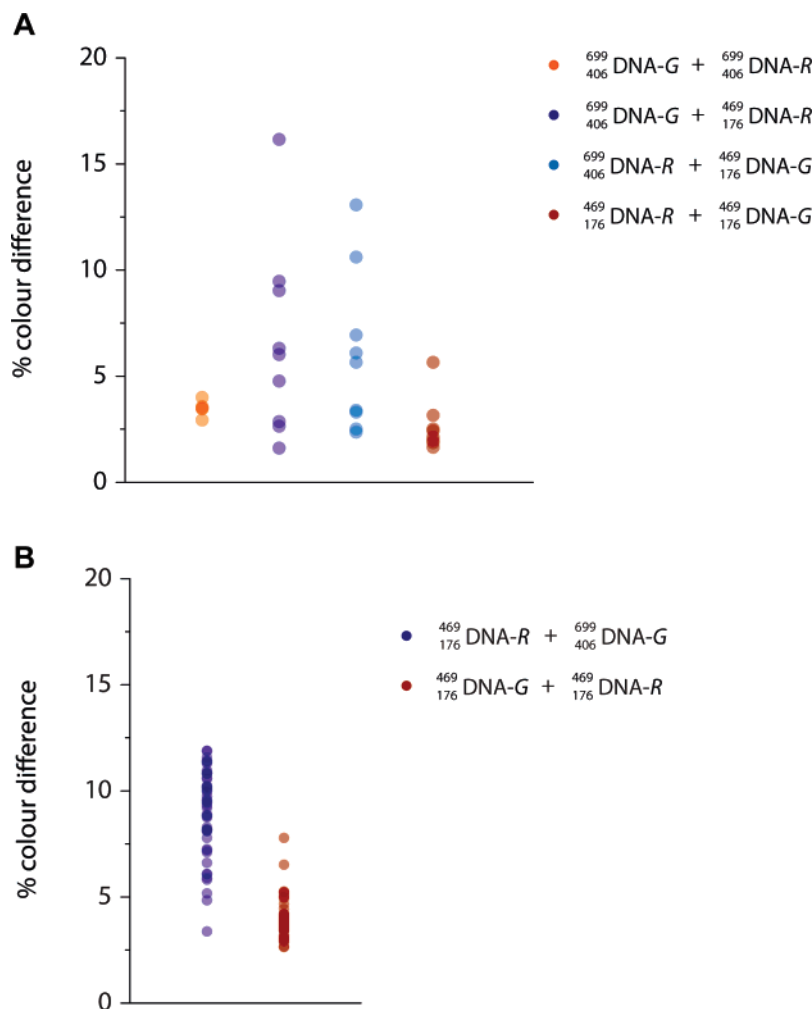
Figure 4. Quantitative analysis of color separation in different DNA mixtures. (A) Color separation fractions (eq 1) for the spherulites shown in Figures 2 and 3. (B) A statistically significant sample of spherulites for $^{469}_{176}$DNA-R + $^{699}_{406}$DNA-G (dark blue, 50 samples, $f = 0.09 \pm 0.02$) and $^{469}_{176}$DNA-G + $^{469}_{176}$DNA-R (red, 50 samples, $f = 0.04 \pm 0.01$) from a different experiment.

In binary mixtures of $^{469}_{176}$DNA-R with $^{699}_{406}$DNA-G and $^{469}_{176}$DNA-G with $^{699}_{406}$DNA-R, a significant degree of color separation was observed (Figure 3). Analysis of a statistically significant sample gave a color separation fraction of $f = 0.09 \pm 0.02$ over the spherulite cross sections, indicating DNA segregation (Figure 4). Within the areas of color separation, values of $f = 0.25-0.47$ were obtained, indicating that fragments with identical sequences were approximately 2 times more likely to be found near each other than the fragments with different sequences. Such segregation suggests that the pair interaction between the same fragments is more favorable than the interaction between fragments with different sequences by ∼1 kT (0.6 kcal/mol) (where k is the Boltzmann constant and T is the absolute temperature).

Amazingly, the forces responsible for the sequence recognition can reach across more than 1 nm of water separating the surfaces of nearest neighbor DNAs in the spherulites. Slight water loss over the equilibration time could reduce the separation from 2 to 3 nm separation expected[19,20] at 5% PEG in 0.5 M NaCl. However, the separation must still have been larger than 1 nm, as the DNA remained in the cholesteric phase rather than a columnar phase expected[20,22] at smaller distances.

We hypothesize that the origin of this recognition may be as follows.[15] In-register alignment of phosphate strands with grooves on opposing DNA minimizes unfavorable electrostatic interactions between the negatively charged phosphates and maximizes favorable interactions of phosphates with bound counterions. DNAs with identical sequences will have the same structure and will stay in register over any juxtaposition length. Nonhomologous DNAs will have uncorrelated sequence-dependent variations in the local pitch that will disrupt the register over large juxtaposition length. The register may be restored at the expense of torsional deformation, but the deformation cost will still make juxtaposition of nonhomologous DNAs unfavorable.[14] The sequence recognition energy, calculated from the corresponding theory is consistent with the observed segregation within the existing uncertainties in the theoretical and experimental parameters (Supplemental Theory). This energy is ∼1 kT under the conditions utilized for the present study, but it is predicted to be significantly amplified, for example, at closer separations, at lower ionic strength, and in the presence of DNA condensing counterions.[10,15]

Presently, we cannot exclude other mechanisms for the observed segregation. For instance, sequence-dependent bending of double helices may also lead to homology recognition by affecting the strand-groove register of two DNA molecules in juxtaposition. The juxtaposition of bent, nonhomologous DNAs may also be less energetically favorable under osmotic stress, since it may reduce the packing density of spherulites. In addition, formation of local single-stranded bubbles and base flipping[23] may cause transient cross-hybridization between the molecules, as proposed to explain Mg^{2+} induced self-assembly

of DNA fragments with the same sequence and length.[16] We consider it to be rather unlikely in this instance, since the probability of bubble formation in unstressed linear DNA of the studied length is very small in contrast to the case where topological strain is relieved by bubble formation in small circular DNA molecules.[23] Furthermore, bubble formation would distort the cholesteric order of spherulites and we see no evidence of this in spherulites composed of a single type of DNA molecule.

However, regardless of the underlying mechanism, the segregation of identical DNAs in highly hydrated cholesteric spherulites provides evidence for homology recognition between intact double helices through physical forces as an intrinsic property of DNA. It is notable that some recognition of unknown origin and pairing between homologous double helices has been proposed as a necessary step preceding double strand breaks in homologous DNA recombination within cells.[24−26] Much work remains to be done before we know the mechanism of the recognition observed in the present study and whether this recognition plays any role within the complex cellular environment, but the possibilities are intriguing. Here, our primary goal was to demonstrate homology recognition in a concentrated mixture of two fragments with the same base pair composition and length but different sequences by using physical experiments in a pure electrolytic solution. We hope that this report will catalyze more detailed studies with DNA sequences of different length and nucleotide composition in different environments.

Supporting Information Available: Supplementary figures and theory. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Minsky, A. *Ann. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 317.
(2) Livolant, F. *Eur. J. Cell Biol.* **1984**, *33*, 300.
(3) Rau, D. C.; Parsegian, V. A. *Biophys. J.* **1992**, *61*, 246.
(4) Podgornik, R.; Strey, H. H.; Rau, D. C.; Parsegian, V. A. *Biophys. Chem.* **1995**, *57*, 111.
(5) Bloomfield, V. A. *Biopolymers* **1997**, *44*, 269.
(6) Strey, H. H.; Podgornik, R.; Rau, D. C.; Parsegian, V. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 309.
(7) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.
(8) Gelbart, W.; Bruinsma, R.; Pincus, P.; Parsegian, V. *Phys. Today* **2000**, *53*, 38.
(9) Besteman, K.; Van Eijk, K.; Lemay, S. G. *Nat. Phys.* **2007**, *3*, 641.
(10) Kornyshev, A. A.; Lee, D. J.; Leikin, S.; Wynveen, A. *Rev. Mod. Phys.* **2007**, *79*, 943.
(11) 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.
(12) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. *J. Mol. Biol.* **1995**, *247*, 34.
(13) Dickerson, R. E. *Methods Enzymol.* **1992**, *211*, 67.
(14) Cherstvy, A. G.; Kornyshev, A. A.; Leikin, S. *J. Phys. Chem. B* **2004**, *108*, 6508.
(15) Kornyshev, A. A.; Leikin, S. *Phys. Rev. Lett.* **2001**, *86*, 3666.
(16) Inoue, S.; Sugiyama, S.; Travers, A. A.; Ohyama, T. *Biochemistry* **2007**, *46*, 164.
(17) Parsegian, V. A.; Rand, R. P.; Rau, D. C. *Methods Enzymol.* **1995**, *259*, 43.
(18) Leforestier, A.; Livolant, F. *Biophys. J.* **1993**, *65*, 56.
(19) Stanley, C. B.; Hong, H.; Strey, H. H. *Biophys. J.* **2005**, *89*, 2552.
(20) 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.
(21) Random fluctuations in the number of green and red-labelled molecules within the focal volume of the microscope will result in ∼0.01 contribution to *f*. The value of *f* will be further increased by the detector and digital noise. Although such noise is reduced by the averaging over 5 × 5 pixel box, it still contributes ∼0.02, as estimated by measuring the dependence of *f* on the averaging box size.
(22) Durand, D.; Doucet, J.; Livolant, F. *J. Phys. II* **1992**, *2*, 1769.
(23) Travers, A. *Curr. Biol.* **2005**, *15*, R377.
(24) Lewin, B. *Genes VIII*; Pearson Prentice Hall: Upper Saddle River, NJ, 2004; p 421.
(25) Weiner, B. M.; Kleckner, N. *Cell* **1994**, *77*, 977.
(26) Zickler, D. *Chromosoma* **2006**, *115*, 158.