HAL open science

Effect of Genomic Long-Range Correlations on DNA Persistence Length: From Theory to Single Molecule Experiments

Julien Moukhtar, Cendrine Faivre-Moskalenko, Pascale Milani, Benjamin Audit, Cédric Vaillant, Emeline Fontaine, Fabien Mongelard, Guillaume Lavorel, Philippe St-Jean, Philippe Bouvet, et al.

To cite this version:

Julien Moukhtar, Cendrine Faivre-Moskalenko, Pascale Milani, Benjamin Audit, Cédric Vaillant, et al.. Effect of Genomic Long-Range Correlations on DNA Persistence Length: From Theory to Single Molecule Experiments. Journal of Physical Chemistry B, 2010, 114 (15), pp.5125-5143. 10.1021/jp911031y . hal-01556946

HAL Id: hal-01556946

https://hal.science/hal-01556946v1

Submitted on 12 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Effect of Genomic Long-Range Correlations on DNA Persistence Length: From Theory to Single Molecule Experiments

Julien Moukhtar,[†,‡,§] Cendrine Faivre-Moskalenko,[†,‡,§] Pascale Milani,[†,‡,§] Benjamin Audit,[‡,§] Cedric Vaillant,[‡,§] Emeline Fontaine,[‡,||] Fabien Mongelard,[‡,||] Guillaume Lavorel,[‡,§] Philippe St-Jean,[‡,§] Philippe Bouvet,[‡,||] Françoise Argoul,[‡,§] and Alain Arneodo*,[‡,§]

Université de Lyon, F-69000 Lyon, France, Laboratoire Joliot-Curie and Laboratoire de Physique, CNRS/ Ecole Normale Supérieure de Lyon, 46 allée d'Italie, F-69007 Lyon, France, and Laboratoire Joliot-Curie and Laboratoire de Biologie Moléculaire de la Cellule, CNRS/Ecole Normale Supérieure de Lyon, 46 allée d'Italie, F-69007 Lyon, France

Sequence dependency of DNA intrinsic bending properties has been emphasized as a possible key ingredient to *in vivo* chromatin organization. We use atomic force microscopy (AFM) in air and liquid to image intrinsically straight (synthetic), uncorrelated (hepatitis C RNA virus) and persistent long-range correlated (human) DNA fragments in various ionic conditions such that the molecules freely equilibrate on the mica surface before being captured in a particular conformation. 2D thermodynamic equilibrium is experimentally verified by a detailed statistical analysis of the Gaussian nature of the DNA bend angle fluctuations. We show that the worm-like chain (WLC) model, commonly used to describe the average conformation of long semiflexible polymers, reproduces remarkably well the persistence length estimates for the first two molecules as consistently obtained from (i) mean square end-to-end distance measurement and (ii) mean projection of the end-to-end vector on the initial orientation. Whatever the operating conditions (air or liquid, concentration of metal cations Mg^{2+} and/or Ni^{2+}), the persistence length found for the uncorrelated viral DNA underestimates the value obtained for the straight DNA. We show that this systematic difference is the signature of the presence of an uncorrelated structural intrinsic disorder in the hepatitis C virus (HCV) DNA fragment that superimposes on local curvatures induced by thermal fluctuations and that only the entropic disorder depends upon experimental conditions. In contrast, the WLC model fails to describe the human DNA conformations. We use a mean-field extension of the WLC model to account for the presence of long-range correlations (LRC) in the intrinsic curvature disorder of human genomic DNA: the stronger the LRC, the smaller the persistence length. The comparison of AFM imaging of human DNA with LRC DNA simulations confirms that the rather small mean square end-to-end distance observed, particularly for G+C-rich human DNA molecules, more likely results from a large-scale intrinsic curvature due to a persistent distribution of DNA curvature sites than from some increased flexibility.

Introduction

The dynamics of folding and unfolding of DNA within living cells plays a major role in regulating a host of biological processes such as gene activity, DNA replication, recombination, and DNA damage repair.[1−4] In particular, the structure and dynamics of nucleosomes which constitute the basic unit of eukaryotic chromatin organization have attracted increasing experimental and theoretical interest.[5,6] High resolution X-ray analyses[7,8] have provided deep insight into the wrapping of about 147 bp of DNA in almost two superhelical turns around a histone octamer to form the nucleosome. An additional fragment of DNA associated with a linker histone separates successive nucleosomes which are disposed as beads-on-a-string along the DNA. This nucleosomal array is further organized into successive high order structures spanning a wide range of scales to reach a full extent of condensation in metaphase chromosomes.[1−4] Recent experiments have shown that nucleosomes are actually highly dynamical structures that can be moved along DNA by chromatin remodeling complexes at the expense of ATP[9−13] but also move autonomously on short DNA segments.[14,15] Different models have been proposed to account for the nucleosome mobility[16−18] extending from the DNA reptation model[19] to the twist defect diffusion model.[20]

Since the discovery of naturally bent DNA,[21] many experiments have established that the local mechanical properties of the DNA double helix depend upon the nucleotide content and are likely to be sensed by DNA binding proteins.[22,23] Hence, the energy cost for nucleosome formation presents a sequence-dependent contribution which may influence the overall formation and positioning of nucleosomes[24] and in turn the access of proteins to their target sites.[25] It has been proposed that certain nucleotide motifs help DNA folding around the core histones, suggesting that the nucleosome packaging can be facilitated by proper DNA sequences.[26−28] In particular, it is known that some genomic sequences presenting a 10 bp periodicity in the distribution of some di or trinucleotides (e.g., AA/TT) show a higher affinity for nucleosomes.[29−33] The periodic positioning

* To whom correspondence should be addressed. E-mail: alain.arneodo@ens-lyon.fr.
† These authors contributed equally to this work.
‡ Université de Lyon.
§ Laboratoire Joliot-Curie and Laboratoire de Physique, CNRS/Ecole Normale Supérieure de Lyon.
|| Laboratoire Joliot-Curie and Laboratoire de Biologie Moléculaire de la Cellule, CNRS/Ecole Normale Supérieure de Lyon.

of these motifs over a few helical pitches would contribute in a coherent manner to a global spontaneous curvature of DNA and favor its wrapping on the histone surface.[34,35] This explains that, for years, genomic 10 bp periodicity has been considered a typical nucleosomal positioning signal.[30,33,36,37] However, complementary *in vitro* experiments have revealed that more than 95% of genomic sequences present binding affinity for core histones that do not differ from random synthetic DNA fragments, the strongest positioning sequences being artificial ones.[38] Recently, tiled microarray experiments have provided nucleosome positioning data along the 16 *S. cerevisiae* chromosomes.[39,40] Statistical analysis[41,42] of these genome-wide nucleosome occupancy profiles has confirmed that a significant set (30−50%) of nucleosome positions found *in vivo* could effectively be related to a "genomic nucleosomal pattern" based on the 10 bp periodicity in the distribution of given dinucleotide steps. However, how could the results of these genome-wide nucleosome positioning experiments[39−42] account for the vast set (≳95%) of genomic signals which apparently do not differ from random sequences as far as binding affinity to core histones is concerned?[38] How could these apparently random sequences condition the positioning of 70% of the bulk nucleosomes as claimed in ref 39? And how could these apparently "noisy" signals facilitate a collective organization of nucleosomes, leading ultimately to the arrangement of about 30% of the nucleosomes into regular arrays?[39,40]

An alternative understanding of the nucleosome positioning issue relies on the observation that, in eukaryotic genomes, sequence motifs favoring the formation of nucleosome present particular persistent correlation properties.[43,44] These correlations, which are referred to as long-range correlations (LRC), extend over large distances and are related to the scale-invariant properties of DNA sequences.[45−49] A comparative statistical analysis of "DNA walks"[47] generated from eukaryotic, eubacterial, and archeal sequences using various coding tables including a structural coding table based on nucleosome positioning data[43,44] has revealed that the corresponding DNA chain bending profiles display scale-invariance properties characterized by two LRC regimes. In the 10−200 bp range, LRC are observed for eukaryotic sequences and quantified by a Hurst exponent value $H \simeq 0.6$ as the signature of the nucleosomal structure. In contrast, for eubacterial sequences, the uncorrelated $H = 0.5$ value is systematically observed. Over larger distances (≳200 bp) up to several kbp, stronger LRC with $H \simeq 0.8$ seem to exist in most genomic DNA sequences and are likely to play a role in the condensation of the nucleosomal string into the so-called 30 nm chromatin fiber.[43,44] The fact that this large-scale regime of LRC is also present in eubacterial sequences shows that it is a possible key to the understanding of the structure and dynamics of both the eukaryotic and prokaryotic chromatin fibers. Actually, the sequence-dependent (LRC) persistent random distribution of local curvature sites is another way to induce some large-scale intrinsic curvature,[50] henceforth favoring the formation of small (i.e., a few hundred bp) DNA loops,[51,52] the larger the correlations, the smaller the size of the loops and in turn the propensity of eukaryotic DNA to interact with histones to form nucleosomes. A recent statistical analysis[53] of tiled microarray data has revealed that the experimental nucleosome occupancy profiles along yeast chromosomes display LRC, corroborating that the positioning of nucleosomes is fundamentally influenced by the sequence over rather long distances. In particular, these data confirm that the $H = 0.8$ LRC regime in the sequence disorder contributes to the overall arrangement of nucleosomes into "crystal-like" phases of confined regularly spaced nucleosomes alternating with "fluid-like" phases of rather nonpositioned nucleosomes. Furthermore, these LRC provide an enlightening interpretation of the nucleosome free regions observed at gene promoters[39−42,54] in terms of very high energy barriers that are absent in the energy landscapes generated from uncorrelated sequences.

In regard to their potential role in regulating the hierarchical structure and dynamics of chromatin, our aim here is to use atomic force microscopy (AFM) to provide the first experimental evidence of the existence of sequence-dependent LRC in naked DNA molecules deposited onto mica under 2D thermodynamic equilibrium. As a reference theoretical guide, we develop some thermodynamic modeling inspired from polymer statistical physics[55] coupled to numerical DNA simulations. In particular, we elaborate on a generalization of the worm-like-chain (WLC) model[56,57] for intrinsically straight polymers, that has been recently proposed[58] to account for the presence of LRC in the intrinsic curvature distribution of genomic DNA. We then proceed to an extensive experimental AFM study of the persistence length of different DNA molecules deposited onto mica under 2D thermodynamic equilibrium. We test the robustness of our results by comparing AFM imaging in air and liquid, and under different ionic conditions. As compared to the 2D equilibrium conformations of synthetic intrinsically straight DNA and of uncorrelated DNA from hepatitis C RNA virus (HCV) that are well described by the WLC model, the conformations adopted by human DNA clearly display a macroscopic intrinsic curvature that confirms that a persistent LRC distribution of DNA curvature sites constitutes a realistic and very attractive alternative to the more popular 10 bp periodicity nucleosome positioning signature. This sequence-dependent large scale intrinsic curvature manifests itself as a significant decrease of the persistence length of human DNA that is well reproduced by the generalized WLC model.[58] We show that this decrease is more pronounced in G+C-rich human DNA regions where the amplitude of the LRC structural disorder is known to be larger than in G+C-poor regions,[59] and that it must not be misinterpreted as the consequence of some increased flexibility.

Experimental Methods

Design and Preparation of the DNA Molecules. Four different types of DNA molecules were designed and prepared for AFM experiments. Purifications were performed by agarose gel electrophoresis without DNA intercalating dyes.

Intrinsically Straight DNA. An intrinsically straight 200 bp DNA fragment was obtained from a plasmid kindly provided by A. Vologodskii.[60] The Vologodskii plasmid was digested with *Hind*III, and 207 bp fragments were gel-purified using a QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. Fragments were ligated using T4 DNA ligase, and concatemers with a size between 800 and 1000 bp were gel fractionated and cloned into a *Hind*III digested and dephosphorylated pBluescript II vector. This ligation product was used to transform the Sure2 *E. coli* host (Stratagene). DNA was prepared from random clones, to select constructs containing an 800 bp insert, as measured after digestion with *Cla*I and *Eco*RI, two sites that flank the *Hind*III site. DNA was finally prepared for AFM using Qiagen's Maxiprep kit, according to the manufacturer's instructions. During the whole procedure, UV irradiation of the sample was not used.

Uncorrelated Hepatitis C RNA Virus Molecule. A 2211 bp fragment (G+C = 56%) was extracted from phCMVcE1E2 plasmid (EMBL: CQ867236), kindly provided by B. Bartosch

and F.-L. Cosset, via *Eco*RI digestion. This fragment corresponds to E1 and E2 Hepatitis C virus (HCV) envelope protein genes and a piece of Sindbis virus junk DNA. HCV and Sindbis virus are single-stranded, positive sense RNA viruses with no DNA stage that were shown to display uncorrelated bending profiles.[44]

LRC Human DNA Fragments of Different GC Contents. Four repeated sequence free loci of size $L \simeq 2200$ bp were delineated along the human genome using UCSC data (May 2004 release): (i) chr 21: 21524828−21527029, $L = 2202$ bp, G+C = 31%; (ii) chr 21: 34394438−34396643, $L = 2206$ bp, G+C = 38%; (iii) chr 21: 37044232−37046437, $L = 2206$ bp, G+C = 43%; (iv) chr 8: 125789398−125791587, $L = 2190$ bp, G+C = 46%. They were extracted by PCR amplification, cloned in pGEMTeasy vector, and digested by *Not*I and *Eco*RI.

Human Genomic DNA Molecules. A mixture of DNA molecules of size ranging from 1000 to 3000 bp representative of human complete genome was obtained by *Dra*I digestion of genomic DNA. *In silico* simulation of digestion experiments showed that *Dra*I digestion of human genome produced a significant proportion of DNA molecules with the size of interest and a G+C content distribution similar to that of the complete genome (data not shown). High molecular weight DNA was purified from exponentially growing HeLa cells using standard procedures, digested using *Dra*I, and fractionated on 1% agarose gel together with a molecular weight marker. DNA was purified from the gel using the High Pure PCR Product Purification Kit (Roche) following manufacturer instructions. Care was taken to protect purified DNA from UV irradiation.

Wavelet-Based LRC Analysis. A wavelet-based LRC analysis of DNA bending profiles obtained when using the PNuc trinucleotide structural table, as previously reported in refs 43 and 44, was performed for fragments that displays a sequence-induced structural disorder. The HCV fragment presented convincing scaling behavior over the entire accessible range of scales with Hurst exponent $H = 1/2$, thus confirming uncorrelated bend angle fluctuations.[44] The four human DNA fragments of different G+C contents displayed a scaling behavior with clear LRC properties ($H > 1/2$), in particular $H = 0.73 \pm 0.03$ for scales greater than 200 bp. This estimate was confirmed when extending the sequence 30 kbp upstream and 30 kbp downstream of the selected human fragments. Actually, this estimate was quite representative of the H values obtained all over the human genome[44] (data not shown).

Atomic Force Microscopy. AFM imaging was performed in air and in solution using a Nanoscope IIIa (Veeco/Digital Instruments, Santa Barbara, CA) equipped with a type-E scanner operating in tapping mode. Freshly cleaved mica (muscovite mica, grade V-I, SPI) served as a support for sample adsorption. DNA preparations were diluted (0.3 ng/μL) in different buffers according to different experimental protocols.

AFM Imaging in Air. In the presence of doubly charged counterions (Mg^{2+}, Ni^{2+}) in the solution, DNA molecules are able to equilibrate on the surface before being captured in a given conformation.[61−63] For scanning in air, we used the following buffer solutions.

- Buffer A1: 10 mM Tris−HCl, pH 7.4, 5 mM $MgCl_2$
- Buffer A2: 10 mM Tris−HCl, pH 7.8, 5 mM $MgCl_2$, 0.05 mM $NiCl_2$
- Buffer A3: 10 mM Tris−HCl, pH 7.8, 0.5 mM $NiCl_2$.

A 5 μL droplet of the solution was deposited for 2 min, and then, the mica disk was rinsed carefully with 1 mL of Milli-Q water and dried under a nitrogen flow prior to imaging. Samples were imaged with silicon tips (type MPP-11100, Veeco Instruments) at a resonance frequency f of 250−350 Hz and set point of 0.6−1.2 V.

AFM Imaging in Liquid. For scanning in liquid, we followed a slightly different protocol with the following buffer (same as buffer A3)

- Buffer L: 10 mM Tris−HCl, pH 7.8, 0.5 mM $NiCl_2$.

As before, a 5 μL droplet of the solution was deposited onto a freshly cleaved mica disk and incubated for 2 min. Then, a 100 μL drop of buffer L was introduced in the liquid cell prior to imaging with a commercial silicon nitride probe (type NP-S, Veeco Instruments) at a resonance frequency f of 9−10 Hz and a set point of 0.3−0.4 V.

The images were recorded at room temperature, both in air and liquid, at a scan size of 1.5×1.5 μm^2, a scan rate of 2 Hz, and a resolution of 512×512 pixel^2. Prior to image processing, images were oversampled to 1024×1024 pixel^2, resulting in a resolution of ~4 bp per pixel. To avoid molecular crowding effects resulting from the interaction of neighboring molecules during their equilibrium process,[61] we used a high dilution rate so that each image contained only a few distant DNA molecules, at the expense of our statistical sampling.

Image Processing. The DNA traces were analyzed using a homemade Matlab (MathWorks Inc., Natick, MA) toolbox. First, the images were flattened to remove the long-term drift of the setup. A few flattened images of DNA molecules are shown in Figures 5a−c, 8, and 10. Then, the DNA chains were identified and digitized using morphological tools such as hysteresis thresholding, erosion, dilation, and skeletonization to estimate their free DNA length using Kulpa and corner chain estimators.[64] As reported in previous studies,[61,64,65] our estimate of the mean contour length of the DNAs under study yields a helical rise of 3.20 ± 0.05 Å (respectively, 3.60 ± 0.05 Å) per bp in air (respectively, in liquid) which underestimates (respectively, overestimates) the 3.38 Å per bp measured by crystallography. For each entropic realization of a DNA molecule of size L (~2200 bp), we estimated both the end-to-end distance $R(s)$ and the scalar product $\vec{u}(0)\cdot\vec{R}(s)$ with a resolution of 1 pixel = 4 bp. Then, by averaging over all of the DNA trajectories obtained from the AFM images, we estimated $\ell_p^{\mathcal{D}}(s) = \langle R^2(s)\rangle/2s$ and $\ell_p^{\int}(s) = \langle \vec{u}(0)\cdot\vec{R}(s)\rangle$ that both converged to the 2D persistence length in the limit $s \to L \to \infty$ (see eqs 9 and 12).

Numerical Simulations of Uncorrelated and LRC DNA Chains. Because of the rather small size ($\lesssim$2200 bp) of the DNA molecules under study, DNA chains were simulated without taking into account excluded volume effects among segments of the same molecule (self-avoiding effects) which were shown to show up for sizes larger than 10−20 ℓ_p only.[64,66] The interaction energy between the DNA chain and the surface was assumed to be zero. A simulated chain consisted of a series of L segments of length 1 pixel and infinitesimal thickness, joined together at one end.

Intrinsically Straight DNA. 2D equilibrium conformations of straight DNA were generated using the centered Gaussian distribution of variance $1/A$ given by eq 7 ($\ell = 1$) to randomly choose the angle deformation $\Delta\theta(n)$ between successive segments.

DNA Chains with Intrinsic Uncorrelated Structural Disorder. To model the presence of some intrinsic uncorrelated curvature disorder induced by the sequence, we first generated a Gaussian white noise $\Delta\theta_o(n)$ of zero mean and variance σ_o^2 from which we constructed a frozen uncorrelated trajectory like the one shown in Figure 2a. Then, 2D equilibrium conformations were generated as before but using for $\Delta\theta(n)$ a Gaussian law of variance $1/A$ which was no longer of zero mean but centered at $\Delta\theta_o(n)$. In Figure 4c, we show for illustrative purposes $N =$

100 trajectories mimicking the effect of the thermal fluctuations ($2A = 490$ bp) on the original uncorrelated frozen chain ($H = 1/2$, $\sigma_o = 0.022$).

DNA Chains with LRC Structural Disorder. Along the lines of our generalized WLC model,[58] we modeled the presence of LRC in the DNA intrinsic curvature properties using a fractional Gaussian noise of zero mean and variance σ_o^2 and whose index ($1/2 < H < 1$) actually controlled (i) the power-law increase of the variance $\sigma_o^2 \ell^{2H}$ of the distribution of intrinsic curvature $\Delta\theta_o(\ell)$ over a distance ℓ (eq 16), and (ii) the slow power-law decay of the intrinsic curvature correlation function (Figure 2b). A typical constructed frozen LRC trajectory with $H = 0.8$ and $\sigma_o = 0.05$ is shown in Figure 2a. Then, 2D equilibrium conformations were generated as before for uncorrelated chains, i.e., using for $\Delta\theta(n)$ a Gaussian law centered at $\Delta\theta_o(n)$ and of variance $1/A$. In Figure 4d, we show $N = 100$ trajectories mimicking the effect of thermal fluctuations on some LRC frozen chains ($H = 0.73$, $\sigma_o = 0.007$) and which, according to the value of the elastic bend rigidity ($2A = 490$ bp), more or less retain the overall large-scale curvature of the original LRC chain.

Theoretical Analysis of the Influence of Sequence-Dependent LRC on the Elastic Properties of 2D DNA Chains

During the 1990s, several experimental and theoretical studies investigated the effect of the sequence on the elastic properties of long DNA chains. In particular, micromanipulation experiments performed on single molecules of DNA have provided very interesting information on the elastic response of these macromolecules to external stretching forces.[67−72] One remarkable feature revealed by these pioneering experiments is the fact that the force vs extension curves are very well described by a simple elastic model, the so-called worm-like-chain (WLC) model[56,57,73−76] with a single elastic constant, the persistence length A^{eff} that characterizes DNA bend rigidity. In physiological salt concentrations, at room temperature, most experiments yield a similar estimate of $A^{\text{eff}} \simeq 50$ nm (~150 bp). However, A^{eff} does not correspond to the genuine bend rigidity (A) of the DNA double helix, since intrinsic structural disorder induced by the sequence contributes to the elastic response to external stresses. In the late 1980s, Trifonov et al.[77] suggested that A^{eff} can be expressed as

$$1/A^{\text{eff}} = 1/A + 1/A_o \tag{1}$$

where A_o is the "static" bend persistence length of the random walk defined by the axis of the DNA double helix in the absence of thermal fluctuations and A is the "dynamic" bend persistence length of a DNA double helix in the absence of intrinsic structural disorder. Equation 1 has received some early theoretical and computational confirmation.[78−80] Experimentally, from the investigation of native[77,81] and "intrinsically straight" synthetic[60,82,83] DNA, eq 1 has led to values of A ranging from 60 up to 210 nm, as compared to the generally accepted value $A^{\text{eff}} \simeq 50$ nm. In a previous work,[59] we have revisited the results of single-molecule DNA stretching experiments using a 3D model that explicitly includes some sequence-dependent intrinsic LRC structural disorder. The investigation of artificial and native sequences has confirmed that the WLC model reproduces quite well the data yet with an effective bend stiffness A^{eff} which underestimates the true elastic bend stiffness according to Trifonov et al. (eq 1). However, the main message of this 3D modeling of the double-helix elastic response to external stress is the fact that this response mainly depends on the amplitude σ_o of the structural disorder and turns out to be rather insensitive to the presence of LRC in the sequence when neglecting the coupling between intrinsic curvature and twist.[59] Our goal in this theoretical section is to show that the presence of LRC structural disorder is likely to become noticeable and important when confining the DNA heteropolymer in a plane (2D).[50−52,58]

In two dimensions, in the continuous limit, the local curvature of a semiflexible chain at the location s will be defined as $C(s) = \mathrm{d}\theta(s)/\mathrm{d}s = \dot{\theta}(s)$. Given a certain DNA tract, the local curvature[55] at base-pair n can be defined by approximating locally the chain at the base-pair resolution by an arc of circle of radius $r(n)$, as illustrated in Figure 1. In a small angle approximation, we get

$$C(n) = \frac{1}{r(n)} = \Delta\theta_n = \theta_n - \theta_{n-1} \tag{2}$$

where θ_n is the angle of the nth unit segment with respect to the x-axis. At larger scales, we will use as a "global curvature" the deflection over a distance ℓ:

$$\mathcal{C}_\ell(s) = \int_{s-\ell/2}^{s+\ell/2} \dot{\theta}(s)\,\mathrm{d}s = \theta(s + \ell/2) - \theta(s - \ell/2) \tag{3}$$

Sequence-Dependent Structural Disorder: "Static" Disorder. As illustrated in Figure 2a, when comparing some realizations of uncorrelated and LRC "unconstrained" 2D frozen trajectories (respectively, Brownian[84] and fractional Brownian (fBm)[85] trajectories; see Experimental Methods), due to the persistence in the intrinsic local curvature fluctuations $\dot{\theta}_o(s)$ of the double helical axis, LRC 2D trajectories are more looped than uncorrelated ones. This can be quantified by looking at the two-point curvature correlation function:[51,52]

$$\Lambda_{o,\ell}(\Delta n) = \overline{\mathcal{C}_{o,\ell}(n)\mathcal{C}_{o,\ell}(n + \Delta n)} \underset{\Delta n \to \infty}{\sim} H(2H - 1)\sigma_\ell^2 \Delta n^{2H-2} \tag{4}$$

where $\overline{\cdot}$ stands for the averaging over intrinsic curvature disorder ($\dot{\theta}_o$) realizations. Actually, $\mathcal{C}_{o,\ell}$ displays Gaussian statistics of zero mean and variance $\sigma_\ell^2 = \sigma_o^2 \ell^{2H}$, where the index H ($0 < H < 1$), the Hurst exponent of the fBm, also controls the power-law decay of $\Lambda_{o,\ell}(\Delta n)$ at large distances $\Delta n \gg \ell$. For uncorrelated chains corresponding to $H = 1/2$ (Brownian trajectories[84]), the curvature correlation function decreases rather fast to zero over a distance of the order of the curvilinear distance ℓ used to define locally the curvature (Figure 2b). In the presence of LRC[85] ($1/2 < H < 1$), this correlation function decreases more slowly at large distances (the larger H, the slower the decrease), and this is in very good agreement with the asymptotic power-law behavior given by eq 4. This persistence in the local curvature distribution is likely to manifest itself at larger scales as some propensity to form intrinsic macroscopic curvature and possibly loops.[51,52]

The 2D Worm-Like Chain Model for Intrinsically Straight Polymers: "Dynamic" (or Entropic) Disorder. In this section, we summarize the main results of the WLC model[55−57,73−76] that accounts for the entropic elastic behavior of an intrinsically straight semiflexible polymer. In 2D, the chain conformations

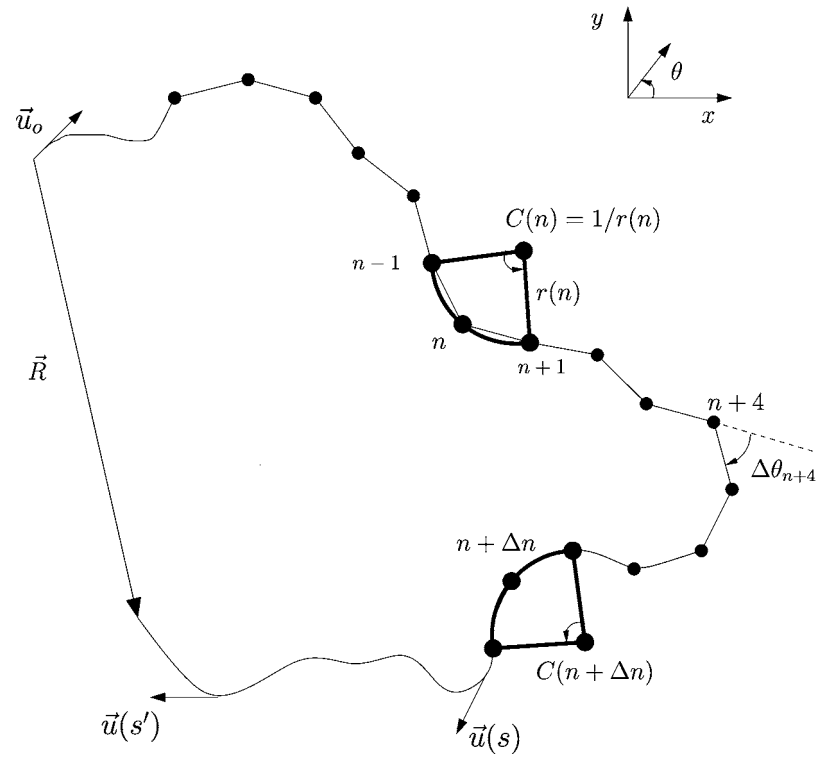

Figure 1. Schematic representation of a DNA chain of total length L with curvature disorder induced by the sequence (intrinsic) and possibly by thermal fluctuations ($T \neq 0$). The local curvature at the base-pair n is defined as $C(n) = 1/r(n) = \Delta\theta_n = \theta_n - \theta_{n-1}$. In the continuous limit, $\vec{R}$ is the end-to-end vector, $\vec{u}(s)$ and $\vec{u}(s')$ are unit vectors tangent to the chain at locations s and s', respectively, and $\vec{u}_o$ is the unit tangent vector at the origin.

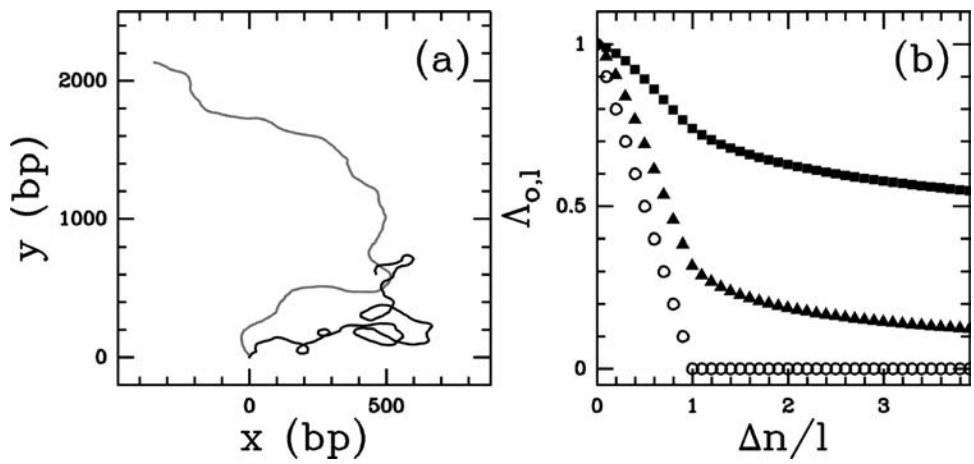

Figure 2. (a) Frozen trajectories for two 2D semiflexible chains of length $L = 3000$ bp with uncorrelated ($H = 0.5$, gray) and LRC ($H = 0.8$, black) structural disorder ($\sigma_o = 0.05$). (b) Normalized intrinsic curvature correlation function $\Lambda_{o,\ell}(\Delta n) = \overline{\mathcal{C}_{o,\ell}(n)\mathcal{C}_{o,\ell}(n+\Delta n)}/\overline{\mathcal{C}_{o,\ell}{}^2(n)}$ vs $\Delta n/\ell$ for uncorrelated $H = 1/2$ (○) and LRC $H = 0.7$ (▲) and 0.9 (■) chains ($N = 1000$).

of length L with local curvatures induced by thermal fluctuations (Figure 1) are controlled by the elastic energy function:[86]

$$\frac{E_{el}}{kT} = \frac{A}{2}\int_0^L \left[\frac{d\theta(s)}{ds}\right]^2 ds \tag{5}$$

where the elastic bend rigidity AkT controls the exponential decay of the orientational correlation between tangent unit vectors to the chain, $\vec{u}(s)$ and $\vec{u}(s')$, separated by a curvilinear distance $\ell = |s' - s|$:

$$\langle \vec{u}(s)\cdot\vec{u}(s+\ell)\rangle = \langle \cos(\theta(s+\ell) - \theta(s))\rangle = \exp\left(-\frac{\ell}{2A}\right), \qquad \forall s \tag{6}$$

where $\langle\cdot\rangle$ stands for averaging over thermal realizations. Thus, $2A = \ell_p$ is nothing but the persistence length, i.e., the decay length through which the memory of the initial orientation of the chain persists. Actually, the probability of having a certain local curvature defined by a bend angle $\Delta\theta(s,\ell)$ over the distance ℓ ($\Delta\theta(s,\ell) = \theta(s+\ell) - \theta(s)$) at any position s follows a Gaussian distribution:[86−88]

$$P_\ell(\Delta\theta) = \sqrt{\frac{A}{2\pi\ell}}\exp\left(-\frac{A}{2\ell}\Delta\theta^2\right) \tag{7}$$

Thus, the smaller $\ell_p = 2A$, the larger the variance $2\ell/\ell_p$ of this distribution, i.e., the larger the amplitude of the local curvatures induced by thermal fluctuations.

Persistence Length Estimate from the Mean Square End-to-End Distance. Then, upon a straightforward integration, one gets the following expression for the mean square end-to-end distance of the polymer:

$$\begin{aligned}\langle R^2\rangle &= \int_0^L \int_0^L \langle \vec{u}(s)\cdot\vec{u}(s')\rangle \,\mathrm{d}s\,\mathrm{d}s' \\ &= 4AL\left[1 - \frac{2A}{L}\left(1 - \exp\left(-\frac{L}{2A}\right)\right)\right]\end{aligned} \tag{8}$$

from which we can estimate the persistence length

$$\ell_\mathrm{p} = \lim_{L\to\infty}\langle R^2\rangle/2L = 2A \tag{9}$$

For finite length polymers, we will use the following definition of the "local persistence length":

$$\ell_p^{\mathscr{D}}(L) = \frac{\langle R^2\rangle}{2L} = 2A\left[1 - \frac{2A}{L}\left(1 - \exp\left(-\frac{L}{2A}\right)\right)\right] \tag{10}$$

where the superscript $\mathscr{D}$ stands for the double integration required to compute $\langle R^2\rangle$ in eq 8.

Persistence Length Estimate from the Mean Projection of the End-to-End Vector on the Initial Orientation. Alternatively, we can estimate the persistence length from the projection of the end-to-end vector $\vec{R}$ on the initial unit tangent vector to the chain,[87] i.e.,

$$\begin{aligned}\langle \vec{u}(0)\cdot\vec{R}\rangle &= \int_0^L \langle \vec{u}(0)\cdot\vec{u}(s)\rangle \,\mathrm{d}s \\ &= 2A(1 - \exp(-l/2A))\end{aligned} \tag{11}$$

from which we get in the limit $L \gg 2A$:

$$\lim_{L\to\infty}\langle \vec{u}(0)\cdot\vec{R}\rangle = 2A = \ell_\mathrm{p} \tag{12}$$

Again, for finite length polymers, we will use the following definition:

$$\ell_p^{\int}(L) = \langle \vec{u}(0)\cdot\vec{R}\rangle = 2A(1 - \exp(-l/2A)) \tag{13}$$

where the superscript $\int$ now stands for the simple integration involved in eq 11. Let us remark that, as compared to the rather slow $1/L$ convergence of $\ell_p^{\mathscr{D}}(L)$ toward the persistence length $\ell_\mathrm{p} = 2A$, $\ell_p^{\int}(L)$ converges exponentially fast to ℓ_p.

Generalized 2D WLC Model for Long-Range Correlated DNA Chains: Superimposing "Static" and "Dynamic" Disorders. In this section, we present a recently introduced generalized 2D WLC model for long-range correlated heteropolymers.[58] As sketched in Figure 2a, DNA is not intrinsically straight and the main limitation of the WLC model is that it does not take into account the intrinsic (sequence-dependent) structural disorder $\dot{\theta}_\mathrm{o}(s)$. This can be done by modifying the elastic energy function to the following new form:

$$\frac{E_\mathrm{el}}{kT} = \frac{A}{2}\int_0^L (\dot{\theta}(s) - \dot{\theta}_\mathrm{o}(s))^2 \,\mathrm{d}s \tag{14}$$

where the local fluctuation $\dot{\theta}(s) = \mathrm{d}\theta(s)/\mathrm{d}s$ for a straight polymer (eq 5) has been replaced by a local fluctuation $\dot{\theta}(s) - \dot{\theta}_\mathrm{o}(s)$ around the intrinsic frozen DNA trajectory (Figure 2a). Let us point out that when modeling the local intrinsic curvature of the chain $\dot{\theta}_\mathrm{o}(s)$ as the realization of a fractional Gaussian noise of zero mean and variance σ_o^2, then[51,52]

$$\Delta\theta_\mathrm{o}(s,\ell) = \int_s^{s+\ell} \dot{\theta}_\mathrm{o}(u)\,\mathrm{d}u = B_H(s+\ell) - B_H(s) = \delta B_{H,\ell}(s) \tag{15}$$

is the increment over a distance ℓ of a fBm B_H,[84,85] which is quite consistent with the Gaussian statistics of fluctuations obtained from wavelet transform analysis of the bending profiles generated from DNA sequences using the PNuc structural coding table in refs 43 and 44 (see Experimental Methods). This means that the intrinsic curvature $\Delta\theta_\mathrm{o}(\ell)$ over a distance ℓ follows a Gaussian law

$$P_\ell(\Delta\theta_\mathrm{o}) = \sqrt{\frac{1}{2\pi\sigma_\mathrm{o}^2\ell^{2H}}}\exp\left(-\frac{\Delta\theta_\mathrm{o}^2}{2\sigma_\mathrm{o}^2\ell^{2H}}\right) \tag{16}$$

of zero mean and variance $\sigma^2(\ell) = \sigma_\mathrm{o}^2\ell^{2H}$, where the Hurst exponent H accounts for the possible existence of LRC ($1/2 < H < 1$) in the distribution of intrinsic curvature along the chain (Figure 2a). Under this assumption, we can analytically compute the directional correlations between tangent unit vectors and eq 6 transforms into

$$\overline{\langle \vec{u}(s)\cdot\vec{u}(s+\ell)\rangle} = \exp\left(-\frac{\ell}{2A} - \frac{\sigma_\mathrm{o}^2\ell^{2H}}{2}\right) \tag{17}$$

where we recall that $\bar{\cdot}$ stands for averaging over intrinsic curvature disorder ($\dot{\theta}_\mathrm{o}$) realizations.

Uncorrelated DNA Chains (H = 1/2). Let us remark that, for uncorrelated (or short-range correlated) DNA chains, introducing $H = 1/2$ in eq 17 yields

$$\overline{\langle \vec{u}(s)\cdot\vec{u}(s+\ell)\rangle} = \exp\left(-\frac{\ell}{2}\left(\frac{1}{A} + \sigma_\mathrm{o}^2\right)\right) \tag{18}$$

We recover the exponential behavior predicted by the WLC model (eq 6) but with an effective persistence length $\ell_\mathrm{p}^\mathrm{eff}$:

$$\frac{1}{\ell_\mathrm{p}^\mathrm{eff}} = \frac{1}{\ell_\mathrm{p}^\mathrm{d}} + \frac{1}{\ell_\mathrm{p}^\mathrm{s}} \tag{19}$$

where $\ell_\mathrm{p}^\mathrm{d} = 2A$ is the dynamical persistence length associated with entropic disorder and $\ell_\mathrm{p}^\mathrm{s} = 2/\sigma_\mathrm{o}^2$, commonly called the static persistence length, is an additional contribution due to the presence of uncorrelated structural disorder (the larger the amplitude σ_o of the local intrinsic curvature fluctuations, the smaller the static persistence length $\ell_\mathrm{p}^\mathrm{s}$). Note that eq 19 is identical to the Trifonov et al. equation (eq 1). From the mean square end-to-end distance of an uncorrelated chain which freely equilibrates on a surface, we can show, under the hypothesis that the chain is a continuous, homogeneous, and isotropic material, that its apparent persistence length is twice the one found for the 3D conformations:[87,88]

$$\ell_p = \ell_p^{2D} = 2\ell_p^{3D} \tag{20}$$

Importantly, WLC eqs 10 and 13 for $\ell_p^{\mathscr{D}}$ and $\ell_p^{\mathscr{S}}$, respectively, still apply to heteropolymers with uncorrelated structural disorder but with an effective elastic bend rigidity that has been renormalized by the amplitude σ_o^2 of the intrinsic curvature fluctuations (Figure 3a−c):

$$A^{\text{eff}} = \left(\frac{1}{A} + \sigma_o^2\right)^{-1} \tag{21}$$

Note that in the weak disorder limit, $\sigma_o^2 \ll 1/A$, eq 19 leads to Nelson's formula[80]

$$\ell_p^{\text{eff}} = 2A(1 - A\sigma_o^2) \tag{22}$$

Long-Range Correlated DNA Chains (H > 1/2). For LRC structural disorder,[58] when using eq 17 to compute both the mean square end-to-end distance and the mean projection of the end-to-end vector on the initial unit tangent vector, then eqs 8 and 11 become, respectively,

$$\overline{\langle R^2\rangle} = \int_0^L \mathrm{d}s' \int_0^{s'} \mathrm{d}s \exp\left(-\frac{s}{2A} - \frac{s^{2H}\sigma_o^2}{2}\right) \tag{23}$$

and

$$\overline{\langle \vec{u}(0)\cdot\vec{R}\rangle} = \int_0^L \mathrm{d}s \exp\left(-\frac{s}{2A} - \frac{s^{2H}\sigma_o^2}{2}\right) \tag{24}$$

Both of these equations can be handled perturbatively and numerically. As shown in Figure 3a,b for $2A = 310$ bp and 490 bp, respectively, when fixing $\sigma_o = 0.01$, we see that, for uncorrelated $H = 1/2$ chains, $\ell_p^{\mathscr{D}}$ and $\ell_p^{\mathscr{S}}$ both converge to the effective persistence length $\ell_p^{\text{eff}} = 305$ bp and 478 bp according to the WLC eq 19 and in good agreement with the weak-disorder perturbative formula (eq 22). When introducing LRC, for both A values, we observe a faster convergence to a smaller persistence length, the larger the H, the smaller the ℓ_p. $\ell_p^{\mathscr{S}}$ still converges faster than $\ell_p^{\mathscr{D}}$ to the persistence length that is well approximated by the following weak disorder formula[58] when $\sigma_o^2 \ll 2/[(2A)^{2H}\Gamma(2H + 1)]$:

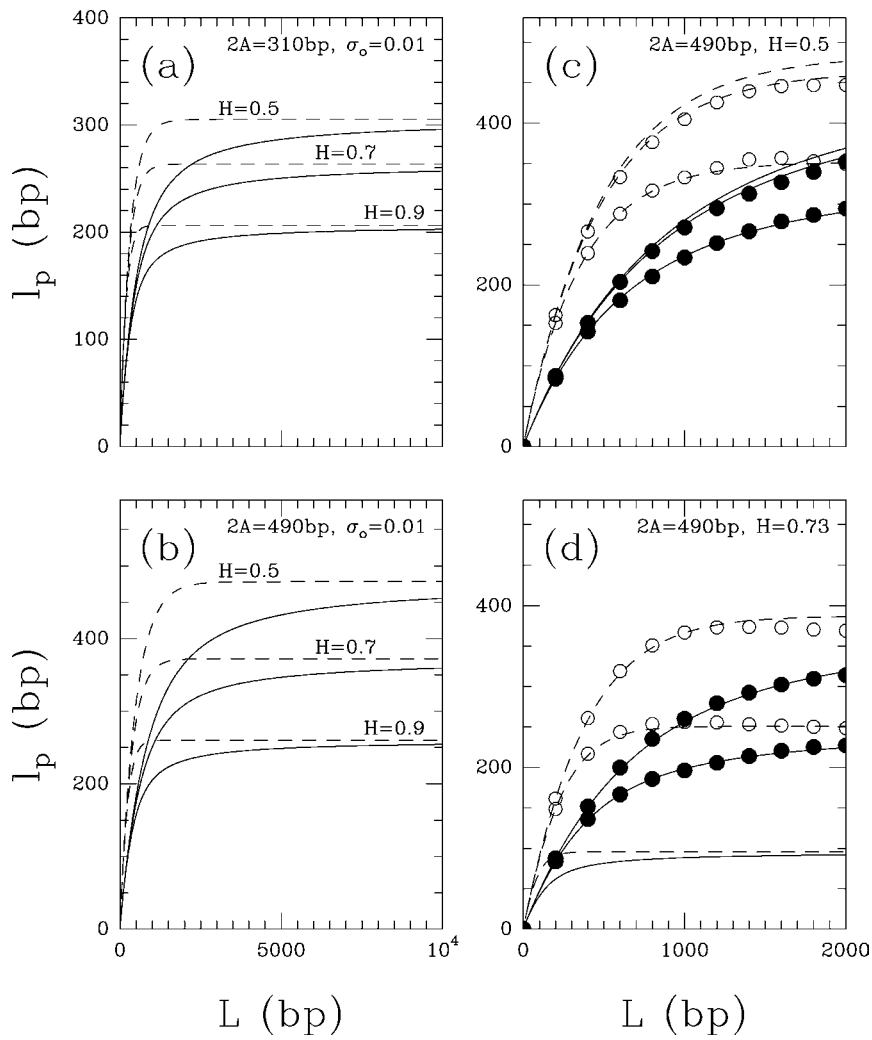

Figure 3. Persistence length $\ell_p^{\mathscr{D}}$ (solid line) and $\ell_p^{\mathscr{S}}$ (dashed line) computed with eqs 23 and 24, respectively: (a) $2A = 310$ bp, $\sigma_o = 0.01$, $H = 0.5$, 0.7, and 0.9 from top to bottom; (b) same as part a but for $2A = 490$ bp; (c) $2A = 490$ bp, $H = 0.5$, $\sigma_o = 0.007$, 0.022, and 0.04 from top to bottom; (d) same as part c but for $H = 0.73$. The symbols $\ell_p^{\mathscr{D}}$ (●) and $\ell_p^{\mathscr{S}}$ (○) correspond to DNA simulations when averaging over one thermal realization for $N = 10^3$ different frozen chains generated with the corresponding parameters as described in Experimental Methods. In parts a and b, the asymptotic convergence of $\ell_p^{\mathscr{D}}$ and $\ell_p^{\mathscr{S}}$ is illustrated up to $L = 10\,000$ bp disregarding excluded volume effects.[64,66]

$$\ell_p^{\text{eff}} = 2A\left(1 - \frac{(2A)^{2H}\sigma_o^{\,2}}{2}\Gamma(2H+1)\right) \qquad (25)$$

where Γ is the Gamma function. This equation generalizes the WLC ($H = 1/2$) Nelson's formula (eq 22) to LRC semiflexible polymers. As shown in Figure 3c for $H = 1/2$ and 3d for $H = 0.73$, when increasing the amplitude σ_o of the intrinsic curvature disorder, the decrease of the persistence length is even more pronounced for LRC chains than for uncorrelated ones. This significant decrease of ℓ_p^{eff} that cannot be accounted for by the WLC model is not the signature of some increased flexibility but rather the consequence of the LRC induced persistence in the distribution of intrinsically bent sites.[50,58]

Comparison with DNA Simulations of Disordered 2D Elastic Chains. The generalized WLC model[58] described in the previous section (note that each H value corresponds to a different LRC WLC model generalization) involves some averaging over the structural disorder, i.e., over many DNA sequences. This is without any doubt a rather prohibitive experimental task that led us, in a first step, to extend the DNA simulations performed in ref 89 for intrinsically straight DNA to uncorrelated and LRC DNAs,[50,58] as explained in Experimental Methods. In Figure 4a and b, we show $N = 100$ frozen DNA chains of length $L = 2200$ bp of uncorrelated ($H = 1/2$, gray) and LRC ($H = 0.73$, black) curvature angle fluctuations of amplitude $\sigma_o = 0.007$ and 0.022, respectively. Most of the LRC chains display some marked macroscopic curvature in contrast to the rather straight uncorrelated chains. In Figure 4c and d, we show $N = 100$ trajectories mimicking the effect of Gaussian thermal fluctuations of variance $1/A$ (eq 7) on some original uncorrelated chain ($\sigma_o = 0.022$) and LRC chain ($\sigma_o = 0.007$) (see Experimental Methods). The values of A, H, and σ_o were chosen to allow for some comparative analysis of persistence length estimates of various DNA fragments using AFM imaging[50] in the Experimental Results section. We report in Figure 3c,d the results of numerical estimates of $\ell_p^{\mathcal{D},f}(L)$ from DNA simulations when averaging over one thermal realization for $N = 10^3$ different uncorrelated ($H = 1/2$) and LRC ($H = 0.73$) frozen chains, respectively. For the two values of $\sigma_o = 0.007$ and 0.022 considered in these simulations, the numerical results obtained for the LRC chains in Figure 3d are in just as good agreement with the theoretical predictions of $\ell_p^{\mathcal{D}}(L)$ (eq 23) and $\ell_p^{f}(L)$ (eq 24), as are the numerical results for the uncorrelated chains with the WLC model predictions in Figure 3c. These results strengthen our generalized WLC model as a very reliable guide to future experiments. They also confirm that some decrease of the persistence length does not necessarily mean enhancement of flexibility but can rather be the signature of some LRC sequence-dependent large-scale intrinsic curvature.[50]

Experimental Results

Various experimental techniques have been used to investigate DNA curvature and flexibility, including X-ray crystallography, gel electrophoresis, electric birefringence, DNA cyclization, NMR, and high resolution microscopy (electron microscopy and

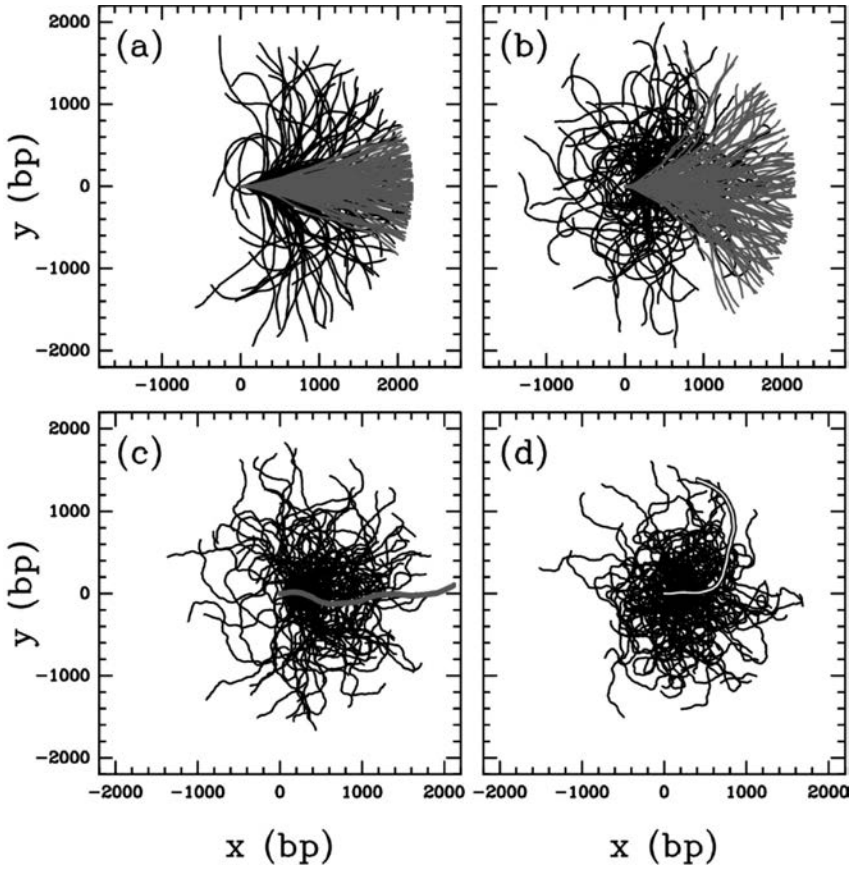

Figure 4. (a) $N = 100$ numerically generated frozen DNA chains of length $L = 2200$ bp with uncorrelated ($H = 1/2$, gray) and LRC ($H = 0.73$, black) intrinsic curvature angle fluctuations of amplitude $\sigma_o = 0.007$. (b) Same as in part a except with $\sigma_o = 0.022$. (c) DNA simulations of $N = 100$ 2D equilibrium conformations (thin black trajectories) generated with an elastic bend rigidity of $2A = 490$ bp from a chain of length $L = 2200$ bp with uncorrelated ($H = 1/2$) intrinsic bend angle fluctuations of amplitude $\sigma_o = 0.022$ (gray trajectory). (d) Same as in part c but for a chain with LRC ($H = 0.73$) intrinsic bend angle fluctuations of amplitude $\sigma_o = 0.007$ (thick black and white).

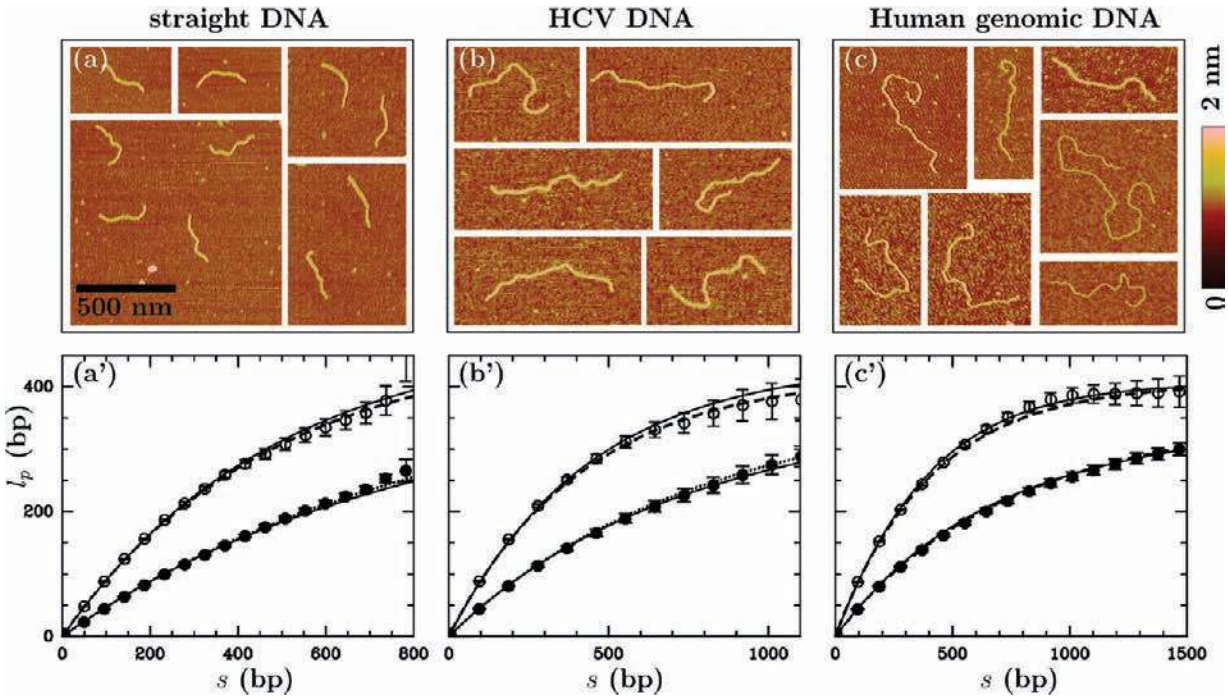

Figure 5. AFM images in tapping mode in air using buffer A1 of (a) $L = 800$ bp intrinsically straight DNA, (b) $L = 2200$ bp uncorrelated HCV DNA, and (c) 1700 bp $\lesssim L \lesssim$ 2500 bp LRC human genomic DNA. Persistence length measurements of $\ell_p^{\mathscr{D}}$ (●) and $\ell_p^{\mathscr{S}}$ (○) when averaging over N imaged molecules of (a′) intrinsically straight DNA ($N = 193$), (b′) uncorrelated HCV DNA ($N = 65$), and (c′) LRC human genomic DNA ($N = 290$). The WLC model predictions (10) and (13) are shown for $2A = 470$ bp (dashed line), 520 bp (dotted line), and 490 bp (solid line) for the straight DNA in part a′, for $2A^{\text{eff}} = 421$ bp (dashed line), 465 bp (dotted line), and 440 bp (solid line) for the HCV DNA in part b′, and for $2A = 410$ bp (dashed line) for human genomic DNA in part c′. In part c′ are shown the predictions of the generalized WLC model (see eqs 23 and 24) for $2A = 490$ bp, $H = 0.73$, and $\sigma_o = 0.007$ (solid line).

AFM).[21−23,61,82,87−91] Among these methods, because it provides a visualization of the 2D conformations of a DNA molecule, AFM has proved to be a very efficient technique, not only because it can be used to estimate the persistence length of DNA via end-to-end distance measurements[61,87−91] but more interestingly because it can give access, via some thermodynamic averaging over a population of these conformations, to both the intrinsic DNA curvature and bend rigidity and this with a surprisingly good resolution of a few double helix pitches.[89,92−95] During the past decade, DNA structures imaged by AFM have been analyzed in the framework of the WLC model.[87−89] In this section, our goal is to use AFM to bring the first experimental evidence of sequence-induced LRC in naked DNA molecules deposited onto mica under 2D thermodynamic equilibrium conditions,[50] thus making the experimental estimate of the persistence length amenable to treatment by the generalized 2D WLC model described in the previous theoretical section. All of the results reported in this section were obtained by AFM imaging in air using buffer A1 (see Experimental Methods).

Intrinsically Straight DNA Molecules. An intrinsically straight DNA fragment of length $L = 800$ bp was designed according to the protocol defined in ref 60 (see Experimental Methods). Figure 5a shows representative AFM images of 2D equilibrium conformations of this straight DNA fragment when submitted to thermal fluctuations. To prevent neighboring molecules from interacting during their equilibration process,[61] we used, as a general experimental strategy, a high dilution ratio so that each image contained no more than one or two DNA molecules, at the expense of our statistical sampling. From a large set of images collected in different experiments, the DNA traces of $N = 193$ molecules were digitized as described in Experimental Methods. An experimental rise of 3.20 ± 0.05 Å·bp^{-1} was determined from the AFM images by dividing the measured average contour length by the total number of base-pairs. In good agreement with previous AFM experimental studies,[61,64,65] this value of the helical rise slightly underestimates the 3.38 Å·bp^{-1} measured by crystallography.[1−4]

Figure 5a′ shows the experimental data for the local persistence length versus s, for s going from 1 to $L = 800$ bp, when using either the mean square end-to-end distance estimate of $\ell_p^{\mathscr{D}}(s)$ or the mean projection of the end-to-end vector onto the initial orientation estimate of $\ell_p^{\mathscr{S}}(s)$. Both data sets are remarkably well fitted by the WLC model predictions (10) and (13) when adjusting $\ell_p = 2A = 520$ and 470 bp, respectively. As a test of the consistency of these results, Figure 6a shows similar estimates of $2A$ when restricting the domain of fit of the data in Figure 5a′ to the first s bp only. For s ranging from 200 bp to the total length 800 bp, all of the data points for $2A^{\mathscr{D}}$ and $2A^{\mathscr{S}}$ fall in the range $\ell_p = 2A = 490 \pm 30$ bp. This yields an estimate of the bend rigidity $A = 245 \pm 15$ bp which corresponds, according to eq 20, to a 3D persistence length $\ell_p^{3D} = A \times 0.32 \simeq 79$ nm, in good agreement with previous measurements by cryo-electron microscopy.[82] For comparison, in Figure 6b,c are shown similar estimates of $2A^{\mathscr{D}}$ and $2A^{\mathscr{S}}$ for two sets of $N = 193$ 2D equilibrium conformations of straight DNA of total length $L = 800$ bp numerically generated by DNA simulations (see Experimental Methods), using a centered bend angle distribution of variance $1/A$ given by eq 7 ($\ell = 1$) with $2A = 490$ bp. It is clear that, for a statistical sample of about 200 DNA trajectories of rather small total length $L = 800$ bp, there is some statistical limitation in the achievable accuracy when estimating the persistence length $\ell_p = 2A$. Results of the numerical simulations in Figure 6b,c show that estimates of $\ell_p^{\mathscr{D}}(s)$ and $\ell_p^{\mathscr{S}}(s)$ can differ by a few tens of bp, which corrob-

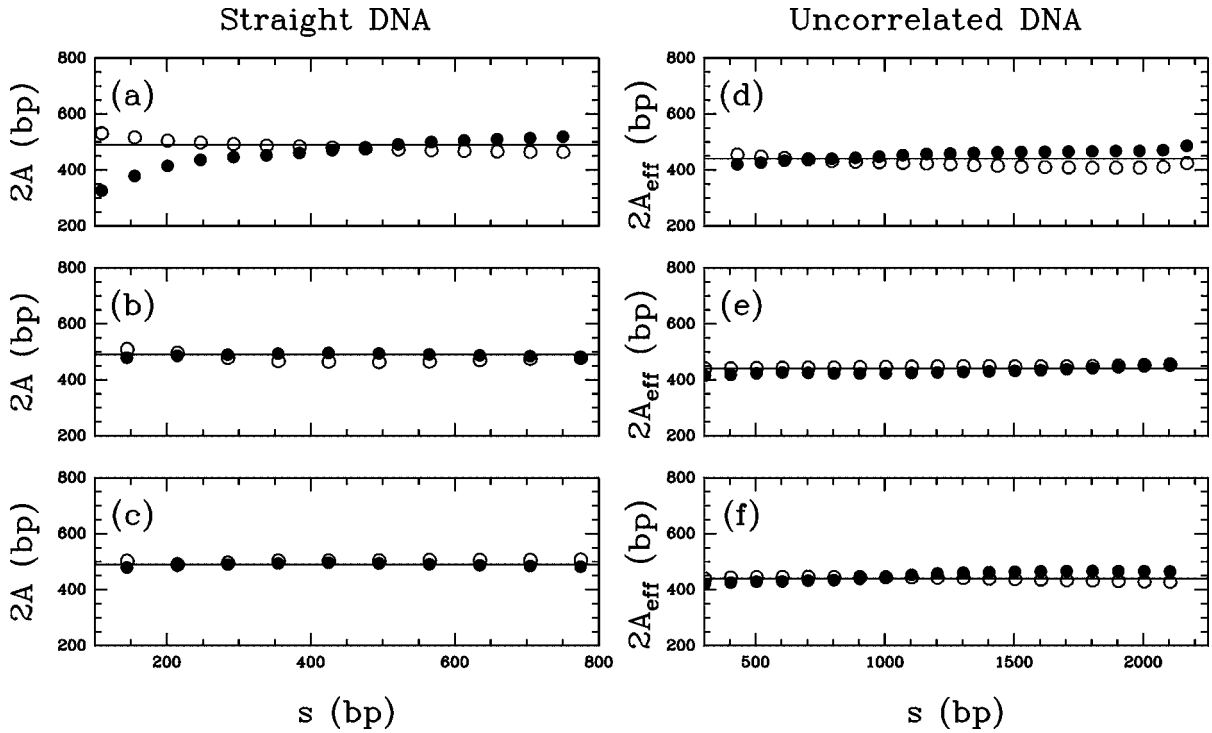

Figure 6. Fit of the data for $\ell_p^{\mathscr{D}}(s)$ (●) and $\ell_p^{\mathscr{S}}(s)$ (○) with the WLC model eqs 10 and 13, respectively: $2A$ estimated when restricting the domain of fit of the data in Figure 5a′ and b′ to the first s bp only. *Intrinsically straight DNA*: (a) AFM study ($N = 193$, $L = 800$ bp), (b and c) DNA simulations ($N = 193$, $L = 800$ bp) with elastic bend rigidity $2A = 490$ bp. *Uncorrelated DNA*: (d) AFM study of HCV molecules ($N = 65$, $L = 2200$ bp), (e and f) DNA simulations ($N = 100$, $L = 2200$ bp) of 2D equilibrium conformations of two uncorrelated random chains ($H = 1/2$, $\sigma_o = 0.022$) generated with an elastic bend rigidity of $2A = 490$ bp and corresponding to $2A^{\text{eff}} = 440$ bp.

orates our estimate of the experimental uncertainty $\ell_p = 2A = 490 \pm 30$ bp in Figure 6a.

Uncorrelated Hepatitis C RNA Virus Molecules. A fragment of $L = 2200$ bp from the HCV genome of total length 9000 bp was extracted via *Eco*RI digestion of phCMVcE1E2 plasmid[96] (see Experimental Methods). HCV belongs to the family of non-retrotranscribed RNA viruses that display uncorrelated bending profiles.[44] Prior to AFM analysis, we performed a LRC analysis of the corresponding DNA chain bending profile generated with the PNuc trinucleotide coding table based on nucleosome positioning data.[43,44] As reported in Experimental Methods, when using the wavelet-based methodology experienced in refs 44−46, we confirmed that, over the entire range of scales, a nice scaling behavior with Hurst exponent $H = 1/2$ is observed as the signature of uncorrelated intrinsic bend angle fluctuations. Figure 5b shows representative AFM images of 2D equilibrium conformations of this uncorrelated $L = 2200$ bp HCV fragment when submitted to thermal fluctuations. From a collection of such images, the DNA traces of $N = 65$ molecules were digitized (see Experimental Methods) for persistence length measurement.

Figure 5b′ shows the experimental data for $\ell_p^{\mathscr{D}}(s)$ and $\ell_p^{\mathscr{S}}(s)$ obtained from our set of 65 HCV molecules. Again, these data are well fitted by the WLC model eqs 10 and 13 yet with an effective persistence length $\ell_p^{\text{eff}} = 2A^{\text{eff}} = 465$ and 421 bp, respectively, i.e., values smaller than the dynamic persistence length estimated just above for intrinsically straight DNA. Indeed, as shown in Figure 6d, when performing a fit of the data with eqs 10 and 13 over different ranges of scales from $s = 200$ to 2000 bp, we get values for $2A^{\mathscr{D}}$ and $2A^{\mathscr{S}}$ that all fall in the interval $\ell_p^{\text{eff}} = 2A^{\text{eff}} = 440 \pm 30$ bp which corresponds to a 3D persistence length $\ell_p^{\text{3D}} = A^{\text{eff}} \times 0.32 \simeq 70$ nm. Since the experimental protocol used to deposit and visualize the different types of DNAs was rigorously the same, we can reasonably assume that the bend rigidities of all molecules are identical. Then, from eq 19, we get the following estimate of the static persistence length $\ell_p^{\text{s}} = 4310$ bp that corresponds to a rather small curvature angle fluctuation amplitude $\sigma_o = (2/\ell_p^{\text{s}})^{1/2} = 0.022 \pm 0.012$, typical of a weak structural disorder. We report in Figure 6e the results of a comparative DNA simulation study of $N = 100$ 2D equilibrium conformations of an uncorrelated random chain ($H = 1/2$, $\sigma_o = 0.022$), generated with an elastic bend rigidity of $2A = 490$ bp as described in Experimental Methods and illustrated in Figure 4c. Like for the HCV molecules, the estimated persistence length definitely underestimates the elastic bend rigidity consistently with Trifonov et al. formula 19. Redoing these DNA simulations for a second realization of uncorrelated random chains with the same curvature disorder amplitude $\sigma_o = 0.022$, then, as shown in Figure 6f, some average estimate of $2A^{\text{eff}} = 440 \pm 20$ bp is obtained in good agreement with both eq 21 and its weak disorder limit eq 22, which predicts $\ell_p^{\text{eff}} = 2A^{\text{eff}} = 436$ bp. These results bring the demonstration that AFM imaging is capable of detecting and quantifying the presence of rather weak curvature structural disorder in short (a few kbp long) DNA molecules.[50]

Long-Range Correlated Human DNA Molecules. DNA fragments of total length 1000 bp $\lesssim L \lesssim$ 3000 bp were purified from exponentially growing HeLa cells using standard procedures, as described in Experimental Methods. We checked that the collected molecules had a G+C content distribution similar to that of the complete human genome. From a genome-wide wavelet-based LRC analysis (see Experimental Methods) of the DNA chain bending profiles generated with the PNuc trinucleotide coding table,[43,44] we confirmed the presence of scaling, in particular for the range of scales (from 200 to 2500 bp) probed by the AFM apparatus, where the Hurst exponent $H = 0.73 \pm 0.03$ was found significantly larger than 1/2 as the signature of

the presence of LRC. Figure 5c shows representative AFM images containing a few 2D equilibrium conformations of these human genomic molecules when submitted to thermal fluctuations. From a large set of images collected in different experiments, the DNA traces of $N = 290$ molecules were digitized with a total length selection: 1700 bp $\lesssim L \lesssim$ 2500 bp. Since each of these imaged DNA molecules most probably correspond to a different DNA sequence, the experimental estimate of the persistence length is likely to account for averaging over both entropic and intrinsic structural disorders, thus making the comparison with the predictions of our generalized WLC model[58] (see the theoretical section) amenable.

Figure 5c′ shows the experimental data for $\ell_p^{\mathcal{D}}(s)$ and $\ell_p^{\mathcal{J}}(s)$ obtained for our set of 290 human DNA molecules. When using the theoretical predictions (eqs 23 and 24) of the generalized WLC model to fit these data, limiting the range of variation of H between 0.7 and 0.8 according to the LRC detected in the genomic sequences, we get a very good agreement for the parameter values ($2A = 487$ bp, $H = 0.73$, $\sigma_o = 0.0077$) and ($2A = 490$ bp, $H = 0.73$, $\sigma_o = 0.0062$), respectively. As a first observation, we recover a value of the elastic bend rigidity in remarkable agreement with the reference value obtained for the intrinsically straight DNA fragment $2A = 490 \pm 30$ bp. As far as the amplitude of the intrinsic curvature disorder $\sigma_o \simeq 0.007$ is concerned, it looks smaller than the one obtained for the uncorrelated HCV molecule.[97] We show in Figure 3d for comparison the numerical data for $\ell_p^{\mathcal{D}}(s)$ and $\ell_p^{\mathcal{J}}(s)$ obtained from DNA simulations of $N = 1000$ 2D equilibrium conformations generated with an elastic bend rigidity $2A = 490$ bp from as many as 1000 different LRC frozen chains of length $L = 2200$ bp with intrinsic disorder parameters $H = 0.73$ and $\sigma_o = 0.007$ (see Experimental Methods). Besides the fact that they provide a very convincing quantitative numerical check of the validity of the generalized WLC model to LRC DNA molecules, these results corroborate the significant lowering of the asymptotic estimate of ℓ_p for the human genomic DNA as compared to the uncorrelated HCV DNA. Indeed, for these parameter values, our generalized WLC model predicts a persistence length $\ell_p = 396$ bp that corresponds to a value of the 3D persistence length $\ell_p^{3D} = \ell_p/2 \times 0.32 \simeq 63$ nm that is significantly smaller than the persistence lengths previously found for the intrinsically straight DNA ($\ell_p^{3D} = 79$ nm) and the uncorrelated HCV DNA ($\ell_p^{3D} = 70$ nm).

Discussion

Experimental Check of 2D Thermodynamic Equilibrium Conditions. A crucial issue in the present interpretation of experimental AFM studies in terms of the WLC model and its generalization to LRC DNA chains is the control of the deposition process in order to avoid kinetic trapping of the DNA molecules onto the surface before equilibrating in a given 2D conformation. As described in Experimental Methods, we followed protocols used in previous studies[61−63,98,99] as attaining 2D thermodynamic equilibrium. This section is devoted to some experimental check on our set of $N = 193$ imaged conformations of the $L = 800$ bp intrinsically straight DNA fragment when imaged in air using buffer A1 (Figure 5a). The results of the statistical analysis of DNA bend angle fluctuations $\Delta\theta(s,\ell)$ over different distances ℓ are reported in Figure 7. In Figure 7a,b are shown the probability density functions (pdf's) $P_\ell(\Delta\theta)$ obtained for $\ell = 230$ and 310 bp, respectively. As a check of 2D thermodynamic equilibrium, we computed the first four moments of $P_\ell(\Delta\theta)$ with the specific goal of testing the Gaussian nature of these pdf's (see eq 7). As shown in Figure 7e,f, the data for $\langle\Delta\theta_\ell\rangle$ and $\langle\Delta\theta_\ell{}^3\rangle$ do not display any significant departure from zero, as expected for the odd moments of a Gaussian distribution. For the second moment, eq 7 predicts

$$\langle\Delta\theta_\ell{}^2\rangle = \frac{2\ell}{2A} = \frac{2\ell}{\ell_p} \tag{26}$$

The data for $\langle\Delta\theta_\ell{}^2\rangle$ in Figure 7g exhibit a very convincing linear behavior when plotted versus ℓ. Accordingly, when plotting $2\ell/\langle\Delta\theta_\ell{}^2\rangle$ vs ℓ in Figure 7h, a well-defined plateau at the value $2A = 490 \pm 30$ bp is observed, in remarkable agreement with the estimate of the bend rigidity previously obtained from persistence length measurements (see Experimental Results). The ratio of the fourth moment to the square of the second moment, the so-called kurtosis of a Gaussian distribution, satisfies the following remarkable relationship:

$$\frac{\langle\Delta\theta_\ell{}^4\rangle}{\langle\Delta\theta_\ell{}^2\rangle^2} = 3, \qquad \forall\ell \tag{27}$$

As shown in Figure 7i, the ratio determined from the experimental AFM data is very close to the theoretical value 3. Overall, the results reported in Figure 7 bring the experimental demonstration that the bend angle fluctuations of the intrinsically straight DNA fragment have Gaussian statistics for distances ℓ between two chain vectors ranging from 100 to 500 bp. For shorter distances, the analysis is biased by the AFM apparatus resolution. For distances larger than 500 bp, the analysis starts being affected by finite sampling effects.

For the sake of comparison, we have reproduced this test of Gaussian statistics for two independent sets of $N = 193$ numerically generated 2D equilibrium conformations of an $L = 800$ bp straight DNA with a bend rigidity of $2A = 490$ bp (see Experimental Methods). As shown in Figure 7, the results obtained for the real and simulated molecules are quite comparable and consistent. In particular, the slight differences observed in the first four moments behavior of $P_\ell(\Delta\theta)$ between our two sets of simulated molecules are as important as the differences observed between the real and numerical molecules. This observation attests that there is no significant departure from Gaussian statistics in the fluctuations of DNA bending angle imaged by AFM corroborating that 2D thermodynamic equilibrium is achieved by our experimental deposition protocol.

Comparing AFM Imaging in Air and in Liquid. One of the most attractive features of the AFM is that it can operate equally well with the cantilever immersed in liquid as in air, making it possible to image biological molecules like DNA and proteins under quasi-physiological conditions.[62,64,65,90,94,95,98,99] Whereas the weak electrostatic attachment of DNA to the mica surface obtained by adding the divalent cations Mg^{2+} into the buffer is efficient enough when imaging in air (Figure 5a−c), it is definitely too loose when scanning in liquid.[62,98] In aqueous buffer, Ni^{2+} ions are generally preferred because of their higher affinity for mica which allows a better confinement of DNA to the mica surface without either immobilizing it or releasing it in the liquid cell. In this section, we reproduced the persistence length measurements reported in the experimental results section when AFM imaging in liquid using the buffer L that contains $NiCl_2$ and no more $MgCl_2$ (see Experimental Methods) (Figure 8b). If we could reasonably anticipate some change in the persistence length estimate of the considered DNA molecules,[62,63,98,100]

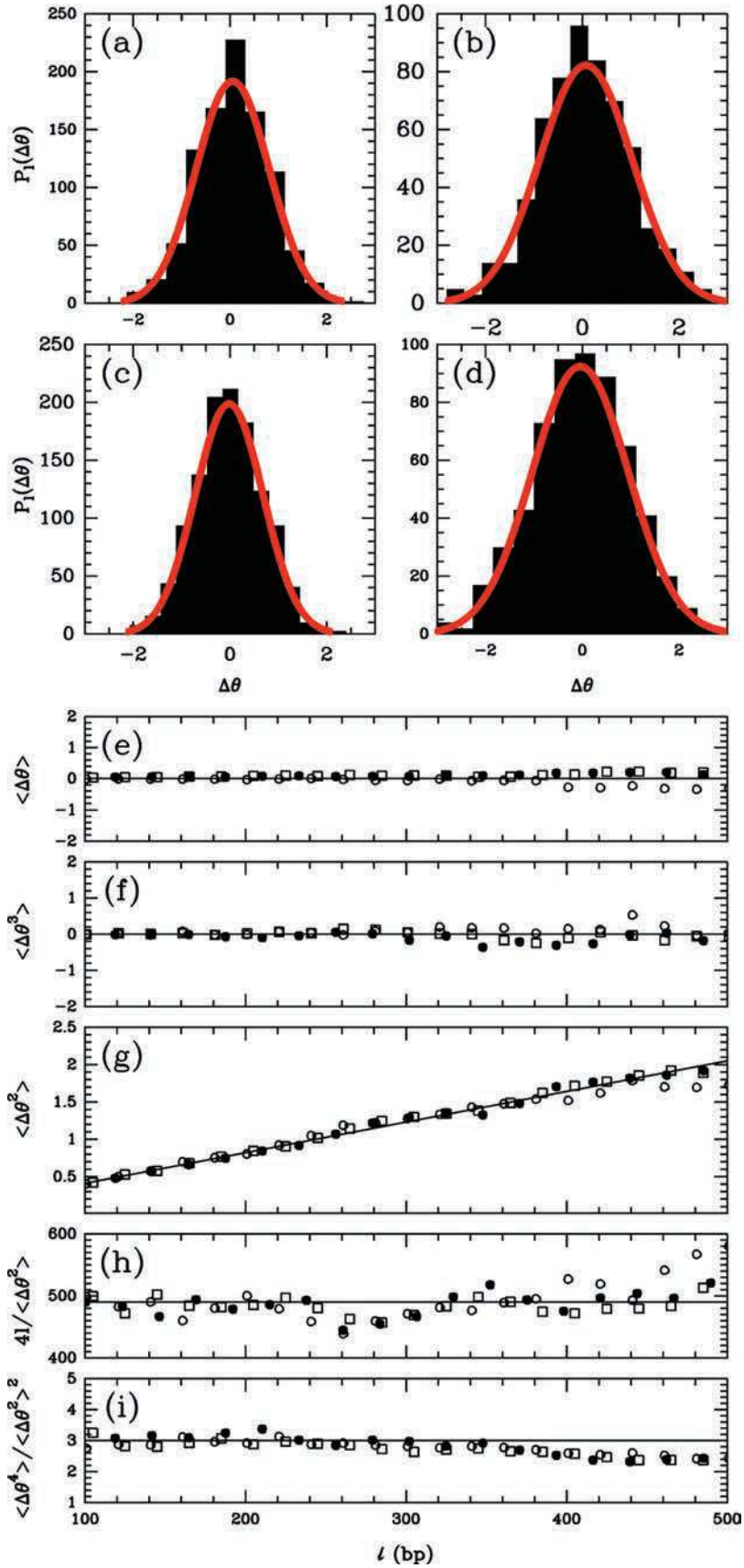

Figure 7. Statistical analysis of DNA bend angle fluctuations for the $N = 193$ intrinsically straight DNA molecules of length $L = 800$ bp imaged in air with buffer A1. Probability density function $P_l(\Delta\theta)$ estimated for different scales l : AFM experiment for $l = 230$ bp (a) and 370 bp (b); DNA simulations with $2A = 490$ bp and for $l = 230$ bp (c) and 370 bp (d). Moment analysis: (e) $\langle\Delta\theta\rangle$, (f) $\langle\Delta\theta^3\rangle$, (g) $\langle\Delta\theta^2\rangle$, (h) $2l/\langle\Delta\theta^2\rangle$, and (i) $\langle\Delta\theta^4\rangle/\langle\Delta\theta^2\rangle^2$ versus l; the symbols ● correspond to AFM experiment and ○ and □ to two DNA simulations, respectively. The horizontal solid lines in parts e, f, h, and i correspond to $\langle\Delta\theta_l\rangle = 0$, $\langle\Delta\theta_l{}^3\rangle = 0$, $2l/\langle\Delta\theta_l{}^2\rangle = 490$ bp, and $\langle\Delta\theta_l{}^4\rangle/\langle\Delta\theta_l{}^2\rangle^2 = 3$, respectively. The solid line in part g corresponds to the slope $2/2A$ with $2A = 490$ bp.

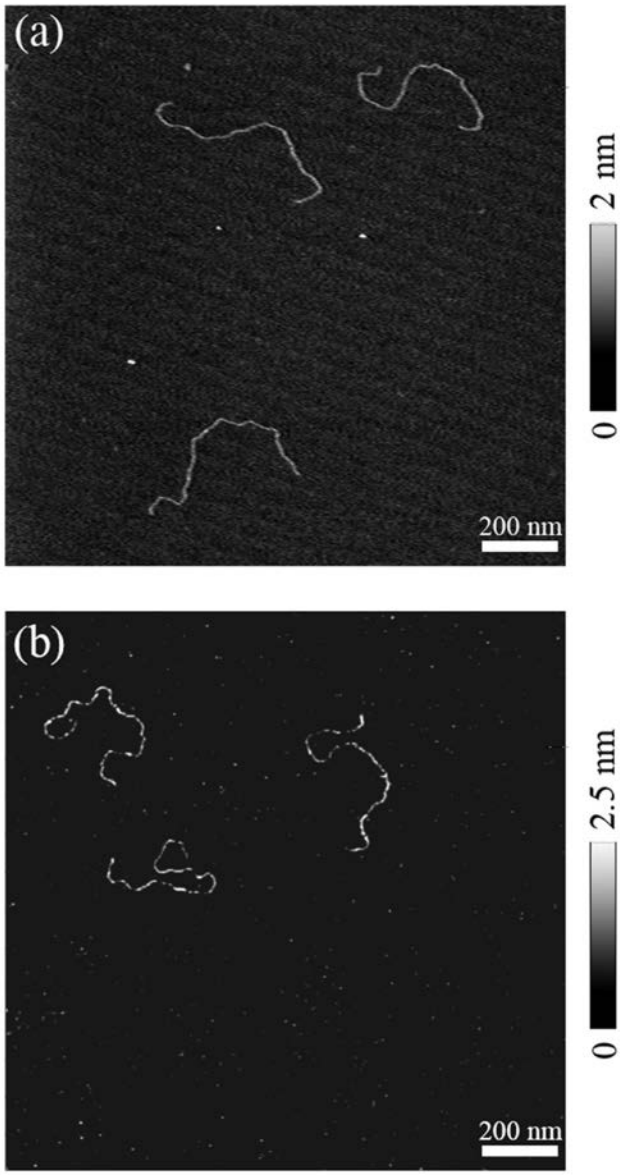

Figure 8. AFM images in tapping mode in air with buffer A3 (a) and in liquid with buffer L (b) of the uncorrelated HCV DNA ($L = 2200$ bp).

our aim was to discuss the validity and robustness of the conclusions drawn when operating in air, i.e., with the risk of some artifacts resulting from the drying of the sample.

In Figure 9a are reported the results of persistence length measurement in liquid for the $L = 800$ bp intrinsically straight DNA fragment. From a large set of AFM images, the DNA traces of $N = 109$ molecules were digitized as described in Experimental Methods. When dividing the measured average contour length by the total number of base-pairs, we got an experimental helical rise of 3.60 ± 0.05 Å·bp^{-1} that slightly overestimates the 3.38 ± 0.05 Å·bp^{-1} measured by crystallography. Note that this contrasting situation with respect to imaging in air (where some underestimate of the rise 3.20 ± 0.05 Å·bp^{-1} was observed) was already noticed in previous AFM studies in liquid.[61,64,65] As shown in Figure 9a for $\ell_p^{\mathcal{S}}(s)$ (similar results were obtained for $\ell_p^{\mathcal{D}}(s)$, data not shown), the experimental data issued from the mean projection of the end-to-end vector onto the initial orientation are again very well fitted by the WLC model prediction (13) when adjusting $\ell_p = 2A = 285 \pm 20$ bp, i.e., a value which is significantly smaller than the value $\ell_p = 490 \pm 30$ bp previously obtained in air (Figure 5a) using buffer A1. A similar result was obtained in Figure 9b when averaging AFM images of $N = 98$ HCV molecules (Figure 8b). Whatever the curvilinear distance s, the data points $\ell_p^{\mathcal{S}}(s)$ (●) fall significantly below the ones (△) previously obtained when imaging in air, and are still well reproduced by the WLC model eq 13 with an effective persistence length $\ell_p^{\text{eff}} = 2A^{\text{eff}} = 276 \pm 20$ bp, that is again much smaller than the value $\ell_p^{\text{eff}} = 440 \pm 30$ bp previously measured in air using buffer A1 (see Table 1).

Overall, the results obtained so far raise several interesting comments. According to the AFM imaging protocol, we got different estimates of the persistence length of DNA chains. In particular, when imaging in liquid using buffer L, we were no longer able to distinguish the HCV from the intrinsically straight DNA molecules, since both were found to have a persistence length of $\ell_p \sim 280$ bp which corresponds to a 3D persistence length of $\ell_p^{3D} = (\ell_p/2) \times 0.36 \simeq 50.5$ nm in good agreement with the 50 nm consensus value obtained for native DNA by independent methods.[22,33,73,82] When plugging $\ell_p^{d} = \ell_p^{\mathcal{S}}$ (straight DNA) = 285 bp and $\ell_p^{\text{eff}} = \ell_p^{\mathcal{S}}$ (HCV) = 276 bp into eq 19, we get an estimate of the curvature angle fluctuation amplitude $\sigma_o = 0.015 \pm 0.034$, as compared to the value 0.022 ± 0.012 previously obtained when imaging in air with buffer A1 (Table 1). The prohibitive size of the error bars explains why the effect of the sequence on the DNA elastic properties was found negligible in many experimental studies performed in quasi-physiological conditions. However, a careful examination of the data in Figure 9 reveals that if the $\ell_p^{\mathcal{S}}(s)$ data points for the intrinsically straight DNA (Figure 9a) and the HCV DNA (Figure 9b) fall in the error bars of each other, nevertheless for each s value, $\ell_p^{\mathcal{S}}(s)$ for HCV molecules is found systematically smaller than for the intrinsically straight DNA molecules. This strongly suggests that, despite the statistical (and experimental) uncertainty, the systematic underestimation of the persistence length observed for HCV relative to intrinsically straight DNA is an indication of the presence of an uncorrelated intrinsic curvature disorder in the former molecules.

Persistence Length Measurement by AFM under Different Ionic Conditions. In order to avoid any misunderstanding of the results reported in the previous section when performing AFM imaging in liquid with buffer L, we reproduced the same persistence length AFM measurement in air using the same buffer, renamed buffer A3 (see Experimental Methods), that contains Ni^{2+} divalent cations instead of Mg^{2+} as previously experienced in air with buffer A1 (Figure 5). As reported in Figure 9, the data for $\ell_p^{\mathcal{S}}(s)$ obtained from the analysis of $N = 101$ intrinsically straight DNA molecules and $N = 71$ HCV molecules cannot be distinguished, up to the statistical and experimental uncertainties, from the corresponding data previously obtained in liquid. Both sets of data are again very well fitted by the WLC prediction (13) with $\ell_p = 2A = 297$ bp for the former and $\ell_p^{\text{eff}} = 291$ bp for the latter, i.e., values that are very close to the estimates $\ell_p = 285$ bp and $\ell_p^{\text{eff}} = 276$ bp previously obtained in liquid. These results show that, when using the same experimental deposition procedure with the same buffer composition, AFM imaging in air and in liquid provides consistent and robust estimates of the persistence length of short DNA chains. The persistence length difference observed in our previous AFM studies was actually the consequence of the use of different ionic conditions and in particular of the change from Mg^{2+} ions in buffer A1 to Ni^{2+} ions in buffer A3 that are known[62,63,98] to be able to bind to DNA major or minor grooves,

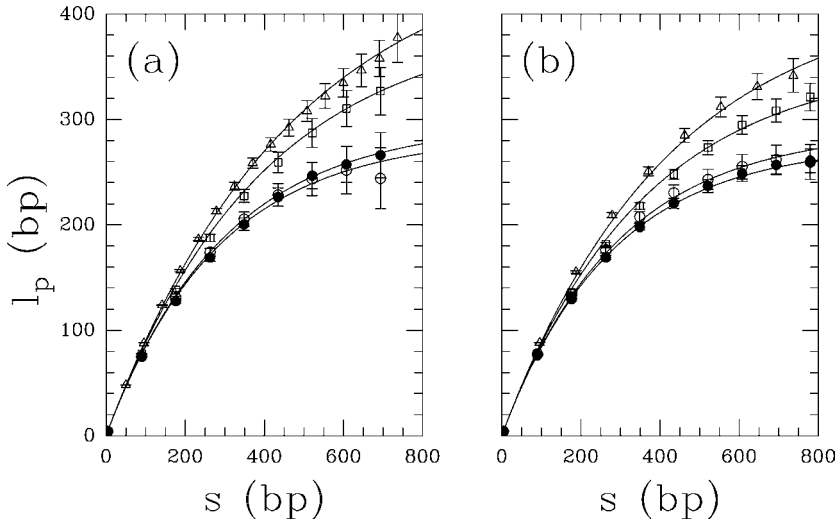

Figure 9. Persistence length measurement of $\ell_p^{\int}$ when averaging over N imaged molecules of (a) intrinsically straight DNA ($L = 800$ bp) and (b) uncorrelated HCV DNA ($L = 2200$ bp). The symbols correspond to the following AFM imaging protocols: air with buffer A1 (△), A2 (□), and A3 (○); liquid with buffer L (●). The solid lines correspond to least-squares fit estimates of the persistence length with the WLC model prediction (13). The so-obtained estimates of ℓ_p and the corresponding numbers of analyzed molecules are reported in Table 1.

TABLE 1: 2D Persistence Length Measurement from the Mean Projection of the End-to-End Vector onto the Initial Orientation of *N* DNA Molecules Deposited onto Mica with Different Procedures[a]

	straight DNA $L = 800$ bp	HCV $L = 2200$ bp	
experimental protocol	$\ell_p(N)$	$\ell_p^{eff}(N)$	σ_o
Air			
buffer A1	490 ± 30 bp (193)	440 ± 30 bp (65)	0.022 ± 0.012
10 mM Tris−HCl,	157 ± 10 nm	141 ± 10 nm	
5 mM $MgCl_2$			
buffer A2	395 ± 25 bp (108)	360 ± 25 bp (99)	0.023 ± 0.015
10 mM Tris−HCl,	126 ± 8 nm	115 ± 8 nm	
5 mM $MgCl_2$,			
0.05 mM $NiCl_2$			
buffer A3	297 ± 20 bp (101)	291 ± 20 bp (71)	0.012 ± 0.038
10 mM Tris−HCl,	95 ± 6 nm	93 ± 6 nm	
0.5 mM $NiCl_2$			
Liquid			
buffer L	285 ± 20 bp (109)	276 ± 20 bp (98)	0.015 ± 0.034
10 mM Tris−HCl,	102 ± 6 nm	99 ± 6 nm	
0.5 mM $NiCl_2$			

[a] The persistence lengths ℓ_p ($=\ell_p^d$) and ℓ_p^{eff} were estimated from a least-squares fit of the corresponding $\ell_p^{\int}(s)$ data for the straight DNA (Figure 9a) and the HCV (Figure 9b) molecules, respectively, using the WLC model equation (13). $\sigma_o = \sqrt{2/\ell_p^s}$ was estimated using eq 19.

partially balancing the surface charge of the molecule, likely increasing its flexibility.

To confirm this interpretation, we performed additional persistence length measurements in air using buffer A2 that contained both Mg^{2+} and Ni^{2+} divalent counterions. More precisely, buffer A2 was obtained by adding 0.05 mM $NiCl_2$ to the 5 mM $MgCl_2$ containing buffer A1. The $\ell_p^{\int}(s)$ data recorded with buffer A2 (□) fall systematically in between the corresponding experimental points previously obtained with buffers A1 (△) and A3 (○) and this for both sets of $N = 108$ intrinsically straight DNA molecules (Figure 9a) and $N = 99$ HCV molecules (Figure 9b). From the very good fits obtained with the WLC eq 13, we got the following estimates $\ell_p = 395 \pm 25$ bp for the former molecules and $\ell_p = 360 \pm 25$ bp for the latter ones, i.e., values that are intermediate between the corresponding ones obtained with buffers A1 and A3 (Table 1). However, what is remarkable is that when introducing the so-obtained values in eq 19, identifying the dynamic persistence length ℓ_p^d to ℓ_p (straight DNA) and the effective persistence length ℓ_p^{eff} to ℓ_p (HCV), we get a value for the intrinsic curvature fluctuation amplitude $\sigma_o \simeq 0.023 \pm 0.015$ which is quantitatively similar to the one obtained in air with buffer A1 ($\sigma_o = 0.022 \pm 0.012$) (see Table 1). Thus, despite the relative large error bars, especially as far as the value $\sigma_o = 0.012 \pm 0.038$ (respectively, 0.015 ± 0.034) obtained with buffer A3 (respectively, L), the different persistence length estimates collected under different buffer conditions, when AFM imaging in air as well as in solution, are all consistent with the existence of an uncorrelated intrinsic curvature disorder of amplitude $\sigma_o \simeq 0.02$ in the HCV molecules. The variations observed in the effective persistence length ℓ_p^{eff} of these HCV molecules when changing the ionic conditions (Figure 9b) are likely to be accounted for by the variations of the dynamic persistence length ℓ_p^d contribution in eq 19, as estimated from the set of intrinsically straight DNA molecules (Figure 9a). As summarized in Table 1, when progressively replacing Mg^{2+} by Ni^{2+} cations, $\ell_p^d = 2A$ decreases from 490 ± 30 bp (buffer A1) to 290 ± 20 bp (buffers A3 and L), which corresponds to a decrease of the DNA bending rigidity A and in turn of the 3D persistence length ℓ_p^{3D} (eq 20) from 79 nm down to 49 nm. This increase in

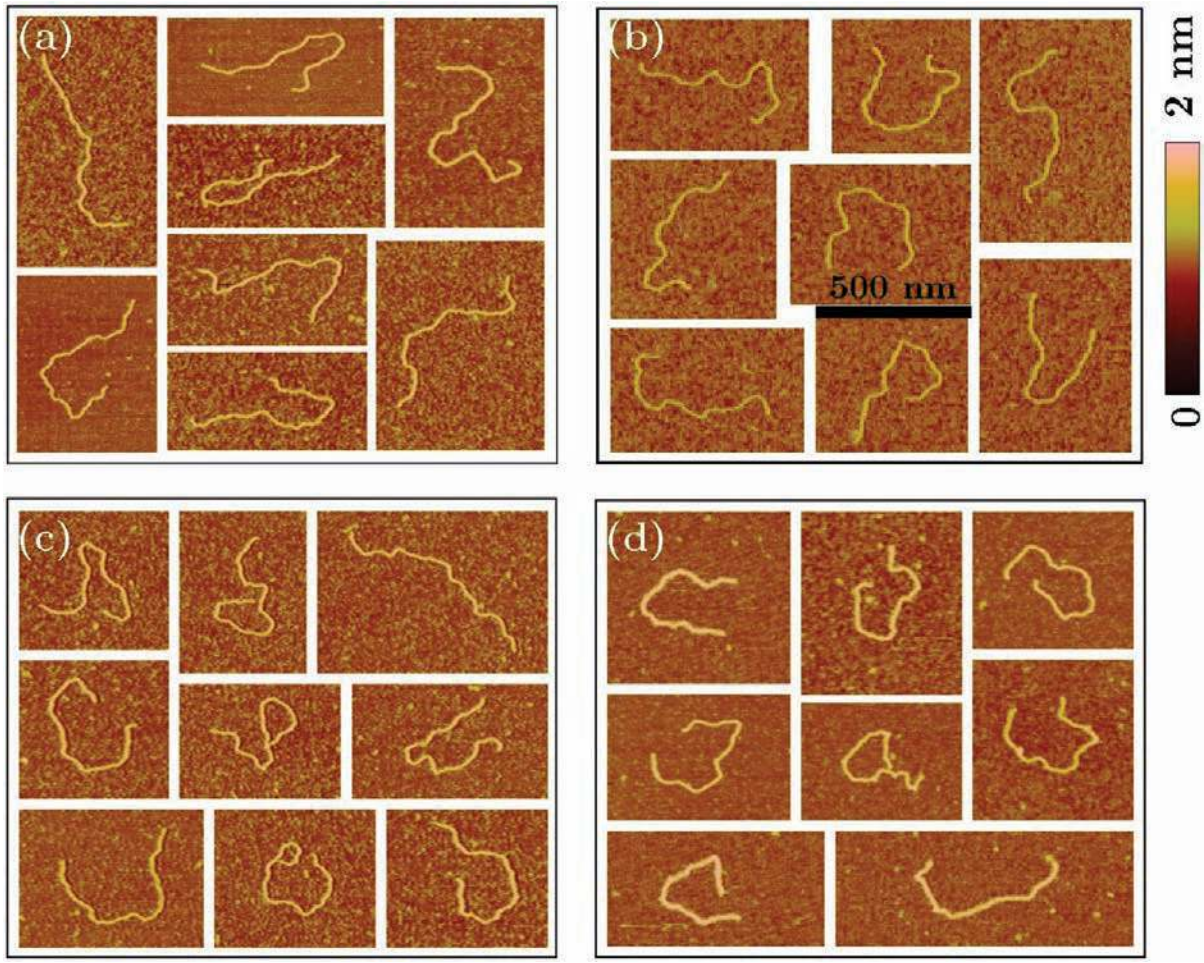

Figure 10. AFM images in tapping mode in air using buffer A1 of human DNA fragments of different G+C contents. (a) chr 21: $L = 2202$ bp, G+C = 31%; (b) chr21: $L = 2206$ bp, G+C = 38%; (c) chr 21: $L = 2206$ bp, G+C = 43%; (d) chr 8: $L = 2190$ bp, G+C = 46%.

flexibility is discernible when comparing HCV molecule traces imaged in air using buffer A2 (Figure 8a) and in liquid using buffer L (Figure 8b). It is likely to result from some charge-induced electrostatic softening of the DNA due to the high affinity of Ni^{2+} ions to DNA that can attenuate repulsions between negative charges on the phosphates.[62,63,98,100] However, what our AFM results suggest is that these changes in ionic conditions have not affected the static persistence length ℓ_p^s of the HCV molecules, which quantifies the intrinsic curvature disorder induced by the HCV DNA sequence with an amplitude as small as $\sigma_o = (2/\ell_p^s)^{1/2} \simeq 0.02$ (Table 1).

The consistency of the AFM results reported in Figure 9 and Table 1 brings into light a remark of practical importance. Since the weak curvature structural disorder evidenced for the HCV molecules $\sigma_o \simeq 0.02$ at the base-pair scale is likely to be typical of native DNA molecules, according to eq 19, the best situation to estimate $\sigma_o^2 = 2(1/\ell_p^{eff} - 1/\ell_p^d)$ as accurately as possible is when the two terms $1/\ell_p^{eff}$ and $1/\ell_p^d$ are both as small as their difference, i.e., when ℓ_p^{eff} and mainly $\ell_p^d = 2A$ are as large as possible. In other words, if one intends to characterize the intrinsic structural disorder induced by the DNA sequence and possibly evidence the presence of LRC, it is better to operate under deposition and AFM imaging conditions such that the considered DNA molecule is as rigid as possible (ideally, for uncorrelated disorder, $\sigma_o^2 = 2/\ell_p^{eff}$ in the limit $\ell_p^d = 2A \rightarrow +\infty$). This justifies *a posteriori* our choice in the Experimental Results section to operate in air using the AFM imaging protocol with buffer A1 (Table 1). This practical choice will be further pursued in the following section.

Experimental Demonstration of the Existence of Long-Range Correlations in (G+C)-Rich Human DNA Molecules. In Figure 5c′, we showed that when following the *a priori* most favorable protocol, namely, AFM imaging in air with buffer A1 (see Experimental Methods), the experimental AFM data for the persistence lengths $\ell_p^{\mathcal{D}}(s)$ and $\ell_p^{\int}(s)$ of human genomic DNA molecules were well fitted by the generalized WLC model (eqs 23 and 24) for the parameter values: $2A = 490$ bp consistently to the bend rigidity found for the intrinsically straight DNA, $H = 0.73$, in good agreement with the LRC exponent value $H = 0.73 \pm 0.02$ found homogeneously along human chromosomes from wavelet analysis of structural bending profiles[43,44] (data not shown) and $\sigma_o = 0.007$.[97] However, the classical WLC model predictions (eqs 10 and 13) also provide an acceptable fit of the persistence length data (Figure 5c′) when fixing $2A^{eff} = 410 \pm 10$ bp, which corresponds, according to eq 21, to a larger value of $\sigma_o = 0.028$. Actually, a close inspection of the behavior of $\ell_p^{\int}(s)$ in the range of scales 600 bp $\lesssim s \lesssim$ 1000 bp reveals that the WLC model indeed fails to account for the observed pronounced "knee" behavior hastening the convergence of $\ell_p^{\int}(s)$. As illustrated in Figure 3 and further discussed in the theoretical section, this faster convergence to a lower value of ℓ_p is likely to be the signature of the presence of LRC in the local DNA curvature fluctuations.[50,58] Unfortunately, for rather weak structural disorder with $\sigma_o \lesssim 0.01$, as observed for the human genomic DNA molecules, this "knee" behavior in $\ell_p(s)$ is not sufficiently pronounced to allow us to definitely discard the WLC model.

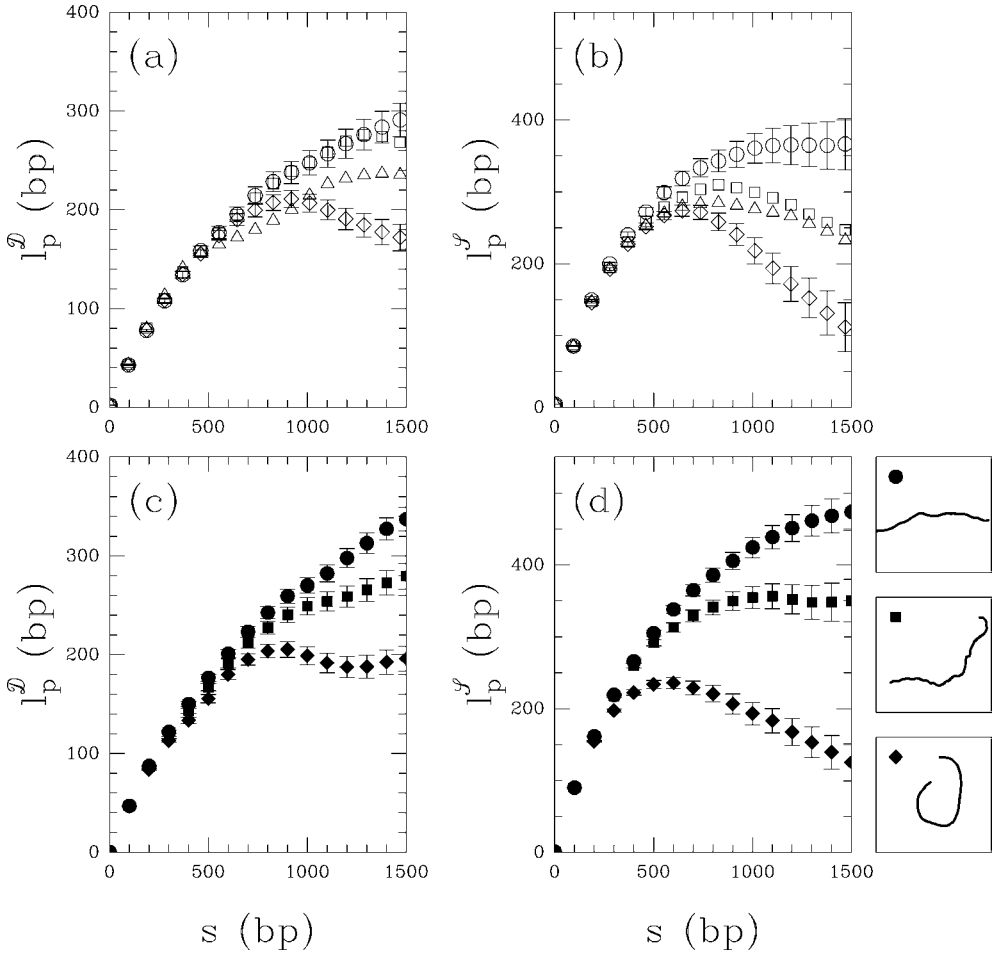

Figure 11. Persistence length measurement of $\ell_p^{\mathcal{D}}$ and $\ell_p^{\mathcal{S}}$ when averaging over N DNA traces. (a) $\ell_p^{\mathcal{D}}(s)$ and (b) $\ell_p^{\mathcal{S}}(s)$ measured from AFM study in air with buffer A1 of human DNA fragments of G+C content (○) 31% ($N = 99$), (△) 38% ($N = 80$), (□) 43% ($N = 87$), and (◇) 46% ($N = 134$). (c) $\ell_p^{\mathcal{D}}(s)$ and (d) $\ell_p^{\mathcal{S}}(s)$ computed from DNA simulations of $N = 200$ 2D equilibrium conformations generated with an elastic bend rigidity of $2A = 490$ bp from the three illustrated frozen LRC chains (●, ■, ◆) illustrated at the right edge of part d and selected as representatives of increasingly macroscopically curved DNA chains generated with $H = 0.73$ and $\sigma_o = 0.01$.

Thankfully, the human genome is well-known to be compartmentalized into wide (~10^5 bp) domains with uniform G+C content, commonly called isochores,[101–103] with appreciable scatter of the average G+C content when comparing different domains. One of the main isochore characteristics is the fact that the variance of the G+C fluctuations increases significantly with the mean G+C content. As noticed in a previous work,[59] when using structural trinucleotide coding tables like the PNuc table, the variance of the generated bending profiles is mainly governed by the variance of the G+C content. More precisely, when comparing G+C-poor (~30%) to G+C-rich (~55%) isochores, the variance σ_o^2 of bend angle fluctuations can be multiplied by a factor of 2 or more. As described in Experimental Methods, along this observation, we selected four repeated-sequence free DNA fragments of size $L \simeq 2200$ bp along the human genome with G+C content ranging from 31 to 46%. Three of these sequences are from chromosome 21, with the last one from chromosome 8. Prior to AFM experiments, we performed a wavelet-based LRC analysis of the corresponding DNA bending profiles obtained when using the PNuc trinucleotide coding table as previously reported in refs 43 and 44. The four human DNA fragments displayed nice scaling behavior characteristic of LRC ($H > 1/2$), in particular $H = 0.73 \pm 0.03$ for scales greater than 200 bp (data not shown). Note that this estimate was confirmed by extending the sequence 30 kbp upstream and downstream of the selected fragments.

Figure 10 shows AFM images of typical 2D equilibrium conformations obtained in air with buffer A1 for these four selected human DNA fragments with respective G+C content of 31% (Figure 10a), 38% (Figure 10b), 43% (Figure 10c), and 46% (Figure 10d). Figure 11a and b show the corresponding experimental data for $\ell_p^{\mathcal{D}}(s)$ and $\ell_p^{\mathcal{S}}(s)$, respectively, as obtained when averaging over statistical samples of about $N \simeq 100$ molecules. Several observations must be emphasized. First, the $\ell_p^{\mathcal{D}}(s)$ and $\ell_p^{\mathcal{S}}(s)$ data for the DNA fragment of lowest G+C content (31%) are very similar to the ones previously obtained for the human genomic DNA molecules in Figure 5c′. According to our generalized WLC model[58] (with $2A = 490$ bp, $H = 0.73$, $\sigma_o = 0.007$), we thus predict for the low G+C human DNA fragment the same persistence length $\ell_p = 396$ bp corresponding to a 3D persistence length $\ell_p^{3D} = \ell_p/2 \times 0.32 = 63$ nm, suggesting that our experimental sample of $N = 290$ human genomic DNA molecules is typical of the rather low average G+C content (~41%) observed in the human genome. For the other DNA fragments with higher G+C contents, both $\ell_p^{\mathcal{D}}(s)$ and $\ell_p^{\mathcal{S}}(s)$ are found significantly lower than for the HCV (Figure 5b′) and the human genomic DNA (Figure 5c′). Indeed, for each of these single DNA fragments, the observed behavior of $\ell_p^{\mathcal{D}}(s)$ and $\ell_p^{\mathcal{S}}(s)$ versus s is no longer comparable to the predictions (10) and (13) of the WLC model. This is particularly apparent for the G+C = 46% human DNA fragment for which

both $\ell_p^{\mathscr{D}}(s)$ and $\ell_p^{\mathscr{S}}(s)$ start decreasing beyond $s \simeq 700-800$ bp as a possible indication of the presence of some large-scale cooperative intrinsic curvature. A simple visual inspection of the AFM images of these human DNA fragments in Figure 10 confirms this interpretation. Whereas molecules can still adopt bent or straighter HCV-like configurations indicating that their flexibility is likely to be retained, some proportion of them displays some macroscopic curvature. However, what is remarkable is that this proportion becomes very significant for the DNA fragment of highest G+C content (Figure 10d). To demonstrate that this persistence in curvature properties is not due to some "hyperperiodic" distribution of curved DNA fragments[104] but rather to the presence of LRC in the random distribution of bend sites, we performed DNA simulations of LRC chains with parameters $2A = 490$ bp, $H = 0.73$, and σ_o ranging from 0.006 to 0.015. We selected three frozen LRC chains that are illustrated at the right edge of Figure 11d as representative of increasing intrinsically macroscopically curved DNA chains. The $\ell_p^{\mathscr{D}}(s)$ and $\ell_p^{\mathscr{S}}(s)$ numerical data obtained when averaging over $N = 100$ 2D equilibrium conformations of these three frozen chains are shown in Figure 11c and d, respectively. These numerical data display the same characteristics as experimentally observed in Figure 11a and b for the four human DNA fragments. Going from the less intrinsically curved synthetic DNA to the most curved one, we recover the WLC-like behavior observed for the G+C-poor DNA as well as the decrease of $\ell_p^{\mathscr{D}}(s)$ and $\ell_p^{\mathscr{S}}(s)$ observed at large s values ($s \gtrsim 800$ bp) for the G+C-rich DNA fragments.

Now, if we look back to the DNA simulations of uncorrelated and LRC DNA chains in Figure 4a and b, we recognize that the frozen LRC chains display some marked macroscopic curvature. Actually, from the Gaussian statistics of intrinsic bend angle fluctuations (eq 16), we can show that the characteristic size of a frozen LRC chain making a half-loop is $\ell^* = (\pi/\sigma_o)^{1/H}$; this yields for $H = 0.73$ the values $\ell^* = 4295$ and 895 bp for $\sigma_o = 0.007$ and 0.022, respectively. Since the internal end-to-end distance is expected to decrease when s reaches half the loop size, this explains that the decrease observed for $s \gtrsim 800$ bp in $\ell_p^{\mathscr{D}}(s)$ and $\ell_p^{\mathscr{S}}(s)$ of G+C-rich DNA fragments is quite representative of the presence of LRC between randomly distributed curvature sites. Actually, in order to get a similar decrease with an uncorrelated ($H = 1/2$) frozen chain, we would need a quite unrealistic large value of $\sigma_o = (\pi^2/\ell^*)^{1/2} \simeq (\pi^2/895)^{1/2} \simeq 0.105$. With a more realistic value of $\sigma_o \simeq 0.02-0.03$, the probability to get by chance an uncorrelated chain that displays a macroscopic curvature similar to the one observed for the G+C-rich DNA fragments is less than 10^{-3}, which again discards a possible interpretation of the AFM data in Figures 10 and 11 in terms of the WLC model. Finally, for $\sigma_o = 0.010$, 0.012, and 0.015, our generalized WLC model[58] for $2A = 490$ bp and $H = 0.73$ yields the following respective asymptotic estimates of the persistence length of G+C-rich DNA fragments: $\ell_p = 339$, 306, and 262 bp. These estimates of ℓ_p correspond to 3D persistence length values $\ell_p^{3D} = 54$, 49, and 42 nm that are consistent with the 50 nm consensus values for native DNA obtained from previous measurements by independent methods.[22,23,73,82]

Conclusion

The persistence length of DNA molecules is indeed sensitive to many experimental parameters like temperature, pH, ionic forces, and excluded volume effects for long DNA fragments, that can explain significant discrepancies between ℓ_p values estimated by different groups using various techniques including light scattering,[105,106] ligase-catalyzed cyclization,[60,107−110] flow dichroism,[111] transient electric birefringence,[112,113] and transient electric dichroism.[114] However, all of these methods require simplifying theoretical models to interpret the data that may affect the results. Usually, DNA is assumed to be a classical intrinsically straight polymer and the local sequence-dependent variations in local curvature and flexibility are neglected; this can significantly affect the efficiency of circularization in biochemical cyclization experiments.[60,107−110] As aforementioned, electron microscopy[115] and atomic force microscopy[61,64,65,87−95] are unique techniques that provide essential information to persistence length measurement, namely, the contour of the individual molecules and a population of such contours. However, these direct visualization techniques require the DNA molecules to be confined in 2D for imaging which can cause some changes in the preferred DNA conformations. As emphasized by Bednar et al.,[82] cryo-electron microscopy overcomes the drawback of classical electron microscopy, since there is no adsorption, no drying, and no staining artifact, the molecules being suspended in a thin layer of cryo-vitrified buffer. As far as AFM is concerned, two extreme cases of DNA molecule adsorption onto supporting films can be distinguished:[61−63] (i) strong kinetic trapping, roughly corresponding to a projection of the 3D conformations, and (ii) weak adsorption so that the molecules have time to freely equilibrate on the surface before being trapped in a particular 2D equilibrium conformation.

In the present study, we have used AFM in tapping mode in air and in liquid to image DNA molecules under different ionic conditions. If the different deposition protocols led to different estimates of the persistence length, they all provided consistent estimates of the so-called static persistence length that quantifies the amplitude (σ_o) of the intrinsic curvature disorder induced by the sequence. The fact that the less flexible the DNA chain, the more accurate the measurement of σ_o led us to adopt a protocol[50,61−63] which allowed imaging in air and which consists of the weak adsorption of DNA molecules onto untreated mica in the presence of divalent ions Mg^{2+}. The ions act as bridges between the negative mica surface and DNA charges; some fine-tuning of the concentration of these cations allows for 2D relaxation. We tested this protocol by checking the Gaussian statistics of bend angle fluctuations of a synthetically designed intrinsically straight DNA fragment of length $L = 800$ bp. Consistently, persistence length measurements were found in good agreement with the WLC model predictions. From the 2D persistence length $\ell_p = 490 \pm 30$ bp, we deduced the value of the elastic bend rigidity $A = 245 \pm 15$ bp and, according to eq 20, the value of the 3D persistence length $\ell_p^{3D} = \ell_p/2 = A \times 0.32 = 79 \pm 4$ nm, in good agreement with the Bednar et al. estimate (80 nm) for similar molecules by cryo-electron microscopy[82] and recent molecular dynamic simulation by Mazur.[116] As a reference AFM study of sequence effects on the structural and mechanical properties of DNA molecules, we first looked for a DNA sequence that did not display any LRC, in particular in the range of scales $\gtrsim$200 bp where most prokaryotic and eukaryotic sequences were shown[43,44] to exhibit universal LRC with $H \simeq 0.8$. This drastically restricted our choice to the family of single-stranded, positive sense RNA viruses with no DNA stage that were shown to display uncorrelated DNA walk and bending profiles.[44] When performing a comparative analysis of $L = 2200$ bp HCV molecules imaged by AFM and DNA simulations of intrinsically uncorrelated chains, we found that the corresponding persistence length data were still very well fitted by the WLC model predictions but with an effective persistence length $\ell_p^{eff} = 440 \pm 30$ bp corresponding to a 3D persistence length $\ell_p^{3Deff} = 70 \pm 4$ nm significantly smaller

than the dynamic persistence length previously estimated for the straight DNA. Under the reasonable assumption that, with the same experimental deposition protocol, the elastic bend rigidity A should be the same for all molecules, by applying Trifonov et al. formula 19, we obtained an estimate of the static persistence length $\ell_p^s = 4310$ bp (= 1380 nm) that corresponds to a rather small bend angle fluctuation amplitude $\sigma_o = (2/\ell_p^s)^{1/2} = 0.022$, typical of a weak structural disorder. Consistent values of σ_o were obtained when using different buffer compositions that allowed quantitative estimates (Table 1). Together, our results show that the contribution of the intrinsic curvature to the persistence length is at least 10 times smaller than the contribution of thermal fluctuations, a result which is in good agreement with cyclization efficiency measurements by Vologodskaia and Vologodskii.[60]

When using the same deposition protocol to image human genomic DNA molecules obtained by standard *Dra*I digestion (see Experimental Methods), we observed a further decrease of the persistence length $\ell_p = 396$ bp corresponding to $\ell_p^{3D} \simeq 62$ nm, that could be well reproduced by the generalized WLC model that accounts for the LRC ($H = 0.73$) observed in human DNA sequences. These results provide a strong indication that this increase in bending does not result from some enhancing of bending flexibility but rather from some large-scale intrinsic curvature induced by a sequence-dependent persistent distribution of local bending sites. Additional AFM imaging assisted by DNA simulations of $L = 2200$ bp human DNA fragments confirmed that, with increasing G+C content, a larger and larger proportion of DNA molecule conformations exhibited a macroscopic curvature contributing to an even more significant lowering of the persistence length down to values $\ell_p \simeq 310$ bp consistent with the $\ell_p^{3D} \sim 50$ nm consensus value obtained for native DNA by independent methods.[50] Since both prokaryotic and eukaryotic DNA sequences were shown to display LRC,[43,44] the results of the present study show that the contribution of the intrinsic LRC curvature disorder to the DNA persistence length is more important than previously thought and is likely to be competitive with the entropic contribution of thermal fluctuations at least under commonly used ionic environments. This raises the issue of developing methodologies capable of separating the local sequence-induced curvature and flexibility along the DNA chain. The very promising results obtained in pioneering AFM experiments in solution[94,95] open new perspectives to a better understanding of the actual role of sequence-dependent DNA mechanical properties in the DNA−protein interactions involved in genome regulation and packaging.

Acknowledgment. We are very grateful to B. Bartosch, F.-L. Cosset, and A. Vologodskii for the kind gifts of plasmids. This work was supported by the Conseil Régional Rhône-Alpes (project "Le rôle de la séquence sur la structure et la dynamique de la chromatine", Emergence 2005), the Contrat Plan Etat Région "Nouvelle Approches Physiques des Sciences du Vivant" (2005), and the Agence Nationale de la Recherche "Programme Physique et Chimie du Vivant 2006" (Project DNAnucl "DNA sequence effects on the structure and dynamics of nucleosome", ANR-06-PCVI-0026).

References and Notes

(1) van Holde, K. E. *Chromatin*; Springer-Verlag: New York, 1988.
(2) Wolffe, A. P. *Chromatin Structure and Function*, 3rd ed.; Academic Press: London, 1998.
(3) Calladine, C. R.; Drew, H. R. *Understanding DNA*; Academic Press: San Diego, CA, 1999.
(4) Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Roberts, K.; Watson, J. D. *Molecular Biology of the Cell*, 3rd ed.; Garland Publishing: New York, 1994.
(5) Widom, J. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 285–327.
(6) Kornberg, R. D.; Lorch, Y. *Cell* **1999**, *98*, 285–294.
(7) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251–260.
(8) Richmond, T. J.; Davey, C. A. *Nature* **2003**, *423*, 145–150.
(9) Peterson, C. L. *FEBS Lett.* **2000**, *476*, 68–72.
(10) Vignali, M.; Hassan, A. H.; Neely, K. E.; Workman, J. L. *Mol. Cell. Biol.* **2000**, *20*, 1899–1910.
(11) Becker, P. B. *EMBO J.* **2002**, *21*, 4749–4753.
(12) Felsenfeld, G.; Groudine, M. *Nature* **2003**, *421*, 448–453.
(13) Montel, F.; Fontaine, E.; St-Jean, P.; Castelnovo, M.; Faivre-Moskalenko, C. *Biophys. J.* **2007**, *93*, 566–578.
(14) Meersseman, G.; Pennings, S.; Bradbury, E. M. *EMBO J.* **1992**, *11*, 2951–2959.
(15) Flaus, A.; Richmond, T. J. *J. Mol. Biol.* **1998**, *275*, 427–441.
(16) Schiessel, H. *J. Phys.: Condens. Matter* **2003**, *15*, R699–R774.
(17) Flaus, A.; Owen-Hughes, T. *Biopolymers* **2003**, *68*, 563–578.
(18) Schiessel, H. *Eur. Phys. J. E* **2006**, *19*, 251–262.
(19) Schiessel, H.; Widom, J.; Bruinsma, R. F.; Gelbart, W. M. *Phys. Rev. Lett.* **2001**, *86*, 4414–4417.
(20) Kulić, I. M.; Schiessel, H. *Phys. Rev. Lett.* **2003**, *91*, 148103.
(21) Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T. *Cold Spring Harbor Symp. Quant. Biol.* **1983**, *47* (1), 279–283.
(22) Hagerman, P. J. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 265–286.
(23) Crothers, D. M.; Haran, T. E.; Nadeau, J. G. *J. Biol. Chem.* **1990**, *265*, 7093–7096.
(24) Rando, O. J.; Ahmad, K. *Curr. Opin. Cell Biol.* **2007**, *19*, 250–256.
(25) Liu, X.; Lee, C. K.; Granek, J. A.; Clarke, N. D.; Lieb, J. D. *Genome Res.* **2006**, *16*, 1517–1528.
(26) Simpson, R. T.; Stafford, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 51–55.
(27) Widom, J. *Nature* **1984**, *309*, 312–313.
(28) Travers, A. A.; Klug, A. *Philos. Trans. R. Soc. London, Ser. B* **1987**, *317*, 537–561.
(29) Satchwell, S. C.; Drew, H. R.; Travers, A. A. *J. Mol. Biol.* **1986**, *191*, 659–675.
(30) Ioshikhes, I.; Bolshoy, A.; Derenshteyn, K.; Borodovsky, M.; Trifonov, E. N. *J. Mol. Biol.* **1996**, *262*, 129–139.
(31) Widlund, H. R.; Cao, H.; Simonsson, S.; Magnusson, E.; Simonsson, T.; Nielsen, P. E.; Kahn, J. D.; Crothers, D. M.; Kubista, M. *J. Mol. Biol.* **1997**, *267*, 807–817.
(32) Stein, A.; Bina, M. *Nucleic Acids Res.* **1999**, *27*, 848–853.
(33) Thaström, A.; Lowary, P. T.; Widlund, H. R.; Cao, H.; Kubista, M.; Widom, J. *J. Mol. Biol.* **1999**, *288*, 213–229.
(34) Trifonov, E. N.; Sussman, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 3816–3820.
(35) Shrader, T. E.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7418–7422.
(36) Widom, J. *Q. Rev. Biophys.* **2001**, *34*, 269–324.
(37) Cohanim, A. B.; Kashi, Y.; Trifonov, E. N. *J. Biomol. Struct. Dyn.* **2005**, *22*, 687–694.
(38) Lowary, P. T.; Widom, J. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1183–1188.
(39) Yuan, G. C.; Liu, Y. J.; Dion, M. F.; Slack, M. D.; Wu, L. F.; Altschuler, S. J.; Rando, O. J. *Science* **2005**, *309*, 626–630.
(40) Lee, W.; Tillo, D.; Bray, N.; Morse, R. H.; Davis, R. W.; Hughes, T. R.; Nislow, C. *Nat. Genet.* **2007**, *39*, 1235–1244.
(41) Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thaström, A. C.; Field, Y.; Moore, I. K.; Wang, J. P. Z.; Widom, J. *Nature* **2006**, *442*, 772–778.
(42) Ioshikhes, I. P.; Albert, I.; Zanton, S. J.; Pugh, B. F. *Nat. Genet.* **2006**, *38*, 1210–1215.
(43) Audit, B.; Thermes, C.; Vaillant, C.; d'Aubenton Carafa, Y.; Muzy, J. F.; Arneodo, A. *Phys. Rev. Lett.* **2001**, *86*, 2471–2474.
(44) Audit, B.; Vaillant, C.; Arneodo, A.; d'Aubenton-Carafa, Y.; Thermes, C. *J. Mol. Biol.* **2002**, *316*, 903–918.
(45) Arneodo, A.; Bacry, E.; Graves, P. V.; Muzy, J. F. *Phys. Rev. Lett.* **1995**, *74*, 3293–3296.
(46) Arneodo, A.; d'Aubenton-Carafa, Y.; Bacry, E.; Graves, P. V.; Muzy, J. F.; Thermes, C. *Physica D* **1996**, *96*, 291–320.
(47) Peng, C. K.; Buldyrev, S. V.; Goldberger, A. L.; Havlin, S.; Sciortino, F.; Simons, M.; Stanley, H. E. *Nature* **1992**, *356*, 168–170.
(48) Li, W.; Kaneko, K. *Europhys. Lett.* **1992**, *17*, 655–660.
(49) Voss, R. F. *Phys. Rev. Lett.* **1992**, *68*, 3805–3808.
(50) Moukhtar, J.; Fontaine, E.; Faivre-Moskalenko, C.; Arneodo, A. *Phys. Rev. Lett.* **2007**, *98*, 178101.

(51) Vaillant, C.; Audit, B.; Arneodo, A. *Phys. Rev. Lett.* **2005**, *95*, 068101.
(52) Vaillant, C.; Audit, B.; Thermes, C.; Arneodo, A. *Eur. Phys. J. E* **2006**, *19*, 263–277.
(53) Vaillant, C.; Audit, B.; Arneodo, A. *Phys. Rev. Lett.* **2007**, *99*, 218103.
(54) Miele, V.; Vaillant, C.; d'Aubenton Carafa, Y.; Grange, T. *Acids Res.* **2008**, *36*, 3746–3756.
(55) Grossberg, A. Y.; Khoklov, A. R. In *Statistical Physics of Macromolecules, AIP series in Polymers and Complex Materials*; Larson, R., Pincus, P. A., Eds.; AIP Press: Woodbury, NY, 1994.
(56) Kratky, O.; Porod, G. *Recl.: J. R. Neth. Chem. Soc.* **1949**, *68*, 1106.
(57) Schellman, J. A. *Biopolymers* **1974**, *13*, 217–226.
(58) Moukhtar, J.; Vaillant, C.; Audit, B.; Arneodo, A. *Europhys. Lett.* **2009**, *86*, 48001.
(59) Vaillant, C.; Audit, B.; Thermes, C.; Arneodo, A. *Phys. Rev. E* **2003**, *67*, 032901.
(60) Vologodskaia, M.; Vologodskii, A. *J. Mol. Biol.* **2002**, *317*, 205–213.
(61) Rivetti, C.; Guthold, M.; Bustamante, C. *J. Mol. Biol.* **1996**, *264*, 919–932.
(62) Pastré, D.; Piétrement, O.; Fusil, S.; Landousy, F.; Jeusset, J.; David, M.-O.; Hamon, L.; Le Cam, E.; Zozime, A. *Biophys. J.* **2003**, *85*, 2507–2518.
(63) Sushko, M. L.; Shluger, A. L.; Rivetti, C. *Langmuir* **2006**, *22*, 7678–7688.
(64) Rivetti, C.; Codeluppi, S. *Ultramicroscopy* **2001**, *87*, 55–66.
(65) Sanchez-Sevilla, A.; Thimonier, J.; Marilley, M.; Rocca-Serra, J.; Barbet, J. *Ultramicroscopy* **2002**, *92*, 151–158.
(66) Valle, F.; Favre, M.; De Los Rios, P.; Rosa, A.; Dietler, G. *Phys. Rev. Lett.* **2005**, *95*, 158105.
(67) Cluzel, P.; Lebrun, A.; Heller, C.; Lavery, R.; Viovy, J. L.; Chatenay, D.; Caron, F. *Science* **1996**, *271*, 792–794.
(68) Smith, S. B.; Cui, Y.; Bustamante, C. *Science* **1996**, *271*, 795–799.
(69) Strick, T. R.; Allemand, J. F.; Bensimon, D.; Bensimon, A.; Croquette, V. *Science* **1996**, *271*, 1835–1837.
(70) Wang, M. D.; Yin, H.; Landick, R.; Gelles, J.; Block, S. M. *Biophys. J.* **1997**, *72*, 1335–1346.
(71) Allemand, J. F.; Bensimon, D.; Lavery, R.; Croquette, V. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14152–14157.
(72) Léger, J. F.; Romano, G.; Sarkar, A.; Robert, J.; Bourdieu, L.; Chatenay, D.; Marko, J. F. *Phys. Rev. Lett.* **1999**, *83*, 1066–1069.
(73) Bustamante, C.; Marko, J. F.; Siggia, E. D.; Smith, S. *Science* **1994**, *265*, 1599–1600.
(74) Vologoskii, A. *Macromolecules* **1994**, *27*, 5623.
(75) Marko, J. F.; Siggia, E. D. *Macromolecules* **1995**, *28*, 8759.
(76) Bouchiat, C.; Wang, M. D.; Allemand, J.; Strick, T.; Block, S. M.; Croquette, V. *Biophys. J.* **1999**, *76*, 409–413.
(77) Trifonov, E. N.; Tan, R. K. Z.; Harvey, S. C. *DNA bending curvature*; Academic Press: Schenectady, NY, 1987; p 243.
(78) Schellman, J. A.; Harvey, S. C. *Biophys. Chem.* **1995**, *55*, 95–114.
(79) Katritch, V.; Vologodskii, A. *Biophys. J.* **1997**, *72*, 1070–1079.
(80) Nelson, P. *Phys. Rev. Lett.* **1998**, *80−102*, 5810–5812.
(81) Song, L.; Schurr, J. M. *Biopolymers* **1990**, *30*, 229–237.
(82) Bednar, J.; Furrer, P.; Katritch, V.; Stasiak, A. Z.; Dubochet, J.; Stasiak, A. *J. Mol. Biol.* **1995**, *254*, 579–594.
(83) Furrer, P.; Bednar, J.; Stasiak, A. Z.; Katritch, V.; Michoud, D.; Stasiak, A.; Dubochet, J. *J. Mol. Biol.* **1997**, *266*, 711–721.
(84) Mandelbrot, B. B. *The Fractal Grometry of Nature*; Freeman: San Francisco, CA, 1982.
(85) Mandelbrot, B. B.; Van Ness, J. W. *SIAM Rev.* **1968**, *10*, 422–437.
(86) Landau, L. D.; Lifshitz, E. M. *Theory of Elasticity*; Pergamon Press: Oxford, New York, 1970.
(87) Rivetti, C.; Walker, C.; Bustamante, C. *J. Mol. Biol.* **1998**, *280*, 41–59.
(88) Anselmi, C.; DeSantis, P.; Scipioni, A. *Biophys. Chem.* **2005**, *113*, 209–221.
(89) Cognet, J. A.; Pakleza, C.; Cherny, D.; Delain, E.; Le Cam, E. *J. Mol. Biol.* **1999**, *285*, 997–1009.
(90) Lyubchenko, Y.; Shlyakhtenko, L.; Harrington, R.; Oden, P.; Lindsay, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2137–2140.
(91) Hansma, H. G.; Revenko, I.; Kim, K.; Laney, D. E. *Nucleic Acids Res.* **1996**, *24*, 713–720.
(92) Zuccheri, G.; Scipioni, A.; Cavaliere, V.; Gargiulo, G.; De Santis, P.; Samori, B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3074–3079.
(93) Scipioni, A.; Anselmi, C.; Zuccheri, G.; Samori, B.; De Santis, P. *Biophys. J.* **2002**, *83*, 2408–2418.
(94) Scipioni, A.; Zuccheri, G.; Anselmi, C.; Bergia, A.; Samori, B.; De Santis, P. *Chem. Biol.* **2002**, *9*, 1315–1321.
(95) Marilley, M.; Sanchez-Sevilla, A.; Rocca-Serra, J. *Mol. Genet. Genomics* **2005**, *274*, 658–670.
(96) Bartosch, B.; Vitelli, A.; Garnier, C.; Goujon, C.; Dubuisson, J.; Pascale, S.; Scarselli, E.; Cortese, R.; Nicosia, A.; Cosset, F.-L. *J. Biol. Chem.* **2003**, *278*, 41670–41676.
(97) Note that the $H = 0.73$ LRC regime is observed for scales $\gtrsim$ 200 bp only, whereas for scales $\lesssim$ 200 bp there is a cross-over to an $H = 0.6$ LRC regime, as documented in refs 43 and 44. Thus, up to scale 200 bp, $\sigma_{\ell} = \sigma_o \ell^{0.6}$ rather than $\sigma_{\ell} = \sigma'_o \ell^{0.73}$ which is only valid above 200 bp; by continuity, $\sigma'_o = \sigma_o 200^{0.6}/200^{0.73}$. Taking this property into account to estimate the intrinsic curvature disorder amplitude at the bp scale, we get $\sigma_o = 0.007 \times 200^{0.73}/200^{0.6} \simeq 0.015$, i.e., a value similar to the one ($\sigma_o = 0.022$) found for the HCV DNA. This is, *a posteriori*, an important observation, since the local curvature fluctuations observed between successive base-pairs is expected to be of the same magnitude whatever the DNA molecules under consideration.
(98) Pietrement, O.; Pastre, D.; Fusil, S.; Jeusset, J.; David, M.-O.; Landousy, F.; Hamon, L.; Zozime, A.; Le Cam, E. *Langmuir* **2003**, *19*, 2536–2539.
(99) Shlyakhtenko, L. S.; Lushnikov, A. Y.; Lyubchenko, Y. L. *Biochemistry* **2009**, *48*, 7842–7848.
(100) Podesta, A.; Indrieri, M.; Brogioli, D.; Manning, G. S.; Milani, P.; Guerra, R.; Finzi, L.; Dunlap, D. *Biophys. J.* **2005**, *89*, 2558–2563.
(101) Bernardi, G. *Gene* **2001**, *276*, 3–13.
(102) Lander, E. S.; et al. *Nature* **2001**, *409*, 860–921.
(103) Li, W.; Bernaola-Galvan, P.; Carpena, P.; Oliver, J. L. *Comput. Biol. Chem.* **2003**, *27*, 5–10.
(104) Moreno-Herrero, F.; Seidel, R.; Johnson, S. M.; Fire, A.; Dekker, N. H. *Nucleic Acids Res.* **2006**, *34*, 3057–3066.
(105) Kam, Z.; Borochov, N.; Eisenberg, H. *Biopolymers* **1981**, *20*, 2671–2690.
(106) Sobel, E. S.; Harpst, J. A. *Biopolymers* **1991**, *31*, 1559–1564.
(107) Zhang, Y.; Crothers, D. M. *Biophys. J.* **2003**, *84*, 136–153.
(108) Cloutier, T. E.; Widom, J. *Mol. Cell* **2004**, *14*, 355–362.
(109) Du, Q.; Smith, C.; Shiffeldrim, N.; Vologodskaia, M.; Vologodskii, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5397–5402.
(110) Czapla, L.; Swigon, D.; Olson, W. K. *J. Chem. Theory Comput.* **2006**, *2*, 685–695.
(111) Rizzo, V.; Schellman, J. *Biopolymers* **1981**, *20*, 2143–2163.
(112) Hagerman, P. J. *Biopolymers* **1981**, *20*, 1503–1535.
(113) Lu, Y.; Weers, B.; Stellwagen, N. C. *Biopolymers* **2001**, *61*, 261–275.
(114) Porschke, D. *Biophys. Chem.* **1991**, *40*, 169–179.
(115) Griffith, J.; Bleyman, M.; Rauch, C. A.; Kitchin, P. A.; Englund, P. T. *Cell* **1986**, *46*, 717–724.
(116) Mazur, A. K. *Biophys. J.* **2006**, *91*, 4507–4518.