

Probing Persistence in DNA Curvature Properties with Atomic Force Microscopy

J. Moukhtar, E. Fontaine, C. Faivre-Moskalenko, and A. Arneodo

*Laboratoires Joliot Curie (USR 3010) et de Physique (UMR 5672), Ecole Normale Supérieure de Lyon,
46 allée d'Italie, 69364 Lyon cedex 07, France*

(Received 1 December 2006; published 23 April 2007)

We elaborate on a mean-field extension of the wormlike chain model that accounts for the presence of long-range correlations (LRC) in the intrinsic curvature disorder of genomic DNA, the stronger the LRC, the smaller the persistence length. The comparison of atomic force microscopy imaging of straight, uncorrelated virus and correlated human DNA fragments with DNA simulations confirms that the observed decrease in persistence length for human DNA more likely results from a sequence-induced large-scale intrinsic curvature than from some increased flexibility.

DOI: [10.1103/PhysRevLett.98.178101](https://doi.org/10.1103/PhysRevLett.98.178101)

PACS numbers: 87.14.Gg, 36.20.Ey, 36.20.Fz, 87.64.Dz

Since the discovery of naturally curved DNA [1], many experiments have demonstrated that DNA exhibits sequence-dependent curvature [2]. In addition, DNA bending and torsional flexibility that are likely to influence DNA-protein interactions, may also be sequence dependent [2]. For instance, some sequence motifs that favor intrinsically curved DNA and in turn the formation and positioning of nucleosomes were found to be regularly spaced [3]. Along that line, recent studies [4] have demonstrated that DNA encodes its packaging into an intrinsic nucleosome organization that can explain up to 50% of the *in vivo* nucleosome positions in the vicinity of *S. cerevisiae* genes. This raises the issue of the possible coherent organization of the bulk nucleosomes of higher eukaryotes for which the genes represent only a small portion of the genome. A statistical analysis of eukaryotic and prokaryotic genome sequences and their corresponding DNA chain bending profiles [5] has revealed the existence of two long-range correlations (LRC) regimes. In the 10–100 bp range, LRC are observed for eukaryotic sequences as quantified by a Hurst exponent value $H \simeq 0.6$ (and not for eubacterial sequences for which $H = 0.5$), as the signature of the nucleosomal structure. These LRC were shown to favor the autonomous formation and dynamics of small 2D DNA loops and likely the propensity of eukaryotic DNA to form nucleosomes [6]. Over larger distances (≥ 200 bp), stronger LRC with $H \simeq 0.8$ seem to exist in any sequences [5]. These LRC are observed in the *S. cerevisiae* nucleosome positioning data [7] suggesting that they are involved in the nucleosome organization into the 30 nm chromatin fiber. As this second LRC regime is also present in eubacterial sequences, it is likely to be a key to the understanding of the structure and dynamics of both the eukaryotic and prokaryotic chromatin fibers. Here we use AFM imaging to bring the first experimental evidence of the existence of sequence-induced LRC in naked DNA molecules.

Various experimental techniques have been used to investigate DNA curvature and flexibility [1,2,8]. By visualizing the 2D contour of a DNA molecule, atomic force microscopy (AFM) has proved to be very efficient to characterize conformational chain statistics [9,10]. So far,

AFM data have been mainly analyzed in the framework of the wormlike chain (WLC) model [11] that accounts for the entropic behavior of an intrinsically straight semiflexible polymer with the 2D elastic energy function:

$$\frac{E_{\text{el}}}{kT} = \frac{A_d}{2} \int_0^L \left[\frac{\partial \theta(s)}{\partial s} \right]^2 ds, \quad (1)$$

where the bending flexibility A_d controls the exponential decay of the correlations between tangent unit vectors $\langle \mathbf{u}(s) \cdot \mathbf{u}(0) \rangle = \langle \cos \theta(s) \rangle = \exp(-s/2A_d)$. Thus $2A_d = l_p$ is nothing but the 2D persistence length. The probability of having a certain bend angle over the distance s follows a Gaussian law [8,9]: $P(\theta) = \sqrt{l_p/4\pi s} \exp(-l_p \theta^2/4s)$. From the polymer mean square end-to-end distance:

$$\begin{aligned} \langle R^2 \rangle &= \int_0^L \int_0^L \langle \mathbf{u}(s) \mathbf{u}(s') \rangle ds ds', \\ &= 2l_p L \left[1 - \frac{l_p}{L} \left(1 - \exp\left(-\frac{L}{l_p}\right) \right) \right], \end{aligned} \quad (2)$$

one can estimate the persistence length $l_p = \langle R^2 \rangle / 2L$ in the limit $L \gg l_p$.

But the WLC model does not take into account the structural disorder $\theta_0(s)$ induced by the DNA sequence. As suggested by Trifonov *et al.* [12] and theoretically confirmed in [13], the intrinsically curved regions contribute to the experimental persistence length l_p^{eff} with a static term l_p^s , whereas thermal fluctuations appear in the dynamic term $l_p^d = 2A_d$ according to

$$1/l_p^{\text{eff}} = 1/l_p^d + 1/l_p^s. \quad (3)$$

This relationship is only valid if the intrinsic local curvature $\theta_0(s)$ of the chain is uncorrelated (or short-range correlated). To account for the presence of sequence-induced LRC in $\theta_0(s)$, the local fluctuation $\theta(s)$ from a straight polymer must be replaced in Eq. (1) by a local fluctuation $\theta(s) - \theta_0(s)$ from the frozen DNA trajectory. This trajectory can be generated by a Gaussian fractional noise of zero mean and variance σ_0^2 [6]; then $\theta_0(s) = \int_0^s \dot{\theta}_0(u) du$ is a fractional Brownian motion of zero mean

and variance $\sigma_0^2(s) = \sigma_0^2 s^{2H}$, where $1/2 < H < 1$ accounts for the existence of LRC in the intrinsic curvature fluctuations ($H = 1/2$ for uncorrelated chains). Under this assumption, one can analytically compute $\langle \mathbf{u}(s) \cdot \mathbf{u}(0) \rangle = \langle \cos(\theta(s) - \theta_0(s)) \rangle = \exp(-s/2A_d - s^{2H}\sigma_0^2/2)$, from which one recovers Eq. (3) for $H = 1/2$ with $l_p^s = 2/\sigma_0^2$. Then Eq. (2) becomes

$$\langle R^2 \rangle = \int_0^L ds' \int_0^{s'} ds \exp\left(-\frac{s}{2A_d} - \frac{s^{2H}\sigma_0^2}{2}\right), \quad (4)$$

that can be handled perturbatively and numerically. As shown in Fig. 1(a), when fixing $2A_d = 540$ bp and $\sigma_0 = 0.01$, we see that for uncorrelated chains ($H = 1/2$), $l_p(L) = \langle R^2 \rangle / 2L$ converges rather slowly to $l_p^{\text{eff}} = 512$ bp [Eqs. (2) and (3)]. When introducing LRC, one observes a faster convergence to a smaller persistence length, the larger the H , the smaller the l_p . As shown in Fig. 1(b), this decrease can be even more significant if one increases the amplitude σ_0 of this LRC structural disorder.

This mean-field generalization of the WLC model involves averaging over many DNA sequences. This rather prohibitive experimental task led us to perform DNA simulations as in [10(a)] for uncorrelated and LRC DNAs. To model the presence of intrinsic uncorrelated disorder, we first generated a Gaussian white noise $\theta_0(s)$ of zero mean and variance σ_0^2 [Fig. 2(a')]. Then 2D equilibrium conformations were generated using a Gaussian law of variance $1/A_d$ but centered at $\theta_0(s)$, to randomly choose the angle of the next unit rod with respect to the preceding one. In Fig. 2(b') are illustrated $N = 100$ trajectories mimicking the effect of thermal fluctuations on the original uncorrelated frozen chain. To model LRC in the intrinsic curvature properties we used a Gaussian corre-

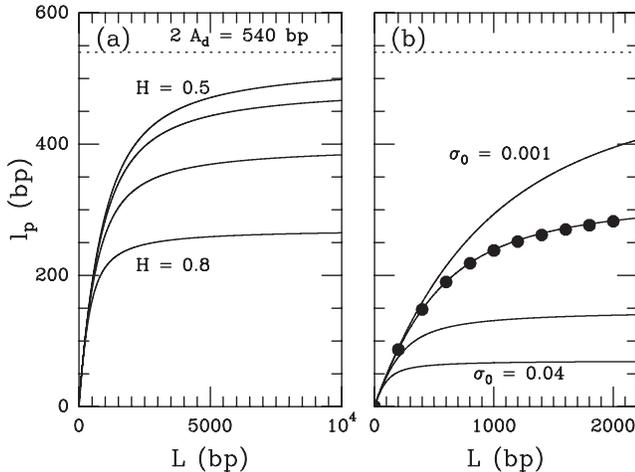


FIG. 1. Persistence length $l_p(L) = \langle R^2 \rangle / 2L$ computed with Eq. (4) for $2A_d = 540$ bp: (a) $\sigma_0 = 0.01$, $H = 0.5, 0.6, 0.7, 0.8$ from top to bottom; (b) $H = 0.8$ and $\sigma_0 = 0.001, 0.008, 0.02, 0.04$ from top to bottom. The black dots correspond to DNA simulations with $\sigma_0 = 0.008$ and $H = 0.8$, when averaging over $N = 10^3$ different frozen chains.

lated noise with index H for $\dot{\theta}_0(s)$ [Fig. 2(a')] and we induced thermal fluctuations as before. A representative $H = 0.8$ LRC frozen chain is illustrated in Fig. 2(c') together with $N = 100$ 2D equilibrium conformations.

AFM imaging was carried out with the Nanoscope III (Digital Instruments, Santa Barbara, CA) operating in tapping mode in air [14]. In the presence of Mg^{2+} in the solution, DNA molecules are able to equilibrate on the surface before being captured in a given conformation [9,15]. To avoid molecules interacting during their equilibrium process, we have used a high dilution rate so that each image contains no more than one or two DNA molecules. The DNA traces were digitized using a homemade MATLAB (MathWorks Inc., Natick, MA) script based on morphological tools. As reported in Refs. [9,16], our estimate of the mean DNA contour length yields an helical rise of 3.21 ± 0.05 Å per bp which underestimates the 3.38 Å per bp measured by crystallography. For each entropic realization of a DNA molecule of size L (~ 2200 bp), we estimated the “internal” end-to-end distance $R(s)$ in a

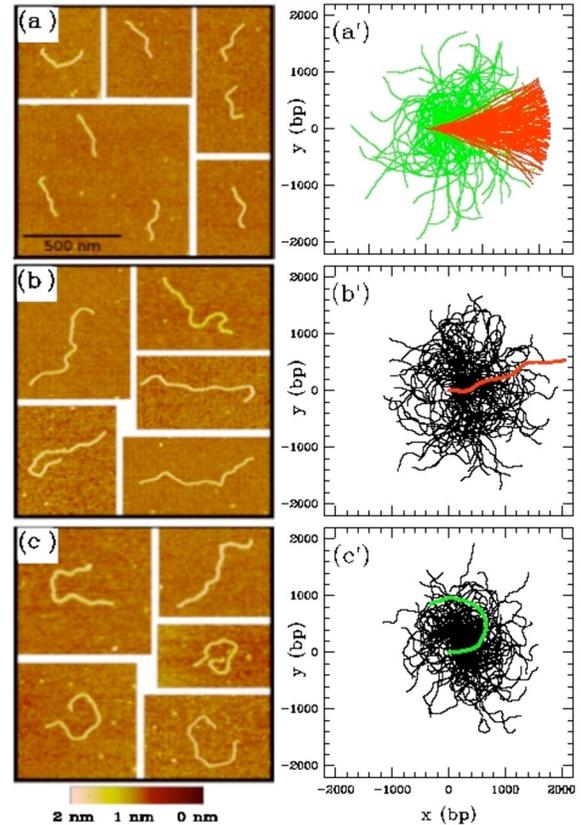


FIG. 2 (color online). AFM images of $L = 800$ bp intrinsically straight (a) and $L = 2200$ bp HCV (b) and human (c) DNA molecules. DNA simulations of $N = 100$ $L = 2200$ bp chains: 2D frozen chains with uncorrelated ($H = 1/2$, dark red) and LRC ($H = 0.8$, light green) curvature angle fluctuations of amplitude $\sigma_0 = 0.008$ (a'); 2D equilibrium conformations (black) generated with $2A_d = 540$ bp from a frozen chain with uncorrelated ($H = 1/2$, $\sigma_0 = 0.02$, red) (b') and with LRC ($H = 0.8$, $\sigma_0 = 0.008$, green) (c').

window of size s , that was slid by 1 pixel (~ 4.5 bp) along the DNA trajectory from one extremity to the other. This symmetric procedure leads to non polar estimate of $R(s)$. We have repeated this step for s going from 20 bp to the total DNA length L . Then by averaging over all the DNA trajectories obtained from the AFM images, we have estimated some internal persistence length $l_p(s) = \langle R^2(s) \rangle / 2s$ which converges to the actual persistence length in the limit $s \rightarrow L \rightarrow \infty$.

We have selected three different types of DNA molecules. (i) An intrinsically straight DNA fragment of total length $L \simeq 800$ bp was designed according to Ref. [17]. (ii) A fragment of $L = 2200$ bp was extracted from the hepatitis C virus (HCV) genome (sequence core, E_1, E_2) [18]. HCV belongs to the family of nonretrotranscribed RNA virus that was shown to display uncorrelated bending profiles [5]. (iii) Two (repeated sequence free) LRC DNA fragments were extracted from intergenic regions in chromosome 8 (125789398-125791587) and 21 (37044232-37046437) of the human genome (NCBI build 35) by PCR amplification and cloning. Prior to the AFM analysis, we have performed a wavelet based analysis of the corresponding DNA chain curvature profiles as in Ref. [5]. While for the HCV sequence we confirm that over the whole range of scales from 10 to 2200 bp, one observes scaling behavior with $H = 1/2$ as the signature of uncor-

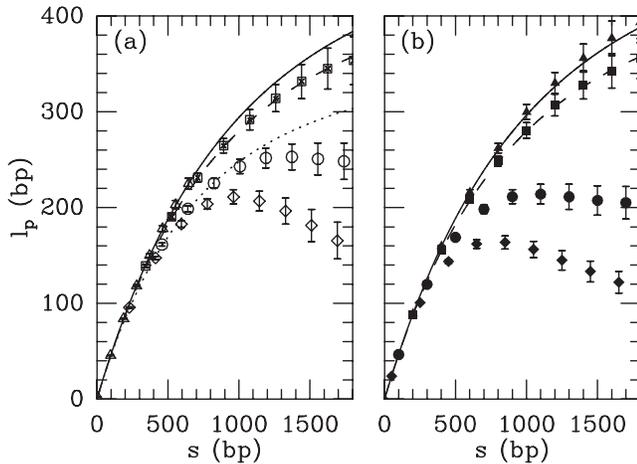


FIG. 3. Internal end-to-end distance measurement of $l_p(s)$ vs s when averaging over N samples of length $L = 2200$ bp when nonspecified. (a) AFM imaging: (Δ) intrinsically straight DNA ($N = 131$, $L = 800$ bp), (\square) HCV DNA ($N = 61$), (\circ) human chromosome 21 DNA ($N = 91$), (\diamond) human chromosome 8 DNA ($N = 81$). The WLC model predictions [Eq. (3)] are in solid ($l_p = 540$ bp), dashed ($l_p = 485$ bp) and dotted ($l_p = 385$ bp) lines. (b) DNA simulations of $N = 100$ 2D equilibrium conformations generated with $2A_d = 540$ bp from a frozen chain that is: (\blacktriangle) straight, (\blacksquare) with uncorrelated ($H = 1/2$, $\sigma_0 = 0.02$) curvature disorder [Fig. 2(b')], (\bullet) with a LRC ($H = 0.8$, $\sigma_0 = 0.008$) curvature disorder [Fig. 2(c')], (\blacklozenge) with another frozen LRC chain. The solid and dashed lines are the same as in (a).

related curvature fluctuations, we definitely detect LRC in both human DNA sequences, in particular, for scales ranging from 200 to 2200 bp which will be probed by the AFM apparatus, where $H = 0.80 \pm 0.05$.

AFM images of the straight DNA are shown in Fig. 2(a). A quantitative analysis of $N = 131$ DNA traces shows a remarkable agreement with the WLC model. As reported in Fig. 3(a), Eq. (2) provides a nice fit of $l_p(s) = \langle R^2(s) \rangle / 2s$, when adjusting $l_p^d = 2A_d = 540 \pm 20$ bp. This yields an estimate of $A_d = 270$ bp that corresponds to a 3D persistence length $\lambda = 0.321 \times l_p / 2 = 86 \pm 3$ nm, in good agreement with previous measurements by cryo-electron microscopy [19] and recent molecular dynamic simulations [20]. In Fig. 2(b) are shown typical HCV molecules visualized with our AFM. The corresponding $l_p(s)$ data reported in Fig. 3(a) are again very well reproduced by the WLC model predictions but with $l_p^{\text{eff}} = 485 \pm 15$ bp ($\lambda = 78 \pm 2$ nm) smaller than the dynamic persistence length estimated just above. Since the experimental protocol used to deposit and visualize the three types of DNAs was rigorously the same, one can assume that the bending flexibility of all molecules are identical; then from Eq. (3), one gets $l_p^s = 4762$ bp that corresponds to a rather small curvature angle fluctuation amplitude $\sigma_0 = (2/l_p^s)^{1/2} = 0.02$, typical of a weak structural disorder. In Fig. 3(b) are reported the results of a comparative numerical study of $N = 100$ 2D equilibrium conformations of length $L = 2200$ bp, generated with $2A_d = 540$ bp from, respectively, a straight chain and an uncorrelated random chain [Fig. 2(b')]. The $l_p(s)$ numerical data look very similar to the corresponding AFM data. They are remarkably well fitted by the same WLC curves giving strong support to our experimental estimate of $\sigma_0 = 0.02$ for the uncorrelated HCV DNA.

In Fig. 3(a) we also report the $l_p(s)$ data obtained from $N = 81$ (91) human chromosome 8 (21) DNA AFM contours. The comparison with the HCV data leads to the following observations. First, for both human DNA fragments, $l_p(s)$ is found significantly smaller than for HCV and this over the whole range of s values. Second, the WLC model [Eq. (2)] fails to fit the human $l_p(s)$ data which suggests that the observed increase in bending is not due to an increase in flexibility which would exhibit WLC behavior, but rather due to the presence of some large-scale cooperative intrinsic curvatures as clearly seen on the AFM images of human DNA in Fig. 2(c). To demonstrate that this persistence in curvature properties is not due to some hyperperiodic distribution of curved DNA sites [21], but rather results from the presence of LRC in the curvature angle fluctuations, we have performed in Fig. 2(c'), DNA simulations of LRC chains with parameters ($H = 0.8$, $\sigma_0 = 0.008$) as estimated from the human DNA sequence analysis. Remarkably, the $l_p(s)$ numerical data shown in Fig. 3(b), for two so-generated frozen LRC chains, display the same characteristics as experimentally observed in

Fig. 3(a) for the two human DNAs. Not only the global lowering of $l_p(s)$ is reproduced but also the observed decrease at large s values is numerically recovered. In Fig. 2(a') are illustrated $N = 100$ frozen DNA chains of length $L = 2200$ bp of uncorrelated ($H = 1/2$) and LRC ($H = 0.8$) curvature angle fluctuations of amplitude $\sigma_0 = 0.008$. One sees that most of the LRC chains intrinsically display some marked macroscopic curvature in contrast to the uncorrelated ones. Actually, one can show that the characteristic size of a frozen LRC chain to make half a loop is $l^* = (\pi/\sigma_0)^{1/H}$ which yields $l^* = 1748$ bp for $H = 0.8$ and $\sigma_0 = 0.008$. Since the end-to-end distance is expected to decrease when s reaches half the loop-size, this explains the decrease observed in $l_p(s)$ at large s in Fig. 3(b). This is a strong indication that the similar decrease of $l_p(s)$ observed experimentally in Fig. 3(a) for both human DNAs is quite representative of the presence of LRC between randomly distributed curvature sites. As a final remark, our mean-field approach predicts some decrease of the persistence length when increasing the LRC exponent H (Fig. 1). For $2A_d = 540$ bp, $H = 0.8$ and $\sigma_0 = 0.008$, one gets $l_p = 315$ bp that corresponds to a value of the 3D persistence length, $\lambda = 50.6$ nm, in quantitative agreement with previous measurements by independent methods [2,19,22]. The fact that this $H = 0.8$ LRC regime is universally observed (for distances $s > 200$ bp) in prokaryotic and eukaryotic DNA sequences [5], might thus explain the 53 nm consensus value for the persistence length of native DNA.

In summary, by assisting AFM imaging with DNA simulations, we have revealed some significant lowering of the persistence length of human DNA molecules as compared to measurements performed on both intrinsically straight and uncorrelated HCV DNAs. The fact that this spectacular decrease cannot be accounted for by the WLC model provides strong indication that this increase in bending does not result from some enhancing of flexibility but rather from some large-scale intrinsic curvature induced by a sequence-dependent persistent random distribution of local bending sites. We have proposed a mean-field generalization of the WLC model that reproduces the decrease of the persistence length when enhancing LRC in the DNA curvature disorder. But as shown in Fig. 1(b), this mean-field model requires the averaging over several hundreds DNA molecules with different sequences. In a work under progress, we hope to achieve this sequence averaging by digesting genomic DNA to produce DNA fragments with blunt ends at random, with the perspective of measuring the LRC parameters (H , σ_0) in various eukaryotic and prokaryotic organisms.

We are very grateful to F. Argoul, B. Audit, G. Lavorel, F. Montel, P. St-Jean, and C. Vaillant for helpful discussions. We thank B. Bartosch, F.-L. Cosset, and A. Vologodskii for the kind gifts of plasmids. This work was supported by the ACI IMPBio 2004, the Conseil Regional Rhône-Alpes (Emergence 2005) and the

Agence Nationale de la Recherche (PCV06_135105).

-
- [1] J.C. Marini *et al.*, Cold Spring Harbor Symposia on Quantitative Biology (1933-) [Proceedings] **47**, 279 (1983).
 - [2] P.J. Hagerman, Annu. Rev. Biophys. Biophys. Chem. **17**, 265 (1988); D.M. Crothers, T.E. Haran, and J.G. Nadeau, J. Biol. Chem. **265**, 7093 (1990).
 - [3] S.C. Satchwell, H.R. Drew, and A.A. Travers, J. Mol. Biol. **191**, 659 (1986); I.P. Ioshikhes *et al.*, J. Mol. Biol. **262**, 129 (1996); J. Widom, J. Mol. Biol. **259**, 579 (1996).
 - [4] G.C. Yuan *et al.*, Science **309**, 626 (2005); I.P. Ioshikhes *et al.*, Nat. Genet. **38**, 1210 (2006); E. Segal *et al.*, Nature (London) **442**, 772 (2006).
 - [5] B. Audit *et al.*, Phys. Rev. Lett. **86**, 2471 (2001); B. Audit *et al.*, J. Mol. Biol. **316**, 903 (2002).
 - [6] C. Vaillant, B. Audit, and A. Arneodo, Phys. Rev. Lett. **95**, 068101 (2005); C. Vaillant *et al.*, Eur. Phys. J. E **19**, 263 (2006).
 - [7] C. Vaillant, B. Audit, and A. Arneodo (to be published).
 - [8] H.G. Hansma *et al.*, Nucleic Acids Res. **24**, 713 (1996); C. Rivetti, C. Walker, and C. Bustamante, J. Mol. Biol. **280**, 41 (1998); C. Anselmi, P. DeSantis, and A. Scipioni, Biophys. Chem. **113**, 209 (2005).
 - [9] C. Rivetti, M. Guthold, and C. Bustamante, J. Mol. Biol. **264**, 919 (1996).
 - [10] (a) J.A.H. Cagnet *et al.*, J. Mol. Biol. **285**, 997 (1999); (b) A. Scipioni *et al.*, Biophys. J. **83**, 2408 (2002); (c) M. Marilley, A. Sanchez-Sevilla, and J. Rocca-Serra, Mol. Genet. Genomics **274**, 658 (2005).
 - [11] O. Kratky and G. Porod, Recueil: J. Roy. Netherlands Chem. Soc. **68**, 1106 (1949); J.A. Schellman, Biopolymers **13**, 217 (1974).
 - [12] E. Trifonov, R. Tan, and S. Harvey, in *DNA Bending and Curvature, Structure and Expression*, edited by W.K. Olson *et al.* (Adenine, Schenectady, NY, 1988), Vol. 3.
 - [13] J.A. Schellman and S.C. Harvey, Biophys. Chem. **55**, 95 (1995).
 - [14] Samples were imaged with silicon tips (resonant frequency 250 kHz) at scanning rate of 2 Hz. The (512×512) images were collected with a scan size of $1.5 \mu\text{m}$. DNA preparations were diluted ($\sim 0.3 \text{ ng}/\mu\text{l}$) in 10 mM Tris-HCl, $\text{pH} = 7.4$ buffer containing 5 mM MgCl_2 . A $5 \mu\text{l}$ droplet was deposited onto a freshly cleaved mica surface, incubated for 2 min, rinsed with MilliQ water and dried under a nitrogen flow prior to imaging.
 - [15] D. Pastre *et al.*, Biophys. J. **85**, 2507 (2003); M. Sushko, A. Shluger, and C. Rivetti, Langmuir **22**, 7678 (2006).
 - [16] C. Rivetti and S. Codeluppi, Ultramicroscopy **87**, 55 (2001); A. Sanchez-Sevilla *et al.*, Ultramicroscopy **92**, 151 (2002).
 - [17] M. Vologodskaia and A. Vologodskii, J. Mol. Biol. **317**, 205 (2002).
 - [18] B. Bartosch *et al.*, J. Biol. Chem. **278**, 41 624 (2003).
 - [19] J. Bednard *et al.*, J. Mol. Biol. **254**, 579 (1995).
 - [20] A.K. Mazur, Biophys. J. **91**, 4507 (2006).
 - [21] F. Moreno-Herrero *et al.*, Nucleic Acids Res. **34**, 3057 (2006).
 - [22] C. Bustamante *et al.*, Science **265**, 1599 (1994).