Influence of the sequence on elastic properties of long DNA chains

C. Vaillant,[1] B. Audit,[2] C. Thermes,[3] and A. Arnéodo[4]
[1]*Institut Bernoulli, EPFL, 1015 Lausanne, Switzerland*
[2]*Computational Genomics Group, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, United Kingdom*
[3]*Centre de Génétique Moléculaire, CNRS, Laboratoire associé à l'Université Pierre et Marie Curie, Allée de la Terrasse, 91198 Gif-sur-Yvette, France*
[4]*Laboratoire de Physique, ENS Lyon, 46 allée d'Italie, 69364 Lyon, France*

(Received 11 June 2002; revised manuscript received 17 December 2002; published 11 March 2003)

We revisit the results of single-molecule DNA stretching experiments using a rodlike chain (RLC) model that explicitly includes some intrinsic structural disorder induced by the sequence. The investigation of artificial and real genomic sequences shows that the wormlike chain model reproduces quite well the data but with an effective bend stiffness A_{eff}, which underestimates the true elastic bend stiffness A, independently of the elastic twist stiffness C. Mainly dominated by the amplitude of the structural disorder, this correction seems rather insensitive to the presence of long-range correlations. This RLC model is shown to remarkably fit the experimental data for λ-DNA when considering $A \simeq 70 \pm 10$ nm ($> A_{eff} \simeq 50$ nm), in good agreement with previous experimental estimates of the "dynamic" persistent length. From the analysis of large human contigs, we speculate about the possible dependence of A_{eff} and/or A upon the $(G+C)$ content of the considered sequence.

DOI: 10.1103/PhysRevE.67.032901 PACS number(s): 87.15.−v, 36.20.Fz, 87.10.+e, 87.14.Gg

The highly compacted organization of chromatin *in vivo* involves DNA coiling up twice around an octamer of core histone proteins to form the nucleosome [1], followed by successive higher-order foldings to reach maximal condensation in metaphase chromosomes [2]. Since the discovery of naturally bent DNA, several works have investigated the possibility that the DNA sequence may facilitate the nucleosome packaging [3]. Very recently, the statistical analysis of DNA chain bending profiles for complete genome sequences has revealed that long-range correlations in the 10–200 bp range are the signature of the nucleosomal structure and that over larger distances ($\gtrsim 200$ bp) they are likely to play a role in the hierarchical packaging of DNA [4]. To which extent sequence-dependent DNA mechanical properties do help to regulate the structure and dynamics of chromatin is an issue of fundamental importance. A possible key to understanding is that the structural disorder induced by the sequence may modify the DNA chain elastic response.

During the last few years, micromanipulation experiments on single DNA molecules have enabled the study of their elastic response to external stretching [5] and twisting [6,7] forces. These pioneering experiments are very well described by simple elastic models. In the absence of a twisting force, the wormlike chain (WLC) model [8] with a single elastic constant, the bend persistence length A_{eff}, is sufficient. From the extension vs force data, most experiments yield similar estimates of $A_{eff} \simeq 50$ nm in physiological conditions. In the presence of a twisting constraint, the rodlike chain (RLC) model [9], which involves an extra parameter, the twist persistence length C_{eff}, reproduces quite well the experimental extension vs supercoiling curves ($f \leqslant 0.5$ pN) with C_{eff} between 75 nm and 110 nm [9]. But, as suggested by Trifonov *et al.* [10], the measurable bend persistence length A_{eff} does not correspond to the bend rigidity A of the double helix. It follows from the joint effect of the "static" bend persistence length A_o of the random walk defined by the axis of the DNA double helix in the absence of any thermal fluctuations and the "dynamic" bend persistence length A of a DNA double helix in the absence of any intrinsic structural disorder:

$$1/A_{eff} = 1/A + 1/A_o . \tag{1}$$

This equation has received some early theoretical and computational confirmation [11]. Experimentally, from the investigation of natural [10,12] and "intrinsically straight" synthetic [13] DNA, Eq. (1) has led to values of A ranging from 60 nm up to 210 nm, as compared to the generally accepted value $A_{eff} \simeq 50$ nm. Recently, under some working hypothesis, Nelson [14] has proved that Eq. (1) is correct in the limit of weak structural disorder, i.e., $A_{eff} = A(1-\lambda)$, for small $\lambda = A/A_o$. This correction differs from $A_{eff} = A(1 - \sqrt{\lambda}/2)$ found by Bensimon *et al.* [15] in a random version of the Kratky-Porod model. Also, according to Nelson the twist persistence length would not suffer such a correction: $C_{eff} = C$.

Our aim here is to take explicitly into account the intrinsic local bend and twist fluctuations of the DNA double-helix reflecting sequence information in the RLC model. The conformations of an inextensible RLC under applied tension in the $\mathbf{z}$ direction at the free end of the chain are controlled by the elastic energy functional [14]:

$$\frac{E}{k_BT} = \int_0^L ds \Bigg[\frac{A}{2}(\Omega_1 - \omega_1)^2 + \frac{A}{2}(\Omega_2 - \omega_2)^2 + \frac{C}{2}(\Omega_3 - \theta_o - \omega_3)^2 - \frac{f}{k_BT}\mathbf{e}_3 \cdot \mathbf{z} \Bigg], \tag{2}$$

up to quadratic order terms in the deformations from the intrinsic quenched ($T=0$) double-helix configuration $\{\omega_1(s), \omega_2(s), \theta_o + \omega_3(s)\}$, where $\theta_o = 2\pi/3.5\ \mathrm{nm}^{-1}$ is the unstressed double-helix twist and $\omega_1,\ \omega_2,\ \omega_3$ are the intrinsic roll, tilt, and twist angles (per unit length), respectively. If one neglects the sequence $[\{\omega_i(s)\} = 0]$, then, in the parti-

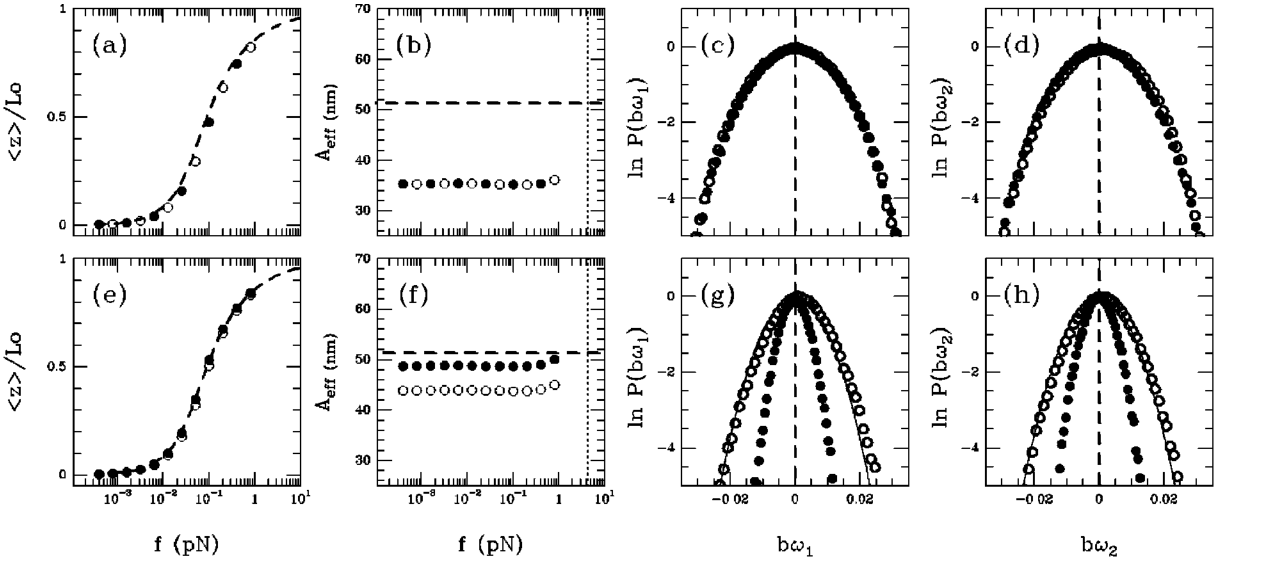

FIG. 1. RLC model calculations ($A=51.3$ nm,$C=0$) for two sets of 20 long-range correlated [(●)$H=0.8$] and 20 uncorrelated [(○)$H=0.5$] artificial sequences of length $L=20\,000$ bp, when using the Ulyanov and James [16] (a)–(d) and the Gorin *et al.* [17] (e)–(h) coding tables. (a),(e) Average extension vs force curve; (b),(f) A_{eff} vs f as obtained when fitting the numerical data by the WLC model prediction [Eq. (3)]; (c),(g) pdf of $\tilde{\omega}_1$; (d),(h) pdf of $\tilde{\omega}_2$. In (a), (b), (e), and (f), the dashed line corresponds to the WLC model prediction for $A=51.3$ nm; the vertical dotted line corresponds to $f_{sup}=k_BTA/b^2$ (see text).

tion function calculation, one can explicitly integrate over the twist variable to end up with the WLC model. As a very accurate (better than 0.01%) interpolation of the exact numerical solution of this model, we will use the Bouchiat *et al.* formula [8(c)]:

$$f=\frac{k_BT}{A}\left\{\frac{1}{4\left(1-\frac{\langle z\rangle}{L}\right)^2}-\frac{1}{4}+\frac{\langle z\rangle}{L}+\sum_{i=2}^{7}a_i\left(\frac{\langle z\rangle}{L}\right)^i\right\}, \qquad (3)$$

where the $\{a_i\}$ are parameters. In the limit of small stretching forces ($f<k_BT/A$), one recognizes the linear extension relation $\langle z\rangle/L=\frac{2}{3}Af/k_BT$; in the limit of large stretching forces ($f>k_BT/A$), one recovers the asymptotic behavior $\langle z\rangle/L=1-(k_BT/4Af)^{1/2}$. When taking into account the intrinsic structural disorder $\{\omega_i(s)\}$, the bend and twist variables no longer decouple and the extension curves *a priori* depend on the two elastic constants A and C.

In the present work, we solve numerically the isotropic RLC model using the transfer matrix techniques. As discussed by Bouchiat and Mézard [9(c)], this model is singular in the limit of a purely continuous chain, and its discretization requires the introduction of a short length scale cutoff $b\simeq 7$ nm (approximately twice the double-helix pitch). Here we only consider a stretching constraint, torsional forcing will be discussed in a forthcoming publication. To account for the effects of the sequence, we use several experimentally established structural tables that code for the intrinsic local bending and flexibility properties of the DNA double helix. We report the results obtained with the dinucleotide coding tables for the intrinsic ω_i angles elaborated by Ulyanov and James [16] from nuclear magnetic resonance data and Gorin *et al.* [17] from crystallographic data.

In Fig. 1 we show the results of our RLC modeling of DNA stretching experiments for DNA sequences displaying long-range correlations associated to a Hurst exponent value $H=0.8$, as observed in real genomic sequences [4]. These sequences were artificially built [18], with the specific goal to generate monofractal bending profiles [4]. For comparison, we have reproduced our analysis, but after having randomly shuffled the nucleotides to suppress the correlations ($H=0.5$). When averaging the relative extension $\langle z\rangle/L$ over our two sets of 20 sequences, one gets extension curves that are remarkably well fitted by the WLC model [Eq. (3)] with an effective bend persistence length $A_{eff}<A$, found to be quite insensitive to the value of the twist persistence length C (less than 1% variation for $C\in[0,150\text{ nm}]$). Therefore we will report results for $C=0$ only. In Figs. 1(a)–1(d), we show the numerical results obtained with the Ulyanov and James table, when fixing $A=51.3$ nm (the λ-DNA persistence length [6]). In Fig. 1(a), the extension curve obtained for the correlated sequences is compared to the WLC model with $A=51.3$ nm. As shown in Fig. 1(b), for a wide range of forces extending almost up to $f_{sup}=k_BTA/b^2$, where discretization effects become significant, this curve is well fitted by Eq. (3) with $A_{eff}=35\pm1$ nm, i.e., a value significantly smaller than the dynamic bend persistence length A. This leads [Eq. (1)] to a "static" bend persistence length $A_o\simeq110$ nm ($\lambda=A/A_o\simeq0.47$). The probability density functions (pdf) of the angles $\tilde{\omega}_1=b\omega_1$ and $\tilde{\omega}_2=b\omega_2$ are shown in Figs. 1(c) and 1(d). In a semilogarithmic representation, these curves fall on the same centered parabola, which is the signature of isotropic zero-mean Gaussian statistics. From value larger than that extracted from our RLC model calculations. In Figs. 1(a) and 1(b), we do not see any notable change on the extension curve computed for the set of uncorrelated sequences. This is consistent with the fact that the

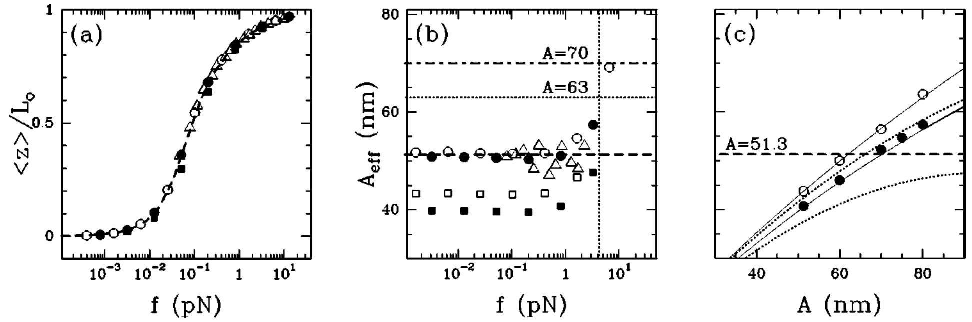

FIG. 2. RLC model calculations $(A,C=0)$ for the λ-DNA chain ($L=48\,502$ bp) when using the Ulyanov and James (black symbols) and Gorin *et al.* (white symbols) tables. (a) Extension vs force curves obtained for $A=51.3$ nm (■,□), 63 nm (○), and 70 nm (●) as compared to the experimental data (△) [6]. (b) A_{eff} vs f as obtained from the WLC equation (3). (c) A_{eff} vs A; the solid lines correspond to Eq. (1) with $A_o=190$ nm (Ulyanov and James) and 295 nm (Gorin *et al.*); the dotted lines correspond to the Nelson perturbative equation $A_{eff}=A(1-\lambda)$ [14]. In (a)–(c) the dashed line corresponds to the WLC model equation (3) for $A=51.3$ nm.

pdfs of $\tilde{\omega}_1$ and $\tilde{\omega}_2$ in Figs. 1(c) and 1(d) superpose on those computed for the correlated sequences.

In Figs. 1(e)–1(h), the results obtained with the Gorin *et al.* table [17] are reported. For the correlated sequences, the extension curve [Fig. 1(e)] is still very well fitted with the WLC equation (3) but with $A_{eff}=49.1\pm0.2$ nm$\lesssim A=51.3$. As shown in Figs. 1(g) and 1(h), this small correction is the consequence of a weaker structural disorder: the pdfs of $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are still Gaussian with zero mean but with a smaller variance, $\sigma_o^2\simeq0.0062$, from which we get $A_o=b/\sigma_o^2\simeq1129$ nm. Equation (1) yields $A_{eff}=49.1\pm0.2$ nm $(\lambda\simeq0.041)$, a value that matches perfectly the estimate obtained from the extension RLC calculations [Fig. 1(f)]. For the uncorrelated sequences, one gets a more sensitive response of the DNA chains to the stretching force, since $A_{eff}=43.8\pm0.2$ nm is smaller than for the correlated sequences. As shown in Figs. 1(g)–1(h), the pdfs of $\tilde{\omega}_1$ and $\tilde{\omega}_2$ are still indistinguishable and approximately Gaussian but with a larger variance $\sigma_o^2\simeq0.0219$. So $A_o=b/\sigma_o^2\simeq320$ nm and Eq. (1) yields $A_{eff}=44\pm0.5$ nm $(\lambda\simeq0.16)$, again in remarkable agreement with the estimate extracted from the extension vs force curve. When using the Gorin *et al.* table, the long-range correlations are associated with some weakening of the structural disorder induced by the sequence, which contrasts the results obtained from the Ulyanov and James table.

In Fig. 2(a), we report the experimental extension vs force data recorded by Strick *et al.* [6] for λ-DNA chains ($L=48\,502$ bp). These data are very well fitted by the WLC model [Eq. (3)] when adjusting the (effective) bend persistence length to $A_{eff}=51.3$ nm. If one uses this value as the dynamic persistent length $A=51.3$ nm in Eq. (2), then as shown in Figs. 2(a) and 2(b), when using the Ulyanov and James table, one gets with the RLC model results that no longer fit the experimental data but that are still well reproduced by the WLC model with an effective bend persistence length $A_{eff}=40\pm1$ nm$<A=51.3$ nm. This sequence disorder correction is represented in Fig. 2(c) when using the Ulyanov and James, and Gorin *et al.* tables. By plotting A_{eff} vs A, for A ranging from 40 to 80 nm, one gets data that are remarkably well fitted by Eq. (1) when setting $A_o=b/\sigma_o^2$, where σ_o^2 is the variance of the pdfs of $\tilde{\omega}_1$ and $\tilde{\omega}_2$ which are found indistinguishable. Hence, we get $A_o=190$ nm with the Ulyanov and James table $(\sigma_o^2=0.0368)$ and $A_o=295$ nm with the Gorin *et al.* table $(\sigma_o^2=0.0237)$. Using an apparent persistent length $A_{eff}=51.3$ nm as observed in the experiments, the inversion of Eq. (1), leads to $A=70\pm2$ nm (Ulyanov and James) and 63 ± 2 nm (Gorin) for the dynamic bend persistence length [see also Fig. 2(b)]. These results are in good agreement with previous experimental estimates ($A=70\pm10$ nm) of the "dynamic" persistence length of natural [10] and intrinsically straight [13] DNA chains. Note that we have reproduced our RLC model calculations after having randomly shuffled the λ-DNA sequence to remove the long-range correlations $(H\simeq0.8)$ observed at scales larger than 200 bp [4], without noticing any quantitative difference from the results reported in Fig. 2.

A very interesting issue, which can be tackled with RLC model simulations, is the possible influence of the sequence

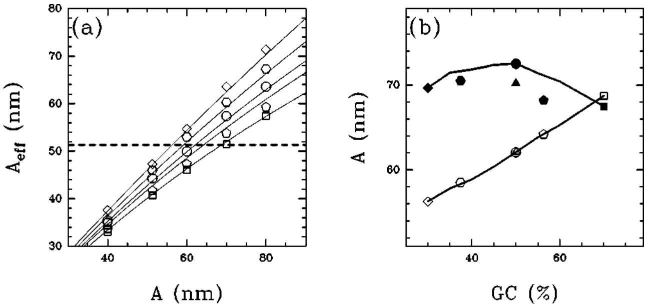

FIG. 3. RLC model calculations $(A,C=0)$ for artificial long-range correlated $(H=0.8)$ sequences ($L=20\,000$) with an average GC percentage equal to 30 (◇,◆), 50 (○,●), and 70 (□,■) and for two human DNA sequences with GC percentage equal to 40 ($L=20\,080$, white and black hexagons) and 56.3 ($L=29\,200$, white and black pentagons). The black (white) symbols correspond to the Ulyanov and James (Gorin *et al.*) table. (a) A_{eff} vs A, the solid lines correspond to Eq. (1) with $A_o=b/\sigma_o^2$. (b) A vs GC percentage, where A is the dynamic persistence length that leads to $A_{eff}=51.3$ nm as observed for λ-DNA chains (△,▲). The continuous curves correspond to the mean value $\overline{A}$ obtained for sets of ten artificial chains with GC percentage ranging from 30 to 70.

composition on the elastic response of the corresponding DNA chain. In particular, in possible relation to the isochore structure of the human genome, it has been clearly shown in Ref. [19] that the long-range correlation properties of human DNA sequences are dependent upon their GC $(=G+C)$ content. In Fig. 3, we report the results of RLC model calculations for several artificial DNA sequences and two real human DNA sequences of different GC contents. From the results obtained for λ-DNA in Fig. 2, we fix $A=70$ nm $(C=0)$. For both Ulyanov and James, and Gorin *et al.* tables, we recover the same agreement with the WLC model prediction [Eq. (3)] with an effective bend persistence length $A_{eff}<A$, which satisfies the Trifonov *et al.* relationship (1) with $A_o=b/\sigma_o^2$. When using the Ulyanov and James table, one gets a value $A_{eff}\simeq 50$ nm as observed for λ-DNA chains and this almost independently of the GC content. Consistently, the $(\tilde{\omega}_1,\tilde{\omega}_2)$ pdfs do not display any significant change when varying the GC content. This is no longer the case when one uses the Gorin and James table. Indeed, σ_o^2 now increases almost linearly with the GC percentage. This enhancing of the sequence induced structural disorder in the GC rich regions of the human genome corresponds to some decrease of A_o and, as shown in Fig. 3(a), results in some systematic (i.e., whatever the value of A) decrease of A_{eff} when increasing the GC percentage. In that respect, the experimental investigation of the human DNA chains looks rather crucial. The observation of no GC content dependence of A_{eff} would seem to be in favor of the Ulyanov and James table but would not exclude the Gorin *et al.* table. In Fig. 3(b), we show the value of A vs the GC percentage required for the RLC model to yield $A_{eff}=51.3$ nm when using the Gorin *et al.* table.

The hypothetical but possible observation of some universal persistence length $A_{eff}\simeq 50$ nm for general DNA chains could be understood from the Gorin *et al.* table by introducing some GC dependence in the dynamic persistent length A in the RLC model. Figure 3(b) reveals higher GC content, larger A, and greater rigidity to the bending of the corresponding chain, a result that would be quite plausible with respect to the actual experimental and numerical knowledge concerning the mechanical properties of the DNA double helix [20]. On the contrary, some experimental observation of a decrease of A_{eff} when increasing the GC percentage would make the Ulyanov and James table rather inadequate, since it would require A to be smaller in GC rich DNA chains.

To conclude, the numerical results reported in this paper show that the extension vs force RLC model predictions are mainly dependent on the amplitude σ_o^2 of the structural disorder and seem rather insensitive to the possible presence of long-range correlations in the sequence. In that respect, the RLC model can provide decisive test simulations of the pertinence of experimentally established dinucleotide and trinucleotide structural coding tables. Our results should encourage further experiments on the sequence-dependent response of DNA chains to external stretching constraints, as well as motivate molecular dynamics studies of the mechanical properties of DNA at the base-pair level [20,21]. The simultaneous knowledge of the intrinsic local structural disorder $\{\omega_i(s)\}$ and of the local bend stiffness $\kappa(s)=A(s)k_BT$ and twist stiffness $\tilde{\kappa}(s)=C(s)k_BT$ at the scale of one or two helical pitches would open the door to parameter free RLC modeling.

We would like to thank J. F. Allemand, D. Bensimon, V. Croquette, J. H. Maddocks, and Y. d'Aubenton-Carafa for valuable discussions. B.A. was financially supported by the European Community (Grant No. HPMF-CT-2001-01321). This work was supported by the Action Bioinformatique 2000 (CNRS).

[1] K. Luger *et al.*, Nature (London) **389**, 251 (1997); J. Widom, Annu. Rev. Biophys. Biomol. Struct. **27**, 285 (1998).

[2] K. E. van Holde, *Chromatin* (Springer, New York, 1988); A. P. Wolffe, *Chromatin Structure and Function*, 3rd ed. (Academic Press, London, 1998).

[3] I. Ioshikhes *et al.*, J. Mol. Biol. **262**, 129 (1996); A. Thaström *et al.*, *ibid.* **288**, 213 (1999).

[4] B. Audit *et al.*, Phys. Rev. Lett. **86**, 2471 (2001); J. Mol. Biol. **316**, 903 (2002).

[5] S.B. Smith *et al.*, Science **258**, 1122 (1992); P. Cluzel *et al.*, *ibid.* **271**, 792 (1996).

[6] T.R. Strick *et al.*, Science **271**, 1835 (1996).

[7] J.F. Léger, *et al.*, Phys. Rev. Lett. **83**, 1066 (1999).

[8] (a) C. Bustamante *et al.*, Science **265**, 1599 (1994); (b) A. Vologodskii, Macromolecules **27**, 5623 (1994); (c) C. Bouchiat *et al.*, Biophys. J. **76**, 409 (1999).

[9] (a) J.F. Marko and E.D. Siggia, Phys. Rev. E **52**, 2912 (1995); (b) J.D. Moroz and P. Nelson, Proc. Natl. Acad. Sci. U.S.A. **94**, 14 418 (1997); (c) C. Bouchiat and M. Mézard, Eur. Phys. J. E **2**, 377 (2000).

[10] E. N. Trifonov *et al.*, in *DNA Bending and Curvature*, edited by W. K. Olson *et al.* (Adenine Press, Schenectady, 1987), p. 243.

[11] J.A. Schellman and S.C. Harvey, Biophys. Chem. **55**, 95 (1995).

[12] L. Song and J.M. Schurr, Biopolymers **30**, 223 (1990).

[13] J. Bednar *et al.*, J. Mol. Biol. **254**, 579 (1995); P. Furrer *et al.*, *ibid.* **266**, 711 (1997).

[14] P. Nelson Phys. Rev. Lett. **80**, 5810 (1998).

[15] D. Bensimon *et al.*, Europhys. Lett. **42**, 97 (1998).

[16] N.B. Ulyanov and T.L. James, Methods Enzymol. **261**, 90 (1995).

[17] A. Gorin *et al.*, J. Mol. Biol. **247**, 34 (1995).

[18] B. Audit *et al.*, IEEE Trans. Info. Theory **48**, 2938 (2002).

[19] A. Arnéodo *et al.*, Eur. Phys. J. B **1**, 259 (1998).

[20] F. Lankas *et al.*, J. Mol. Biol. **299**, 695 (2000).

[21] R.S. Manning *et al.*, J. Chem. Phys. **105**, 5626 (1996).