Base Pair Openings and Temperature Dependence of DNA Flexibility

Nikos Theodorakopoulos[1,2,3] and Michel Peyrard[3]

[1]*Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation, Vasileos Constantinou 48, 116 35 Athens, Greece*

[2]*Fachbereich Physik, Universität Konstanz, 78457 Konstanz, Germany*

[3]*Ecole Normale Supérieure de Lyon, Laboratoire de Physique CNRS, 46 allée d'Italie, 69364 Lyon Cedex 7, France*

(Received 4 October 2011; published 16 February 2012)

The relationship of base pair openings to DNA flexibility is examined. Published experimental data on the temperature dependence of the persistence length by two different groups are well described in terms of an inhomogeneous Kratky-Porot model with soft and hard joints, corresponding to open and closed base pairs, and sequence-dependent statistical information about the state of each pair provided by a Peyrard-Bishop-Dauxois (PBD) model calculation with no freely adjustable parameters.

DOI: 10.1103/PhysRevLett.108.078104 PACS numbers: 87.15.−v, 87.10.Pq, 87.14.gk

The bending flexibility of DNA is essential in key biological processes, such as its packing in chromatin or in viruses. While single strands of DNA (or RNA) are highly flexible owing to an almost free rotation around single chemical bonds, the double helical structure with its hydrogen-bonded stacked base pairs favors a very rigid configuration. One may ask what are the natural limits of this rigidity. This question has been addressed almost since the discovery of DNA structure and has led to controversial results.

A measure of the flexibility of the molecule is provided by its *persistence length*, which is the correlation length for the thermal fluctuations of a vector tangent to the helix axis. Early measurements were made by light scattering methods [1]. The temperature dependence of persistence length was measured by Gray and Hearst [2] in sedimentation experiments. The analysis of these results by a model that describes DNA as a flexible rod, the so-called wormlike chain (WLC) model, led the authors to suggest that a statistically homogeneous WLC description did not apply, and to suggest that, instead of a smooth bending, the flexibility could come from sharp bends, at ''flexible joints.'' This appeared consistent with the observation of ''DNA breathing'' by von Hippel and co-workers [3]. A denaturation of the double helix, forming local single strands, would provide those ''joints.'' Another hint of the possibility of sharp bends came from cyclization measurements of short chains [4] which gave a much higher probability to find the two ends of a short DNA molecule next to each other than predicted by a homogeneous WLC model leading to smooth bendings. A theoretical analysis [5] showed that a localized single-stranded bubble mechanism was a possible explanation, compatible with the energetics of the thermal fluctuations of DNA. However, the interpretation of cyclization experiments was criticized in later work [6] identifying the high ligase protein used in [4] as the source of a possible artefact.

As the local opening of DNA can be thermally induced, a study of the temperature dependence of the persistence length should provide a clue. In fact recent measurements [7] performed between 5 and 60 C reveal a much stronger variation of the persistence length than that either found in earlier work [2] or predicted by a WLC model with a fixed stiffness constant. The authors of Ref. [7] ruled out the possibility that local openings could be responsible for this behavior on the basis of the low opening probabilities derived from the rate of exchange of the protons involved in the base-pair hydrogen bonds when DNA is in solution in deuterated water [8].

A promising experimental route to resolve the issue of whether local openings are relevant to increased DNA flexibility would be to examine the strongly fluctuating premelting and melting regimes. It turns out that this was done many years ago in a pioneering study of magnetic birefringence of DNA solution by Maret and co-workers [9]. In this work we interpret the results of [9] in terms of base pair openings whose populations can be determined by a lattice-dynamically motivated mesoscopic model of DNA. Furthermore, we apply these ideas to a direct calculation of the average end-to-end distance which allows us to compare predicted effective persistence length with the results of [7]. Our analysis of the two experiments in terms of standard model ideas of heterogenous polymer elasticity and DNA bubbles describes the full temperature dependence of the two experimental results and suggests that a quantitative link between base pair openings and local flexibility can be detected in measured properties of DNA solutions.

We model the elastic behavior of a DNA chain of N base pairs in a magnetic field B along the z axis in terms of a heterogeneous Kratky-Porod chain [10]

$$H_{\mathrm{KP}} = -\sum_{j=1}^{N-1} J_{j,j+1}\vec{S}_j \cdot \vec{S}_{j+1} - \frac{1}{2}\Delta\chi \sum_{j=1}^{N}(BS_j^z)^2 \qquad (1)$$

where $\vec{S}_j$ is the unit vector describing the direction of the jth base-pair plane. The first term is simply a discrete version of the WLC model, with a local stiffness constant $J_{j,j+1}$. The second term describes the contribution of the magnetic energy. $\Delta\chi = \chi_{||} - \chi_\perp$ is the anisotropy of the magnetic susceptibility of a single base pair, known to be negative (diamagnetic) in a fixed molecular frame where the parallel direction is along the axis of the helix. The $(S_j^z)^2 \equiv \cos^2\theta_j$ factor follows from the Euler (θ, ϕ) rotation which transforms the susceptibility tensor to the laboratory frame if one integrates over all possible azimuthal angles (axial symmetry). The stiffness constants will be taken as either J (stiff) or J' (soft) with an appropriate probability distribution (cf. below).

The magnetically induced optical birefringence $\Delta n = n_{||} - n_\perp$ of the refractive index along directions parallel and perpendicular to the field, measured in [9], is produced by the difference in the (optical frequency) dielectric constant which originates in the (electrically) anisotropically polarizable units, the DNA base pairs. Standard electromagnetic theory implies

$$\Delta n = \frac{2\pi}{n}\rho\Delta\alpha\hat{Q} \tag{2}$$

where n is the average index of refraction, $\Delta\alpha = \alpha_{||} - \alpha_\perp$ the anisotropy of the electronic polarizability tensor of a single base pair in the molecular frame, ρ the number of monomers (base pairs) per unit volume, and

$$\hat{Q} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{2}\langle(3\cos^2\theta_j - 1)\rangle \tag{3}$$

is an averaged orientational factor whose origin can be traced, in analogy with the second term in the magnetic Hamiltonian (1), to the anisotropy of the axially symmetric polarizability tensor in the laboratory frame. Following [11] we consider any local field corrections to be incorporated in effective values of the polarizability. The thermal averaging is performed over conformations of the molecule according to (1). Since, even for high magnetic fields, $\Delta\chi B^2$ is much smaller than the typical thermal energies k_BT, the averaging of the orientational factor can be evaluated to leading order in B^2

$$\hat{Q} = \Delta\chi\frac{B^2}{2k_BT}Q,$$

$$\text{where } Q = \frac{1}{N}\sum_{i,j=1}^{N}\langle\frac{1}{2}(3\cos^2\theta_j - 1)\cos^2\theta_i\rangle_0$$

and the 0 subscript implies averaging over field-free configurations of the KP chain (1). For a homogeneous KP chain of ν monomers the value of Q is [12]

$$Q_\nu = \frac{2}{15}\left\{\nu\frac{1+u}{1-u} - 2\frac{u(1-u^\nu)}{(1-u)^2}\right\} \tag{4}$$

where $u = 1 - (3/K)(\coth K - 1/K)$, $K = J/k_BT$. For the case of DNA, the orientational averages are computed in an heterogeneous KP ensemble where the stiffness constants depend on whether the molecule is locally in a double-stranded or open configuration

$$J_{j,j+1} = (1 - P_0^{j,j+1})J + P_0^{j,j+1}J'. \tag{5}$$

$P_0^{j,j+1}$ is the joint probability that base pairs (j) and $(j+1)$ are open. It is obtained from an exact calculation of the partition function of the mesoscopic PBD Hamiltonian [13], which associates a transverse coordinate y_n with each base pair and is known to describe the thermal denaturation of long DNA chains with known sequence and salt content, with the method described in Ref. [14]. The model parameters, determined in earlier analysis of melting profiles [15] have *not* been adjusted for this study.

Figure 1 compares theoretical predictions as outlined above with the experimental data on calf thymus DNA. The experimental values of the dimensionless quantity Q have been extracted from the measured birefringence values [9] using a mean index of refraction $n = 1.33$ [11], and the values $\Delta\chi = -1.55 \times 10^{-20}$ erg/T^2 [16] and $\Delta\alpha = -18.2$ Å^3 [17] for the anisotropies, respectively, of the diamagnetic susceptibility and electronic polarizability per base pair. The parameter J has been set to $J = 7.02 \times 10^{-12}$ erg which corresponds to a room temperature persistence length of 59 nm. Consistently with the parameters of the PBD model [15] which introduce a ratio of 50 between the base-pair interactions for stacked or unstacked bases, we set $J' = J/50 = 0.14 \times 10^{-12}$ erg, which gives a persistence length of 1.87 nm for single-stranded DNA. For reference purposes, we show on Fig. 1 the experimental calf thymus melting

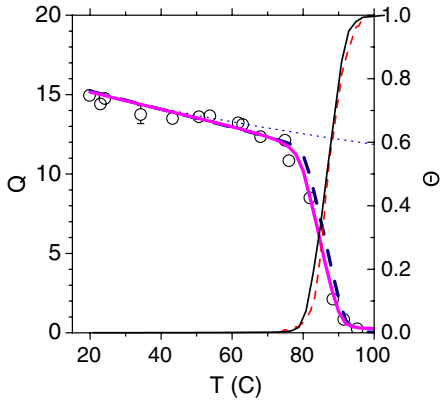

FIG. 1 (color online). The quantity Q (left y axis), as extracted from measurements of magnetic birefringence (circles, adapted from [9]), and calculated (i) on the basis of Eq. (7) (dashed curve), (ii) from the heterogeneous KP model (full line); the dotted curve shows a $1/T$ behavior for comparison. Also shown (right y axis) is the measured melting fraction for a similar sample [18] (dotted line) and as calculated from a 290 kb sequence of *bos taurus* [19] (thin full line).

profile [18] and a PBD-model computed profile from a 290 kbp long segment of bovine chromosome with a similar melting temperature [19]. The computed profile has been obtained with the model parameters and the numerical procedure described in [15].

To complete this study we would like to present an alternative way of computing the dimensionless quantity Q using PBD-model input on base-pair openings, following the analysis used recently in the interpretation of neutron scattering experiments on B-DNA [20]. Although slightly less accurate than the previous numeric evaluation, the alternative route illustrates some of the underlying properties of open vs closed base pairs. According to that picture, neutron diffraction—and in the present case magnetic birefringence—is produced from oriented intact clusters of bound base pairs. An intact cluster of length ν (defined as enclosed by open base pairs on both sides) occurs with a probability P_ν which satisfies the sum rules [14] $\sum_{\nu=0}^{\infty} P_\nu = \Theta$, $\sum_{\nu=1}^{\infty} \nu P_\nu = 1 - \Theta$, where Θ is the overall fraction of open base pairs. It follows that the average size of intact clusters (measured in base pairs)

$$s = \frac{\sum_{\nu=1}^{\infty} \nu P_\nu}{\sum_{\nu=1}^{\infty} P_\nu} = \frac{1-\Theta}{\Theta - P_0} \tag{6}$$

can be determined in terms of Θ and the "zero size cluster probability" P_0 (cf. Fig. 2); note that the latter quantity is in fact the joint probability of two successive base pairs being in the open state. Moreover, the P_ν's behave exponentially, i.e., $P_\nu = Ae^{-\sigma\nu}$, $\nu \geq 1$ where $A = (1-\Theta)/[s(s-1)]$ and $\sigma = -\ln(1-1/s)$, to a high degree of accuracy [14,20].

It is therefore possible, in an approximate sense which neglects contributions to birefringence from open base pairs, to evaluate Q as an average over all intact clusters, treated as if they were freely jointed

$$Q = \sum_{\nu=1}^{\infty} P_\nu Q_\nu \sim \frac{4}{45} K(1-\Theta) \frac{1}{1+(K/3s)}, \tag{7}$$

where it should be recognized that for typical ds-DNA parameters, $K \gg 1$ is very nearly equal to the persistence length λ expressed in units of the monomer distance $a = 0.34$ nm. The dashed curve in Fig. 1 expresses Eq. (7) for the same $J = 7.02 \times 10^{-12}$ erg as above. The agreement between the theoretical predictions and the magnetic birefringence experiments (Fig. 1), in the whole temperature range from room to melting temperatures, validates this approach which implies that fluctuational openings and persistence length of DNA are intimately linked, even at temperatures significantly below melting.

We now turn our attention to the more recent set of persistence length measurements which were based on cyclization of λ-phage DNA fragments [7]. The overall temperature dependence of the persistence length appears (Fig. 3) to be stronger than the $1/T$ predicted by the homogeneous KP model. In what follows we will show that such a strong temperature dependence is entirely consistent with the bubble-based scenario proposed above for the description of the birefringence data. We consider a heterogeneous KP model with stiffness constants J or J' determined with the probabilities given by the PBD model [using Eq. (5), $J' = 0.14 \times 10^{-12}$ erg and slightly varying J]. It can be used to compute a theoretical end-to-end distance, based on the bubble-based scenario, for the 200 bp sequence of the λ phage for a Na^+ concentration of 0.004 M (cf. [7]), from

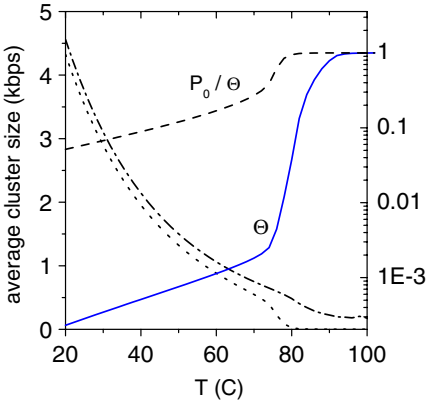

FIG. 2 (color online). Some average properties of the bovine sequence [19]. Left y-scale: the average size (in bps) of intact clusters (dash-dotted curve); the inverse of the melting fraction (dotted curve). Right y scale: the melting fraction Θ (solid curve) and the conditional probability P_0/Θ that if a base pair is in the open state, the next one will also be in the open state (dashed curve); note that the latter quantity is much larger than Θ (bubble aggregation tendency).

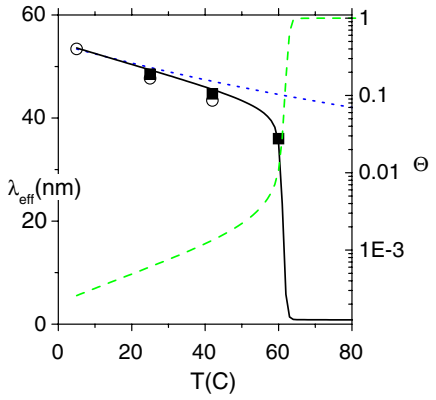

FIG. 3 (color online). Persistence length obtained by cyclization of DNA fragments (filled squares, open circles, redrawn from Ref. [7]). The solid curve shows the estimate obtained from the inhomogeneous KP chain with stiff and soft joints distributed according to (5) and probabilities from a PBD calculation. Also shown is a $1/T$ reference curve (dotted line), and the melting fraction (dashed line, right y axis) obtained from the PBD calculation.

$$\langle R^2 \rangle = \frac{1}{N} \sum_{i,j=1}^{N} \langle \vec{S}_i \cdot \vec{S}_j \rangle. \quad (8)$$

A subsequent analysis based on the homogeneous WLC model as in [7,21], with the same total length $L = Na$, i.e., with

$$\langle R^2 \rangle = 2L\lambda - 2\lambda^2(1 - e^{-L/\lambda}) \quad (9)$$

allows us to extract an effective persistence length which can be directly compared with the experimental data. The resulting effective persistence length, obtained for $J = 6.13 \times 10^{-12}$ erg, plotted in Fig. 3, shows that the data of [7] are consistent with a bubble-based analysis.

We have shown that the results of the two experiments which measured DNA persistence length in a temperature range sufficiently broad to detect deviations from a $1/T$ behavior can be fitted by adopting a *single* approach combining a discrete version of the WLC model with the statistical physics of the base-pair fluctuational openings. Three comments are in order here.

First, models which do not include fluctuational openings lead to a persistence length that varies as $1/T$, in significant disagreement with experiments. The data of Ref. [7] show that, in the 5–42 °C range the observed variation is 1.45 times bigger than predicted by an $1/T$ law.

The second comment concerns compatibility of our results with the low densities ($< 10^{-5}$) of fluctuational openings and the associated long lifetimes (~ 10 msec) of base pairs measured by imino proton exchange [8]. The results of Ref. [8] show very clearly that the activation enthalpy for the opening, which stays roughly constant in the 0–25 °C range, increases significantly above 25 °C. In other words there is an acceleration of the exponential rate at which lifetimes decrease with increasing temperature. A straightforward extrapolation of data appearing in Fig. 12 of [8b] suggests an increase by a factor of 10 between 25 °C and 40 °C. Thus, events considered as exceptional at room temperature become more plausible at a temperature close to physiological temperature. Moreover it should be noted that relatively few disruptions are sufficient to significantly change the persistence length. Perhaps the most striking indirect evidence of the role of small numbers of fluctuational openings in the premelting regime of genomic DNA comes from viscosity measurements [22].

The third comment refers to possible extensions and applications of the present work. The approach used here, owing to its computational efficiency, can be exploited in its present form for a systematic study of sequence effects on local flexibility [23]. On the other hand, it is likely that extended soft regions of the DNA molecule may require a less minimalist approach than the one presented here. In particular, it is possible that a more complex interplay between the torsional, bending, and local melting [24] effects may determine the details of the premelting regime of cyclization experiments; as more experimental data become available, generalizations of the present work in that direction may become fruitful.

As a final remark, it may be noted that the quantitative link between fluctuational openings and local flexibility explored in this work provides further evidence that the dynamical character of DNA should not be underestimated, even away from its melting transition.

N. T. acknowledges support by the program Accueil-Pro of Région Rhône-Alpes.

[1] A. Peterlin, Nature (London) **171**, 259 (1953); J. Polym. Sci. **10**, 425 (1953).
[2] H. B. Gray and J. E. Hearst, J. Mol. Biol. **35**, 111 (1968).
[3] M. P. Printz and P. H. von Hippel, Proc. Natl. Acad. Sci. U.S.A. **53**, 363 (1965); P. H. von Hippel and Kwok-Ying Wong, J. Mol. Biol. **61**, 587 (1971).
[4] T. E. Cloutier and J. Widom, Mol. Cell **14**, 355 (2004).
[5] J. Yan and J. F. Marko, Phys. Rev. Lett. **93**, 108108 (2004).
[6] Q. Du, C. Smith, N. Shiffeldrim, M. Vologodskaia, and A. Vologodskii, Proc. Natl. Acad. Sci. U.S.A. **102**, 5397 (2005).
[7] S. Geggier, A. Kotlyar, and A. Vologodskii, Nucleic Acids Res. **39**, 1419 (2011).
[8] M. Gueron and M. Kochoyan, J-L Leroy, Nature (London) **328**, 89 (1987); J-L Leroy, M. Kochoyan, T. Huynh-Dinh, and M. Gueron, J. Mol. Biol. **200**, 223 (1988).
[9] G. Maret, M. v. Schickfus, A. Meyer, and K. Dransfeld, Phys. Rev. Lett. **35**, 397 (1975).
[10] O. Kratky and G. Porod, Recl. Trav. Chim Pays-Bas **68**, 1106 (1949).
[11] G. Maret and G. Weill, Biopolymers **22**, 2727 (1983).
[12] R. W. Wilson, Biopolymers **17**, 1811 (1978); J.-G. Hagmann, K. K. Kozlowski, N. Theodorakopoulos, and M. Peyrard, J. Stat. Mech. (2009) P04011.
[13] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989); T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
[14] N. Theodorakopoulos, J. Nonlinear Math. Phys. **18**, 429 (2011).
[15] N. Theodorakopoulos, Phys. Rev. E **82**, 021905 (2010).
[16] D. L. Bryce, J. Boisbouvier, and A. Bax, J. Am. Chem. Soc. **126**, 10 820 (2004).
[17] N. C. Stellwagen, Biopolymers **20**, 399 (1981).
[18] J. Marmur and P. Doty, Nature (London) **183**, 1427 (1959).
[19] Gene Bank, Bos taurus breed Hereford chromosome 12 (NW_001848864.1).
[20] A. Wildes, N. Theodorakopoulos, J. Valle-Orero, S. Cuesta-López, J-L Garden, and M. Peyrard, Phys. Rev. Lett. **106**, 048101 (2011); Phys. Rev. E **83**, 061923 (2011).
[21] M. Vologodskaia and A. Vologodskii, J. Mol. Biol. **317**, 205 (2002).
[22] A-M. Freund and G. Bernardi, Nature (London) **200**, 1318 (1963).
[23] G. Weber, J. W. Essex, and C. Neylon, Nature Phys. **5**, 769 (2009).
[24] H. Chen, Y. Liu, Z. Zhou, L. Hu, Z-C Ou-Yang, and J. Yan, Phys. Rev. E **79**, 041926 (2009).