

Dependence on temperature and guanine-cytosine content of bubble length distributions in DNA

G. Kalosakas^{1,a)} and S. Ares²

¹*Department of Materials Science, University of Patras, GR-26504 Rio, Greece*

²*Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, D-01187 Dresden, Germany and Grupo Interdisciplinar de Sistemas Complejos (GISC)*

(Received 16 January 2009; accepted 14 May 2009; published online 18 June 2009)

We present numerical results on the temperature dependence of the distribution of bubble lengths in DNA segments of various guanine-cytosine (GC) concentrations. Base-pair openings are described by the Peyrard–Bishop–Dauxois model and the corresponding thermal equilibrium distributions of bubbles are obtained through Monte Carlo calculations for bubble sizes up to the order of a hundred base pairs. The dependence of the parameters of bubble length distribution on temperature and the GC content is investigated. We provide simple expressions which approximately describe these relations. The variation of the average bubble length is also presented. We find a temperature dependence of the exponent c that appears in the distribution of bubble lengths. If an analogous dependence exists in the loop entropy exponent of real DNA, it may be relevant to understand overstretching in force-extension experiments. © 2009 American Institute of Physics.

[DOI: [10.1063/1.3149859](https://doi.org/10.1063/1.3149859)]

I. INTRODUCTION

Local openings of the DNA double helix are required in several biological functions, for instance, transcription and replication. These local separations of the two DNA strands are mediated by a specific machinery in the cell. In order to deal with such complex processes, it is necessary first to understand the interactions keeping together the two complementary strands within a single DNA duplex, as well as the properties of fluctuating DNA openings in thermal equilibrium. This fact has been realized long time ago and base-pairing interactions and the thermal stability of the double helix have been conventionally probed by increasing temperature up to the DNA denaturation transition.^{1,2}

Base-pair openings (bubbles) occur in DNA due to thermal fluctuations even at temperatures well below the melting transition. It has been speculated that they may play a role in the recognition of specific DNA sites by DNA-binding proteins.^{3–6} Recent experiments using a novel hairpin quenching technique have been able to show this bubble formation in the premelting regime and characterize it quantitatively.^{7–9} By increasing temperature these bubbles grow and more bubbles are nucleated, thus leading to the complete separation of the two strands at the denaturation transition. Therefore statistical properties of DNA bubbles in a wide temperature regime, extending from biological temperatures up to the melting transition, are of particular interest. The purpose of this work is the investigation of the distribution of bubble lengths and its temperature dependence for DNA sequences containing different percentages of guanine-cytosine (GC) base pairs. The study is performed in

the framework of the Peyrard–Bishop–Dauxois (PBD) model,¹⁰ from where conjectures are drawn about the actual distributions of bubbles in DNA.

If our results on the temperature dependence of this distribution, obtained for finite bubble lengths, remain qualitatively valid in the asymptotic regime (for very large values of bubble lengths), then there would be a connection with the ongoing discussion about the interpretation of the DNA overstretching observed in force-extension experiments. In this case the possibilities of either a force-induced melting or the existence of a double-stranded elongated DNA phase (the so-called S-DNA) have been proposed to explain the abrupt elongation of DNA at forces ~ 65 pN.¹¹ This is briefly discussed in the concluding section.

In a recent study we have presented the distribution of bubble lengths in the PBD model of DNA at 310 K and we found that it can be described by a power-law modified exponential.¹² Anharmonic interactions between complementary bases forming base-pairs are responsible for the observed nonexponential distribution. The same form of distribution has been derived in the framework of the Poland–Scheraga model,^{13–15} which represents a completely different theoretical approach of DNA denaturation than the PBD model that we use in our calculations. This distribution is also found for a primitive version of the PBD model, viz., the Peyrard–Bishop model¹⁶ with linear stacking interactions, but in this case the characteristic values of the parameters of the distribution are different.¹⁷

Here we examine how the bubble length distribution varies with temperature and present its complete dependence on both temperature and the GC fraction of the DNA segment. The PBD model¹⁰ is used for the description of base-pair openings, where a set of continuous variables y_n represent the base-pair displacements from equilibrium distance and

^{a)}Electronic mail: georgek@upatras.gr.

the index n labels the base-pairs along the DNA chain. The potential energy of the system consists of two parts: the on-site interaction $V(y_n)$ within each base-pair and the stacking interaction $U(y_n, y_{n+1})$ between adjacent base-pairs. A Morse potential is used for the on-site energy,

$$V(y_n) = D_n(e^{-a_n y_n} - 1)^2, \quad (1)$$

where the parameters D_n and a_n distinguish between G-C and A-T base-pairs ($D_{GC}=0.075$ eV, $a_{GC}=6.9$ Å⁻¹ for a G-C base-pair and $D_{AT}=0.05$ eV, $a_{AT}=4.2$ Å⁻¹ for an A-T pair), while a nonlinear potential describes the stacking interaction,

$$U(y_n, y_{n+1}) = \frac{K}{2}(1 + \rho e^{-b(y_n + y_{n+1})})(y_n - y_{n+1})^2, \quad (2)$$

with $K=0.025$ eV/Å², $\rho=2$, and $b=0.35$ Å⁻¹. We use parameter values from previous works.^{4,5,18,19} These parameters have been originally obtained from empirical fits to experimental data¹⁸ and are also able to successfully describe other experimental situations.^{4,19} Entropic effects of DNA are described from the PBD Hamiltonian when studying its thermodynamics, leading to an entropy driven melting transition.^{10,20}

The efficiency of the rather simple PBD model to describe base-pair openings in DNA has led to its extensive use in literature.^{4-6,12,18-32} In particular, our choice of the PBD model for the study of bubble length distributions is motivated by the success of the model¹⁹ in reproducing experimental measurements of bubble formation,^{7,8} as well as the accurate description at a quantitative level of melting curves in short DNA segments.¹⁸ The coarse-grained description of the model allows the possibility to perform calculations with long sequences of up to tens of thousands of base-pairs.³⁰ Moreover, the popular nearest-neighbor model,³³ which is very successful in describing the melting of short oligomers, does not reproduce well experimental data on intermediate states for longer sequences.^{8,9,34,35} Compared to the more conventional modified Ising-type models used in the study of DNA melting,^{2,13} the PBD model is qualitatively similar: the on-site potential of the PBD model is an extension of the *magnetic field* in the Ising-type models, while the nonlinear stacking potential in Eq. (2) corresponds to the interaction between neighboring spins and loop entropies in Ising models. We expect our results for the PBD model to hold qualitatively also for the Ising-type models, since both models yield the same shape for the distribution of bubble sizes.¹²

II. RESULTS

Considering a random DNA sequence of a given GC percentage, x_{GC} , at equilibrium at temperature T , we calculate the distribution per base-pair of bubble lengths l , $P(l)$, by counting during Monte Carlo simulations of the PBD model the average occurrences of openings (base-pair displacements) larger than a fixed threshold y_{thres} at l successive base-pairs. In each simulation a random sequence of 1000 base-pairs is used in which the AT or GC identity of each base-pair is generated randomly under the constraint of a specified GC percentage. The statistical properties of such

random long sequences are not different from natural sequences, as we showed in previous work¹² comparing the results for a segment from *Escherichia coli*'s *gal* promoter with those from a randomly generated sequence with the same GC percentage.

The Monte Carlo simulation is performed using the Metropolis algorithm,³⁶ which is used first for thermalization and then for measurement runs. Results are averaged over several realizations using the same random sequence and different initial conditions for the interbase distance within each base-pair and for the random number generator. Other details of the simulation (number of realizations and Monte Carlo steps) are as in Ref. 12. Since the used sequences are so long, the results are independent of the precise realization of the random sequence. We have run further simulations using different random sequences to assure this point. There is an issue about the dissociation observed in the PBD model in Monte Carlo simulations. In particular, due to the upper bound of the on-site potential for large positive displacements, for long enough simulations a complete dissociation would eventually occur at any temperature, even below the melting transition.^{20,37} However, because of the large DNA segments considered here, the probability of a complete dissociation during the simulation time is so low that no such events were observed at any of the studied temperatures below the melting temperature.

The threshold value considered for the openings is $y_{\text{thres}}=1.5$ Å. Results for different values of this threshold are qualitatively similar. We use periodic boundary conditions in DNA segments of length 1000 base-pairs (sufficiently larger than the studied bubble sizes) and therefore our results refer to internal bubbles in long DNA chains. The results are independent of the length of the studied sequence, provided that it is longer than the longest observed bubbles.¹² This has been verified by simulations on DNA segments of different length, resulting in the same distributions $P(l)$ for the bubble sizes studied here, i.e., with l up to about a hundred base-pairs. For short molecules, the actual sequence does play a role, since it can trigger local openings in different regions at the end or in the middle of the sequence.^{7,8} In this case, the border effects associated cannot be characterized only by the GC percentage of the short sequence, but this problem is out of the scope of the present work.

In Fig. 1 we show bubble length distributions at various temperatures for a GC percentage of $x_{GC}=50\%$ and 87.5% . Similar plots have been obtained for nine different values of x_{GC} : 0%, 12.5%, 25%, 37.5%, ..., 100%. Note here that although the PBD model takes into account cooperativity in bubble formation through the nonlinear stacking interaction, it does not contain any parameter describing a bubble nucleation size. Thus the distributions in Fig. 1 show results even for bubbles of length $l=1$, implying that the nucleation size is 1. This should not be confused with the minimum length of a short DNA sequence necessary to sustain bubble states, which has been shown experimentally^{7,8} and theoretically within the PBD model¹⁹ to be greater than one.

Except for the cases of $l=1$ at relatively higher tempera-

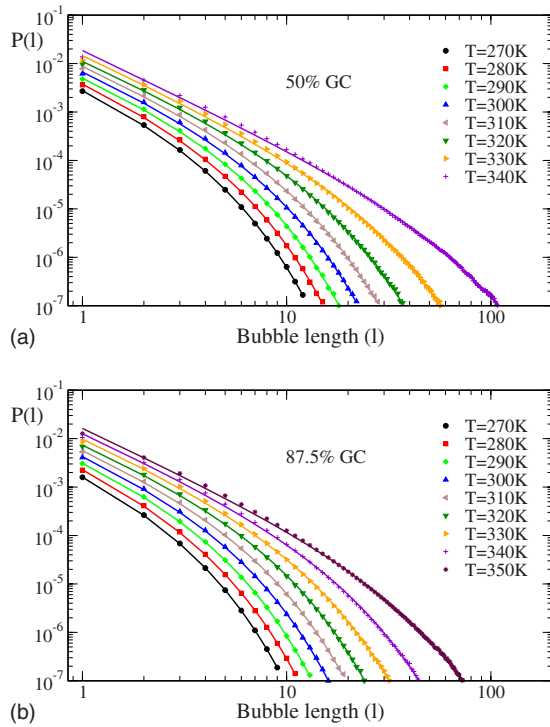


FIG. 1. (Color online) Distribution per base-pair of bubble lengths l (in number of base-pairs), $P(l)$, for different temperatures (points, as indicated in the plots) at random DNA sequences with a GC content of (a) 50% and (b) 87.5%. Continuous lines are fits with the distribution of Eq. (3). $y_{\text{thres}} = 1.5 \text{ \AA}$.

tures (closer to the melting transition), the numerical results obtained for the distributions can be rather well described by the power-law modified exponential¹²

$$P(l) = W \frac{e^{-l/c}}{l^c}, \quad \text{for } l > 1. \quad (3)$$

In the following we characterize the dependence of the parameters of distribution on T and x_{GC} , and provide approximate expressions for the relations $\xi(T, x_{\text{GC}})$, $c(T, x_{\text{GC}})$, and $W(T, x_{\text{GC}})$. The distribution parameters are obtained through fittings of plots like those of Fig. 1 with Eq. (3), using a weight proportional to $1/P(l)^2$.

We note here that Eq. (3) can be derived from polymer physics^{13–15} as an *asymptotic* expression for very large bubble sizes, $l \gg 1$, in DNA. However, we find that the same expression describes well the bubble length distribution in the PBD model of DNA, even for small bubble sizes l . We emphasize that we are interested here in bubble lengths up to about a hundred, or a few hundred at most, base-pairs (which are relevant for any practical purpose) and *not* for the asymptotic behavior of the distribution. In this context we use Eq. (3) as an empirical formula valid for finite bubble lengths l (in the non-asymptotic regime), and describe the variation of its parameters on T and x_{GC} . If one is interested in the asymptotic behavior of the distribution, this can be obtained by proper scaling analysis and, as it has been shown in Ref. 29, it may be described by different values of parameter c . Therefore we are *not* concerned with the order of the melting transition and the exponent c presented here is not indicative of the kind of the transition,^{14,15,22,29,35} as it also

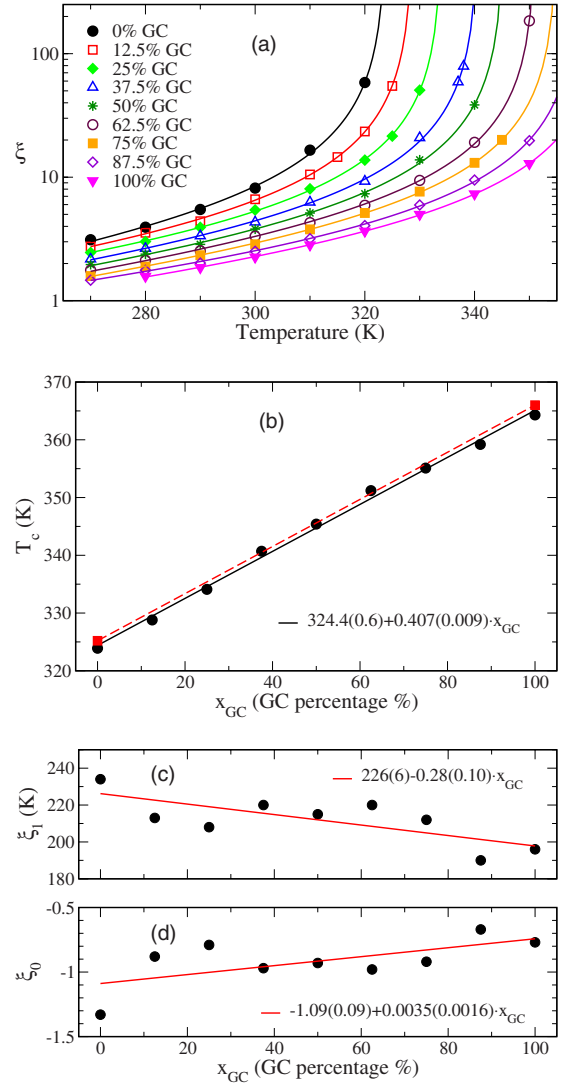


FIG. 2. (Color online) (a) Dependence of the decay length ξ of the distribution (3) on temperature, for different values of the GC content of the DNA sequence (points). Lines show fits with the function of Eq. (4). (b) Dependence of the critical temperature T_c , as obtained from the fitting of the $\xi(T)$ data with Eq. (4), on the GC content of the DNA sequence (circles). Solid line represents a least square fit according to a linear dependence. Squares show exact results of the critical temperatures for the homogeneous cases of 0% GC and 100% GC, obtained from transfer integral calculations, while the dashed line connects these two points. [(c) and (d)] Dependence of the parameters ξ_1 and ξ_0 , respectively, of the fit of the $\xi(T)$ data with Eq. (4), on the GC content of the sequence (circles). Solid lines represent linear fits of the corresponding data. Equations of straight lines resulting from the corresponding fittings are shown in (b)–(d), where the values in parentheses represent errors of the fitting parameters.

depends on the somehow arbitrary value of the threshold chosen to consider a base-pair open.

Figure 2 presents the dependence of the decay length ξ of Eq. (3). In Fig. 2(a) the variation of ξ with temperature is shown (points) for different values of x_{GC} . The T -dependence is accurately described by the divergent function

$$\xi(T) = \xi_0 + \frac{\xi_1}{T_c - T}, \quad (4)$$

where T_c is the denaturation temperature. Such a relation is also valid for the Poland–Scheraga model.¹⁵ Lines in Fig. 2(a) show fittings of the $\xi(T)$ data with Eq. (4), using a

weight proportional to $1/\xi^2$. The critical temperature T_c as obtained from the fitting at different values of x_{GC} is presented with circles in Fig. 2(b). A linear dependence of T_c on the GC content is found, in accordance with known experimental results³⁸ and calculations from simplified models.³⁹ A least square fitting of the $T_c(x_{GC})$ data results in the continuous line shown in Fig. 2(b). Regarding the homopolymer cases of poly(dA)-poly(dT) ($x_{GC}=0\%$) and poly(dG)-poly(dC) ($x_{GC}=100\%$), the critical temperatures for the transition can be independently calculated through the numerically exact transfer integral technique,^{40–42} and the corresponding results are $T_c(x_{GC}=0)=325.2$ K and $T_c(x_{GC}=100)=366.0$ K. These values are shown with squares joined by a dashed line in Fig. 2(b), where the line lies inside the error interval of the Monte Carlo results. The actual denaturation temperatures obtained through the Monte Carlo simulations are in agreement with those presented in Fig. 2(b). Therefore, the critical temperatures T_c shown in Fig. 2(b) provide a guide of how far from the melting transition are the data at various temperatures presented in this work.

The other parameters ξ_1 and ξ_0 resulting from the fitting of the $\xi(T)$ data with Eq. (4) are shown with circles in Figs. 2(c) and 2(d), respectively. Their dependence on x_{GC} can be approximately considered as linear [continuous lines in Figs. 2(c) and 2(d)]. Therefore, the relation

$$\xi(T, x_{GC}) = a_1 + a_2 x_{GC} + \frac{a_3 + a_4 x_{GC}}{a_5 + a_6 x_{GC} - T}, \quad (5)$$

where a_1, a_2, \dots, a_6 are constants, can approximately provide the dependence of the decay length ξ on T and x_{GC} .

The variation of the exponent c of distribution (3) is presented in Fig. 3. Points in Fig. 3(a) show the temperature dependence of c for different GC percentages (for clarity of the plot the corresponding results for $x_{GC}=12.5\%$, 37.5% , 62.5% , 87.5% have been omitted). This dependence may approximately be described by a linear function

$$c(T) = c_0 + c_1 T. \quad (6)$$

Lines in Fig. 3(a) show fittings of the numerical results with the above formula. The parameters c_1 and c_0 obtained from the fitting at different values of GC content are shown with points in Figs. 3(b) and 3(c), respectively, while the corresponding error bars are derived from the fitting procedure. The latter plots indicate a linear dependence of c_1 and c_0 on x_{GC} , implying the approximate relation

$$c(T, x_{GC}) = b_1 + b_2 x_{GC} + (b_3 + b_4 x_{GC})T, \quad (7)$$

where b_1, b_2, b_3 , and b_4 are constants independent of T and x_{GC} .

In Fig. 4 is shown the dependence of the coefficient W of distribution (3). Points in Fig. 4(a) display the variation of W with T for various GC contents. A quadratic function

$$W(T) = W_0 + W_1 T + W_2 T^2 \quad (8)$$

can describe rather well the temperature dependence of W , at least in the studied temperature regime. The corresponding fittings with Eq. (8) are shown by lines in Fig. 4(a). The

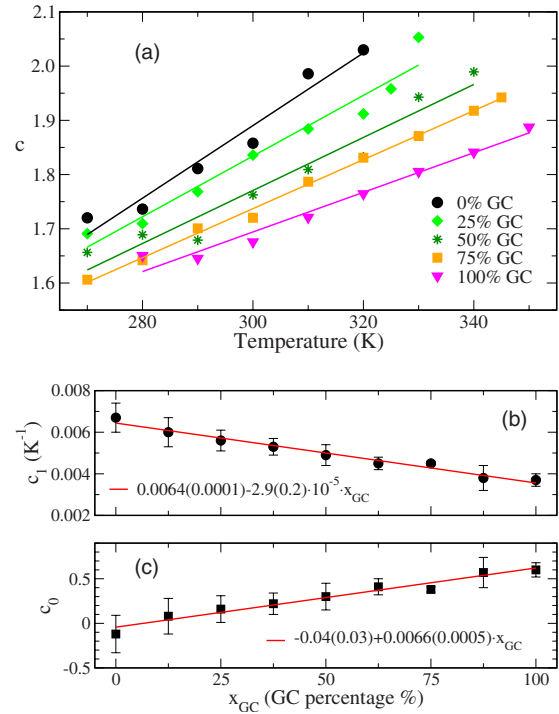


FIG. 3. (Color online) (a) Dependence of the exponent c of distribution (3) on temperature, for different values of the GC content of the DNA sequence (points). Lines show linear fits with Eq. (6). [(b) and (c)] Dependence of the parameters c_1 and c_0 , respectively, of the fit of the $c(T)$ data with Eq. (6), on the GC content of the sequence (filled points). Error bars are standard errors resulting from the fitting procedure. Solid lines represent linear fits of the corresponding data and the resulting equations are also shown.

resulting fitting parameters at different GC percentages are plotted by points in Figs. 4(b)–4(d), respectively. These plots can approximately be described by linear dependences of W_2 , W_1 , and W_0 on x_{GC} (solid lines). As a result the expression

$$W(T, x_{GC}) = d_1 + d_2 x_{GC} + (d_3 + d_4 x_{GC})T + (d_5 + d_6 x_{GC})T^2 \quad (9)$$

can approximate the dependence of W on T and x_{GC} , where d_1, d_2, \dots, d_6 are constants.

The detailed results of this investigation do not confirm a bilinear dependence of the parameters of the distribution on the GC content at a fixed temperature.¹² Instead, a linear dependence of the exponent c and the coefficient W on x_{GC} arises at constant T . We note that the results presented in Fig. 2 of Ref. 12 can be described well by Eqs. (5), (7), and (9) for $T=310$ K and using the values of constants as they provided in Figs. 2(b)–2(d) (for a_i), Figs. 3(b) and 3(c) (for b_i), and Figs. 4(b)–4(d) (for d_i).

We finally present the variation of the average bubble length, L_B , on the two parameters of interest, viz., temperature and GC percentage. L_B is obtained through the total number of base-pairs in bubble states divided by the total number of bubbles:¹²

$$L_B = \frac{\sum_l l P(l)}{\sum_{l \geq 1} P(l)}. \quad (10)$$

Substituting $P(l)$ by the power-law modified exponential function of Eq. (3), the sums in Eq. (10) yield

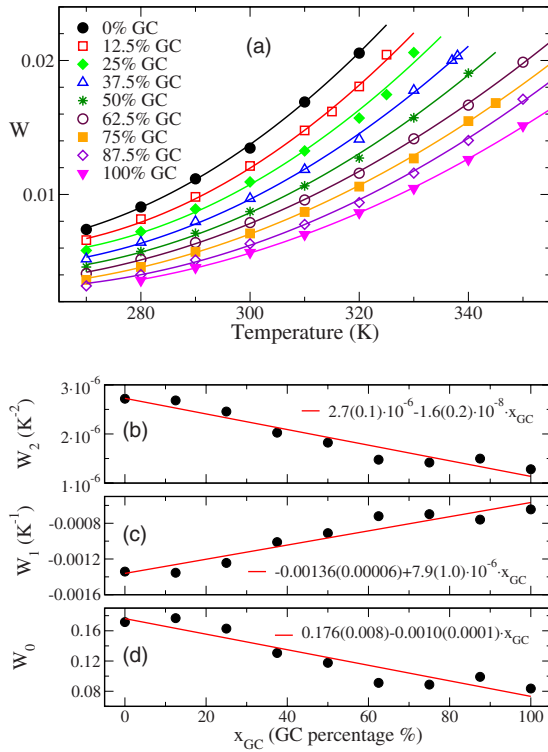


FIG. 4. (Color online) (a) Dependence of the pre-exponential coefficient W of distribution (3) on temperature, for different values of the GC content of the DNA sequence (points). Lines show quadratic fits with Eq. (8). [(b)–(d)] Dependence of the parameters W_2 , W_1 , and W_0 , respectively, of the fit of the $W(T)$ data with Eq. (8), on the GC content of the sequence (circles). Solid lines represent linear fits of the corresponding data and the resulting equations are also shown.

$$L_B(c, \xi) = \frac{\text{Li}_{c-1}(e^{-1/\xi})}{\text{Li}_c(e^{-1/\xi})}, \quad (11)$$

where $\text{Li}_s(z)$ is the first branch of the polylogarithm function,⁴³ defined as $\text{Li}_s(z) = \sum_{k=1}^{\infty} z^k / k^s$. The average bubble length for a given temperature and GC percentage, $L_B(T, x_{GC})$, can be obtained substituting in Eq. (11) the parameters ξ and c , calculated for the corresponding T and x_{GC} through Eqs. (5) and (7), respectively, using the values of constants determined in Figs. 2 and 3.

Figure 5(a) depicts the dependence of L_B on x_{GC} for different temperatures, obtained directly from our numerical simulations (points). Equation (11), shown by dotted lines in the plot, reproduces correctly the behavior of the data. However, the errors in the estimated parameters in Figs. 2 and 3 add up to produce a small deviation between the numerical points and the values obtained by Eq. (11). We also show, using solid lines in the plot, more accurate fits of the numerical data with a simpler phenomenological expression, namely, the exponentially decaying function

$$L_B = \Gamma_0 + \Gamma_1 \exp(-\Gamma_2 x_{GC}). \quad (12)$$

These empirical fits generalize the apparent exponential decay seen earlier at 310 K.¹² Presenting these data on a different way, more appropriate for experimental investigations, the temperature dependence of L_B for fixed GC contents is shown in Fig. 5(b) (points), for temperatures sufficiently below the melting transition. Here for clarity of the plot we do

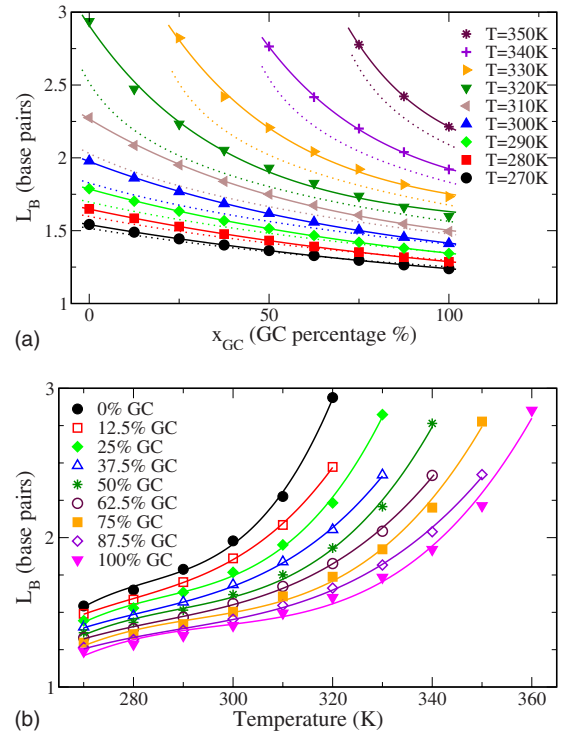


FIG. 5. (Color online) (a) Dependence of the average bubble length L_B , Eq. (10), on the GC content, for different temperatures (points). Dotted lines show the analytical result of Eq. (11), using the values of ξ and c from Eqs. (5) and (7). Solid lines show empirical fits with exponential functions, Eq. (12). (b) Dependence of the average bubble length L_B on temperature for different GC contents (points). Lines represent empirical fits with cubic functions.

not show the prediction of Eq. (11), but instead we show with lines fits with a simple phenomenological function, namely a cubic polynomial, which describes approximately the data in the investigated temperature regime. The coefficients of the cubic polynomials show an exponential dependence, of form (12), on the GC percentage.

III. CONCLUSIONS

In conclusion, we have presented the dependence of bubble length distributions in the PBD model of DNA on temperature and the GC content. The investigated temperature regime was extended from biologically relevant values up to values below the melting transition. Approximate expressions have been obtained for the parameters of the power-law modified exponential distribution in the non-asymptotic regime (for bubble lengths up to about a hundred base-pairs). The exponent c behaves linearly both in temperature and GC content, Eq. (7), while the coefficient W shows a quadratic dependence on temperature and a linear dependence on the GC fraction, Eq. (9). The decay length ξ is described by the relatively simple Eq. (5). The constants a_i , b_i , and d_i appearing in these expressions depend on the amplitude y_{thres} of the considered base-pair openings, and for $y_{\text{thres}} = 1.5 \text{ \AA}$ are given by the values shown in Figs. 2–4. Using the expressions of the exponent c and the decay length ξ , the average bubble length is analytically given by Eq. (11). Our results may be useful in biotechnological applications

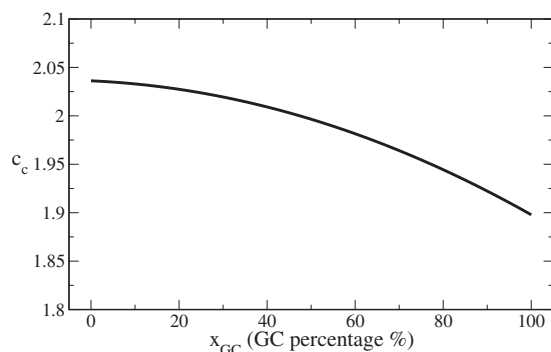


FIG. 6. Extrapolated dependence of the exponent c at the critical temperature (c_c) on the GC content.

which involve thermally induced DNA denaturation, as well as in recent hairpin quenching experiments⁷⁻⁹ for studying bubble formation.

An important result of this work is the prediction of a dependence both in temperature and genetic sequence of the exponent c of the bubble distribution (3). This prediction can be experimentally verified using the method proposed in Ref. 44: based on the measurement of certain correlation functions of base-pair openings using fluorescence correlation spectroscopy,⁴⁵ the exponent c can be experimentally calculated.

The sequence dependence of c is especially relevant at the melting temperature. At this temperature very large bubbles appear and the asymptotic behavior of the bubble length distribution is necessary. Assuming that the behavior revealed in our findings qualitatively holds in the asymptotic regime (which needs to be investigated since there is not *a priori* any reason for this to be true), then the particular dependence of c can be extrapolated. Combining an expression for c as that given in Eq. (7) and the dependence of critical temperature on the GC percentage presented in Fig. 2(b), we show in Fig. 6 estimated values of c at the melting temperature, c_c , as a function of the GC percentage. We see that c_c is not constant, but decreases as the GC fraction increases. Although, in view of the above assumptions, this result should be taken with caution, such a behavior is consistent with the observation, coming from lattice-models simulations that consider DNA strands as self-avoiding random walks, of a decrease in c_c with increasing stiffness of the strands.⁴⁶

Under the same assumption that our results for the temperature dependence of c qualitatively hold in the asymptotic regime, then this may be useful to understand experimental results in force-extension experiments. When performing single molecule experiments, force-extension curves of double-stranded DNA display hysteresis.^{47,48} The hysteresis observed is more pronounced at higher temperatures. The existence of an elongated double-stranded form of DNA, called S-DNA, has been proposed to explain the abrupt elongation of DNA at a force of ~ 65 pN.⁴⁹ Numerical studies of the force-extension problem⁵⁰ suggest that the existence of S-DNA is necessary to explain the temperature dependence of the observed hysteresis. However, an alternative explanation without the appearance of an S-DNA phase is possible if

the exponent c is temperature dependent.⁵⁰ The order of variation of c we find here is in the right direction (c increases with temperature, implying larger hysteresis loops at high temperatures according to the experiments) and of the necessary magnitude⁵⁰ to explain the hysteresis effects observed. Hence, if the qualitative behavior of our results is valid in the asymptotic regime, this would have an effect on the ongoing debate about S-DNA, favoring force-induced melting as an explanation for the DNA elongation over the formation of an S-DNA phase. Interestingly, recent experimental results support the force-induced melting explanation,⁵¹ in agreement with the expectation from the temperature dependence of c .

ACKNOWLEDGMENTS

We thank J. Bois for critical reading of this manuscript and J. Bois and N. Theodorakopoulos for enlightening discussions. G.K. acknowledges the hospitality of MPI-PKS in Dresden and support from the C. Carathéodory program C155 of University of Patras. S.A. acknowledges financial support from Ministerio de Educación y Ciencia (Spain) through grant MOSAICO. Both authors (G.K. and S.A.) contributed equally to this work.

- ¹R. M. Wartell and A. S. Benight, *Phys. Rep.* **126**, 67 (1985).
- ²D. Poland and H. A. Scheraga, *Theory of Helix Coil Transition in Biopolymers* (Academic, New York, 1970).
- ³A. Banerjee and H. M. Sobell, *J. Biomol. Struct. Dyn.* **1**, 253 (1983); H. M. Sobell, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5328 (1985).
- ⁴C. H. Choi, G. Kalosakas, K. Ø. Rasmussen, M. Hiromura, A. R. Bishop, and A. Usheva, *Nucleic Acids Res.* **32**, 1584 (2004).
- ⁵G. Kalosakas, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi, and A. Usheva, *Europhys. Lett.* **68**, 127 (2004).
- ⁶C. H. Choi, Z. Rapti, V. Gelev, M. R. Hacker, B. Alexandrov, E. J. Park, J. S. Park, N. Horikoshi, A. Smerzi, K. O. Rasmussen, A. R. Bishop, and A. Usheva, *Biophys. J.* **95**, 597 (2008).
- ⁷A. Montrichok, G. Gruner, and G. Zocchi, *Europhys. Lett.* **62**, 452 (2003); Y. Zeng, A. Montrichok, and G. Zocchi, *Phys. Rev. Lett.* **91**, 148101 (2003).
- ⁸Y. Zeng, A. Montrichok, and G. Zocchi, *J. Mol. Biol.* **339**, 67 (2004).
- ⁹Y. Zeng and G. Zocchi, *Biophys. J.* **90**, 4522 (2006).
- ¹⁰T. Dauxois, M. Peyrard, and A. R. Bishop, *Phys. Rev. E* **47**, R44 (1993).
- ¹¹M. C. Williams and I. Rouzina, *Curr. Opin. Struct. Biol.* **12**, 330 (2002).
- ¹²S. Ares and G. Kalosakas, *Nano Lett.* **7**, 307 (2007).
- ¹³D. Poland and H. A. Scheraga, *J. Chem. Phys.* **45**, 1464 (1966).
- ¹⁴Y. Kafri, D. Mukamel, and L. Peliti, *Eur. Phys. J. B* **27**, 135 (2002).
- ¹⁵B. Coluzzi and E. Yeramian, *Eur. Phys. J. B* **56**, 349 (2007).
- ¹⁶M. Peyrard and A. R. Bishop, *Phys. Rev. Lett.* **62**, 2755 (1989).
- ¹⁷W. Sung and J.-H. Jeon, *Phys. Rev. E* **69**, 031902 (2004); J.-H. Jeon, W. Sung, and F. H. Ree, *J. Chem. Phys.* **124**, 164905 (2006); J.-H. Jeon, P. J. Park, and W. Sung, *ibid.* **125**, 164901 (2006).
- ¹⁸A. Campa and A. Giansanti, *Phys. Rev. E* **58**, 3585 (1998).
- ¹⁹S. Ares, N. K. Voulgarakis, K. Ø. Rasmussen, and A. R. Bishop, *Phys. Rev. Lett.* **94**, 035504 (2005).
- ²⁰T. Dauxois and M. Peyrard, *Phys. Rev. E* **51**, 4027 (1995).
- ²¹D. Cule and T. Hwa, *Phys. Rev. Lett.* **79**, 2375 (1997).
- ²²N. Theodorakopoulos, T. Dauxois, and M. Peyrard, *Phys. Rev. Lett.* **85**, 6 (2000).
- ²³N. K. Voulgarakis, G. Kalosakas, K. Ø. Rasmussen, and A. R. Bishop, *Nano Lett.* **4**, 629 (2004).
- ²⁴T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, *Phys. Rev. Lett.* **95**, 218104 (2005); C. H. Choi, A. Usheva, G. Kalosakas, K. Ø. Rasmussen, and A. R. Bishop, *ibid.* **96**, 239801 (2006); T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, *ibid.* **96**, 239802 (2006).
- ²⁵Z. Rapti, A. Smerzi, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi, and A. Usheva, *Europhys. Lett.* **74**, 540 (2006).
- ²⁶B. S. Alexandrov, L. T. Wille, K. Ø. Rasmussen, A. R. Bishop, and K. B.

- Blagoev, *Phys. Rev. E* **74**, 050901 (2006).
- ²⁷ G. Kalosakas, K. Ø. Rasmussen, and A. R. Bishop, *Chem. Phys. Lett.* **432**, 291 (2006).
- ²⁸ N. K. Voulgarakis, A. Redondo, A. R. Bishop, and K. Ø. Rasmussen, *Phys. Rev. Lett.* **96**, 248101 (2006).
- ²⁹ N. Theodorakopoulos, *Phys. Rev. E* **77**, 031919 (2008).
- ³⁰ F. de los Santos, O. Al Hammal, and M. A. Muñoz, *Phys. Rev. E* **77**, 032901 (2008).
- ³¹ T. Das and S. Chakraborty, *Europhys. Lett.* **83**, 48003 (2008).
- ³² B. Alexandrov, N. K. Voulgarakis, K. Ø. Rasmussen, A. Usheva, and A. R. Bishop, *J. Phys.: Condens. Matter* **21**, 034107 (2009).
- ³³ J. SantaLucia, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460 (1998).
- ³⁴ R. Gonzalez, Y. Zeng, V. Ivanov, and G. Zocchi, *J. Phys.: Condens. Matter* **21**, 034102 (2009).
- ³⁵ R. Everaers, S. Kumar, and C. Simm, *Phys. Rev. E* **75**, 041918 (2007).
- ³⁶ N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- ³⁷ Y.-L. Zhang, W.-M. Zheng, J.-X. Liu, and Y. Z. Chen, *Phys. Rev. E* **56**, 7100 (1997).
- ³⁸ J. Marmur and P. Doty, *J. Mol. Biol.* **5**, 109 (1962).
- ³⁹ S. Ares and A. Sánchez, *Eur. Phys. J. B* **56**, 253 (2007).
- ⁴⁰ D. J. Scalapino, M. Sears, and R. A. Ferrell, *Phys. Rev. B* **6**, 3409 (1972).
- ⁴¹ S. Aubry, *J. Chem. Phys.* **62**, 3217 (1975); J. A. Krumhansl and J. R. Schrieffer, *Phys. Rev. B* **11**, 3535 (1975).
- ⁴² S. Ares and A. Sánchez, *Phys. Rev. E* **70**, 061607 (2004).
- ⁴³ L. Lewin, *Polylogarithms and Associated Functions* (North-Holland, New York, 1981).
- ⁴⁴ A. Bar, Y. Kafri, and D. Mukamel, *Phys. Rev. Lett.* **98**, 038103 (2007).
- ⁴⁵ G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, *Phys. Rev. Lett.* **90**, 138101 (2003).
- ⁴⁶ E. Carlon, E. Orlandini, and A. L. Stella, *Phys. Rev. Lett.* **88**, 198101 (2002).
- ⁴⁷ M. Rief, H. Clausen-Schaumann, and H. E. Gaub, *Nat. Struct. Biol.* **6**, 346 (1999).
- ⁴⁸ H. Mao, J. R. Arias-Gonzalez, S. B. Smith, I. Tinoco, and C. Bustamante, *Biophys. J.* **89**, 1308 (2005).
- ⁴⁹ P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J. L. Viovy, D. Chatenay, and F. Caron, *Science* **271**, 792 (1996).
- ⁵⁰ S. Whitlam, S. Pronk, and P. L. Geissler, *Biophys. J.* **94**, 2452 (2008).
- ⁵¹ L. Shokri, M. J. McCauley, I. Rouzina, and M. C. Williams, *Biophys. J.* **95**, 1248 (2008).