Simplified Langevin approach to the Peyrard-Bishop-Dauxois model of DNA

F. de los Santos, Omar Al Hammal, and Miguel A. Muñoz
Instituto de Física Teórica y Computacional Carlos I and Departamento de Electromagnetismo y Física de la Materia, Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain
(Received 4 October 2007; published 31 March 2008)

A simple Langevin approach is used to study stationary properties of the Peyrard-Bishop-Dauxois model for DNA, allowing known properties to be recovered in an easy way. Results are shown for the denaturation transition in homogeneous samples, for which some implications, so far overlooked, of an analogy with equilibrium wetting transitions are highlighted. This analogy implies that the order parameter, asymptotically, exhibits a second-order transition even if it may be very abrupt for nonzero values of the stiffness parameter. Not surprisingly, we also find that, for heterogeneous DNA, within this model the largest bubbles in the premelting stage appear in adenine-thymine-rich regions, while we suggest the possibility of some sort of not strictly local effects owing to the merging of bubbles.

DOI: 10.1103/PhysRevE.77.032901 PACS number(s): 87.14.G−, 02.50.−r, 64.60.Ht

The DNA thermal melting transition (also called denaturation, coiling, or unzipping) occurs when, above a certain critical temperature, the double-stranded DNA molecule unravels into two separate coils, while for smaller temperatures (premelting stage) only localized openings or bubbles exist [1]. This phase transition is of importance for DNA duplication and transcription, and many studies have scrutinized its nature (whether first or second order), trying to pin down the relevant traits of the rich phenomenology experimentally observed (a nonexhaustive list of references is [2–7]). Moreover, it has been suggested that the dynamics of a DNA molecule in its premelting stage may play a role in its own transcription initiation. Indeed, bubbles are determined by sequence specificity and they have been reported to occur with high probability in the neighborhood of the functionally relevant *transcription start site* (TSS) and near other regulatory sites, facilitating further microbiological activity [7–9].

This relation between thermal dynamics and biological functionality has been claimed to be borne out by experimental data from real promoter DNA sequences and is supported by results from a theoretical model (see below) [8,9]. Even if this might differ from biological, protein-mediated processes, studies of thermal properties of the DNA by itself are a first step forward in understanding more complex situations [1] (see [10] for a different view).

Let us mention some observations in this context, which have been the object of recent analyses. Even though one would expect that adenine-thymine- (AT-)rich regions should be more prone to sustain bubbles than guanine-cytosine- (GC-)rich ones (as AT pairs bind the two strands more weakly than GC ones [1]), counterintuitive situations in which this is not the case have been reported [7,11]. In the same vein, the dependence of bubble formation on the specific base-pair sequence was reported to be highly nonlocal: Upon mutation of two AT base pairs into two (stronger) GC base pairs near the TSS, rendering a specific promoter sequence completely inactive for transcription, the opening profiles of the original sequence and its mutant variant differed not only in the expected suppression of the large thermal opening near the TSS, but also in a sizable increase in the probability of formation of a bubble at a distant base pair [9]. However, subsequent studies using more efficient methods for the calculation of bubble statistics in the Peyrard-Bishop-Dauxois (PBD) model [12,13] did not confirm the above nonlocal scenario, and pointed to more localized effects. See [14,15] for recent developments on this interesting problem.

Many of these and other relevant issues have been investigated by employing the Peyrard-Bishop-Dauxois model [4] (see below). The model phenomenology has been profusely analyzed by means of various analytical and numerical techniques: transfer integral calculations, Monte Carlo simulations, molecular dynamics, and Langevin dynamics, and the results have been found to properly describe experiments on the melting transition [16], premelting bubbles [15], etc. Let us caution that, under certain circumstances, torsional effects (absent in the PBD model) should be included to properly account for some of the described phenomenology [7,10,17].

In this Brief Report we reconsider the DNA thermal denaturation problem, analyzing the PBD model [4] by means of a different, simplified Langevin approach. This strategy allows us to (i) reproduce numerically in a relatively easy way the stationary bubble probability distribution and other statistical properties for both homogeneous and heterogeneous sequences; (ii) establish an analogy with well-known *equilibrium wetting* problems, deeper than previously thought, permitting us to infer results about the order of the denaturation transition.

In the PBD model the stretching of hydrogen bonds between corresponding base pairs is represented by a set of continuous variables $\{h_n\}$ (at positions $n=1,\ldots,N$ where N is the chain length). The model is defined by the following Hamiltonian [4]:

$$H=\sum_{n=1}^{N}\left(\frac{1}{2}m\dot{h}_n^2+V(h_n)+W(h_n,h_{n-1})\right). \tag{1}$$

The first term is the kinetic energy for bases of mass m. The second one stands for the interaction between opposite bases as described by the Morse potential

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

where D_n is the dissociation energy of the nth base pair and a_n denotes the spatial range of the potential. Standard, empirically found pair-base-dependent parameter values are customarily employed: $D_n(\mathrm{AT})=0.05$ eV, $D_n(\mathrm{GC})=0.075$ eV, $a_n(\mathrm{AT})=4.2$ Å^{-1}, and $a_n(\mathrm{GC})=6.9$ Å^{-1} [16]. Finally, the third *stacking* term arises from the interaction between adjacent bases along the DNA molecule [4]. It reads

$$W(h_n,h_{n-1}) = \frac{k}{2}(1+\rho e^{-\alpha(h_n+h_{n-1})})(h_n - h_{n-1})^2, \tag{3}$$

where the values of k, ρ, and α are determined from fittings of experimental DNA denaturation curves [16]: $k=0.025$ eV Å^2, $\rho=2$, $\alpha=0.35$ Å^{-1}. The nonvanishing *stiffness parameter* ρ captures the fact that the double-stranded backbone is more rigid than the unwound strands (controlled by a standard elastic interaction). Note that this model includes only transverse degrees of freedom for nucleotides.

The average stretching at each site $\langle h_n \rangle$ and its space-averaged counterpart $\langle h \rangle$, as well as $\langle e^{-h} \rangle$, which can be interpreted as the density of closed base pairs, are the standard order parameters.

Different scenarios have been reported for the denaturation transition depending on the stiffness parameter ρ and the randomness of the DNA sample. In the simplest case $\rho=0$ [4], the stacking term is harmonic and a smooth (second-order) denaturation transition is known to occur for both homogeneous and heterogeneous DNA [6,18]. On the contrary, nonvanishing ρ and heterogeneous sequences lead to very abrupt thermal denaturation curves that exhibit a multistep behavior in line with experimental observations [6].

The case of nonzero ρ and homogeneous DNA is still unsettled as the transition has been reported to be (i) first-order-like yet with a diverging correlation length in [18,19], and (ii) second order although very sharp in appearance [6]. We shall return to this issue below. Let us also remark that, as pointed out in [18], a continuous transition for the order parameter $\langle h \rangle$ with associated critical exponents and a diverging length scale could be compatible (if $\rho \neq 0$) with the number of bound pairs $\langle n \rangle$ exhibiting a discontinuity at the transition.

In evaluating the partition function associated with the Hamiltonian Eq. (1), the kinetic terms factorize and, as a result, can be dropped out if the focus is only on equilibrium configurational properties. In such a case, the equilibrium state can be recovered from the configurational part H' of H (including only V and W terms) and, therefore, can be reproduced from the stationary solution of the associated Langevin equation,

$$\frac{\partial h_n(r,t)}{\partial t} = -\frac{\partial H'(h_n)}{\partial h_n} + \sigma\eta(r,t), \tag{4}$$

where η is a Gaussian white noise and σ its amplitude. In the following, Eq. (4) is taken as the starting point for study, and an Euler algorithm is used to solve it. This differs from previous Langevin studies in that inertial terms do not appear, enabling slightly faster computational studies. A similar approach was used in [20]. Let us stress that the dynamics imposed by Eq. (4) is a fictitious one, not related to real DNA dynamics (which is not purely relaxational), but leads to the same stationary probability distribution as the original one.

Homogeneous DNA. We begin by studying the case of homogeneous samples with only GC base pairs. The temperature T is the control parameter, and the value of σ is obtained from the fluctuation-dissipation relation. We have run simulations in systems of size 2^{17}, initializing all the base pairs to $h(t=0)=2$ and letting them evolve until a stationary state is reached. $\langle h \rangle$ was monitored as a function of time for zero and nonzero values of ρ. At low temperatures $\langle h \rangle$ saturates to a finite value whereas at high enough temperatures it diverges as $t^{1/4}$ (see below), signaling a phase transition. While for $\rho=0$ a smooth (continuous) transition is observed, for $\rho=2$ it is rather abrupt (results not shown), being apparently first order. The same picture, in line with previous numerical results [4], can also be drawn by monitoring $\langle e^{-h} \rangle$, but our results are not fully conclusive.

As originally argued in [6], as $h_n \approx h_{n-1}$, the exponential factor in Eq. (3) can be approximated by $e^{-\alpha h_n}$ without provoking any significant effect. If $\rho=0$, H' is readily recognized (apart from constant terms) as a discretized version of the continuous Hamiltonian

$$H_{ew} = \int dx \left(\frac{k}{2}(\nabla h)^2 + w_1 e^{-ah} + w_2 e^{-2ah} \right), \tag{5}$$

where w_1, w_2, and k are generic parameters. H_{ew} is the standard interfacial Hamiltonian for equilibrium critical wetting transitions in the presence of short-ranged forces, i.e., the unbinding of the interface separating two coexisting phases from a wall, which occurs upon increasing the temperature [21]. At this point, we recall that in wetting phenomena continuum models are valid approximations to lattice models as long as T is above the roughening temperature T_R, which is $T_R=0$ in $d=1$ ($d=2$ bulk).

Although the connection between wetting and DNA denaturation has already been recognized (see, for instance, [6,20]), some of its consequences have not been fully appreciated. For instance, the set of recently reported [18] critical exponents characterizing the DNA denaturation transition in the homogeneous case, $\langle h \rangle \sim |\delta|^{-\beta}$ and $\xi \sim |\delta|^{-\nu}$ [where $\delta=(T-T_c)/T_c$], with $\beta=-1$, ξ the correlation length, and $\nu=2$, are nothing but the two-dimensional critical wetting exponents dating back to the early 1980s [21]. Furthermore, the density of closed base pairs scales as $\langle h^{-1} \rangle \sim |\delta|$ (see [6]), as corresponds to the surface order parameter in a wetting context [21]. Additionally, since in equilibrium wetting the dynamic critical exponent z, defined by $\xi \sim t^{1/z}$, is $z=2$, the thickness of the wetting layer grows as $t^{1/4}$ [22], in agreement with the value reported above for the PBD model. To the best of our knowledge, these correspondences have not been established before.

More interestingly, the implications of the wetting analogy can be extended to the nonzero-ρ case. In the wetting context, a long-standing problem regarding the order of the transition in three-dimensional systems has been recently solved [23]. The original renormalization-group calculations

led to the prediction of nonuniversal results, in blatant disagreement with computational studies [24] and experiments [25], both of which yield a mean-field-like second-order phase transition. An early attempt to reconcile theory and experiments questioned the validity of the effective Hamiltonian Eq. (5) to describe equilibrium wetting and concluded that k in Eq. (5) should be replaced by a position-dependent *stiffness coefficient* $k(h)=k+w_1'e^{-\alpha h}+w_2'ahe^{-2\alpha h}+\cdots$ [26]. Curiously enough, with only the leading correction included in $k(h)$, this Hamiltonian is the continuous counterpart of the PBD one.

In critical wetting the parameter w_1' vanishes at the transition point and, according to a linear renormalization-group study, only the term proportional to w_2' is capable of destabilizing the critical wetting transition, driving the transition weakly first order in $d=3$ [26]. A subsequent investigation allowed the analysis to be extended, with the conclusion that a first-order transition can appear only for dimensions $d \gtrsim 2.41$ [27]. Remarkably, it has been shown [23] that by including the whole series expansion the experimental and computational results can be finally reproduced.

These results can be adapted for homogeneous DNA melting. Indeed, by switching on a nonvanishing w_1' and truncating the series to first order, we do not expect the above conclusions to change qualitatively, since it is naively expected that w_1' plays a similar role to w_2' (the detailed proof of this is not straightforward and will be published elsewhere). Therefore, using the wetting analogy, the one-dimensional melting transition for homogeneous DNA sequences should be *asymptotically* continuous for $\langle h \rangle$, in agreement with some previous transfer integral analyses [6], but in partial disagreement with other calculations [18,19]. Reconciling all these results remains an open challenging task.

Our conclusion about the order of the transition might change if we consider versions of the PBD model embedded in a three-dimensional space [17] where bubble entropic effects are expected to play a crucial role [3]. Note also that for such three-dimensional models the analogy with wetting problems breaks down.

Heterogeneous DNA. Following the recent literature, we have simulated our model for two particular sequences of 69 base pairs: the adeno-associated viral P5 (AAVP5) promoter and a mutation of it inactive for transcription [8]. In the mutant sequence two AT bases located near the TSS at positions 48 and 49 are replaced by (more tightly bound) GC base pairs. In our analyses a bubble is defined as a group of adjacent sites that satisfies the condition $h>1.5$. To avoid finite-size effects, we use periodic boundary conditions on lattices of sizes $L=690$ and 6900 consisting of 10 and 100 replicas, respectively, of the same AAVP5 sequence, After sufficient ensemble averaging, indistinguishable long-time results are obtained for both sizes. The bubble distributions for the AAVP5 sequence and its mutant are shown in Fig. 1. It can be seen that the large bubbles forming around the TSS (top panel) are suppressed in the mutant sequence (bottom panel) in agreement with experimental observations [8]. The effect of the mutation is quite local, in line with that obtained in [12] and in contrast to the first claims [8]. Observe, also, that bubbles in the DNA sequence form more frequently where AT bases are more abundant, as naively expected

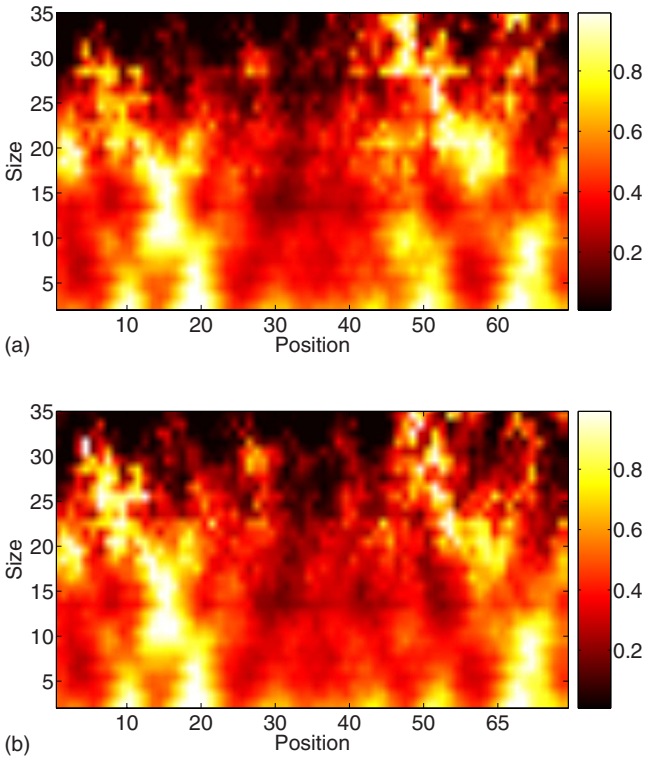

FIG. 1. (Color online). Probability of bubble opening as a function of position and bubble size for the AAVP5 promoter (top panel) and the mutant P5 promoter (bottom panel) at $T=310$ K. Probabilities in each row are normalized to their maximum value as in [12]. The results are very similar to those in [12].

[14,15]. Situations in which this is not the case (like those reported in [11]) are likely to be physically ascribable to torsional effects [7,10]. Our conclusion is that the local bubble-opening probability within the PBD model is controlled by the relative density of AT base pairs, in accordance with [14,15].

To explore the possibility of having some sort of nonlocal effect in bubble formation within the present model, consider an artificial chain with a GC-rich region separating two AT-rich zones (see Fig. 2). Small bubbles formed in the two AT-rich regions might eventually merge together, bridging across the GC region as illustrated in Fig. 2. This can induce the largest possible bubble to be centered around a GC-rich zone, and non-strictly-local effects could be generated upon

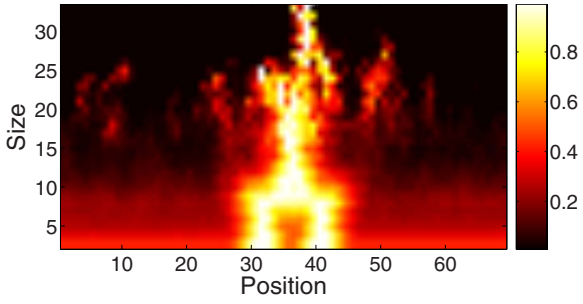

FIG. 2. (Color online). Bubble merging over a GC region from the openings above two small AT regions.

introducing mutations. Further research is needed to quantify this mechanism and to assess if it is capable of inducing nonlocal effects by repetition of the above scenario, which has already been discussed in the literature in various forms [28].

In summary, the simple Langevin equation (4) gives relatively quick access to the stationary properties of the PBD model for DNA denaturation. It reproduces many known results for the homogeneous case, e.g., for $\rho=0$ a continuous transition is obtained. Moreover, we have pointed out that the (recently obtained) critical exponents are well known for the wetting problem. The analogy with equilibrium critical wetting can be extended using very recent developments to the $\rho \neq 0$ case, where also a continuous transition is predicted (even if it might be a very abrupt one [6,18]). We have also employed the Langevin approach to study the bubble statistics in heterogeneous real sequences, confirming the tendency for creation of thermal openings around AT-rich regions. According to our observations mutations modify the statistics of bubbles only in a local way. However, nonstrictly-local effects due to the merging of bubbles could induce large openings in locally GC-rich regions.

It is our hope that this simple Langevin approach will be useful to elucidate other aspects of this fascinating field.

We acknowledge financial support from the Spanish MEyC-FEDER, Project No. FIS2005-00791 and from Junta de Andalucía as group FQM-165.

[1] S. R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985), and references therein.
[2] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1456 (1966).
[3] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000); Eur. Phys. J. B **27**, 135 (2002); E. Carlon, E. Orlandini, and A. L. Stella, Phys. Rev. Lett. **88**, 198101 (2002); M. S. Causo, B. Coluzzi, and P. Grassberger, Phys. Rev. E **62**, 3958 (2000); S. Cocco and R. Monasson, Phys. Rev. Lett. **83**, 5178 (1999).
[4] T. Dauxois, M. Peyrard, and A. R. Bishop Phys. Rev. E, **47**, R44 (1993); see also M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989); M. Peyrard, Nonlinearity **17**, R1 (2004).
[5] L.-H. Tang and H. Chaté, Phys. Rev. Lett. **86**, 830 (2001); T. Hwa *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 4411 (2003); V. Ivanov, Y. Zeng, and G. Zocchi, Phys. Rev. E **70**, 051907 (2004); M. Y. Azbel, Phys. Rev. A **20**, 1671 (1979); Phys. Rev. E **68**, 050901(R) (2003).
[6] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[7] C. J. Benham, Proc. Natl. Acad. Sci. U.S.A. **90**, 2999 (1993); R. M. Fye and C. J. Benham, Phys. Rev. E **59**, 3408 (1999).
[8] C. H. Choi, A. Usheva, G. Kalosakas, K. O. Rasmussen, and A. R. Bishop, Nucleic Acids Res. **32**, 1584 (2004); C. H. Choi *et al.*, Phys. Rev. Lett. **96**, 239801 (2006).
[9] G. Kalosakas *et al.*, Europhys. Lett. **68**, 127 (2004).
[10] C. J. Benham and R. R. P. Singh, Phys. Rev. Lett. **97**, 059801 (2006).
[11] U. Dornberger, M. Leijon, and H. Fritzsche, J. Biol. Chem. **274**, 6957 (1999).
[12] T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, Phys. Rev. Lett. **95**, 218104 (2005); T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, *ibid.* **96**, 239802 (2006); T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, *ibid.* **97**, 059802 (2006).
[13] Z. Rapti *et al.*, Europhys. Lett. **74**, 540 (2006).
[14] T. S. van Erp, S. Cuesta-López, and M. Peyrard, Eur. Phys. J. E **20**, 421 (2006).
[15] S. Ares, N. K. Voulgarakis, K. O. Rasmussen, and A. R. Bishop, Phys. Rev. Lett. **94**, 035504 (2005); Z. Rapti *et al.*, Phys. Rev. E **73**, 051902 (2006); B.S. Alexandrov, L. T. Wille, K. O. Rasmussen, A. R. Bishop, and K. B. Blagoev, *ibid.* **74**, 050901(R) (2006); S. Ares and G. Kalosakas, Nano Lett. **7**, 307 (2007).
[16] A. Campa and A. Giansanti, Phys. Rev. E **58**, 3585 (1998).
[17] M. Barbi, S. Cocco, and M. Peyrard, Phys. Lett. A **253**, 358 (1999).
[18] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, Phys. Rev. Lett. **85**, 6 (2000).
[19] S. Buyukdagli and M. Joyeux, Phys. Rev. E **73**, 051910 (2006); M. Joyeux and S. Buyukdagli, *ibid.* **72**, 051902 (2005)..
[20] S. Ares and A. Sánchez, Eur. Phys. J. B **56**, 253 (2007).
[21] S. Dietrich, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. Lebowitz (Academic, New York, 1988), Vol. 12; M. Schick, in *Liquids at Interfaces*, edited by J. Charvolin, J. F. Joanny, and J. Zinn-Justin, Proceedings of the Les Houches Summer School of Theoretical Physics XLVIII (North-Holland, Amsterdam, 1990).
[22] R. Lipowsky, J. Phys. A **18**, L585 (1985).
[23] A. O. Parry, J. M. Romero-Enrique, and A. Lazarides, Phys. Rev. Lett. **93**, 086104 (2004).
[24] K. Binder, D. P. Landau, and D. M. Kroll, Phys. Rev. Lett. **56**, 2272 (1986).
[25] D. Ross, D. Bonn, and J. Meunier, Nature (London) **400**, 737 (1999).
[26] A. J. Jin and M. E. Fisher, Phys. Rev. B **48**, 2642 (1993).
[27] C. J. Boulter, Phys. Rev. Lett. **79**, 1897 (1997).
[28] T. Novotny *et al.*, Europhys. Lett. **77**, 48001 (2007).