PHYSICAL REVIEW E **82**, 031916 (2010)

Thermal and mechanical properties of a DNA model with solvation barrier

R. Tapia-Rojo,[1,2] J. J. Mazo,[1,3] and F. Falo[1,2]
[1]*Departamento de Física de la Materia Condensada, Universidad de Zaragoza, 50009 Zaragoza, Spain*
[2]*Instituto de Biocomputación y Física de Sistemas Complejos, Universidad de Zaragoza, 50009 Zaragoza, Spain*
[3]*Instituto de Ciencia de Materiales de Aragón, C.S.I.C.–Universidad de Zaragoza, 50009 Zaragoza, Spain*
(Received 19 May 2010; published 27 September 2010)

We study the thermal and mechanical behaviors of DNA denaturation in the frame of the mesoscopic Peyrard-Bishop-Dauxois model with the inclusion of solvent interaction. By analyzing the melting transition of a homogeneous A-T sequence, we are able to set suitable values of the parameters of the model and study the formation and stability of bubbles in the system. Then, we focus on the case of the P5 promoter sequence and use the principal component analysis of the trajectories to extract the main information on the dynamical behavior of the system. We find that this analysis method gives an excellent agreement with previous biological results.

DOI: 10.1103/PhysRevE.82.031916 PACS number(s): 87.15.H−, 87.15.A−, 05.10.−a

I. INTRODUCTION

In the last years there exist an increasing interest in the description of complex biological systems using simple physical models [1]. Physical models for biological problems should contain the key ingredients to explain their basic phenomenology. In this sense, DNA melting can be modeled using very simple statistical-mechanics models [2]. One of the most successful is the one-dimensional Peyrard-Bishop-Dauxois (PBD) model [3,4], which with very few assumptions is able to reproduce the melting curves for different DNA sequences. The PBD model undergoes an entropic phase transition between the native closed state and a denatured one in which the DNA strands are separated. The nature of this transition strongly depends on model parameter, so different versions of the model have been proposed in the last years.

A virtue of the PBD model is that it allows us to study not only the equilibrium properties of the molecule but also dynamical ones, for instance, the formation and stability of the so-called DNA bubbles (short open segments of the DNA chain) [5,6]. Due to its versatility, this simple model has been applied in other contexts beyond the study of the equilibrium melting curves. Thus, it has been used to investigate the DNA mechanical denaturation [7,8], and it has been especially successful in the modeling of the open regions in short DNA hairpins [6,9,10]. This fact has motivated to extend the model to study nonhomogeneous sequences and to try to correlate the *dynamical* behavior of the model to *functional* aspects of the DNA chain. To be specific, it has been suggested that the localized dynamical excitations (bubbles) of the model are directly related to the protein-DNA binding sites [11,12]. This point has been controversial due to the simplicity of the PBD model assumptions [13,14].

One of the most interesting improvements to the model is the inclusion of a barrier in the intrabase interaction term [15]. Such a barrier accounts for solvent interactions in the system. The addition of barriers modifies deeply the nature of the denaturation transition and the dynamics of the molecule. Depending on the parameter values, the transition become sharper, even in the harmonic stacking energy version. The transition width is not a trivial behavior and it deserves a more careful study. On the dynamical side, the bubble lifetime increases dramatically with the inclusion of this term and approaches the experimental observations.

The purpose of this paper is twofold. First, we characterize the melting transition of homogeneous sequences and the formation of bubble phenomenon for a wide range of parameters (Sec. IV). This allows us to set suitable parameter values for next research. Then, we focus on the problem of the relation between dynamics and function. In Sec. V, we study a heterogeneous sequence which contains known information on the transcription process of DNA to messenger RNA (mRNA) (a P5 virus promoter). We use principal component analysis (PCA) of trajectories to identify the regions in the sequence which mostly contribute to the dynamical fluctuations of the molecule. These regions are the softest ones, from a mechanical point of view, and they correlate fairly well with the relevant biological sites. Given its simplicity and efficiency, we propose to extend this analysis to other sequences in order to get insight over functionality of the different genome regions.

II. MODEL

We study a modification of the Peyrard-Bishop-Dauxois model. In the PBD model the complexity of DNA is reduced to the study of the dynamics of the N base pairs of the molecule. For each base pair we define the variable y_n associated with the distance between the bases. The total energy of the system is then approached by

$$H = \sum_{n=1}^{N} \left[\frac{p_n^2}{2m} + V(y_n) + W(y_n, y_{n-1}) \right]. \tag{1}$$

Here, $p_n = m dy_n/dt$, n is the index of a base pair, and m is its reduced mass.

In this equation, we identify two energy potential terms: $V(y_n)$, an on-site potential one, and $W(y_n, y_{n-1})$, which accounts for interpair interactions. The standard PBD model corresponds to a particular choice for these energy terms.

The potential $V(y_n)$ describes the interaction between the two bases of a pair. The PBD model uses a Morse potential

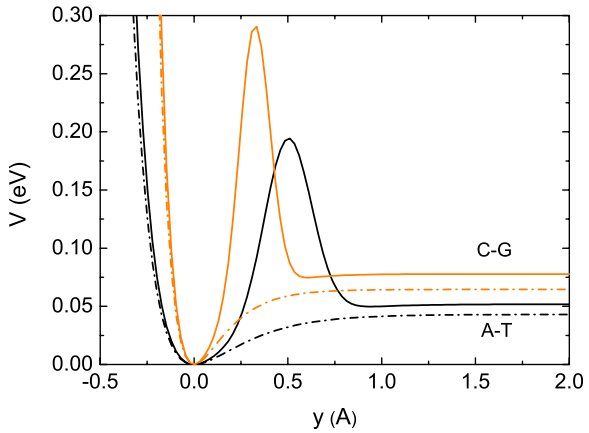

FIG. 1. (Color online) Morse potential with (solid lines) and without (dashed lines) solvation barrier for A-T (black) and C-G (orange/light gray) base pairs. For A-T pair without barrier $\alpha=4$ Å^{-1} and $D=0.043$ eV. For A-T with barrier $\alpha=4$ Å^{-1}, $D=0.051\,85$ eV, $G=3D$, $y_0=2/\alpha$, and $b=1/2\alpha^2$ (the value of D has been adjusted in each case to give the desired melting temperature for the A-T uniform chain). For C-G pairs $D_{CG}=1.5D_{AT}$ and $\alpha_{CG}=1.5\alpha_{AT}$.

to account for such an interaction (see Fig. 1). This is a standard approximation for chemical bonds,

$$V(y) = D(e^{-\alpha y} - 1)^2, \tag{2}$$

where D corresponds to the dissociation energy of the pair $[V(0)=0$ and $V(\infty)=D]$ and α sets the amplitude of the potential well $[V''(0)=2D\alpha^2]$. See also that the variable y measures deviations of the base distance with respect to equilibrium; thus, $V(y)$ has a minimum for $y=0$.

An important contribution to the stability of DNA molecule comes from the stacking of the bases. The potential $W(y_n,y_{n-1})$ describes the interaction between base pairs along the DNA strand. The simplest approximation is to consider just harmonic interaction between adjacent pairs in the molecule. However, this choice does not reproduce the observation of a sharp thermal denaturation curve. To solve this, the PBD model includes a position-dependent nonlinear coupling constant

$$W(y_n,y_{n-1}) = \tfrac{1}{2}K(1+\rho e^{-\delta(y_n+y_{n-1})})(y_n-y_{n-1})^2. \tag{3}$$

The effect of this term, whose intensity is governed by ρ, is to change the effective coupling constant from $K(1+\rho)$ to K when one of the base pairs is displaced away from its equilibrium position. The parameter δ sets the scale length for this behavior.

DNA melting or denaturation refers to the separation of the two strands of the DNA molecule to generate two single strands. The great merit of the PBD model defined in Eqs. (2) and (3) is to reproduce the different melting curves observed in DNA molecules. Thus, it has been used as a solid starting point for many other statistical-mechanics studies of DNA.

The melting process involves the breaking of hydrogen bonds between the bases. In order to study different DNA sequences the PBD model also can include sequence-dependent Morse potential parameters: D_n and α_n. It can be intuitively deduced from the fact that A-T pairs are linked by two hydrogen bonds; meanwhile, C-G ones are linked by three, thus forming a more stable link. Following [16] in our simulations we will use $D_{CG}=1.5D_{AT}$ and $\alpha_{CG}=1.5\alpha_{AT}$. Although recent works have considered more complex sequence dependence [6,17], we will not introduce more complexity at this level in the model.

We will work with a modified version of the PBD model to include solvent interaction. Solvent interaction stabilizes open pair states by means of the hydrogen bonds that the base pairs may form with the solvent when opened [15]. Such bonds have to be broken before the pair closes again. This effect can be included in a simple way with an effective barrier in the Morse potential. The addition of the barrier has to avoid, as much as possible, any other effects in the Morse well. Hence, we have chosen the following definition for the on-site potential $V(y)$:

$$V(y) = D(e^{-\alpha y} - 1)^2 + Ge^{-(y-y_0)^2/b}. \tag{4}$$

This potential results from the addition of a Gaussian barrier, whose height is controlled by G, the position is given by y_0, and its width is given by b. A reasonable selection for such parameters is $G=3D$, $y_0=2/\alpha$, and $b=1/2\alpha^2$. Figure 1 plots the intrabase potential V for an A-T and a C-G base pair.

III. METHODS

A. Langevin dynamics simulations

In order to study the behavior of the system we have performed molecular-dynamics numerical simulations of the Langevin equation

$$m\frac{d^2y_n}{dt^2} + m\gamma\frac{dy_n}{dt} + \frac{\partial V(y_n)}{\partial y_n} + \frac{\partial[W(y_n,y_{n-1}) + W(y_{n+1},y_n)]}{\partial y_n} = \xi_n(t), \tag{5}$$

where m is the mass of the pair, γ is the effective damping of the system and $\xi(t)$ accounts for thermal noise, $\langle\xi_n(t)\rangle=0$, and $\langle\xi_n(t)\xi_k(t')\rangle=2m\gamma k_BT\delta_{nk}\delta(t-t')$, with T as the bath temperature.

The equations were numerically integrated using the stochastic Runge-Kutta algorithm [18,19]. Simulation of large A-T chain (Sec. IV) used periodic boundary conditions in order to avoid any terminal effect. For the P5 promoter (Sec. V) we used fixed boundary conditions which will be described below.

To best characterize the phase transition we have computed the mean energy $\langle u\rangle$ and mean displacement $\langle y\rangle$ defined as

$$\langle u\rangle = \frac{1}{Nt_s}\sum_{n,t}^{N,t_S}[W(y_n,y_{n-1}) + V(y_n)], \tag{6}$$

$$\langle y\rangle = \frac{1}{N}\sum_{n=1}^{N}\langle y_n\rangle = \frac{1}{Nt_s}\sum_{n,t}^{N,t_S} y_n(t). \tag{7}$$

Here, N is the number of base pairs in the sequence to be studied and t_s is the total simulation time.

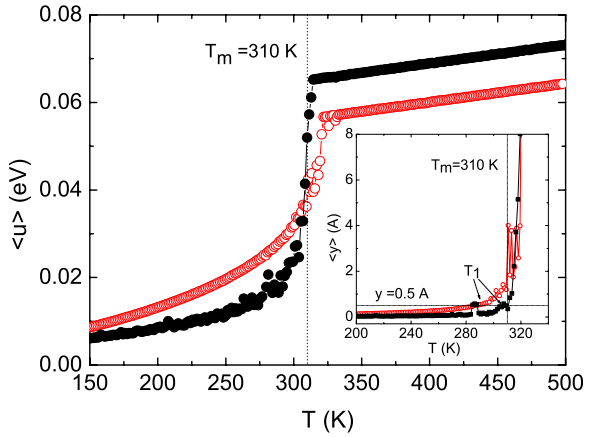

FIG. 2. (Color online) The melting transition. Mean energy and displacement (inset) as functions of temperature of a homogeneous chain for the standard (open circles) PBD model and the modified one (full circles). Parameters are given in the text.

We are also interested in computing the probability of being opened for each base. We define this as follows:

$$P_n(y_{\text{th}}) = \frac{1}{t_s}\sum_{t}^{t_s} \Theta(y_n(t) - y_{\text{th}}), \tag{8}$$

where y_{th} is a threshold for opening that we choose just behind the barrier of the Morse potential and $\Theta(x)$ is the Heaviside step function [$\Theta(x)=0$ for $x<0$ and $\Theta(x)=1$ for $x\geq 0$]. Thus, P_n indicates the fraction of time when the nth base is opened during the time t_s.

B. PCA

PCA is a statistical method to extract information from a large set of data in a multidimensional phase space, allowing us to reduce the dimensionality of the variables to those that include most of the fluctuations of the original system [20]. This is achieved by a change of coordinates in the phase space, and the new coordinates are the so-called principal components (PCs). This method has become a standard tool

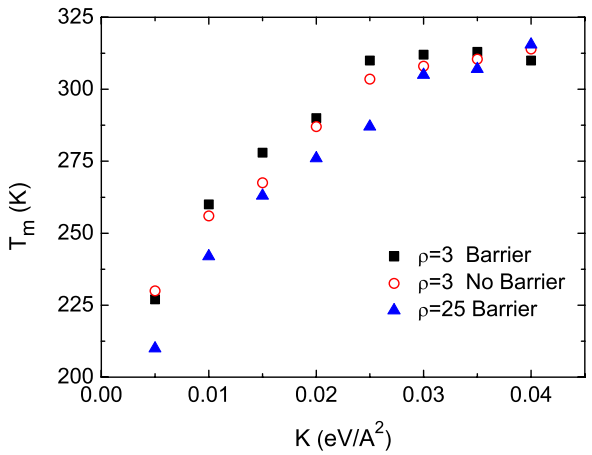

FIG. 3. (Color online) Transition temperature versus the stacking constant K for various ρ values.

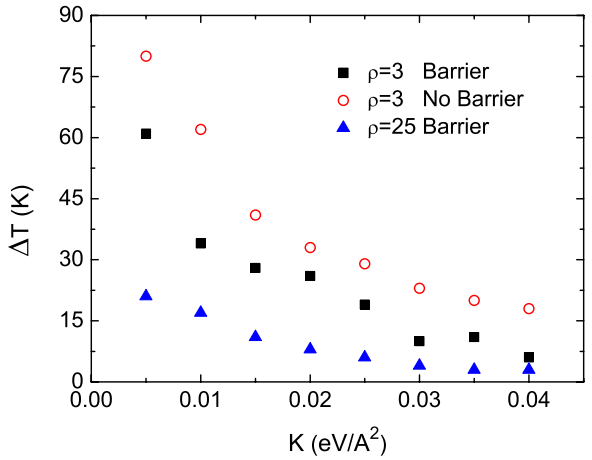

FIG. 4. (Color online) Transition width versus the stacking constant K for various ρ values.

in the analysis of trajectories in "all-atom" simulations of macromolecules [21].

Operationally, we have to build the $N\times N$ correlation matrix

$$C(i,j) = \langle y_i y_j\rangle - \langle y_i\rangle\langle y_j\rangle. \tag{9}$$

Diagonalizing this matrix we obtain an ordered set of eigenvalues $(\lambda_1>\lambda_2>\lambda_3\cdots)$ with their corresponding eigenvectors $(v_1,v_2,v_3,\ldots)$. Eigenvalue λ_i gives the amount of fluctuations, which corresponds to eigenvector v_i; thus, the new coordinates are ordered in such a way that the few first ones retain most of the fluctuations of the system. It is useful to define a frequency ω_i associated with each eigenvalue λ_i and given by

$$\omega_i = \sqrt{\frac{k_B T}{\lambda_i}}. \tag{10}$$

From this point of view, the larger fluctuations will correspond to the lowest frequencies. It can be proved that in the low-temperature limit these frequencies converge to the normal-mode frequencies of the system. Normal-mode analysis has also used in the detection of relevant motions in coarse-grained model of proteins [22].

IV. UNIFORM CHAIN

A. Fitting the phase transition: Parameters of the model

In this section we will study the melting transition of a homogeneous chain of A-T pair bases. Although there are some studies on the influence of the different parameters in the transition, up to our knowledge, there is no systematic scan over the parameter space. In particular, when the barrier is included on the Morse potential a clear narrowing of the transition is observed (see Fig. 2). The same effect is expected when the ρ parameter is increased. Thus, one of the first problems when dealing with this modified PBD model is to find suitable model parameters. This is mainly done by adjusting the theoretical melting temperatures to the experimental ones. The melting temperature T_m is usually defined

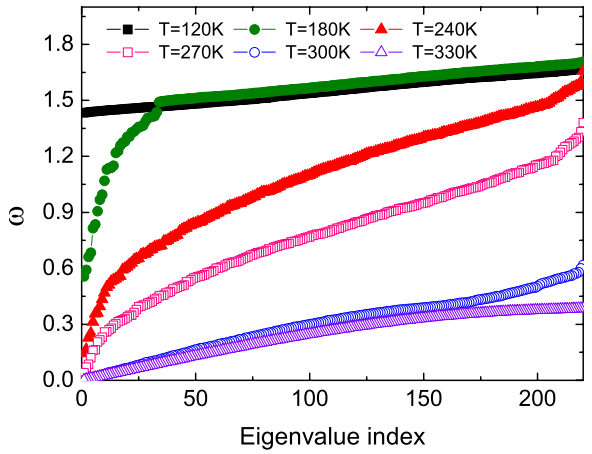

FIG. 5. (Color online) PC frequency spectrum at different temperatures between 120 and 330 K. Frequency units are $(D/m)^{1/2}\alpha=5.15\times10^{12}$ s^{-1}.

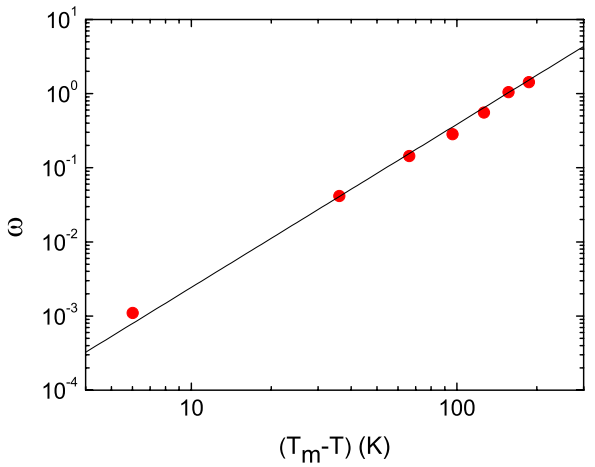

FIG. 6. (Color online) Temperature dependence of the lowest PC frequency. The straight line corresponds to the best fit with the parameters given in the text. Frequency units are $(D/m)^{1/2}\alpha=5.15\times10^{12}$ s^{-1}.

as the temperature at which the DNA strands are half-denatured. For a homogeneous A-T chain, the melting transition has been reported to be around $T_m=310$ K [17,23].

One should be aware that, in general, the melting temperature depends on the length of the sequence, base composition, topological structure, and salt concentration. Thus, it is difficult to predict the exact transition temperature of a given sequence. In the case of the uniform chain the melting temperature depends strongly on salt concentration, which is not explicitly included in this mesoscopic model. Usual parameters for the standard PBD model can be found in Refs. [16,17], for instance. Following previous findings and guided by our numerical simulation results we choose the following set of values for studying the standard model: $m=300$ Da, $D=0.043$ eV, $\alpha=4$ Å^{-1}, $K=0.03$ eV Å^2, $\rho=3$, and $\delta=0.8$ Å^{-1} (these values are slightly different from those given in [17]). If a barrier is included we modify the value of D to obtain the same T_m. Thus, in this case $D=0.051\,85$ eV, and $G=3D=0.1556$ eV, $y_0=2/\alpha=0.5$ Å, and $b=1/2\alpha^2=0.031\,25$ Å^2.

The melting transition curve for the model with and without solvation barrier is shown in Fig. 2 for a homogeneous A-T sequence with $N=220$. At high temperature is observed the expected linear behavior of a free Gaussian polymer chain, i.e., the completely unzipped state.

The numerical determination of the melting temperature is not a trivial issue. For the case of a uniform chain, we use the following computational criteria to determine the melting temperature and transition width from the $\langle u(T)\rangle$ and $\langle y(T)\rangle$ curves. First we define two temperatures. The larger one, T_2, gives the onset of the linear behavior in $\langle u(T)\rangle$ which indicates that the chain is completely melted. The other one, T_1, estimates the beginning of the transition defined as the temperature for which $\langle y(T)\rangle=y_0$, i.e., when the chain can be considered to be by average in the barrier position. Then we define the transition width as $\Delta T=T_2-T_1$ and the melting temperature as $T_m=(T_1+T_2)/2$.

It is interesting to briefly discuss the effect of the main model parameters in the melting transition and the dynamical behavior of the system. High values of D, the Morse potential dissociation energy, produce also high T_m's. Parameters K and ρ set the stacking interaction between adjacent bases. Its effect in T_m and ΔT can be seen in Figs. 3 and 4. Suitable melting temperature and transition width are obtained at high K values (thus, we chose $K=0.03$ eV/Å^2) and moderate ρ's ($\rho\sim3$). Moderate ρ's are biologically satisfactory and avoid the very narrow bubbles observed for high ρ values. With respect to the solvation barrier its presence makes bubbles live longer. Thus, a complete separation of the strands is also facilitated leading to a decrease in the melting temperature (this change can be counterbalanced by increasing the D value). The inclusion of the barrier also reduces the width of the transition, allowing the choice of moderate ρ values.

B. PCA of the phase transition

PCA provide us another method to analyze the phase transition. Figure 5 shows the spectrum of effective frequencies as defined in Eq. (10) for the uniform A-T chain. At low temperatures, a sharp band of linear wave excitations close to the ground state is clearly identified. At very low temperatures, the frequencies are given by the dispersion relation [24]

$$m\omega^2 \simeq 2D\alpha^2 + 2K(1+\rho)[1-\cos(\pi n/N)]. \qquad (11)$$

Thus, the gap for the lowest mode is controlled by the intrabase potential parameters. In this limit the PC eigenvectors are the normal modes of a homogeneous array. As temperature increases, the nonlinear excitations are more important. Such nonlinear modes are responsible of the larger fluctuations and correspond to the larger eigenvalues or lower frequencies. As we approach the melting temperature a soft mode goes to zero (see Fig. 5), the chain is not pinned by an on-site potential, and every strand moves freely. Hence, at high temperature the frequencies are those of a free Gaussian chain with coupling given by the stacking parameter K:

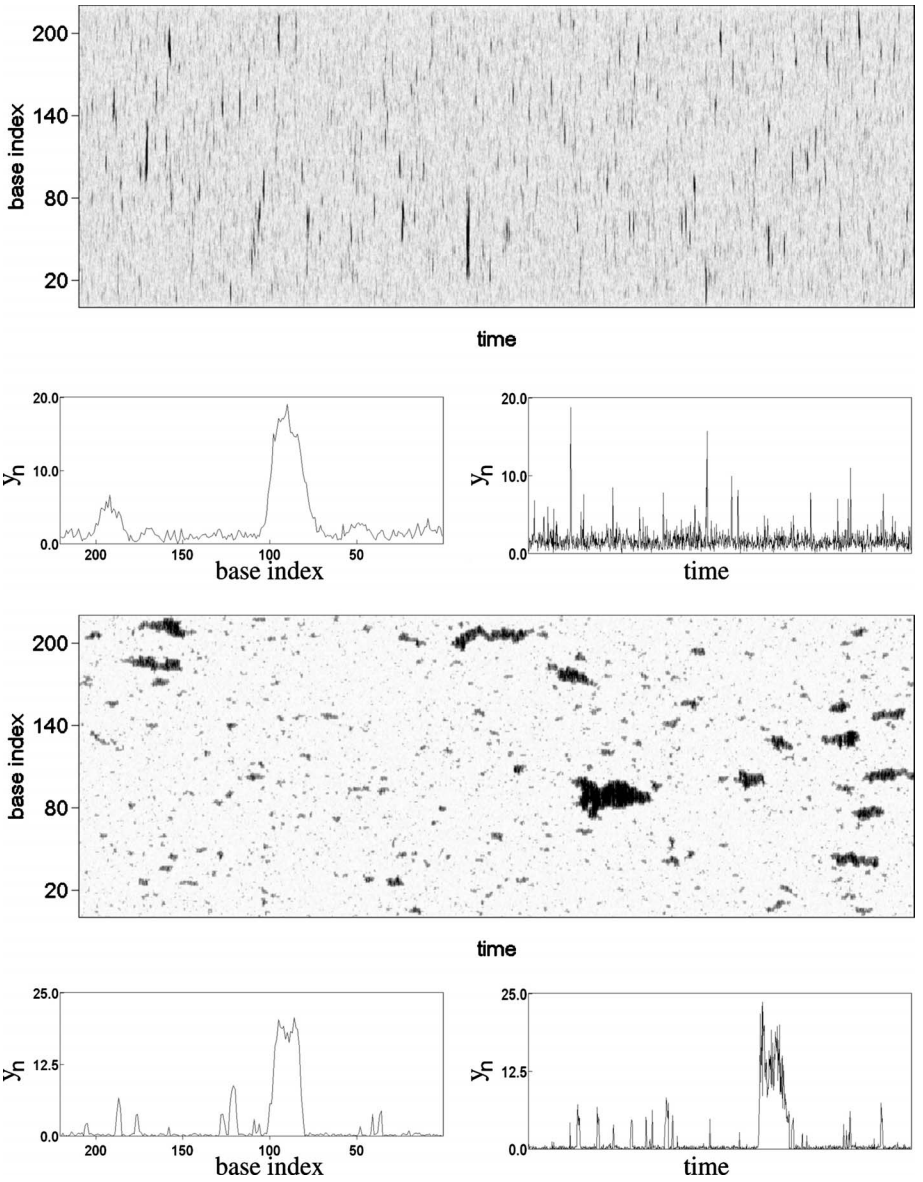

FIG. 7. Typical molecular-dynamics simulation trajectories for the PBD model of a homogeneous sequence without (upper figures) and with (lower ones) barrier. Dark areas correspond to open base pairs. Long-living bubbles are clearly observed when solvation barrier is included. Small figures show a base pair configuration at a given time (left) and the trajectory of a given base as a function of time (right). Trajectory time is 200 ns and y_n is given in units of $\alpha^{-1}=0.25$ Å.

$$m\omega^2 = 2K[1-\cos(\pi n/N)]. \tag{12}$$

Note that the zero mode corresponds to the largest wavelength.

Figure 6 shows the temperature evolution of this mode frequency. The curve can be fitted using a critical behavior function $\omega \propto (T_m - T)^{\nu}$, with $T_m=0.51$ and $\nu=2.2$. The characterization of the order of the phase transition is a difficult issue [25,26]. To our knowledge dynamic exponents for this family of models are not known.

C. Bubble formation

One of the more interesting aspects of DNA molecule amenable to be studied in the framework of the PBD mode is the formation and stability of the so-called DNA bubbles (short open segments of the DNA chain). The onset of such states has an important role in the understanding of the operation of DNA molecules.

We observe that the inclusion of a solvation barrier in the model has a dramatic effect in the dynamics of the system, increasing the lifetime of bubbles. Figure 7 compares molecular-dynamics numerical simulations for a model with barrier to another without barrier and at $\rho=3$. As expected the presence of barriers modifies the kinetics of base opening and further closing. The individual trajectories of the different bases clearly show this behavior. Although the profile of the bubbles is similar in the two cases, the nonbarrier dynamics is completely different from that with barrier. Without a barrier the opening of the base pair corresponds to a large-amplitude oscillation along the Morse potential, and an easy closing is favored. With a barrier, the kinetics is controlled by the presence of two equilibrium states separated by the solvation barrier. Closing events are more difficult, and bubbles live longer. Although in this way we approach experimental values of lifetime bubbles (on the order of ns), this fact makes the simulations more difficult since it requires very long runs to have good statistics.

We see that parameter ρ affects the cooperativity between open bases and then the bubble lifetime and width. A large ρ produces long-live bubbles but extremely narrow ones (one or two bases), as those shown in Ref. [27] with $\rho=25$. With this value of ρ, the transition is extremely narrow and, after a long transient, bubbles nucleate to drive the unzipping of the whole chain. A moderate value, $\rho=3$, is good enough to get both longer and wider bubbles (four to ten bases), as shown in figures.

V. P5 PROMOTER SEQUENCE

A promoter is a region of the DNA sequence where the transcription to mRNA is initiated and controlled. In prokaryotic cells, regions of −10 and −35 bp (upstream from the transcription start site) are the most important sites for transcription regulation. In this section we analyze the P5 core promoter, which has been widely studied in the literature [12,28]. We show that the PCA of Langevin trajectories of the molecule clearly identifies the biologically relevant sites.

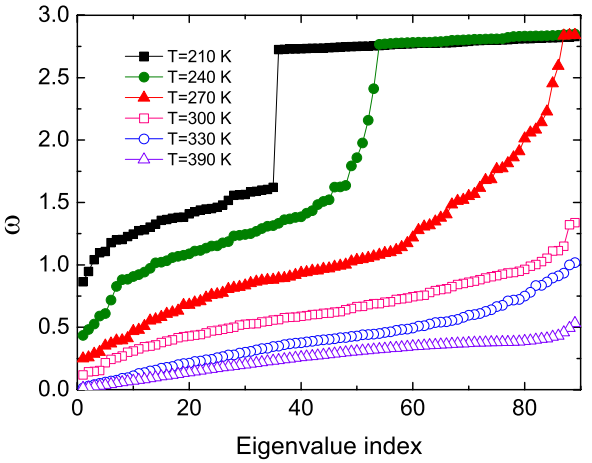

FIG. 8. (Color online) PC frequency spectrum at different temperatures for the P5 promoter sequence. Frequency units are $(D/m)^{1/2}\alpha=5.15\times10^{12}$ s^{-1}.

The sequence of the P5 promoter is given by the 69 pb:

5′-GTGCCCATTTAGGGTATATATGGCCGAGTGAGCGAGC AGGATCTCCATTTTGACCGCAAATTTGAACG-3′.

We will perform numerical simulations of this sequence (parameter values for C-G and A-T base pairs have been given previously). In order to avoid unphysical denaturation events at low temperature due to finite-size effects [6], we apply the following boundary conditions: First we add 10 C-G bp sequences to the ends of the promoter to create *hard* boundaries. Second, we set the extremes to zero (closed) to avoid the complete opening of the chain.

Figure 8 shows the evolution of the PC frequency spectrum with temperature. Several features distinguish these curves from the homogeneous case. For low temperatures, two bands corresponding to A-T (lower band) and C-G (upper band) links are clearly identified. The eigenvectors are localized around rich regions in both kinds of complementary pairs. At intermediate temperatures the gap between both bands disappears and the eigenvectors broaden. C-G pairs surrounded by A-T ones are more likely to open and frequencies diminish.

Close to transition ($T\approx345$ K), several modes detach to low frequencies. The first three modes are strongly localized and show peaks in four regions (see Fig. 9). These modes represent the "softest" regions of the sequence, are related to high probability of opening, and drive the unzipping.

This picture is validated by the study of the first three eigenvectors (the large eigenvalue or small frequency ones). Figure 9 shows the measured probability of opening in the sequence, defined by Eq. (8), and the computed eigenvectors. An excellent correlation between both figures is observed. Region +1 corresponds to the transition starting site (TSS), whereas regions −30 and −40 correspond to the binding sites of transcription factors, i.e., a region rich in A-T pairs like the TATA box. Localized eigenvectors span over regions of ten base pairs, which fairly correspond to the width of the bubbles.

Figure 10 plots a typical trajectory, showing a few bubbles mainly at the regions pointed by the PC eigenvectors. Note that at this temperature very few opening events occur, so a precise statistics on bubble formation becomes difficult. However, PC analysis gives a good account of these

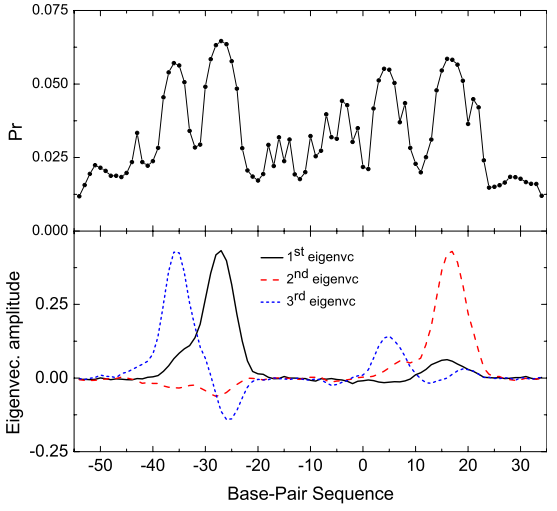

FIG. 9. (Color online) Top figure: probability (not normalized) of opening for sequence of P5 promoter. Bottom figure: the three first PC eigenvectors at $T=290$ K.

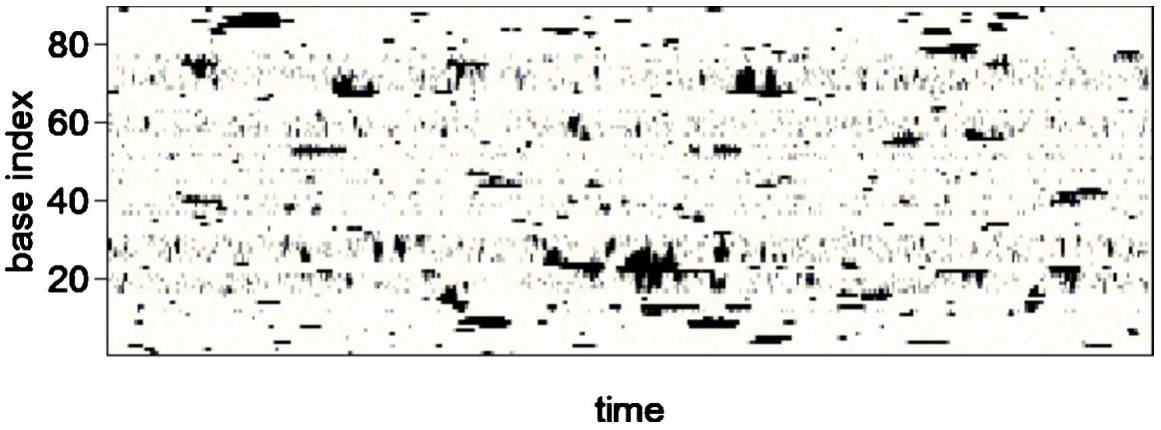

FIG. 10. (Color online) Trajectory of the P5 sequence promoter at $T=290$ K. The larger bubbles appear mainly in the region of the localized PC eigenvectors shown in Fig. 9. The time of the shown trajectory is 400 ns.

sites even with the presence of a few bubbles. Interestingly, the third eigenvalue shows a clear correlation between the region close to the TATA box and the TSS. This represents motion fluctuations that involve both regions. Thus, modification of one of them could interfere in the fluctuations in the other, opening a channel for the regulation of transcription.

VI. CONCLUDING REMARKS

We have studied the dynamics of a DNA molecule using the PBD model with solvation interaction. Solvation interaction is modeled by including a barrier in the usual on-site Morse potential. First we have analyzed the melting transition in an A-T homogeneous chain. We conclude that the inclusion of the barrier not only modifies the phase transition but also has a great influence in the dynamics of localized excitations (bubbles). In combination with the nonlinear parameter ρ, the width and lifetime of bubbles can be tuned. We find a set of parameters which will be suitable for future studies with this model.

We have applied the model to an inhomogeneous sequence with biological meaning, a virus promoter. The study of trajectories using principal component analysis allows us to detect the biologically relevant sites without the need of long molecular-dynamics runs. We have found that the softest modes of the PC spectrum are highly localized in those sites. Even more, these sites have been also detected as the most likely to be opened in bubbles [29]. In this work we cannot elucidate the controversy on whether the DNA can direct its own transcription as was suggested in [12] and then make use of theoretical methods to obtain the functional regions. However, we can obtain from the mesoscopic PBD model some regions which are more sensible to the formation of bubbles (as extracted from the eigenvectors of the PCA). Of course, the simplicity of the model cannot take into account other effects, like the flexibility of DNA molecule, but it is able to tackle those related to the sequence. Finally, we stress the use of a tool, the PCA, to the analysis of statistical-mechanics simulations of simple models to obtain useful information on linear and nonlinear excitations and the relation to its phase transition.

ACKNOWLEDGMENTS

We thank L. M. Floría and D. Prada-Gracia for helpful comments and discussion. R.T.R. acknowledges BIFI for financial support. This work was supported by the Spanish MICINN Project No. FIS2008-01240, co-financed by FEDER funds.

[1] K. Sneppen and G. Zocchi, *Physics in Molecular Biology* (Cambridge University Press, Cambridge, England, 2005).
[2] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[3] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
[4] M. Peyrard, Nonlinearity **17**, R1 (2004).
[5] M. Peyrard, Nat. Phys. **2**, 13 (2006).
[6] M. Peyrard, S. Cuesta-López, and D. Angelov, J. Phys.: Condens. Matter **21**, 034103 (2009).
[7] S. Cuesta-López, J. Errami, F. Falo, and M. Peyrard, J. Biol. Phys. **31**, 273 (2005).
[8] N. K. Voulgarakis, A. Redondo, A. R. Bishop, and K. O. Rasmussen, Phys. Rev. Lett. **96**, 248101 (2006).
[9] S. Ares, N. K. Voulgarakis, K. O. Rasmussen, and A. R. Bishop, Phys. Rev. Lett. **94**, 035504 (2005).
[10] S. Ares, Ph.D. thesis, Universidad Carlos III, 2005.
[11] C. H. Choi *et al.*, Nucleic Acids Res. **32**, 1584 (2004).
[12] G. Kalosakas, K. O. Rasmussen, A. R. Bishop, C. H. Choi, and A. Usheva, EPL **68**, 127 (2004).
[13] T. S. van Erp, S. Cuesta-López, and M. Peyrard, Eur. Phys. J. E **20**, 421 (2006).
[14] T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, Phys. Rev. Lett. **97**, 059802 (2006).
[15] G. Weber, EPL **73**, 806 (2006).
[16] A. Campa and A. Giansanti, Phys. Rev. E **58**, 3585 (1998).
[17] B. S. Alexandrov *et al.*, Nucleic Acids Res. **37**, 2405 (2009).
[18] E. Helfand, Bell Syst. Tech. J. **58**, 2289 (1979).
[19] H. S. Greenside and E. Helfand, Bell Syst. Tech. J. **60**, 1927 (1981).
[20] I. T. Jolliffe, *Principal Components Analysis*, 2nd ed.

(Springer-Verlag, New York, 2002).
[21] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen, Proteins **17**, 412 (1993).
[22] I. Bahar and A. J. Rader, Curr. Opin. Struct. Biol. **15**, 586 (2005).
[23] R. D. Wells *et al.*, J. Mol. Biol. **54**, 465 (1970).
[24] The addition of the barrier to the standard Morse potential that we use barely modifies the original position and curvature of the potential minimum [Eq. (4)].
[25] J. M. Romero-Enrique, F. de los Santos, and M. A. Muñoz, EPL **89**, 40011 (2010).
[26] M. Joyeux and A. M. Florescu, J. Phys.: Condens. Matter **21**, 034101 (2009).
[27] M. Peyrard, S. Cuesta-López, and G. James, Nonlinearity **21**, T91 (2008).
[28] B. S. Alexandrov *et al.*, PLOS Comput. Biol. **5**, e1000313 (2009).
[29] B. Alexandrov, N. K. Voulgarakis, K. O. Rasmunssen, A. Usheva, and A. R. Bishop, J. Phys.: Condens. Matter **21**, 034107 (2009).