On the discrete Peyrard-Bishop model of DNA: Stationary solutions and stability

Sara Cuenda and Angel Sánchez

Articles you may be interested in

Equilibration of complexes of DNA and H-NS proteins on charged surfaces: A coarse-grained model point of view
J. Chem. Phys. **141**, 115102 (2014); 10.1063/1.4895819

Mathematical modeling of DNA's transcription process for the cancer study
AIP Conf. Proc. **1494**, 161 (2012); 10.1063/1.4764631

The self-assembly of DNA Holliday junctions studied with a minimal model
J. Chem. Phys. **130**, 065101 (2009); 10.1063/1.3055595

Computer simulation study of probe-target hybridization in model DNA microarrays: Effect of probe surface density and target concentration
J. Chem. Phys. **127**, 144912 (2007); 10.1063/1.2787618

All-or-none switching of transcriptional activity on single DNA molecules caused by a discrete conformational transition
Appl. Phys. Lett. **86**, 223901 (2005); 10.1063/1.1937990

On the discrete Peyrard-Bishop model of DNA: Stationary solutions and stability

Sara Cuenda[a)]
Grupo Interdisciplinar de Sistemas Complejos (GISC) and Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain

Angel Sánchez[b)]
Grupo Interdisciplinar de Sistemas Complejos (GISC) and Departamento de Matemáticas, Universidad Carlos III de Madrid, Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain and Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Facultad de Ciencias, Universidad de Zaragoza, 50009 Zaragoza, Spain

As a first step in the search of an analytical study of mechanical denaturation of DNA in terms of the sequence, we study stable, stationary solutions in the discrete, finite, and homogeneous Peyrard-Bishop DNA model. We find and classify all the stationary solutions of the model, as well as analytic approximations of them, both in the continuum and in the discrete limits. Our results explain the structure of the solutions reported by Theodorakopoulos *et al.* [Phys. Rev. Lett. **93**, 258101 (2004)] and provide a way to proceed to the analysis of the generalized version of the model incorporating the genetic information. © *2006 American Institute of Physics.*
[DOI: 10.1063/1.2194468]

DNA, the molecule that constitutes the basis of the genetic code, is of utmost importance. In particular, its mechanical properties are crucial, as opening the double helix structure of DNA is needed to read the genetic code and for replication of the molecule for reproduction. The complete separation of the double helix is called replication, and can be achieved by heating, or mechanically, by pulling the two strands of the molecule apart. We here address the mathematical description of mechanical denaturation in terms of a simple model. We determine and classify the solutions of the model equations and study their stability properties. We also provide an approximate but very accurate way to deal analytically with those solutions. Beyond mechanical features, our results are relevant for studies of the thermodynamic properties of the DNA chain, and may have genomic applications, in so far as mechanical denaturation experiments that give information about DNA composition can be modeled by our model and solutions.

I. INTRODUCTION

Nonlinear models supporting coherent excitations appear in many fields of science since the pioneering discoveries of Fermi, Pasta, and Ulam.[1] This work, in the field of physics, has led many scientist to use nonlinear models in the study of complex systems[2] in other subjects. Nonlinear models entered into DNA physics with Englander and co-workers[3] (see Ref. 4 for a review on nonlinear models of DNA), in 1980, when they modeled the dynamics of DNA with a sine-Gordon equation. Since then, a lot of work has been devoted

to nonlinear excitations in DNA, both from the dynamics and the statistical mechanics points of view. Among this body of work, a particularly successful model is the Peyrard-Bishop (PB) one,[5,6] which will be our starting point in this paper.

One problem of special interest in the framework of DNA was the thermal denaturation transition, which takes place at temperatures around 90 °C, when the two strands of the DNA molecule separate. On the other hand, mechanical denaturation, which occurs when one of the strands of the molecule is separated from the other by pulling it in single-molecule experiments, was achieved in the last few years.[7] In order to model these phenomena, most of the research done so far refers to *homopolymers*; i.e., homogeneous DNA molecules consisting entirely of A-T or C-G base pairs. When the issue under discussion is genomics, or gene identification, which is very much related to the above-mentioned problems, models of *heteropolymers* are required: The distribution of A-T and C-G base pairs follows nonuniform, non-homogeneous sequences obtained from genome analysis. The heterogeneous PB model is also being used for identifying relevant sites, such as promoters,[8,9] in viral sequences, and also for analyzing the thermal denaturation process.[10]

The main motivation of this work is the study of the effects of the sequence heterogeneity on the dynamics of the mechanical denaturation process. We began that research program by analyzing the Englander (basically, the sine-Gordon equation[2]) model. The results we obtained[11,12] showed that the Englander model was much too simple to reproduce the phenomena observed in experiments, and therefore we decided to focus on the PB model (see Ref. 13 for an authoritative review about this model). In that context, our immediate aim was to obtain a tool in this model similar the effective potential proposed for the Englander model by

[a)]Electronic mail: scuenda@math.uc3m.es
[b)]Electronic mail: anxo@math.uc3m.es

     16, 023123-1

Salerno and Kivshar[14–16] in order to study the relation between the dynamics of these excitations and gene identification. To that end, it is necessary to obtain stationary states of the homogeneous model. Those were available for the continuous version of the PB model, but, in fact, DNA is quite a discrete system, and the discretization parameter depends on experimental measurements used as parameters in the model. For instance, in the PB model, the dimensionless parameter that defines the effective discretization of the system can go from $R=10.1$ (see the next section for a definition of R), used in Refs. 13 and 17 to $R \simeq 75$ used in Ref. 18, or even to $R=100$ in Ref. 19. In all cases, these R values correspond to systems that are far from the continuum limit. Therefore, as a first step towards our chief goal of understanding sequence effects on denaturation, our immediate purpose is to study stable, stationary states in the discrete PB model, with a special focus on their dependence on the effective discretization.

In this paper, we aim to find stationary solutions of the PB model and their corresponding stability conditions. These issues are addressed in Sec. II. Subsequently, we discuss the validity of the continuum limit and the domain wall approximation in Sec. III, while in the main part of the paper, Sec. IV, we propose analytical approximations for the discrete case and compare our results with the ones obtained in Ref. 17. Finally, Sec. V concludes the paper by summarizing our main results and their possible implications.

II. DISCRETE SOLUTIONS AND STABILITY

In the following we will use the dimensionless PB model, defined by the Hamiltonian

$$H = \sum_{n=0}^{N-1} \left\{ \frac{1}{2}\dot{Y}_n^2 + \frac{1}{2R}(Y_{n+1} - Y_n)^2 + V(Y_n) \right\}, \tag{1}$$

where $V(Y)=(1-e^{-Y})^2$ is the Morse potential, which stands for the attraction between the two bases of a base pair, and R is a positive, dimensionless constant that refers to the intensity of the coupling of two consecutive bases. This constant plays the role of an effective discretization, $a=\sqrt{R}$, so that $R \gg 1$ stands for a large discretization and $R \ll 1$ is the continuous limit.

Static solutions of Hamiltonian (1) must satisfy $\partial H / \partial Y_n = 0$, which turns out to be the recurrence relation

$$Y_{n+1} = 2Y_n - Y_{n-1} + RV'(Y_n), \tag{2}$$

for $n=1,2,\ldots,N$. Hence, solutions (2) are uniquely defined by the initial condition $\{Y_0, Y_1\}$. If we restrict ourselves to solutions with $Y_0=0$, which we can do without loss of generality, then each Y_n will depend only on the value $Y_1=y$, so that Eq. (2) can be rewritten in terms of y, introducing the notation $Y_n(y)$ instead of Y_n. From now on, we will discuss the behavior of the solutions $Y_n(y)$ as a function of y.

Equation (2) describes stable and unstable solutions of Hamiltonian equation (1). In order to assess the stability properties of the solutions, we need to study the hessian of the system:

$$\mathcal{H}_N(y) = \begin{pmatrix} d_1(y) & -1 & 0 & \ldots & 0 \\ -1 & d_2(y) & -1 & \ldots & 0 \\ 0 & -1 & d_3(y) & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & d_N(y) \end{pmatrix}, \tag{3}$$

where $d_n(y) = 2 + RV''(Y_n(y))$. Calling $\Delta_n(y)$ the determinant of the principal minor of order n of the hessian $\mathcal{H}_N(y)$, i.e.,

$$\Delta_n(y) = \det(\mathcal{H}_n(y)), \tag{4}$$

a stable solution must satisfy $\Delta_n(y) > 0$ for all $n = 1, 2, \ldots, N$. As the hessian is a tridiagonal matrix, there is a recursive relation between different Δ_n,

$$\Delta_{n+1}(y) = d_{n+1}(y)\Delta_n(y) - \Delta_{n-1}(y), \tag{5}$$

with $\Delta_1 = d_1$ and $\Delta_2 = d_1 d_2 - 1$.

Expression (5) above can be rewritten in terms of $Y_n'(y)$. By deriving expression (2) with respect to y, we find

$$\begin{aligned} Y_{n+1}'(y) &= \frac{dY_{n+1}(y)}{dy} \\ &= [2 + RV''(Y_n(y))]Y_n'(y) - Y_{n-1}'(y) \\ &= d_n(y)Y_n'(y) - Y_{n-1}'(y), \end{aligned} \tag{6}$$

with $Y_2'(y) = \Delta_1(y)$ and $Y_3'(y) = \Delta_2(y)$. Therefore, it has to be

$$\Delta_n(y) = Y_{n+1}'(y), \tag{7}$$

for $n = 1, 2, \ldots, N$, and hence the stability region of solutions (2) are the points that satisfy $Y_n'(y) > 0$ for all $n = 2, 3, \ldots, N+1$.

This far, no approximations were needed to obtain these results, which are still valid for any $V(Y)$. From now on, we will focus on the PB model by choosing the Morse potential as our $V(Y)$, in order to search for an analytic expression of the solutions (2), as well as to find the stability in terms of the initial condition y.

III. CONTINUUM LIMIT OF THE PEYRARD-BISHOP MODEL

This limit corresponds to taking $R \ll 1$, which means that we can use the approximation $Y_n(y) \rightarrow Y_{\text{cont}}(x, y)$ with $x = n\sqrt{R}$ in Eq. (2). By so doing, we obtain the following differential equation:

$$\frac{\partial^2 Y_{\text{cont}}}{\partial x^2} = \frac{dV}{dY_{\text{cont}}}, \tag{8}$$

which can be easily solved using the initial conditions $Y_{\text{cont}}(x_0, y) = 0$ and $\partial Y_{\text{cont}}(x_0, y)/\partial x = y/\sqrt{R}$, where x_0 stands for the initial site of the model. This solution is

$e^{Y_{cont}(x,y)}$

$$= \frac{y\sqrt{\dfrac{2}{R}}\sinh\left[\sqrt{2+\dfrac{y^2}{R}}(x-x_0)+\sinh^{-1}\left(\dfrac{y}{\sqrt{2R}}\right)\right]+2}{2+\dfrac{y^2}{R}}$$

$$(9)$$

for $x \geq x_0$. This result differs from the domain wall obtained in the continuum limit in the PB model,[13]

$$e^{Y(x)} = 1 + e^{\sqrt{2}(x-x_0)}, \tag{10}$$

mainly because the integration has been done with different boundary conditions. Whereas Eq. (9) was obtained integrating Eq. (8) for a *semi-infinite* system starting on x_0 with the initial conditions stated above, Eq. (10) is valid for an *infinite* system with $Y(x)=0$ and $Y'(x)=0$ as x goes to $-\infty$. This yields two different pictures of DNA: The one of Eq. (9), describing a system with a starting point in x_0, and with an opening depending on the value of the derivative $\partial Y_{cont}(x_0,y)/\partial x$ [if this value is zero then $Y_{cont}(x_0,y)=0$ and the DNA chain is closed]; and the one of Eq. (10), which describes an infinite system translationally invariant.

In this respect, we believe that Eq. (10) should not be used as an approximation of a finite system because it is not a critical point of the continuum version of Eq. (1) and, therefore, we cannot speak of stability in this case as long as we consider DNA as a finite lattice. The case of Eq. (9) is different because, although still in a semi-infinite system, the stability of a finite chain can be stated as long as we consider fixed boundary conditions: Indeed, if free boundary conditions are imposed, then the only possible solution for N finite is $Y_n=0$ for all $n=0,\ldots,N$, because an additional $Y_{N-1}=Y_N$ condition is imposed. Thus, from now on, we will consider Y_N as obtained from the recurrence relation (2) and then froze on that value when considering dynamics; i.e., integrating in time with fixed boundary conditions.

An important feature of the continuum approximation is that, in Eq. (9), Y_{cont} can be written as a function of x and ξ, where $\xi=y/\sqrt{R}$. This implies a scaling relation between these two parameters, a relation that is absent for solutions of the discrete limit. This behavior can be seen in Fig. 1, where we represent Y_{cont} with respect to x (with $x_0=0$) for two values of ξ, compared to the exact result $Y_n(y)$ (recall that $x=n\sqrt{R}$) for different values of R, and with $y=\xi\sqrt{R}$. We clearly observe that for the largest values of R ($R=1$ and, mostly, $R=10$) the scaling relation is not fulfilled, indicating the crossover to the discrete limit regime.

To analyze the stability of these solutions, using the result of Sec. II, it is enough to study the sign of $\partial Y_{cont}(x,y)/\partial y$ for all $x\geq 0$. As the derivative of expression (9) is quite cumbersome, we prefer to show plots of the result for different values of x, which we collect in Fig. 2. It can be shown in general that $\partial Y_{cont}(x,y)/\partial y>0$ for all $x>0$, and therefore Eq. (9) is a stable solution of (2).

In order to check these results, we used the equations of motion of the model (1) in order to simulate the dynamics of initial data given by (9) with small perturbations, and then
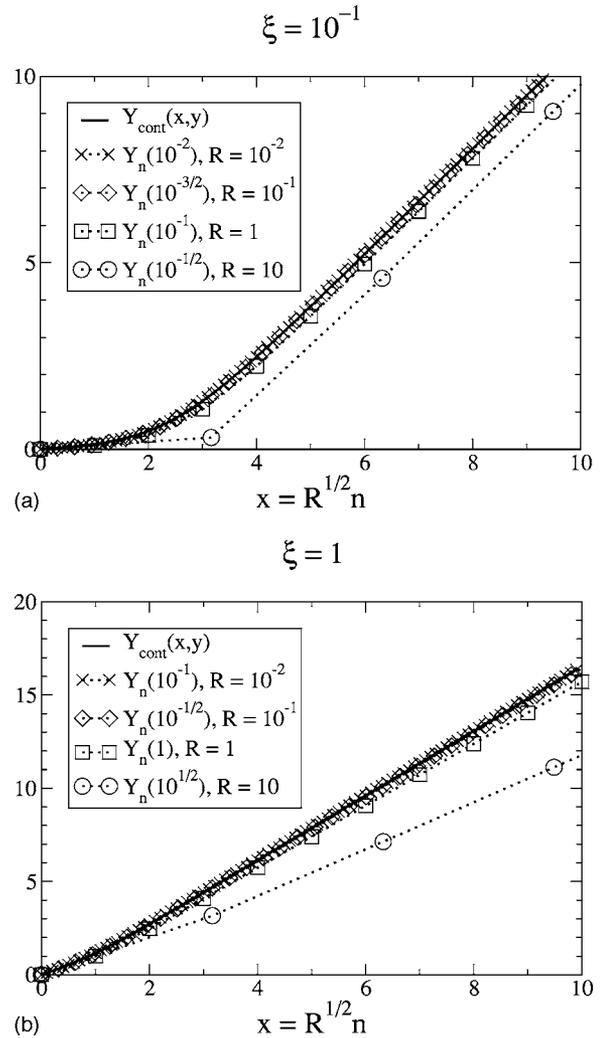

FIG. 1. Plots of $Y_{cont}(x,y)$ [obtained from expression (9)] for $\xi=0.1$ (a) and $\xi=1$ (b), where $\xi=y/\sqrt{R}$, compared to the exact solutions $Y_n(y)$ of the discrete recurrence relation (2) calculated for the same quotient $\xi=y/\sqrt{R}$ but for different values of R (and, therefore, different values of y).
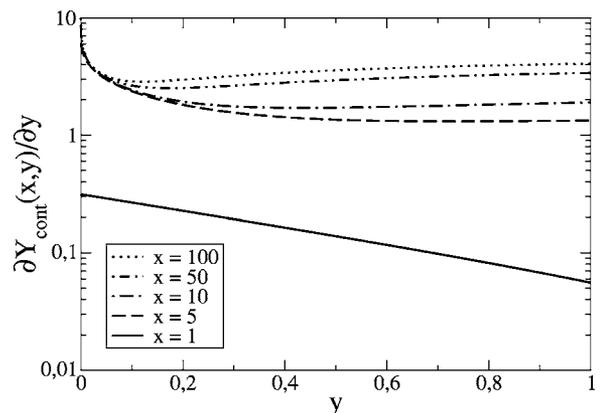

FIG. 2. Plots of $\partial Y_{cont}(x,y)/\partial y$ for (from lower to higher) $x=1$, $x=5$, $x=10$, $x=50$, and $x=100$. All of them are positive for any $y>0$, so that all solutions of the form (9) are stable for any $y>0$.

monitored the evolution of this curves in time. Fixed boundary conditions at both ends of the simulated interval were used, in order to prevent the chain from spontaneously closing, as explained above. Therefore, in order to verify our results we had to restrict the simulations to the sector of open-chain solutions by choosing those boundary conditions. With that caveat, our simulations fully confirm the predicted stability of solutions. We stress that such solutions are the ones that are relevant to the mechanical denaturation problem, where the spontaneous closing of the chain is prevented by the force exerted on the open end.

IV. DISCRETE LIMIT OF THE PEYRARD-BISHOP MODEL

A. Solutions

The discrete limit of the PB model corresponds to letting $R \gg 1$, and can be obtained following a few steps. Using a telescopic summation of $Y_{n+1} - Y_n$, and noting the initial conditions $Y_0 = 0$ and $Y_1 = y$, Eq. (2) can be rewritten as

$$Y_{n+1}(y) = (n+1)y + R\sum_{k=1}^{n} (n+1-k)V'(Y_k(y)). \quad (11)$$

We now define

$$f_k(y) \equiv V'(Y_k(y)) = 2e^{-Y_k(y)}(1 - e^{-Y_k(y)}) = f_1(Y_k(y)). \quad (12)$$

These functions are plotted for different values of R in the discrete limit in Fig. 3. As can be seen, these f_k are strongly localized, their overlapping depending on R. In fact, in the discrete limit we are working on, which implies low overlapping of the curves, we can write

$$f_k(y) \simeq f_k^{(1)}(y) \equiv f_1(b_k y), \quad (13)$$

where b_n is calculated as follows. Let $f_1(y) = 2e^{-y}(1 - e^{-y})$, and $y_1 = \log 2$ the position of the maxima of f_1. The expression of $f_2(y)$ obtained from (12) can then be approximated by $f_1(y) \simeq 2y$ for $y \ll y_1$, finding $b_2 = 2R + 2$. Repeating the process for f_{n+1} in terms of $f_n, f_{n-1}, \ldots, f_1$, we find the linear recurrence relation $b_{n+1} = (2R+2)b_n - b_{n-1}$, valid for $y \ll y_n$, which can be easily solved from $b_1 = 1$, $b_2 = 2R+2$ to find

$$b_n = \frac{1}{2\sqrt{R(R+2)}}\left\{\left[R+1+\sqrt{R(R+2)}\right]^n \right. $$
$$\left. - \left[R+1-\sqrt{R(R+2)}\right]^n\right\}. \quad (14)$$

With these values of b_n we can now compare the exact f_n and the approximate $f_n^{(1)}$ for different values of n and R. Figure 4 shows the validity of the approximation (see caption for more details).

This result allows us to obtain an analytic, approximate expression of the solutions for different y in the discrete limit. Substituting it in Eq. (11), we find that

$$Y_{n+1}^{(1)}(y) = (n+1)y + R\sum_{k=1}^{n} (n+1-k)f_k^{(1)}(y). \quad (15)$$

is a good approximation of the exact solution $Y_n(y)$ for large values of R. Figure 5 confirms the accuracy of this approximation: for $R \gtrsim 100$, the approximation is very accurate,
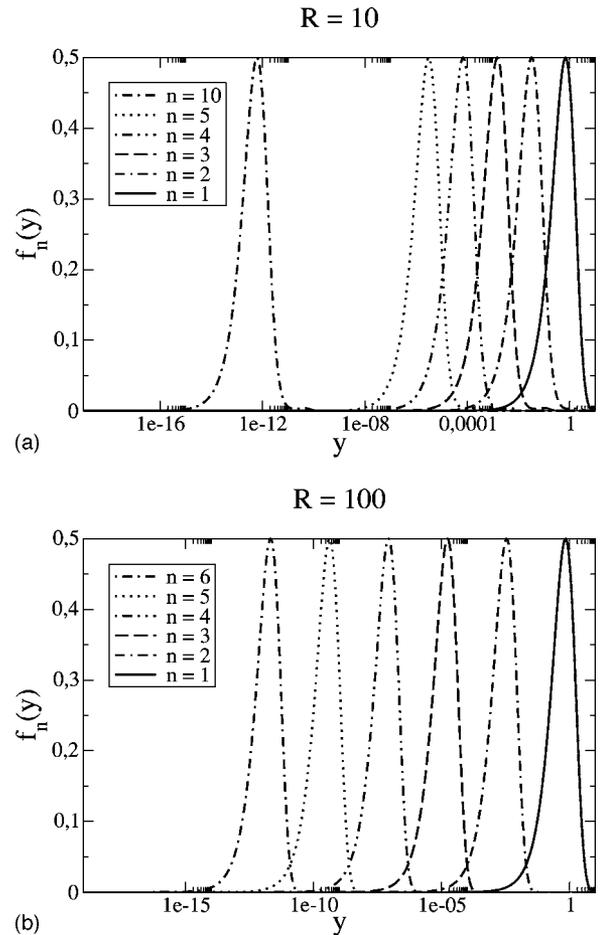
R = 10

(a)

R = 100

(b)

FIG. 3. Functions $f_n(y)$ for $R=10$ (a) and $R=100$ (b), for different values of n.

whereas the smaller R the worse the approximation. For smaller values of R, the approximation can be improved by resorting to the next function, $f_2(y)$, instead of $f_1(y)$, as a substitute for the rest of the $f_k(y)$, by defining

$$f_k(y) \simeq f_k^{(2)}(y) \equiv f_2\left(y\frac{b_k}{b_2}\right) \quad (16)$$

for $k=3,4\ldots$, and approximating $Y_n(y)$ by

$$Y_{n+1}^{(2)}(y) = (n+1)y + nRf_1(y) + R\sum_{k=2}^{n} (n+1-k)f_k^{(2)}(y). \quad (17)$$

In this case, the approximation is even better than for $Y_n^{(1)}(y)$, and even for $R=10$ the results are very close to the exact ones (see Fig. 6 for details).

The errors of these approximations depend on the values of R and n. For instance, $Y_n^{(1)}(y)$ is exact for $n=1$ and $n=2$, for any value of R, whereas $Y_n^{(2)}(y)$ is exact up to $n=3$ for any R. For low values of n, the main difference between the exact $f_n(y)$ (which can be easily obtained numerically) and $f_n^{(1)}$ is located around the secondary heap of $f_n(y)$, with a maximum error $E_{\max}^{(1)} \simeq 0.06$ for $R=10$ and $E_{\max}^{(1)} \simeq 0.006$ for $R=100$. For the second-order approximation based on $f_n^{(2)}$, the maximum error corresponds to the third heap with $E_{\max}^{(2)} \simeq 2.7 \times 10^{-3}$ for $R=10$ and $E_{\max}^{(2)} \simeq 3.2 \times 10^{-5}$ for R
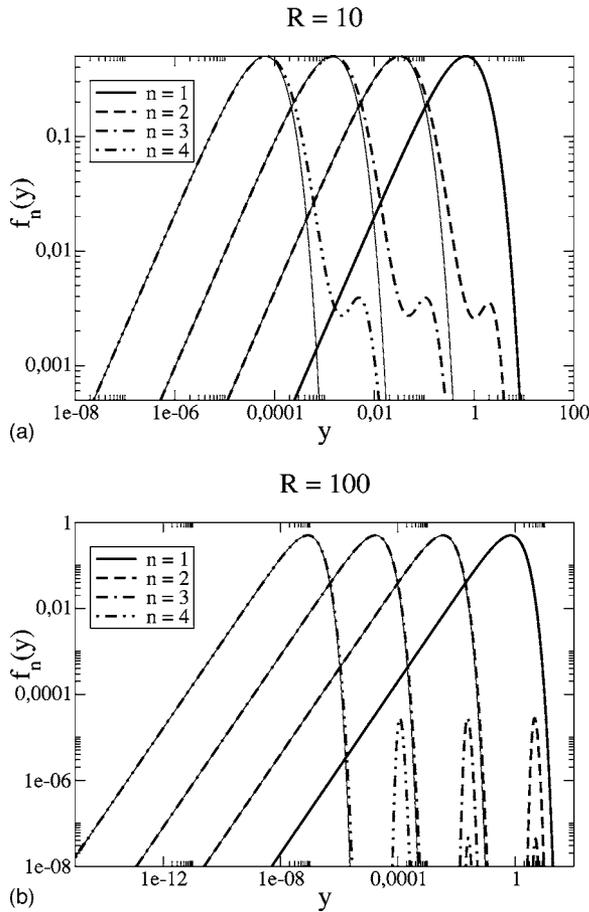
R = 10

(a)

R = 100

(b)

R = 10

(a)

R = 100

(b)

FIG. 4. Approximation $f_n^{(1)}(y)$ [see Eq. (13)] vs exact $f_n(y)$, for $R=10$ (a) and $R=100$ (b) and different values of n. Exact solutions are drawn with thick lines (solid or dashed as indicated), whereas the corresponding approximations are drawn in thin, solid lines. The vertical axis is in logarithmic scale to enlarge the differences between each pair of curves. It is clearly shown that the approximation is better for higher values of R (notice the different range of the vertical axis in both plots). The plots also show the main difference between f_n and f_{n-1}, which is the growing of a new, much smaller heap for higher values of y. In this way, f_1 would show just one heap, f_2 would show a secondary heap, f_3 three, and so on.

$=100$; at the same time, another discrepant region, much less so than the main one, appears around the position of the next heap. The same calculation can be done for higher orders of the approximating function $f_n^{(k)}(y)$, and it can be seen that the reduction of the error using $k+1$ instead of k is at least of one order of magnitude. There is a computational limit near the precision of the machine, which does not allow us to check the validity of this assumption further than a certain k and n, depending of the value of R, but, as far as we know, it is reasonable to expect that the same behavior will take place for higher values of k and n. Therefore, we conjecture that higher orders of functions $f_k(y)$ can be used as approximations of $f_n(y)$, as $f_n(y) \simeq f_n^{(k)}(y)$, with

$$f_n^{(k)}(y) = f_k\left(y\frac{b(n)}{b(k)}\right) \tag{18}$$

for $n > k$, in order to obtain better approximations $Y_n^{(k)}(y)$ of the exact solution $Y_n(y)$, and that the error of an approxima-
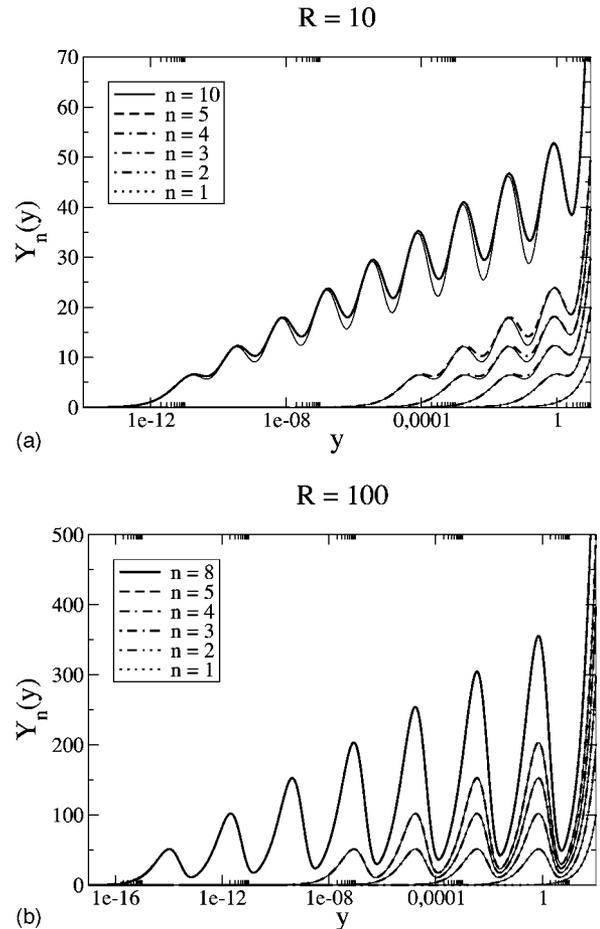
FIG. 5. Approximation $Y_n^{(1)}(y)$ [see Eq. (15)] vs exact $Y_n(y)$, for $R=10$ (a) and $R=100$ (b), for different values of n. Exact solutions are drawn with thick lines, whereas the corresponding approximations are drawn in thin, solid lines.

tion of order k, $E_n^{(k)}(y) = Y_n(y) - Y_n^{(k)}(y)$, can be estimated as the difference

$$E_n^{(k)}(y) \simeq Y_n^{(k+1)}(y) - Y_n^{(k)}(y) + \mathcal{O}(E_n^{(k+1)}(y)), \tag{19}$$

with $E_n^{(k+1)}(y) \ll E_n^{(k)}(y)$.

B. Stability

The approximations defined in (15) and (17), as well as the ones mentioned in the above section, allow us to calculate very accurately the solution $Y_n(y)$ for any value of n and y. This is important because in the exact, numerical calculation of Eqs. (2) and (11) there are problems for values of y close to zero, due to the numerical precision of the computer (see also the discussion below). Therefore, for analyzing the stability in the discrete limit, we proceed to use the approximations $Y_n^{(k)}(y)$ previously discussed. By these means, we can work with systems of much larger size than the ones that could be studied solving numerically the original recurrence relations. For comparison, in the study of stability we will show results for systems of small size, where the derivative $Y_n'(y)$ can be calculated without approximations for each n without high errors of the precision of the computer. In Fig. 7 we show the dependence of $\partial Y_n(y)/\partial \log_{10}(y)$ on function
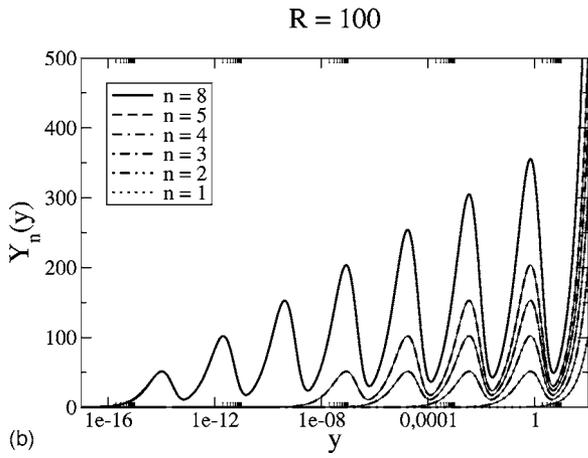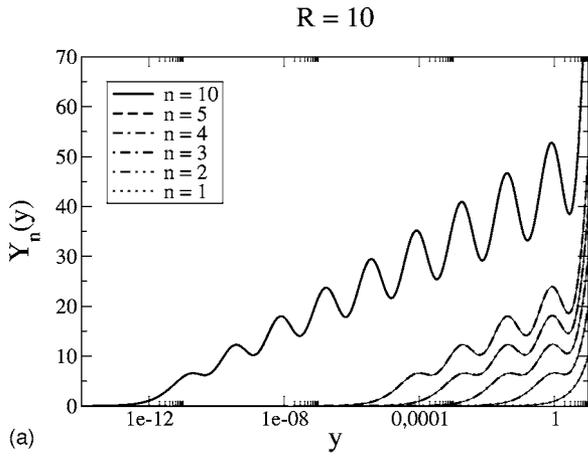
R = 10

(a)

R = 100

(b)

FIG. 6. Approximation $Y_n^{(2)}(y)$ [see Eq. (17)] vs exact $Y_n(y)$, for $R=10$ (a) and $R=100$ (b), for different values of n. Exact solutions are drawn with thick lines, whereas the corresponding approximations are drawn in thin, solid lines.
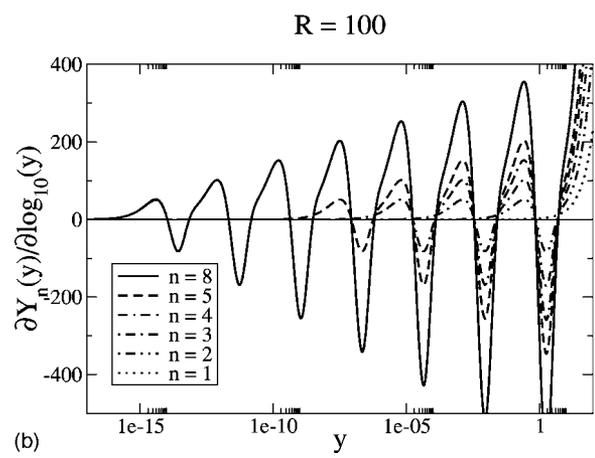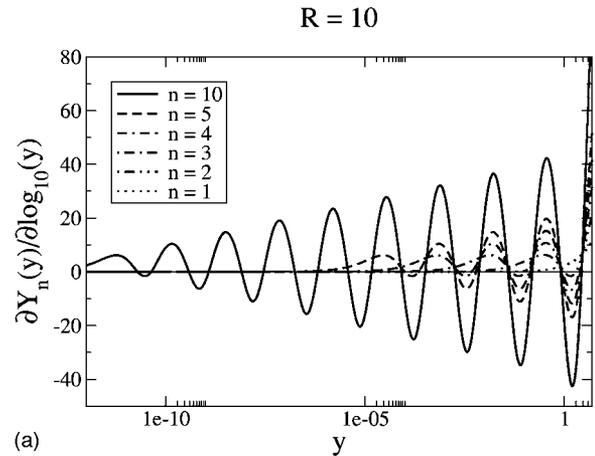
R = 10

(a)

R = 100

(b)

FIG. 7. Derivative of $Y_n(y)$ with respect to $\log_{10}(y)$ for different values of n and for $R=10$ (a) and $R=100$ (b), with logarithmic x axis. The stability region of a system of size N is the intersection of all the points that satisfy $\partial Y/\partial y > 0$ for $n=1,2,\ldots,N+1$. From the figures, we find that the stability region corresponds to the points that satisfy the condition for $n=N+1$, as the stability region of a system of size N seems to be embedded in the stability region of a system of size $N-1$ (see text for an explanation).

of the initial condition y, in logarithmic scale, for different values of the size of the system, N. We chose $\partial Y_n(y)/\partial \log_{10}(y)$ instead of $Y_n'(y)$ in order to obtain a smooth curve: The direct plot of $Y_n'(y)$ would make it very difficult to observe the intervals with $Y_n'(y) > 0$. As the sign of both derivatives is the same for all $y > 0$, we have resorted to the logarithmic one. With this change, a modulated "sinusoidal" structure reveals itself in Fig. 7 for each n, with $(n-1)$ maxima and minima around $Y_n'(y)=0$. From that figure, it is apparent that a new interval of instability for lower values of y appears in systems of size n as compared to systems of size $n-1$. In addition to this, Fig. 7 also suggests that the set of unstable points of a system of size n contains the set of unstable points of a system of size $n-1$. A plausibility argument for this statement goes as follows: Let us look at points that satisfy $Y_n'(y_0)=0$, in the extremes of an interval of unstable points. Then, if $Y_{n-1}'(y_0) > 0$, it must be $Y_{n+1}'(y_0) < 0$ [see Eq. (6)], and therefore the interval of unstable points for a system of size n will be larger than for a system of size $n-1$. This condition is satisfied by all the new unstable intervals that appear for each $Y_n(y)$, starting on $Y_2(y)$, and therefore, by induction, it can be applied to all

systems. Therefore, all stable points of a system of size n are those that satisfy $Y_{n+1}'(y) > 0$.

As an independent check of the validity of the results shown in this section, we compared our results with the ones recently reported in Ref. 17. By studying the discrete, stationary problem with fixed boundary conditions, they found eight stable and seven unstable solutions of a system of size $N=28$. The specific boundary conditions they used were $Y_0 = 0$ and $Y_N = 80$ for $R=10.1$. The exact $Y_N(y)$ and the approximate solution $Y_N^{(2)}(y)$ of that system are in Fig. 8. The plot now makes clear the precision problem we mentioned above, namely, when we tried to calculate the exact, numerical solution for low values of y. On the other hand, the approximate solution $Y_N^{(2)}(y)$ was calculated without any problem in a wide range of y. It is also shown that $Y_N^{(2)}(y)$ gives the same number of both stable and unstable solutions as in Ref. 17 (see explanation in caption), which implies that the structure of peaks of $Y_N(y)$ gives a good explanation of the number and structure of solutions. We think that this method can be applied for larger systems with the way to estimate errors that we explained in this section.
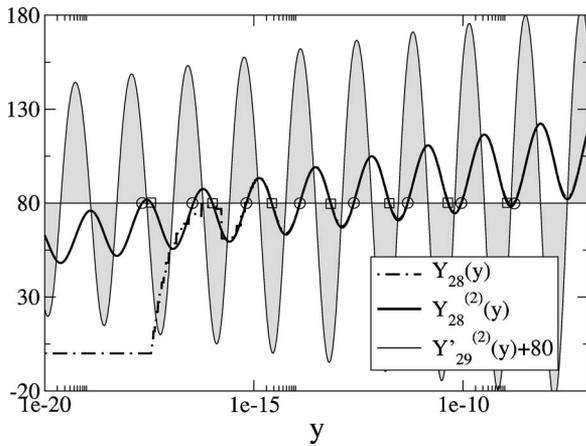
FIG. 8. Stable (circles) and unstable (squares) solutions given by approximation $Y_N^{(2)}$ for a system with $N=28$ sites, $Y_0=0$ and $Y_N=80$. The solutions are the intersections of $Y_{28}(y)$ (dot-dashed line) with the line $Y_{28}=80$ (horizontal line). Due to computer precision, approximation $Y_{28}^{(2)}(y)$ (solid thick line) is used instead. In order to establish the stability, it is necessary to study the sign of Y'_{29} (solid thin line) on each solution. Once again, $dY_{29}(y)/d\log_{10}(y)$ is plotted for clarity, and it is shifted upwards 80 units to make easier the study of the sign and, therefore, the stability of solutions. Stable solutions (circles) are the ones with positive value of Y'_{29}, whereas unstable solutions (squares) have negative Y'_{29}. Our approximation gives eight stable and seven unstable solutions, exactly as in Ref. 17.

V. CONCLUSIONS

In this paper, we have reported a study of the stationary solutions of the PB model, obtaining exact and analytical approximations of the continuum and the discrete limit. We have been able to obtain all the stationary solutions and to classify them according to their stability by considering the stationary equation as an initial value problem. We have also found that, in the discrete limit, the exact solutions can be approximated to the desired degree of accuracy by using the functions $f_k(y)$ as explained above. We have compared our results obtained in the discrete limit with Ref. 17, finding very good agreement with the number of stable and unstable solutions of a PB system with fixed boundary conditions, thus giving an explanation of the multiple solutions of the problem and the stability. In fact, our results show that every solution of the initial value problem, which is unique for every choice of y, corresponds to exactly one of the problem with fixed boundary conditions,[17] which does not have a unique solution. This is the reason the picture we are providing here is much more comprehensive, allowing us to understand fully the space of solutions of the problem. On the other hand, the method explained in this paper to obtain stable solutions of a system of size N and opening L allows to work with larger systems, as solutions and their stability are calculated by evaluating a function, instead of numerically (as in Ref. 17, with $N=28$). We also believe that this study may be extended to the more accurate description of DNA given by the Peyrard-Bishop-Dauxois model,[6] where the coupling between two consecutive bases of the DNA molecule has an anharmonic term that affects the general behavior of the openings.[8,18]

As stated in Sec. I, this stems from previous studies in the sine-Gordon (Englander) model of DNA,[11,12] where the

relation between the dynamics of soliton-like excitations and the inhomogeneity of the DNA sequence was studied. In fact, what we are reporting here is only the first step towards the study of an effective potential that may explain the dynamics of these stationary, stable solutions in presence of heterogeneities in the sequence and an external force. In this regard, it is important to realize that these heterogeneities change the Hamiltonian (1), introducing two new site-dependent parameters that would affect the shape of the Morse potential $V(Y_n)$ by changing the width and the depth on each site between two possible configurations. Therefore, the mapping we have considered in this work is not valid any longer. However, much as we have done in previous work about the sine-Gordon model,[11,12] we would proceed as follows: Once an analytical expression of the stationary solutions of the *homogeneous* model is found, as we have just done, we will use it as an *Ansatz* in a collective coordinate technique to find an *effective potential* description of the dynamics. The final aim of such a program is to find out whether this approach allows to identify important sites from the genomic viewpoint along any given sequence. While work along these lines is in progress, we believe that the richness of the structure of the stationary solutions we have found and their stability is of interest in itself and can motivate further research in these and related models. Finally, we believe our solutions can be exploited to analyze the statistical mechanics of the PB model along the lines of Refs. 17 and 19, in particular because of the advantage of having an approximate, analytical expression.

[1]E. Fermi, J. R. Pasta, and S. Ulam, Los Alamos Report LA-UR-1940 (1955); reprinted in *Collected Papers of Enrico Fermi*, edited by E. Segré (University of Chicago Press, Chicago, 1965).
[2]A. C. Scott, *Nonlinear Science* (Oxford University Press, Oxford, 1999).
[3]S. W. Englander, N. R. Kallenbach, A. J. Heeger, J. A. Krumhansl, and A. Litwin, Proc. Natl. Acad. Sci. U.S.A. **77**, 7222 (1980).
[4]L. V. Yakushevich, *Nonlinear Physics of DNA*, 2nd ed. (Wiley-VCH, Weinheim, 2004).
[5]M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[6]T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993); T. Dauxois and M. Peyrard, *ibid.* **51**, 4027 (1995).
[7]A good summary of the experimental advances can be found in the collection of articles in Nature (London) **421**, 396 (2003).
[8]G. Kalosakas, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi, and A. Usheva, Europhys. Lett. **68**, 127 (2004); C. H. Choi, G. Kalosakas, K. Ø. Rasmussen, M. Hiromura, A. R. Bishop, and A. Usheva, Nucleic Acids Res. **32**, 1584 (2004).
[9]T. S. van Erp, S. Cuesta-López, J.-G. Hagmann, and M. Peyrard, Phys. Rev. Lett. **95**, 218104 (2005).
[10]S. Ares, N. K. Voulgarakis, K. Ø. Rasmussen, and A. R. Bishop, Phys. Rev. Lett. **94**, 035504 (2005).
[11]S. Cuenda and A. Sánchez, Fluct. Noise Lett. **4**, L491 (2004).

[12] S. Cuenda and A. Sánchez, Phys. Rev. E **70**, 051903 (2004).
[13] M. Peyrard, Nonlinearity **17**, R1 (2004).
[14] M. Salerno, Phys. Rev. A **44**, 5292 (1991).
[15] M. Salerno, Phys. Lett. A **167**, 49 (1992).
[16] M. Salerno and Yu. S. Kivshar, Phys. Lett. A **193**, 263 (1994).
[17] N. Theodorakopoulos, M. Peyrard, and R. S. MacKay, Phys. Rev. Lett. **93**, 258101 (2004).
[18] A. Campa and A. Giansanti, Phys. Rev. E **58**, 3585 (1998).
[19] H. Qasmi, J. Barr, and T. Dauxois, preprint arXiv: cond-mat/0407662 (2004).