Fluctuation and Noise Letters
Vol. 4, No. 3 (2004) L491–L504
© World Scientific Publishing Company

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

DISORDER AND FLUCTUATIONS IN NONLINEAR EXCITATIONS IN DNA

SARA CUENDA AND ANGEL SÁNCHEZ*

Grupo Interdisciplinar de Sistemas Complejos (GISC) and Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain
**Instituto de Biocomputación y Física de Sistemas Complejos (BIFI) Universidad de Zaragoza, 50009 Zaragoza, Spain*

Received 31 May 2004
Revised 16 July 2004
Accepted 20 July 2004
Communicated by Derek Abbott

We study the effects of the sequence on the propagation of nonlinear excitations in simple models of DNA, and how those effects are modified by noise. Starting from previous results on soliton dynamics on lattices defined by aperiodic potentials [23], we analyze the behavior of lattices built from real DNA sequences obtained from human genome data. We confirm the existence of threshold forces, already found in Fibonacci sequences, and of stop positions highly dependent on the specific sequence. Another relevant conclusion is that the effective potential, a collective coordinate formalism introduced by Salerno and Kivshar [21] is a useful tool to identify key regions that control the behaviour of a larger sequence. We then study how the fluctuations can assist the propagation process by helping the excitations to escape the stop positions. Our conclusions point out to improvements of the model which look promising to describe mechanical denaturation of DNA. Finally, we also consider how randomly distributed energy focus on the chain as a function of the sequence.

Keywords: DNA; mechanical denaturation; unzipping; solitons; noise; genome; correlated disorder; localization

1. Introduction

After a quiet, 30 year revolution in the way we look at natural phenomena, nonlinear science is nowadays a well established, respected body of knowledge with ubiquitous applications [1]. Restricted at first to the realm of physics, nonlinear models are becoming a versatile tool in other, unrelated fields. Such is the case of biology, where the nonlinear viewpoint was successfully introduced more than 20 years ago [2] (see also Ref. [3] for references). Ever since the early years, one of the subjects where the use of nonlinear models has been more productive is the modelling of DNA physics [4–7]. Indeed, the first paper on this issue dates back to 1980, when Englander and coworkers published a report entitled "Nature of the open state in long

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

polynucleotide double helices: possibility of soliton excitations" [9]. Subsequently, a large number of researchers contributed to this field, which has by now reached maturity; the reader is referred to Ref. [6] for a historical account and an extensive summary of the available results.

Models such as that proposed by Englander and coworkers (hereafter, the Englander model) [9] as well as other simple models proposed by Peyrard and Bishop [10] (see also the generalizations and improvements proposed in Refs. [11, 12]) or Causo *et al.* [13] have demonstrated that, in spite of their simplicity, they can accurately (sometimes even quantitatively) capture the phenomenology experimentally observed in DNA (see, e.g. Ref. [14]). Most of the research done in the framework of these models refers to homopolymers, i.e. homogeneous DNA molecules. In additions, a great deal of the experimental tests were carried out from a more thermodynamical or statistical mechanics viewpoint, focusing, e.g. on the characterization of the thermal denaturation transition (the phase transition that takes place at temperatures around 70°C when the two strands of the DNA molecule separate). However, the recent theoretical advances in nonlinear theory, related to self-localization in nonlinear networks, [1] along with the capability to carry out experiments on single molecules (to be discussed below) achieved in the last few years [15] have impulsed a shift on the emphasis of the modelling to the dynamics of single molecules. On the other hand, the decoding of the genome has also led to new questions, and, in particular, to the necessity for a better understanding the relationship between sequence, physical properties, and biological function.

In this work, we aim to contributing a step in the above described direction by revisiting the Englander model. Our main goal is to include in the system the effects of the sequence heterogeneity in order to assess whether the model contains the necessary ingredients to reproduce the behavior of real DNA molecules. In addition, we want to keep in mind that actual experiments (or biological functioning) will unavoidable take place in the presence of noise (at least from thermal origin) and therefore another of our objectives here will be to verify the robustness of the observations in the model when noise is considered. In the following sections, we review the previous research on which our work is based on, then present our results for the deterministic case, to close finally with the report on the stochastic model.

2. Background

2.1. *Model definition*

The model proposed by Englander *et al.* [9] is schematically represented in Fig. 1, and is nothing but a chain of damped pendula, i.e. a discrete, damped sine-Gordon model, given by

$$\ddot{u}_n - \frac{1}{a^2}(u_{n+1} - 2u_n + u_{n-1}) + V_n \sin u_n + \alpha \dot{u}_n + F = 0, \tag{1}$$

where a is the lattice spacing, V_n is a site dependent constant which arises from the specific parameters for the pendulum at site n, α is the damping coefficient, and F is a driving term possibly acting on the chain. When $V_n = V$ (i.e. we have a homopolymer) and the lattice spacing is very small, the system of ordinary differential equations (1) can be very well approximated by its continuum limit.

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

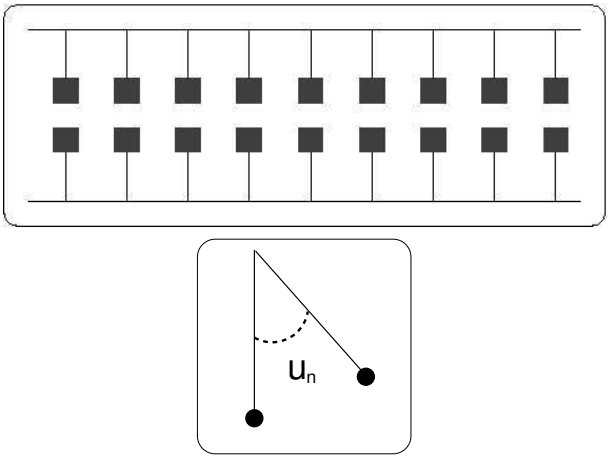

Fig. 1. Sketch of the model. The two strands of the DNA are depicted as two straight lines connected by the base pairs (bases are the squares). The lower strand is assumed to be fixed and as showed on the right, the angle u_n is the deviation of the upper base of the n-th pair with respect to the lower one.

Letting $u_n(t) \longrightarrow u(x,t)$ when $a^2 \ll V$, we find the driven, damped sine-Gordon equation:

$$\partial_t^2 u - \partial_x^2 u + V \sin u + \alpha \partial_t u + F = 0. \tag{2}$$

It is well known [1] that in the absence of dissipation and force ($\alpha = F = 0$), Eq. (2) possesses soliton solutions of the kink type, whose expression is

$$\phi_\pm(x,t) = 4 \arctan \left\{ \exp \left[\pm \sqrt{V} \left(\frac{x - x_0 - vt}{\sqrt{1-v^2}} \right) \right] \right\} \tag{3}$$

where the plus or minus sign stands for kinks or antikinks respectively. If dissipation is included, the soliton is asymptotically stopped, whereas the force changes the minima of the onsite potential and correspondingly the values of the kink "wings". If both are present, as we will see below, the soliton acquires a limit velocity as a result of the balance of both effects. Figure 2 sketches the physical meaning of a kink solution in the context of DNA modelling.

A comment is in order regarding the physical origin of the force term F. As stated in the Introduction, the experimental capabilities have been improved in the last ten years to the point that it is possible to manipulate single DNA molecules with high accuracy. In particular, in the so-called mechanical unzipping experiments, one of the DNA strands is attached to a glass bead, which is pulled by a glass micro-needle, while the other strand is attached to a glass plate, which serves as fixed reference point. The glass bead is then pulled at constant velocity, opening or unzipping the double chain by consecutively breaking the hydrogen bonds of the base pairs, and the force used to pull it is recorded. The first experiments along this line were carried out in 1997 by Essevaz-Roulet and cowokers [16] and largely improved in accuracy in 2002 [17]. Other experiments have been carried out; the interested reader may find a review by Bustamante, leader of one of the most important groups in the field, in [15]. The use of a model as ours is in fact very appropriate

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

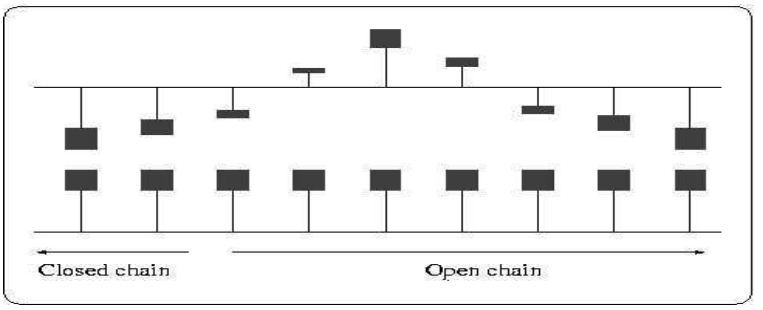

Fig. 2. Kink soliton in the sine-Gordon model. The kink joins a sector of the chain where bases are closed $u_n = 0$ to another one where bases have performed a complete turn $u_n = 2\pi$. In this last part the chain is said to be open if the model is to represent mechanical denaturation.

since an analysis of the experimental results shows that such an experiment can not distinguish individual base pairs; instead, it observes the breaking of groups of a few tens of base pairs, which is correctly represented by the displacement of one kink [8]. In this context, the force F in Eq. (2) represents the action of the pulling on the end of one strand on the region where unzipping is taking place. We will see in the Conclusion that our model indeed reproduces qualitatively the experiments.

2.2. *Previous results*

Following the pioneering report of Ref. [9], an important extension and application was proposed and studied by Salerno [18]. He was the first one to consider the effects of the sequence by studying spontaneously travelling kinks in $T7A_1$ promoter (a region of the genome preceding a gene where the transcription activity starts) of the bacteriophage T7. Salerno studied further other regions of the bacteriophage [19], finding that the dynamical activity of kinks should be very special in the promoters. Recently, the problem was revisited in Ref. [20] following the sequencing of the whole genome of the bacteriophage T7. The main conclusion of this work was that there indeed is a significantly higher degree of activity in promoter regions, even in the presence of noise.

The results of the simulations reported in Refs. [18, 19] were analyzed theoretically by Salerno and Kivshar [21], who, using a collective coordinate approach (see, e.g. Ref. [22] for a review on collective coordinate techniques for soliton-bearing equations), developed a description of the kink dynamics in terms of an effective potential. We will not review in detail the procedure to obtain this effective potential but, instead, we will simply outline the main steps. The idea is to consider the undamped model, which can be derived from the Hamiltonian

$$H^{sG}[\{u\}] = \sum_{n=1}^{N} \{\frac{1}{2}\dot{u}_n^2 + \frac{1}{2a^2}[u_{n+1} - u_n]^2 + V_n(1 - \cos u_n) + F u_n\}. \tag{4}$$

We now insert the following *Ansatz*

$$\phi_n(X(t)) = 4 \arctan \left\{ \exp \left[\sqrt{V_{avg}}\,(na - X(t)) \right] \right\}, \tag{5}$$

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

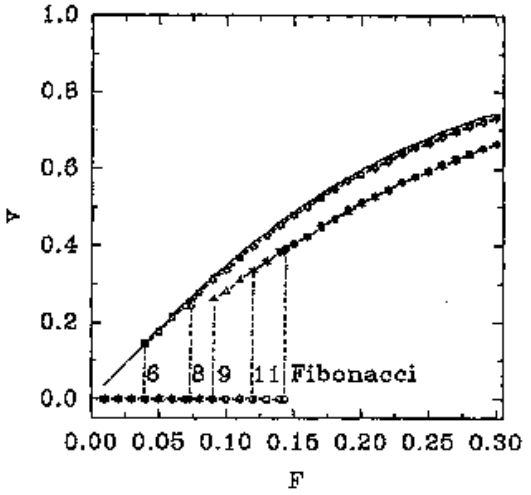

Fig. 3. Steady state velocity versus applied force for chains with different orderings. Points correspond to numerical results, numbers indicate the Fibonacci generation constituting the unit cell of the chain. Reprinted from Ref. [23]; see the original paper for more details.

where V_{avg} is the average value of V_n over the chain, in the Hamiltonian (4). We thus arrive at

$$E_{eff} = \frac{1}{2}\dot{X}^2 + V_{eff}(n, \{V_n\}), \tag{6}$$

which is formally equivalent to the energy of a particle subjected to the action of the effective potential V_n. The exact formula for this potential is quite cumbersome and we do not need to reproduce it here. We refer the reader interested in the details of the calculation and the full result to Refs. [21, 23, 24].

2.3. *Aperiodic chains*

In this situation, in 1995, when the genome was not yet available, F. Domínguez-Adame and coworkers [23] propose to mimic the behavior of DNA heteropolymers of biological relevance by replacing the sequence dependent values V_n by an aperiodic, but fully deterministic sequence. As examples, they considered mostly the Fibonacci sequence, that is generated according to the rules $A \to AB$, $B \to A$ repeatedly applied to the initial seed A. The results obtained in that work (subsequently confirmed in Ref. [24]) pointed out the existence of an intriguing phenomenon, namely the existence of a threshold force for a kink to start moving along the chain. Once the force is above the threshold, the kink moves with an asymptotically constant given by the balance of damping and driving. An analytical expression can be easily derived to predict the asymptotic velocity from energy conservation arguments (see Ref. [25]). Although the original argument is valid for a homogeneous chain, repeating it here with the average value V_{avg} leads to a prediction which is very accurate for periodically ordered chains while overestimating (by approximately a 10%) the velocity for the Fibonacci chains. Even more interestingly, it was also found (see Fig. 3) that the threshold value depended on the length of the quasiperi-

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

odic chain considered. This was seen by forming periodic chains with unit cell F_n, F_n being the Fibonacci chain in iteration n. As shown in Fig. 3, increasing the size of the unit cell leads to an increasing of the threshold, reaching a limiting value for the whole Fibonacci chain. This behavior was well described with the help of the effective potential approach summarized above.

Based on this results and on the fact that, for the same parameters, random chains exhibited much larger threshold values (around $F \simeq 0.5$), Domínguez-Adame *et al.* conclude that long range order effects give rise to measurable consequences on kink dynamics in aperiodic chains. Further, they suggested that the fact that there are long range correlations in DNA and, in any event, that it contains information and is not purely random, could lead to similar phenomenology in the propagation of nonlinear coherent excitations along the molecule. At this point is where we take that previous research, with the body of genomic data available nowadays.

3. Effects of the Sequence

In order to check whether or not the phenomenology we have summarized in the previous section carries over to real DNA data, we have simulated Eq. (1) with V_n chosen according to sequences obtained from the human genome. We want to stress that this description of DNA, arising from the original model of Englander and coworkers [9], intends to be only qualitatively correct. Therefore, the parameters can be freely chosen, trying, of course, to mimic the real ones. Therefore, what we do is choose $V_n = 2, 3$ according to what we have at site n is an A-T pair, linked by two hydrogen bonds, or a C-G pair, with three hydrogen bonds. The other parameters are chosen as in Ref. [23], namely lattice spacing $a = 0.1$ and damping $\alpha = 0.1$. With respect to the lattice spacing, we want to point out that the chosen value leads to a width of the kink which is comparable to that of spontaneous openings of real DNA chains. In what follows, we use data obtained from the National Center for Biotechnology Information for the human genome (`http://www.ncbi.nlm.nih.gov`).

We have carried out simulations on many different sequences of different chromosomes, choosing both coding and non-coding regions. Typical results from our simulations are plotted in Figs. 4 and 5; the outcome from all the regions analyzed is always qualitatively the same. Figure 4 shows the kink dynamics on a real dna sequence when the kink is forced with two different drivings. For reference, we plot in the same graphs the effective potential obtained as described in the previous section. It is important to note that the effective potential does not contain the contribution of the force F, and therefore its most relevant information is the position of the peaks and valleys. With this in mind, we see that the dynamics is basically as in the case of the Fibonacci chain. There is a threshold for the kink to propagate along the whole chain and, for forces below threshold, it ends up being trapped at some potential well.

Figure 5 collects our observations regarding the existence of thresholds in two examples: a coding region and a non-coding region. We again see the existence of threshold forces, in agreement with a description in terms of an effective potential. However, we do not observe any qualitative or otherwise relevant difference between the kink dynamics in the two regions. This is the case with all the regions we have

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

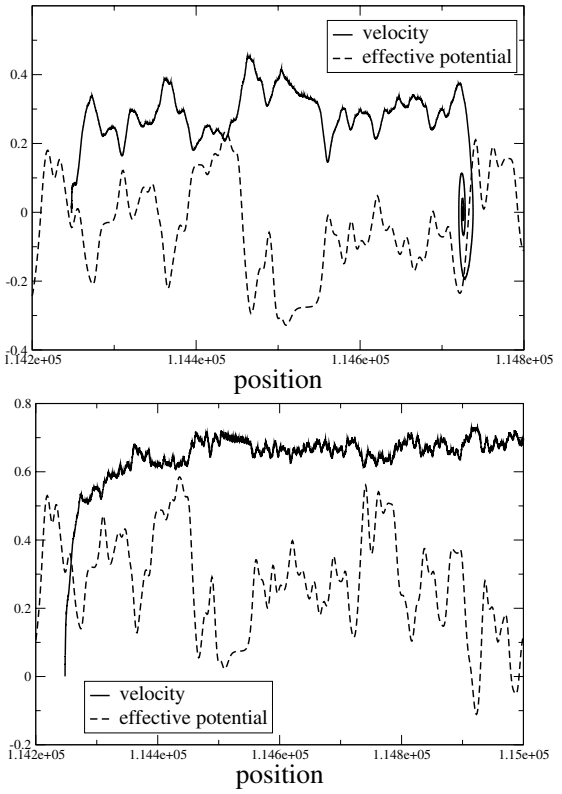

Fig. 4. Simulations of the kink soliton in the sine-Gordon model with genome data. Shown are the velocity and the (properly scaled to fit in the plot) effective potential vs the position along the chain. Left: $F = 0.06$; the kink travels along the sequence but stops at an effective potential well. Right: $F = 0.07$; the kink travels along the whole chain. The DNA sequence corresponds to contig NT_028395.2 of human cromosome 22, between positions 114 100 and 115 100, part of a gen.

analyzed. Therefore, the hint from the Fibonacci results that information may lead to different kink dynamical properties is, at least at the level of our simple model, not in agreement with the simulation results.

In order to assess further how good is the comparison between the previous work on the Fibonacci chain and the present results, we have simulated periodic systems with unit cell built from pieces of a genetic sequence, repeated to complete a longer chain. The sizes for the unit cells were chosen to mimic the sizes for the Fibonacci iterations. The results are collected in Fig. 6. From this plot, we see that the most important difference between the two cases is that in the actual DNA chain, the threshold value does not depend monotonously on the size of the unit cell (note that the threshold for the periodic chain with unit cell of size 55 is the smallest one). It is still true, however, that the threshold is largest for the whole sequence. Therefore, we are led to conclude that the information effect on the kink propagation along DNA chains is certainly different of what we expected in terms of our simple Fibonacci approach. Notwithstanding, we want to stress that the

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

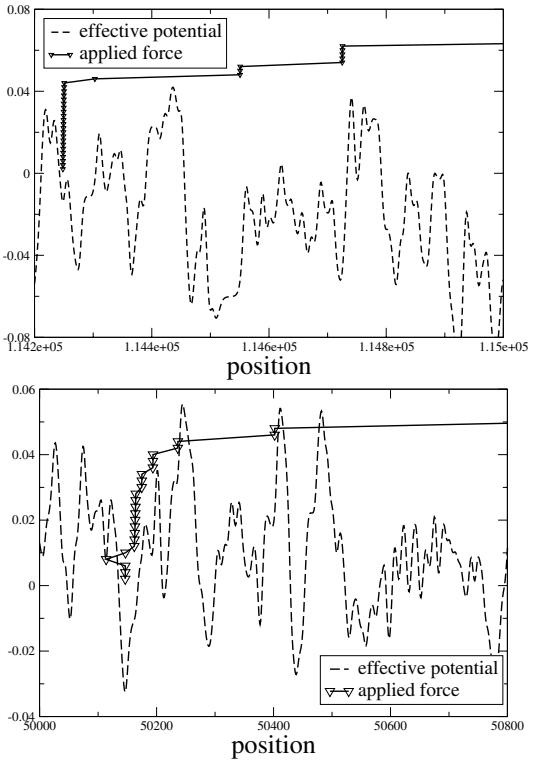

Fig. 5. Simulations of the kink soliton in the sine-Gordon model with genome data. As shown are the force required to reach a certain position and the (properly scaled to fit in the plot) effective potential vs the position along the chain. Thresholds correspond to the largest force for which there are points in the plot. Comparison of propagation along a coding region (left, same contig as in Fig. 4, positions 114 000 to 115 000) and a non-coding region (right, same contig as in Fig. 4, positions 50 000 to 51 000).

analytical approach in terms of collective coordinates and an effective potential is a very good picture of the observed phenomenology even in this, more realistic setting. On the basis of this scenario, we suggest that the reason for the difference between the two models lies in the larger diversity available for possible DNA chains. The fact that sites of type B in Fibonacci chains are always isolated is very restrictive as to the shapes and sizes of the peaks of wells of the potential felt by the kinks, and therefore the Fibonacci model cannot capture the richness of the DNA model. As for the possible different behaviors between coding and non-coding regions, we conclude that the Englander model does not reflect the information content of the chain in the kink dynamics, which in principle prevents the use of this simple model as a genome sequencing tool.

4. Effects of Noise

The results discussed so far correspond to the purely deterministic case, i.e. they are related only to the "disorder" (understood as inhomogeneity, even if it contains

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

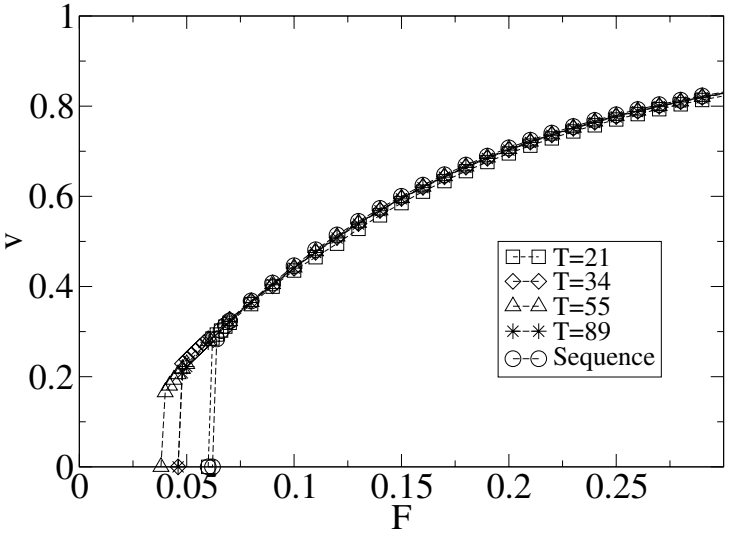

Fig. 6. Velocity vs applied force for different periodic systems formed by repeating as a unit cell n sites of the same contig as in Fig. 4 beginning at site 114 241, with $n = 21$, 34, 55 and 89 (as in the Fibonacci case) and for the whole sequence going from 114 100 to 115 100.

information) present in the DNA chains. However, a fully deterministic approach is not really relevant to the DNA properties in so far as life takes place at nonzero temperature, and the same applies to single molecule experiments, usually carried out at room temperature. It goes without saying that there are many aspects of the influence of noise on the dynamics of the nonlinear excitations in DNA, and that our purpose is simply to present the first results of two lines of research we are beginning to explore in this context. Below we discuss the robustness of the deterministic results presented in the preceding section when thermal noise is considered, and how nonlinear coherent excitations arise from unstructured, noisy distortions of the DNA chain.

4.1. *Thermal noise*

Our first aim is to check the validity of the general description of the kink dynamics in terms of the effective potential in the presence of thermal noise. The reason for this is that, unless the scenario holds at temperatures of the order of room temperature, the insights provided by the effective potential about special places in the chain will not be relevant for the actual behavior of DNA, either *in vivo* or in single molecule experiments. To this end, we considered the Langevin version of Eq. (2), namely

$$\partial_t^2 u - \partial_x^2 u + V \sin u + \alpha \partial_t u + F + \xi(x,t) = 0, \tag{7}$$

where $\xi(x,t)$ is a gaussian white noise, of zero mean and variance

$$\langle \xi(x,t)\xi(x',t')\rangle = 2D\alpha\delta(x-x')\delta(t-t'), \tag{8}$$

verifying the fluctuation-dissipation theorem.

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

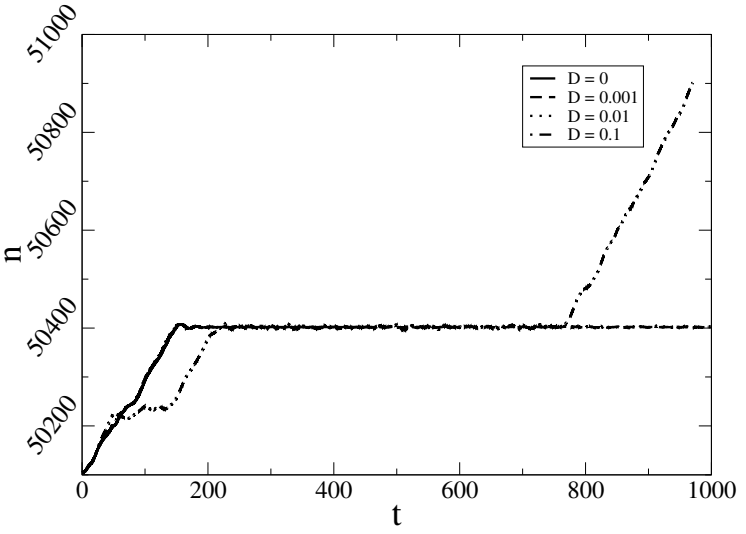

Fig. 7. Kink soliton propagation in the presence of noise. The sequence is as in Fig. 4. The applied force is $F = 0.048$ (deterministic threshold for this sequence is $F = 0.05$). Noise intensities D are as indicated in the plot.

The results obtained in the presence of noise are illustrated by the example in Fig. 7. The plot represents the trajectory of the center of a kink for different values of the noise intensity D. The applied force F is below threshold, so for $D = 0$ we observe that the kink is trapped at a position around the base pair 50400. As we may see, it is necessary to add a large amount of noise (note that the value of D is twice the value of the threshold force) to allow the kink to escape that well, and even then this takes place after a rather long residence time. Other runs yield similar results. Therefore, our preliminary conclusion is that the effective potential information is relevant even in the presence of moderate or large noises. Pending a more detailed comparison with *in vivo* temperatures, we feel that the available data is enough to ascertain that the model will behave qualitatively as single molecule, mechanical denaturation experiments carried out at room temperature.

4.2. *Formation of nonlinear excitations and transcription initiation*

In this subsection, we deal with the problem of the formation of nonlinear excitations in the model. This is an important issue in view of the fact that we are studying kink propagation using a kink as an initial condition, but we have not justified whether or not these kinks can build up from a generic excitation of the chain. For this study, we have considered one of the sequences studied in Ref. [26]. Specifically, we have taken the sequence of adenoassociated virus 1 (AAV1), consisting of 4718 base pairs. The sequence contains the adenoassociated viral P5 promoter studied in Ref. [26]. Kalosakas *et al.* performed Langevin simulations of a very short chain containing only the promoter and found that the chain exhibited preferential opening precisely at the so-called transcription start site. From their data, they concluded that DNA dynamically directs its own transcription. By looking at the same sequence data

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

these authors have analyzed, we simultaneously address the issue of the origin of nonlinear excitations and the confirmation of the results of Ref. [26].

In our simulations, we have again considered the deterministic equation (2), which we now integrate starting with random initial data given by $u_n = u_0 + v_n$, where v_n are independently generated gaussian random numbers. We have set the applied force F to zero in order to check whether breathers and kink can be formed in the absence of directed external pulling, only from randomly distributed energy. Dissipation is also zero in these simulations because otherwise all breathers would be suppressed in the long term (while they survive in the presence of thermal noise, see [26]). We have taken free boundary conditions, which implies that energy is not conserved. The reason for this choice is twofold: On the one hand, if periodic boundary conditions are used, a large amount of radiation is all the time present in the system, forming a background that masks the static breathers, formed and pinned by the sequence (this is not the case with kinks of course, but in many simulations we obtain only breathers). Free boundary conditions let this radiation go away from the system allowing to identify the relevant pinned breathers. On the other hand, in experiments with single DNA molecules the energy in a single molecule is not conserved, as the chain deformations may be resolved through the free ends of the molecules. Therefore, by using free boundary conditions we are mimicking more closely the experimental situation. We note that in doing so we are not in the microcanonical ensemble of statistical mechanics; this is not a problem in so far as we are only interested in statistics as a way to quantify how often breathers or kinks form at specific places. Finally, we mention that our chain consists of the base pairs from position 200 to 400 (the P5 promoter goes from position 256 to 324, and the transcription start site is at position 296).

Figure 8 presents an example of the time evolution of random initial data. Although the plots correspond to a specific set of parameters, the different outcomes represent those of typical simulations. The case represented here illustrates the effect of changing the mean value of the initial data, u_0. In general, what we observe is that the initial data may or may not give rise to localized, coherent excitations, and that these can be breather-like or kinks (these last ones appear only if the value of u_0 is large, typically close to π). Our statistics so far is not large enough to give reliable values for the relative frequencies of each event, but we have seen that none of these outcomes is rare, i.e. all of them are statistically relevant. Hence, these simulations demonstrate that it is indeed possible to have nonlinear coherent excitations in DNA chains, formed spontaneously from unstructured initial data.

As regards the comparison with Ref. [26], we have found that when breathers are formed, they preferently locate in the transcription start site or its neighborhood. Actually, in all the simulations we have carried out, we observe that breathers only build up in three different regions of the studied sequence. This agrees with the findings reported in Ref. [26], although we need to do a much larger statistical analysis. On the other hand, preliminary analytical results show that the collective coordinate approach leading to the effective potential for kinks can be (approximately) extended to breathers. This extension, which will be reported elsewhere, predicts with good accuracy the existence of three potential wells roughly at the places where we observe the formation of breathers in the simulations. If this is indeed the case, the process would develop in two stages: first, breathers are formed

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

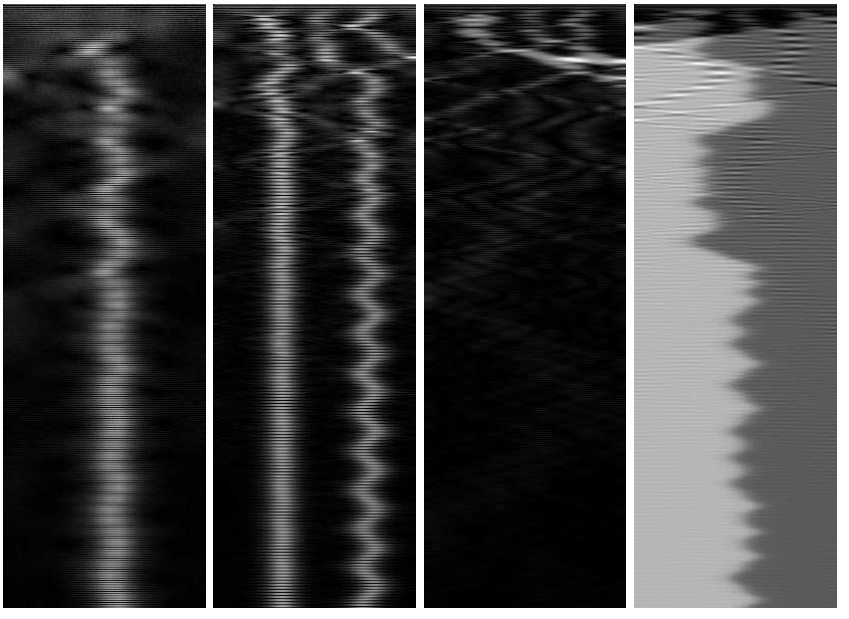

Fig. 8. Focusing of random initial data. Grey levels code the absolute value of u along the chain horizontally, and time proceeds downwards. Left to right, the initial data is $u_0 = 0.9$, 2.0, 2.5, and 3.0. Superimposed noise variance is 0.1. Damping is set to zero. Sequence corresponds to AAV1 from position 200 to 400 (see text).

by focusing of the initial data, and second, these breathers experience the effective potential and end up located at one of its wells. While this is an appealing picture, much more analytical work is needed to confirm it, and this preliminary report should still be taken with some degree of caution.

5. Conclusions

In this paper, we have reported the first conclusions of our work on the effects of the genetic sequence on the propagation of nonlinear excitations along (simple models of) DNA chains. The research was motivated by previous results on chains given by aperiodic sequences [23], which lead to the hypothesis that the information content of the inhomogeneities might be relevant for the dynamical behavior. While this general statement cannot be ruled out at this stage, we believe that at the level of the Englander model [9] the simulations presented here do not confirm that hypothesis. Specifically, we have not observed qualitative difference in the dynamics of kinks propagating along coding or non-coding regions. In this respect, it is important to realize that the function and information content of non-coding regions are far from well know, and it may well be that the dynamics in the two types of sequences is the same because the quantity of information is similar. This is certainly an interesting issue that deserves further attention. On the positive side, we have seen that a model as simple as ours is able to reproduce the main stylized features observed in mechanical unzipping experiments [8, 15–17], such as the existence of a threshold to begin the unzipping process and the irregular time behavior of the force, which

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

closely resembles stick-slip dynamics. This agreement suggests that, in modelling DNA, the most relevant factor is the correct inclusion of the "disorder" given by the sequence rather than the mechano-chemical details of the interactions.

Another conclusion that we deduce from this research is the validity of the effective potential approach [21, 23, 24] to the dynamics of nonlinear excitations in DNA. Although the precise dynamics of kinks is difficult to predict, this analytical approximation allows to identify the possible stop positions as well as the barriers that control the total opening (mechanical denaturation) of the chain. Furthermore, the scenario that arises from considering the effective potential a good description of the kink dynamics leads us to speculate that the results of mechanical denaturation experiments on single molecules, with their stick-slip-like time evolution, could be also understood within simple models like the one studied here. This impression is reinforced by the fact that the effective potential approach works equally well in the presence of (thermal) noise, and by our preliminary report of its validity for breathers. Work along these lines is in progress.

Finally, a few words are in order regarding our findings about the preferential opening of DNA chains. We have qualitatively confirmed the results reported in Ref. [26] with a much more elaborated and realistic model [10–12]. We want to stress that our study is not a Langevin simulation and is not even a microcanonical simulation because of the free boundary conditions. With those caveats in mind, we can conclude that in the sequence we have studied there are three preferential places for breather (opening) formation, one of the most visited ones being the neighborhood of the transcription start site. In this respect, we want to point out that those locations may also be predicted by the effective potential approach. If that is indeed the case, it could well be that the analyses of longer chains would also reveal other preferential places whose significance should then be understood. In any event, we stress that our conclusions on this subject are very preliminary and that we believe that much more research is needed before we can reach definitive conclusions.

Acknowledgments

We thank Michel Peyrard for many discussions about this work and for his patient explanations on DNA dynamics. We also thank Alan R. Bishop for making the results in Ref. [26] available prior to publication. This work has been supported by the Ministerio de Ciencia y Tecnología of Spain through grant BFM2003-07749-C05-01. S.C. is supported by a fellowship from the Consejería de Educación de la Comunidad Autónoma de Madrid and the Fondo Social Europeo.

References

[1] A. C. Scott, *Nonlinear Science*, Oxford University, Oxford (1999).

[2] A. S. Davydov, *Phys. Scripta* **20** (1979) 387.

[3] A. S. Davydov, *Solitons in Molecular Systems*, 2nd edition, Kluwer, Dordrecht (1991).

[4] L. V. Yakushevich, *Q. Rev. Biophys.* **26** (1993) 201.

[5] G. Gaeta, C. Reiss, M. Peyrard and T. Dauxois, *Riv. Nuovo Cimento* **17** (1994) 1.

[6] L. V. Yakushevich, *Nonlinear Physics of DNA*, Wiley, Chichester (1998).

[7] G. Gaeta, *J. Biol. Phys.* **24** (1999) 81.

Fluct. Noise Lett. 2004.04:L491-L504. Downloaded from www.worldscientific.com
by UNIVERSIDAD CARLOS III DE MADRID LIBRARY on 04/07/20. Re-use and distribution is strictly not permitted, except for Open Access articles.

[8] M. Peyrard, *Nonlinearity* **17** (2004) R1.

[9] S. W. Englander, N. R. Kallenbach, A. J. Heeger, J. A. Krumhansl and A. Litwin, *Proc. Natl. Acad. Sci. USA* **77** (1980) 7222.

[10] M. Peyrard and A. R. Bishop, *Phys. Rev. Lett.* **62** (1989) 2755.

[11] T. Dauxois, M. Peyrard and A. R. Bishop, *Phys. Rev. E* **47** (1993) R44.

[12] T. Dauxois and M. Peyrard, *Phys. Rev. E* **51** (1995) 4027.

[13] M. S. Causo, B. Coluzzi and P. Grassberger, *Phys. Rev. E* **62** (2000) 3958 .

[14] A. Campa and A. Giansanti, *Phys. Rev. E* **58** (1998) 3585.

[15] A good summary of the experimental advances can be found in the collection of articles *The Double Helix — 50 Years*, *Nature* **421** (2003) 396.

[16] B. Essevaz-Roulet, U. Bockelmann and F. Heslot, *Proc. Natl. Acad. Sci. USA* **94** (1997) 11935.

[17] U. Bockelmann, Ph. Thomen, B. Essevaz-Roulet, V. Viasnoff and F. Heslot, *Biophys. J.* **82** (2002) 1537.

[18] M. Salerno, *Phys. Rev. A* **44** (1991) 5295.

[19] M. Salerno, *Phys. Lett. A* **167** (1992) 49.

[20] E. Lennholm and M. Hörnquist, *Physica D* **177** (2003) 233.

[21] M. Salerno and Yu. S. Kivshar, *Phys. Lett. A* **193** (1994) 263.

[22] A. Sánchez and A. R. Bishop, *SIAM Rev.* **40** (1998) 579.

[23] F. Domínguez-Adame, A. Sánchez and Yu. S. Kivshar, *Phys. Rev. E* **52** (1995) R2183.

[24] E. Lennholm and M. Hörnquist, *Phys. Rev. E* **59** (1999) 381.

[25] D. W. McLaughlin and A. C. Scott, *Phys. Rev. A* **18** (1978) 1652.

[26] G. Kalosakas, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi and A. Usheva, `arXiv:cond-mat/0309157` (2003); C. H. Choi, G. Kalosakas, K. Ø. Rasmussen, M. Hiromura, A. R. Bishop and A. Usheva, *Nuc. Acids Res.* **32** (2004) 1584.