Available online at www.sciencedirect.com
ScienceDirect
Physica D 223 (2006) 214–221
PHYSICA D
www.elsevier.com/locate/physd

Does the dynamics of sine–Gordon solitons predict active regions of DNA?

Sara Cuenda[a,*], Angel Sánchez[a,b], Niurka R. Quintero[c]

[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*
[b] *Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, 50009 Zaragoza, Spain*
[c] *Universidad de Sevilla, Departamento de Física Aplicada I, E.U.P., Virgen de África 7, 41011 Sevilla, Spain*

Received 19 June 2006; received in revised form 6 September 2006; accepted 16 September 2006
Available online 25 October 2006
Communicated by A. Mikhailov

Abstract

In this work we analyze the possibility that the soliton dynamics in a simple nonlinear model allows functionally relevant predictions of the behaviour of DNA. This suggestion was first put forward by Salerno [M. Salerno, Phys. Rev. A 44 (1991) 5292] by showing results indicating that sine–Gordon kinks were set in motion at certain regions of a DNA sequence that include promoters. We revisit that system and show that the observed behaviour has nothing to do with promoters; on the contrary, it originates from the bases at the boundary, which are not part of the genome studied. We explain this phenomenology in terms of an effective potential for the kink center. This is further extended to disprove recent claims that the dynamics of kinks [E. Lennholm, M. Hörnquist, Physica D 177 (2003) 233] or breathers [J.D. Bashford, J. Biol. Phys. 32 (2006) 27] has functional significance. We conclude that no such information can be extracted from this simple nonlinear model or its associated effective potential.

Keywords: DNA; Genome; Nonlinear dynamics; Solitons; Collective coordinates

1. Introduction

Nonlinear models supporting coherent excitations have appeared in many fields of science since the pioneering discoveries by Fermi, Pasta et al. [1] more than 50 years ago. The success of this approach in modeling complex systems has encouraged its application in other fields. That is the case for biology, where nonlinear models were widely applied in many subjects, such as in the study of the DNA molecule (see, for example, [2–4]). To realize the relevance of these models it should be noticed that, nowadays, the computational cost of molecular dynamics for realistic models of DNA molecules with a few tens of base pairs allows simulation times up to tens of nanoseconds at most. Nonlinear models allow the study of such a complex system with very many degrees of freedom by drastically reducing this amount up to one degree of freedom per base pair, the most relevant for the process under study. It goes without saying that the reduction of a very complicated object such as the DNA duplex to a polymer formed by base pairs, each one with just one degree of freedom (sometimes a few more), helps enormously in the theoretical and computational study of these models. Nevertheless, although simplified, these models can yield important results. An example of these models is the Peyrard–Bishop model of DNA [5], which achieved an important goal when describing the denaturation process of DNA in terms of just the radial distance of the bases on each base pair [6].

Among all these approaches we focus here on the work of Englander et al. [7], who introduced the sine–Gordon (sG) equation as a model for DNA in 1980. The existence of sG solitons in the DNA molecule has been surrounded by controversy, as expected in a field were biology and physics do not always meet in a fruitful way [8,9]. When Englander and co-workers introduced the sG model of DNA, they based their hypothesis on experimental results that showed unexpectedly

* Corresponding author.
E-mail addresses: scuenda@math.uc3m.es (S. Cuenda), anxo@math.uc3m.es (A. Sánchez), niurka@euler.us.es (N.R. Quintero).

0167-2789/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2006.09.005

long lifetimes of open states of DNA duplexes [10]. In spite of the fact that, later, Guéron et al. [11] found more reasonable lifetimes, smaller by one or two orders of magnitude than the ones reported in previous works, a vast amount of literature is still based on the Englander model. On the other hand, the very existence of solitons in DNA is questionable, as the viscous critical force of water is about a thousand times larger than the typical scale of forces in DNA (piconewton range). In fact, the effect of water friction damps out any inertial effect in the world of the cell and, consequently, in DNA [12]. In particular, this raises questions about the applicability of the sine–Gordon model of Englander et al. and related ones, which contain an inertial term at least of the order of the dissipative one: Dominance of the dissipative term would lead to pinning of sine–Gordon solitons and to annihilation of sine–Gordon breathers. Therefore, we want to stress that the approach we are dealing with here, namely DNA models with soliton-like excitations, is purely phenomenological and does not imply any claim concerning the true existence and character of such excitations.

In this context, and keeping in mind the above caveats, the aim of this work is to analyze in depth part of the literature related to the work of Englander et al., providing new results that give insight into a number of important questions. Specifically, we will study the relation between the dynamics of sG solitons and the position of promoters in the genome of the bacteriophage $T7$. This line of work began with Salerno [13–15] at the beginning of the 1990s and was subsequently continued in several works [16–20]. We stress that this is a very important issue: Indeed, if the Englander model behaviour could be connected to functionally relevant positions in the sequence, it would provide a cheap and efficient tool for genomics. Claims in this direction have already been presented [20]. Note that the serious doubts about the existence of solitons in DNA discussed above would have no consequences for this application of the model, because nonlinear models might somehow phenomenologically correlate with important genomic features of sequences. However, as we will show below, the main result of the present work is that, unfortunately, such a connection cannot be substantiated for reasons intrinsic to the nonlinear models themselves.

The structure of the paper is as follows. In Section 2 we discuss the methodology and the results of the first two papers concerning this issue [13,14] in terms of the effective potential introduced by Salerno in collaboration with Kivshar in [15]. In Section 3 we describe the main features of the promoters of the $T7$ genome, and analyze the simulation results of the work of Lennholm and Hörnquist [17] in terms of the effective potential. In Section 4 we discuss recent work concerning breathers in the sG model [20]. Finally, Section 5 concludes the paper by summarizing our main results and their implications.

2. Early work on $T7$: A_1, A_0 and A_3 promoters

More than a quarter of a century ago, Englander and co-workers [7] introduced solitonic excitations into the DNA world as an initial step towards understanding the stability of open segments of DNA molecules [10]. They suggested the well known sG model, that describes the dynamics of a line of pendula in a vertical gravitational field with torsional spring coupling between units, as an effective description of DNA molecules. In this way, the double helix is approximated by two parallel rods on which pendula (base pairs) are attached, and bonding to the opposite base is represented by a "gravitational" potential of each pendulum. Calling ϕ_i the twist angle of the ith base, this model has static soliton (kink) solutions given by

$$\phi_i = 4\arctan(\mathrm{e}^{ai}), \tag{1}$$

valid for $a \ll 1$, where the continuum approximation applies. In Eq. (1), a is a dimensionless parameter representing the parameters of the model, and acts as an effective discretization parameter of the continuum sG problem. In spite of such a great oversimplification of the real problem, the model contained the main feature of breaking a bond around $\phi = 0$. In addition to this, the results were consistent with available data [10] although Englander et al. were aware of the lack of evidence of solitonic excitations.

Salerno, in his pioneering and interesting work [13], tried to find a relation between relevant sites in the $T7$ genome and the dynamics of sG kinks moving along the inhomogeneous DNA sequence under study; the main difference with respect to previous works was the introduction of the inhomogeneity of the sequence in the model. To do so, he took the static kink solution (1), with center at n_0, and used it as initial condition of the equations of motion of the discrete, *inhomogeneous* sG (or Englander) model,

$$\ddot{\phi}_i = \phi_{i+1} - 2\phi_i + \phi_{i-1} - q_i \sin\phi_i, \tag{2}$$

q_i being the parameter that carries all the information of the sequence under study. It is defined as $q_i = \beta\lambda_i/K$, where K is the torsional spring constant between consecutive bases, β is the energy of a hydrogen bond and λ_i is the number of hydrogen bonds in a base pair, which is $\lambda_i = 2$ for AT base pairs and $\lambda_i = 3$ for CG base pairs. Considered as a discrete version of the continuum sG equation, the effective discretization of the lattice used in [13] was $a = \bar{q}^{1/2}$, where $\bar{q} = \frac{1}{N}\sum_{i=1}^{N} q_i$ (N being the number of bases of the sequence). This value is around $a \simeq 0.07$, which is small enough to avoid spurious discretization effects when numerically integrating Eq. (2). In fact, taking Eq. (1) as an ansatz in Eq. (2) was a good choice, as the kink is a very robust object even in inhomogeneous sequences and its center can be well defined by interpolating the position where $\phi = \pi$ [18,19].

Once the model was defined, Salerno built a sequence $\{q_i\}$ to introduce it in (2). He was interested in the genomic sequence of the $T7$ A_1 promoter but, instead of using the original DNA sequence, he built a "synthetic" one from the original. We will review all the details of this process as this will be the key to understanding the results of [13]. He took a sequence S of 168 bases containing the so-called A_1 promoter (further details on $T7$ promoters will be given in the next section) which