Discrete model for DNA-promoter dynamics

Mario Salerno
Dipartimento di Fisica Teorica, Università di Salerno, I84100 Salerno, Italy
(Received 6 May 1991)

We introduce a discrete model for DNA that takes into account the information about specific base sequences along the double helix. We use this model to study nonlinear wave dynamics of the $T7A_1$ DNA promoter. As results we show the existence in the promoter of a dynamically active region in which static solitons acquire finite velocities, which contrasts with regions where solitons simply remain static. Furthermore, when they pass through this region moving solitons are accelerated, decelerated, or reflected, depending on their initial velocities. The possibility that these dynamical effects play a role in the mechanism of genetic activation is suggested.

PACS number(s): 87.10.+e

I. INTRODUCTION

The possibility that vibrational energy in DNA might be trapped into solitary-wave excitation has received considerable attention in the past decade. This idea was initially suggested by Englander *et al.* [1] to explain the existence of transiently open states in DNA and was further developed in terms of mechanical systems consisting of chains of pendula (bases) coupled by springs (the sugar-phosphate backbone) by several authors [2–10]. In these models the attention was directed to the degrees of freedom characterizing base rotations in the plane perpendicular to the helical axis around the backbone structure (plane base rotator). In the continuum limit these models lead to the well-known sine-Gordon equation, or to equations closely related to it, in terms of which several dynamical properties of DNA were investigated [2–10]. In particular, the effects of weak inhomogeneities in simple DNA fragments, consisting of uniform base sequences of a given type followed by uniform base sequence of the other type, were considered in terms of a parametrically perturbed sine-Gordon equation [8]. Other nonlinear models of DNA, based on different choices of the dominant degrees of freedom involved and modeling different aspects of DNA dynamics, were also developed [11–15]. In all these models, however, little attention was devoted to specific base sequences of natural DNA.

In the present paper we propose a model for DNA that uses the information about the specificity of base sequences of real DNA to investigate the possible connection existing between base sequences, dynamics, and functioning of the molecule. We expect this connection to play an important role in the mechanism of genetic activation, where the specificity of DNA promoters is crucial both for the binding of specific proteins (RNA-polymerase) to DNA and for the transport of such proteins from the promoter to the fragments corresponding to the genes. It has been suggested that the binding process may involve a conformational distortion of DNA double helices that can be described by a solitary-wave excitation (also called a soliton) [9]. Such excitation can travel along DNA without dispersion, carrying the bonded protein in the coding region corresponding to the gene. The distance from the promoter to the target (gene) might be quite large, sometimes more than 1000 bases, so that a question naturally arises: Where does this excitation find the energy to cover such a large distance? Although the activation mechanism is not yet fully understood, it is known that there exists in the promoter region an activation region that gives energetic input to the transport process. To investigate this problem, we think it is crucial to take into account the nonuniformity of the DNA induced by specific base sequences characterizing promoter regions. To this end we consider the base-rotational motion around the sugar-phosphate backbone structure in terms of a discrete sine-Gordon equation modeling a chain of pendula connected by springs, each pendulum representing a specific base pair, and including base information through the potential function modeling the hydrogen bonds between base pairs. By noting that the hydrogen bond involved in the pairings is double for adenine-thymine (A-T) and triple for guanine-cytosine (G-C), we obtain a simple rule to construct a chain corresponding to a specific DNA sequence, i.e., we fix the ratio between the strength of the potential functions of A-T and G-C pairs to be $\frac{2}{3}$, while the ratio between anharmonicity (base rotation) and dispersion (backbone springs) is kept as a free parameter to be fixed by experimental data. As a result we show the existence in the chain corresponding to $T7A_1$ promoter of a region in which a static solitary-wave excitation acquires, as a consequence of the nonuniform background, a finite velocity along the chain, in contrast with what happens in the rest of the chain, in which the wave simply remains static. Furthermore, nonstatic solitons passing through this active region are accelerated, decelerated, or reflected depending on their initial velocities. The possibility that these dynamical effects may play a role in the mechanism of genetic activation is suggested.

The paper is organized as follows. In Sec. II we introduce the discrete system we used to model DNA promoter dynamics. In Sec. III we numerically investigate the dynamics of a soliton in the $T7A_1$ promoter region, and the results are qualitatively described in terms of a

 ©1991 The American Physical Society

continuous parametrically perturbed sine-Gordon equation. In the last section we summarize the main results of the paper.

II. MODEL

Let us start by considering the so-called B form of the DNA molecule, whose conventional structure is schematically shown in Fig. 1. In this figure the thick lines represent the backbone double-helix structure, the horizontal lines terminated by conjugated letters denote complementary bases pairing, with the double and triple symbol between them representing hydrogen bonds. We concentrate our attention on the degrees of freedom characterizing base rotations in the plane perpendicular to the helical axis around the backbone structure. This dynamics plays an important role for DNA functioning since, under certain circumstances, it can open the hydrogen bonds between conjugated pairs, exposing the unpaired bases to the action of external ligands. To model this motion we assume that each base of a strand is coupled with the next-neighbor bases of the same strand by the elastic backbone restoring forces, and with the complementary base in the opposite strand by the anharmonic potential used to model the hydrogen bond. We take as canonical variables the deflection angles ψ_i and θ_i that two complementary bases form with the line passing from the attaching points of the bases to the strands. For the potential between base pairs we use a simple cosine function. The Hamiltonian for such a model is then written as

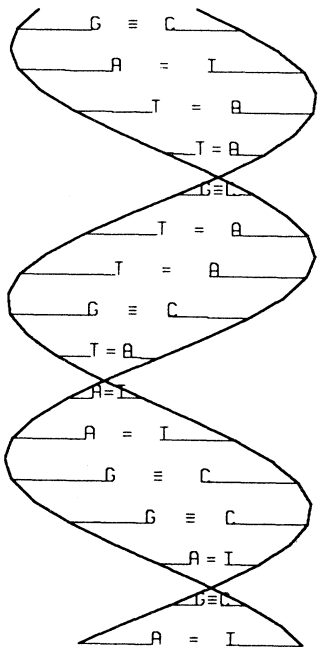

FIG. 1. Schematic structure of the B form of the DNA. Double and triple symbols denote hydrogen bonding between complementary bases.

$$H=\sum_{i=1}^{N}\tfrac{1}{2}\{I_i(\dot{\psi}_i^2+\dot{\theta}_i^2)+k_i(\psi_{i+1}-\psi_i)^2 +\bar{k}_i(\theta_{i+1}-\theta_i)^2+\eta_i[1-\cos(\psi_i-\theta_i)]\}\ , \quad (1)$$

where k_i and $\bar{k}_i$ denote the backbone spring constants along the two helices, I_i is the moment of inertia of the individual bases, N is the number of base pairs in the chain, and η_i is a nonlinear parameter used to model the strength of hydrogen bonds between complementary bases. As mentioned before, we choose the coefficients η_i in Eq. (1) according to the rule $\eta_i=\lambda_i\beta$ with $\lambda_i=2$ if it refers to A-T or T-A pairs, $\lambda_i=3$ otherwise, with β a free parameter to be determined later. For simplicity in the following we consider only uniform restoring forces and uniform moments of inertia along the two strands of DNA, so we fix $k_i=\bar{k}_j=K$, $I_i=I$, and $i,j=1,\ldots,N$. The equations of motion obtained from Eq. (1) are then

$$\begin{aligned} I\ddot{\psi}_i&=K(\psi_{i+1}-2\psi_i+\psi_{i-1})-\frac{\beta}{2}\lambda_i\sin(\psi_i-\theta_i)\ ,\\ I\ddot{\theta}_i&=K(\theta_{i+1}-2\theta_i+\theta_{i-1})-\frac{\beta}{2}\lambda_i\sin(\theta_i-\psi_i)\ , \end{aligned} \quad (2)$$

from which we obtain the following equation for the angle difference $\phi_i=\psi_i-\theta_i$ between complementary bases:

$$\ddot{\phi}_i=\phi_{i+1}-2\phi_i+\phi_{i-1}-\frac{\beta}{K}\lambda_i\sin(\phi_i)\ . \quad (3)$$

In this equation time has been scaled according to $t\rightarrow\sqrt{I/k}\,t$, so as to leave as a parameter in the equation just the ratio β/K between anharmonicity and dispersion. To get an estimate of this parameter we use the values of β and K reported in Ref. [7], which were derived by combining information from the sine-Gordon model with experimental measurements. This leads to a value of β/K between 10^{-3} and 10^{-4}. These values are found to be consistent with the requirement of existence and stability of solitary-wave solutions for Eq. (3), as shown in the next section. Finally, we note that in the case of uniform nonlinear parameters, $\lambda_i\equiv\lambda$, $i=1,\ldots,N$, Eq. (3) reduces, in the continuum limit, to the well-known sine-Gordon equation

$$\phi_{xx}-\phi_{tt}-\sin(\phi)=0\ , \quad (4)$$

with exact soliton solutions

$$\phi(x,t)=4\tan^{-1}\{\exp[\gamma(x-vt-x_0)]\}\ ,\qquad \gamma=(1-v^2)^{-1/2}\ . \quad (5)$$

In the next section Eq. (5) will be used as an initial condition to integrate Eq. (3) with the λ values corresponding to the $T7A_1$ promoter.

III. NUMERICAL EXPERIMENT AND ANALYSIS

In order to investigate dynamical effects due to the specificity of base sequences, we integrate Eq. (3) with λ_i values corresponding to the $T7A_1$-promoter base sequence S reported in Table I with β/K fixed to 2×10^{-3}. From biochemical studies it is known that RNA polym-

TABLE I. Sequence of 168 bases corresponding to promoter $T7A_1$. The double arrows delimit the region in which RNA polymerase can interact with the double helix, while the numbers in parentheses refer to the standard numeration used in biology.

```
T T G T C T T T A T T A A T A C A A C T C A C T A T A A G G

A G A G A C A A C T T A A A G A G A C T T A A A A G A T T A
                                  ↑    ⟹
A T T T A A A A T T T A T C A A A A A G A G T A T T G A C T
                  50                                   (−1)
                                                        ↓
T A A A C T C T A A C C T A T A G G A T A C T T A C A G C C
(+1)
↓
A T C G A G A G G G A C A C G G C G A A T A G C C A T C C C
                              ⟸     ↑
                                    140
A A T C G A C A C C G G G G T C A A
```

lymerase can bind to DNA in the region going from base pair (BP) 51 to BP 140, so we expect this region to be dynamically active. To avoid the influence of boundary conditions on promoter dynamics we have constructed from sequence S in Table I a longer sequence of 1000 bases according to the rule

$$S(1,5)+9S(1,50)+S(51,140)$$
$$+16S(140,168)+S(162,168)\ ,$$

where the symbol $kS(i,j)$ denotes the subsequence of S going from BP i to BP j repeated k times. In this longer sequence the promoter region extends from BP 455 to BP 545, so that we could safely use reflexive boundary conditions in the numerical simulation. To investigate differences in soliton dynamics in this region we have performed several integrations of Eq. (3) with the initial position of the static soliton varied inside the promoter region. In Fig. 2 we show the time evolution of a solitary wave placed outside the promoter in correspondence to BP 415, from which we see that the initial solitary wave remains static. Increasing the initial position through the promoter region from BP 415 up to BP 505 by increments of 1 does not introduce significant differences in dynamics, except for small oscillations around the initial base-pair value. In Fig. 3 we report the time evolution of an initial soliton placed inside the promoter region at BP 510. From this figure it is clear that the wave acquires a velocity $v=0.18$ towards the left end of the chain, is reflected without loss of energy at the boundary, and is reflected again at the promoter region with velocity $v=0.18$. This behavior is enhanced when the initial position is increased from BP 510 to BP 535. In Figs. 4 and 5 we report the soliton time evolution for initial positions at BP 525 and BP 535, respectively. From these figures we see that the effect is stronger at BP 535, where the wave also reached the maximum velocity $v=0.3$. Beyond BP 535 this dynamical behavior is drastically re-

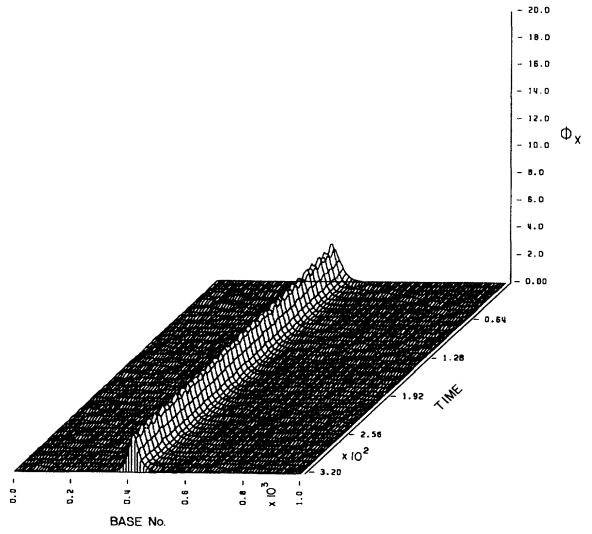

FIG. 2. Time evolution of a solitary wave with initial velocity $v=0$ placed outside the promoter region at BP 415.

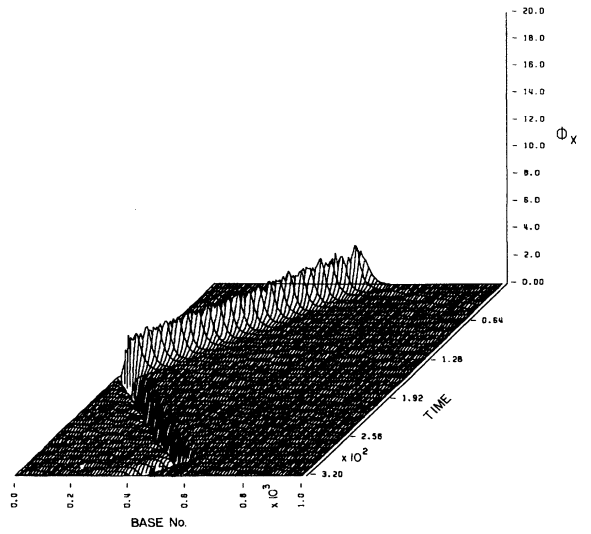

FIG. 3. Time evolution of a solitary wave with initial velocity $V=0$ placed inside the promoter region at BP 510.

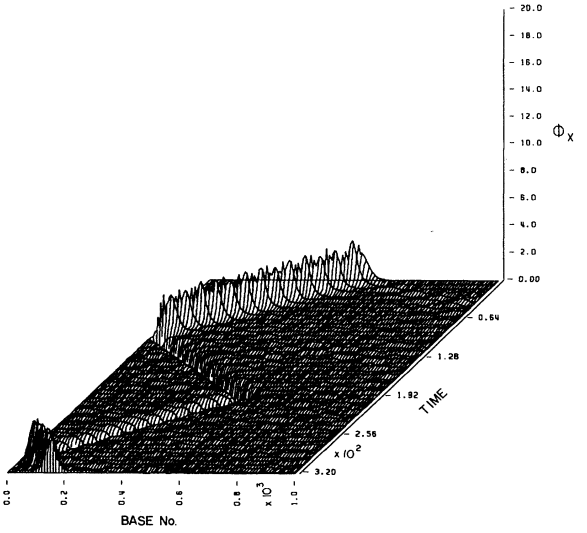

FIG. 4. Same as in Fig. 3 but with intial position at BP 525.

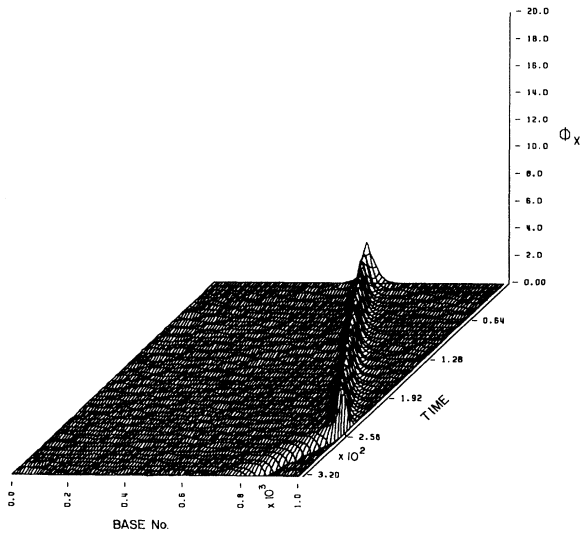

FIG. 6. Same as in Fig. 3 but with initial position at BP 540.

duced. In Figs. 6 and 7 , we report the time evolution of an initially static soliton placed at BP 540 and BP 555, respectively, from which we see that at BP 540 the wave acquires a small velocity ($v \cong 0.08$). towards the right end of the chain, while at BP 555 the wave simply remains static. In Fig. 8 we show the effect of background inhomogeneity on a moving soliton placed at BP 900 with an initial velocity $v = 0.3$. From this figure we see that the soliton is accelerated by the promoter region when it travels from the right to the left, and is deccelerated when it travels in the opposite direction.

These results show the existence of a dynamically "active" region going from BP 510 to BP 540 inside the $T7A_1$ promoter. When an initially static soliton is inside this region, the wave receives an energetic input to travel along the DNA chain, the direction of the motion depending on the initial position of the soliton and the effect being weaker when the wave is placed to the left of BP 535. Furthermore, the above dynamical behavior disappears as the wave moves far away from this region. The above dynamical behavior can be understood in terms of a chain of pendula, with lighter and heavier masses corresponding to A-T and to G-C base pairs, respectively. A soliton placed on the top of the heavier-pendulum region has a bigger rest mass than the one placed on the lighter one. On a nearly uniform back-

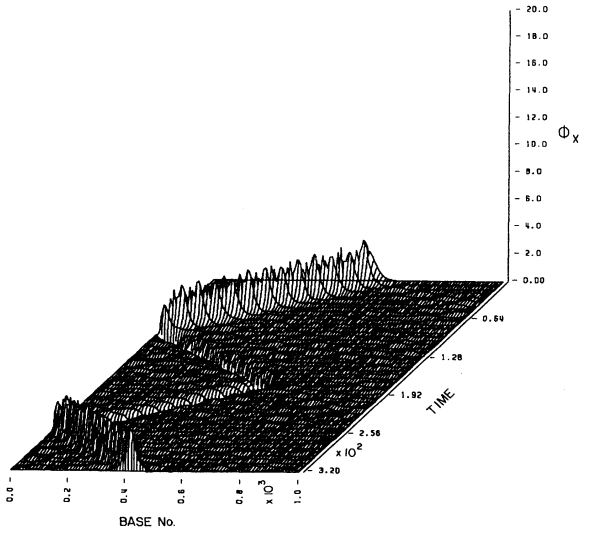

FIG. 5. Same as in Fig. 3 but with initial position at BP 535.

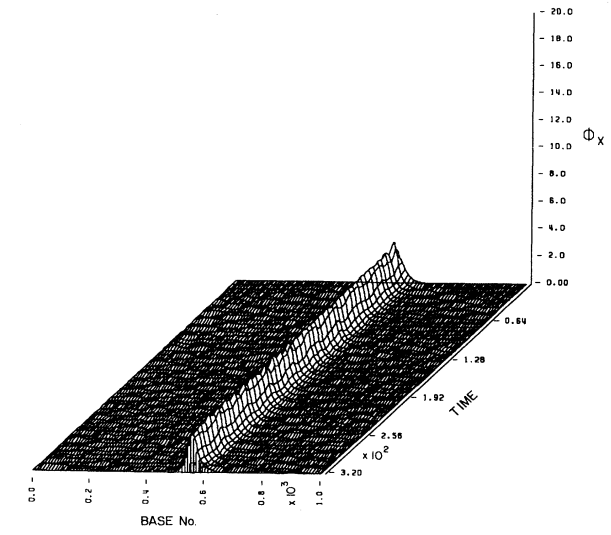

FIG. 7. Same as in Fig. 3 but with initial position at BP 555.

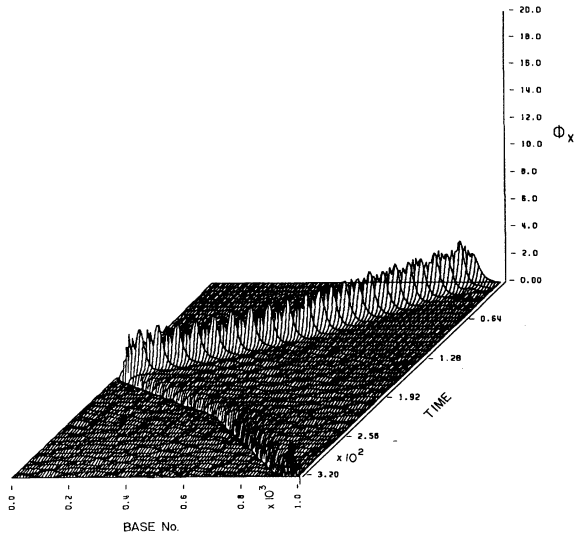

FIG. 8. Time evolution of a solitary wave with initial velocity $v=0.3$ placed inside the promoter region at BP 900.

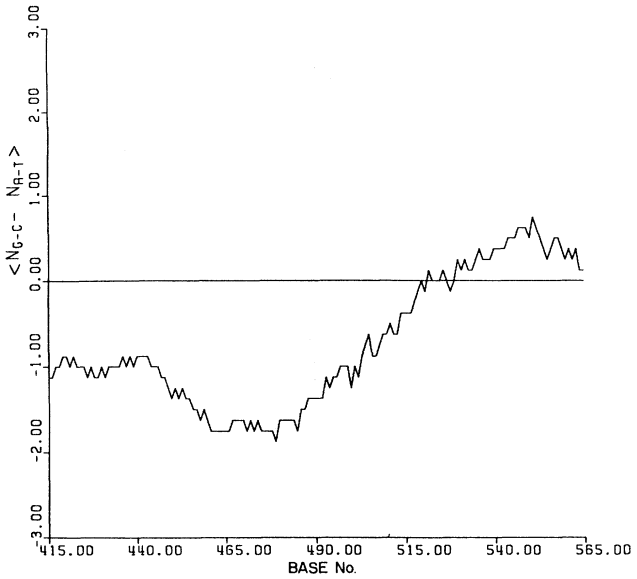

FIG. 9. Averaged difference between the numbers $N_{\text{G-C}}$ and $N_{\text{A-T}}$ of G-C and A-T pairs contained in a DNA fragment of N_s BP inside the promoter region. The average is calculated with respect to N_s and N_s is increased from 10 to 50 in increments of 1.

ground the solitary wave remains static or oscillates around some equilibrium position, while in the transition region where the wave partially overlaps the "heavier" part of the chain there is an effective potential that imparts motion to the wave. The kinetic energy acquired by the wave can be approximated by the difference between its initial and final rest masses. From Table I we see that the base sequence of the $T7A_1$ promoter contains a region going from BP 110 to BP 140 (from BP 515 to BP 545 in the long chain, respectively) in which G-C pairs are more frequent, in contrast with the rest of the sequence, in which the A-T pairs are equal in number or are more abundant. The parameter that plays a crucial role in the above dynamical behavior is N_S, the number of base pairs along the DNA double helices over which the wave extends, which in turn depends on the ratio between the anharmonicity and the dispersion parameters. In our simulation the solitary wave extends over 35 base pairs, i. e., almost three turns of DNA double helices. We checked that the above dynamical behavior is qualitatively preserved if we decrease the size of the wave down to 20 base pairs (two turns of DNA helices). In Fig. 9 we reported the averged difference between the numbers $N_{\text{G-C}}$ and $N_{\text{A-T}}$ of G-C and A-T pairs, respectively, contained in a DNA fragment of N_s base pairs inside the promoter region, versus the site i around which N_s is centered. Here the average is calculated with respect to N_s, with N_s increased from 10 to 50 in increments of 1. From this figure the existence of a transition around site 515 from a region dominated by A-T pairs to a region in which G-C pairs are more frequent is evident.

Although these phenomena refer to a discrete chain, they can be qualitatively described in terms of a continuous sine-Gordon equation with a smooth spatial homogeneity modeling the transition from a uniform A-T region to a uniform G-C region [8]. This leads to consideration of the continuous parametrically perturbed sine-Gordon equation [8,16]

$$\phi_{xx}-\phi_{tt}-\sin(\phi)=\epsilon(x)\sin(\phi)\ , \tag{6}$$

where $\epsilon(x)$ represents a smooth function of space going from $\epsilon=0$ to some finite value $\epsilon=\delta$ as x goes from $-\infty$ to $+\infty$ (here the smoothness of ϵ is assumed to avoid the creation of background radiation). As shown in Ref. [16], the energy

$$\int_{-\infty}^{+\infty}\{\tfrac{1}{2}(\phi_x^2+\phi_t^2)+[1+\epsilon(x)][1-\cos(\phi)]\}dx \tag{7}$$

is a conserved quantity and therefore

$$\frac{\gamma(v_i)}{\gamma(v_f)}=\sqrt{1+\delta}\ , \tag{8}$$

where v_i and v_f denote, respectively, the initial and final velocity of the soliton, γ is the Lorentz contraction factor, and $\sqrt{1+\delta}$ is the ratio between the soliton rest masses corresponding, respectively, to the regions $\epsilon=0$ and $\epsilon=\delta$. From Eq. (8) it is clear that the soliton will be reflected with velocity $v_f=-v_i$ or transmitted with velocity

$$v_f=[v_i^2(1+\delta)-\delta]^{1/2} \tag{9}$$

according to whether v_i is greater or less than $[\delta/(1+\delta)]^{1/2}$. Furthermore, we have $v_f>v_i$ or $v_f<v_i$ depending on whether $\delta<0$ or $\delta>0$; i.e., if the soliton moves from a uniform region of A-T bases into a uniform region of G-C bases or vice versa [8,16]. Finally, we note

that this analysis is in qualitative agreement with the above numerical results.

IV. CONCLUSION

In summary, we have introduced a discrete model for DNA that takes into account the information about specific base sequences along DNA double helices. In the case of the $T7A_1$ promoter we have demonstrated the existence of a region in which a static solitary-wave excitation acquires a finite velocity, in contrast with regions in which it remains static. Furthermore, it has been shown that moving solitons passing through the $T7A_1$-promoter region experience an acceleration, deceleration, or reflection, depending on the modulus and on the direction of the incoming velocity. A preliminary investigation indicates that this result will generalize to other DNA promoters of the $T7$ family, as will be reported elsewhere. We think the above dynamical behavior is of particular importance because it can explain the functioning of DNA promoters as energetic activators of the RNA-polymerase transport process.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge Dr. Ludmila Yakushevich for interesting and stimulating discussions and Dr. S. G. Kamzolova and Dr. O. N. Ozoline for sending me information about the promoter I used in the paper. Financial support from INFM (Italy) is also acknowledged.

[1] S. W. Englander, N. R. Kallenbach, A. J. Heeger, J. A. Krumhansl, and S. Litwin, Proc. Natl. Acad. Sci. U.S.A. **77**, 7222 (1980).
[2] S. Yomosa, J. Phys. Soc. Jpn. **51**, 3318 (1982); **52**, 1866 (1983).
[3] S. Takeno and S. Homma, Prog. Theor. Phys. **70**, 308 (1983).
[4] S. Yomosa, Phys. Rev. A **27**, 2120 (1983); **30**, 474 (1984).
[5] S. Homma and S. Takeno, Prog. Theor. Phys. **72**, 679 (1984).
[6] Chung-Ting Zhang, Phys. Rev. A **35**, 886 (1987).
[7] V. K. Fedyanin and L. V. Yakushevich, Stud. Biophys. **103**, 171 (1984).
[8] L. V. Yakushevich, Stud. Biophys. **121**, 201 (1987).
[9] R. V. Polozov and L. V. Yakushevich, J. Theor. Biol. **130**, 423 (1988).
[10] L. V. Yakushevich, Phys. Lett. A **136**, 413 (1989).
[11] J. A. Krumhansl and D. M. Alexander, in *Structure and Dynamics: Nucleic Acids and Proteins,* edited by E. Clementi and R. H. Sarma (Adenine, New York, 1983), p. 61.
[12] E. Balonovski and P. Baconsfield, Phys. Lett. **93A**, 52 (1982); **95A**, 454 (1983); Phys. Rev. A **32**, 3059 (1985).
[13] A. C. Scott, Phys. Rev. A **31**, 3518 (1985); Phys. Scr. **32**, 617 (1985).
[14] V. Muto, Ph.D. thesis, The Technical University of Denmark, Lyngby, 1988.
[15] V. Muto, J. Halding, P. L. Christiansen, and A. C. Scott, J. Biomol. Struct. Dyn **5**, 873 (1988).
[16] M. Salerno, M. R. Samuelsen, P. S. Lomdahl, and O. H. Olsen, Phys. Lett. **108A**, 241 (1985).