Journal of Biological Physics **24:** 97–114, 1999.

A Twist Opening Model for DNA

MARIA BARBI[1,3,5], SIMONA COCCO[2,3], MICHEL PEYRARD[3] and STEFANO RUFFO[4,5]

[1] *Dipartimento di Fisica, Università degli Studi di Firenze, Largo E. Fermi, 2 – 50125 Firenze, Italy*
[2] *Dipartimento di Scienze Biochimiche P.le A. Moro, 5 – 00185 Roma, Italy*
[3] *Laboratoire de Physique, CNRS URA 1325, École Normale Supérieure de Lyon, 46, allée d'Italie, 69364 Lyon Cedex 07, France*
[4] *Dipartimento di Energetica 'S. Stecco' Via S. Marta, 3 – 50139 Firenze, Italy*
[5] *INFN, Sezione di Firenze, Italy*

Abstract. The real mechanisms of several biological processes involving DNA are not yet understood. We discuss here some aspects of the initiation of transcription, in particular the formation of the open complex and the activation mechanism associated to enhancer binding proteins. Transcription activation seems to be governed by underlying dynamical mechanisms related to several distortions of the double chain structure: a dynamical approach on a mesoscopic description level could then allow a deeper understanding of this complex process. Starting from the Peyrard Bishop (PB) model, that considers only the hydrogen bond stretching of each base pair, we describe here an extended DNA model, proposed in [1], that allows a rather good representation of the double helix geometry and of its structural features by the introduction of angular variables related to the twist angle. Using a generalized multiple scale expansion for the case of vectorial lattices derived elsewhere [2], we derive analytically small amplitude approximate solutions of the model which are movable and spatially localized: we present here the results of this calculation and show how the special shape of the solutions is in good agreement with what can be expected for coupled angular radial distortions in the real molecule.

Key words: DNA modeling, helix, hydrogen bond stretching, nonlinear dynamics, solitons, transcription initiation, twist

1. Introduction: transcription initiation processes and double helix structure

We illustrate a geometrical dynamical model of the DNA helical structure. The main aim of the model is to help in studying some physical features that could be implied in the actual mechanisms of some biological processes which are not yet well understood, such as replication, transcription and its initiation, in which the structural and dynamical distortions of DNA play a major role. The model takes into account explicitly the geometrical constraints acting during these processes. As it should be clear from the following discussion, biologists pay more and more attention to the involved structural mechanisms and to their direct implications in allowing the processes themselves.

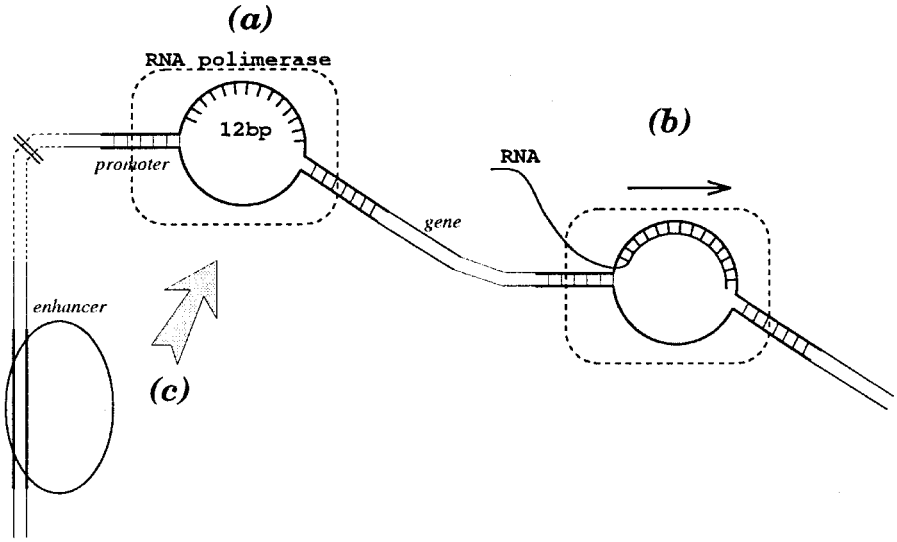

Figure 1. Schematic representation of the transcription process. (a): open complex formation; (b): elongation phase; (c): enhancer linking protein activation.

We have focused our attention on the transcription process (Figure 1), i.e. the process by which the instructions coded in a gene can be read and transcribed on a RNA chain, that will then be used as a template for the synthesis of proteins. Transcription is a complex process that involves a large number of enzymes. The main enzyme responsible for transcription, RNA-polymerase, has first to recognize a specific region in the DNA chain, called *promoter*, that is located immediately upstream with respect to the gene, and to bind to it, in many cases together with some other proteins or *transcription factors* that are needed to maximize the rate of transcription, forming the *initiation complex* [3, 4]. The next step is the formation of the *open complex*: a quite small segment (about 15–20 base pairs (bps)) of the helix, in correspondence to the promoter region, is 'melted' with the separation of the two strands and the formation of a *transcription bubble.* Then the RNA-polymerase and the transcription bubble start to move along the gene, copying the coding sequence into RNA (*elongation phase*).

We shall consider in particular transcription initiation, i.e. the group of processes that allows the transcription to start. The formation of the open complex and its motion along the chain are often regulated by a set of activation/inhibition processes that function by means of the binding of other proteins to different DNA regions that could be very far along the chain either upstream or downstream with respect to the promoter. This starting activating sites can be upstream and close to the promoter region (usually with a distance along the chain of about 100–200 bps), or far from it (up to several kilobases). In the latter case they are called *enhancers*. The mechanisms of activation are still mostly unknown, particularly for what concerns the latter.

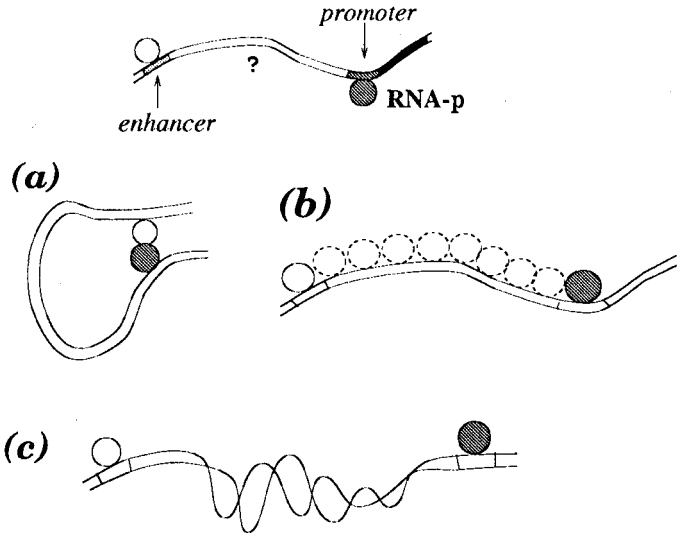

Figure 2. Proposed enhancer activation mechanisms: (a) loop formation; (b) 'oozing'; (c) distortion transmission.

Atomic force microscope technology has allowed the visualization of one mechanism of activation for transcription factors bound to DNA sites distant from the promoter: they loop the molecule so to bring the promoter and enhancer regions closer (Figure 2a) [5]. In this case a direct interaction between the various bound proteins could be the main way of activation.

However, the formation of a loop is not the unique way for activation, and two other mechanisms for these long range activation effects have been proposed, namely the induction of the cooperative binding of transcription factors, so that all the DNA between the enhancer and the promoter is covered (*'oozing'* model, Figure 2b), and the transmission of altered DNA structures from the enhancer, along the helix, to the transcription complex (Figure 2c) [3]. Among these two proposed models, the latter seems to correspond better to experimental results. Some experimental works seem in fact to confirm that the structural modifications induced in the enhancer sites are actually much more important for activation, than the specificity of the linked proteins themselves. The enhancer region is known to be bent by the activator factors; furthermore, it has been shown that activation can be achieved, in some cases, by replacing the enhancer region by an intrinsically bent DNA sequence that is not a protein binding site, so that there is no more protein mediation: in this case, then, the structural deformation acts by itself as an activator [6, 7, 8, 9].

It is interesting to mention also the proposed *'hit and run'* mechanism for DNA binding proteins, according to which, immediately before the open complex formation, activator factors bind just for a short time the DNA, locally modify it, and then leave the binding site [10]. This could suggest that the action of these activator

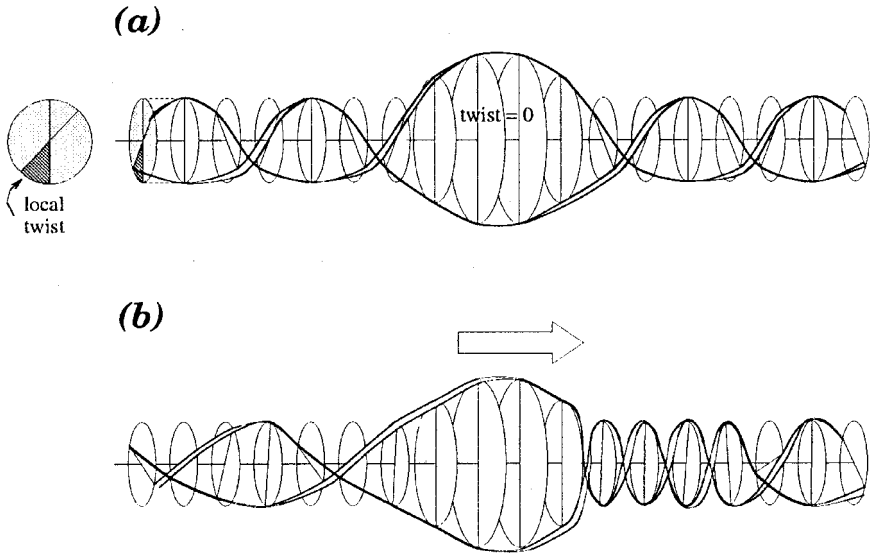

Figure 3. Local twist effects. (a): bubble local opening is possible only if accompanied by an untwisting; (b): in the elongation phase, the movement of the untwisted bubble along the helix produces some overtwist upstream and some undertwist downstream.

factors could be that of 'launching' a structural distortion that could travel along the chain toward the promoter region, where it will eventually act as activator by inducing the right conformational change.

These are not the only cases in which structural deformations can be shown to be relevant for biological functioning in transcription, in long range activation effects as well as in open complex formation, and bending is not the unique kind of 'active' deformation. Besides the cited results there are for instance several evidences on the enhancing effects of intrinsically superhelical sequences, i.e. regions where the rotation between the neighboring base pair, or *twist* angle (Figure 3), is changed with respect to its value in equilibrium conditions [11, 12]. Because twist deformations are strongly related, for geometrical reasons, with bubble formation, this kind of effects has from our point of view a great interest, as we will discuss immediately hereafter.

Local changes in twist are in fact necessarily involved in the transcription bubble formation: simple geometrical considerations allow to understand that the stretch of the hydrogen bonds in some local region of the chain is possible only if the twist angle is decreased with respect to its equilibrium value Θ_0 in that region, in order to bring the steps of the double helix 'ladder' toward a common vertical plane (Figure 3a) [4]. This constraint is well known by biologists because in the elongation phase, when the bubble starts to move, it causes topological problems, giving rise to a positive excess of twist in front of the RNA-polymerase and to a negative excess behind (Figure 3b). Special enzymes, the *topoisomerases*, are

designed to release these structural stresses by cutting one of the two strands and rejoining it after having made some turns around the other. If the topoisomerases cannot act efficiently enough, the excess of twist could sometimes be so important that it prevents the transcription, showing how the twist deformations are important in all the processes in which hydrogen bond opening is implied.

In DNA, changes in the twist angle can be always put in relation with modifications in its three-dimensional shape by the well known White law $Lk = Tw + Wr$, where Tw (twisting number) is the number of turns that one strand makes around the other, and is then the ratio between the overall twist angle change along the considered DNA segment and 2π; Wr (writhing number) is the number of times the molecular main axis crosses over itself, giving rise to a coiled three-dimensional configuration; Lk (linking number) is the total number of times the two strands wrap around, and is a fixed number in closed circular chains [4]. The phenomenology which relates to this interdependence between local angular variables (twist) and three-dimensional secondary structures is usually referred to as *supercoiling*, and many interesting papers have been devoted to the study of their influence on transcription and on other processes.

Besides the topological constraint related to the helix geometry, open complex formation presents finally other interesting features from the physical point of view, which could be in turn related to the mentioned deformational effect. In this phase no chemical energy is required by the RNA-polymerase to open the two strands. It is interesting to investigate the actual mechanism that allows the concentration of enough energy to break hydrogen bonds; furthermore it is well known that the bound RNA-polymerase induces a strong bending in the promoter region, that could be in relation with the opening [3].

From what we have discussed so far it should be clear that transcription initiation is in many aspects strongly dependent on structural modifications. Consequently it is quite natural to suppose that these modifications could act directly in the process itself. Our aim is to look at these mechanisms from a physical point of view, and to investigate for instance how the RNA-polymerase can collect enough energy in the promoter region to melt the DNA in a bubble, how bending and twist modifications are implied in this process, if there is any structural mechanism which is responsible for the activation induced by the enhancer linked proteins, and how structural modifications are responsible for some features of transcription activation and initiation. To do this, it is necessary to simplify in some way our description of the double helix, by building a model that should be simple enough to be treated in a mathematical way, but that should contain the most essential geometrical constraints of the molecule to allow a quite realistic description of the main dynamical features of the various processes involved.

We will describe in the next section how this can be made, according to our results [1]. The outlines of the paper is the following. In Section 2 we will initially recall briefly the results of the planar Peyrard Bishop model, then we shall describe our extended model, the chosen degrees of freedom, the geometrical constraints

introduced, and we will then write the model Lagrangian. The model produces the correct helical shape and leads to a correct coupling between torsional deformation and hydrogen bond opening: we will confirm this fact by numerical integrations of the equations of motion performed by simple initial conditions. In Section 3, we will deal with the problem of finding small amplitude soliton like solutions for the model. We briefly recall the extended expansion technique presented in [2] and the results obtained by applying it to the model [1]. The analytical shape of the solution will be presented, compared with the results of numerical integrations and briefly commented. In the Conclusions we will present some possible future investigations allowed by the new model and we will discuss some problems related to the choice of some physical parameters.

2. A geometrical model of DNA

A first attempt to describe physically the dynamics of the formation of the transcription bubble has been made with the Peyrard Bishop (PB) model [13, 14]. The authors focus mainly on the energetic aspects of the process and build a planar model, where each base pair is described by a single degree of freedom y_n that corresponds to the stretching of the hydrogen bond linking the two bases. The strength of the bond is described by an on site Morse potential $V(y_n) = D(\exp(-\alpha y_n) - 1)^2$, where the constants D and α are related respectively to the depth and the width of the potential well. The neighboring bases are coupled by an elastic potential with the elastic constant K which represents the stacking interaction between neighboring base pairs.

The Lagrangian for the PB model is then

$$\mathcal{L}_{PB} = \sum_n m\dot{y_n}^2 - \sum_n D\left(e^{-\alpha y_n} - 1\right)^2 - \sum_n K\left(y_n - y_{n-1}\right)^2 . \tag{1}$$

One of the main results of this model is that it is possible to find analytically, applying a standard Multiple Scale Expansion technique [15], special (approximate) soliton-like solutions of the form of breathers that can eventually move along the chain and which are characterized by a localized envelope and an internal oscillation [16]. The internal oscillation is asymmetrical, corresponding to a positive mean value of the stretching y_n. These localized oscillations are well stable in the small amplitude limit, in the sense that they can propagate along the chain for long times with a very small loss in energy and without changing their shape.

Having a very small amplitude, the PB moving breathers probably cannot be considered as the traveling activation distortions of the type suggested in the introduction. The authors suggest anyway that breathers could be considered as precursors for the transcription bubble formation. They show in fact that, in a thermalized chain, breathers form spontaneously by energy localization [13, 14]; simulating the DNA thermal denaturation by increasing the bath temperature, one can see that they have a tendency to grow by collecting energy from smaller excitations [17],

up to the formation of bigger denaturation bubbles and then lead to a complete strand separation. Among the main interesting properties of these kind of solutions it is even shown that they can eventually be trapped by local inhomogeneities of the chain [16]. This allows then the authors to make a hypothesis for a possible mechanism of the open complex formation: since we know that RNA-polymerase bends the promoter region, one can imagine that this results in a local change of the coupling constant K, and thus in the introduction of a local inhomogeneity that acts as a breather trap and collects in that region enough energy to allow bubble formation.

The PB model represents a good starting point because it is able to reproduce some interesting features of the energetics and dynamics of DNA. However, as we have discussed, there are constraints related to the molecule geometry which are very important in determining the efficiency of several biological functions. For this reason we have proposed [1] a more realistic model, based on the introduction of the helical structure into the description of the molecule, by the addition of one more degree of freedom to the hydrogen bond stretching, namely: twist angle. 'Twist' can be introduced as a local degree of freedom remaining on the one dimensional lattice model, without needs of three dimensional description. By constraining local twists and hydrogen bonds stretching in an appropriate, geometrical way, we are able to obtain a model that contains the fundamental features of the helical structure and that mimics quite well its possible dynamical deformations.

We stress furthermore that the twist angle can maybe model indirectly other DNA structural deformations. In fact, there exists strong geometrical links between the various local helix deformations in the three-dimensional configuration – bending, supercoiling – as well as in local helix deformation – stretch of the hydrogen bonds, changes in twist. It is especially interesting to remark that the bending of the chain implies that, in the same region, the twist angle tends to decrease [4]: this is probably what happens when RNA-polymerase binds to the molecule; we know furthermore how a decreasing of the twist is a necessary condition for the opening of the chain, so that maybe an opening could be directly induced by the bending deformation in the promoter region. We introduced the twist angle becauses it describes the molecule helicity and we consider it as the degree of freedom that is coupled in the most direct way to the opening deformations; at the same time we have then to keep in mind that it is influenced by deformations in the three dimensional configuration as bending and supercoiling, so that it could perhaps be used in the future in such a way to contain informations on some three dimensional properties.

The two degrees of freedom per site in the improved model are the radius r_n of the base pair, which is related to the opening, and the angle φ_n defined with respect to an external fixed reference frame (Figure 4). The twist is then defined by the difference between neighboring base pair angles. The kinetic energy term in the Lagrangian will be written in these polar coordinates.

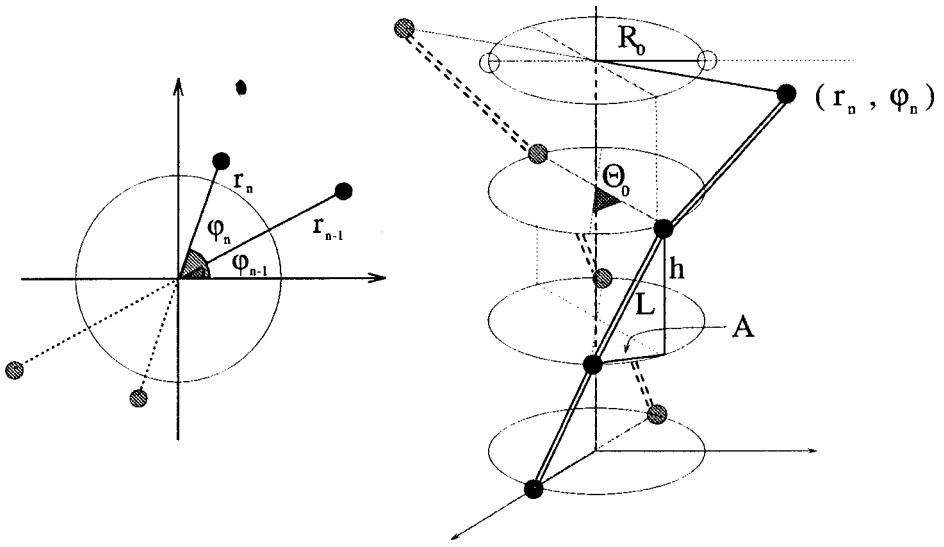

Figure 4. Choice of variables (left) and the schematic view (right) of the model. The n-th base-pair is represented by its radial and angular displacements (r_n, φ_n) with respect to a fixed external reference frame. Neighboring base-pairs are linked along the two strands by elastic rods with equilibrium length L. R_0, Θ_0 and h are the geometrical parameters of the model, with values deduced from the real B-DNA geometry.

As in the PB model we describe the hydrogen bonds by a Morse potential, with depth $D = 0.04$ eV and width $\alpha = 4.45$ Å^{-1}, depending on the stretch variable $(r_n - R_0)$, $D\big(\exp[-\alpha(r_n - R_0)] - 1\big)^2$.

To obtain the correct coupling between radii and angles, we have then explicitly referred to the helical geometry. The helical shape of DNA is essentially due to the competition between two forces: the stacking interaction and the backbone rigidity. The stacking interaction arises from the hydrophobic character of bases: this tends to eliminate water from the core of the molecule by bringing the base pairs close to one another. At the same time bases along the same strand are connected by a sugar phosphate backbone segment which is quite rigid, so that their distance along the strand is almost fixed: to bring base pairs closer it is then necessary to rotate them, inclining the backbone in the helical structure [4]. To reproduce this effects we allow the base pairs to move on planes whose distance h is fixed at the approximate B-DNA base pair step, $h = 3.4$ Å; then we describe the link between neighboring bases along the strands, due to the rigid backbone segments in between, by elastic rods. We then fix the elastic rod length to a value L greater then h, in such a way that the rods have to be inclined in their equilibrium positions (Figure 4). The constant L is chosen as in B-DNA

$$L = \sqrt{h^2 + 4R_0^2 \sin^2\left(\frac{\Theta_0}{2}\right)} > h \tag{2}$$

and the resulting geometry is such that the corresponding rotation of neighboring base pairs is $\Theta_0 = 36°$, i.e. the B-DNA equilibrium twist value. The elastic rods have in a generic, distorted configuration, lengths equal to

$$\sqrt{h^2 + r_{n-1}^2 + r_n^2 - 2r_{n-1}r_n \cos(\varphi_n - \varphi_{n-1})} \tag{3}$$

and correspondingly, we introduce in the Lagrangian of the system quadratic terms depending on the differences between these lengths and L.

Because there is no constraint on the direction of the rotation for each rod, we could have at this point a zig-zag behavior with the rods randomly inclined left- or right-handed; to have a true helical shape, we have then to add a curvature term which essentially constrains each twist angle to have almost the same value as that of the previous one.

Our final Lagrangian [1] then reads

$$\begin{aligned}\mathcal{L} = &\textstyle\sum_n \left(m\dot{r_n}^2 + m{r_n}^2\dot{\varphi_n}^2\right) - D\left(e^{-\alpha(r_n - R_0)} - 1\right)^2 \\ &- \sum_n K\left(\sqrt{h^2 + r_{n-1}^2 + r_n^2 - 2r_{n-1}r_n \cos(\varphi_n - \varphi_{n-1})} - L\right)^2 \\ &- \textstyle\sum_n G_0 \left(\varphi_{n+1} + \varphi_{n-1} - 2\varphi_n\right)^2 .\end{aligned} \tag{4}$$

For the elastic constants K and G_0 we use the values $K = 1$ eV Å^{-2} and $G_0 = K R_0^2/2$. We have to stress that, as we shall discuss again in the conclusions, at this stage this choice is quite arbitrary and it is useful just in the context of this first study. A more appropriate choice will have to be made after all the properties of the model will be completely characterized, if one wants to refer to real biological features.

We have now a model whose equilibrium state is the helix. Furthermore, we obtain, as a consequence of the introduced geometrical constraint, a good coupling between the two degrees of freedom. It is in fact easy to understand that trying to stretch a single base pair, the neighboring pairs are pulled by the rods in such a way that the twist angles tend to decrease, as expected from the geometrical considerations sketched in the introduction; if instead we rotate the neighboring base pairs in opposite direction, bringing them towards a common vertical plane, the rigidity of the backbone rods acts on the direction of pushing the central base out of its equilibrium position, breaking the hydrogen bonds.

This can be easily verified by numerical integration of the equation of motion of the system with special initial conditions. In all the graphs presented in Figure 5 we report the difference between the radial, angular and twist variables with respect to the equilibrium, and we show the time evolution, performed by numerical integration of the equations of motion derived from Lagrangian 4, of a chain initially at rest with the exception of a small region in which the radius is stretched (Figure 5a) or the twist is decreased (Figure 5b). In the first case the

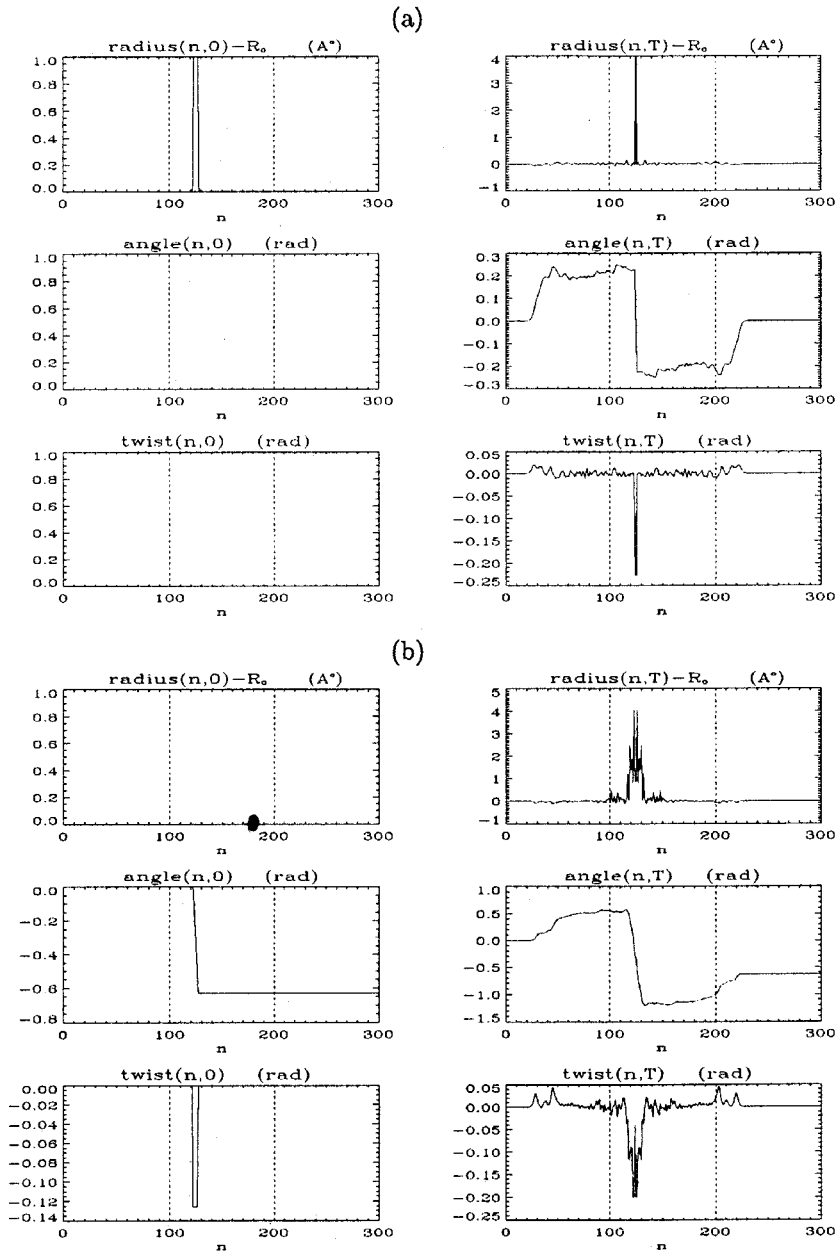

Figure 5. Simple initial distortions and consequent model behavior. We integrate the complete equations of motion of the system with, as initial conditions: (a) a local radial stretched region, (b) a local untwisted region. The system configuration is then shown after integrating for 120 time units (1 t.u. = 1.02 10^{-14} s). The expected untwist effect of an initial hydrogen bond stretching is clearly visible in (a); in (b), the coupling between opening and untwisting is evident.

deformation induces a rotation of all the angles of a positive value upstream and of a negative value downstream the initial distortion, so that there is a reduction of twist in correspondence to the stretched region. In the case of an initial untwisted region, we see how a positive distortion in the radial variables is induced.

3. Small amplitude approximate solutions of the model

The existence of moving breathers in the planar PB model suggests to look for analogous solutions in the generalized helicoidal model.

In order to look for soliton-like solutions we need an extension of a mathematical tool, namely the multiple scale expansion technique, to the case of more than one degree of freedom per site, which is not a trivial point. This can be made in a general way, as explained in [2]. The aim is the derivation of small amplitude soliton-like solutions whose envelopes obey a Non-Linear Schrödinger (NLS) equation: the central point of the derivation is to look for a wave packet solution where a weak dispersion effect is balanced by a weak nonlinearity. The weakly dispersive wave packet is built as a superposition of normal modes whose wave vectors lie in a small interval around a central value q_0, so that one can write for each wave number q contributing to the wave packet $q = q_0 + \epsilon q_1$: the resulting wave packet is then given by the central mode oscillation modulated by a slowly varying envelope function. The expansion in multiple scales consists essentially in introducing sets of independent variables to describe on one hand the slow envelope time-space behavior and on the other the central mode fast oscillatory motion; it can be proved [2] that the same result can be obtained in a more efficient way from the expansion, in small variations ϵq_1, of the linear system that gives, in the wave vector space, the dispersion relations $\omega(q)$ for the normal modes. With respect to the case of systems with only one degree of freedom per site, in the case of vectorial fields the linear part corresponds to an eigenvalue-eigenvector problem, so that one has to perform a perturbative expansion instead of a Taylor series expansion.

Besides this perturbative expansion that treats the dispersive part of the equations of motion, one has to consider small amplitude solutions in order to introduce the nonlinear terms into the equations as successive corrections at increasing orders of accuracy. To combine the two effects of dispersion and nonlinearity one has then to perform two parallel expansions with the common small parameter ϵ; they can be treated in an unique scheme to obtain the NLS equation for the envelope function in the general case of vectorial lattices with nonlinear on-site potentials and eventually, non-cartesian coordinate systems [2].

The general solution one finds with this technique has a main $O(\epsilon)$ wave packet like contribution, proportional to $\exp(q_0 t - \omega(q_0)x)$, q_0 and $\omega(q_0)$ being a chosen central wavenumber and the corresponding frequency on one chosen branch. In addition one has some smaller $O(\epsilon^2)$ terms corresponding to the eigenvector perturbative correction and to the d.c. and second harmonic contributions produced by the nonlinearity. According to [2] this solution can be written in the form

$$\vec{E}(n) = \epsilon e^{i(q_0 x - \omega(q_0)t)}(\vec{V} - i\epsilon \vec{V}^{(1)} \frac{\partial}{\partial(\epsilon x)}) A(\epsilon x, \epsilon t, \epsilon^2 t)$$

$$+\epsilon^2 e^{2i(q_0 x - \omega(q_0)t)} \vec{\gamma}(\epsilon x, \epsilon t, \epsilon^2 t) + \epsilon^2 \vec{\mu}(\epsilon x, \epsilon t, \epsilon^2 t) \; + c.c. \tag{5}$$

Here $\vec{V}$ and $\epsilon \vec{V}^{(1)}$ are respectively the central mode eigenvector and its first order perturbative correction. The slowly varying amplitudes $\vec{\gamma}$ and $\vec{\mu}$ can be solved as functions of A which is the amplitude for which the NLS equation holds.

With this extended technique we are able to solve for approximate analytical solutions, well stable for sufficiently small amplitudes, which represent the generalization of the breather solutions of the planar model [1]. We first rescale in an appropriate way the system variables passing to adimensional variables that represent small displacements with respect to the equilibrium configuration:

$$y_n = \alpha(r_n - R_0) \tag{6}$$

$$\phi_n = \alpha R_0(\varphi_n - n\Theta_0) \tag{7}$$

with a rescaled time $t = \sqrt{D\alpha^2/m}\; t$; then we approximate the Lagrangian rod terms and the Morse potentials respectively to the second and to the fourth order in y_n and ϕ_n so that, finally, the approximated equations of motion can be written as

$$\ddot{y}_n = \left(1 + \frac{y_n}{R_0}\right)\frac{1}{R_0}\dot{\phi}_n^2 - (y_n - \frac{3}{2}y_n^2 + \frac{7}{6}y_n^3)$$

$$-K_{yy}\check{\mathrm{e}}(y_{n+1} + y_{n-1} + 2y_n) - \frac{K_{y\phi}}{2}(\phi_{n+1} - \phi_{n-1}) \tag{8}$$

$$\ddot{\phi}_n = -\frac{2}{R_0}\dot{y}_n\dot{\phi}_n - \frac{2}{R_0}y_n\ddot{\phi}_n - \frac{2}{R_0^2}y_n\dot{y}_n\dot{\phi}_n - \frac{1}{R_0^2}y_n^2\ddot{\phi}_n$$

$$+K_{\phi\phi}\check{\mathrm{e}}(\phi_{n+1} + \phi_{n-1} - 2\phi_n) + \frac{K_{y\phi}}{2}(y_{n+1} - y_{n-1})$$

$$-G(\phi_{n+2} + \phi_{n-2} - 4\phi_{n+1} - 4\phi_{n-1} + 6\phi_n) \tag{9}$$

where

$$K_{yy} = [KR_0^2/(D\alpha^2 L^2)](1 - \cos\Theta_0)^2 \tag{10}$$

$$K_{\phi\phi} = [KR_0^2/(D\alpha^2 L^2)](\sin^2\Theta_0) \tag{11}$$

$$K_{y\phi} = 2[KR_0^2/(D\alpha^2 L^2)](\sin\Theta_0)(1 - \cos\Theta_0) \tag{12}$$

$$G = G_0/(D\alpha^2 R_0^2)\,. \tag{13}$$

We have applied the results of [2] to these equations of motion, looking for a general envelope soliton solution. Their special form is such that an $O(\epsilon)$ d.c. second component exists, so we have added a term of the type $\epsilon\vec{\sigma}(\epsilon x, \epsilon t, \epsilon^2 t)$ to the general solution (5).

The final solution we have derived for the small amplitude approximate excitations are again of the form of a breather in the radial variables, which is now accompanied by a non oscillating, $O(\epsilon)$ kink-like angular distortion, depending on the spatial integral of the squared modulus of the envelope function:

$$y(x,t) = \epsilon(V_1 - i\epsilon V_1^{(1)}\frac{\partial}{\partial(\epsilon x)})Ae^{i\theta} + c.c. + \epsilon^2\gamma_{1_c}A^2e^{2i\theta} + c.c. + \epsilon^2\mu_{1_c}|A|^2$$

$$+O(\epsilon^3) \tag{14}$$

$$\phi(x,t) = (\epsilon\sigma_c + \epsilon^2\mu_{2_c})\int |A|^2 d(\epsilon x) + \epsilon(V_2 - i\epsilon V_2^{(1)}\frac{\partial}{\partial(\epsilon x)})Ae^{i\theta} + c.c.$$

$$+\epsilon^2\gamma_{2_c}A^2e^{2i\theta} + c.c. + O(\epsilon^3). \tag{15}$$

Here γ_{1_c}, γ_{2_c}, μ_{1_c}, μ_{2_c}, σ_c are constants, $\theta = (q_0x - \omega(q_0)t)$ and the envelope function A is

$$A(x,t) = \mathcal{A}\ \mathrm{sech}[\frac{\epsilon}{L_e}(x - V_et)]\exp[i\frac{u_e}{2P}(x - V_ct)] \tag{16}$$

with

$$\mathcal{A} = \sqrt{\frac{u_e^2 - 2u_eu_c}{2PQ}} \tag{17}$$

$$L_e = \frac{2P}{\sqrt{u_e^2 - 2u_eu_c}} \tag{18}$$

$$V_e = V + \epsilon u_e \tag{19}$$

$$V_c = V + \epsilon u_c\ , \tag{20}$$

u_e and u_c are arbitrary (small) constants, V is the group velocity for the wave packet and P, Q are the dispersion and nonlinearity NLS equation parameters [1].

Figure 6 shows the time evolution of the theoretical solutions (14), (15) in the case of a wave packet central mode with $q_0 = 0.1$ and with the corresponding frequency chosen on the upper dispersion branch. The solution is shown in the original unrenormalized variables in a chain of 300 bps. The total time interval correspond to a few internal oscillations. The spatially localized, oscillating radial solution, and the coupled, non oscillating angular part can be recognized. There

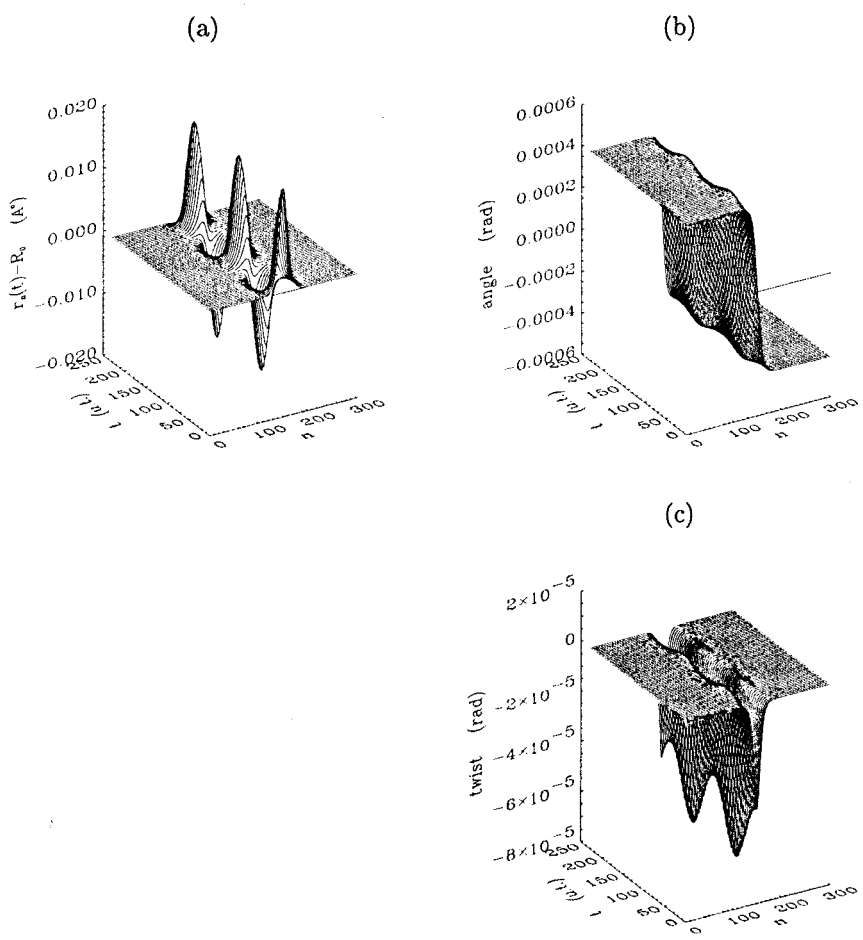

Figure 6. Analytical solution (14), (15). We show here (a): the radial displacement $r_n(t) - R_0$, obtained with the model parameters indicated below and with solution parameters $\epsilon = 0.1$, $u_e = -0.1$, $u_c = 0.1$. The typical breather oscillation is clearly visible; its period is of 105.1 time units. The amplitude $2\epsilon\mathcal{A}/\alpha$ and the half height width, which is about $2.77L_e/\epsilon$, are respectively 0.016 Å and 20 b.p. (b): the corresponding angular displacement $\varphi_n(t)$. The main contribution to this solution is the kink-like term $\epsilon\sigma_c L_e \mathcal{A}^2 \tanh(\eta(x - V_e t))$ of (15). (c): the twist angle, given by the difference between neighboring angles, $\Delta\varphi_n(t)$. We stress that the breather opening mode in the radial variables corresponds to an untwisted region.

are, in addition, smaller oscillations corresponding to the next orders corrections. Other details are reported in the figure caption.

As the twist is defined as the difference between neighboring base pairs angles φ_n, it is immediately clear that the kink-like deformation in the angles corresponds to an untwist, that moves together with the breather along the chain: we can see this negative twist distortion in Figure 6c where the differences $\varphi_n - \varphi_{n-1}$ have been plotted. Since the radial breather represents in average an opening of the helix, the

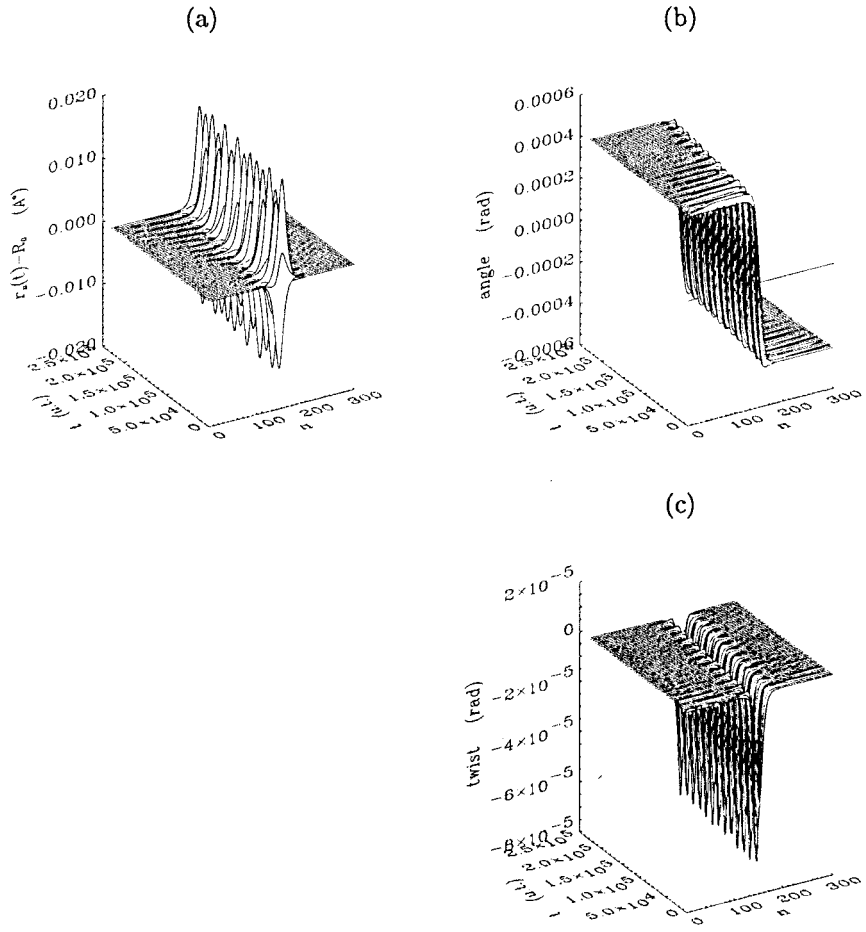

Figure 7. Numerical integration (obtained with a Runge Kutta algorithm) of the initial condition which corresponds to the analytical solution of Figure 6. The amplitude and width parameters preserve their initial values $2\epsilon_e A \sim 0.016$ Å and $2.77 L_e/\epsilon \sim 20$ b.p. with good accuracy, as well as the oscillation period. The total simulation time corresponds to 2234 oscillation periods. We introduce here absorbing boundary conditions to avoid noise effects produced by reflection of the small amount of radiated energy (less then 10^{-5} of the total energy). (a): radial displacement $r_n(t) - R_0$; (b): angular displacement $\varphi_n(t)$ generated by the numerical integration of the same initial conditions; (c): twist angle.

fact that the twist is decreased in the breather core, confirms that our model leads to a correct coupling between radii and angles, inducing a behavior that fits quite well with the known DNA geometrical effects discussed in the introduction. The result even shows how the localized excitations of the planar PB model can actually be extended to the more realistic helicoidal model.

If, as in the case reported in Figure 6, we deal with small enough amplitudes, the analytical solution can be used as an initial condition for a direct integration of

the real equations of motion of the model, derivable from the complete Lagrangian (4). We report in Figure 7 the result of this integration with initial conditions which correspond to the analytical solution of Figure 6. This shows that the derived analytical functions are actually quasi-solutions of the model, in the sense that they maintain their stability for long times with a very small loss of energy: in the case reported in Figure 7 the total integration time is of more than $2\,10^3$ periods of oscillation, and the total energy loss is less then 10^{-5} of the total initial excitation energy.

Going to bigger amplitudes, the analytical solutions are no longer almost exact solutions of the model. They tend to radiate energy and stabilize to a smaller solution of the type described in [1]. It is however important to notice that the model does sustain localized solutions, and some preliminary tests starting from a broad initial excitation have even shown that energy tends to self-localize in the system.

4. Conclusions

We described here a geometrical DNA model which aims at providing a rather realistic representation for the structural and dynamical behavior of the helix while staying simple enough to allow the study of long DNA segments which would not be possible within the molecular dynamics simulations of a full DNA model. We then showed the results of an analytical study of the dynamical properties of the model, solving in particular for approximate small amplitude envelope soliton solutions, and discussed how their shape could be well interpreted by the known helical constraints.

These solutions can be considered as a generalization of the breather solutions of the planar PB model, for which there are accurate studies regarding their formation in the presence of a thermal bath or by modulational instability, their mutual interactions in collisions, their interactions with chain local inhomogeneities. One possible future direction for our work is to study further the breather/kink solutions to see if they can be thermally induced, how they interact, if they could be trapped by local inhomogeneities; more generally, we will study in more details their dynamical behavior.

Besides this, we can focus our attention on the possible features arising from the geometrical properties of our two degrees of freedom model, considering e.g. large amplitude distortions, or the possibility of different propagating solutions. *We found, for example, a different kind of distortion,* that looks quite stable and could be built in a numerical way starting from special initial conditions: it is a localized overtwist (but also a symmetrical undertwist) propagating along the chain together with a very small radial deformation, which is not oscillating but with constant profile. *This solution does not belong to the same class as the breather/kink solution discussed above because it is topologically different.* It can perhaps also be obtained analytically as an approximated solution of the equations of motion.

Clearly the model, with one more degree of freedom per site, is richer than the PB model because it allows a different combination of distortions which could propagate along DNA. Furthermore one can consider the fact that our model allows transitions from a right-handed to a left-handed form to investigate the effects of Z-DNA segments on the transcription process, or use the known relation which links the various forms of coiling of the molecule to try to take into account, at least in an indirect way, the supercoling effects.

Before going on with these projects, we have however to improve our model for what concerns the choice of the various parameters. While the parameters of the Morse potential can be quite well chosen, the elastic constant K and the curvature constant G_0 are not easy to determine. For what concerns the choice of the constant G_0 the main difficulty arises from the fact that the corresponding Lagrangian curvature term is zero for homogeneous twist distortions, i.e. for distortions equally distributed along the chain, that is always the case in today‘s experimental works on the elastic properties of the molecule [18, 19, 20, 21]. It is true, on the other hand, that for the same reason this term is not too relevant as long as we deal with sufficiently smooth distortions. The elastic constant K is instead more important, and we can refer to different experimental and theoretical data but each one probes one aspect of the properties of DNA and they lead to different values of K so that we don’t have yet a well determined value for K. Numerical experiments could be performed on the model to make comparisons with biological and experimental data (in particular thermal denaturation) in order to chose more accurately this parameter, before going on in studying the properties of our model.

Acknowledgements

We gratefully thank Dr Pietro Liò for his helpful collaboration.

References

1. Barbi, M., Cocco. S. and Peyrard, M.: Physics Letters A **253** (1999), 358.
2. Cocco, S., Barbi, M. and Peyrard, M.: Physics Letters A **253** (1999), 161.
3. Travers, A.A.: *DNA-Protein Interactions*. Chapman and Hall, London, 1993.
4. Calladine, C. and Drew, H.: *Understanding DNA*. Academic Press, London, 1992.
5. Lyubchenko, Y.L., Shlyakhtenko, L., Aki, T. and Adya, S.: *Nucleic Acid Research*, **25** (1997), 873.
6. Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. and Buc, H.: *Embo J.* **8** (1989), 4289.
7. Gartenberg, M.R. and Crothers, D.M.: *J. Mol. Biol.* **219** (1991), 217.
8. Kim, J. and Shaphiro, D.J.: *Nucleic Acid Research* **24** (1996), 4341.
9. Pareh, B. and Hatfield, G.W.: *Proc. Natl. Acad. Sci. USA* **92** (1996), 10550.
10. Rigaud, G., Roux, J., Pictet, R. and Grange, T.: *Cell*, 1991.
11. Brahms, G., Brahms, S. and Magasanik, B.: *J. Mol. Biol.* **246** (1995), 35.
12. Albert, A.C., Spirito, F., Figueroa-Bossi, N., Bossi, L. and Rahmouni, A.R.: *Nucleic Acid Research* **24** (1996), 4289.
13. Peyrard, M. and Bishop, A.R.: *Phys. Rev. Lett.* **62** (1989), 2755.

14. Dauxois, T. and Peyrard, M.: *Phys. Rev. E* **51** (1995), 4027.
15. Remoissenet, M.: *Phys. Rev. B* **33** (1996), 1835.
16. Ting, J.L. and Peyrard, M.: *Phys. Rev. E* **53** (1996), 1011.
17. Bang, O. and Peyrard, M.: *Phys. Rev. E* **53** (1996), 4143.
18. Marko, J.F. and Siggia, E.D.: *Macromolecules* **27** (1994), 981.
19. Marko, J.F. and Siggia, E.D.: *Science* **265** (1994), 506.
20. Smith, S., Cui, Y. and Bustamante, C.: *Science* **271** (1996), 795.
21. Strick, T.R., Allemand, J.F., Bensimon, D., Bensimon, A. and Croquette, V.: *Science* **271** (1996), 1835.