Nonlinear waves in double-stranded DNA

Natalia L. Komarova[1,2] and Avy Soffer[1,2]

[1] Department of Mathematics, Rutgers University, 110 Frelinghuysen Rd., Piscataway, NJ 08854S

[2]Institute for Advanced Study, Einstein Dr., Princeton, NJ 08540

We propose a nonlinear model derived from first principles, to describe bubble dynamics of DNA. Our model equations include a term derived from the dissipative effect of intermolecular vibrational modes. Such modes are excited by the propagating bubble, and we term it "curvature dissipation". The equations we derive allow for stable pinned localized kinks which form the bubble. We perform the stability analysis and specify the energy requirements for the motion of the localized solutions. Our findings are consistent with properties of DNA dynamics, and can be used in models for denaturation bubbles, RNA and DNA transcription, nucleotide excision repair and meiotic recombination.

Keywords: Denaturation bubble, transcription, soliton, kink, dissipation.

1 Introduction

Recent experimental breakthrough works [Altan-Bonnet et al, 2003], [Wang et al, 1998], [Poglitsch et al, 1999], [Ashkin, 1997], [Hansma, 1996], [Koch and Wang, 2003] on DNA strand separation in transcription, denaturation and other processes, have made it possible to develop a detailed understanding of such fundamental steps of life (e.g. protein production from the DNA template, see below). There has been a parallel extensive work on theoretical models to describe the corresponding "bubble" dynamics. One line of research concentrated on the nonlinear nature of the denaturation opening, see e.g. [Englander et al, 1980], [Fedyanin et al, 1986], [Yakushevich, 1989], [Muto et al, 1990], [Peyrard and Bishop, 1992], [Dauxois et al, 1992], [Bogolubskaya and Bogolubsky, 1994], [Barbi et al, 1999A], [Bhattacharjee and Seno, 2003], [Campa, 2001]. In these models, the denaturation bubble is described either as a breather or a kink of nonlinear Klein-Gordon (NLKG) or of sine-Gordon type equations. Statistical models to include temperature and noise were also developed [Kafri et al, 2000], [Theodorakopoulos et al, 2000]. A number of detailed numerical studies of local DNA opening have been carried out, which are detailed enough to be compared with experimental data on long chains of DNA strands. These include, among others, studies of Lavery and colleagues, by methods of molecular mechanics [Ramstein and Lavery, 1988], [Bernet et al, 1997], and molecular/Brownian dynamics [Briki et al, 1991], [Giudice, 2003]; molecular dynamics simulations of DNA by the group of Beveridge [Beveridge et al, 2004], [McConnell and Beveridge, 2000], [McConnell and Beveridge, 2001] and Langowski [Bussiek et al, 2002], [Lankas et al, 2004]. In a series of works, Schlick and her group developed comprehensive models of DNA dynamics. This ap-

proach uses a bead model of DNA chains, where the energy of the chain depends on twist, stretch, bend and hydrodynamic-mediated inter-base interactions. This system is then assumed to obey the Langevin equation, and both inertial and overdamped cases are studied in detail, [Beard and Schlick, 2002], [Yang et al, 2002], [Ramachandran and Schlick, 1995], [Schlick, 1995].

The simulation of dynamics of DNA on relevant time scales of transcription and folding is extremely computationally extensive. One of the important features is that the relaxation times are at the picosecond level, while we need to follow the dynamics in the millisecond and second level, [Schlick, 2001]. Therefore it is desirable to find a nonlinear equation that governs the bubble dynamics, while taking into account all the physics, including the double stand nature of DNA and sequence dependence, curvature, solvent effects etc.

Our model attempts to understand the bubble in the double stranded DNA using a classical mechanical model, which is in many ways similar to models discussed above. We start by looking at a chain of connected masses, and take the by-layer structure into account. Dissipation of the molecular dynamics in fluid is usually derived from the interactions between the DNA and the polarization and kinematic properties of the water molecules surrounding it. The result is a friction force proportional to the velocity and stochastic term to describe the effect of thermal motion. However, this assumes that the DNA is a point particle, with no shape and internal degrees of freedom. The effect of internal degrees of freedom may also be relevant. It can be shown that coupling of a system to a *conservative* big system leads to a form of dissipation for the small system. In fact, a simple motivating

example was solved exactly by Lamb in 1900, where he showed that coupling an oscillator to an infinite string leads to dissipation. For a review of recent works in this direction see [Soffer, 2001]. We then ask what will be the effect of the internal vibrational and other modes of each base in the DNA. By modeling each such base as a string we derive the leading order effect, and show it is dissipative, yet can not be incorporated in the usual friction terms as it is curvature dependent. Hence we refer to this contribution as *curvature dissipation*. Just like the usual friction it depends only on the first derivative in time. On the other hand it also contains the derivatives of the amplitude with respect to x, the position along the chain.

It is interesting to note that curvature effects are sometimes relevant to friction, see e.g. [Lighthill, 1975], [Wiggins and Goldstein, 1998], [Goldstein et al, 2000]. However, the kind of dissipation considered in these and other works comes from the interaction of a curved object with the surrounding fluid. The curvature dissipation introduced in this paper comes from the *internal* motions of the molecule and would enter the equation of motion even if the motion took place in vacuum.

Our main result is an equation of motion for a double-stranded DNA which allows for stable, pinned, localized solutions. These solutions (a kink and an anti-kink) can be used to model denaturation bubbles in many biological systems. At this first stage of developing a new theory, we addressed the following questions: What are the static parameters of denaturation bubbles (their steepness and their longitudinal size)? How much energy does a bubble require to be moved along the DNA chain (or, alternatively, how difficult is it to keep it in place?) What defines the direction of motion of a bubble?

How can a bubble collapse?

This paper is organized as follows. We start by presenting several biological systems where denaturation bubbles play an important role (Section 2). In Section 3 we outline the derivation of the new equation of motion for a double-stranded DNA. In Section 4 we find relevant solutions of this equation and study their stability; we prove that our model supports *stable, pinned* localized solutions. We also present numerical stability results for a more complicated, spatially inhomogeneous system. In Section 5 we discuss properties of our model in the context of several biological scenarios. We present model predictions on the size and shape of the bubble, energetic requirements for bubble motion, directionality of bubble motion, and bubble collapse. We identify the parameters that have to be measured to validate the model. We also define the relative importance of curvature dissipation. Section 6 is reserved for conclusions.

2 Denaturation bubbles in biological systems

Here we list several examples of biological phenomena where denaturation bubbles are essential. It is remarkable that bubbles are found at the very basis of life: reproduction (both mitosis and meiosis) and protein synthesis.

RNA transcription. A denaturation bubble plays the central role in RNA transcription, the first step in protein synthesis [Alberts et al, 2002], see figure 1. The process begins when an RNA polymerase (RNAP) enzyme molecule binds to the promoter sequence of the DNA. It starts the transcription by opening up a local region of about 20 base-pairs on the double helix

to expose the nucleotides. One of the two strands serves as a template for complimentary base-pairing with incoming monomers, which begin an RNA chain. The RNAP molecule then moves stepwise along the DNA, unwinding the DNA helix just ahead to expose a new region for base-pairing, and rewinding the region just behind. In this process, a short region (about 8-9 base-pairs) of DNA-RNA helix is formed briefly, after which the newly-built region of the single-stranded DNA molecule is released to allow the rewinding of the DNA-DNA helix. The rate of transcription at 37 C is about 30 nucleotides per second. A typical size of a completed RNA chain is between 70 and $10,000$ nucleotides.

Nucleotide excision repair. A denaturation bubble plays central role in a repair process called *nucleotide excision repair*. There, a damaged site of the DNA is recognized, and then a bubble is formed around it, which is about 25 [Evans et al, 1997] or 20 [Mu et al, 1997] base pairs long for humans, and is shorter ($\sim$ 6 base pairs) in E. coli [Zou and Houten, 1990]. The bubble is created by a helicase which plays a similar part at the initiation of RNA transcription. The repair then proceeds by single strand incision at both sides of the lesion, a removal of the damaged part from the bubble area, DNA repair synthesis to replace the gap and ligation of the remaining single stranded nick.

Other biological systems. A DNA bubble occurs in a variety of situations besides transcription and nucleotide excision repair. An expanding bubble is formed at DNA replication. Also, we will mention the process of meiotic recombination, where a type of a helicase (RecBCD) propels itself

along the DNA [Bianco and Kowalczykowski, 2000] creating a bubble, until a recognition site is encountered, where the traveling loop of DNA is cut, which initiates the genetic recombination event.

3 The new equation: motivation and derivation outline

The motion of the double strand is usually modeled by means of some non-linear equation of the form,

$$\dot{z} = bz_{xx} - \partial V/\partial z + \text{random forcing} + \text{higher order friction terms},$$

where z, the transversal displacement of the nucleotides, is a function of space, x, and time, t, and the dot stand for its time-derivative. The coefficient b is the "spring constant" of the longitudinal interactions modeled as (non)-linear springs, and V is the potential. Different authors proposed various shapes of the nonlinearity corresponding to the hydrogen-bond potential, V, introduced nonlinearity in the "elasticity properties" of the sugar-phosphate backbone, and included extra degrees of freedom coming from the secondary structure of the DNA, as well as chiral forces and torques. The energy terms could be very sophisticated, and often include twist, stretch and bend of the molecule.

One common feature of nonlinear models of DNA dynamics can be identified as follows: they rely on a somewhat *ad hoc* assumption that the coupling between neighboring nodes of the lattice occurs by means of non-material springs. The main point of this paper is to argue that the interaction be-

tween oscillators has a different form and is better described as coupling by "strings", rather than "springs". Roughly speaking, we can say that some of the energy of transversal oscillations of the double strand gets absorbed in the motion of the material connecting the neighboring nodes. The relevant forces are proportional to the momentum, that is, to the time-derivative of the displacement, z. To leading order, this leads to a mixed-derivative term in the master equation,

$$a(\dot{z}_{m-1} - 2\dot{z}_m + \dot{z}_{m+1}).$$

The continuous version of the equation, in its simplest form, will read,

$$\dot{z} = bz_{xx} + a\dot{z}_{xx} - \partial V/\partial z + \text{random forcing}. \tag{1}$$

In equation (1) we assume that the constants, $a, b > 0$. Here we outline the main ideas behind the derivation of equation (1). This is an equation of motion for the node z_m in the direction perpendicular to the molecule, in the overdamped limit, where the influence of the second time-derivative term can be neglected (see the end of Section 5 for the inertial limit). The expression that multiplies the constant b comes from the forces acting on each node from its neighbors due to stretching (the vibrations are not taken into account). The term $-\partial V/\partial z$ comes from the potential forces of interaction of the two units across the double strand. The term $\dot{z}$ represents the usual friction, since the motion takes place in a viscous medium. The third derivative term multiplying the constant a, the curvature dissipation term, reflects the loss of energy due to vibrational modes of the longitudinal connections among the nodes. In what follows, we will outline the derivation of this new term.

In modeling double stranded molecules, one should consider the fact that each longitudinal link is in fact a many-particle molecule, and therefore has a large number of degrees of freedom. Such a molecule should then be described by a dispersive system with many degrees of freedom. In the simplest classical approximation we treat it as a string (rather than a massless spring) of some fixed length. Note that a more general dispersion relation than the usual string will not change qualitatively our analysis.

The critical difference between a "spring" and a "string" is that a string will effectively act as a reservoir which absorbs some of the oscillatory energy. Therefore, we expect the motion of the ends of the links to obey an equation that contains a dissipative correction. Such a correction can, to the leading order, be approximated by the terms with $\dot{z}$. To see the origin of the $\dot{z}$ terms, let us consider the node z_m and solve the wave equation to the left and to the right from it, see Fig. 2. Let $u(x,t)$ denote the position of the string at point x at time t. We have, to the left of the node z_m,

$$u_{tt} - g^2 u_{xx} = 0, \quad u(0,t) = z_{m-1}(t), \quad u(L,t) = z_m(t),$$

where L is the length of the connection and g is the speed of sound in the string. Similarly, to the right from the node z_m we have,

$$\tilde{u}_{tt} - \tilde{g}^2 \tilde{u}_{xx} = 0, \quad \tilde{u}(0,t) = z_m(t), \quad \tilde{u}(\tilde{L},t) = z_{m+1}(t),$$

where the constants do not have to be the same.

The force exerted on the node from the right in the direction perpendicular to the string is proportional to $\tilde{u}_x|_{x=0}$, and the force from the left is proportional to $u_x|_{x=L}$. By examining the solution of the wave equation, we can show that the spatial and temporal derivatives are linearly dependent;

therefore, the force can be defined in terms of $\tilde{u}_t|_{x=0}$ and $u_t|_{x=L}$. In turn, the solution $\tilde{u}(x,t)$ is a linear functional of the boundary conditions, $z_m(t)$ and $z_{m+1}(t)$, thus $\tilde{u}_t|_{x=0}$ is a linear functional of $\dot{z}_m(t)$ and $\dot{z}_{m+1}(t)$, which we denote $G(\dot{z}_m(t), \dot{z}_{m+1}(t))$. Similarly, $u_t|_{x=L}$ is a linear functional of $\dot{z}_{m-1}(t)$ and $\dot{z}_m(t)$. Therefore, the force from the moving springs can be expressed as

$$\tilde{G}(\dot{z}_m(t), \dot{z}_{m+1}(t)) - G(\dot{z}_{m_1}(t), \dot{z}_m(t)).$$

In equation (1) we used a very simple model for G and $\tilde{G}$, where they were just linear functions of their variables. This gave rise to the term $a(\dot{z}_{m-1} + \dot{z}_{m+1}) - 2a'\dot{z}_m$. Setting $a' = a$ and taking the continuous limit, leads to equation (1), which corresponds simply to $G(y_1, y_2) = \tilde{G}(y_1, y_2) = a(y_2 - y_1)$. A complete derivation of the functionals G, $\tilde{G}$ will be presented elsewhere.

4 Localized solutions and their stability

In the literature, the DNA denaturation bubble is often modeled in terms of kinks or breathers. However, both types of localized solutions have several problems [Campa, 2001]. Breathers generically lose stability as the level of discretization of the lattice becomes lower [Aubry, 1997]. Kinks, on the other hand, are very difficult to pin, even on a lattice. As the degree of discretization decreases, the Peierls-Nabarro barrier that keeps a kink from moving decreases exponentially [Willis et al, 1986], [Joos and Duesbery, 1997]. Therefore, a very small amount of energy can set a kink in motion.

In this section, we will describe solutions of equations of type (1) which have properties relevant to many biological systems. Namely, we will study

equation

$$\dot{z} = bz_{xx} + a\dot{z}_{xx} - \partial V/\partial z, \tag{2}$$

and prove that it supports *stable pinned localized solutions*. The effects of random forcing (equation (1)), which is an integral feature of dynamics on the relevant scales, has to be analyzed separately. Stability of localized solutions in the system without random forcing is a necessary condition for their stability once the temperature effects have been added.

Kinks, solitons and the energy functional. The exact shape of a stationary localized solution, $\bar{z}(x)$, is found from the equation

$$b\bar{z}_{xx} - \partial V(z)/\partial z|_{z=\bar{z}} = 0. \tag{3}$$

Note that the nature of the solution $\bar{z}(x)$, will depend on the form of the potential, V, as a function of z. Let us suppose that $V(z)$ is a smooth function. Integrating equation (3) in x, we can see that the quantity $C = b\bar{z}_x^2/2 - V(\bar{z})$ is a constant along the solution for $-\infty < x < \infty$. Using this property, we can see that a *topological kink* (or antikink) solution exists only if the potential, $V(z)$, as a function of z, has at least two minima, say, at points z_1 and z_2, such that $V(z_1) = V(z_2)$, see figure 3; here $z_1 < z_2$ are some real numbers. The kink will satisfy the conditions at infinity,

$$\lim_{x\to-\infty} \bar{z}(x) = z_1, \quad \lim_{x\to+\infty} \bar{z}(x) = z_2.$$

For the antikink, we have

$$\lim_{x\to-\infty} \bar{z}(x) = z_2, \quad \lim_{x\to+\infty} \bar{z}(x) = z_1.$$

A different type of localized solutions is a soliton. A soliton solution is possible whenever the function $V(z)$ has a local minimum, say, at a point $z = z_0$. There will be another point, z_0', with $V(z_0) = V(z_0')$. The soliton solution will have a maximum (or minimum) value of z_0', and the following condition at infinity:

$$\lim_{x\to-\infty} \bar{z}(x) = \lim_{x\to+\infty} \bar{z}(x) = z_0.$$

Let us define distance between functions, $z(x)$ and $v(x)$, as

$$d(z, v) = \int_{-\infty}^{\infty} [(z - v)^2 + (z_x - v_x)^2]\, dx.$$

The energy functional of equation (2), is given by

$$E\{z\} = \int_{-\infty}^{\infty} \left[b z_x^2/2 + V(z) - V_\infty \right] dx. \tag{4}$$

Here, V_∞ is some constant; we subtract this constant in order to make sure that a localized solution has finite energy. For solitons we take $V_\infty = V(z_0)$. For kinks, we have $V_\infty = V(z_1)$, see figure 3. With this choice of the constant, in each case, the function $E\{z\}$ is defined for solutions $z(x, t)$, such that $d(z, \bar{z}) < \infty$, where $\bar{z}(x)$ is the localized solution. Using equation (2), it is easy to show that

$$\frac{dE\{z\}}{dt} = -a \int_{-\infty}^{\infty} z_{xt}^2\, dx - \int_{-\infty}^{\infty} z_t^2\, dx \le 0, \tag{5}$$

which means that starting from any initial conditions (for which $E\{z\}$ is defined), the solution will always decrease the energy functional.

Stability of kinks. Here we will show that kinks are stable. The analysis for antikinks is similar. Let us prove that the energy functional, $E\{z\}$, has

a local minimum at the point $z = \bar{z}$, where $\bar{z}$ is a kink satisfying stationary equation (3). Let us calculate the gradient and the curvature of $E\{z\}$ at $\bar{z}$. We have,

$$\left(\frac{\delta E\{z\}}{\delta z}, \psi\right)_{z=\bar{z}} = \int_{-\infty}^{\infty} \left(b\bar{z}_x\psi_x + \frac{\partial V(\bar{z})}{\partial \bar{z}}\psi\right) dx = 0$$

for all test functions, $\psi(x)$, in the appropriate space. Here and below we use the short-hand notation, $\partial Q(\bar{z})/\partial\bar{z} \equiv \partial Q(z)/\partial z|_{z=\bar{z}}$, where $Q(z)$ is a function of z. Next, we evaluate

$$\left(\psi, \frac{\delta^2 E\{z\}}{\delta z^2}\psi\right)_{z=\bar{z}} = \int_{-\infty}^{\infty} \left(b\psi_x^2 + \frac{\partial^2 v(\bar{z})}{\partial \bar{z}^2}\psi^2\right) dx = \int_{-\infty}^{\infty} (\psi, H\psi)\, dx,$$

where the self-adjoint operator H is given by

$$H = -b\frac{\partial^2}{\partial x^2} + \frac{\partial^2 V(\bar{z})}{\partial \bar{z}^2}.$$

Using Weyl's theorem,see e.g [Reed and Simon,1978], it is easy to show that this operator has a positive continuous spectrum. Indeed, we have

$$\lim_{x\to\infty} \frac{\partial^2 V(\bar{z})}{\partial \bar{z}^2} = V''(z_1) > 0, \quad \lim_{x\to-\infty} \frac{\partial^2 V(\bar{z})}{\partial \bar{z}^2} = V''(z_2) > 0,$$

that is, for large values of x, the potential V'' approaches the value of the curvature at its minima, see figure 3. The continuous spectrum must therefore be positive. The only negative contribution could come from the discrete spectrum. In order to exclude this possibility, let us consider the eigenfunction $\bar{z}_x$, corresponding to the horizontal translation of the kink. This eigenvector corresponds to the eigenvalue zero. Indeed, differentiating equation (3) in x, we obtain $H\bar{z}_x = 0$. On the other hand, this eigenfunction is positive for monotonically increasing kinks. Using Sturm's oscillation theorem (or,

more generally, the Perron Frobenius theory, see e.g. [Sigal,2000]) we conclude that this eigenfunction is the ground state of the operator, which means that all other localized solutions, if they exist, have nonnegative eigenvalues. Therefore, $(\psi, H\psi) \geq 0$.

We have proved that the function $\bar{z}$ is a local minimum of the energy functional $E\{z\}$. Therefore, starting from any solution in a vicinity of the kink, the system will return to the kink. This concludes the stability analysis.

Instability of solitons. The above argument breaks down in the case of solitons. It will remain the same up to the point where we look at the second derivative of $E\{z\}$ at the point $z = \bar{z}(x)$, the soliton solution. The operator H satisfies,

$$\lim_{|x|\to\infty} \frac{\partial^2 V(\bar{z})}{\partial \bar{z}^2} = V''(z_0) > 0,$$

so the continuous spectrum is positive. The minimum of $H(x)$ is negative, because between the points z_0 and z_0' there must be a point, z_* such that $V''(z_*) < 0$, see figure 3. The discrete spectrum must have a negative eigenvalue, because the translational mode with a zero eigenvalue, $\bar{z}_x$, is not a positive function in the case of a soliton. Therefore, by Sturm's oscillation theorem, there will be another eigenfunction (the ground state) with an eigenvalue between $V''(z_*) < 0$ and zero. This suggests that the soliton solution is a saddle point for the energy functional. An infinitesimal perturbation in the "right" direction will destabilize the solution and bring the system to a different stationary state, with a lower energy, e.g. the solution $z(x) = z_0 = const$.

Nonhomogeneous chains: numerical stability results. In the model discussed so far, we treated all base-pairs as if they were identical. A more accurate model for a double-stranded DNA will distinguish between two types of hydrogen bonds, A–T and G–C. The two bonds are characterized by different potentials, namely, $D_{A-T} = 0.05\,eV$ and $D_{G-C} = 0.075\,eV$. In order to model a non-homogeneous DNA sequence, we can use a version of a discretized equation,

$$\dot{z}_m = b(z_{m-1} - 2z_m + z_{m+1}) + a(\dot{z}_{m-1} - 2\dot{z}_m + \dot{z}_{m+1}) - \frac{\partial V_m(z)}{\partial z}\Big|_{z=z_m}, \quad (6)$$

where $z_m(t)$ is the vertical displacement of each nucleotide, and instead of one potential $V(z)$, we have functions $V_m(z)$ which represent interaction between two nucleotides in each base pair. The interaction potentials are allowed to differ from site to site.

We have performed numerical experiments where the functions $V_m \in \{V_{A-T}, V_{G-C}\}$ were taken from some distribution. It appears that the stability properties of the kink are not affected by this type of randomness as long as the values V_{A-T} and D_{G-C} are not too far apart.

Inertial systems. Finally, we consider systems where the motion is not overdamped, such as DNA molecules in non-soluble media. In this situation we need to include the kinematic terms corresponding to elastic modes, $\ddot{z}$. Such cases arise in many applications [Porath et al, 2004], including nanowires made of DNA strands, DNA on dry surfaces, DNA held by electric fields and other nanodevices. In these cases, the propagation of bubbles is governed by the following equation:

$$\kappa\ddot{z} = a\dot{z}_{xx} + bz_{xx} - \dot{z} - \partial\tilde{V}/\partial z + f(x,t), \quad (7)$$

where $f(t)$ is the external force corresponding to the tweezers, electric/magnetic fields etc. Our stability analysis holds almost exactly as before. In the absence of $f(x,t)$, we need to introduce the energy functional,

$$\tilde{E}\{z\} = \int_{-\infty}^{\infty} \left[\frac{1}{2} \left(\kappa z_t^2 + b z_x^2 \right) + V(z) - V_\infty \right] dx.$$

Let $\bar{z}$ be the stationary localized solution, as before. It is clear that if $\bar{z}$ is a local minimum of $E\{z\}$, then it is also a local minimum of $\tilde{E}\{z\}$, since $z_t^2 \geq 0$ and $\bar{z}_t = 0$. Therefore, the stability results for the kink hold in this case.

The potential term in the absence of RNA polymerase is different, and a bubble life-time analysis can be performed using equation (7). In this case, the lifetime may be non-small, due to the absence of stochastic noise.

5 Biological applications

The equation of motion derived here, with a curvature dissipation term, can serve as a starting point to design detailed models of many biological systems where a denaturation bubble plays a role, see Section 2. However, this is not the goal of the present paper. In fact, at this stage we are still quite far from grasping all the features of such complex biological phenomena as RNA transcription, or nucleotide excision repair. For example, in order to describe RNA transcription, a model must contain information on the RNA polymerase molecule. In the present work we are mostly concerned with properties of the double-stranded DNA molecule. A natural question is, what is the value of this modeling for studies of real biological systems?

We will answer this question by using the following analogy. Let us suppose that we need to model a cruise ship. In order to accomplish this task, we need to include all the details of the ship's design. However, no such model would be any good unless we understand the basic properties of water! So a reasonable start for modeling a cruise ship is good old fluid mechanics. Would water be able to hold a massive object without sinking it? Can a (generic heavy) object move along in water? How much energy does such motion take? And so on. In the case of modeling RNA transcription, we first need to understand how double-stranded DNA moves, and how a denaturation bubble forms, before we can begin talking about details of the transcription process itself. The model developed here addresses the following questions: What is the generic shape of a denaturation bubble? What is its size? Can a denaturation bubble be stable? How much energy does it take to move it along the DNA molecule? How can a bubble collapse?

Shape and size of the bubble. A denaturation bubble can be modeled as a solution of (1) which consists of a kink and an antikink, see Fig. 4. If the kink and the antikink are sufficiently far apart, we can say that they do not interact and can coexist for a long time. The width of a kink is roughly given by

$$w = \sqrt{b/\Delta V}, \tag{8}$$

where ΔV is the potential barrier of the interaction energy of nucleotides across the double strand, given by the difference between $V(z)$ at its maximum, and at its minimum.

The longitudinal size of the bubble, n, is given by the distance between

the kink and the antikink. We must require that

$$n \gg \sqrt{b/\Delta V} \tag{9}$$

in order for the bubble to be stable, see Fig. 4. Note that the size of the bubble in this model is not defined by intrinsic properties of the DNA molecule (except for the constraint that a bubble cannot be too small, to satisfy condition (9) above). This means that the bubble size can be different under different circumstances. For instance, in RNA transcription process it is defined by the RNA polymerase molecule. The size of the bubble created in the process of nucleotide excision repair is defined by the appropriate helicase. Finally, the denaturation region formed during DNA replication or meiotic recombination does not have a fixed size, as it is created by a moving helicase which opens up the DNA double helix on one side of the bubble. This suggests that modeling denaturation bubbles as a pair of two independent localized solutions (the kink and the antikink of Fig. 4) is consistent with biological reality, more so than using one localized solution like a breather or a soliton.

In order to relate the model's prediction, equation (8), to biological systems, we need to know numerical values for b and ΔV.

Measurements of "static" parameters of the bubble. The quantities relevant for the shape of the bubble (formula (8)) are given by

$$b = \frac{Kh^2}{D}, \quad \Delta V = \alpha^2 h^2,$$

where D is the depth of the hydrogen bond potential, h is the longitudinal distance between nucleotide pairs, α is the width of the potential well and

K is the "spring constant" of the DNA sugar-phosphate backbone. The first three parameters can be measured relatively accurately, whereas K presents a problem.

The depth of the hydrogen bond potential, D, has been estimated to be $D_{A-T} = 0.05\,eV$ and $D_{G-C} = 0.075\,eV$ for the two types of pairing. The parameter α that defines the width of the potential well is taken to be $\alpha = 2.55\,A$ in [Peyrard and Bishop, 1992], $\alpha = 4.45\,A^{-1}$ in [Barbi et al, 1999A], $\alpha = 4\,A^{-1}$ in [Campa, 2001]. The distance between pairs is $h = 3.4\,A$ ([Campa, 2001] and [Barbi et al, 1999A]).

A more difficult quantity to measure is the "spring constant" K, of the DNA sugar-phosphate backbone.[1] In [Peyrard and Bishop, 1992] it was merely estimated from the model to give a realistic denaturation temperature; the corresponding value is $3.0 \times 10^{-3}\,eV/A^2$. However, the paper by [Kamien et al, 1997] suggests that this value is much larger, the measured parameter is $K = 0.22\,eV/A^2$. An even larger value, $K = 1.0eV/A^2$, is quoted in [Barbi et al, 1999A]. The paper by [Gerland et al, 2001] uses the value $K = 0.026\,eV/A^2$ (however, this value has been estimated for RNA and includes effects of the secondary structure). Note that other experimental measurements give very different values, see [Smith et al, 1996], [Bensimon et al, 1995], where the spring constant is found to be very small, of the order of $10^{-6}\,eV/A^2$. However, it must be noted that in those experi-

[1] Large discrepancies in the values of the spring constant are not surprising. Our models suggests that the "spring" properties of the DNA, that is, the coefficient b in equation (1), is not the entire story. Energy losses due to vibrational modes of the nucleotides have to be taken into account, which can in principle be done by measuring the spectrum of vibrational modes.

ments the spring constant of the DNA molecule as a whole was measured as opposed to local elasticity properties of the sugar-phosphate backbone, and it is the latter quantity which is of interest to us.

With the information that we have so far, we can obtain the value of w between 0.18 (for $K = 0.2\,eV$, $\alpha = 4.45\,A$ and $D = 0.33\,eV$) and 1.96 (for $K = 1\,eV$, $\alpha = 2.55$ and $D = 0.4\,eV$). This means that the number of nodes in the "knee" of the kink is of the order one. This estimate is consistent with the picture of RNA transcription (Section 2) where an RNA polymerase enzyme molecule opens up only a few base-pairs to complete the transcription of a small portion of the DNA template, with 3 or fewer nucleotides forming the "sides" of the bubble.

Energy needed to move the bubble. In potential systems, such as nonlinear Klein-Gordon equation, a whole family of moving kinks, $\bar{z}(x - vt)$, exists for any velocity v. Therefore, moving a kink along a lattice does not take any energy. In the new equation, this is not the case. In order to move the bubble along the DNA molecule, an external force must be applied.

This has relevance for many biological systems involving denaturation bubbles. In the context of RNA polymerase, we can ask: how strong a push does a transcription bubble need to travel along the DNA? The RNAP molecule is thought to be a molecular motor, which uses the energy of ribonucleoside triphosphades to propel itself in the 3'-5' direction along the coding strand of the DNA molecule [Gelles and Landick, 1998]. Our model implies that the RNAP "drags" the transcription bubble (consisting of a kink and and antikink) along, using the appropriate fraction of its total energy. The

same holds for expanding denaturation regions during the process of DNA replication and meiotic recombination. How much energy does a helicase need to propel a traveling loop of DNA?

In the context of nucleotide excision repair, one can ask the opposite question: how stable is the bubble? How easy is it to keep it in place for as long as it takes to perform the repair?

Theoretically we can address these questions in the framework of our model. Using equation (5), we can calculate how much energy it takes to move a kink with velocity v for time Δt:

$$\Delta E = v^2 \Delta t \int_{-\infty}^{\infty} \left(\bar{z}_x^2 + a \bar{z}_{xx}^2 \right) \, dx. \tag{10}$$

We can see that energy losses come from two sources: the first term under the integral is the usual dissipation. The second term is the curvature dissipation, that is, the loss due to internal vibrational modes of the DNA molecule. This is the novel contribution of the present model.

In order to obtain a quantitative prediction, several detailed measurements must be performed. First of all, the shape of the bubble has to be identified, to find the slope, z_x and the curvature, z_{xx} along the bubble. Then, the velocity of motion, v, has to be estimated during a time-interval, Δt. Finally, the contribution of dissipation and curvature dissipation must be identified. This is the most difficult task. Measuring the spectrum of vibrational modes will eventually lead to the information necessary to estimate the coefficient a in equations (1) and (10). This is discussed next.

Direction of the bubble motion. According to our model, the bubble motion direction is defined externally, by the "motor" which propels the

kink along the DNA chain. In biological systems, bubble motion happens in a fixed direction. For example, in RNA transcription, the process of elongation always proceeds in the 5'–3' direction (i.e. the RNA polymerase moves along the template strand of DNA in the 3'–5' direction). Therefore, the "polarity" of the coding strand defines the arrow of motion. Our suggestion is that it is the molecular motor (RNAP) that recognizes the directionality of the DNA strand, and the bubble itself can be moved in either direction.

Relative importance of curvature dissipation. Recent measurements of both intermolecular and intramolecular vibrational modes of nucleotides show their significance for DNA dynamics. In the works of [Lee et al, 2000], [Lee et al, 2001], the Raman spectrum of nucleotides is measured in the range from 200 to 4000 cm^{-1}. These modes correspond to the internal vibrations within the molecule, and they are in the same energy range as the hydrogen bonds between the strands. Moreover, when the measurements are done at low temperature (10–20K) one observes that broad absorption lines are in fact many resonances, fused together due to thermal fluctuations. Other modes correspond to vibrations of two coupled nucleotides; they have a lower energy and therefore are easier to excite. These modes have also been measured and are typically in the range from 30 to 150cm^{-1}. Some of them are measured by [Fisher et al, 2002] in experiments on crystals, and by [Boland and Ratner, 1995] on the molecular level; see also [Olson, 2004]. It is clear from this abundance of the modes at the relevant energy scales, that a realistic temporal description, as required for example for bubble motion in transcription, must adequately incorporate the corresponding contributions.

We can use simple energetic considerations to estimate the effect of curvature dissipation on the DNA dynamics. As the bubble propagates through the DNA, it excites many internal modes. Therefore, we need to compare the energies of the motion of the bubble with the vibrational modes. The number of relevant modes of one nucleotide is multiplied by the number of points where the curvature is not zero, about 6 (that is, 12 nucleotides). Using the fact that each vibrational mode of nucleotides, as well as nucleotide-nucleotide couplings, is of the order of $10^{-3}\ eV$ [Olson, 2004], we can see that each vibrational mode of the 6 involved base-pairs contributes an amount which is about 10-20% of the difference of the base-coupling energies ($0.25\, eV$).

It is now possible to estimate the value of the coefficient a in equation (1). Let us suppose that the bubble moves with a constant speed, v. Then, we assume that it excites 12 nucleotides in each strand, with the total excitation energy e. Then the energy change per unit time is given by $4ev/d$, where d is the size of the kink (say, $d = 3$ base pairs)The factor 4 appears since there are 4 kinks, on two strands per bubble. Our formula for the rate of energy change related to internal modes is given by $av^2 \int \bar{z}_{xx}^2\, dx$. Therefore

$$a \sim \frac{4e}{vd \int \bar{z}_{xx}^2\, dx},$$

where the integral is completely determined by the shape of the kink and has support ($\sim d$) of a few base points. Now, notice that we can eliminate the dependence on v using the fact that E, the total energy is proportional to v^2. Therefore $\Delta E/E = 4ev/dE \sim a \int \bar{z}_{xx}^2\, dx$ where ΔE stands for the energy loss per unit time.

Bubble collapse. Our approach can find applications in modeling transcription termination. There are two ways in which transcription is terminated. ρ-independent termination involves a specific sequence prone to forming a hairpin. ρ-dependent termination requires a subunit of RNAP which utilizes the energy of ATP to stop the transcription. In order to model this, we can use equation (6), a discrete version of the equation of motion where the two different types of hydrogen bonds, A-T and G-C, are taken into account.

The process of ρ-independent termination can be modeled by introducing a large perturbation in the sequence of V_m (hydrogen bonds). Simulations show that a particularly *small* value of V_m at one site can lead to a collapse of the kink and the antikink on each other. The ρ-dependent termination can be modeled by adding a large perturbation somewhere between the kink and antikink. Say, if the value of z outside the bubble is z_1, and it is z_2 inside, setting several (strategically chosen) nodes inside the kink back to the value z_1 may cause a collapse of the kink-antikink pair.

The behavior of our model is in qualitative agreement with reality. At this stage, we can only suggest that the equations of motion that we derived for the dynamics of a double-stranded DNA molecule allow for a bubble collapse if appropriate forcing is applied. A more detailed model, based on particular sequences, must be devised to give quantitative predictions.

6 Conclusions

We have introduced a nonlinear equation of motion describing the dynamics of double-stranded DNA. Along with the usual dissipation term, it contains a curvature dissipation term, corresponding to the loss of energy to the many vibrational modes of the DNA molecule. This equation allows for a localized, pinned solution which can be relevant for modeling DNA denaturation bubble because of the following useful properties:

(i) It is not a breather, that is, its existence does not depend on the fast transversal vibrations;

(ii) It is pinned, that is, it will not travel along the DNA when perturbed in the longitudinal direction; in fact, it requires finite energy to move;

(iii) It is stable, and its stability can be proved rigorously.

There are many biological processes involving DNA denaturation bubbles, such as RNA transcription, nucleotide excision repair, DNA replication and meiotic recombination. When modeling these and other processes, the basic equation of motion for the double-stranded molecule must allow for stable solutions corresponding to local opening of the DNA. In this first paper we have suggested a framework for such modeling.

The dynamical formulation we use, allows the incorporation of other important structural factors. The first thing we need to include is the effect of chain content. This is easily done by making the strength of the coupling between the strands change value according to whether it is GT or TC. We can also include the effect of content by changing the "string" constant as

we move from type base to base along the chain; for this we can use the information on relevant excited modes of each such molecule. We note also that the effect of stacking, which was considered by many authors before (see e.g. [Barbi et al, 1999B], [Barbi et al, 2003]) can be included in a similar way by adding "spin" degrees of freedom. There is no reason to believe that existence and stability results would change in the modified system.

Most importantly, the effect of curvature of the DNA molecule, can also be implemented by making the coefficients of the discrete Laplacian position-dependent. It is not easy to see how this can be done in a nonhamiltonian, energy landscape type models. The implications of these modifications may be very important, and will be studied in a forthcoming work. Here we only mention that the *curvature effects* play a central role in the DNA dynamics, and can have important consequences for the regulation and the dynamics of the transcription process. For example, when the DNA region is tightly bound, curled around a chromatin, transcription initiation is impossible. But when the curvature is lowered, by the action of appropriate enzymes, the process can begin. Once the process has started, the curvature will affect the velocity of propagation of the bubble. Most importantly, we conjecture that, in some places, it will also change the effective energy landscape, to the point of creating, or moving of *arrest points*. Such points are critical to understanding transcription regulation.

Finally, we would like to describe details of biochemical reaction which include more players. In upcoming papers we will concentrate on the process of RNA transcription and show how the equation of motion for the double-stranded DNA can be coupled with an explicit equation for the RNA

polymerase molecule. Unfortunately, complicated models like this do not often allow for a clear and rigorous mathematical analysis. The advantage of the present model is its transparent behavior. It will serve an a building block for more complicated systems.

References

[Alberts et al, 2002] B. Alberts et al. 2002. Molecular biology of the cell. Taylor & Francic, Inc.

[Altan-Bonnet et al, 2003] G. Altan-Bonnet, A. Libchaber and O. Krichevsky. 2003. Bubble dynamics in double-stranded DNA. PRL. 90: 138101.

[Ashkin, 1997] A.Ashkin. 1997. Optical trapping and manipulation of neutral particles using lasers. PNAS 94: 4853-60.

[Aubry, 1997] S. Aubry. 1997. Breathers in nonlinear lattices: Existence, linear stability and quantization. Physica D. 103: 201-250.

[Barbi et al, 1999A] M. Barbi, S. Cocco, and M. Peyrard. 1999. Helicoidal model for DNA openings. Phys. Let. A. 253: 358-69.

[Barbi et al, 1999B] M. Barbi, S. Cocco, M. Peyrard, and S. Ruffo. 1999. A twist opening model for DNA. Jour. Biol. Phys. 24: 358-369.

[Barbi et al, 2003] M. Barbi, S. Lepri, M. Peyrard and N. Theodorakopoulos. 2003. Thermal denaturation of a helicoidal DNA model. Phys. Rev. E. 68: 061909-23.

[Beard and Schlick, 2002] Beard, D. and Schlick, T. 2000. Inertial stochastic dynamics. I. Long-time-step methods for Langevin dynamics. J. Chem. Phys. 112(17) 7313-22. Beard, D. and Schlick, T. 2000. Inertial stochastic dynamics. II. Influence of inertia on slow kinetic processes of supercoiled DNA. J. Chem. Phys. 112(17) 7323- 38.

[Bensimon et al, 1995] Bensimon, D, Simon, A.J., Croquette, V. and Bensimon, A. 1995. Stretching DNA with a receding miniscus: experiments and models. Phys. Rev. Lett. 74: 4754-7.

[Bernet et al, 1997] Bernet,J., Zakrzewska,K. and Lavery,R. (1997) Modelling base pair opening: the role of helical twist. J. Mol. Struct. (Theochem), 398-399, 473-482.

[Beveridge et al, 2004] Beveridge DL, Dixit SB, Barreiro G, Thayer KM. Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. Biopolymers. 2004, 73(3):380-403.

[Bhattacharjee and Seno, 2003] S. M. Bhattacharjee and F. Seno. 2003. Helicase on DNA: a phase coexistence based mechanism. J. Phys. A 36, L181-7.

[Bianco and Kowalczykowski, 2000] P.R.Bianco and S.C.Kowalczykowski. 2000. Translocation step size and mechanism of the RecBC DNA helicase. Nature. 405:368-72.

[Bogolubskaya and Bogolubsky, 1994] A.A. Bogolubskaya and I.L. Bogolubsky. 1994. Two-component localized solutions in a nonlinear DNA model. Phys. Lett. A. 192: 239-46.

[Boland and Ratner, 1995] Boland, T and Ratner, B.D. 1995. Direct measurement of hydrogen bonding in DNA nucleotide bases by atomic force microscopy. Proc. Nat. Acad. Sci, 92: 5297-5301.

[Briki et al, 1991] Briki,F., Ramstein,J., Lavery,R. and Genest,D. (1991) Evidence for the stochastic nature of base pair opening in DNA: a Brownian dynamics simulation. J. Am. Chem. Soc., 113, 2490-2493.

[Bussiek et al, 2002] Bussiek M, Klenin K, Langowski J. Kinetics of site-site interactions in supercoiled DNA with bent sequences. J Mol Biol. 2002;322(4):707-18.

[Campa, 2001] A. Campa. 2001. Bubble propagation in a helicoidal molecular chain. Phys. Rev. E. 63: 021901-10.

[Dauxois et al, 1992] T. Dauxois, M. Peyrard and C.R. Willis. 1992. Localized breather-like solution in a discrete Klein-Gordon model and application to DNA. Physica D. 57:267-282.

[Englander et al, 1980] S.W. Englander, N.R. Kallenbach, A.J. Heeger, J.A. Krumhansl, and S. Litwin. 1980. Nature of the open state in long polyonucleotide double helices: possibility of soliton excitations. Proc. Natl. Acad. Sci. 77: 7222-6.

[Evans et al, 1997] E. Evans, J. Fellows, A. Coffer, and R.D. Wood. 1997. Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. EMBO Jour. 16:625-38.

[Fedyanin et al, 1986] V.K. Fedyanin, I Gochev and V. Lisy. 1986. Nonlinear dynamics of bases in continual model of DNA double helices. Stud. Biophys. 116: 59-64.

[Fisher et al, 2002] Fischer, B.M., Walther, M. and Jepsen, P.U. 2002. Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy. Phys. Med. Biol. 47: 3807-3814.

[Gelles and Landick, 1998] J. Gelles and R. Landick. 1998. RNA polymerase as a molecular motor. Cell. 93: 13-16.

[Giudice, 2003] Giudice, E., Varnai, P. and Lavery, R.(2003) Base pair opening within B-DNA: free energy pathways for GC and AT pairs from umbrella sampling simulations. Nucleic Acids Research, 2003, Vol. 31, No. 5 1434-1443.

[Goldstein et al, 2000] Goldstein, R.E., Goriely, A. and Wolgemuth, C.W. (2000) Bistable helices, Phys. Rev. Let. 84(7), 1631-34.

[Gerland et al, 2001] Gerland, U., Bundschuh, R. and Hwa, T. 2001. Force-induced denaturation of RNA. Biophys. Jour. 81: 1324-32.

[Hansma, 1996] H.G. Hansma. 1996. Atomic force microscopy of biomolecules. J. Vasc. Sci. Technol. B14: 1390-94.

[Joos and Duesbery, 1997] B. Joos and M.S. Duesbery. 1997. Dislocation kink migration energies and the Frenkel-Kontorowa model. Phys. Rev. B. 55: 11161-66.

[Kafri et al, 2000] Y. Kafri, D. Mukamel and L. Peliti. 2000. Why is the DNA denaturation transition first order? PRL. 85: 4988-91.

[Kamien et al, 1997] R.D. Kamien, T.V. Lubensky, P. Nelson and C.S. O'Hern. 1997. Direct determination of DNA twist-stretch coupling. Europhys. Lett. 38:237-242.

[Koch and Wang, 2003] S.J.Koch and M.D.Wang. 2003. Dynamic force spectroscopy of protein-DNA interactions by unzipping DNA. PRL 91: 028103.

[Lankas et al, 2004] Lankas F, Sponer J, Langowski J, Cheatham TE 3rd. 2004. DNA deformability at the base pair level. J Am Chem Soc. 126(13): 4124-5.

[Lee et al, 2000] S.A. Lee, A. Anderson, W. Smith, R.H. Griffey, V. Mohan. 2000. Temperature-dependent Raman and infrared spectra of nucleosides Part I-adenosine. J.Raman Spectroscopy 31: 891-896

[Lee et al, 2001] S.A. Lee, J. Li, A. Anderson, W. Smith, R.H. Griffey and V. Mohan . 2001. Temperature-dependent Raman and infrared spectra of nucleosides: II, cytidine. J. Raman Spectroscopy 32: 795-802

[Lighthill, 1975] Lighthill, M.J. *Mathematical biofluiddynamics*, Philadelphia : Society for Industrial and Applied Mathematics, 1975.

[McConnell and Beveridge, 2000] McConnell KJ, Beveridge DL. DNA structure: what's in charge? J Mol Biol. 2000, 304(5):803-20.

[McConnell and Beveridge, 2001] McConnell KJ, Beveridge DL. Molecular dynamics simulations of B-DNA: sequence effects on A-tract-induced bending and flexibility. J Mol Biol. 2001, 314(1):23-40.

[Muto et al, 1990] V. Muto, P.S. Lomdahl and P.L. Christiansen. 1990. Two-dimensional discrete model for DNA dynamics: longitudinal wave propagation and denaturation. Phys. Rev. A. 42: 7452-58.

[Mu et al, 1997] D. Mu, M. Wakasugi, D.S. Hsu, and A. Sancar. 1997. Characterization of reaction intermediates of human excision repair nuclease. J. Biol. Chem. 272: 28971-79.

[Olson, 2004] Olson, W. 2004 Private communication.

[Peyrard and Bishop, 1992] M.Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation 1989. Phys. Rev. Lett. 62: 2755-58.

[Poglitsch et al, 1999] C.L. Poglitsch, G.D. Meredith, A.L. Gnatt, G.J. Jensen, W.H. Chang, J. Fu, and R.D. Kornberg. 1999. Electron crystal structure of an RNA polymerase II transcription elongation complex. Cell 98: 791-8.

[Porath et al, 2004] Porath, Cuniberti and Di Felice. 2004. Charge Transport in DNA-based Devices, cond-mat/0403640.

[Prohofsky, 1988] E. W. Prohofsky. 1988. Solitpons hiding in DNA and their possible significance in RNA transcription. Phys. Rev. A. 38: 1538-41.

[Ramachandran and Schlick, 1995] G. Ramachandran and T. Schlick, "Solvent Effects on supercoiled DNA explored by Langevin dynamics simulations", Phys. Rev. E 51. 6188-6203, 1995.

[Ramstein and Lavery, 1988] Ramstein,J. and Lavery,R. (1988) Energetic coupling between DNA bending and base pair opening. Proc. Natl Acad. Sci. USA, 85, 7231-7235.

[Reed and Simon,1978] Reed,M. and Simon,B. (1978) Methods of Modern Mathematical Physics 4, Analysis of Operators. Academic Press,New-York

[Schlick, 1995] Schlick, T. Modeling Superhelical DNA: Recent Analytical and Dynamical Approaches 1995, in Current Opinion in Structural Biology vol 5.2 on "Theory and Simulation", B. Honig (Ed.)

[Schlick, 2001] Schlick, T. Time-trimming tricks for dynamic simulations: Splitting force updates to reduce computational work. Structure 9:R45-R53, 2001.

[Sigal,2000] I.M. Sigal, Perron Frobenius Theory and Symmetry of solutions to nonlinear PDEs,Letters in Math. Phys.53, 313-320 (2000).17

[Smith et al, 1996] Smith, S.B., Cui, Y. and Bustamante, C. 1996. Overstretching B-DNA: the Elastic Response of Individual Double Stranded and Single Stranded DNA Molecules, Science. 271:795-799.

[Soffer, 2001] Soffer, A. 2001. Dissipation through dispersion. CRM Proceedings and Lecture Notes 27: 175-184.

[Takeno and Homma, 1983] S. Takeno and S. Homma. 1983. Topological solitons and modulated structure of bases in DNA double helices. Prog. Theor. Phys. 70:308-311.

[Theodorakopoulos et al, 2000] N. Theodorakopoulos, T. Dauxois, and M. Peyrard. 2000. Order of the phase transition in models of DNA thermal denaturation. PRL. 85:6-9.

[Wang et al, 1998] M.D.Wang, M.J. Schnitzer, H. Yin, R. Landick, J. Gelles, and S.M. Block. 1998. Force and velocity measured for single molecules of RNA polymerase. Science 282:902-7.

[Wiggins and Goldstein, 1998] Wiggins, C.H. and Goldstein, R.E. (1998) Flexive and propulsive dynamics of elastica at low Reinolds number, Phys. Rev. Let. 80(17), 3897-82.

[Willis et al, 1986] C. Willis, M. El-Batanouny and P. Stancioff. 1986. Sine-Gordon kinks on a discrete lattice. I. Hamiltonian formalism. Phys. Rev. B. 33:1904-11.

[Yakushevich, 1989] L.V. Yakushevich. 1989. Nonlinear DNA dynamics: a new model. Phys. Let. A. 136: 413-17.

[Yang et al, 2002] L. Yang, W. A. Beard, S. H. Wilson, B. Roux, S. Broyde, and T. Schlick, "Local Deformations Revealed by Dynamics Simulations of DNA Polymerase with DNA Mismatches at the Primer Terminus", J. Mol. Biol., 321:459-478 (2002)

[Yomosa, 1983] S. Yomosa. 1983. Soliton excitations in deoxyribonucleic acid (DNA) double helices. Phys. Rev. A. 27: 2120-25.

[Ch-T Zhang, 1987] Ch-T Zhang. 1987. Soliton excitations in deoxyribonucleic acid (DNA) double helices. Phys. Rev. A. 35:886-891.

[Zou and Houten, 1990] Y. Zou and B.V.Houten. 1999. Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. EMBO Jour. 18: 4889-901.

Figure legends

Figure 1. A schematic of the RNA transcription process, showing a DNA denaturation bubble.

Figure 2. The geometry of the DNA molecule and the variables used to derive the equation of motion. The nodes represent the nucleotides.

Figure 3. The shape of the potential which allows for a topological kink solution and a soliton solution. Note that in our model, the potential $V(z)$ defines interactions between nucleotides via the hydrogen bond. We can assume that the first minimum, z_1, corresponds to the equilibrium distance between the strands of the double helix. The other minimum, z_2, is one

of many in a complicated potential landscape corresponding to separated strands. In the context of RNA transcription, the complex shape of the potential can be a consequence of the presence of the RNAP as well as the newly-formed RNA chain. It is not unreasonable to assume that at least one of the minima will be sufficiently close to the minimum $V(z_1)$ to guarantee a long-lived topological kink.

Figure 4. A model of a denaturation bubble as a kink-antikink pair.

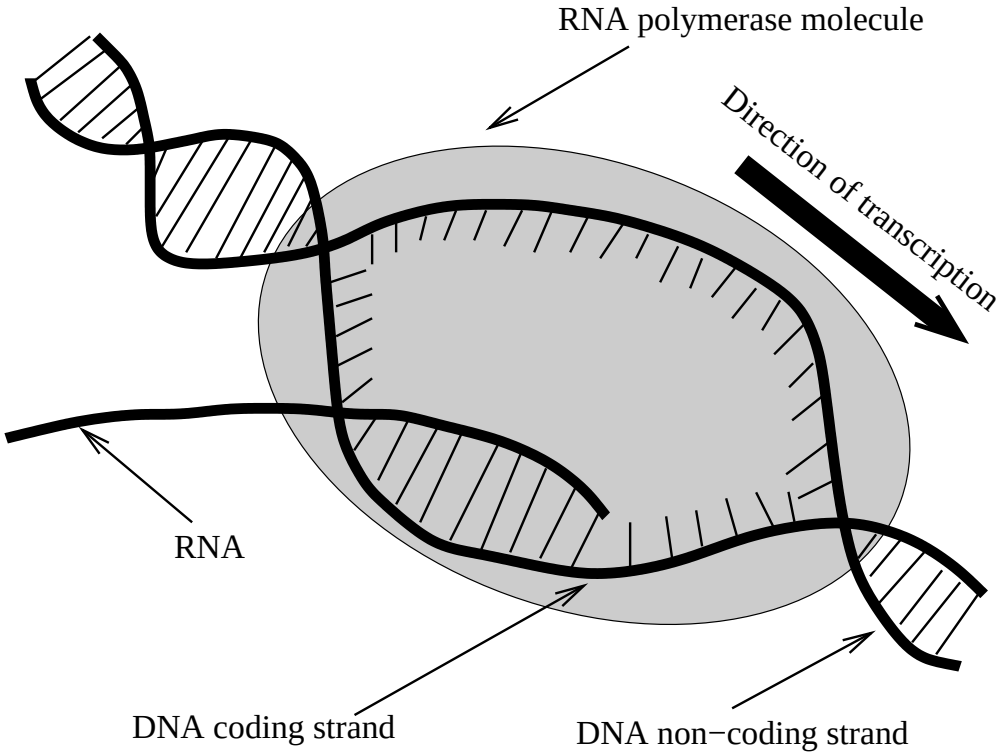
RNA polymerase molecule
Direction of transcription
RNA
DNA coding strand
DNA non−coding strand

Figure 1:

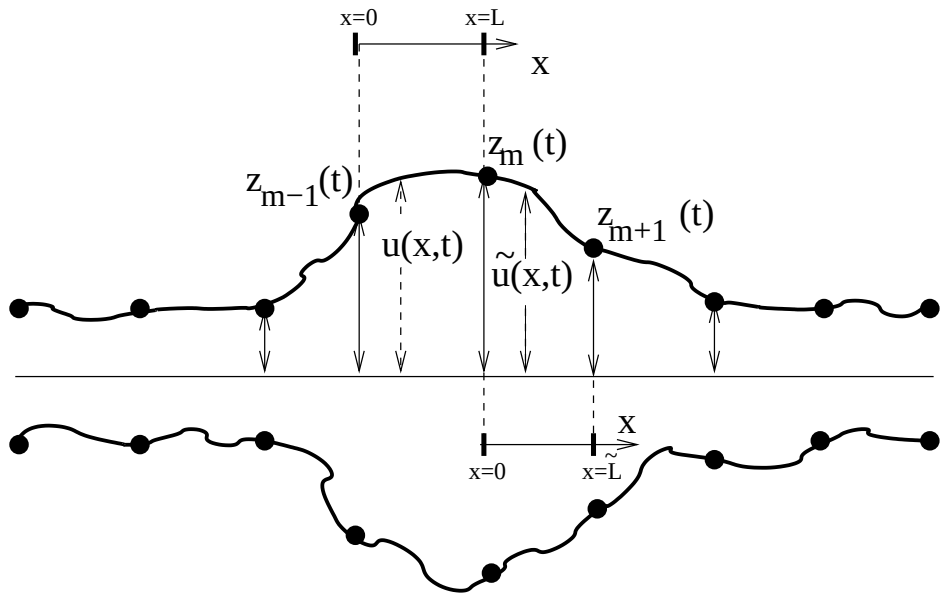

Figure 2:

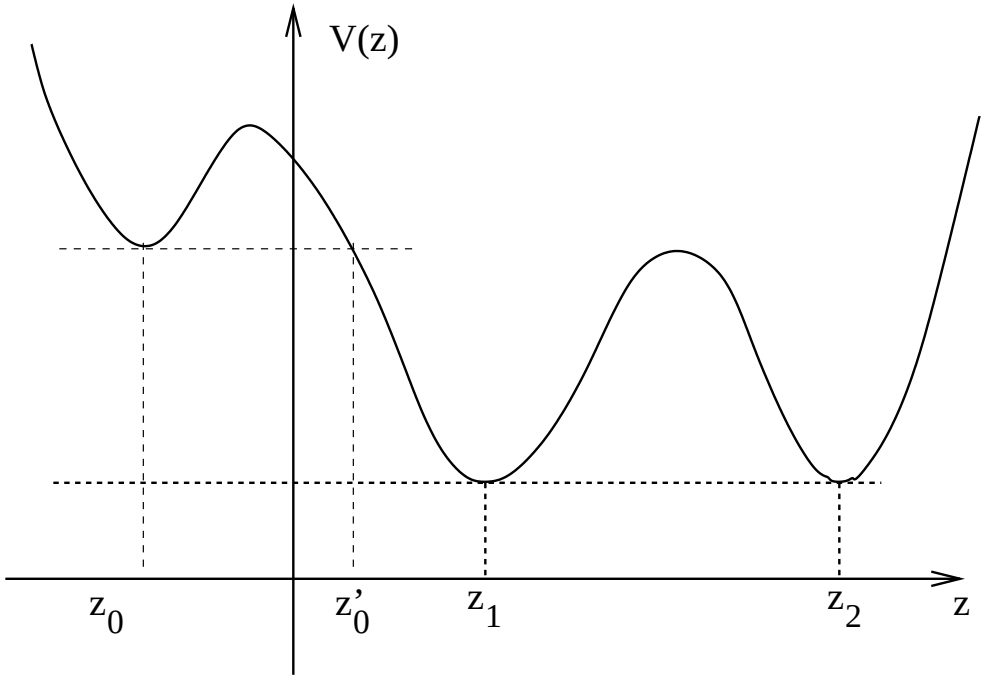

Figure 3:

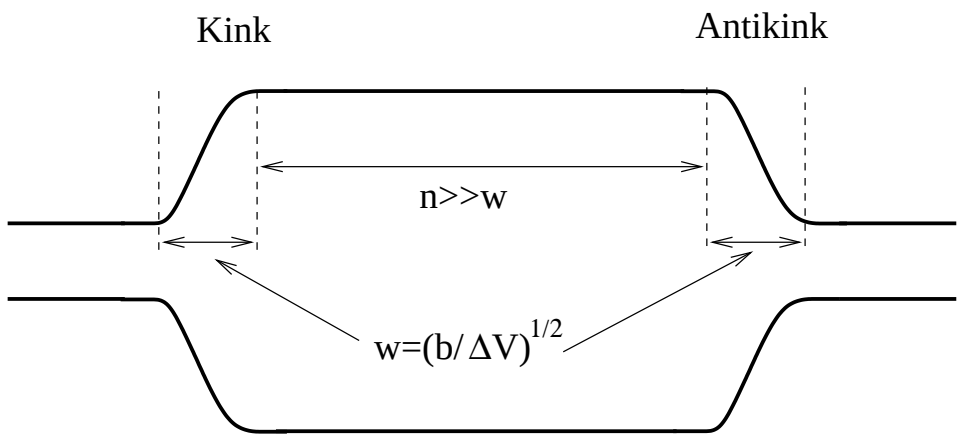

Figure 4: