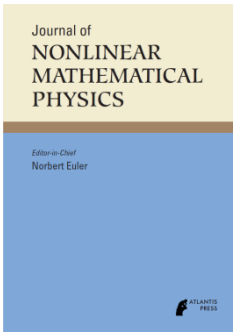

ATLANTIS PRESS

Journal of Nonlinear Mathematical Physics

ISSN (Online): 1776-0852 **ISSN (Print): 1402-9251**
Journal Home Page: https://www.atlantis-press.com/journals/jnmp

The Transfer Integral Operator Method in the Study of DNA Unzipping and Bubble Formation

Z. Rapti, K. Ø. Rasmussen, A. R. Bishop

To cite this article: Z. Rapti, K. Ø. Rasmussen, A. R. Bishop (2011) The Transfer Integral Operator Method in the Study of DNA Unzipping and Bubble Formation, Journal of Nonlinear Mathematical Physics 18:Supplement 2, 381–396, DOI: https://doi.org/10.1142/S1402925111001581

To link to this article: https://doi.org/10.1142/S1402925111001581

Published online: 04 January 2021

ARTICLE

World Scientific
www.worldscientific.com

ATLANTIS PRESS

Journal of Nonlinear Mathematical Physics, Vol. 18, Suppl. 2 (2011) 381–396

DOI: 10.1142/S1402925111001581

THE TRANSFER INTEGRAL OPERATOR METHOD IN THE STUDY OF DNA UNZIPPING AND BUBBLE FORMATION

Z. RAPTI

Department of Mathematics
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
zrapti@illinois.edu

K. Ø. RASMUSSEN* and A. R. BISHOP†

Theoretical Division, Los Alamos National Laboratory
Los Alamos, NM 87545, USA
**kor@lanl.gov*
†*arb@lanl.gov*

Received 28 October 2010
Revised 6 January 2011
Accepted 11 January 2011

In this work we reexamine the unzipping and bubble formation of DNA in the context of the Peyrard–Bishop–Dauxois DNA model using the transfer integral operator method. After a brief overview of the method, we use it to calculate the probabilities that consecutive base-pairs are stretched beyond a threshold amplitude. We compare the continuous Peyrard–Bishop–Dauxois model with an Ising-type model and demonstrate their similarities. For the Peyrard–Bishop–Dauxois model, we derive an expression for the force that is required to open a double-stranded sequence at one end while the other end is kept closed. In the thermodynamic limit, we use this expression to study the dependence of the unzipping force on the temperature. We also calculate the opening probabilities in the case of zero force and in the case where a force is applied in the middle of the sequence. Analytically, the case where the force is applied in the middle is similar to the case of a homogeneous DNA sequence of strong GC pairs with a defect, in the form of a weaker AT base pair, in the middle. Numerical results verify this similarity.

Keywords: DNA; Peyrard–Bishop–Dauxois model; transfer integral operator method.

2000 Mathematics Subject Classification: 92D20, 82B05

1. Introduction

1.1. *DNA models*

The genetic properties of an organism are preserved by storing the bases that encode its genetic information within the double helical DNA molecule. Since the discovery of the DNA double-helical structure in 1952 [41], various models have been proposed to describe

its structure and have attempted to explain its function. While the structure of course is essential, it is becoming clear that structural information alone is insufficient to understand function. It is now widely accepted that the dynamical conformational changes of DNA play a central role for some of the molecular functions in which it participates. For instance, in two of the most fundamental processes of life, transcription and replication, DNA has to locally open in order to allow the appropriate enzymes to engage and operate. In transcription the double strand is separated to allow the RNA polymerase to read the genetic code, while replication begins with a partial unzipping of the DNA end where DNA polymerase binds and starts assembling and reforming the molecule. Separation of the double strand of the DNA molecule is also technologically important. In the polymerase chain reaction (PCR), for example, DNA double strands are routinely converted to separate single strands by melting at elevated temperature. This approach uses thermal energy to disrupt the base pairing and base stacking interactions that otherwise stabilize the double helix at room temperature. From this perspective it is important to understand DNA "breathing" — the dynamical process of base pair opening and closing due to thermal fluctuations. Two established modeling methods invoke thermodynamics [22, 26] and dynamics [17, 31, 36, 42] in order to capture the mechanism of the disruption of the bonds and the breathing of DNA.

One of the most studied models is the Peyrard–Bishop–Dauxois model [15, 26], which consists of two spring-mass chains representing the DNA strands connected by Morse potentials representing the hydrogen bonds. This model was introduced to investigate the mechanism of DNA denaturation, but has since been used to study the dynamics of DNA [1]. The model introduced by Prohofsky *et al.* [36] to study dynamical aspects is essentially the same as the original Peyrard–Bishop–Dauxois model with the important difference that the nearest neighbor coupling is harmonic. Effects of long-range interactions and additional on-site nonlinearities were also considered in this context of this model.

Englander and collaborators [17] first considered the possibility that structures in DNA of 10 or more adjacent open base pairs correspond to thermally induced soliton excitations. This idea was further developed by Salerno [31, 32] who introduced nonhomogeneities in the model and studied the dynamics of DNA promoters in the context. His results have attracted attention [5, 12] and other nonlinear excitations [1, 34, 35], such as discrete breathers, are emerging as important for DNA dynamics and function. These works take advantage of the large amount of research [3, 4, 13, 14, 21, 23, 24] that has been concerned with understanding discrete breathers in nonlinear systems in recent years.

Recently, in [22] the prediction of sequence dependent, thermally induced base pair opening probabilities was based on an Ising-type model taking into account the lateral pairing between the complementary bases and the stacking of the pairs along the longitudinal axis. In this model each base pair is assumed to be in one of two states: closed or open. In the later state, the hydrogen bonds are assumed to be completely disrupted and the bases flipped out of the helical stack. The ten stacking interaction parameters, as well as the two hydrogen bond parameters, were obtained by studying the electrophoretic mobility in gel of DNA fragments with nicks [28].

Other dynamical aspects, such as the twist motions of the DNA strands, have also attracted interest. Yakushevich [42] for example has introduced a model that consists of two chains of long, elastic, and weakly interacting parallel rods. This model describes the

motions of DNA on the nanosecond time scale. The soliton dynamics of this model has been studied by Gaeta [19].

1.2. *DNA bubble formation*

Whenever the available entropic energy in the system equals the energy that is required to disrupt the hydrogen bonds between complementary base pairs, temporary localized single-stranded regions may form. These regions are often called *bubbles*. This form of thermal denaturation has some similarities with what is required for transcription, due to the localized nature of the thermally induced bubbles. Hence, the dynamical behavior and statistics of bubbles and their relation to DNA function have been the subject of numerous studies.

In the context of the Peyrard–Bishop–Dauxois molecular dynamics model the dynamics of bubbles have been investigated using various thermostated molecular dynamics and equilibrium statistical mechanics techniques [10, 15, 29, 30, 39, 43]. Other approaches directed towards the study of bubbles have been based on Ising-type models [20] or stochastic simulations [6] based on the Poland-Scheraga helix-coil model [27].

1.3. *DNA unzipping*

Only recently have direct measurements of the mechanical properties of DNA been made [8]. These studies have led to a better understanding of the interactions between DNA and proteins and enzymes. Mechanical unzipping forces were first measured in [7, 18].

In [18], the mechanical unzipping of DNA was achieved by attaching one end of the strand to a sliding surface through a linker arm, and the other end to a bead attached to a cantilever. The cantilever served as a force sensor, as the deflection of its tip was monitored as a function of the lateral displacement of the surface.

Bockelman *et al.* [7] introduced a molecular construction anchored between a glass microscope slide and a silica bead held in an optical trap. As the surface is laterally displaced, the double helix progressively opens. The force acting on the bead is derived by the position difference between the bead and the trap center. Very significant increase in base pair sensitivity was reported, and sequence features appearing at a scale of 10 base pairs were observed.

In the framework of the Peyrard–Bishop–Dauxois model, the micro-mechanical unzipping was studied using Monte–Carlo simulations and agreement with experimental results was achieved for homogeneous DNA sequences [40].

It has been noted [18] that a simple relation between the formation of bubbles at high temperature and the zones of sudden mechanical unzipping has not been established. However, such an understanding is imperative for further development of technologies [40] based on mechanical unzipping. Here we show how statistical mechanics calculations based on the transfer integral operator method can produce predictions on both bubble formation and the effect of forces. We also investigate how the effect of weak base pairs in an otherwise homogeneous sequence compares with the effect of a force exerted in the middle of a homogeneous sequence.

This article is organized as follows. In Sec. 2 we briefly describe the model. In Sec. 3 we give details on the transfer integral operator method and show a comparison between the Peyrard–Bishop–Dauxois model [26] and the two-state "Ising" model introduced in [22]. In

Sec. 4 we apply the method to calculate the base pair opening probabilities in sequences with and without applied forces. Finally, in Sec. 5 we discuss the results and future directions.

2. DNA Model of Coupled Oscillators

The Peyrard–Bishop–Dauxois (PBD) model [15, 26] has been studied extensively since it was introduced to study DNA denaturation. Recently, has also proven useful in the study of DNA dynamics [1] and mechanical unzipping [40]. In the framework of the PBD, DNA is modeled as a system of coupled oscillators with an on-site Morse potential modeling the hydrogen bonds between each pair of complementary bases, and anharmonic coupled springs modeling the stacking of adjacent bases. The direct interactions between bases are assumed to extend only to nearest-neighbors, and for each base pair n a single degree of freedom, y_n, is used to represent the displacement of complementary bases. Here, $n = 1, \ldots, N$, where N is the length of the sequence.

The configurational energy of the system reads

$$E = \sum_{n=1}^{N} \{V(y_n) + W(y_n, y_{n-1})\} = \sum_{n=1}^{N} \mathcal{E}(y_n, y_{n-1}), \tag{2.1}$$

where

$$V(y_n) = D_n(e^{-a_n y_n} - 1)^2$$

is a Morse potential and

$$W(y_n, y_{n-1}) = \frac{k}{2}(1 + \rho\, e^{-b(y_n + y_{n-1})})(y_n - y_{n-1})^2$$

accounts for the nearest-neighbor coupling. The model includes two sets of parameters for the Morse potential, depending on whether a strong GC or a weaker AT base pair is being modeled. The GC pair contains three hydrogen bonds while the AT pair contains just two. D_n represents the depth of the Morse potential and a_n sets the range of the interaction. The stacking interaction was originally assumed to be homogeneous, but a set of ten parameters has been recently derived that accounts for nonhomogeneous stacking has recently been derived [2]. The anharmonic stacking interaction was introduced in [16] to reproduce the experimentally observed sharp denaturation curves, and serves the purpose of modifying the spring constant so that it effectively models the entropy-driven effect of melting.

3. Transfer Integral Operator Method

All of the equilibrium thermodynamical properties and quantities of a system can be obtained from the partition function. In our case of DNA, we will focus on the following thermodynamic quantities

- the average displacement of complementary bases
- the probability that a number of consecutive base pairs are open, i.e., the bonds in these pairs are stretched beyond a threshold, and
- unzipping forces, i.e the forces required to open base pairs in double-stranded DNA.

The average displacement from equilibrium as a function of temperature can be used in the calculation of melting curves. We evaluate the number of open base pairs (stretched beyond a threshold) as a function of the temperature, and the temperature at which half the base pairs are open is defined as the melting temperature.

The probability that consecutive base pairs are open is related to the formation of bubbles, i.e. temporary localized openings along the double strand. These have been shown in a number of studies to reveal information regarding DNA function [1].

The calculation of unzipping forces is relevant to the role of enzymes during transcription and other processes when DNA has to be partially open and the bases exposed.

The partition function is defined as

$$Z = \int \prod_{n=1}^{N} dy_n \exp(-\beta(V(y_n) + W(y_n, y_{n-1}))). \tag{3.1}$$

There are several possible ways of treating the chain ends, y_0 and y_N. For the case of periodic boundary conditions, one has $y_0 = y_N$. For the case of open boundary conditions, one has to impose $y_0 = y_1$ with no restriction on y_N. Another possibility is to keep the ends at zero displacement $y_0 = 0$ and $y_N = 0$.

3.1. *Partition function evaluation*

The partition function can be evaluated using the transfer integral operator (TIO) method. In order to apply this method, one must first solve two integral eigenvalue equations — one for GC and the second for AT parameter values — with symmetric kernel

$$\int dy\, S(x, y)\, \phi(y) = \lambda \phi(x), \tag{3.2}$$

where the kernel is

$$S(x, y) = \exp\left(-\frac{\beta}{2}\left(V(x) + V(y) + 2\,W(x, y)\right)\right) = S(y, x).$$

Then repeated application of (3.2) can be used to replace consecutive integrations by eigenvalue multiplications. We illustrate the method by obtaining an expression for the partition function in the case of a homogeneous and a nonhomogeneous sequence.

In the homogeneous case, for open boundary conditions, one obtains

$$\begin{aligned} Z &= \int dy_0 \delta(y_0 - y_1) \prod_{n=1}^{N} dy_n \, \exp(-\beta(V(y_n) + W(y_n, y_{n-1}))) \\ &= \int dy_1 \, \exp(-\beta V(y_1)) \prod_{n=2}^{N} dy_n \, \exp(-\beta(V(y_n) + W(y_n, y_{n-1}))) \\ &= \int dy_1 \, \exp\left(-\frac{\beta}{2} V(y_1)\right) \prod_{n=2}^{N} dy_n S(y_n, y_{n-1}) \, \exp\left(-\frac{\beta}{2} V(y_N)\right) \end{aligned}$$

$$
\begin{aligned}
&= \int dy_1 \, \exp\left(-\frac{\beta}{2}V(y_1)\right) dy_2 d\eta S(y_2,\eta)\delta(y_1-\eta) \\
&\quad \times \prod_{n=3}^{N} dy_n S(y_n,y_{n-1}) \, \exp\left(-\frac{\beta}{2}V(y_N)\right) \\
&= \int \sum_i dy_1 \, \exp\left(-\frac{\beta}{2}V(y_1)\right) \phi_i(y_1) d\eta S(y_2,\eta)\phi_i(\eta) \\
&\quad \times \prod_{n=3}^{N} dy_{n-1} S(y_n,y_{n-1}) dy_N \, \exp\left(-\frac{\beta}{2}V(y_N)\right) \\
&= \sum_i \int dy_1 \, \exp\left(-\frac{\beta}{2}V(y_1)\right) \phi_i(y_1)\lambda_i dy_2 S(y_3,y_2)\phi_i(y_2) \\
&\quad \times \prod_{n=4}^{N} dy_{n-1} \, S(y_n,y_{n-1}) dy_N \, \exp\left(-\frac{\beta}{2}V(y_N)\right) \\
&= \sum_i \left(\int dy \phi_i(y) e^{-\frac{\beta}{2}V(y)} \right)^2 \lambda_i^{N-1}.
\end{aligned}
$$

In the third equality we have symmetrized the kernel, and in the fourth we introduce an auxiliary variable η to allow the expansion of the corresponding δ function in terms of the eigenfunctions of the integral eigenvalue equation. This auxiliary integration acts as a catalyst that initiates the repeated application of the transfer integral operator.

In the nonhomogeneous case, for open boundary conditions, the partition function is

$$
\begin{aligned}
Z &= \int \prod_{n=1}^{N} dy_n dy_0 \delta(y_0-y_1) S(y_n,y_{n-1}) e^{\frac{\beta}{2}(V(y_0)-V(y_N))} \\
&= \sum_{i,j} \int dy_1 \phi_i(y_1) e^{-\frac{\beta}{2}V(y_1)} \lambda_i^{r-1} \int dy_r d\eta S(\eta,y_r)\phi_i(y_r)\phi_j(\eta) \\
&\quad \times \lambda_j^{N-r-1} \int dy_N \phi_j(y_N) e^{-\frac{\beta}{2}V(y_N)}.
\end{aligned} \tag{3.3}
$$

Here, we assume that an inhomogeneity exists at base pair r, so that r is not of the same type as $r-1$ and $r+1$. Hence, neither of the integral eigenvalue equations can be used to propagate the integration and one needs to introduce a new auxiliary variable in order to overcome the integration at $r+1$ and move to the next homogeneous region. This step is used whenever a nonhomogeneity occurs.

3.2. *Comparison with an Ising-type model*

The expression for the partition function obtained from the TIO method exposes similarities between the PBD model and an Ising-type model introduced to study DNA thermal fluctuations [22]. According to the Ising-type model, the partition function for a sequence

consisting of N base pairs reads

$$Z = \sum_{\sigma_1=0,1} \sum_{\sigma_2=0,1} \cdots \sum_{\sigma_N=0,1} \prod_{i=1}^{N} (\delta_i)^{\sigma_i} (\alpha_i)^{\sigma_i} (\delta_{i+1})^{f(\sigma_i,\sigma_{i+1})} (\xi)^{f(\sigma_i,\sigma_{i+1})}, \tag{3.4}$$

where $\delta_1 = 1, \delta_{N+1} = 1$ and $\sigma_{N+1} = 1$. Here, $\sigma_i = 0, 1$ corresponds to the closed and open state of the base-pair, respectively, in position i,

$$f(\sigma_i, \sigma_{i+1}) = \begin{cases} \sigma_i & \text{if } \sigma_{i+1} = 0 \\ 0 & \text{if } \sigma_{i+1} = 1, \end{cases}$$

and α_i, δ_i, and ξ are the base-pairing, base-stacking and entropic penalty parameters, respectively. The base-pairing and base-stacking parameters are given in terms of the free energies as

$$\alpha_i = \exp\left(\frac{\Delta G_i^{BP}}{RT}\right) \quad \text{and} \quad \delta_i = \exp\left(\frac{\Delta G_{i-1,i}^{ST}}{RT}\right),$$

where R is the gas constant and T is the temperature.

The similarity between (3.1) and (3.4) are apparent if we rewrite (3.4) as

$$\begin{aligned} Z = & \sum_{\sigma_1=0,1} \sum_{\sigma_2=0,1} \cdots \sum_{\sigma_N=0,1} \prod_{i=1}^{N} \\ & \times \exp\left(\frac{1}{RT}(\sigma_i \Delta G_i^{BP} + \sigma_i \Delta G_{i-1,i}^{ST} + f(\sigma_i, \sigma_{i+1}) \Delta G_{i,i+1}^{ST})\right) (\xi)^{f(\sigma_i,\sigma_{i+1})} \\ = & \sum_{\sigma_1=0,1} \sum_{\sigma_2=0,1} \cdots \sum_{\sigma_N=0,1} \prod_{i=1}^{N} \\ & \times \exp\left(\frac{1}{RT}\{\sigma_i \Delta G_i^{BP} + \sigma_i \Delta G_{i-1,i}^{ST} + f(\sigma_i, \sigma_{i+1})(\Delta G_{i,i+1}^{ST} + RT \ln \xi)\}\right) \end{aligned}$$

In comparison to (3.1), the integrations of the continuous variables y_n have been replaced by summations over the two discrete states, the first term corresponds to the base pairing interaction, and the second and third ones to the stacking interaction. The entropic penalty factor ξ in (3.4) modifies the stacking interaction, which is the dominant stability contribution according to this model, so that the numerically calculated probability values are in agreement with the experimentally measured ones.

3.3. *Opening probabilities*

In the framework of the Peyrard–Bishop–Dauxois model, we define the probability $P_k(s)$ for formation of a "bubble" spanning k base-pairs, starting at base-pair s as

$$P_k(s) = Z^{-1} \int_t^{\infty} Z_k(s) \prod_{n=s}^{s+k-1} dy_n e^{-\beta \mathcal{E}(y_n, y_{n-1})}. \tag{3.5}$$

Here, t is the threshold separation of the double strand above which we define a base-pair to be open. The quantities Z_k are defined as

$$Z_k(s) = \int \left(\prod_{n \neq s,\ldots,s+k-1} dy_n e^{\beta \mathcal{E}(y_n,\ y_{n-1})} \right) dy_0 \delta(y_0 - y_1), \tag{3.6}$$

We note that our working definition of a "bubble" does not require that the open base-pairs s to $s+k-1$ are flanked by closed ones [29, 30]. Similarly, in the context of the Ising-type model introduced in [22] the probability $P_k(s)$ for formation of a "bubble" spanning k base-pairs, starting at base-pair s is given by

$$P_k(s) = Z^{-1} \sum_{\sigma_1=0,1} \cdots \sum_{\sigma_{s-1}=0,1} \sum_{\sigma_{s+k}=0,1} \cdots \sum_{\sigma_N=0,1} \prod_{i=1}^{N} (\delta_i)^{\sigma_i} (\alpha_i)^{\sigma_i} (\delta_{i+1}\xi)^{f(\sigma_i,\sigma_{i+1})}, \tag{3.7}$$

with $\delta_1 = 1, \delta_{N+1} = 1, \sigma_{N+1} = 1$, and $\sigma_j = 1$ for $j = s, \ldots, s+k-1$.

In Fig. 1, we show the probabilities — calculated by Eq. (3.5) — that at least 1 or 7 consecutive base pairs are open beyond a threshold set to 1.5 Å, together with the same probabilities — calculated by Eq. 3.7. The sequence that we used is that of the major late promoter for the adenovirus found in [33] from base pair −62 to base pair +24: 5′ gccacgt gaccgggtgttcctgaaggggggctataaaaggggtgggggcgcgttcgtcctcactctcttcccctccataccttc 3′ . The parameter values for the PBD model can be found in [9], while the parameter values for the Ising-type model can be found in [22]. The value for the entropic factor parameter was taken to be $\xi = 0.001$. In all cases the temperature was $37°C$.

We notice that for single base pair openings the locations of the peaks in the opening probability profiles coincide for both models, and that the largest probability corresponds

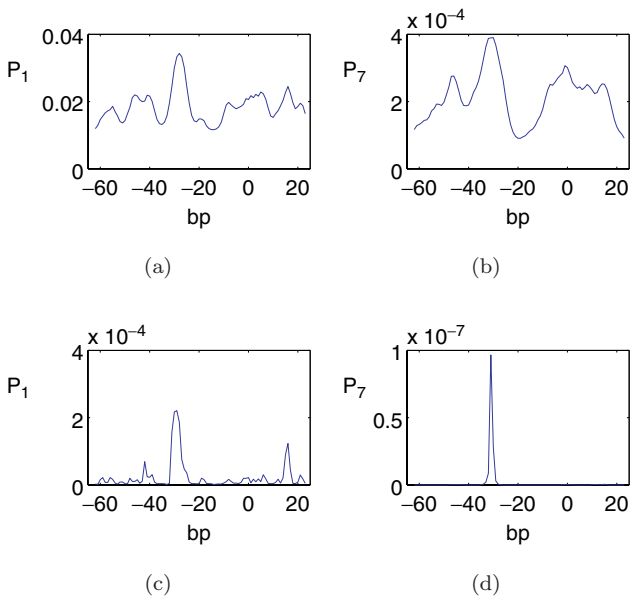

Fig. 1. Panels (a) and (b) show the probability based on the PBD model that at least 1 (P_1) or 7 (P_7), consecutive base pairs are open beyond a threshold of 1.5 Å. In panels (c) and (d), we show the same probability derived based on Ising-type model.

to position −30, where the so-called TATA-box resides. The TATA-box is a core promoter sequence that is the binding site of either transcription factors or histones and has the sequence 5′ TATAAAA 3′ in this case. In the case of the Ising-type model, especially for P_7, there is a strong bias towards high opening probabilities in soft regions, that is in regions containing AT base pairs. Also, the overall opening probabilities are smaller in the Ising-type model than in the PBD model.

4. Applications to Unzipping and Bubble Formation

4.1. *Calculation of the force required to keep apart one end of the chain*

Following [37] we can calculate the unzipping force as a function of L. We assume that $y_0 = 0$ and either $y_{N+1} = 0$ or $y_{N+1} = L > 0$. Then, using the transfer integral operator method one can obtain the following expression for the partition function of a homogeneous chain:

$$\begin{aligned} Z(y_{N+1}) &= \int dy_1 e^{-\frac{\beta}{2}(V(y_1)+V(0)+2W(y_1,0))} \\ &\quad \times \prod_{n=2}^{N} dy_n S(y_n, y_{n-1}) e^{-\beta(V(y_{N+1})+\frac{1}{2}V(y_N)+W(y_{N+1},y_N))} \\ &= \sum_i \int dy_1 S(y_1, 0) d\eta dy_2 \phi_i(\eta)\phi_i(y_1) S(y_2, \eta) \\ &\quad \times \prod_{n=3}^{N} dy_n S(y_n, y_{n-1}) e^{-\beta(V(y_{N+1})+\frac{1}{2}V(y_N)+W(y_{N+1},y_N))} \\ &= \sum_i \phi_i(0)\lambda_i^{N+1}\phi_i(y_{N+1}) e^{-\frac{\beta}{2}V(y_{N+1})}. \end{aligned}$$

Hence,

$$Z(L) = \sum_i \phi_i(0)\lambda_i^{N+1}\phi_i(L) e^{-\frac{\beta}{2}V(L)}$$

and

$$Z(0) = \sum_i \phi_i(0)^2 \lambda_i^{N+1}.$$

The free energy difference between the two states is

$$\Delta G = -\frac{1}{\beta} \ln\left(\frac{Z(L)}{Z(0)}\right),$$

which in the thermodynamic limit $N \to \infty$ becomes

$$\Delta G = -\frac{1}{\beta} \ln\left(\frac{\phi_1(L)}{\phi_1(0)}\right) + \frac{1}{2}V(L), \tag{4.1}$$

where ϕ_1 the eigenfunction corresponding to the largest eigenvalue λ_1. We note that (4.1) is similar to Eq. (6) in [11] that was obtained through a continuum transfer matrix technique

for a semi-microscopic model. In that model a single degree of freedom is used for every base-pair, and besides the interactions with the nearest neighbors and the hydrogen bonding, also a torque energy term was included.

The unzipping force can be calculated as the derivative

$$\frac{\partial \Delta G}{\partial L} = -\frac{1}{\beta}\frac{\phi_1'(L)}{\phi_1(L)} + \frac{1}{2}V'(L). \tag{4.2}$$

It is worth pointing out that, at zero temperature, the unzipping force is

$$\frac{\partial E}{\partial L} = \frac{\partial W(y_N, L)}{\partial L} + V'(L).$$

However, there is a way to rewrite (4.2) so that $V'(L)$ has unity as its prefactor: We write

$$Z(y_{N+1}) = \sum_i \phi_i(0)\lambda_i^N \int dy_N \phi_i(y_N) e^{-\frac{\beta}{2}(V(y_N)+2W(y_{N+1},y_N))} e^{-\beta V(y_{N+1})}.$$

Then, as $N \to \infty$

$$\Delta G = -\frac{1}{\beta}\ln\left(\frac{\int dy_N \phi_1(y_N) e^{-\frac{\beta}{2}(V(y_N)+2W(L,y_N))} e^{-\beta V(L)}}{\phi_1(0)\lambda_1}\right)$$

and

$$\frac{\partial \Delta G}{\partial L} = \frac{\int dy \phi_1(y) e^{-\frac{\beta}{2}(V(y)+2W(L,y))}\frac{\partial W(y,L)}{\partial L}}{\int dy \phi_1(y) e^{-\frac{\beta}{2}(V(y)+2W(L,y))}} + V'(L). \tag{4.3}$$

Both (4.2) and (4.3) are useful because (4.2) reveals the almost linear behavior in temperature, while (4.3) reveals the resemblance with the force at zero temperature.

In Fig. 2, we show the force given by (4.2) vs. the stretching in the last base pair $N+1$. The stretching length L is given in Å and the force in pN. The temperature is 309.15 K.

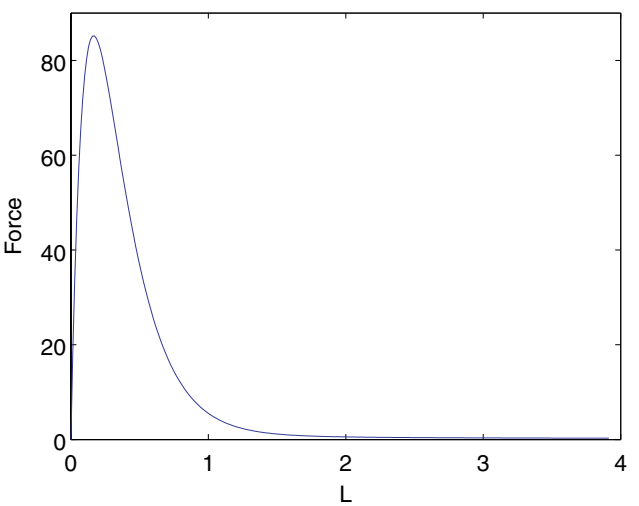

Fig. 2. The force vs. the stretching L in base pair $N+1$. The length is given in Å and the force in pN.

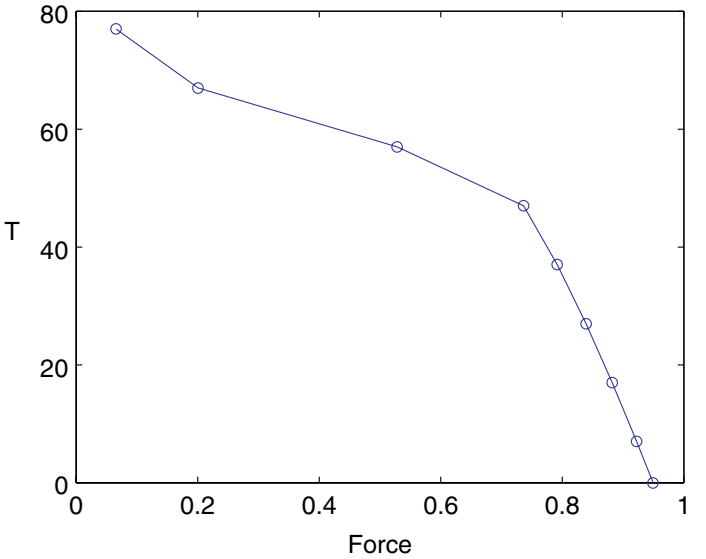

Fig. 3. The temperature vs. the unzipping force at the end of the chain for $L = 1.7$ Å. The force is given in pN and the temperature in ${}^0C$.

The force barrier is of the same order of magnitude as the one reported in [40], where a force was applied at the end of a chain 300 base-pairs long pulling the last base-pair apart at a constant velocity. The parameter values used here correspond to a homogeneous AT sequence.

In Fig. 3, we show force-temperature diagram for a homogeneous AT sequence. For the numerical evaluation of the data points we used Eq. (4.2) and assumed that the base pair at the end of the sequence was stretched by $L = 1.7$ Å. The shape of the $T(Force)$ curve is in agreement with previously reported experimental and numerical studies [40].

4.2. *Effect of a force applied in the middle of a sequence*

We now assume that there is a force applied at base pair r within a homogeneous sequence and study the effect this has on the opening profile. We compare this profile with that of a homogeneous sequence with a single "defect" base pair r of another type. We compare the influence that the force and the nonhomogeneity have on the average opening displacements. The case with the force applied on base pair r can be used to explain the effect of restriction proteins on DNA [38]. We also compare these cases with the one when no force or inhomogeneity is present and show the range of influence of the force, namely the number of neighboring base pairs that are affected.

We consider a homogeneous sequence of N base pairs with a force applied to base pair $1 < r < N$. The modified configurational energy now reads

$$E = \sum_{n=1}^{N} \{V(y_n) + W(y_n, y_{n-1}) - V_{pull}(y_r)\},$$

where

$$V_{pull}(y_r) = \frac{k_p}{2}(y_r - Y_r)^2 \tag{4.4}$$

and k_p, Y_r are the spring constant and displacement of the elastic spring that deforms base pair r.

Then, using the same ideas as above, the partition function can be evaluated from the expression:

$$Z = \sum_{i,j} \int dy_1 \phi_i(y_1) e^{-\frac{\beta}{2}V(y_1)} \lambda_i^{r-1} \int dy_r \, dy_{r+1} S(y_{r+1}, y_r) \phi_i(y_r) \phi_j(y_{r+1}) e^{-\beta V_{pull}(y_r)}$$
$$\times \lambda_j^{N-r-1} \int dy_N \phi_j(y_N) e^{-\frac{\beta}{2}V(y_N)}. \tag{4.5}$$

The expressions (3.3) and (4.5) suggest a similarity between the action of forces and nonhomogeneity on a DNA sequence.

In Fig. 4, we show the opening probability profile for a homogeneous GC sequence, with a weak pair AT that acts as a defect at base pair 43, and also with a force applied at base pair 43. As a reference, the purely homogeneous sequence is also shown. All sequences contain the same number of base pairs $N = 87$ and are flanked by 50 base pairs at each side to avoid opening due to the boundary effect. The contribution to the energy of the applied force is given by (minus) (4.4) with parameter values $k_p = 0.0016\,\text{pN/Å}$ and $Y_r = 0.5\,\text{Å}$. For these parameter values, the range of effect of the force is slightly larger than that of the defect. There are base pairs with opening propensity greater than that of the homogeneous GC sequence that are located further from base pair 43 in the case of the force than in the case of the defect. Another difference is in the opening propensity of base pairs 42 and 44: in the presence of the force, their opening probabilities are significantly lower than

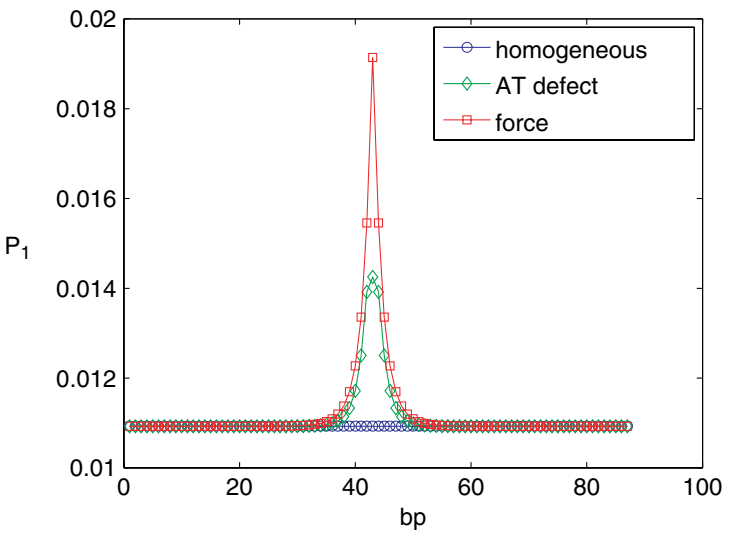

Fig. 4. Probability that one or more base-pairs are opened a beyond threshold 1.5 Å, for a homogeneous GC sequence, a homogeneous GC sequence with an AT defect at base pair 43, and a homogeneous GC sequence with a force applied at base pair 43.

in the presence of the defect. This is due to the fact that when the defect is present, the nearest neighbor coupling also affects the immediate neighbors and effectively increase their amplitude.

To understand the dependence of the the opening profile on the force parameters, we repeated the calculation with k_p varying from 1.6×10^{-4} to 1.6×10^{-2} pN/Å and Y_r varying from 0.5 to 1.5 Å. As Y_r increases, the opening probabilities decrease slightly, but the overall profile is otherwise unaffected. On the other hand, keeping $Y_r = 1$ Å and varying k_p results in significant changes in the opening profiles. These results are shown in Figs. 5 and 6.

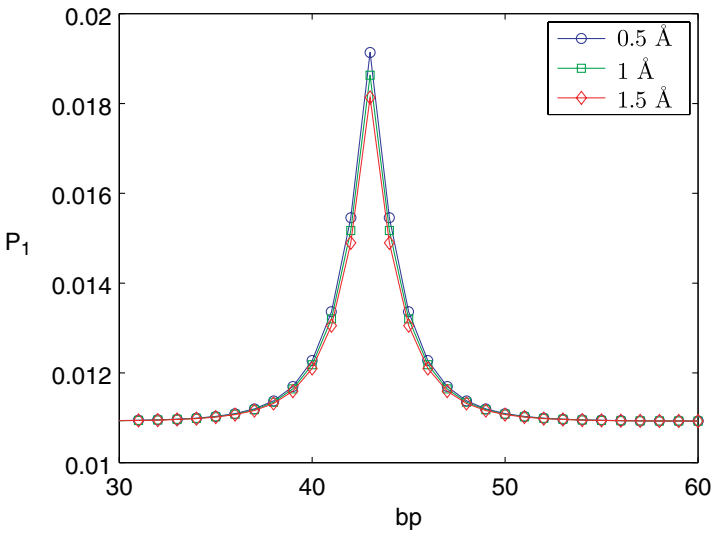

Fig. 5. Probability that one or more base-pairs are open beyond threshold 1.5 Å, for a homogeneous GC sequence with a force applied at base pair 43, with various values of Y_r.

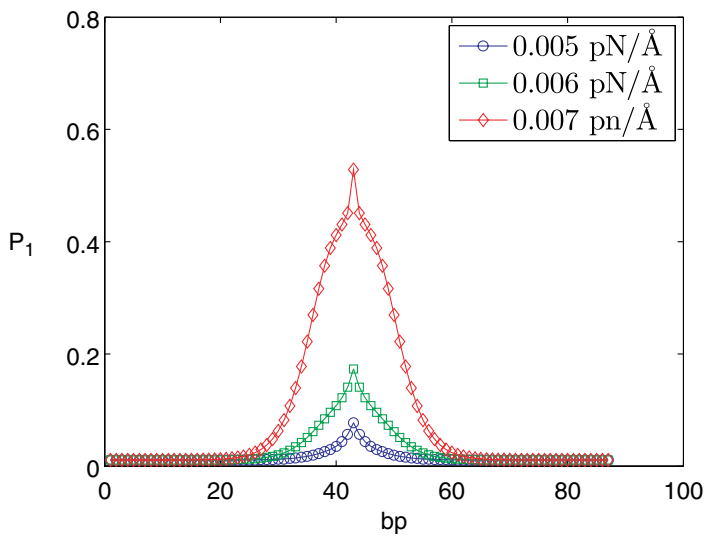

Fig. 6. Probability that one or more base-pairs are open beyond threshold 1.5 Å, for a homogeneous GC sequence with a force applied at base pair 43, with various spring constants k_p.

5. Discussion and Future Directions

We have shown that a model of DNA with a single degree of freedom per base pair gives valuable results capturing essential dynamical features of the molecule. Our studies strongly suggest that this level of modeling even has biological relevance [1]. As this kind of mesoscale modeling continues to improve [2, 25], and maintain a strong coupling to experiments, it has the potential to enable significant insight into several of DNAs biomolecular functionalities

Acknowledgments

Z. R. acknowledges support from the NSF under grant number DMS-0708421. Research is carried out at Los Alamos National Laboratory under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy under Contract No. DE-AC52-06NA25396.

References

[1] B. S. Alexandrov, V. Gelev, S. W. Yoo, A. R. Bishop, K. Ø. Rasmussen and A. Usheva, Toward a detailed description of the thermally induced dynamics of the core promoter, *PLoS Comput. Biol.* **5** (2009) e1000313.

[2] B. S. Alexandrov, V. Gelev, Y. Monisova, L. B. Alexandrov, A. R. Bishop, K. Ø. Rasmussen and A. Usheva, A nonlinear dynamic model of DNA with a sequence-dependent stacking term, *Nucleic Acids Res.* **37** (2009) 2405–2410.

[3] J. F. R. Archilla, J. Cuevas, B. Sánchez-Rey and A. Alvarez, Demonstration of the stability or instability of multibreathers at low coupling, *Physica D* **180** (2003) 235–255.

[4] S. Aubry, Breathers in nonlinear lattices: Existence, linear stability and quantization, *Physica D* **103** (1997) 201–250.

[5] J. D. Bashford, Salerno's model of DNA Re-Analyzed: Could Breather solitons have biological significance? *J. Biol. Phys.* **32** (2006) 27–47.

[6] S. K. Batik, T. Ambjörnsson and R. Metzler, Stochastic approach to DNA breathing dynamics, *Europhys. Lett.* **71** (2005) 852–858.

[7] U. Bockelmann, Ph. Thomen, B. Essevaz-Roulet and F. Heslot, Unzipping DNA with optical tweezers: High sequence sensitivity and force flips, *Biophys. J.* **82** (2002) 1537–1553.

[8] C. Bustamante, Z. Bryant and S. B. Smith, Ten years of tension: Single-molecule DNA mechanics, *Nature* **421** (2003) 423–427.

[9] A. Campa and A. Giansanti, Experimental tests of the Peyrard–Bishop model applied to the melting of very short DNA chains, *Phys. Rev. E* **58** (1998) 3585–3588.

[10] C. H. Choi, G. Kalosakas, K. Ø. Rasmussen, M. Hiromura, A. R. Bishop and A. Usheva, DNA dynamically directs its own transcription initiation, *Nucleic Acids Res.* **32** (2004) 1584–1590.

[11] S. Cocco, R. Monasson and J. F. Marko, Force and kinetic barriers to unzipping of the DNA double helix, *Proc. Natl. Acad. Sci. USA* **98** (2001) 8608–8613.

[12] S. Cuenda, A. Sanchez and N. R. Quintero, Does the dynamics of sine-Gordon solitons predict active regions of DNA? *Physica D* **223** (2006) 214–221.

[13] J. Cuevas, V. Koukouloyannis, P. G. Kevrekidis and J. F. R. Archilla, Multibreather and vortex breather stability in Klein–Gordon lattices: Equivalence between two different approaches (2010) arXiv:1006.0346v1.

[14] I. Daumont, T. Dauxois and M. Peyrard, Modulational instability: First step towards energy localization in nonlinear lattices, *Nonlinearity* **10** (1997) 617–630.

[15] T. Dauxois, M. Peyrard and A. R. Bishop, Dynamics and thermodynamics of a nonlinear model for DNA denaturation, *Phys. Rev. E* **47** (1993) 684–697.
[16] T. Dauxois, M. Peyrard and A. R. Bishop, Entropy-driven DNA denaturation, *Phys. Rev. E* **47** (1993) R44–R47.
[17] S. W. Englander, N.R. Kallenbach, A. J. Heeger, J. A. Krumhansl and S. Litwin, Nature of the open state in long polynucleotide double helices: Possibility of soliton excitations, *Proc. Natl. Acad. Sci. USA* **77** (1980) 7222–7226.
[18] B. Essevaz-Roulet, U. Bockelmann and F. Heslot, Mechanical separation of the complementary strands of DNA, *Proc. Natl. Acad. Sci. USA* **94** (1997) 11935–11940.
[19] G. Gaeta, Solitons in planar helicoidal Yakushevich model of DNA dynamics, *Phys. Lett. A* **168** (1992) 383–390.
[20] D. Jost and R. Everaers, Genome wide application of DNA melting analysis, *J. Phys. Condens. Matter* **21** (2009) 034108.
[21] Y. S. Kivshar and M. Peyrard, Modulational instabilities in discrete lattices, *Phys. Rev. A* **46** (1992) 3198–3207.
[22] A. Krueger, E. Protozanova and M. D. Frank-Kamenetskii, Sequence dependent base-pair opening in DNA double helix, *Biophys. J.* **90** (2006) 3091–3099.
[23] R. S. MacKay and S. Aubry, Proof of existence of breathers for time-reversible or Hamiltonian networks of weakly coupled oscillators, *Nonlinearity* **7** (1994) 1623–1643.
[24] R. S. MacKay and J-A Sepulchre, Effective Hamiltonian for traveling discrete breathers, *J. Phys. A* **35** (2002) 3985–4002.
[25] M. Peyrard, S. Cuesta-Löpez and G. James, Nonlinear analysis of the dynamics of DNA breathing, *J. Biol. Phys.* **35** (2009) 73–89.
[26] M. Peyrard and A. R. Bishop, Statistical mechanics of a nonlinear model for DNA denaturation, *Phys. Rev. Lett.* **62** (1989) 2755–2758.
[27] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).
[28] E. Protozanova, P. Yakovchuk and M. D. Frank–Kamenetskii, Stacked-unstacked equilibrium at the nick site of DNA, *J. Mol. Biol.* **342** (2004) 775–785.
[29] Z. Rapti, A. Smerzi, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi and A. Usheva, Lengthscales and cooperativity in DNA bubble formation, *Europhys. Lett.* **74** (2006) 540–546.
[30] Z. Rapti, A. Smerzi, K. Ø. Rasmussen, A. R. Bishop, C. H. Choi and A. Usheva, Healing length and bubble formation in DNA, *Phys. Rev. E* **73** (2006) 051902.
[31] M. Salerno, Discrete model for promoter dynamics, *Phys. Rev. A* **44** (1991) 5292–5297.
[32] M. Salerno, Dynamical properties of DNA promoters, *Phys. Lett. A* **167** (1992) 49–53.
[33] M. Sawadogo and R. G. Roeder, Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region, *Cell* **43** (1985) 165–175.
[34] C. B. Tabi, A. Mohamadou and T. Kofane, Formation of localized structures in the Peyrard–Bishop–Dauxois model, *J. Phys. Condens. Matter* **20** (2008) 415104.
[35] C. B. Tabi, A. Mohamadou and T. Kofane, Modulational instability in the anharmonic Peyrard–Bishop–Dauxois model of DNA, *Eur. Phys. J. B* **74** (2010) 151–158.
[36] M. Techera, L. L. Daemen and E. W. Prohofsky, Analysis of the breakdown of continuum and semidiscrete approximation of a nonlinear model for the DNA double helix, *Phys. Rev. A* **41** (1990) 4543–4546.
[37] N. Theodorakopoulos, M. Peyrard and R. S. MacKay, Nonlinear structures and thermodynamic instabilities in a one-dimensional lattice system, *Phys. Rev. Lett.* **93** (2004) 258101.
[38] M. R. Tock and D. T. F. Dryden, The biology of restriction and anti-restriction, *Current Opinion Microbiology* **8** (2005) 466–472.
[39] T. S. van Erp, S. Cuesta-Lopez, J.-G. Hagmann and Michel Peyrard, Can one predict DNA transcription start sites by studying bubbles? *Phys. Rev. Lett.* **95** (2005) 218104.

[40] N. K. Voulgarakis, A. Redondo, A. R. Bishop and K. Ø. Rasmussen, Probing the mechanical unzipping of DNA, *Phys. Rev. Lett.* **96** (2006) 248101.
[41] J. D. Watson and F. H. C. Crick, Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid, *Nature* **171** (1953) 737–738.
[42] L. V. Yakushevich, Nonlinear DNA dynamics: A new model, *Phys. Lett. A* **136** (1989) 413–417.
[43] Y.-l. Zhang, W.-M. Zheng, J.-X. Liu and Y. Z. Chen, Theory of DNA melting based on the Peyrard-Bishop model, *Phys. Rev. E* **56** (1997) 7100–7115.