Force induced unzipping of DNA with long range correlated noise

Pui-Man Lam and Yi Zhen

Physics Department, Southern University, Baton Rouge, LA 70813, USA
E-mail: puiman_lam@subr.edu and yi_zhen@subr.edu

Received 15 March 2011
Accepted 14 June 2011
Published 30 June 2011

Online at stacks.iop.org/JSTAT/2011/P06023
doi:10.1088/1742-5468/2011/06/P06023

Abstract. We derive and solve a Fokker–Planck equation for the stationary distribution of the free energy, in a model of unzipping of double-stranded DNA under external force. The autocorrelation function of the random DNA sequence can be of a general form, including long range correlations. In the case of Ornstein–Uhlenbeck noise, characterized by a finite correlation length, our result reduces to the exact result of Allahverdyan et al, with the average number of unzipped base pairs going as $\langle X \rangle \sim 1/f^2$ in the white noise limit, where $f$ is the deviation from the critical force. In the case of long range correlated noise, where the integrated autocorrelation is divergent, we find that $\langle X \rangle$ is finite at $f = 0$, with its value decreasing as the correlations become of longer range. This shows that long range correlations actually stabilize the DNA sequence against unzipping. Our result is also in agreement with the findings of Allahverdyan et al obtained using numerical generation of the long range correlated noise.

Keywords: fractal growth (theory), kinetic roughening (theory), roughening transition (theory), self-affine roughness (theory)
1. Introduction

In the past two decades, micromanipulation techniques have become important tools in the repertoire of biophysicists and structural biologists, complementing more traditional scattering and spectroscopic measurements. In the DNA molecule, the two individual strands are bonded by hydrogen bonds while they themselves are constructed with much stronger covalent bonds. In addition, the two double strands are wrapped around each other in a double-helix structure. For simplicity, this last aspect of the DNA structure will be neglected in our study. Each single strand is a polymer formed from nucleotides which can be of two types: purines, consisting of adenine (A) and guanine (G), and pyrimidines, consisting of cytosine (C) and thymine (T). The hydrogen bonds between the two opposite strands can only be of two types, either A–T bases or G–C bases, with different formation energies. The GC base pairs are made of three hydrogen bonds while AT base pairs are made of two hydrogen bonds only. These base pairs, since they are hydrophobic, are located at the core of the double helix. The polymerase, whose function is to read the genetic code encoded in the DNA, must first unzip the two strands in the DNA in order to get to it. This makes force induced unzipping of DNA by an external force an important mechanism in the functioning of all living organisms. Force induced unzipping of DNA has been actively investigated in the last decade [1]–[9]. For a very recent thorough review on biomolecules under mechanical force, see the article by Kumar and Li [10].

It is known that the concentrations of AT and GC base pairs are approximately equal, especially in higher organisms [11]. The difference between the AT and GC formation energies for one hydrogen bond is of the same order as the average formation energy of 2.5 hydrogen bonds. For certain bulk properties, this difference may not be relevant and DNA can be considered as a homogeneous base sequence. However for situations where the unzipping energy is of the order of the formation energy, the heterogeneity of the base sequence becomes relevant. Lubensky and Nelson [4] took the first step in this direction. They showed that for a homogeneous sequence, the number of unzipped base pairs $\langle X \rangle$ diverges as $\langle X \rangle \sim (\mathcal{F}_c - \mathcal{F})^{-1}$ for external force $\mathcal{F}$ near the critical force $\mathcal{F}_c$.
while for a heterogeneous sequence with short range white noise correlation, it diverges as $\langle X \rangle \sim (3c - 3)^{-2}$.

It is known that DNA sequences in fact display long range correlations [12]–[15], both in the non-coding (intron) and coding regions of the DNA: two base pairs separated by thousands of pairs appear to be statistically correlated. Despite the ubiquity of these long range correlations, the biological reasons for them are largely unexplored. For long range correlation with the correlation function $\langle \eta(m)\eta(m') \rangle \sim |m - m'|^\alpha$, where $\eta(m)$ is the noise or deviation of the binding energy from its average value at the base-pair position $m$ along the sequence, $\alpha < 1$, Lubensky and Nelson [4] had predicted, using heuristic arguments, that the divergence of $\langle X \rangle$ near the critical unzipping force should go as $\langle X \rangle \sim (3c - 3)^{-2/\alpha}$. This is a stronger divergence than for the short range, white noise case.

Allahverdyan et al [7] studied the unzipping of DNA with correlated base sequences using a somewhat simplified version of the model of Lubensky and Nelson. For the case of finite range correlation characterized by a finite correlation length, they derived and solved a Fokker–Planck equation for the distribution, from which they calculated the average number of unzipped base pairs $\langle X \rangle$. The white noise limit could be obtained by taking the limit of zero correlation length. In this limit, they recovered the result $\langle X \rangle \sim (3c - 3)^{-2}$ of Lubensky and Nelson [4]. For the case of long range correlated noise, they could not derive a Fokker–Planck equation. Rather, the long range correlated noise with the correct behavior of the correlation function was numerically generated and used to calculate the average number of unzipped base pairs $\langle X \rangle$. Contrary to the prediction of Lubensky and Nelson, they found that long range correlations actually stabilize the DNA against unzipping.

Here we study the force induced unzipping of DNA with both finite range and long range correlations using an approximate Fokker–Planck equation. For finite range correlations we reproduce exactly the results of Allahverdyan et al. For long range correlations, we also find that the DNA is more stable as compared to the short range correlated case.

In section 2 we present the details of the model. In section 3 we will concentrate on the derivation of a Fokker–Planck equation from which we can obtain the average free energy and the average number of unzipped base pairs. In section 4 we present the solution of the stationary Fokker–Planck equation for the case of Ornstein–Uhlenbeck noise, where the autocorrelation function of the noise $K(t)$ has a finite correlation length. For this case we show that the stationary probability distribution function for the free energy reduces to the exact result of Allahverdyan et al [7]. In section 5, we present the solution for the case of a long range correlated noise. We find that long range correlations actually stabilize the DNA against unzipping. Our findings corroborate the findings of Allahverdyan et al [7] obtained using numerical simulation and stand in contradiction to the predictions of Lubensky and Nelson [4]. There has been very little work on the effect of long range correlation on DNA unzipping besides that of the references [4, 7]. Our Fokker–Planck equation approach may provide a new perspective. Section 6 is a conclusion.

2. The model

We will now review the model studied by Allahverdyan et al. The physical basis of the model has been described in detail in their paper and will not be repeated here. In section 3 we will concentrate on the derivation of a Fokker–Planck equation that can be
applied to both short range and long range correlated noise in the DNA sequence. The DNA lies along the x-axis between \( x = a \) and \( L \). The base pairs are located at points \( x_i, a \leq x_i \leq L, i = 1, 2, \ldots, M \). They can be in one of two states: bound or disconnected. A disconnected base pair at point \( x_i \) contributes a binding energy \( \phi(x_i) \), while bound pairs contribute nothing. The binding energy \( \phi(x_i) \) is a random quantity with an average \( \langle \phi \rangle \):

\[
\phi(x_i) = \langle \phi \rangle + \eta(x_i)
\]

where \( \eta(x_i) \) is a random deviation from the average value at point \( x_i \). An external force \( \mathfrak{F} \) is acting on the left end \( x = a \), pulling apart the two strands. If a bond at point \( x_i \) is broken, then all base pairs with \( j < i \) are broken as well.

The Hamiltonian is given by

\[
H(x) = -\mathfrak{F} x + \sum_{i=1}^{x} \phi(x_i) = -\mathfrak{F} x + \sum_{i=1}^{x} [\langle \phi \rangle + \eta(x_i)]
\]

\[
= (\langle \phi \rangle - \mathfrak{F}) x + \sum_{i=1}^{x} \eta(x_i) \equiv f x + \sum_{i=1}^{x} \eta(x_i),
\]

where \( x \) is the number of broken base pairs and

\[
f \equiv \langle \phi \rangle - \mathfrak{F}.
\]

Here \( \mathfrak{F} \) denotes both the force and the force multiplied by the base-pair separation, which is taken as unity. The units in equations (2) and (3) may look strange. The form is a result of the base-pair separation being taken as unity. On this basis, force and energy have the same notation. Similarly, \( x, x_i \) can be either numbers or distances in units of the base-pair separation. In this Hamiltonian, the two unzipped single strands exert no restoring force. This is a simplification of the model studied by Lubensky and Nelson. Of course in both models, the long range, excluded volume interaction of the polymers has been neglected. However, using this model, Allahverdyan et al reproduced the results of Lubensky and Nelson on the divergence of the average number of unzipped base pairs, when the external force is close to the critical value, for the case when the \( \eta(x_i) \) are short range correlated. This shows that this model can also be used to study the effect of long range correlation on the average number of unzipped base pairs.

In the continuum limit, the Hamiltonian becomes

\[
H(x) = (x - a) f + \int_{a}^{x} ds \eta(s).
\]

From this we can calculate the partition function and the free energy:

\[
Z = \int_{a}^{L} dx e^{-\beta H(x)}, \quad F = -T \ln Z,
\]

with \( \beta = 1/T \) the inverse temperature and the Boltzmann’s constant \( k_B \equiv 1 \).

The order parameter is the number of broken base pairs \( X \):

\[
X = \partial_f F, \quad \langle X \rangle = \partial_f \langle F \rangle.
\]
It remains to specify the properties of the noise $\eta$. We assume an autocorrelation function of the noise of the form

$$K(t - t') \equiv \langle \eta(t) \eta(t') \rangle, \quad K(t) = K(-t). \quad (7)$$

Depending on the behavior of $K(t)$ for large $t$, two cases are distinguished: the finite range correlated situation and the long range correlated situation.

In the finite range situation, the total intensity of the noise is finite:

$$D = \int_0^\infty ds K(s). \quad (8)$$

In particular, the white noise case

$$K(t) = D \delta(t) \quad (9)$$
describes completely uncorrelated noise.

The Ornstein–Uhlenbeck noise is characterized by a finite correlation length $\tau$:

$$K(t) = \frac{D}{\tau} e^{-|t|/\tau}. \quad (10)$$

The power law correlated noise is given by

$$K(t) \sim |t|^{-\delta}, \quad |t| \geq 1, \quad \delta > 1. \quad (11)$$

Long range correlated noise is characterized by

$$K(t) \sim \sigma |t|^{-\alpha}, \quad |t| \geq 1, \quad 0 < \alpha < 1. \quad (12)$$

Here we take $K(t)$ to be

$$K(t) = \begin{cases} \sigma, & |t| \leq 1 \\ \sigma |t|^{-\alpha}, & |t| > 1, \quad \alpha < 1 \end{cases}. \quad (13)$$

A Langevin equation can be obtained by differentiating equation (5) with respect to $a$:

$$\frac{dZ}{da} = -e^{-\beta H(a)} - \int_a^L dx \beta e^{-\beta H(x)} \frac{\partial H}{\partial a} = - \exp(0) + \beta \int_a^L e^{-\beta H(x)} [f + \eta(a)]. \quad (14)$$

With the substitution $t = -a$, this can be rewritten as

$$\frac{dZ}{dt} = 1 - \beta f Z - \beta \eta(t) Z, \quad -L < t < 0. \quad (15)$$

In terms of the free energy $F$, this can be written as

$$\frac{dF}{dt} + T e^{\beta F} - f = \eta(t), \quad -L < t < 0. \quad (16)$$

doi:10.1088/1742-5468/2011/06/P06023
3. Derivation of a Fokker–Planck equation

In this section we will derive a Fokker–Planck equation for the probability distribution \( P(F, t) \) of the free energy at time \( t \), corresponding to the Langevin equation (16). We will follow the functional integral method of Fox [16]. Any Gaussian noise has the functional form

\[
P[\eta] = \mathcal{R} \exp \left[ -\frac{1}{2} \int ds \int ds' \eta(s)\eta(s')C(s - s') \right],
\]

where \( \mathcal{R} \) is a normalization factor:

\[
\mathcal{R}^{-1} = \int D\eta \exp \left[ -\frac{1}{2} \int ds \int ds' \eta(s)\eta(s')C(s - s') \right].
\]

The function \( C(s) \) is the functional inverse of the autocorrelation function \( K(t) \) defined in equation (7), i.e.,

\[
\int ds C(t - s)K(s - s') = \delta(t - s).
\]

This can be shown by the following analysis (as suggested by an anonymous referee):

\[
0 = \mathcal{R} \int D\eta \frac{\delta}{\delta\eta(t)} \left\{ \eta(s) \exp \left[ -\frac{1}{2} \int ds' \int ds'' \eta(s')\eta(s'')C(s' - s'') \right] \right\} \\
= \mathcal{R} \int D\eta \left\{ \delta(t - s) - \eta(s) \int ds' ds'' \delta(t - s')C(s' - s'')\eta(s'') \right\} \\
\times \exp \left[ -\frac{1}{2} \int ds' \int ds'' \eta(s')\eta(s'')C(s' - s'') \right] \\
= \delta(t - s) - \int ds''C(t - s'')K(s - s'').
\]

The probability for having free energy \( F \) at time \( t \) is given by

\[
P(F, t) = \int D\eta P[\eta(t)]\delta(F - F[\eta, t]).
\]

The time derivative of this is given by

\[
\frac{dP}{dt} = \int D\eta P[\eta(t)] \left[ -\frac{\partial}{\partial F} \delta(F - F[\eta, t]) \right] \frac{dF[\eta, t]}{dt}.
\]

Using equation (16), this can be written as

\[
\frac{dP}{dt} = -\frac{\partial}{\partial F}(f - Te^F/T) \int D\eta P[\eta]\delta(F - F[\eta, t]) - \frac{\partial}{\partial F} \int D\eta P[\eta]\delta(F - F[\eta, t])\eta(t).\]

From the form of \( P[\eta] \) in (17) and equation (19) one can easily show that

\[
\eta(t)P[\eta] = -\int \delta P[\eta]\frac{\delta}{\delta\eta(u)}K(u - t)du.
\]

doi:10.1088/1742-5468/2011/06/P06023
Substituting this into the second term in equation (22) one has

\[- \frac{\partial}{\partial F} \int D\eta \delta(F - F[\eta, t]) P[\eta] \eta(t) = \frac{\partial}{\partial F} \int_D \delta F \frac{\delta P[\eta]}{\delta \eta(u)} \delta(F - F[\eta, t])\]

\[= - \frac{\partial}{\partial F} \int K(u - t) du \int D\eta \frac{\delta}{\delta \eta(u)} \delta(F - F[\eta, t])\]

\[= \frac{\partial}{\partial F} \int K(u - t) du \int D\eta \frac{\delta}{\delta F} \delta(F - F[\eta, t]) \frac{\delta F[\eta, t]}{\delta \eta(u)}\]

\[= \frac{\partial^2}{\partial F^2} \int_{-L}^0 ds K(t - s) \left\langle \delta(F - F[\eta, t]) \frac{\delta F[\eta, t]}{\delta \eta(s)} \right\rangle\]

where the second equality was obtained by a functional integration by parts.

The functional derivative \( \delta F[\eta, t]/\delta \eta(s) \) involved in the last equality in equation (24) can be obtained using the Langevin equation (16):

\[\frac{\delta F[\eta, t]}{\delta \eta(s)} = \theta(t - s) \exp \left[ \int_t^s e^{F[n,t']/T} dt' \right]\]

where \( \theta(t) \) is the Heaviside step function. Using this in the last equality in equation (24) one can write the second term in equation (22) as

\[- \frac{\partial}{\partial F} \int D\eta \delta(F - F[\eta, t]) P[\eta] \eta(t)\]

\[= \frac{\partial^2}{\partial F^2} \int_{-L}^0 ds K(t - s) \left\langle \delta(F - F[\eta, t]) \theta(t - s) \exp \left[ \int_t^s e^{F[n,t']/T} dt' \right] \right\rangle. \tag{26}\]

We now make the following approximation to the integral on the right-hand side of equation (26):

\[\int_{-L}^0 ds K(t - s) \left\langle \delta(F - F[\eta, t]) \theta(t - s) \exp \left[ \int_t^s e^{F[n,t']/T} dt' \right] \right\rangle\]

\[\approx \int_{-L}^t ds K(t - s) P(F, t) \exp[-e^{F/T}(t - s)]\]

\[= P(F, t) \int_{-L}^t ds K(t - s) \exp[-e^{F/T}(t - s)]. \tag{27}\]

This approximation consists in replacing \( F[\eta, t] \) within the expectation value by the constant \( F \). The justification for this is the existence of the Dirac delta function \( \delta(F - F[\eta, t]) \). Using the property \( K(t) = K(-t) \), this can be written as \( P(F, t) \int_{-L}^{L+L} ds K(s) \exp[-e^{F/T}s] \). In the limit \( L \to \infty \), this becomes \( P(F, t) \int_{-L}^{L+L} ds K(s) \exp[-e^{F/T}s] = P(F, t)G(F) \), with

\[G(F) = \int_{-L}^{L+L} ds K(s) \exp[-e^{F/T}s]. \tag{28}\]
With this approximation, the second term in equation (22) can finally be written as

$$\frac{-\partial}{\partial F} \int D\eta \delta(F - F[\eta,t])P[\eta] \eta(t) \approx \frac{\partial^2}{\partial F^2} G(F) P(F,t)$$

with $G(F)$ given in equation (28). Equation (22) can now be written as

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial F} \left[ (f - T e^{F/T}) P(F,t) \right] + \frac{\partial^2}{\partial F^2} \left[ G(F) P(F,t) \right]$$

which is the Fokker–Planck equation for $P(F,t)$.

The stationary distribution is given by $\frac{\partial P}{\partial t} = 0$, or

$$\frac{\partial}{\partial F} \left\{ \frac{\partial}{\partial F} [G(F) P(F)] - (f - T e^{F/T}) P(F) \right\} = 0.$$

This equation can be solved as

$$P(F) = \text{const} \frac{G(F)}{\int_0^\infty dF' \left( f F - T e^{F'/T} \right) / G(F')}.$$

Knowing $P(F)$, we can then calculate the average free energy $\langle F \rangle$ as

$$\langle F \rangle = \int_{-\infty}^\infty dF P(F) F = \frac{\int_{-\infty}^\infty dF(F/G(F)) \exp\left[ \int_0^F \left( (f - T e^{F'/T})/G(F') \right) dF' \right]}{\int_{-\infty}^\infty dF \exp\left[ \int_0^F \left( (f - T e^{F'/T})/G(F') \right) dF' \right] / G(F)}$$

and the average number of unzipped base pairs $\langle X \rangle$ as

$$\langle X \rangle = \partial_f \langle F \rangle.$$

In the following sections we will present solutions of the stationary Fokker–Planck equation for the cases of Ornstein–Uhlenbeck noise, frozen noise and long range correlated noise.

4. Ornstein–Uhlenbeck noise

The autocorrelation function for Ornstein–Uhlenbeck noise is given in equation (10). Substituting this function into equation (28) for the function $G(F)$ we find

$$G(F) = \frac{D}{1 + \tau e^{F/T}}.$$

Using this in equation (32) for the stationary distribution of the free energy, we find

$$P(F) \sim [1 + \tau e^{F/T}] \exp \left\{ \frac{1}{D} \left[ f F - T(T - f \tau) e^{F/T} - \frac{T^2 \tau e^{2F/T}}{2} \right] \right\}.$$

This agrees exactly with the result (34) of Allahverdyan et al [7], which is an exact result. They have shown in their paper that in the white noise limit, where the correlation length $\tau \to 0$, this distribution of the free energy would lead to the average number of unzipped
5. Long range correlated noise

We will now consider the case where the noise autocorrelation function $K(t)$ is long range and is given by equation (13), with $\alpha < 1$. Substituting equation (13) into (28), we find

$$G(F) = -\sigma e^{-F/T}[1 - \exp(-e^{F/T})] + \sigma(e^{-F/T})^{1-\alpha}\Gamma(1 - \alpha, e^{F/T})$$

(38)

where $\Gamma(a, x)$ is the incomplete $\Gamma$-function

$$\Gamma(a, x) = \int_x^\infty dt t^{a-1}e^{-t}.$$  

(39)

In order to use equations (32) and (38) to calculate the stationary free energy distribution, we need to calculate the integral

$$\int_{F_0}^F \frac{f - Te^{F'/T}}{G(F')} \frac{dF'}{G(F')} = \frac{1}{\sigma} \int_{F_0}^F \frac{e^{F'/T}}{1 - \exp(-e^{F'/T}) + e^{\alpha F'/T}\Gamma(1 - \alpha, e^{F'/T})} dF'.$$

(40)

With the substitution $x' = e^{F'/T}$, this integral becomes

$$\int_{F_0}^F \frac{f - Te^{F'/T}}{G(F')} \frac{dF'}{G(F')} = \frac{T}{\sigma} \int_0^x \frac{(f - Tx')}{1 - e^{-x'} + x'^{\alpha}\Gamma(1 - \alpha, x')} dx'.$$

(41)

Using this and equation (32), the stationary probability distribution function for the free energy becomes

$$P(F) \sim \frac{x}{1 - e^{-x} + x^{\alpha}\Gamma(1 - \alpha, x)} \exp\left\{\frac{T}{\sigma} \int_0^x \frac{f - Tx'}{1 - e^{-x'} + x'^{\alpha}\Gamma(1 - \alpha, x')} dx'\right\},$$

$$x \equiv e^{F/T}.$$  

(42)

Using this, the average free energy is given by

$$\langle F \rangle = \int P(F) F dF = \left[\int_0^\infty T dx (\ln x) \frac{1}{1 - e^{-x} + x^{\alpha}\Gamma(1 - \alpha, x)} \right] \times \exp\left\{\frac{T}{\sigma} \int_0^x \frac{f - Tx'}{1 - e^{-x'} + x'^{\alpha}\Gamma(1 - \alpha, x')} dx'\right\}$$

$$\times \left[\int_0^\infty dx \frac{1}{1 - e^{-x} + x^{\alpha}\Gamma(1 - \alpha, x)} \right] \times \exp\left\{\frac{T}{\sigma} \int_0^x \frac{(f - Tx')}{(1 - e^{-x'} + x'^{\alpha}\Gamma(1 - \alpha, x'))} dx'\right\}^{-1}.$$  

(43)
The average number of unzipped base pairs can then be calculated as

\[
\langle X \rangle = \left[ T \int_0^\infty \frac{dx}{1-e^{-x}} \frac{R(x)}{1-e^{-x}+x^{\alpha}\Gamma(1-\alpha, x)} \right. \\
\times \left. \exp \left\{ \frac{T}{\sigma} \int_0^x \frac{dx'}{1-e^{-x}+x^{\alpha}\Gamma(1-\alpha, x')} \right\} \right] \\
\times \left[ \int_0^\infty \frac{dx}{1-e^{-x}+x^{\alpha}\Gamma(1-\alpha, x)} \right]^{-1}
\]

where

\[
R(x) = \frac{T}{\sigma} \int_0^x \frac{dx'}{1-e^{-x'}+x^{\alpha}\Gamma(1-\alpha, x')}
\]

Equation (44) can be used to calculate the average number of unzipped base pairs numerically.

In figure 1 we show the result for the average number of unzipped base pairs \(\langle X \rangle\) as a function of \(f\), for different values of the parameter \(\alpha < 1\). We can see that for small \(f\), \(\langle X \rangle\) increases as \(\alpha\) increases, and for all \(\alpha < 1\), it is finite. This shows that the DNA double strand is more stable for longer range correlations, in agreement with the finding of Allahverdyan et al [7] and in contrast to the prediction of Lubensky and Nelson [4]. In order to check for any power law dependence of \(\langle X \rangle\) on \(f\), we show in figure 2 a plot of \(-\ln\langle X \rangle\) versus \(\ln f\). Any power law behavior would manifest itself as a straight line. We can see that for small \(f\), we have a straight line, but with zero slope.

Our Fokker–Planck equation does not describe the case of completely frozen noise, \(\eta\) is a constant and equations (4) and (5) for the Hamiltonian and free energy can be solved exactly for a particular value of the noise strength:

\[
H(x) = (x-a)(f+\eta)
\]

\[
Z = \int_a^L dx e^{-\beta(x-a)(f+\eta)} = \frac{1}{\beta(f+\eta)[1-e^{-\beta(f+\eta)(L-a)}]}
\]

\[
-\frac{F}{T} = \ln Z = -\ln[\beta(f+\eta)] + \ln\{1-e^{-\beta(f+\eta)(L-a)}\}.
\]
Force induced unzipping of DNA with long range correlated noise

Figure 1. Average number of unzipped base pairs as a function of \( f \), for different values of \( \alpha \).

Figure 2. \(-\ln\langle X \rangle\) versus \( \ln f \), for different values of \( \alpha \).

For this particular noise strength, the number of unzipped base pairs is given by

\[
X = \partial_f F = \frac{T}{f + \eta} - \frac{L(1 - (a/L))e^{-\beta(f + \eta)(1-(a/L))L}}{1 - e^{-\beta(f + \eta)L(1-(a/L))}}.
\]  

(49)

In the limit \( L \to \infty \) this becomes

\[
X = L\theta(-(f + \eta))
\]

(50)

doi:10.1088/1742-5468/2011/06/P06023
where $\theta(x)$ is the Heaviside step function. Averaging this over the noise, we have

$$
\langle X \rangle = L \int_{-\infty}^{\infty} \frac{d\eta}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{\eta^2}{2\sigma} \right] \theta(-f + \eta) = L \int_{0}^{\infty} \frac{d\eta}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{\eta^2}{2\sigma} \right] \theta(\eta - f)
$$

$$
= \frac{L}{2} \left[ 1 - \text{erf} \left( \frac{f}{\sqrt{2\sigma}} \right) \right].
$$

(51)

This shows that the average number of unzipped base pairs $\langle X \rangle$ depends explicitly on $L$ over the whole range of physical parameters. A similar result was obtained in [7] by means of a more elaborate calculation. This means that a Fokker–Planck approach cannot be used to study the case of frozen noise because in the Fokker–Planck equation the thermodynamic limit $L \to \infty$ has already been taken. There is therefore no dependence on $L$ in the stationary free energy distribution $P(F)$ and consequently also no dependence on $L$ in the quantity $\langle X \rangle$ obtained using this distribution and equations (33) and (34). That the thermodynamic limit does not exist for $\langle X \rangle$ in the case of completely frozen noise is probably related to the unphysical nature of the completely frozen noise itself. We are assuming that for long range noise with $1 > \alpha > 0$, but not completely frozen, the thermodynamic limit for $\langle X \rangle$ exists and is independent of $L$. Otherwise a Fokker–Planck approach cannot be applied at all.

6. Conclusion

In this paper we have presented a detailed study of a model of double-stranded DNA being unzipped by an external force. The model was originally proposed by Allahverdyan et al [7] and is a simplified version of that proposed by Lubensky and Nelson [4], in that it neglects the restoring force due to the single strands after the DNA is partially unzipped. The neglect of configurational entropy contribution of the DNA and also the bubble contribution in the model make it impossible to derive qualitatively correct features of the phase diagram in the temperature–force plane [1]–[3]. Even then, the model reproduces the behavior of the average number of unzipped base pairs $\langle X \rangle \sim 1/f^2$ of the Lubensky–Nelson model in the white noise limit. The simplification in the model facilitates its study and, we hope, will produce an answer as regards the effect of long range correlation on unzipping.

Still, there has been very little work done on this latter problem, besides the heuristic argument study of Lubensky and Nelson [4] and the numerical simulation of Allahverdyan et al [7]. Our Fokker–Planck equation approach may provide a new perspective.

We derive and solve a Fokker–Planck equation for the stationary distribution of the free energy for a model in which the autocorrelation function of the random DNA sequence can be of a general form, including long range correlations. The only approximation that we have made in the derivation is that shown in equation (27). In the case of Ornstein–Uhlenbeck noise, characterized by a finite correlation length, our result reduces to the exact result of Allahverdyan et al, with the average number of unzipped base pairs going as $\langle X \rangle \sim 1/f^2$ in the white noise limit, where $f$ is the deviation from the critical force. In the case of long range correlated noise, where the integrated autocorrelation is divergent, we find that $\langle X \rangle$ is finite at $f = 0$, with its value decreasing as the correlations become of longer range. This shows that long range correlations actually stabilize the DNA sequence against unzipping. Our result is also in agreement with the findings of Allahverdyan et al [7], obtained using numerical generation of the correlated noise, but contradicts the
prediction of Lubensky and Nelson [4], \(\langle X \rangle \sim 1/f^{2/\alpha}, \alpha < 1\). Our result for the long range correlated noise is based on the assumption that for long range noise with \(1 > \alpha > 0\), but not completely frozen, the thermodynamic limit for the average unzipped base pair \(\langle X \rangle\) exists and is independent of \(L\). Otherwise a Fokker–Planck approach cannot be applied at all. Such is the case for completely frozen noise.

Acknowledgments

This research was supported by NSF Epscor LASIGMA and the Louisiana Board of Regents Support Fund Contract Number LEQSF(2007-10)-RD-A-29. PML would like to thank for hospitality the Institute of Theoretical Physics, Chinese Academy of Sciences, where part of this work was carried out.

References

[1] Bhattacharjee S M, 2000 J. Phys. A: Math. Gen. 33 L423
Sebastian K L, 2000 Phys. Rev. E 62 1128
Mukamel E A and Shakhnovich E I, 2001 arXiv:cond-mat/0108447
Orlandini E, Bhattacharjee S M, Marenduzzo D, Maritan A and Seno F, 2001 arXiv:cond-mat/0109521
Cocco S, Marko J F and Monasson R, 2002 arXiv:cond-mat/0206238
[2] Sen P and Bhattacharjee S M, 2002 J. Phys. A: Math. Gen. 35 L141
Bhattacharjee S M and Marenduzzo D, 2001 arXiv:cond-mat/0106110
[3] Marenduzzo D, Trovato A and Maritan A, 2001 Phys. Rev. E 64 031901
Marenduzzo D, Bhattacharjee S M, Maritan A, Orlandini E and Seno F, 2002 Phys. Rev. Lett. 88 028102
[4] Lubensky D K and Nelson D R, 2000 Phys. Rev. Lett. 85 1572
Lubensky D K and Nelson D R, 2002 Phys. Rev. E 65 031917
[5] Essevaz-Roulet B, Bockelmann U and Heslot F, 1997 Proc. Nat. Acad. Sci. 94 11935
Perkins T T, Smith D E and Chu S, 1994 Science 264 819
Merkel R, Nassy P, Leung A, Ritchie K and Evans E, 1999 Nature 397 50
[6] Strick T R, Dessinges M-N, Charvin G, Dekker N H, Allemand J-F, Bensimon D and Croquette V, 2003 Rep. Prog. Phys. 66 1
[7] Allahverdyan A E, Gevorkian Zh S, Hu C-K and Wu M-C, 2004 Phys. Rev. E 69 061908
[8] Hatch K, Daniłowicz C, Coljee V and Prentiss M, 2007 Phys. Rev. E 75 051908
[9] Danilowicz C, Lee C H, Coljee V W and Prentiss M, 2007 Phys. Rev. E 75 030902(R)
[10] Kumar S and Li M S, 2010 Phys. Rep. 486 1
[11] Malacinski G M and Freiefelder D, 1993 Essentials of Molecular Biology (Boston, MA: Jones and Bartlett)
[12] Lu, W, 1992 Int. J. Bifurcation Chaos 2 137
Amato I, 1992 Science 257 74
Li W and Kaneko K, 1992 Eur. Phys. Lett. 17 655
Peng C K, Buldyrev S V, Goldberger A L, Havlin S, Sciortino F, Simons M and Stanley H E, 1992 Nature 356 168
Buldyrev S V, Goldberger A L, Havlin S, Peng C K, Simons M and Stanley H E, 1993 Phys. Rev. E 47 4514
Buldyrev S V, Goldberger A L, Havlin S, Mantegna R N, Matsa M E, Peng C K, Simons M and Stanley H E, 1995 Phys. Rev. E 51 5084
[13] Voss R F, 1992 Phys. Rev. Lett. 68 3805
Lu X, Sun Z R, Chen H M and Li Y D, 1998 Phys. Rev. E 58 3578
[14] Chatzidimitriou-Dreismann C A and Larhammar D, 1993 Nature 361 212
Prabhulu V V and Claverie J M, 1992 Nature 359 782
Mohanty A K and Narayana Rao A V S S, 2000 Phys. Rev. Lett. 84 1832
Audit B, Thermes C, Vaillant C, d’Aubenton-Carafa Y, Muzy J F and Arneodo A, 2001 Phys. Rev. Lett. 86 2471
Guharay S, Hunt B R, Yorke J A and White O R, 2000 Physica D 146 388
[15] Li W, 1997 Comput. Chem. 21 257
[16] Fox R F, 1986 Phys. Rev. A 33 467

doi:10.1088/1742-5468/2011/06/P06023