Dynamics of a bubble formed in double-stranded DNA

Shikha Srivastava and Yashwant Singh(a)

Department of Physics, Banaras Hindu University - Varanasi-221 005, India

received 8 October 2008; accepted in final form 8 January 2009
published online 4 February 2009

PACS 87.14.gk – DNA
PACS 87.15.Ya – Fluctuations
PACS 87.15.ad – Analytical theories

Abstract – We study the fluctuational dynamics of a tagged base pair in double-stranded DNA. We calculate the drift force which acts on the tagged base pair using a potential model that describes interactions at base pairs level and use it to construct a Fokker-Planck equation. The calculated displacement autocorrelation function is found to be in very good agreement with the experimental result of Altan-Bonnet et al. (Phys. Rev. Lett., 90 (2003) 138101) over the entire time range of measurement. We calculate the most probable displacements which predominately contribute to the autocorrelation function and the half-time history of these displacements.

DNA double-stranded helical structure is stabilized by the hydrogen bonding between complementary bases and the stacking between neighbouring bases [1]. In physiological-solvent conditions the average value of these interactions for a base pair is of the order of few $k_B T$ (thermal energy) [2] and thermal fluctuations can lead to local and transitory unzipping of the double strands [3,4]. The co-operative opening of a sequence of consecutive base pairs leads to formation of local denaturation zones (bubbles). The DNA unzipping represents the classic competition between energy and entropy, with denaturation bubbles embodying the entropically favoured, energetically costly state. As an AT base pair connected by two hydrogen bonds needs less energy to unzip compared to a GC base pair which is connected by three hydrogen bonds, initiation of a bubble generally takes place in an AT-rich region. A DNA bubble consists of flexible single-stranded DNA and its size fluctuates by zipping and unzipping of base pairs at the two zipper forks where the bubble connects to the intact double strands. The average size of a bubble depends on the sequence of base pairs, temperature and ionic strength and varies from few broken base pairs at room temperature to few hundred open base pairs close to melting temperature [5,6].

The formations of bubble at room or physiological temperatures are rare and intermittent with life times that may vary from few nanoseconds [7] to few tens of microseconds [4]. The occurrence of such bubble domains is important as the opening of dsDNA base pairs by breaking the hydrogen bonds between complementary bases disrupts the helical stack and may initiate biological processes of transcription, replication and protein binding [8,9]. From the physics point of view, DNA bubbles offer a quasi-one-dimensional system for the study of fluctuational dynamics [4].

In an experiment by Altan-Bonnet et al. (ALK) [4] the dynamics of a single bubble in three synthetic DNA constructs having the same GC-rich region but different AT base pairs regions have been investigated by fluorescence correlation spectroscopy (FCS). In the middle of the AT region a T base pair was tagged with a fluorophore while the neighbouring T base of the other strand was tagged with a quencher. The correlation spectrum of fluctuating base pairs was monitored. The dynamics was found to follow a multi-state relaxation kinetics in a wide temperature range with a characteristic time scale in the range of 20–100 $\mu$s. Several theoretical models [4,10–12] have recently been proposed to explain the observed multi-state breathing dynamics. In one of these models [10] the bubble free energy that corresponds to a bubble of infinitely large size [13,14] and whose accuracy for a bubble of few broken base pairs, to best of our knowledge, is not established has been used. Other theoretical models include discrete master equations and stochastic Gillespie schemes [4,11,12].

In this letter, we develop a general theory to study the fluctuational dynamics of a tagged base pair by means of a Fokker-Planck equation based on a potential field which acts on the base pair and which we obtain by integrating out the degrees of freedom of all base pairs of

(a)E-mail: singhyas_in@yahoo.co.in
a dsDNA except those associated with the tagged one. We use the simple potential model of Peyrard-Bishop-Dauxois (PBD) [15] to represent the interactions in dsDNA at base pairs level.

The PBD model reduces the degrees of freedom of DNA to a one-dimensional chain of effective atom compounds describing the relative base pair separation $y_i$ from the ground state position $y_i = 0$. In spite of its one-dimensional character, the model has been found to give quantitative description of thermodynamic properties of DNA [16–18] indicating that the stretching motion of base pairs is primarily important in the denaturation process. The potential of the model is written as

\[
U(y^N) = \sum_i [D_i(e^{-2\alpha_i y_i} - 2e^{-\alpha_i y_i})] + \frac{\kappa}{2(1 + \rho e^{-\alpha(y_i+y_i-1)}(y_i-y_{i-1})^2)]},
\]

where $N$ is the number of base pairs, summation on the r.h.s. is over all base pairs of the molecule and $y^N = \{y_i\}$, the set of relative base pair separations. The first term of eq. (1) is the Morse potential that represents the hydrogen bonds between the bases of the opposite strands and the second term represents the stacking interaction between adjacent base pairs. The values of the parameters found by Campa and Giansanti [16] are $\kappa = 0.025$ eVÅ$^{-2}$, $\rho = 2$ and $\alpha = 0.35$ Å$^{-1}$ for the stacking part, while for the Morse potential $D_{AT} = 0.05$ eV, $a_{AT} = 4.2$ Å$^{-1}$ for an AT base pair and $D_{GC} = 0.075$ eV and $a_{GC} = 6.9$ Å$^{-1}$ for a GC base pair.

We now consider one of the DNA molecules (named $A_{18}$) investigated by ALK [4] and take the 17th base pair counted from the 5′-end as the tagged base pair. The interactions in the molecule is represented by the PBD model. We add a harmonic potential $u_h(y_N) = \theta(y_N - 2\pi \theta/2)$ where $\theta(y)$ is a Heaviside step function at the terminal GC base pair to avoid the complete separation of the two strands. In experiment [4] this was achieved by attaching a hairpin loop of 4T. It may be noted that apart from preventing the complete separation of the two strands the harmonic potential has no effect on the results reported below. We have checked this by varying the value of $c$ from $0.5 \times 10^{-2}$ Å$^{-2}$ to $2.0 \times 10^{-2}$ Å$^{-2}$. The potential felt by the tagged base pair at a separation $y$ from the ground state $y = 0$ is found from the relation

\[
V(y) = -k_BT[\ln Z_n(y) - \ln Z_n(0)],
\]

where

\[
Z_n(y) = \int \prod_{i=1}^N dy_i \delta(y_n-y) \exp[-\beta U(y^N)],
\]

\[
Z_n(0) = \int \prod_{i=1}^N dy_i \delta(y_n-0) \exp[-\beta U(y^N)],
\]

are the constrained partition function integrals, $\delta$ is the Dirac function and $\beta = (k_BT)^{-1}$.

For the PBD model the calculation of a partition function integral reduces to multiplication of $N$ matrices. The discretization of the co-ordinate variable and introduction of a proper cut-off on the maximum values of $y$'s determines the size of the matrices. We have taken $-2\AA$ and $120\AA$ as the lower and upper limit of integration for each co-ordinate variable and discretized space using the Gaussian-Legendre method with number of grid points equal to 900. Note that the values of the partition function integrals and therefore the values of $V(y)$ are independent of the limit of integration. We show in fig. 1 the value of $V(y)$ as a function of $y$ at $45^\circ$. At a separation $y$ the base pair feels a drift force $F = -\partial V(y)/\partial y$ towards the origin $y = 0$. As shown in the inset of fig. 1 this force has a minimum at $y = 0.2\AA$. This minimum corresponds to a force barrier which has been observed in theoretical investigation of force induced unzipping of a dsDNA in the constant extension ensemble [19–21] and is attributed to a combination of the force needed to break the hydrogen bonds and the force needed to overcome the entropic barrier of the stacking interaction [19]. In addition to a minimum at $y = 0.2\AA$ the drift force has a very swallow minimum at $y \sim 1.2\AA$. This minimum may play a role in determining the size of the most probable separation of the base pair and its survival probability.

The relaxation dynamics of a bubble as found in [4] is slow and occurs on the time scale of tens of microseconds ($\mu$s) which is much slower than the relaxation of the polymer degrees of freedom (Rouse modes) of the bubble. This slow bubble dynamics has been interpreted as a random walk in 1D co-ordinate describing the number of denatured base pairs [4,10–12]. Since in ALK [4] experiment the autocorrelation function of base pair fluctuations has been measured, a theory which describes the base pair dynamics is needed. In describing the stochastic dynamics of zipping and unzipping of a base pair one should, however, note that the motions of bases of the pair are intimately connected to the motions of the two zipper forks which connect the two ends of the bubble with the intact double strands and move along the length of the DNA molecule. Therefore in the Langevin equation which

![Fig. 1: The effective potential felt by the tagged base pair in a dsDNA molecule ($A_{18}$ of [4]) of 29 base pairs at separation $y$ at $45^\circ$. In the inset the drift force $F(y) = -\partial V(y)/\partial y$ which drives the base pair to the equilibrium position is plotted.](image-url)
may be used to describe the dynamics of the base pair,
\[
\frac{dy}{dt} = -\Gamma \frac{dV(y)}{dy} + \xi(t) ; \quad \langle \xi(t) \rangle(t) = 2T\Gamma \delta(t),
\]
the transport coefficient \( \Gamma \) accounts not only the frictional forces encountered by the bases of the pair but also the forces which hinder the motion of the two zipper forks.

We therefore note that \( \Gamma \) is not simply the mobility of an isolated base pair but represents the complex dynamics of the bubble. The time scale of the dynamics is set by \((\Gamma T a^2)\sim \mu s \) [4]. We use \( a_{AT} \) and \( \Gamma T a^2_{AT} \) to make, respectively, distance and time dimensionless. The Fokker-Planck equation corresponding to (3) is found to be
\[
\frac{\partial P}{\partial t} = \frac{\partial}{\partial y} \left[ -\beta V(y) \frac{\partial P}{\partial y} \right] + \frac{\partial^2 P}{\partial y^2},
\]
where \( P(y, y_0; t) \) is the probability density of the random walkers.

We assume that if separation \( y \) reduces to zero at time \( t' \), it will not contribute to autocorrelation function defined as \( C(t) = \langle y(t)y(0) \rangle - \langle y \rangle^2 \) for \( t > t' \) and similarly any new fluctuational opening which appear after \( t = 0 \) will not contribute to \( C(t) \). Thus for purposes of computing the autocorrelation function we place an absorbing wall at \( y = 0 \), i.e. \( P(y=0, t) = 0 \). In addition to this we may require \( P(y=L, t) = 0 \), where \( L \) depends on the size of the dsDNA molecule or on any other condition which limits the size of the bubble. The problem of calculating the autocorrelation function \( C(t) \) therefore reduces to finding how many walkers of an ensemble of random walkers distributed according to thermal equilibrium distribution at \( t=0 \) are still present at time \( t \) and have not been absorbed by the wall at \( y = 0 \) [22].

When a substitution \( P = \exp[-\beta V(y)/2] \psi \) is used eq. (4) reduces to
\[
-\frac{\partial \psi}{\partial t} = H\psi, \quad H = -\frac{\partial^2}{\partial y^2} + v(y),
\]
where
\[
v(y) = \frac{1}{4} \left[ \frac{\partial \beta V(y)}{\partial y} \right]^2 - \frac{1}{2} \left[ \frac{\partial^2}{\partial y^2} V(y) \right].
\]

This is the imaginary time Schrödinger equation for a particle of mass 1/2 in the potential \( v(y) \). Let \( \phi_m(y) \) denote the eigenfunctions of the operator \( H \), \( H\phi_m = E_m\phi_m \), with \( \phi_m(y=0) = 0 \) and \( \int dy \phi^\star_m(y) \phi_m(y) = \delta_{nm} \).

Then expanding \( \psi(y, t) \) in terms of eigenfunctions \( \phi_m \) and using the initial condition \( P(y, y_0, t=0) = \delta(y-y_0) \) the transition probability from initial separation \( y_0 \) to a final separation \( y \) at time \( t \) is found to be
\[
P(y, y_0, t) = \exp \left[ -\int_0^t \beta V(y) + V(y_0) \right] \sum_m e^{-E_m t} \phi_m(y_0) \phi^\star_m(y) \]
\[
\times \sum_m e^{-E_m t} \phi_m(y_0) \phi^\star_m(y_0).
\]

For initial distribution of separation \( y_0 \) we choose the Boltzmann factor \( B(y) = A \exp(\beta V(y)) \), where \( A = 1/\int_0^L dy \exp(\beta V(y)) \) is a normalization factor. If we start with the equilibrium distribution function at \( t=0 \), the distribution function at time \( t \), \( P(y, t) \) is
\[
P(y, t) = A \sum_m e^{-E_m t} \int_0^L dy_0 \exp \left[ -\beta V(y) - \frac{V(y_0)}{2} \right] \phi_m(y) \phi^\star_m(y_0); \quad (8)
\]
\( P(y, t) \) measures the survival probability.

For the autocorrelation function we get
\[
C(t) = \int_0^L P(y, t) dy
\]
\[
= A \sum_m e^{-E_m t} \int_0^L e^{-\beta V(y)/2} \phi_m(y) dy \right)^2 \].
\]

The values of \( \phi_m(y) \) and \( E_m \) of the operator \( H \) in eq. (5) are determined numerically using a method developed by Sethia et al. [23]. As shown in fig. 2(a), \( v(y) \) is attractive at small \( y \), rises to a (repulsive) maximum at \( y = 0.2 \) \( \AA \) and then decays to zero as \( y \) increases. The maximum in \( v(y) \) corresponds to the minimum in \( F \) shown in fig. 1. For small values of \( m, \phi_m(y) \) remains confined (see fig. 2(b)) in a region of separation whose values are smaller than the length of the molecule \( L \). After the first three eigenvalues
Shikha Srivastava and Yashwant Singh

Fig. 3: Rescaled autocorrelation function \( g(u) = G(t/t_{1/2}) \) where \( G(t) = C(t)/C(0) \) and \( t_{1/2} \) is such that \( G(t_{1/2}) = 0.5 \) as a function of \( t/t_{1/2} \) at 45 °C. The dashed line represents the experimental results [4] (without error bars) of the \( A_{18} \) molecule. The notations used here are same as in [4].

![Rescaled autocorrelation function](image)

In fig. 3 the rescaled autocorrelation function \( g(u) = G(t/t_{1/2}) \) where \( G(t) = C(t)/C(0) \) and \( t_{1/2} \) is such that \( G(t_{1/2}) = 0.5 \) [4], is plotted as a function of rescaled time \( t/t_{1/2} \) and compared with the experimental results of \( A_{18} \) molecule [4]. Large fluctuations particularly in the short- and long-time range have been found in experimental data. The results reproduced here from fig. 3(B) of [4] are the average values without the error bars. Without using any fitting parameters, we find a very good agreement between theory and experiment over the entire time range of measurement. If we choose \( \Gamma k_B T a_{\Delta T}^2 = 10 \) and plot \( G(t) \) as a function of time \( t \) in ms the resulting curve is found to be in very good agreement with the corresponding curve given in [4].

In fig. 4(a) we show the distribution function \( P(y, t) \) which gives the probability of separation \( y \) of the tagged base pair at time \( t \). From the figure we find that the most probable separation is \( y^* \sim 1 \) Å, although the term “most probable” makes less and less sense because the peak gets broader and broader. Thus, initially as well as presently small separation of the order of 1 Å make the most contributions to the autocorrelation function \( C(t) \) at all times. This can be understood from the nature of the drift force \( F(y) \) (shown in fig. 1) which favours small separation. Since small separations have larger Boltzmann weights initially, they dominate \( C(t) \) at all times. If we plot \( \ln y^* \) vs. \( t \) we find a straight line having a slope equal to 1/6. The slow growth of the most probable displacement with time indicated by the exponent 1/6 can be associated with the shape of the repulsive part of \( v(y) \) for \( y \geq 0 \) which pushes the displacement to grow. To check this we calculated the dependence of the most probable displacement on time taking \( v(y) \) equal to constant independent of \( y \) for \( y \leq 0 \) and found the value of exponent equal to zero.

The half-time history of a random walker that is at \( y^* \) at \( t = 0 \) and at \( t \) is defined as [22]

\[
H(y, t/2|y^*, t; y^*, t = 0) = P(y^*, t|y, t/2)P(y, t/2; y^*, t = 0)
\]

\[
= \sum_m e^{-E_m t} \phi_m(y^*) \phi_m(y)^2. \tag{10}
\]

We plot the half-time distribution as a function of \( y \) for \( t = 5, 10 \) and 20 in fig. 4(b). While values of \( y^* \) corresponding to these times are 0.84, 1.01 and 1.24 the peaks in \( H \) are found, respectively, at 0.90, 1.12 and 1.48 which are somewhat larger than the corresponding values of \( y^* \). The half-width of the distribution \( H(y, t/2|y^*, t; y^*, 0) \) is found to be narrower than that of \( P(y, t) \). Therefore the most probable way for a displacements of size \( y^* \) formed at \( t = 0 \) to survive until a time \( t \) is that they first grow larger than \( y^* \) and then shrink back to the original size.

In conclusion; we developed a theory to describe the multi-state relaxation dynamics of a tagged base pair of dsDNA. We used a potential model which describes interactions in dsDNA at base pairs level and calculated the drift force which acts on the base pair and drives it to its equilibrium position. The dynamics is governed
by the Langevin equation with Gaussian white noise and a transport coefficient associated with the complex dynamics of the bubble centered around the tagged base pair. We derived the associated Fokker-Planck equation and with suitable transformation reduced it to an imaginary time Schrödinger equation for a particle of mass 1/2. We found the eigenvalues and eigenfunctions of the operator using a numerical method described in [23]. The calculated displacement autocorrelation function is found to agree with experimental result for the entire time range of measurement. The most probable displacements which contribute predominately to short as well as long times are found to be small, of the order of 1 Å. The half-time distribution of these displacements which show how the most probable displacements behave between time \( t = 0 \) and \( t \) are calculated. The method developed here is equally applicable to homogeneous and heterogeneous DNA molecules.

***

We thank N. Singh for his help in computation and A. K. Ganguly for useful discussions. This work is supported by a research grant from DST of the Government of India, New Delhi.

REFERENCES

[1] Saenger W., Principle of Nucleic Acid Structure (Springer Verlag, Berlin) 1984.
[2] Sontalucia J. Jr., Proc. Natl. Acad. Sci., U.S.A., 95 (1998) 1460; Pincet F., Perez E. and Bryant G. et al., Phys. Rev. Lett., 73 (1994) 2780; Krueger A., Protozanova E. and Frank-Kamenetskii M. D., Biophys. J., 90 (2006) 3091.
[3] Campa A., Phys. Rev. E, 63 (2001) 021901; Peyrard M., Europhys. Lett., 44 (1998) 271.
[4] Altan-Bonnet G., Libchaber A. and Krichevsky O., Phys. Rev. Lett., 90 (2003) 138101.
[5] Guéron M., Kochoyan M. and Leroy J. L., Nature (London), 328 (1987) 89.
[6] Wartall R. M. and Benight A. S., Phys. Rep., 126 (1985) 67.
[7] Guéron M. and Leroy J. L., Methods Enzymol., 261 (1995) 383; Coman D. and Russu I. M., Biophys. J., 89 (2005) 3285.
[8] Kornberg A. and Baker T. A., DNA Replication (W. H. Freeman, New York) 1992.
[9] Robert R. J. and Chang X., Annu. Rev. Biochem., 67 (1998) 181; Stivers J. T., Nucleic Acid. Res. Mol. Biol., 77 (2004) 37; Léger J. F. et al., Proc. Natl. Acad. Sci., U.S.A., 95 (1998) 12295.
[10] Fogedby H. C. and Metzler R., Phys. Rev. Lett., 98 (2007) 070601; Phys. Rev. E, 76 (2007) 061915.
[11] Bicout D. J. and Kats E., Phys. Rev. E, 70 (2004) 010902.
[12] Ambjörnson T. et al., Phys. Rev. Lett., 97 (2006) 128105; Biophys. J., 92 (2007) 2674.
[13] Poland D. and Scheraga H. A., Theory of Helix-Coil Transition in Biopolymers (Academic Press, New York) 1970.
[14] Vanderzande C., Lattice Models of Polymers (Cambridge University Press, Cambridge) 1998.
[15] Peyrard M. and Bishop A. R., Phys. Rev. Lett., 62 (1989) 2755; Dauxois T., Peyrard M. and Bishop A. R., Phys. Rev. E, 47 (1993) 684.
[16] Campa A. and Giansanti A., Phys. Rev. E, 58 (1998) 3585; Singh N. and Singh Y., Phys. Rev. E, 64 (2001) 042901.
[17] Weber G. et al., Nat. Phys., 2 (2006) 55.
[18] Hanne J. et al., Phys. Rev. E, 76 (2007) 011909; Alexander B. S. et al., Phys. Rev. E, 74 (2006) 050901(R).
[19] Cocco S., Monasson R. and Marko J. F., Proc. Natl. Acad. Sci., U.S.A., 98 (2001) 8608; Phys. Rev. E, 65 (2002) 041907.
[20] Voulgarakis N. K. et al., Phys. Rev. Lett., 96 (2006) 248101.
[21] Singh N. and Singh Y., Eur. Phys. J. E, 17 (2005) 7; 19 (2006) 233.
[22] Tang C., Nakanishi H. and Langer J. S., Phys. Rev. A, 40 (1989) 995.
[23] Sethia A., Sanyal S. and Singh Y., J. Chem. Phys., 93 (1990) 7268.