PRAMANA
— journal of
physics

Vol. 61, No. 2
August 2003
pp. 345–352

Statistical mechanics of thermal denaturation of DNA oligomers

NAVIN SINGH and YASHWANT SINGH
Department of Physics, Banaras Hindu University, Varanasi 221 005, India
Email: ysingh@bhu.ac.in

Abstract. Double stranded DNA chain is known to have non-trivial elasticity. We study the effect of this elasticity on the denaturation profile of DNA oligomer by constraining one base pair at one end of the oligomer to remain in unstretched (or intact) state. The effect of this constraint on the denaturation profile of the oligomer has been calculated using the Peyrard–Bishop Hamiltonian. The denaturation profile is found to be very different from the free (i.e. without the constraint) oligomer. We have also examined how this constraint affects the denaturation profile of the oligomer having a segment of defect sites located at different parts of the chain.

Keywords. DNA; elasticity; defect; denaturation.

PACS Nos 87.10.+e; 63.70.+h; 64.70.-p

1. Introduction

DNA is one of the most complex and important biomolecules as it is central to all living beings. It contains all the information needed for birth, development, living and probably sets the average life. Structurally it is a giant double stranded linear molecule with length ranging from 2 μm for simple viruses to 3.5×10^7 μm for more complex organisms [1]. How such a molecule came into being during the evolution of life and how it acquired the ability to store and transmit the genetic information are still a mystery. The recent progress in genome mapping and availability of experimental techniques to study the physical properties of a single molecule [2,3] has, however, made the field very active from both biological and physical points of view.

A DNA molecule is not just a static object but a dynamical system having a rather complex nature of internal motions [4]. The structural elements such as individual atoms, groups of atoms (bases, sugar rings, phosphates), fragments of double chain including several base pairs, are in constant movement and this movement plays an important role in the functioning of the molecule. The solvent in which the molecule is immersed acts as a thermal bath and provides energy for different motion. In addition, collision with the molecules of the solution which surrounds DNA, local interactions with proteins, drugs or with some other ligands also lead to internal motion. These motions are distinguished by activation energies, amplitudes and characteristic times. The motions which are of our interest here are opening of base pairs, formation of bubbles along the chain and unwinding of helix (denaturation). The energy involved for these motions is of the order of 5–20 kcal/mol. These

motions are activated by increasing temperature, increasing pH of the solvent, action of denaturation agents etc. The time scale of these motions are of the order of microseconds and is therefore generally unobservable in atomistic simulations as these simulations are restricted to time scale of nanoseconds because of computational cost.

The nature of thermal denaturation leading to separation of two strands has been investigated for several decades [5,6]. Experimentally a sample containing molecules of a specific length and sequence is prepared. Then the fraction of bound base pairs as a function of temperature, referred to as the melting curve, is measured through light absorption, typically at about 260 nm. For heterogeneous DNA, where the sequence contains both AT and GC pairs, the melting curve exhibits a multistep behaviour consisting of plateaus with different sizes separated by sharp jumps. These jumps have been attributed to the unwinding of domains characterized by different frequencies of AT and GC pairs. The sharpness of the jump in long DNA molecules suggests that the transition from bound to unbound is first-order. The understanding of this remarkable one-dimensional cooperative phenomenon in terms of standard statistical mechanics, i.e., a Hamiltonian model with temperature independent parameters is a subject of current interest [7].

2. Model Hamiltonian and its properties

Since the internal motion that is basically responsible for denaturation is the stretching of bases from their equilibrium positions along the direction of the hydrogen bonds that connect the two bases, a DNA molecule can be considered as a quasi-one-dimensional lattice composed of N base pair units. The forces which stabilize the structure are the hydrogen bonds between complementary bases on opposite strands and stacking interactions between nearest-neighbour bases on opposite strands. Each base pair is in one of the two states: either open (non-hydrogen bonded) or intact (hydrogen bonded). A Hamiltonian model that has been found appropriate to include these interactions and to describe the displacement of bases leading to denaturation is the Peyrard–Bishop model (PB model) [8]. The PB model is written as

$$H = \sum_{i=1}^{N} H(y_i, y_{i+1})$$

$$H(y_i, y_{i+1}) = \frac{p_i^2}{2m} + W(y_i, y_{i+1}) + V(y_i), \tag{2.1}$$

where m is the reduced mass of the base pair, y_i denotes the stretching of the hydrogen bonds connecting the two bases of the ith pair and

$$p_i = m\left(\frac{\mathrm{d}y_i}{\mathrm{d}t}\right).$$

The on-site potential $V(y_i)$ describes the interaction of the two bases of the ith pair. The Morse potential

$$V(y_i) = D_i(\mathrm{e}^{-ay_i} - 1)^2, \tag{2.2}$$

which is taken to represent the on-site interaction, represents not only the hydrogen bonds connecting two bases belonging to opposite strands, but also the repulsive interactions

of the phosphates and the surrounding solvent effects. The flat part at large values of the displacement of this potential emulates the tendency of the pair 'melt' at high temperatures as thermal phonons drive the molecules outside the well and towards the flat portion of the potential.

The stacking interactions are contributed by dipole–dipole interactions, π-electron systems, London dispersion forces and in water solution, the hydrophobic interactions. These forces result in a complex interaction pattern between overlapping base pairs, with minimum energy distance close to 3.4 Å in the normal DNA double helix. The following anharmonic potential model mimics these features of the stacking energy:

$$W(y_i, y_{i+1}) = \frac{1}{2}k[1+\rho\exp(-\alpha(y_i+y_{i+1}))](y_i-y_{i+1})^2, \tag{2.3}$$

where k is the coupling constant and the second term in the bracket represents the anharmonic term. When the hydrogen bonds connecting the bases break due to stretching, the electronic distribution on the bases is modified causing the stacking interactions to decrease. This is taken into account by the exponential term in eq. (2.3). It may be noted that the effective coupling constant decreases from $k(1+\rho)$ to k when either of the two interacting base pairs is stretched. This decrease in coupling provides a large entropy in the denaturation. The parameter α in eq. (2.3) defines the 'anharmonic range'.

The model described above can be viewed as a model of one-dimensional monoatomic lattice (see figure 1) with each atom having mass m and nearest-neighbour interaction given by eq. (2.3). Furthermore, each atom is subjected to 'external' potential given by eq. (2.1) whose effect is to confine the chain in the potential well. The melting takes place because of the competition between the thermal energy which leads to displacement and 'external' field and the nearest-neighbour interactions which lead to the confinement. The model, therefore, represents a one-dimensional system that differs from the usual one-dimensional systems that do not show phase transitions.

The model Hamiltonian of eq. (2.1) has extensively been used to study the melting profile of a very long ($N \to \infty$) homogeneous DNA chain using both statistical mechanical calculations and constrained temperature molecular dynamics [9,10]. Analytical investigation of non-linear dynamics of the model suggests that intrinsic energy localization can initiate the denaturation [11]. The model for a long homogeneous chain exhibits a peculiar type of first-order transition with finite melting entropy, a discontinuity in the fraction of bound pairs and divergent correlation lengths. However, as the value of the stacking parameter α increases and the range of the 'entropy barrier' becomes shorter than or comparable to the range of the Morse potential, the transition changes to second order. The cross-over is seen at $\alpha/a = 0.5$ [7]. Though the PB model seems capable of explaining the multi-step melting in a sequence-specific disorder [12], how this disorder affects the nature of transition has yet to be understood. In other work the PB model has been used to understand the melting profile of short chains [13] and the effect of defects on this profile [14].

3. Denaturation profile

In a given system of DNA chains, the average fraction θ of bonded base pairs can be written as $\theta = \theta_{\text{ext}}\theta_{\text{int}}$. θ_{ext} represents the average fraction of the strands forming duplexes (double strands), while θ_{int} is the average fraction of unbroken bonds in the duplexes. The

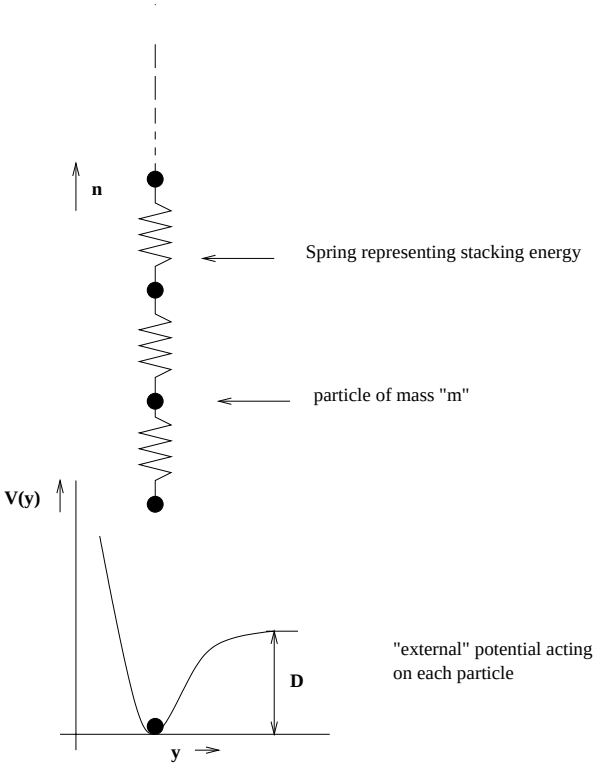

Figure 1. Schematic of quasi-one-dimensional model of DNA.

equilibrium dissociation of the duplexes C_2 to single strand C_1 may be represented by the relation $C_2 \rightleftharpoons 2C_1$. The dissociation equilibrium can be neglected in the case of long chains as θ_{ext} is practically 1 while θ_{int} and therefore θ goes to zero. This is because in the case of long DNA chains when θ goes practically from 1 to zero near the denaturation transition, while most bonds are disrupted and the DNA has denatured, the few bonds still remaining prevent the two strands from getting apart from each other. There will be real separation only at $T \gg T_{\text{m}}$ (T_{m} being the melting temperature at which half of the bonds are broken). Therefore, at the transition the double strand is always a single molecule and in calculation based on PB model one has to calculate only $\theta_{\text{int}} (\equiv \theta)$. On the contrary, in the case of short chains the processes of single bond disruption and strand dissociation tend to happen in the same temperature range, therefore, the computation of θ_{ext} in addition to θ_{int} is essential.

Unfortunately, at present, we do not have any reliable method for calculating θ_{ext}. The method which has been used is based on the partition function of rigid molecules and adjustable parameters [5,13] to be determined from experimental data. To avoid this shortcoming of the theory we in the present article discuss the denaturation profile of oligonucleotides of a given sequence with a base pair at one end of the chain held in such a way that it remains at its equilibrium seperation (i.e. there is no stretching) at all temperatures. This can be done by creating a deep potential well for this base pair or attaching one end of both strands to a substrate. This will be referred to as chain with constraint in order to

distinguish it from the 'free chain'. One of the advantage of having a constraint of this type is obvious; the problem of divergence of the partition function for the PB model for short chains no more exists.

The DNA molecule is known to have non-trivial elastic properties: When the two strands of the DNA molecule are pulled apart by applying a force at one end of the chain, a novel phase transition is found (in case of infinitely long chain) to take place at which the two strands are pulled completely apart [15]. The phase digram plotted in the plane of temperature and force reveals the elastic properties of the DNA chain. Our study reported in this paper differs from the situation just described in two ways: (i) a short chain of 21 given base sequence is considered and (ii) instead of pulling the chain apart, the end base pair is constrained to be in unstreched or intact position.

The oligonucleotide which we consider has the following sequence given by:

$$\begin{aligned} &^{5'}\mathrm{A\,C\,G\,C\,T\,A\,T\,A\,C\,T\,C\,A\,C\,G\,T\,T\,A\,A\,C\,A\,G}_{3'} \\ &_{3'}\mathrm{T\,G\,C\,G\,A\,T\,A\,T\,G\,A\,G\,T\,G\,C\,A\,A\,T\,T\,G\,T\,C}^{5'}, \end{aligned} \tag{3.1}$$

The denaturation profile of this oligonucleotide has been studied by Campa and Giansanti [13] and by us [14]. We take the same parameters as in the previous study. Thus, $D_{\mathrm{AT}} = 0.05$ eV, $D_{\mathrm{GC}} = 0.075$ eV, $a_{\mathrm{AT}} = 4.2$ Å^{-1}, $a_{\mathrm{GC}} = 6.9$ Å^{-1}, $k = 0.025$ eV Å^{-2}, $\rho = 2$ and $\alpha = 0.35$ Å^{-1}. When one end of the chain is held fixed at $y = 0$ distance, the fraction of intact bonds is found using the relation

$$\theta = \frac{1}{N}\sum_{i=1}^{N}\langle\vartheta(y_0 - y_i)\rangle, \tag{3.2}$$

where $\vartheta(y)$ is Heaviside step function, N the total number of base pairs and the canonical average $\langle\cdot\rangle$ is evaluated by considering all the configuration of the chain with one end fixed. The ith bond is considered bound if the value of y_i is smaller than a chosen threshold y_0. For y_0 we have taken a value of 2 Å. Since the model Hamiltonian of eq. (2.1) couples only the nearest neighbours the calculation of canonical average reduces to multiplication of finite matrices. The discretization of the coordinate variables and introduction of a proper cut-off on the matrices and the number of base pairs in the chain fixes the number of matrices to be multiplied.

In figure 2 we compare the value of θ as a function of temperature for the free and the constrained chain. The influence of the constraint on the denaturation profile is enormous. While in case of free chain the breaking of bonds takes place in a narrow temperature range, in the constrained chain it is over a wide temperature range. The value of temperature at which $\mathrm{d}\theta/\mathrm{d}T$ is maximum shifts by about 45°C and the peak is much wider and smaller compared to that of free chain.

Next we calculated the denaturation profile for the following two nucleotides having a segment of chain with defect sites:

$$\begin{aligned} &\text{(a)}\ ^{5'}\mathrm{A\,C\,G\,C\,T\,A\,T\,A\,C\,T\,C\,A\,C\,G\,T\,T\,A\,A\,C\,A\,G}_{3'} \\ &\quad _{3'}\mathrm{T\,C\,G\,C\,T\,A\,T\,A\,C\,T\,C\,T\,G\,C\,A\,A\,T\,T\,G\,T\,C}^{5'} \end{aligned} \tag{3.3}$$

$$\begin{aligned} &\text{(b)}\ ^{5'}\mathrm{A\,C\,G\,C\,T\,A\,T\,A\,C\,T\,C\,A\,C\,G\,T\,T\,A\,A\,C\,A\,G}_{3'} \\ &\quad _{3'}\mathrm{T\,G\,C\,G\,A\,T\,T\,A\,C\,T\,C\,A\,C\,G\,T\,T\,T\,T\,G\,T\,C}^{5'}. \end{aligned} \tag{3.4}$$

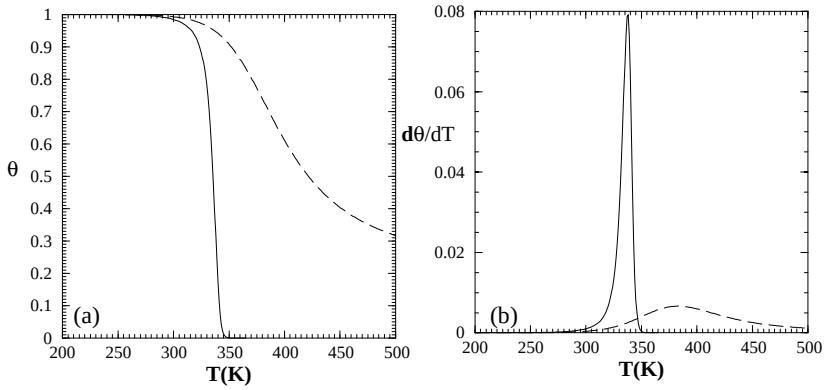

Figure 2. Variation of θ and $\mathrm{d}\theta/\mathrm{d}T$ as a function of temperature for free (solid line) and constrained chain (dashed line).

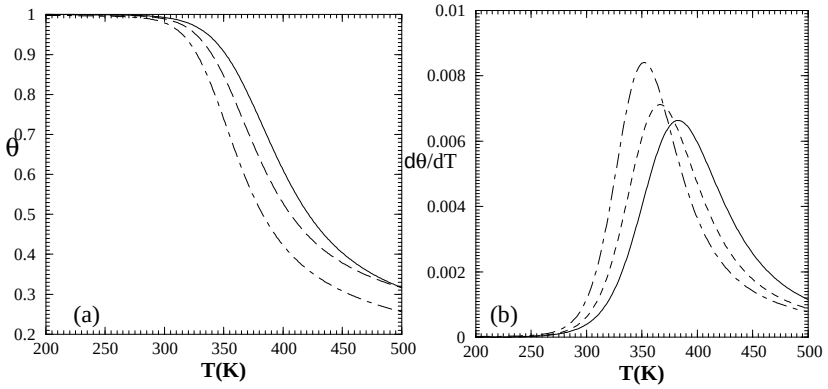

Figure 3. (**a**) Variation of θ as a function of temperature for the constrained chain without defects (solid line) with 10 defects at one end of the chain (dashed line) and 10 defects in the middle (dot–dashed line). (**b**) Plot of $\mathrm{d}\theta/\mathrm{d}T$ for the same.

While both oligonucleotides have 10 defect sites, their locations differ. In (a) the defects are on the left end from site 2 to 11 while in (b) it is in the middle from site 6 to 16. The position of the base pair is counted from left. The purpose is to see how these defects affect the denaturation profile and the formation of loop and stem as often seen in a single strand DNA or RNA. In figure 3 we plot the variation of θ and $\mathrm{d}\theta/\mathrm{d}T$ as functions of temperature. While the denaturation takes place at lower temperatures the shift is more when the defects are in the middle.

The other quantities of interest are the mean value of displacement of nth base pairs defined as

$$\langle y_n \rangle = \frac{1}{Z} \int \left(\prod_{i=1}^{N} \mathrm{d}y_i \right) y_n \exp \left[-\beta \sum_{i=1}^{N} H(y_i, y_{i+1}) \right] \tag{3.5}$$

and its fluctuations

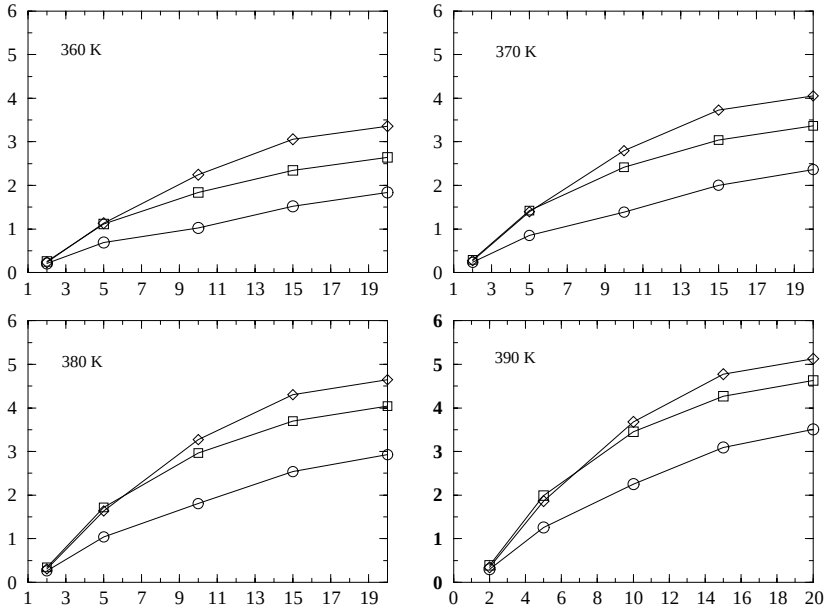

Figure 4. Plot of $\langle y_n \rangle$ vs. n (site position) at four different temperatures. ∘ denotes the constrained chain without defect, □ denotes the chain with 10 defects at one end while ◇ denotes the chain with 10 defects in the middle.

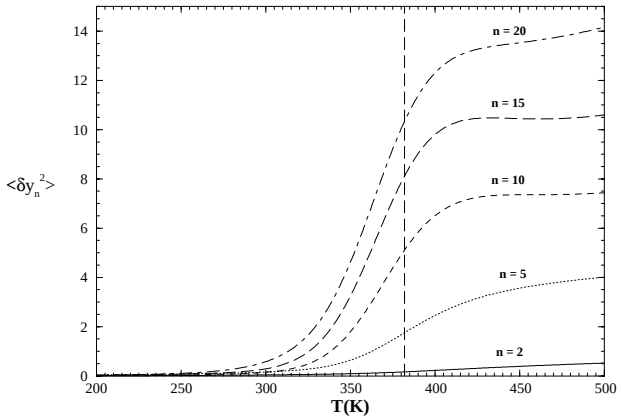

Figure 5. Plot showing the variation of transverse correlation length with temperature for different values of n.

$$\langle |\delta y_n|^2 \rangle = \frac{1}{Z} \int \left(\prod_{i=1}^{N} \mathrm{d}y_i \right) (y_n - \langle y_n \rangle)^2 \exp \left[-\beta \sum_{i=1}^{N} H(y_i, y_{i+1}) \right]. \tag{3.6}$$

Here $Z = \int \left(\prod_{i=1}^{N} \mathrm{d}y_i \right) \exp[-\beta \sum_{i=1}^{N} H(y_i, y_{i+1})]$ is the partition function of the chain.

Because of the constraint that the first base pair of the chain is in unstretched condition, the value of $\langle y_n \rangle$ as well as $\langle |\delta y_n|^2 \rangle$ depend on the site n. In figure 4 we plot the value of $\langle y_n \rangle$ as a function of n at several temperatures for all three constrained chain as described above. We find in all cases that the opening of the chain starts from the open end of the chain. In figure 5 we plot $\langle |\delta y_n|^2 \rangle$ as a function of temperature for several base pairs. The quantity $\langle |\delta y_n|^2 \rangle$ measures the transverse correlation length for the base pair n. This correlation length remains almost zero for all n when the oligonucleotide is in the native state but at the denaturation its value increases. At a given temperature the value of $\langle |\delta y_n|^2 \rangle$ depends on n and it increases with n. For a long chain we expect it to diverge for large value of n.

In conclusion we wish to emphasize that the effect of constraining a base pair at one end of a given oligonucleotide has a very significant effect on the denaturation profile.

Acknowledgements

This work has been supported through research grants by Department of Science and Technology and Council of Scientific and Industrial Research, New Delhi, Government of India.

References

[1] L Stryer, *Biochemistry* (W H Freeman and Company, New York, 1995)
[2] S B Smith, L Finzi and C Bustamante, *Science* **258**, 1122 (1992)
T R Strick *et al*, *Science* **271**, 1835 (1996)
[3] U Bockelmann, B Essevaz-Roulet and F Heslot, *Phys. Rev. Lett.* **79**, 4489 (1997) and references therein
[4] L V Yakushevich, *Non-linear physics of DNA* (John-Wiley & Sons, 1998)
[5] R M Wartell and A S Benight, *Phys. Rep.* **126**, 67 (1985)
[6] N Theodorakopoulos, cond-mat/0210188 and references therein
[7] N Theodorakopoulos, T Dauxois and M Peyrard, *Phys. Rev. Lett.* **85**, 6 (2000)
[8] M Peyrard and A R Bishop, *Phys. Rev. Lett.* **62**, 2755 (1989)
[9] T Dauxois and M Peyrard, *Phys. Rev. Lett.* **70**, 3935 (1993)
[10] T Dauxois, M Peyrard and A R Bishop, *Phys. Rev.* **E47**, R44 (1993)
[11] T Dauxois and M Peyrard, *Phys. Rev.* **E51**, 4027 (1995)
[12] D Cule and T Hwa, *Phys. Rev. Lett.* **79**, 2375 (1997)
[13] A Campa and A Giansanti, *Phys. Rev.* **E58**, 3585 (1998)
[14] Navin Singh and Yashwant Singh, *Phys. Rev.* **E64**, 42901 (2001)
[15] David K Lubensky and David R Nelson, *Phys. Rev. Lett.* **85**, 1572 (2000) and references therein