Investigation of thermal denaturation of DNA molecules based on non-equilibrium transport approach

1994 J. Phys. A: Math. Gen. 27 6147

(http://iopscience.iop.org/0305-4470/27/18/023)

J. Phys. A: Math. Gen. 27 (1994) 6147–6156. Printed in the UK

Investigation of thermal denaturation of DNA molecules based on non-equilibrium transport approach

Huijie Yang, Yizhong Zhuo and Xizhen Wu
Institute of Atomic Energy, PO Box 275(18), Beijing, 102413, People's Republic of China

Received 1 February 1994

Abstract. In this paper, by using non-equilibrium transport theory, the thermal denaturation of DNA molecules is investigated as a preliminary step to clarify the dynamical process of DNA transcription. The distribution functions of the displacement of base pairs at different temperatures are calculated. A modified model is also proposed which can reproduce the essential features of the denaturation process. Calculations show that the Langevin equation is an effective tool for describing the dynamical process of DNA molecules, and it seems to be more advantageous than the Nosé method.

1. Introduction

The discovery of the double-helix structure of the DNA molecule has established a strong relationship between its structure and function. However, there is a growing feeling that a static structure alone cannot determine entirely the biological function of a DNA molecule and that dynamical processes have an essential role [1]. The structure, which is so well designed to include the genetic codes in two complementary strands and protect it against external perturbations, would also prevent the expression of the genetic codes if the molecule were static. A typical example is DNA transcription, during which a segment of the genetic code is copied into single-stranded RNA. In order to expose the coding base pairs to chemical reactions, the double helix unwinds locally and forms a 'bubble' which is about 20 base pairs long and moves along the molecule as transcription proceeds. Therefore, it is extremely important to study the dynamical behaviour of the DNA molecule to understand its biological function. But the actual dynamical process of the DNA molecule is very complex because it is activated by an enzyme, and it remains beyond the reach of physical analyses. Experiments show that at low temperatures some base pairs open for a very short time and then close, known as 'fluctuational opening'. With increasing temperature, the number of base pairs broken increases more and more, and small denaturated regions similar to the transcription 'bubble' are formed. When the temperature is above a critical temperature, the DNA molecule splits into two strands, which is called 'thermal denaturation'. In a sense, thermal denaturation represents the dynamical process of the DNA molecule. Investigation of this process theoretically is a valid preliminary step for understanding DNA transcription. The problem has been extensively investigated experimentally, and an Ising-like model has also been suggested to explain the denaturation curves found in the experiments [2]. However, the Ising-like model cannot describe the dynamical process of the DNA molecule and it relies on many parameters which are difficult to derive from a first-principle calculation. Dauxos *et al* have studied this problem, using the constant-temperature molecular dynamics method (the Nosé method) to treat the dynamical process [3]. Both [4] and our work [5]

0305-4470/94/186147+10$19.50

indicate that the Nosé method can describe the dynamical behaviour of a physical system in contact with a thermal bath only under certain conditions, but it cannot describe the full dynamical process, especially the relaxation process, e.g. the process from non-equilibrium to equilibrium. In this paper, the Langevin equations are employed to describe the process of thermal denaturation of DNA molecules, and some preliminary results are presented.

2. Dynamical model of the DNA molecule and the theoretical method

2.1. A simplified model

To verify the effectiveness of the Langevin method and compare it with the Nosé method, we first consider the same simple model as proposed in [3], which is called the simplified model in this paper. The underlying features of this model are as follows:

(a) Asymmetry of the DNA molecule due to the difference in base pairs is neglected. Each strand of the DNA molecule is represented by a set of point masses which correspond to the nucleotides.

(b) The longitudinal displacements of point masses are not considered because their typical amplitudes are significantly smaller than those of the transverse ones. The displacement of the nth point mass relative to their equilibrium positions is denoted W_n for one strand and V_n for the other.

(c) Two neighbouring point masses in each chain are connected by a harmonic potential to keep the model as simple as possible, while the interaction potential within each base pair is assigned to be of Morse form. Because the base pairs can stretch with large amplitudes when the double helix opens locally, the nonlinearity of the internal potential cannot be ignored. The Hamiltonian of a DNA molecule with N base pairs can be written as

$$H = \sum_{n=1}^{N} \frac{m}{2}(\dot{V}_n^2 + \dot{W}_n^2) + \frac{K}{2}[(V_n - V_{n-1})^2 + (W_n - W_{n-1})^2] + D\{\exp[-\sqrt{2}a(V_n - W_n)] - 1\}^2 \tag{1}$$

where N is the number of base pairs for each DNA molecule, and n is the order number of base pairs. Introducing the variables

$$x_n = \frac{V_n + W_n}{\sqrt{2}} \qquad y_n = \frac{V_n - W_n}{\sqrt{2}}$$

the Hamiltonian can be separated into two parts:

$$H = H_r + H_c$$

where

$$H_r = \sum_{n=1}^{N} \left(\frac{1}{2}\dot{y}_n^2 + \frac{K}{2}(y_n - y_{n-1})^2 + D[\exp(-ay_n) - 1]^2 \right) \tag{2}$$

and

$$H_c = \sum_{n=1}^{N} \left(\frac{1}{2}m\dot{x}_n^2 + \frac{K}{2}(x_n - x_{n-1})^2 \right). \tag{3}$$

It is seen that there is no coupling between the variables x_n and y_n. Because we are only interested in the stretching motions of base pairs, we can ignore H_c and consider only the part of H_r which depends on y_n.

2.2. *Improvement upon the simplified model*

It is indicated in [3] that while the model mentioned above can qualitatively describe some aspects of the thermal denaturation of the DNA molecule, it cannot reproduce some essential features of the process, such as the phase transition, the sharp interval of temperature over which the phase transition occurs, etc. Experiments show that the phase transition occurs very abruptly over a temperature region which is only a few kelvins, or even less for short DNA molecules (base pairs < 600), while the interval obtained with the simplified model is about 100 K. The phase transition in the DNA molecule poses a fundamental problem: since the DNA molecule is basically a one-dimensional system, it is not expected to have a phase transition. The phase transition has been examined within the Ising-like model approach by Poland *et al* [6], who concluded that it can be attributed to cooperativity effects and to the role of the winding entropy released when the two strands separate. In fact, there are several aspects which remain to be considered. For example when one base pair stretches enough, or it breaks, the potential within the base pair should vanish. However, in the model presented above, the internal potential of hydrogen bonds is considered to tend to a constant while the coupling term between two neighbouring base pairs increases in a harmonic form. This model cannot properly take account of the cooperativity effects and cannot have a phase transition as discussed in [3]. In order to improve the model, an additional decreasing factor is proposed to be applied to the harmonic term in the present work, and the result is called the improved model below, the form of which now becomes

$$\Phi = \sum_{n=1}^{N} \left(\frac{K}{2} \exp[-\alpha(y_n + y_{n-1})](y_n - y_{n-1})^2 + D[\exp(-ay_n) - 1]^2 \right). \quad (4)$$

We find that the potential between the two strands of the DNA molecule now has a repulsive term when the stretching of one base pair exceeds a certain threshold, as is discussed in [7], and once the displacement of one base pair exceeds a threshold the potentials of neighbouring base pairs will decrease quickly with the stretching of the base pair; therefore, one can expect that the cooperativity effects would be produced by the motion of base pairs. The decreasing of the potentials gives the base pairs more freedom to move, causing an entropy increase which drives a sharp transition. It is noted that the prefactor of the quadratic term proposed here is somewhat different from the one discussed in [3]. In that paper, a simple extension of the simplified model is suggested as follows:

$$\Phi = \sum_{n=1}^{N} \frac{K}{2}\{1 + \rho \exp[-\alpha(y_n + y_{n-1})]\}(y_n - y_{n-1})^2 + D[\exp(-ay_n) - 1]^2 \quad (5)$$

which is based upon the fact that when the hydrogen bonds connecting the bases break, the electronic distribution on the bases is modified, causing the stacking interaction with adjacent bases to decrease. In equation (5) this effect is enforced by the prefactor of the usual quadratic term $(y_n - y_{n-1})^2$. This prefactor depends on the sum of the stretchings of the two interacting base pairs and decreases from $(K/2)(1 + \rho)$ to $K/2$, when either (or both) base pair stretches. This anharmonic model is considerably better than the simplified model. It can not only provide cooperativity effects which induce a 1D phase transition and a very sharp melting transition in good agreement with experimental results, but can also avoid a too high denaturation temperature with more realistic parameters being employed, as mentioned in [3]. But it seems that there is still a problem which should be studied further. When one base pair stretches enough or even is broken, not only does the interaction

between the two bases of the base pair vanish, but also the stacking effects between either one of the two bases and the nearest bases on the same chain should decrease. In particular, when the distance between two bases on the same chain becomes large, the stacking effects will eventually vanish. The effective potential should decrease and even become repulsive. This is why we adopt another form of the stacking interaction which is somewhat different from the anharmonic model suggested in [3]. Of course, it is too early to say definitely which model is more realistic until we have done more systematic studies.

2.3. Non-equilibrium transport equation

DNA molecules in a salt solution can be considered as a physical system in contact with a thermal bath. The motion of a base pair is acted upon by two forces, i.e. drift force due to the potential field and random force from the medium. This is a typical Brownian motion problem and can be more naturally described by the Langevin equation. Thus we suggest use of the Langevin equation instead of the Nosé method in our approach. Introducing a set of random forces

$$A(t) = (A_1(t), A_2(t), \ldots, A_n(t))$$

the equations of motion of a DNA molecule with N base pairs are as follows. Namely, for the simplified model

$$\begin{aligned} \dot{y}_n &= \frac{P_n}{m} \\ \dot{P}_n &= K(y_{n+1} + y_{n-1} - 2y_n) + 2aD[\exp(-ay_n) - 1]\exp(-ay_n) - \eta P_n + mA_n(t) \end{aligned} \tag{6}$$

and for the improved model

$$\begin{aligned} \dot{y}_n &= \frac{P_n}{m} \\ \dot{P}_n &= K\Big((y_{n+1} - y_n)\exp[-\alpha(y_n + y_{n+1})] - (y_n - y_{n-1})\exp[-\alpha(y_n + y_{n-1})] \\ &\quad + \frac{\alpha}{2}(y_n - y_{n-1})^2 \exp[-\alpha(y_n + y_{n-1})] \\ &\quad + \frac{\alpha}{2}(y_n - y_{n+1})^2 \exp[-\alpha(y_n + y_{n+1})]\Big) \\ &\quad + 2aD[\exp(-ay_n) - 1]\exp(-ay_n) - \eta P_n + mA_n(t) \end{aligned} \tag{7}$$

where $n = 1, 2, 3, \ldots, N$, and η is the friction coefficient. The random forces $A_n(t)$ satisfy the following Einstein relations:

$$A_n = \sqrt{\eta \frac{kT}{m}}\Gamma_n(t) \qquad \langle \Gamma_n(t) \rangle = 0 \qquad \langle \Gamma_n(t)\Gamma_{n'}(t') \rangle = 2\delta(t - t')\delta_{nn'}. \tag{8}$$

The Langevin equations can be solved by the Monte Carlo simulation method. Introducing a function,

$$K(y_n) = \frac{\partial K}{\partial y_n} \tag{9}$$

and taking the Taylor expansion to the first order, i.e.

$$K(y_n + \Delta y_n) = K(y_n) + \left.\frac{\partial K}{\partial y_n}\right|_{y_n} \Delta y_n \tag{10}$$

the iterated solution is

$$\begin{aligned} y_n(t+\Delta t) &= y_n(t) + V(t)\Delta t + \frac{1}{2}(K(y_n) - \eta V(t))(\Delta t)^2 + \frac{1}{6}V(t)\left.\frac{\partial K}{\partial y_n}\right|_{y_n}(\Delta t)^3 \\ &\quad + g\sqrt{2\Delta t}\Delta t\left(\frac{1}{2}\omega_1 + \frac{1}{2\sqrt{3}}\omega_2\right) \\ P_n(t+\Delta t) &= P_n(t) - \eta m(y_n(t+\Delta t) - y_n(t)) + \sqrt{mKT\eta}\sqrt{2\Delta t}\omega_1 \\ &\quad + K(y_n)\Delta t + \frac{1}{2}V(t)\left(\frac{\partial K}{\partial y_n}\right)(\Delta t)^2 \\ V_n(t) &= \frac{P_n(t)}{m} \end{aligned} \tag{11}$$

where ω_1, ω_2 are Gaussian random numbers.

Energy, mass and length are expressed in electronvolts, atomic mass units (amu) and angstroms, respectively. The corresponding time unit is 1 tu = 1.0214×10^{-14} s. In the literature [8], the parameter K, which denotes the strength of a harmonic oscillator consisting of the two neighbouring point masses on the same strand is chosen to be 0.06 eV Å^{-2}. This is consistent with the realistic speed for a transverse or torsional acoustic phonon: 100 cm s^{-1}. As for the depth D and width a of the Morse potential, they are rather diverse in the literature. Some experiments [8] show that the single-bond energy for each hydrogen bond in a base pair is at least 0.044 eV. In [9] the self-consistent phonon (SCP) method is employed to study the hydrogen bonds in the DNA molecule, and it has been shown that the single-bond energy for each hydrogen bond can reach 0.06 eV. Therefore, it should be noted that the value of D employed in [3] is questionable. In order to compare our Langevin equation approach with the Nosé method, we choose the same parameters as in [3] with the simplified model, i.e. $a = 4.45$, $k = 0.06$ eV Å^{-2}, $m = 300$ amu, $D = 0.04$ eV. In the improved model the parameters are chosen as follows: $a = 4.45$, $k = 0.06$ eV Å^{-2}, $m = 300$ amu, $D = 0.12$ eV. It is noted that we set $\alpha = 0.5$, which is due to the ansatz that a base pair is considered to be broken when $y_n > 1$ Å, and η is taken to be 0.03 tu^{-1} to make the physical system become stationary or quasi-stationary quickly.

3. Simulation of the denaturation process of the DNA molecule

A physical ensemble which contains 20 000 DNA molecules is considered. Each DNA molecule contains 95 base pairs. Langevin equations for one DNA molecule with initial values of displacements and velocities of the base pairs $(y_1, y_2, \ldots, y_{95}, V_1, V_2, \ldots, V_{95}) = (0., 0; \ldots, 0.)$ are simulated for 20 000 time steps, and the 20 000 sets of values for $(y_1, y_2, \ldots, y_{95}, V_1, V_2, \ldots, V_{95})$ generated at the different times are chosen to be the initial values of displacements and velocities for the molecules in the ensemble. An appropriate time step length has been found to be 0.5 tu, which makes values of $\langle\sum_{n=1}^{95} y_n^2\rangle$ and $\langle\sum_{n=1}^{95} V_n^2\rangle$ stable with the change of time step length in a certain interval around the above value [10]. At different temperatures the times needed for the system to become stationary or quasi-stationary are different, the range of which is about 100 000 to 200 000 time steps.

3.1. Distribution function

In order to examine the motion of base pairs at different temperatures, the distribution function of the displacements of base pairs are shown at temperatures of 200, 400 and 600 K, respectively (see figure 1). In each case the Langevin equations for the physical system with the simplified model are simulated for a sufficiently long time so that it can become stationary or quasi-stationary. Then we estimate the numbers of base pairs whose displacements are in the interval −0.01 to +0.01 around the values of −0.3, −0.2, . . . , 9.0 Å, respectively. From the curves, we found that at low temperatures (200 K, 400 K) the distribution function can be expressed in an exponential form, i.e.

$$N = N_0 * \exp(-bx)$$

and most of the base pairs collect in a small interval around $x = 0$, while at high temperatures the distribution function becomes Gaussian, i.e.

$$N = N_0 * \exp[-c(x - x_0)^2]$$

which is reasonable because at high temperatures the harmonic terms in the potential of the simplified model tend to be dominant. Therefore, the form of the distribution function has a close relation with the dynamical process of the system. When the displacement of a base pair becomes negative, e.g. the base pair is compressed, the potential of the base pair will increase rapidly, therefore there are only a few base pairs in the region with $y < 0$.

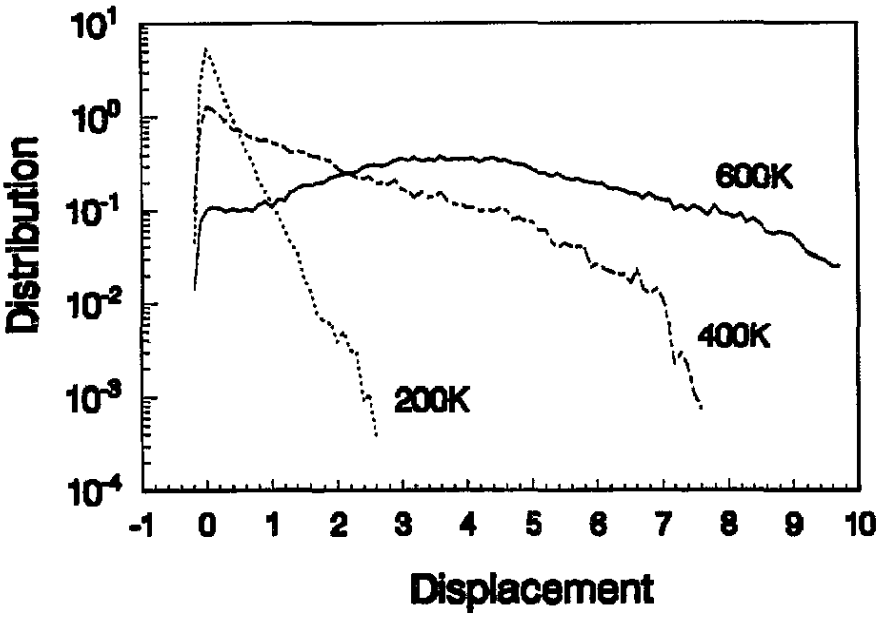

Figure 1. Distribution functions of the displacements of base pairs at different temperatures.

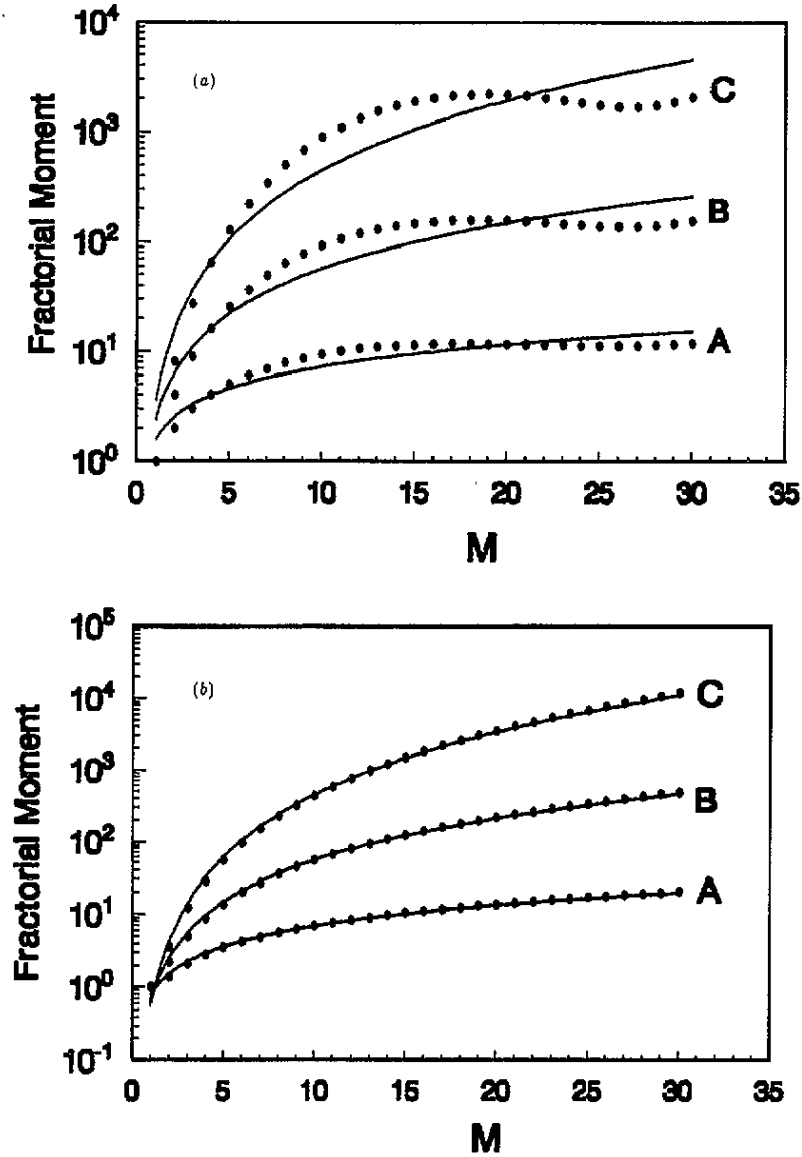

Figure 2. (*a*) Fractorial moments at 200 K obtained with the simplified model proposed in [3]. (*b*) Fractorial moments at 320 K obtained with the improved model proposed in this paper: A, the second fractorial moments; B, the third fractorial moments; C, the fourth fractorial moments. Full curves: regression curves of the second, third and fourth fractorial moments with geometric formulae.

3.2. Fractorial moments

Experiments show that thermal denaturation of DNA molecules is a phase transition process. To verify whether the simplified model and the improved model can describe the phase transition or not, the fractorial moments at different temperatures are calculated with both models and presented in figures 2(*a*) and (*b*). The formula for the fractorial moments employed in this paper is [11]

$$F_i = M^{i-1} \sum_{n=1}^{M} \left\langle \frac{K_m(K_m-1)\cdots(K_m-i+1)}{K(K-1)\cdots(K-i+1)} \right\rangle \tag{12}$$

where F_i is the ith fractorial moment, K the number of base pairs in the interval of displacement considered and M the number of subintervals of the whole interval being divided into. K_m is the number of base pairs in the mth bin and $\langle\cdots\rangle$ denotes the ensemble average over many events. It is indicated in [12] that the fractorial moments can dismiss the statistical fluctuations of a physical system, that is to say, for a system in which there is no dynamical fluctuations, F_i do not depend on the partition of the considered interval of displacement into M subintervals. If the phase transition occurs, the fractorial moments follow a poorer law,

$$F_i \propto \left(\frac{1}{M}\right)^{-\phi i}$$

due to the existence of strong dynamical fluctuations, which has been proved by many experiments [13]. The interval of displacement considered in our paper is -0.3 to 10 Å, which is divided into $1, 2, \ldots, 30$ subintervals, respectively, and the K_m are estimated, from which the fractorial moments are calculated. Figure 2(*a*) presents the results calculated with the simplified model at 200 K, and the results at other temperatures are similar to it. The curves obtained with the improved model at 320 K are shown in figure 2(*b*). We find that the simplified model cannot describe the phase transition in the thermal denaturation of the DNA molecule, while the improved model can reproduce fractorial moments, which can be almost exactly expressed with geometric formulae as follows:

$$\begin{aligned} \mathrm{Ln}\, F_2 &= 0.7507\, \mathrm{Ln}\, M \\ \mathrm{Ln}\, F_3 &= 1.2870\, \mathrm{Ln}\, M \\ \mathrm{Ln}\, F_4 &= 1.6737\, \mathrm{Ln}\, M. \end{aligned} \tag{13}$$

Therefore, there are strong dynamical fluctuations in the process, and a phase transition occurs.

3.3. Denaturation curves

In figure 3 the curves of N, the number of broken base pairs, versus temperature are presented, which are obtained by simulating the Langevin equations of the system for a sufficiently long time for the system to become stationary or quasi-stationary and estimating the number of base pairs broken at each temperature with each model. The ansatz that a base pair is considered to be broken when $y_n > 1$ Å is the same as that suggested in [3]. The results obtained with the simplified model are shown in figure 3. The results presented in [3] are also shown. It is found that the results given by the two methods are similar and can describe qualitatively some aspects of the experiments on thermal denaturation. But there are still some problems for the simplified model, as seen from the curves. One is that the melting temperature T_m, at which the denaturation rate is 0.5, is much larger than the actual one, which is in the range 320–350 K. Another is that the temperature interval over which thermal denaturation occurs is about 100 K, which is unreasonably large compared to the experimental value (several kelvins for a short chain of DNA molecule and about 10 K for a long one). The results obtained with the improved model are much better than with the simplified model. For example, the temperature interval in which the phase transition occurs

becomes about 10 K, which is close to the experimental value. Because the improved model is still very simple and the ansatz that a base pair is broken when $y_n > 1$ Å is subjective, we do not expect that an exact value of T_m can be predicted with the improved model, although the value of $T_m = 320$ K calculated with our model is quite close to experimental data. During the simulations we have found that at high or low temperatures it is difficult for the physical system to become stationary or quasi-stationary; therefore, the errors in the results are rather large when the temperatures are high or low.

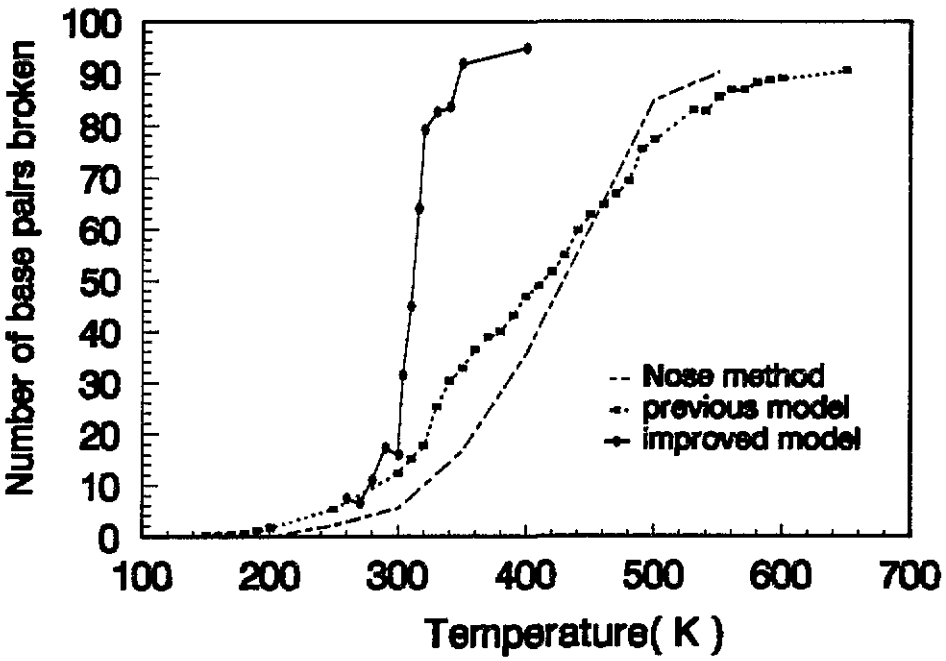

Figure 3. Denaturation curves.

4. Conclusion

Langevin equations have been employed to describe the thermal denaturation process of the DNA molecule, and it can not only reproduce the dynamical behaviour of a physical system in equilibrium but can also provide us with some important information on the relaxation process; therefore, it is an effective method for studying the dynamical behaviour of DNA molecules. As the Nosé method can also describe the dynamics of a system in equilibrium [4], we shall compare the two methods in detail in a separate paper. The improved model can describe the phase transition and give the narrow temperature interval in which denaturation occurs and some other essential dynamical properties. Therefore, the improvement is necessary and the model can be further employed to investigate the dynamical process of DNA molecules. Fractorial moments are used to study the dynamical fluctuations in thermal denaturation, and some typical results are obtained. As a concise model, the investigation of the thermal denaturation of DNA molecules can offer some useful results for our understanding of the essential problem of dynamical fluctuations.

The parameters of the models indicate the influences of the different physical factors on the dynamical process of DNA molecules, so how the parameters affect the results is an important problem to be further studied. The improved model and the Langevin equation can be easily extended to investigate the dynamical problems of triplex DNA molecules [14]. The improved model can describe the phase transition in the thermal denaturation process of DNA molecules, but how the phase transition really occurs is still a problem to be investigated.

Acknowledgments

One of the authors (Yizhong Zhuo) would like to acknowledge Professor C R Willis for the stimulating discussions during his visit to Boston University which led to the present investigations.

References

[1] Karplus M and Petsko G A 1990 *Nature* **347** 631
[2] Wartell R M and Benight A S 1985 *Phys. Rep.* **126** 67
[3] Peyard M and Bishop A 1989 *Phys. Rev. Lett.* **62** 2755
Dauxois T, Peyrard M and Bishop A R 1993 *Phys. Rev.* E **47** 1; 1993 *Phys. Rev.* E **47** 684
[4] Posch H A and Hoover W G 1986 *Phys. Rev.* A **33** 4253
[5] Yang H, Zhuo Y and Wu X 1994 Comparison between Langevin equations and Nosé method (in preparation)
[6] Poland D and Scheraga H 1966 *J. Chem. Phys.* **45** 1456
[7] Dauxois T, Peyard M and Willis C R 1992 *Physica* **57D** 267
[8] Freier S M, Sugimoto N, Sinclair A, Alkema D, Neilson T, Kierzek R, Caruthers M and Turner D 1986 *Biochem.* **25** 3214
Saenger W 1984 *Principles of Nucleic Acid Structure* (Berlin: Springer) p 117
[9] Prabhu V V, Young L and Prohofsky E W 1989 *Phys. Rev. B* **39** 5436
[10] Bao Jing-dong, Zhuo Yi-zhong and Wu Xi-zhen 1990 *Commun. Theor. Phys.* **14** 197
[11] YuanFang Wu, LianShou Liu 1992 *High Energy Phys. Nucl. Phys.* **16** 235 (in Chinese)
[12] Bialas A and Peschanski R 1986 *Nucl. Phys.* B **273** 703; 1988 *Nucl. Phys.* B **308** 857
[13] Copper F, Hwa R C and Sarcevic I (ed) 1990 *Intermittency in High Energy Collisions, Proc. Santa Fe Workshop* (Singapore: World Scientific)
[14] Bai Chunli, Zhang Pingcheng 1992 *J. Natural* **2** 6 (in Chinese)