Statistical Mechanical Theory of DNA Denaturation

SHIGEO HOMMA

School of Engineering, Gunma University, Kiryu 376-8515, Japan

Abstract. Thermodynamic quantities of a coupled sine-lattice chain, which is a simplified model of DNA (rotator model) are calculated for both in low and high temperature regions. In the high temperature region those quantities are expressed as a series expansion in terms of modified Bessel functions of an integer order. The results of numerical calculations are presented in connection with DNA denaturation (melting).

Key words: Denaturation, DNA, rotator model, sine-lattice

1. Introduction

It is widely accepted that the understanding of the mechanism of the transcription of the genetic information from DNA, that is, how it is initiated and transcribed, is one of the important research fields in life science. It is a complicated and interdisciplinary problem and thus needs a challenge of molecular biologists and physicists. From a view point of physics DNA could be regarded as an one dimensional classical system. It is a highly nonlinear system in its dynamical behaviours and random in its spatial configuration [1]. Thus it is expected that strong non-linearity results in nonlinear excitations such as solitary waves and randomness results in energy localization. As a subject of statistical mechanics a denaturation or melting of DNA is deeply related to the problem of a possibility of a phase transition in one-dimensional systems [2].

One fascinating aspect of a transcription is that an initiation is followed by a formation of a local opening of base pairs over 10 base pairs. For nonlinear dynamics this local opening might be interpreted as a solitary waves. And at a finite temperature this might be regarded as a formation of a cluster in a nucleation theory.

Several models were proposed and examined to study DNA dynamics and statistical mechanics [3]. Among them the following two models are to be noticed. One is a stretching model, where a configuration of DNA is described by the separation of two sugar-phosphate backbones [4–6]. The other is a rotator model, first proposed by Englander and others in 1980 [7]. In the rotator model each base is regarded as a rigid body and rotates or librates on the plane perpendicular to the helical axis. Thus the configuration of DNA could be described by an angle of rotation of each base. The rotator model (RM) was modified and improved by

Figure 1. Schematic feature of the B-form of DNA. The sugar-phosphate backbones of the double helix are presented as ribbons S and S' and the rung-like base pairs connect them as arrows.

many authors [8–14]. However previous works on RM were mainly concerned with nonlinear dynamics such as solitary waves, breather modes and so on. There is little work on statistical mechanics and dynamics at finite temperatures of RM [15–18].

It is the purpose of this paper to make detailed study of statistical mechanics of RM both at low- and high-temperature regions. This is done by using continuing approximation in low-temperature region and proposing a new method of calculating the partition function of RM in a form of a series expansion in high-temperature region. In the next section we define the Hamiltonian of RM of DNA. In Section 3 we calculate the partition function of RM in low-temperature region by using a continuum approximation. In Section 4 we propose a new method of a calculation of the partition function of RM in a form of a series expansion and perform a explicit calculation of it. The last section is devoted to results and discussions.

Figure 2. Projection of the n -th base (arrows) in the xy -plane.

2. A Rotator Model as a Model DNA System

We consider the B-form of DNA with its helical axis taken in z -direction. We show B-form in Figure 1, where each base is represented by arrow and complementary base pairs are indicated by conjugated arrows directed inward. The following fundamental assumptions are used to study dynamics and statistical mechanics of rotator model.

(1) The rotational degree of freedom is chosen out of many degrees of freedom which each base can take into its dynamical and statistical mechanical behaviours.

(2) Each base rotates on the plane perpendicular to the helical axis.

Let us define u_n and u'_n are the angles of rotation of base for n -th base pair from the B-form (see Figure 2), where circles represent planes on which each base rotates or librates. Under fundamental assumptions features of DNA could be described by a set of u_n and u'_n . We associate the interstrand base-base interaction $v_1(u_n, u'_n)$ of the n -th base pair to be the form

$$v_1 = -h_n \cos(u_n) - h'_n \cos(u'_n) - g_n \cos(u_n - u'_n). \quad (1)$$

Here g_n is a coupling constant. The terms proportional to h_n and h'_n correspond to a symmetry breaking potential. We next associate the intrastrand base-base interaction or stacking energy v_2 of bases as

$$v_2 = -J_n \cos(u_n - u_{n+1}) - J'_n \cos(u'_n - u'_{n+1}). \quad (2)$$

Here J_n and J'_n are coupling constants in each strand. The Hamiltonian H of the model rotator system of DNA is the sum of the kinetic and potential energy. Thus H is written as

$$H = K + V, \quad (3)$$

where K and V are given by

$$K = \frac{1}{2} \sum_{n=1}^N \left(\frac{p_n^2}{I_n} + \frac{p_n'^2}{I_n'} \right), \quad (4)$$

$$V = \sum_{n=1}^N (v_1 + v_2). \quad (5)$$

Here I_n and I_n' are the moment of inertia of the n -th bases and p_n and p_n' are momentum conjugate to u_n and u_n' . The total number of the base pairs is N . We use a periodic boundary condition. As a first step, we ignore the dependence of parameters such as h_n and others on n , that is, we treat the model rotator system as an uniform one.

3. The Partition Function in Low-Temperature Region

In the low temperature region the potential energy term V in Equation (3) could be well described by a quadratic form of u_n and u_n' . Thus the Hamiltonian H is approximated as, which is written by H_l

$$H_l = \frac{1}{2I} \sum_{n=1}^N (p_n^2 + p_n'^2) + V_q + E_0, \quad (6)$$

where

$$E_0 = -N(g + 2h + 2J), \quad (7)$$

$$\begin{aligned} V_q = \frac{1}{2} \sum_{n=1}^N [& h(u_n^2 + u_n'^2) + g(u_n - u_n')^2 \\ & + J((u_n - u_{n+1})^2 + (u_n' - u_{n+1}')^2)]. \end{aligned} \quad (8)$$

Introducing new variables x_n and y_n , defined by

$$x_n = \frac{1}{\sqrt{2}}(u_n + u_n'), \quad (9)$$

$$y_n = \frac{1}{\sqrt{2}}(u_n - u_n'). \quad (10)$$

We rewrite H_l as a sum of independent quadratic forms H_s and H_a . That is

$$H_l = E_0 + H_s + H_a, \quad (11)$$

where

$$H_s = \frac{1}{2I} \sum_{n=1}^N p_n^2 + \frac{1}{2} \sum_{n=1}^N (hx_n^2 + J(x_n - x_{n+1})^2), \quad (12)$$

$$H_a = \frac{1}{2I} \sum_{n=1}^N p_n'^2 + \frac{1}{2} \sum_{n=1}^N ((h + 2g)y_n^2 + J(y_n - y_{n+1})^2). \quad (13)$$

The partition function Z_l in the low temperature region is given by

$$\begin{aligned} Z_l &= \int (dp_n) \int (dp_n') \int (dx_n) \int (dy_n) [\exp(-\beta H_l)] \\ &= \exp(-\beta E_0) (2\pi I k T)^N \cdot Z_s \cdot Z_a, \end{aligned} \quad (14)$$

where $\beta = (k_B T)^{-1}$ and the quantities Z_s and Z_a are given by

$$Z_s = \int \exp \left[-\frac{\beta}{2} \sum_n (hx_n^2 + J(x_n - x_{n+1})^2) \right] (dx_n), \quad (15)$$

$$Z_a = \int \exp \left[-\frac{\beta}{2} \sum_n ((h + 2g)y_n^2 + J(y_n - y_{n+1})^2) \right] (dy_n). \quad (16)$$

The quantities Z_s and Z_a are calculated by using the Fourier transform of x_n and y_n . The results are

$$Z_s = \prod_{k>0} \left(\frac{2\pi k_B T}{h + 2J(1 - \cos(k))} \right), \quad (17)$$

$$Z_a = \prod_{k>0} \left(\frac{2\pi k_B T}{h + 2g + 2J(1 - \cos(k))} \right). \quad (18)$$

The free energy f_l per site is, in unit of $I = k_B = J = 1$,

$$\begin{aligned} f_l &= -\frac{T}{N} \log(Z_l) \\ &= -(g + 2h + 2) - 2T \log(2\pi T) \\ &\quad - \frac{T}{N} \sum_{k>0} \left[\log \left(\frac{(2\pi T)^2}{(h + 2(1 - \cos(k))) \cdot (h + 2g + (1 - \cos(k)))} \right) \right]. \end{aligned} \quad (19)$$

By changing a sum over k to an integral, we get f_e as

$$\begin{aligned}
f_l &= -(g + 2h + 2) - 2T \log(2\pi T) \\
&\quad + \frac{1}{2} \log \left(\left(\frac{1}{2}h + 1 + \sqrt{\left(\frac{1}{2}\right)^2 + h} \right) \cdot \right. \\
&\quad \left. \left(\frac{1}{2}h + g + 1 + \sqrt{\left(\frac{1}{2}h + g\right) \left(\frac{1}{2}h + g + 2\right)} \right) \right). \tag{20}
\end{aligned}$$

We define an order parameter σ as

$$\sigma = \langle \cos(u_n - u'_n) \rangle, \tag{21}$$

where $\langle A \rangle$ stands for an ensemble average. Then σ in the low temperature region, which is written as σ_l , is calculated as

$$\begin{aligned}
\sigma_l &= -\frac{\partial f_l}{\partial g} \\
&= 1 - \frac{T}{\sqrt{(h + 2g)(h + 2g + 4)}}. \tag{22}
\end{aligned}$$

An internal energy ϵ_l is obtained as

$$\begin{aligned}
\epsilon_l &= \frac{\partial(\beta f_l)}{\partial \beta} \\
&= -(g + 2h + 2) + 2T. \tag{23}
\end{aligned}$$

Here the second term represents an energy equipartition. We define another order parameter σ'_l , which is defined and calculated as

$$\begin{aligned}
\sigma'_l &= \frac{1}{2} \langle \cos(u_n) + \cos(u'_n) \rangle \\
&= -\frac{1}{2} \frac{\partial f_l}{\partial h} \\
&= 1 - \frac{T}{4} \left(\frac{1}{\sqrt{h(h + 4)}} + \frac{1}{\sqrt{(h + 2g)(h + 2g + 4)}} \right). \tag{24}
\end{aligned}$$

4. The Partition Function in the High Temperature Region

A contribution from a kinetic energy to the partition function Z_k is

$$Z_k = (2\pi T)^N. \tag{25}$$

A contribution from potential energy, which is written as Z_V , is defined by

$$Z_V = \int_{-\pi}^{\pi} du_1 \dots \int_{-\pi}^{\pi} du'_N \exp(-\beta V), \quad (26)$$

where V is given by Equation (5). In order to calculate Z_V we use the following identities:

$$\exp(H \cos(u_n)) = I_0(H) \sum_{p=-\infty}^{\infty} (I_p(H)/I_0(H)) \exp(ipu_n), \quad (27)$$

$$\exp(K \cos(u_n - u_{n+1})) = I_0(K) \sum_{k=-\infty}^{\infty} (I_k(K)/I_0(K)) \exp(ik(u_n - u_{n+1})), \quad (28)$$

$$\exp(L \cos(u_n - u'_n)) = I_0(L) \sum_{l=-\infty}^{\infty} (I_l(L)/I_0(L)) \exp(il(u_n - u'_n)), \quad (29)$$

where $H = \beta h$, $K = \beta J$ and $L = \beta g$. $I_p(H)$, $I_k(K)$ and $I_l(L)$ are modified Bessel functions of an integer order. Let us define renormalized modified Bessel functions of an integer order as

$$A(l) = I_l(L)/I_0(L) < 1, \quad |l| \geq 1, \quad (30)$$

$$B(k) = I_k(K)/I_0(K) < 1, \quad |k| \geq 1, \quad (31)$$

$$C(p) = I_p(H)/I_0(H) < 1, \quad |p| \geq 1. \quad (32)$$

They satisfy a symmetry relation: $A(l) = A(-l)$, $B(k) = B(-k)$ and $C(p) = C(-p)$.

By substituting Equations (27–32) into Equation (26) we rewrite Z_V as

$$\begin{aligned} Z_V &= (I_0(L)I_0^2(H)I_0^2(K))^N \\ &\sum_{(p)} \sum_{(p')} \sum_{(l)} \sum_{(k)} \sum_{(k')} \left[\prod_{j=1}^N A(l_j)C(p_j)C(p'_j)B(k_j)B(k'_j) \cdot \right. \\ &\int_{-\pi}^{\pi} du_j \exp(i(p_j + l_j + k_j - k_{j-1})u_j) \cdot \\ &\left. \int_{-\pi}^{\pi} du'_j \exp(i(p'_j - l_j + k'_j - k'_{j-1})u'_j) \right]. \end{aligned} \quad (33)$$

We perform the integration over u_j and u'_j to obtain Z_V as

$$Z_V = Z_{V0} \cdot Z'_V \quad (34)$$

where

$$Z_{V0} = ((2\pi)^2 I_0(L) I_0^2(H) I_0^2(K))^N. \quad (35)$$

This term corresponds to a molecular field approximation, for it does not depend on each lattice site. The quantity Z'_V describes contribution from fluctuations and discreteness effects of lattice, and is given by

$$Z'_V = \sum_{(l, \dots, k')} \left[\prod_{j=1}^N A(l_j) C(p_j) C(p'_j) B(k_j) B(k'_j) \cdot \delta(p_j + l_j + k_j - k_{j-1}) \cdot \delta(p'_j - l_j + k'_j - k'_{j-1}) \right]. \quad (36)$$

Here δ stands for Kronecker's delta. Because of the product of δ , the terms which make finite contributions to Z'_V need to satisfy the following constraints;

$$\sum_{j=1}^N (p_j + l_j) = 0, \quad (37)$$

and

$$\sum_{j=1}^N (p'_j - l_j) = 0. \quad (38)$$

From them we have

$$\sum_{j=1}^N (p_j + p'_j) = 0. \quad (39)$$

This is a global constraint. We approximate Equation (39) by a local constraint defined by

$$p_j + p'_j = 0. \quad (40)$$

We will give in the last section the meaning of the approximation given by Equation (40). From Equation (40), Equations (37) and (38) are reduced to the one

$$\sum_{j=1}^N (p_j + l_j) = 0. \quad (41)$$

We use Equation (41) to calculate Z'_V . We introduce n -site expansion in which values of $(p_j + l_j)$ are different from zero on arbitrary n lattice sites. Using it we perform a lattice sum systematically by increasing n from zero one by one under the constraint of Equation (41).

Remembering the fact that modified Bessel function $I_m(z)$ ($m \neq 0$) approach to zero, whereas $I_0(z)$ to 1 when z goes to zero. Therefore renormalized modified Bessel functions defined by Equations (30–32) decrease to zero by increasing temperature. Thus n -site expansion here corresponds to high temperature expansion in a theory of a phase transition. In the case $n = 0$, $(p_j + l_j)$ of every lattice site is zero. We, consequently, obtain $Z_V(n = 0)$ ($= Z_V(0)$) as

$$\begin{aligned} Z_V(0) &= \sum_{l_n} \left[\prod_{n=1}^N A(l_n) C^2(l_n) \right] \sum_{k,k'} [B(k) B(k')]^N \\ &= \eta^N (1 + 2 \sum_{k=1}^{\infty} (B(k))^N)^2, \end{aligned} \quad (42)$$

where the quantity η is defined by

$$\eta = \sum_{l=-\infty}^{\infty} A(l) C^2(l). \quad (43)$$

Remembering Equation (31), we have $Z_V(0)$, in the thermodynamic limit, as

$$Z_V(0) = \eta^N. \quad (44)$$

The term $Z_V(n = 1)$ does not contribute to Z'_V because of the constraint Equation (41). We proceed to 2-site expansion, in which on arbitrary two lattice sites, the i -th and j -th site ($i < j$), $(p_i + l_i)$ and $(p_j + l_j)$ are not zero. Taking into account Equation (41) we set $(p_i + l_i) = -(p_j + l_j) = g$ and the remaining $(p_m + l_m)$, ($m \neq i, j$) equal to zero. We substitute these relations into Equation (36), we have the $Z_V(n = 2)$ ($= Z_V(2)$) as

$$\begin{aligned} Z_V(2) &= \eta^{N-2} \sum_{g \neq 0} \left[\left(\sum_{l=-\infty}^{\infty} A(l) C^2(l + g) \right) \left(\sum_{l=-\infty}^{\infty} A(l) C^2(l - g) \right) \right] \\ &\quad \sum_k \sum_{k'} \sum_{i < j} (B(k) B(k'))^{N-(j-i)} (B(k - g) B(k' + g))^{j-i}. \end{aligned} \quad (45)$$

Performing the sum over i and j , we have terms which are proportional to

$$\frac{B(k - g) B(k' + g) (B(k) B(k'))^N - B(k) B(k') (B(k - g) B(k' + g))^N}{B(k) B(k') - B(k - g) B(k' + g)}. \quad (46)$$

In the thermodynamic limit the terms $k = k' = 0$ and $k = -k' = g$ make a finite contribution to $Z_V(2)$. We have to note that these two terms are topologically identical. Remembering that we have to sum over i after the sum over j , we need to include only one of these two terms in the sum over j . By this we are able to avoid the double counting of identical terms. We set $k = k' = 0$ in Equation (45) and obtain the result for $Z_V(2)$, in the thermodynamic limit as

$$Z_V(2) = \eta^N N \sum_{g \neq 0} D(-g) \psi(g) D(g). \quad (47)$$

Here the quantities $D(g)$ and $\psi(g)$ are defined as

$$D(g) = \eta^{-1} \sum_{l=-\infty}^{\infty} A(l) C^2(g-l) = D(-g). \quad (48)$$

$$\psi(g) = \frac{B^2(g)}{1 - B^2(g)}. \quad (49)$$

The interpretation of $Z_V(2)$ is following; a fluctuating field $D(g)$ creates a quasi-particle. It propagates through $\psi(g)$ and is absorbed by another fluctuating field $D(g)$ ($= D(-g)$). Thus $Z_V(2)$ is also expressed by the graph represented by

$$Z_V(2) = \eta^N N \sum_{g \neq 0} (\bullet \rightarrow \bullet), \quad (50)$$

where $\bullet$ and $\rightarrow$ stands for $D(g)$ (or $D(-g)$) and $\psi(g)$, respectively. Thus $D(g)$ plays both source and sink term, whereas $\psi(g)$ a propagator of quasi-particle. The term $Z_V(n \geq 3)$ are calculated similarly. For $n = 3$, we have

$$\begin{aligned} Z_V(3) &= \eta^N \sum_{l_1 \neq 0} \sum_{l_2 \neq 0} D(-l_2) \psi(l_2) D(l_2 - l_1) \psi(l_1) D(l_1) \\ &= \eta^N \sum_{l_1} \sum_{l_2} (\bullet \rightarrow \bullet \rightarrow \bullet). \end{aligned} \quad (51)$$

For $n = 4$, besides ‘connected graph’, which is expressed by

$$Z_V(4)_c = \sum_{(l_i)} (\bullet \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet). \quad (52)$$

there appears ‘disconnected graph’. After taking the lattice sum as well as thermodynamic limit, the contribution from ‘disconnected graph’ is calculated as

$$\begin{aligned} Z_V(4)_{dc} &= \eta^N \left[-\frac{1}{2} N \sum_{l_1 \neq 0} \sum_{l_2 \neq 0} D(l_1) \psi(l_1) D(l_1) \cdot D(l_2) \psi(l_2) D(l_2) \cdot \right. \\ &\quad \left. (1 + \phi(l_1) + \phi(l_2)) + \frac{N^2}{2} \left(\sum_l D(l) \psi(l) D(l) \right)^2 \right], \end{aligned} \quad (53)$$

which is expressed by a graph as

$$\begin{aligned}
 Z_V(4)_{dc} &= \eta^N \left[-\frac{N}{2} \sum_{(l_i)} ((\bullet \rightarrow \bullet \quad \bullet \rightarrow \bullet) + 2(\bullet \rightarrow \bullet \quad \bullet \rightarrow \bullet)) \right. \\
 &\quad \left. + \frac{N^2}{2} \left(\sum_l (\bullet \rightarrow \bullet) \right)^2 \right], \tag{54}
 \end{aligned}$$

where the quantity $\phi(l)$ is given by

$$\phi(l) = (1 - B^2(l))^{-1}, \tag{55}$$

and is represented by $\rightarrow$ and plays a role of ‘bare’ propagator. We note that in Equations (53) and (54) there appear terms which are proportional to N^2 . Further calculation of $Z_V(n \geq 5)$ could, in principle, be carried out in the line described above. The general procedures for calculating terms, which are proportional to N , associated with n -site expansion, is given in the appendix. The expression of free energy f_h per site in the high temperature region thus obtained as

$$\begin{aligned}
 \beta f_h &= -3 \log(2\pi) + \log \beta - \log(I_0(L)I_0^2(H)I_0^2(K)) - \log \eta \\
 &\quad - \sum_l (\bullet \rightarrow \bullet) - \sum_{(l_i)} (\bullet \rightarrow \bullet \rightarrow \bullet) \\
 &\quad - \sum_{(l_i)} (\bullet \rightarrow \bullet \rightarrow \bullet \rightarrow \bullet) + \frac{1}{2} \sum_{(l_i)} ((\bullet \rightarrow \bullet \quad \bullet \rightarrow \bullet) + \dots) \\
 &\quad + \dots \tag{56}
 \end{aligned}$$

Note that the terms, except those proportional to N , cancel each other exactly in the expression of $\log Z$. The order parameter σ_h in the high-temperature region is calculated as

$$\begin{aligned}
 \sigma_h &= -\frac{\partial f_h}{\partial g} \\
 &= D(1) + \frac{\partial}{\partial L} \left(\sum (\bullet \rightarrow \bullet) + \dots \right) \\
 &= D(1) + \sum_{(l_i)} ((\circ \rightarrow \bullet) + (\bullet \rightarrow \circ) + \dots), \tag{57}
 \end{aligned}$$

where a symbol $(\circ)$ stands for

$$(\circ) = \frac{\partial(\bullet)}{\partial L} = \frac{\partial D(g)}{\partial L} \propto D(g). \tag{58}$$

Similarly σ'_h is given by

$$\begin{aligned}
\sigma'_h &= -\frac{1}{2} \frac{\partial f_h}{\partial h} \\
&= R(0) + \frac{1}{2} \sum_{(l_i)} ((* \rightrightarrows \bullet) + (\bullet \rightrightarrows *) + \dots) + \dots, \tag{59}
\end{aligned}$$

where a symbol $(*)$ is defined by

$$(*) = \frac{\partial(\bullet)}{\partial H} = \frac{\partial D(g)}{\partial H} \propto D(g). \tag{60}$$

5. Numerical Results and Discussion

In this study we calculated the partition function of a rotator model (coupled sine-lattice model), which might well simulate dynamical as well as statistical mechanical behaviours of DNA macromolecules. In the low temperature region a continuum approximation is used to the model Hamiltonian, thus the thermodynamic quantities are easily calculated.

In the high temperature region, thermodynamic quantities of model Hamiltonian are expressed in a form of series expansion in terms of modified Bessel functions of an integer order. The procedures to associate mathematical expressions to each term which appear in series expansion in graph are presented. This might enable us to give the thermodynamic quantities in an arbitrary order.

In those calculations we used the approximation defined by Equation (40). Remembering the fact that the set (p_j, p'_j) , which appear in Equation (39), describes correlations of fluctuations in symmetry breaking fields h_n and h'_n (see Equations (1) and (32)), the procedure of an approximation from Equation (39) to Equation (40) means that correlations in a given base pair makes dominant contributions to Z'_V . The model rotator system we discuss here is one-dimensional, so that the path which arbitrarily connects two lattice sites is limited only on a line. Thus, except in a same base pair, it is supposed that correlations in a given set (p_j, p'_j) becomes less important in a calculation of Z'_V .

We calculated the order parameters defined by Equations (57) and (59), numerically up to $n = 9$, which are shown in Figures 3 and 4. The result thus obtained here could be compared to those obtained by Peyrard and his research group [4–6]. In a series of works, where they used Morse potential for inter- and intra-strand base-base interactions, they calculated nonlinear excitation modes and the thermodynamic quantities such as free energy and mean square distance of an interstrand base-base pair. In the calculation of the thermodynamic quantities they solved an integral equation both numerically and analytically by invoking the continuum approximation. From these calculations they found the abrupt increase of the mean square distance by increasing temperature, indicating a certain kind of phase transition there.

Figure 3. The temperature dependence of σ' in the high temperature region for $h = 0.1$. Numbers in the figure represent numerical value of g .

Figure 4. The temperature dependence of σ in the high temperature region for $g = 0.05$. Numbers in the figure represent numerical value of h .

From the results of the numerical calculation done in this paper, we could not find any trace of phase transition (see Figures 3 and 4). Rather, we see a gradual increase of relative angles of a base pair. This means that a formation of an open state will be a modest one, similar to a formation of a droplet or a cluster in a nucleation theory. In order to clarify a process of a formation of those open states, we have to study dynamical properties of our model DNA system. This will be the next problem.

Appendix

We briefly describe, in an itemized form, the method of calculating the terms in $Z_V(n)$ which are proportional to N , associated with n -site expansion.

- (1) Draw all the topologically distinct graphs composed of n -sites, whether connected or not.
- (2) To the connected graph, associate ψ and D , which are defined by Equations (49) and (48), to $\Rightarrow$ and $\bullet$, respectively.
- (3) To the disconnected graph;
 - (3-1) Associate ψ and D to the connected part of a given graph.
 - (3-2) To each disconnected part of a given graph associate the quantity defined by

$$\Pi = 1 + \sum_i \phi(l_i), \quad (61)$$

where $\sum_i$ means the sum over $\phi(l_i)$ ($= \phi_{l_i}$), which represents the connected parts of a given graph.

- (3-3) Make a product of a set of connected graph and $(\Pi)^t$, where t denotes a number of disconnected parts in a given graph. In $(\Pi)^t$ a product of ϕ_{l_i} should be counted only once. As an example we take the disconnected graph $(\bullet \Rightarrow \bullet \bullet \Rightarrow \bullet \bullet \Rightarrow \bullet)$ appearing for $n = 6$. Here Π is equal to $(1 + \phi_{l_1} + \phi_{l_2} + \phi_{l_3})$, and $t = 2$. Thus $(\Pi)^t$ should be

$$\begin{aligned} (\Pi)^2 = & 1 + \phi_{l_1} + \phi_{l_2} + \phi_{l_3} + \phi_{l_1}^2 + \phi_{l_2}^2 + \phi_{l_3}^2 \\ & + \phi_{l_1} \cdot \phi_{l_2} + \phi_{l_1} \cdot \phi_{l_3} + \phi_{l_2} \cdot \phi_{l_3}. \end{aligned} \quad (62)$$

- (3-4) Associate $(-1)^t$ to the result of (3-3).
- (3-5) If all the connected parts of a given graph are identical, divide the result of (3-4) by a number of connected parts. This gives the desired expression of the terms associated with a given disconnected graph.

The sum of (2) and (3-5) gives the terms of n -site expansion.

References

1. Peyrard, M. (ed.): *Nonlinear Excitation in Biomolecules*, Springer, 1994.
2. Huang, K.: *Statistical Mechanics*, J. Wiley and Sons, 1987.
3. Yakushevich, L.V.: Nonlinear DNA dynamics, *Physica* **D79** (1994), 77–86 and *Nonlinear Physics of DNA*, Wiley and Sons, 1998.
4. Peyrard, M. and Bishop, A.R.: Statistical mechanics of a nonlinear model for DNA denaturation, *Phys. Rev. Lett.* **62** (1989), 2755–2758.
5. Dauxois, T., Peyrard, M. and Bishop, A.R.: Entropy-driven DNA denaturation, *Phys. Rev.* **E47** (1993), R44–R47.
6. Dauxois, T. and Peyrard, M.: Entropy-driven transition in a one-dimensional system, *Phys. Rev.* **E51** (1995), 4027–4044.
7. Englander, S.W., Kallenbach, N.R., Heeger, A.J., Krumhansl, J.A. and Litwin, S.: Nature of the open state in long polynucleotide double helices, *Proc. Natl. Acad. Sci.* **77** (1980), 7222–7226.
8. Yomosa, S.: Soliton excitations in DNA double helices, *Phys. Rev.* **A27** (1983), 2120–2125 and **A30** (1984), 474–480.
9. Takeno, S. and Homma, S.: Topological solitons and modulated structure of bases in DNA, *Prog. Theor. Phys.* **70** (1983), 308–311.
10. Homma, S. and Takeno, S.: A coupled base-rotator model for structure and dynamics of DNA, *Prog. Theor. Phys.* **72** (1984), 679–693.
11. Zhang, C.T.: Soliton excitations in DNA double helices, *Phys. Rev.* **A35** (1987), 886–891.
12. Takeno, S. and Homma, S.: Kink and breathers associated with collective sugar puckering in DNA, *Prog. Theor. Phys.* **77** (1987), 548–562.
13. Yakushevich, L.V.: Nonlinear DNA dynamics; a new model, *Phys. Lett.* **A136** (1989), 413–417.
14. Takeno, S. and Homma, S.: Self-localized anharmonic rotational modes of bases in DNA, *J. Phys. Soc. Jpn.* **59** (1990), 1890–1901.
15. Homma, S.: Statistical mechanics of the one-dimensional plane rotator model in an external field, *Phys. Lett.* **A133** (1988), 275–278.
16. Homma, S.: Statistical mechanics of the one-dimensional plane rotator model in an external field, *J. Phys. Soc. Jpn.* **59** (1989), 1559–1569.
17. Homma, S.: Statistical mechanics of one-dimensional sine-lattice model in an external field, in A.R. Bishop, S. Jimenez and L. Vazquez (eds): *Fluctuation Phenomena, Disorder and Nonlinearity*, World Scientific (1995), 236–243.
18. Homma, S.: Thermodynamic properties of coupled sine-lattice, I – A method of calculation of a partition function, *Physica* **D113** (1998), 202–211.