

THEORY OF COIL-BALL TRANSITION IN LONG MOLECULES

A. A. VEDENOV and A. M. DYKHNE

Submitted February 21, 1968

Zh. Eksp. Teor. Fiz. 55, 357-362 (July, 1968)

We consider the problem of the intramolecular "melting" of a long molecule consisting of randomly arranged links of two kinds (DNA model). The thermodynamic functions of the molecule are calculated under the assumption that the temperature interval of the transition is narrow. It is shown that allowance for the randomness in the location of the links leads to a significant change in the picture of the "melting," particularly to a considerable broadening of the transition region.

1. Two-filament biopolymers (e.g., desoxyribonucleonic acid-DNA), which form an elongated helical structure at low temperatures, can go over upon heating into a state of a statistical ball. In the helical (unmelted) state, links of two parallel filaments are connected by weak transverse bonds; on going over to the ball (molten) state, these bonds are broken. This phenomenon of intramolecular "melting" is called the coil-ball transition.

For molecules with links of the same kind (homopolymer), the theory of such a transition was considered in various models<sup>[1]</sup>. Attempts at solving the problem of the coil-ball transition for a heteropolymer (a molecule consisting of links of different types randomly arranged along the filaments)<sup>[2-5]</sup> did not lead to success. This problem was considered numerically in<sup>[6]</sup>. In the present paper, within the framework of the universally assumed model (described below), we obtain an analytic solution of the problem.

2. We consider a chain made up of links of two kinds. Each link can be in two states—bound ( $\sigma = -1$ ) and broken ( $\sigma = +1$ ); the breaking energy  $2\bar{U}$  depends on the type of link and does not depend on the surroundings of the link.

The transition of an individual link from a bound state into a broken state would occur as a result of a change of  $\bar{U}$  ( $\bar{U}$  reverses sign at a certain temperature  $T_0$ , which differs for the two types of links). In Appendix I we present a physical justification of such a model. Strong interaction of neighboring links (which is assumed to depend on the state of the links and not on their kind) prevents these links from being situated in different states. Therefore the coil-ball transition takes place in a narrow temperature interval (compared with the melting temperature  $T_0$ ), and when the thermodynamic functions of the chain are calculated with the aid of the partition function  $Z = \sum e^{-E/T}$ , where

$$-E = \bar{V} \sum_k \sigma_k \sigma_{k+1} + \sum U_k \sigma_k \tag{1}$$

it is possible to take into account the temperature dependence only in the quantities  $\bar{U}_k$ . We introduce also the notation  $V = \bar{V}/T_0$  and  $U = \bar{U}/T_0$ .

3. Our problem has been reduced to a one-dimensional Ising model—a chain of interacting spins in an inhomogeneous magnetic field that depends randomly on the number of the link. We solve this thermodynamic problem by a method similar to the methods used in the

solution of dynamic problems in one-dimensional random systems.

We introduce the two-row matrices

$$\hat{A}^{(k)} = \begin{pmatrix} e^{V+U_k} & e^{-V+U_k} \\ e^{-V-U_k} & e^{V-U_k} \end{pmatrix},$$

$$\hat{Z}^{(k)} = \prod_{l=1}^k \hat{A}^{(l)}.$$

The partition function is

$$Z = \hat{Z}_{11}^{(N)} + \hat{Z}_{22}^{(N)} \approx \hat{Z}_{11}^{(N)}$$

( $N$ —number of links in the chain). From the equation

$$\hat{Z}^{(k+1)} = \hat{Z}^{(k)} \hat{A}^{(k+1)}$$

it is easy to obtain the recurrence relations

$$\zeta_{k+1} = \lambda_k + \mu_k / \zeta_k, \tag{2}$$

where

$$\zeta_k = e^{-U_k} Z_{11}^{(k+1)} / Z_{11}^{(k)},$$

$$\lambda_k = e^V (1 + e^{-2U_k}), \quad \mu_k = e^{-2U_k} (e^{-2V} - e^{+2V}). \tag{3}$$

For the logarithm of the partition function we get

$$f \equiv \frac{\ln Z}{N} = \frac{1}{N} \sum_k (U_k + \ln \zeta_k) =$$

$$= \langle U \rangle + \int F(\zeta) \ln \zeta d\zeta, \tag{4}$$

where  $F(\zeta) = N^{-1} \sum_k \delta(\zeta - \zeta_k)$  is the normalized distribution function of the quantity  $\zeta$ , and satisfies the equation

$$F(\zeta) = - \int F\left(\frac{\mu}{\zeta - \lambda}\right) \varphi(U) dU \frac{\mu}{(\zeta - \lambda)^2} \tag{5}$$

(for a derivation see Appendix II); here  $\varphi(U)$ —fraction of the links with binding energy  $2U$ . Equation (5) and expression (4) solve the problem in principle.

We consider below the most interesting case

$$\varphi(U) = c_1 \delta(U - U_1) + c_2 \delta(U - U_2),$$

where  $c_{1,2}$ —fraction of links with energy  $U_{1,2} \equiv U \pm W$ . If we introduce the variables

$$\tau_k = U - V + \ln \zeta_k - 1/2 \ln(1 - \Delta^2), \quad \Delta = e^{-2V},$$

then the relations (2) and (4) take the form

$$\exp\{\tau_{k-1} \pm W\} = \frac{2ch(U \pm W)}{(1 - \Delta^2)^{1/2}} - \exp\{-\tau_k \pm W\}, \tag{6}$$

$$\tilde{f} = V + \frac{1}{2} \ln(1 - \Delta^2) + \int \tau \tilde{\Phi}(\tau) d\tau, \quad (7)$$

where  $\tilde{f} = f - U_0$ ,  $U_0 = (1 - 2c)W$ ,  $\tilde{\Phi}(\tau)$ —distribution function for the variables  $\tau$ .

We note that the substitutions  $U \rightarrow -U$ ,  $W \rightarrow -W$ , and  $\tau \rightarrow -\tau$  do not change relations (6) and (7). It follows therefore that when  $c_{1,2} = 1/2$  the function  $f(U)$  is even.

When  $W = 0$  (homopolymer),  $f = V + \ln(\cosh U + \sqrt{\sinh^2 U + \Delta^2})$ , and the excess of the fraction of the molten links over the unmolten ones

$$v(U) \equiv \partial f / \partial U \quad (8)$$

equals

$$v(U) = \text{sh } U / \sqrt{\text{sh}^2 U + \Delta^2}.$$

The characteristic dimension of the  $v(U)$  curve is  $\delta U \sim \Delta$ ; it follows from experiment that  $\Delta \ll 1$  ( $\Delta \sim 10^{-2}$ ).

4. As will be shown below, near the melting point at  $\Delta^2 \ll 1$  the significant values of  $\tau$  are those for which

$$\Delta^2 \ll \tau \ll U \ll 1. \quad (9)$$

Under these conditions, the recurrence relation (6) can be written approximately in the form:

$$y_{k+1} = y_k - 2(U \pm W), \quad (10)$$

where

$$\text{formula} \quad y_k = \ln(\tau_k - U).$$

In terms of the variables  $y$  we have

$$\tilde{f} = V + U + \int_{\ln \Delta^2}^0 e^y \Phi(y) dy. \quad (11)$$

In our problem, all the calculations that follow are carried out accurate to the small parameter  $1/\ln \Delta^2$ , and therefore the limits in (11) are determined accurate to a term on the order of unity.

Inasmuch as the interval of variation of  $y$  ( $\approx \ln \Delta^2$ ) is large compared with unity, the random-walk problem, described by relation (10), reduces to a differential (Fokker-Planck) equation for the distribution function

$$\begin{aligned} \alpha \frac{d\Phi}{dy} + \beta \frac{d^2\Phi}{dy^2} &= 0, \\ \alpha &= c(U+W) + (1-c)(U-W), \\ \beta &= c(U+W)^2 + (1-c)(U-W)^2. \end{aligned} \quad (12)$$

It is physically obvious that the random-walk problem (12) must be solved under the condition that the total flux vanish:

$$\alpha\Phi + \beta d\Phi/dy = 0$$

(this follows more rigorously from the exact recurrence relations (6)).

At the melting point  $U = U_0$  we have  $\alpha = 0$  and  $\beta = 4W^2c(1-c)$ . The solution of (12) is of the form  $\Phi = Ae^{K y}$ , where in the vicinity of the melting point

$$\kappa = -\frac{U - U_0}{4W^2c(1-c)}.$$

With the aid of the obtained distribution functions we get from (11), at the assumed accuracy,

$$\tilde{f} = V + U + A(\kappa) \int_{\ln \Delta^2}^0 e^{\nu y} dy, \quad (13)$$

where  $A(\kappa)$ —function with characteristic dimension on the order of unity in  $\kappa$ . With the same accuracy we have

$$\tilde{f} = V + U + A(0)\kappa / (1 - \Delta^{2\kappa}).$$

To determine the constant  $A(0)$  we use the condition that the excess of the fraction of the molten links  $\nu(U)$ , determined with the aid of (8), tends to  $\pm 1$  as  $U \rightarrow \pm \infty$ . It turns out that

$$A(0) = 8W^2c(1-c).$$

5. We obtain ultimately

$$\begin{aligned} f &= V + (U - U_0) \text{cth} \frac{U - U_0}{\delta}, \\ v(U) &= \frac{\text{sh } 2x - 2x}{2\text{sh}^2 x}, \quad x = \frac{U - U_0}{\delta}, \end{aligned}$$

where  $\delta$ —width of the transition, equal to

$$\delta = 4W^2c(1-c) \left/ \ln \frac{1}{\Delta} = \frac{2W^2}{V} c(1-c) \right.$$

The quantity  $\delta$  has a maximum  $\delta = W^2/2V$  at  $c = 1/2$ .

Inasmuch as the obtained width of the coil-ball transition is small compared with the temperature of the transition,  $U - U_0$  can be regarded as proportional to  $T - T_0$ :

$$\begin{aligned} U - U_0 &= (T - T_0) \frac{q_+ + q_-}{2T_0^2}, \\ T_0 &= \frac{cq_+ + (1-c)q_-}{q_+/T_+ + q_-/T_-}, \end{aligned}$$

where  $q_{\pm}$  are the melting heats and  $T_{\pm}$  the melting temperatures of the homopolymers from the links of the first and second kind.

Transforming from dimensionless to ordinary variables, we obtain for the width of the transition

$$\tau = 12 \frac{T_0^2}{q_+ + q_-} \frac{W^2}{V} c(1-c),$$

where

$$2W = q_+ \left( \frac{1}{T_0} - \frac{1}{T_+} \right) - q_- \left( \frac{1}{T_0} - \frac{1}{T_-} \right).$$

Knowledge of the value of  $f$  makes it possible to obtain the thermodynamic mean number of molten sections  $n_L$

$$\begin{aligned} \frac{n_L}{N} &= -\frac{1}{4} \frac{\partial}{\partial V} (f - V) = \\ &= \left[ (U - U_0) \left/ \text{sh} \frac{U - U_0}{\delta} \right. \right]^2 \frac{1}{8W^2c(1-c)} \end{aligned}$$

The average length ( $l$ ) of one molten section at the melting point is

$$l = \frac{V^2}{W^2c(1-c)}.$$

It is of interest to compare the results with the corresponding results for homopolymers. Using the same notation, we have for homopolymers

$$\delta \approx e^{-2V}, \quad n_L/N \sim e^{-2V}.$$

These values are exponentially small compared with the corresponding values for heteropolymers. The reason for the difference lies in the different character of the melting in the two cases: in the heteropolymer, by vir-

tue of the random distribution of the links in the chain, sections with high and low melting points are encountered, and the coil-ball transition consists of successive melting of different sections. This leads to a broadening of the region of the transition into an increase in the number of boundaries between phases (molten and un-molten sections) compared with the homopolymer, where, by virtue of the homogeneity, the transition takes place in a narrow temperature interval.

#### APPENDIX I

We present a more detailed justification of the assumed model. We assume that each link can be in a bound state ( $\sigma = -1$ ) or in a broken state ( $\sigma = +1$ ). The state  $\sigma = +1$  represents in fact a group of states characterized by a set of variables, which we denote by  $\tau$ . We assume that the interaction energy ( $V\sigma\sigma'$ ) of the neighboring links does not depend on the variables  $\tau$ , i.e., depends only on whether these links are in the bound state or not. All that depends on  $\tau$  is the energy of the individual link. This makes it possible to sum over the variables  $\tau$  at fixed  $\sigma$ . It is assumed here that the number of states of a given link does not depend on the states of the neighboring links. Such a summation, under the assumptions described above, leads, accurate to terms independent of  $\sigma$ , to the Hamiltonian (1), where  $2U_k T$  is the difference of the free energies of the  $k$ -th link in the states  $\sigma = \pm 1$ .

#### APPENDIX II

We start from the definition

$$F(\zeta) = \frac{1}{N} \sum_{k=1}^N \delta(\zeta - \zeta_k).$$

Using the recurrence relations (2) we obtain, neglecting the end effects,

$$F(\zeta) = \frac{1}{N} \sum_{k=1}^N \delta\left(\zeta - \lambda_k - \frac{\mu_k}{\zeta_k}\right)$$

$$= -\frac{1}{N} \sum_{k=1}^N \frac{\mu_k}{(\zeta - \lambda_k)^2} \delta\left(\zeta_k - \frac{\mu_k}{\zeta - \lambda_k}\right). \quad (\text{A.1})$$

As seen from (3),  $\lambda_k$  and  $\mu_k$  are functions of  $U_k$ .

The summation in (A.1) will be carried out in two stages: we first sum over the aggregate of the nodes  $\{U\}$ , in which  $U_k$  assumes a given value  $U$  (the number of such nodes is  $N\varphi(U)$ ), and then over different values of  $U$ :

$$F(\zeta) = -\frac{1}{N} \sum_U \frac{\mu(U)}{(\zeta - \lambda(U))^2} \sum_{(U)} \delta\left(\zeta_k - \frac{\mu(U)}{\zeta - \lambda(U)}\right) \quad (\text{A.2})$$

Owing to the random disposition of the nodes with different values of  $U$ , the following obvious relation holds for any sufficiently large sequence  $l$  of nodes:

$$\sum_{(l)} \delta(\zeta_l - \zeta) = \frac{l}{N} F(\zeta). \quad (\text{A.3})$$

From (A.2) and (A.3) we obtain Eq. (5).

<sup>1</sup>T. M. Birshteĭn and O. B. Ptitsyn, *Konformatsii makromolekul (Conformations of Macromolecules)*, Nauka, 1964.

<sup>2</sup>P. J. Flory and W. G. Miller, *J. Mol. Biol.* **15**, 284 (1966).

<sup>3</sup>H. Reiss et al., *J. Chem. Phys.* **44**, 4567 (1966).

<sup>4</sup>E. W. Montroll and N. S. Goel, *Biopolymers* **4**, 855 (1966).

<sup>5</sup>D. M. Crothers and N. R. Kallenbach, *J. Chem. Phys.* **45**, 917 (1966).

<sup>6</sup>A. A. Vedenov, A. M. Dykhne, A. D. Frank-Kamenetskiĭ, and M. D. Frank-Kamenetskiĭ, *Mol. biol.* **1**, 313 (1967).

Translated by J. G. Adashko