PHYSICAL REVIEW A VOLUME 20, NUMBER 4 OCTOBER 1979

Phase transitions in DNA

M. Ya. Azbel*

Department of Physics and Astronomy, Tel Aviv University, Tel Aviv, Israel

(Received 22 November 1977; revised manuscript received 29 January 1979)

Experiments on light absorption in DNA solutions in the region of 260 nm demonstrate specific multiple sharp oscillations of dA/dT vs temperature T, A being the optical density. A one-dimensional Ising model with a long-range polyspin interaction in an inhomogeneous "magnetic field" is used to present a quantitative description of these oscillations. An explicit analytical formula, derived in the paper, provides a surprisingly good agreement between the theory and experiments. It also alows one to obtain from these experiments important information about the sequence of the "magnetic fields," which represents the DNA component sequence. This is demonstrated by an example of a particular DNA. The long-range interaction in the Ising model implies a phase transition. Its nature is shown to depend crucially on well-defined properties of the DNA sequence.

I. PHYSICAL NATURE OF DNA MELTING

From the physical viewpoint DNA is a unique physical substance. It is a ready-made macroscopic one-dimensional system. The total length of a single mammal DNA is about 1.8 m; it contains about five billion sites.[1] DNA consists of four types of nucleotide molecules (adenine, thymine, cytosin, and guanine), which form two types of complementary base pairs [adenine-thymine (AT) and guanine-cytosin (GC)]. The sequence of these pairs is specific for each living being, as it represents DNA genetic information. Obviously, this sequence is neither random nor ordered in any sense, nor can it be described by any correlation relations. The precise description of the DNA sequence can be provided only by the explicit indication of all its "components" one by one; this is actually done when the DNA sequence is determined, as for bacteriophages MS-2,[2] ϕX-174,[3,4] FD,[5,6] and virus SV-40.[7] Of course, such a situation is very unusual in physics.

DNA consists of two strands bound by a hydrogen binding ("helix" state) with an energy[8] of about 3000° (i.e., about 6 kcal/mol). When the temperature of a solvent in which DNA is dissolved increases, these strands may unbind; this process is denoted as "coiling", or "melting", of DNA (see Fig. 1). The light absorbtion of bound and unbounded sites is different for different base pairs in the wavelength range 250–290-nm. This allows[1,9-11] experimental determination, of, e.g., the number N_c of melted sites and its derivative dN_c/dT (T is temperature) directly from the optical density measurements. Characteristic experimental plots for dN_c/dT ["differential melting curves," (DMC)] are presented in Fig. 2. These plots clearly demonstrate the following typical features of DNA melting, which should be explained: (i) The DNA melting temperature is about 350–400 °K, i.e., essentially less than the binding energy (~3000 °K) of DNA strands. (ii) DNA DMC's (i.e., dN_c/dT vs T) exhibit multiple oscillations. (iii) Each of the oscillations is extremely narrow, its relative half-width (i.e., the half-width related to the absolute temperature) is on the order of 10^{-3} thus resembling a singularity in a phase transition. The plots also give rise to the following questions: (a) How does the transition to completely separated DNA strands occur? Can it be a kind of a phase transition? (b) DMC's are very specific and, as is clearly demonstrated by Fig. 2, very different for different DNA sequences. How does the shape of the curves depend on the DNA component sequence? (c) DMC's depend both qualitatively and quantitatively on the DNA sequences which determine the number position, and shape of the DMC oscillations. Can we learn something from the melting curves about the DNA sequence? Can the corresponding information be relied on despite inevitable experimental errors?

To answer all these questions we discuss DNA melting in more detail (see Fig. 1). Nucleotide molecules, forming DNA base pairs, are huge organic molecules. When they are separated a large number of degrees of freedom is released, thus providing[1,9,10] an entropy per site $s \sim 10$ (s is expressed in units of the Boltzmann constant).

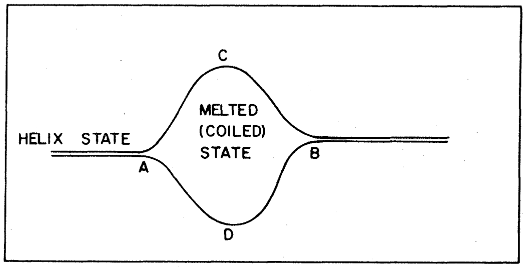

FIG. 1. DNA melting.

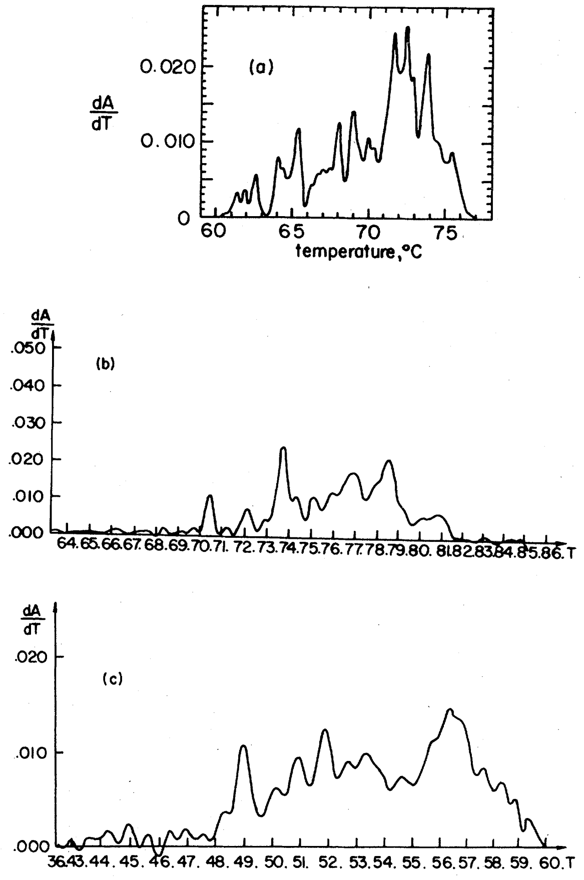

FIG. 2. Examples of differential melting curves (dA/dT vs T, A is the light absorbance at 260 nm, T is temperature) for different DNA's: (a) λDNA in the standard buffer, $[Na^+] = 0.0120$, according to Blake and Lefoley (Ref. 11); (b) Physarum Polycephalum DNA in SSC/10; and (c) Adenovirus II DNA in SSC/100, according to Reiss and Arpa-Gabarro (Ref. 20).

Thus the entropy contribution ($-Ts \sim -10T$) to the free energy per site compensates the loss of the binding energy ($\epsilon_e \sim 3000$ °K) when $T \sim 300$ °K, i.e., at room temperature. Therefore the low temperature of DNA melting is related to the organic nature of its components.

In the immediate vicinity of the melting temperature, $\epsilon_e \approx -Ts$ and the free energy per site is almost the same in the bound and unbound states. But a base pair at the boundary between a helix and a melted region (A or B in Fig. 1) is neither completely bound nor entirely separated, so a certain decrease in the binding energy is only partially compensated by an incomplete entropy release. Thus each phase boundary provides[1,9,10] a large (compared to temperature) boundary energy J of order of $|\epsilon_e| \sim 3000$ °K.

Different base pairs, AT and GC, have slightly different melting temperatures $T^{(1)}$ and $T^{(2)}$, respectively; $T^{(2)} - T^{(1)} \sim 40$ °K. When the temperature T is above $T^{(1)}$, but below $T^{(2)}$, the "fusible" AT wants to melt while the "refractory" GC wants to remain bounded. But a large phase-boundary energy prevents their independent behavior. As a result, only those DNA portions which are sufficiently rich with AT and long enough to compensate for boundary energies will melt. So, for instance, long AT protions will melt first and long GC portions will melt last.

Low temperature ($T \ll J$) implies little fluctuation. Therefore, when it becomes energetically preferable, certain DNA domains melt as a whole in a very narrow (determined by the fluctuations) temperature region. Thus the number N_c of melted DNA sites as a function of temperature T changes almost by jumps, related to the melting of these domains. The quasijumps imply[12] narrow peaks in dN_c/dT, observed in numerous experiments[13-22,11] (see, e.g., Fig. 2).

The higher the temperature is, the more sites are melted, and the longer the melted domains are. The entropy of melted domains, bordering on helix ones, is reduced by the condition which quarantees the meeting of strands, at B in Fig. 1, separated at A. If we neglect the elasticity of the strands and their self-avoidance, then $ACBDA$ in Fig. 1 is a random walk and the probability of such a closed loop is proportional to $L^{-3/2}$,[23] L being the length of ACB. The corresponding entropy decrease ("loop entropy") $\ln L^{-3/2} = -1.5 \ln L$ provides the long-range contribution $1.5T \ln L$ to the free energy and the phase transition[24-26] from helix to melted state. The phase transition in a one-dimensional system[27] (which is obviously affected by the elasticity of the strands and their self-avoidance—see below) is of special physical interest. Also, this transition allows us to observe the dependance of the nature of the transition on the component sequence; later we shall prove that the transition may vary from the first-order one to the essential singularity, when no thermodynamic quantities are discontinuous, but at a certain finite temperature DNA becomes completely melted.

The Landau-Lifshitz impossibility[28] of the phase transition in one-dimensional systems is related to the fact that each excitation creates the phase boundary and divides the system into two separate parts. Thus the phase transition can be possible only in the case of a complete ban on any excitations above the transition temperature (in the infinite system, of course). Therefore, contrary to usual phase transitions, such a transition is related to the change in the temperature-dependent ground state rather than to fluctuations whose relative contribution may tend to zero when the temperature approaches the phase-transition point. In this sense the high-temperature phase is completely ordered, as it contains no fluctuations. (We do not account in this way for the component degrees of freedom released while melting.)

To describe the shape of the melting curve and to determine the nature of the transition one must construct the thermodynamics for the corresponding Hamiltonian. This Hamiltonian[1,9,10] is related to the two possible states of each site (bounded and unbounded), which can be formally described by the spin "up" ($S=+\frac{1}{2}$ for an unbounded state) or "down" ($S=-\frac{1}{2}$ for a bounded state). The Hamiltonian accounts for the difference in the energies of bounded and unbounded states, which depends on the temperature, is of different signs for different components in the inverval $T^{(1)}<T<T^{(2)}$, and can be described by the "effective magnetic field", which is different for different components and is therefore related to the component sequence. Of course, the dependence of the effective Hamiltonian on the temperature is related to its being in fact the free energy for a given set of "spins" (i.e., of bounded and unbounded states); the summation over all other degrees of freedom is supposed to be already performed.

In this model the phase boundary is the boundary between opposite spins, so a large phase-boundary energy is identical to a large (compared to the temperature) exchange interaction in a ferromagnet. The "loop entropy" provides an additional long-range interaction in the Hamiltonian. Thus DNA melting is formally equivalent to a one-dimensional Ising ferromagnet at low temperatures in an inhomogeneous external magnetic field (related to the component sequence) with long-range interaction. As we already mentioned, the DNA component sequence cannot be described in any analytic way. So the question arises: Is it possible to derive the analytical formulas for, e.g., the free energy of a sequence which itself cannot be described analytically?

Though the answer seems to be certainly "no", the answer is "yes" if one somewhat refines a common physical approach. Usually one starts with a certain model of a physical system, this model being *precisely* known (e.g., a component sequence is considered to be random), and then derives an *approximate* formula (for, e.g., the free energy). In the case of DNA there exists no analytical description of the sequence so the exact formula for the DNA free energy F can relate F only to the precise sequence (and, e.g., just to write down the mammal DNA sequence of 5×10^9 sites one would need several million pages). But as we are always interested in the approximate F, we should determine the *approximate* analytical description of the sequence, which allows us to obtain F within the given accuracy. This description will essentially depend on the accuracy and can be found out only *simultaneously* and *self-consistently* with F. In this paper I develop such an approach, which was first proposed in Refs. 12, 26, 29–31, and 65, and which related F to certain DNA-sequence distribution functions depending on very specific and unusual variables. The theoretical formulas provide remarkably accurate agreement with the existing experiments.

The very existence of the relation between F and the sequence indicates the possibility of the inverse-problem solution, i.e., of the determination of certain DNA-sequence characteristics from the melting curves. However, in thermodynamics the inverse problem is usually unstable, i.e., infinitesimal experimental errors provide a significant change in the obtained solution.[32] The solution in our case is also unstable in this sense. However, the relative error in the solution, though finite, is small together with $T/J\sim0.1$ and may be as small as $(T/J)^3\sim10^{-3}$. Practically this allows for a very accurate solution, which may even be precise for a finite digital system. (For instance, the accuracy of 0.1% in concentration is enough to determine accurately the numbers of two components in a domain containing 400 sites.)

At present there exist just two cases which allow us to verify the theory. Only for the phages ϕX-174 and FD are both sequences[5,6] and melting curves[21,22,18(a)] known. In these cases both the computation of the melting curves for known sequences according to the formulas of this paper and the inverse problem solution, i.e., the determination of certain sequence characteristics from melting curves, demonstrate the coincidence with experimental data within experimental accuracy. This may be the first case of a quantitative agreement between theory and experiment for a natural biological system.

The theory of DNA melting is in no case simple or trivial despite its one-dimensional Ising character. This may be seen, for example from Eqs. (51)–(52) in Sec. VII, which describe the melting of a random sequence.

II. DNA HAMILTONIAN AND GROUND STATE

Suppose a DNA component sequence is $\{j_r\}\equiv j_1j_2\cdots$, where a given $j_r=1,2$ indicates the component (the first or the second) at the rth site. The state of this sequence is described by the set of spins $\{S_r\}\equiv S_1S_2S_3\cdots$, where a spin at the rth site $S_r=+\frac{1}{2}$, $-\frac{1}{2}$ describes the state of the site (unbounded or bounded, correspondingly).

The energy H of such a sequence

equals[1,9,10,26,30,31,33]

$$H=-\sum_r h_r S_r - J\sum_r S_r S_{r+1} + \sum_{r,L}(\tfrac{1}{2}-S_r)(\tfrac{1}{2}+S_{r+1})(\tfrac{1}{2}+S_{r+2})\cdots \times(\tfrac{1}{2}+S_{r+L})(\tfrac{1}{2}-S_{r+L+1})bT\ln(L\chi^2)\,, \qquad (1)$$

$$h_r \equiv h^{(j_r)} = s(T-T^{(j_r)}) \qquad (2)$$

(energy is measured here in degrees).

The "effective local magnetic field" h_r is the energy difference between the bounded and unbounded states of rth site; the "exchange" interaction J is the phase boundary energy per segment (with two phase boundaries). The last sum in the Hamiltonian represents the long-range polyspin interaction, related to the loop-entropy contribution of an L-length melted segment (with $S_{r+1}=S_{r+2}=\cdots=S_{r+L}=\frac{1}{2}$, otherwise the last term in H equals zero), bordering on helix sites [as $S_r=S_{r+L+1}=-\frac{1}{2}$, or else the last term in Eq. (1) is zero]. A quantity χ is the characteristic winding angle between adjacent sites, determined by the elasticity of the strands.[30,36]

As I have already mentioned, $T \ll J$. Thus the effective temperature is low, fluctuations are small, and the free energy differs only slightly from the ground-state energy $E=\min H$, where the minimum is determined with respect to all possible sets $\{S_r\}$. So we start with the determination of the ground state. To make the reasoning more vivid, we demonstrate it first with the example of $b=0$, where

$$H=-\sum_r h_r S_r - J\sum_r S_r S_{r+1}\,. \qquad (3)$$

When $T<T^{(1,2)}$ or $T>T^{(1,2)}$, i.e., when, by Eq. (2), $h_r<0$ or $h_r>0$ for all r, the ground state is obviously a homophase one: all $S_r<0$ or all $S_r>0$, correspondingly, as this minimizes both terms in Eq. (3). So we assume $T^{(1)}<T<T^{(2)}$, and thus, by Eq. (1), $h_r>0$ at the first component and $h_r<0$ at the second component.

First let us consider a simple example. Suppose $h^{(1)}=1, h^{(2)}=-1$, and the component sequence $\{j_r\}$ is 111111222221111111. Let us plot the difference ΔH_ρ in the "helix" and "melted" energies of the first ρ sites against ρ; by Eq. (3),

$$\Delta H_\rho=\sum_{r=1}^{\rho} h_r \equiv \sum_{r=1}^{\rho} h^{(j_r)}\,. \qquad (4)$$

Such a plot for our example is presented in Fig. 3, where each first component provides the ascent $h^{(1)}=1$, while each second component provides the descent $h^{(2)}=-1$. For the whole sequence $\Delta H_{18}=8>0$, i.e., the completely-helix sequence has a larger energy than the completely melted sequence. Thus a completely melted state is energetically preferable to a completely helix one.

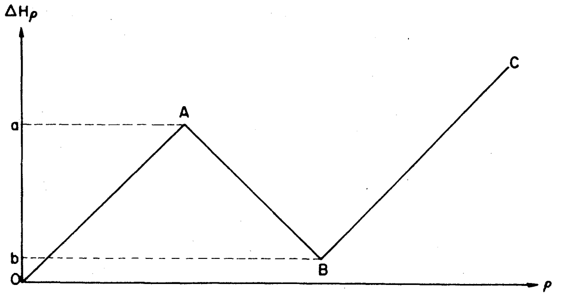

FIG. 3. Difference ΔH_ρ of helix and melted energies of the first ρ sites, plotted against ρ for $h^{(1)}=-h^{(2)}=1$ and the component sequence 111111222221111111, where 1 and 2 denote correspondingly the first and the second components.

According to Fig. 3, ΔH_ρ decreases at a segment AB, i.e., the helix state of this segment (containing only the refractory component) would decrease the energy by $ba=5$ if it were not for the exchange-interaction energy J contributed by two antiparallel spin boundaries. When $J>5$ the helixing of AB increases the energy, and the ground state of this sequence is completely melted; when $J<5$ the ground state is provided by helix AB and melted $0A$ and BC. The considerations in a general case are similar. Suppose ΔH_ρ against ρ has the shape presented by Fig. 4, where $\Delta H_{BE}>|\Delta H_{CD}|>2\Delta H_{0A}$ (and, e.g., $\Delta H_{BE}\equiv\Delta H_E-\Delta H_B$).

As $\Delta H_F<0$, a completely helix state is preferable to a completely melted one. The largest as-

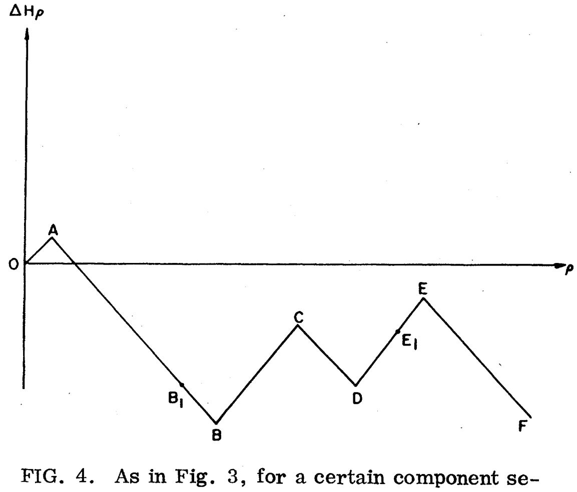

FIG. 4. As in Fig. 3, for a certain component sequence.

cent, i.e., the largest possible energy decrease in the case of melting, is achieved at BE. Thus if $J > \Delta H_{BE}$ the ground state is completely helix. When $J = \Delta H_{BE} - 0$, BE melts, i.e., the ground state then consists of melted BE and helix $0B$ and EF. In the latter case, at melted BE there is a descent CD, which (as it was with AB in Fig. 3) indicates the lower energy of the helix state. Similar to the situation in Fig. 3, the CD helixing occurs when $J = |\Delta H_{CD}| - 0$. Finally, when $\frac{1}{2}J = \Delta H_{0A} - 0$, then $0A$ melts (note that $0A$ in this case has only one phase boundary) and the ground state consists of melted $0A$, BC, DE, and helix AB, CD, EF.

Thus the plot of ΔH_ρ against ρ allows us[12,29] to determine the ground state for any J (starting with a sufficiently large J and then consequently decreasing it). Of course, the ground state may be degenerate. This happens when helix and melted energies of a certain domain are equal. In such a case we may exclude the ambiguity, e.g., by "preferring" the melted state [i.e., the one which becomes energetically preferable when the temperature T, and therefore $h^{(1)}$ and $h^{(2)}$ by Eq. (2), increase infinitesimally].

The previous algorithm can easily be formulated analytically. For example, an arbitrary ground-state melted domain PR, bordering (at P and R) on the ground-state helix domain, has a height $\Delta H_{PR} \geqslant J$ and does not contain (in the plot of ΔH_ρ) any descent exceeding J or any point lower than P or higher than R.

When the temperature, and thus $h^{(1)}$ and $h^{(2)}$, increases, ΔH_ρ, by Eqs. (2) and (4), increases monotonically, thus implying consequent DNA melting. A melting domain may appear in five possible ways: (ii) amidst helix domains, thus creating $n = 2$ phase (i.e., "antiparallel spin") boundaries and contributing the exchange energy J; (ii) amidst melted domains, thus annihilating two phase boundaries ($n = -2$) and decreasing the exchange energy by J; (iii) adjacent to an already melted domain, thus just shifting the boundary ($n = 0$) and leaving the exchange energy unchanged; (iv) bordering on the DNA end and on the melted domain ($n = -1$); (v) bordering on the DNA end and on the helix domain ($n = 1$).

If the melting domain PR contains $i^{(1)}$ first-component sites and $i^{(2)}$ second component sites, then its melting is described by the equation

$$\Delta H_{PR} = h^{(1)} i^{(1)} + h^{(2)} i^{(2)} = \tfrac{1}{2} nJ. \tag{5}$$

Thus, accounting for Eq. (2), the ground-state domain melts at the temperature[36(a)] $T = T_m$,

$$T_m = T^{(1)} X^{(1)} + T^{(2)} X^{(2)} + nT_b/l, \tag{6}$$

where

$$X^{(1)} = \frac{i^{(1)}}{l}, \quad X^{(2)} = \frac{i^{(2)}}{l} = 1 - X^{(1)}, \quad T_b = \frac{J}{2S}. \tag{7}$$

Now it is possible to describe the statistical structure of the component sequence with respect to its ground state. According to Eq. (3),

$$H \equiv \tfrac{1}{2} \Delta h \cdot \tilde{H}, \quad \tilde{H} = -\sum \tilde{h}_r S_r - w \sum_r S_r S_{r+1}, \tag{8}$$

where, by Eq. (2),

$$\Delta h \equiv h^{(1)} - h^{(2)} = s\Delta T,$$

$$\tilde{h}_r \equiv \tilde{h}^{(j_r)}, \quad \tilde{h}^{(1)} = 1 + p, \quad \tilde{h}^{(2)} = p - 1, \tag{9}$$

$$\Delta T = T^{(2)} - T^{(1)}, \quad \overline{T} = \tfrac{1}{2}(T^{(1)} + T^{(2)}),$$

$$p = 2(T - \overline{T})/\Delta T, \quad w = 2J/s\Delta T. \tag{10}$$

By Eqs. (8) and (9) the plot of $\Delta\tilde{H}_\rho$ is specified by the parameter p; together with the parameter w it uniquely determines the ground state for an arbitrary sequence.

Suppose we monotonically increase p keeping w fixed. Then the ground state melts monotonically. Suppose the number of domains, which for a given w melt in the interval $(p, p + dp)$ having length l and creating n new phase boundaries,[37] is dN,

$$dN = g_n(l; p, w)dp; \quad n = 0, \pm 1, \pm 2. \tag{11}$$

When a domain melts at p' its helix and melted energies are equal. When $p > p'$ increases the difference ϵ_G between its melted and helix energies changes as

$$\epsilon_G = \tfrac{1}{2} \Delta hl(p' - p), \tag{12}$$

since any change in the number of phase boundaries is related to domains which melt later. Thus the difference between the ground-state energy E for a given p and the energy H_h of a completely helix state is

$$E - H_h = -\tfrac{1}{2} \Delta h \sum_{n,l} l \int^p (p - p') g_n(l; p', w) dp',$$

$$H_h = \tfrac{1}{2} \Delta h(pN + \Delta N) - \tfrac{1}{4} NJ, \tag{13}$$

$$N = N_1 + N_2, \quad \Delta N = N_1 - N_2.$$

Here N_1 and N_2 are the total numbers of the first and the second component sites of DNA. Thus, except for the "end" effect of $n = \pm 1$, the ground-state energy of an *arbitrary* given component sequence is related to three distribution functions g_0, g_2, and g_{-2}, depending on three variables each. These functions are related to a very special plot; they describe very special domains, which may be of any length and in a general case can be related to a set, *however large*, of many-site correlation functions.

Obviously, Eq. (13) is the final formula in the case of an arbitrary sequence, since then the ground state should just be related to the statistical structure of the sequence (and no further progress is possible unless some assumptions are made about the sequence). Then one may investigate what information about the sequence can be obtained from corresponding experiments (see below).

In the case of a random sequence the plot of ΔH_ρ against ρ represents a random walk. Then Eq. (13) allows us to obtain the explicit formula for the ground-state energy E, which coincides with that of Refs. 34, 29, and 38. Even in this simplest case the formula[38] for E is far from trivial:

$$\frac{E}{N}=\tfrac{1}{2}\bar{h}-\tfrac{1}{4}J-\bar{h}\frac{\bar{h}+\lambda h^{(1)}h^{(2)}}{\bar{h}(1-e^{-\lambda J})+\lambda h^{(1)}h^{(2)}}, \tag{14}$$

where

$$\bar{h}=h^{(1)}\bar{X}^{(1)}+h^{(2)}\bar{X}^{(2)},$$
$$\bar{X}^{(1)}=N_1/N, \quad \bar{X}^{(2)}=N_2/N, \tag{14a}$$

and λ is the root of the equation

$$\bar{X}^{(1)}\exp(-\lambda h^{(1)})+\bar{X}^{(2)}\exp(-\lambda h^{(2)})=1. \tag{14b}$$

In the leading approximation over $\bar{h}/\Delta h$, Eqs. (14) and (14b) reduce to the Vedenov-Dykhne-Lifshitz formula[34,29]

$$E=\tfrac{1}{2}\bar{h}-\tfrac{1}{4}J-\bar{h}/[1-\exp(-\bar{h}J/\beta)],$$
$$\beta=\tfrac{1}{2}(\bar{X}^{(1)}h^{(1)2}+\bar{X}^{(2)}h^{(2)2}). \tag{14c}$$

Now let us consider the Hamiltonian (1) with $b\neq 0$. In this case the change in energy, provided by the melting of a domain, is no longer independent of its environment. Suppose, for instance, that in Fig. 5 domains $A'A$, BB', and CC' are melted and contain L sites each, while domains AB and $B'C$ are helix and contain $l \ll L$ sites each. Then, if AB melts first, the "loop energy", given by Eq. (1) changes by

$$\epsilon_1=bT\ln(2L^*+l^*)-2bT\ln L^*\approx -b_1\ln L^*,$$
$$b_1\equiv bkT, \quad L^*\equiv L\chi^2, \quad l^*\equiv l\chi^2. \tag{15a}$$

If $B'C$ melts first (changing the "loop energy" by ϵ_1) and AB melts second, then AB melting changes the "loop energy" by

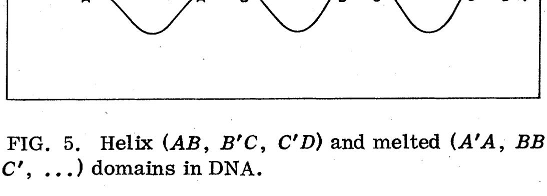

FIG. 5. Helix (AB, $B'C$, $C'D$) and melted ($A'A$, BB', CC', ...) domains in DNA.

$$\epsilon_1'=b_1\ln(3L_1^*)-b_1\ln(2L^*+l^*)-b_1\ln L^*$$
$$\approx b_1\ln L^*+b_1\ln 1.5=\epsilon_1+b_1\ln 1.5. \tag{15b}$$

Meanwhile, if AB and $B'C$ melt simultaneously, they change the "loop energy" by

$$\epsilon_2=b_1\ln(3L^*+2l^*)-3b_1\ln L^*$$
$$\approx -2b_1\ln L^*+b_1\ln 3=\epsilon_1+\epsilon_1'+b_1\ln 2. \tag{15c}$$

Thus the nonadditivity in Eqs. (15b) and (15c) is on the order of $b_1 \ll J$, i.e., small, and can be accounted for as a perturbation. This is a general situation, as either $b_1\ln L \ll J$ or $b_1 \ln L \sim J$ (and therefore L is very large), and its change (which is responsible for the energy nonadditivity) with L is thus small compared to J. Therefore in the leading approximation one can ignore the nonadditivity of energies and determine the ground state according to the algorithm described earlier, just accounting for the "loop-energy" contribution while the domain is melting. For instance, if a domain contains l sites and melts between melted domains containing L_1 and L_2 sites and not bordering on DNA ends, then Eq. (5) (with $n=-2$) should be replaced by the equation

$$h^{(1)}i^{(1)}+h^{(2)}i^{(2)}$$
$$=-J-bT\ln[L_1L_2\chi^2/(L_1+L_2+l)]. \tag{16}$$

III. FREE ENERGY AND MELTING CURVE

In the ground state each domain melts at its "own" melting temperature. Once the domain is melted, the energy difference $-\epsilon_c$ between its helix and melted energies increases with the temperature increase δT, by Eqs. (1) and (2), as $ls\delta T$. It becomes large compared with T very quickly, thus making the fluctuation helixing of this domain exponentially improbable. Similarly, the fluctuation melting of a domain somewhat below its melting temperature is also exponentially improbable. Therefore, at any given temperature only relatively few domains may be in their narrow "melting intervals." So the low-energy excitations (which are the only ones of importance at low temperatures) can be of two types. They refer either to a slight shift of the domain boundaries (e.g., from B to B_1 or from E to E_1 in Fig. 4) or to the change in the state of those (relatively rare) domains which are in their melting intervals; I shall refer to such domains as "flexible," as opposed to "rigid," with $|\epsilon_c| \gg T$. As flexible domains are rare, the DNA free energy can be expressed through a successively decreasing contribution from "separate" flexible domains, their adjacent pairs, triplets, etc.

Let us demonstrate, for instance, the term of this expansion related to separate flexible domains. Suppose the temperature increases monotonically, implying the successive melting of domains. If we measure the energy and free energy from the completely helix state, then a flexible domain, situated between rigid helix domains (e.g., BE in Fig. 4), contributes zero to the energy when it is helix, and the energy ϵ_c when it is melted. For instance, for the segment B_1E_1 in Fig. 4, by Eqs. (1) and (2)

$$\begin{aligned}\epsilon_c(B_1E_1) &= -h^{(1)}i^{(1)}(B_1E_1) - h^{(2)}i^{(2)}(B_1E_1) \\ &\quad + J + bT\ln[\chi^2 L(B_1E_1)] \\ &\equiv \epsilon_c(BE) + \epsilon_c(BB_1) + \epsilon_c(E_1E) . \end{aligned} \tag{17}$$

Here ϵ_c is the change in the energy related to the melting of the segment in the brackets (providing BE melts first):

$$\begin{aligned}\epsilon_c(BE) &= -h^{(1)}i^{(1)}(BE) - h^{(2)}i^{(2)}(BE) + J \\ &\quad + bT\ln[\chi^2 L(BE)] \equiv s(T_m - T) , \end{aligned} \tag{18}$$

$$\begin{aligned}\epsilon_c(B_1B) &\approx -h^{(1)}i^{(1)}(B_1B) - h^{(2)}i^{(2)}(B_1B) \\ &\quad + bTL(B_1B)/L(BE) , \\ \epsilon_c(EE_1) &\approx -h^{(1)}i^{(1)}(EE_1) - h^{(2)}i^{(2)}(EE_1) \\ &\quad + bTL(EE_1)/L(BE) . \end{aligned} \tag{19}$$

Here $i^{(1)}$, $i^{(2)}$, and L correspondingly denote the number of sites of the first and second component and the length of the segment in brackets [we account for the sign in L; so, e.g., $L(E_1E) < 0$ in Fig. 4]; T_m is the melting temperature of the domain.[39]

As the segments can only be in two states, their contribution ΔF^+ to the free energy F (measured from the completely helix state) is of the "Fermi type"; for B_1E_1 the contribution is

$$\begin{aligned}\Delta F^+ &= -T\ln\Big(1 + \sum_{B_1,E_1} \exp[-\epsilon_c(B_1E_1)/T]\Big) \\ &\approx -T\ln(1 + Q_1Q_2t) , \end{aligned} \tag{20}$$

$$\begin{aligned} Q_1 &= \sum_{B_1} \exp[-\epsilon_c(BB_1)/T] , \\ Q_2 &= \sum_{E_1} \exp[-\epsilon_c(EE_1)/T] , \end{aligned} \tag{21}$$

$$t = \exp[-\epsilon_c(BE)/T] . \tag{22}$$

The summation in Eqs. (20) and (21) refers to possible boundary shifts, the superscript "+" indicates that the change n in the number of phase boundaries is positive, $n = 2$, and the temperature is measured in energy units. Boundaries B_1 and E_1 should not meet each other [if they meet, it means that the whole flexible domain is helix, which is accounted for separately by the first term in Eq. (20)]. However, within the accuracy of our approximation, we should not care about the large boundary shift anyway, as it essentially increases the energy (the shift of the boundaries, which makes them close to each other, increases the energy by approximately J) and is therefore exponentially improbable.

When the melted domain becomes rigid, i.e., $-\epsilon_c \gg T$, then, by Eqs. (20) and (22),

$$\Delta F^+ \approx \epsilon_c(BE) - T\ln(Q_1Q_2) . \tag{20a}$$

Now suppose a flexible domain is situated between rigid melted domains. Then the reasoning is exactly the same, but we should take into account that the adjacent rigid domains melted at lower temperatures. Therefore the contribution $-T\ln(Q_1Q_2)$ of the shift of their boundaries "1" and "2" [see Eq. (20a)] has already been considered and should not be accounted for twice. Therefore (as a helix flexible domain now contributes only the energy related to the phase boundary shift, while a melted flexible domain contributes the energy ϵ_c) the corresponding contribution F^- to F equals

$$\begin{aligned}\Delta F^- &= -T\ln(Q_1Q_2 + t) + T\ln(Q_1Q_2) \\ &= -T\ln(1 + t/Q_1Q_2) , \end{aligned} \tag{23}$$

where Q_1, Q_2, and t are calculated according to Eqs. (21) and (22) for the corresponding flexible domain.[40]

Analogous considerations for a flexible domain, which borders on rigid helix and melted domains, provide the following formula for its contribution ΔF^0 to the free energy:

$$\begin{aligned}\Delta F^0 &= -kT\ln(Q_1 + Q_2t) + kT\ln Q_1 \\ &= -kT\ln(1 + Q_2t/Q_1) , \end{aligned} \tag{24}$$

where Q_1 refers to the boundary with the melted rigid domain (which has already been taken into account at lower temperatures), and ϵ_c, which determines t, refers to the flexible domain.

A flexible domain which borders on the DNA end is accounted for in the same way, but since it has only one phase boundary, only one Q enters the corresponding formula; the other Q should be replaced by 1.

The total free energy F equals the sum of contributions (20), (23), and (24) from all consequently melting domains. If m is the ordinal number of the melting domain in the given type (i.e., with $n_m = 2, -2, 0$) then

$$\begin{aligned} F - H_h &= -kT\sum_m \{\ln(1 + Q_{1m}Q_{2m}t_m) \\ &\quad + \ln(1 + t_m/Q_{1m}Q_{2m}) + \ln(1 + Q_{2m}t_m/Q_{1m})\} , \end{aligned} \tag{25}$$

where H_h is provided by Eq. (13).

Thus the procedure of the determination of the DNA free energy is as follows. According to the algorithm of the previous section we determine all consequently melting ground-state domains, evaluate their Q_1, Q_2, and ϵ_c, and then write down formula (25). Equation (25) can in no way be related to any perturbation or mean-field theory, as the ground state and its excitations (in particular, their locations) depend, according to Sec. II, on the whole $\{h_r\}$ sequence and its detailed structure. Note also that all our considerations were based only on a large value of J. Therefore they are readily generalized to any refinement of the Hamiltonian (1) (e.g., to that accounting for the interaction of adjacent sites).

Now that we know the free energy of the two-component ferromagnet (1), evaluation its magnetic moment $M=\frac{1}{2}N_c-\frac{1}{2}(N-N_c)=N_c-\frac{1}{2}N$, is no problem ($N_c$ is the number of "spin up," i.e., of melted, sites) and $dM/dT=dN_c/dT$, which are the experimentally measured quantities.[1,9,11] In the leading approximation over T/J we obtain

$$\frac{dM}{dT}=\frac{dN_c}{dT}\approx\sum_m\frac{\beta l_m^2}{2\cosh^2[\beta l_m(T-T_m)]}\,,\quad \beta=\frac{s}{2T}\,, \tag{26}$$

where all types of domains are included and T_m is by definition determined by the relation

$$\epsilon_c\equiv s(T_m-T)\,. \tag{27}$$

In relatively short DNA's with short melted domains, the loop-entropy term is small compared to J and can be neglected; then T_m is provided by Eq. (6). Using parameters $T^{(1)}=52.5\,^\circ\mathrm{C}$, $T^{(2)}=94.9\,^\circ\mathrm{C}$ from Ref. 50; $\beta=0.0106\,^\circ\mathrm{C}^{-1}$, $T_b=100\,^\circ\mathrm{C}$, and the known DNA sequence from Ref. 4, we obtain agreement with the plots of Ref. 21 (see, e.g., Fig. 6) within limits of experimental accuracy.

Let us estimate the accuracy of Eq. (25). It is related to two factors: we neglected the contribution of adjacent flexible domains and of high-energy fluctuations. Now we demonstrate the contribution of high-energy fluctuations in the example of a homopolymer (which is similar to a large ground-state domain with average composition) and the contribution of the adjacency of flexible domains in the example of a periodic polymer, where all flexible helix domains are adjacent.

IV. HOMOPOLYMER AND PERIODIC POLYMER

The Gibbs probability ω_{lL} of $l\geq 1$ helix sites, which follow a melted site and are followed by $L\geq 1$ melted sites, in a homopolymer equals

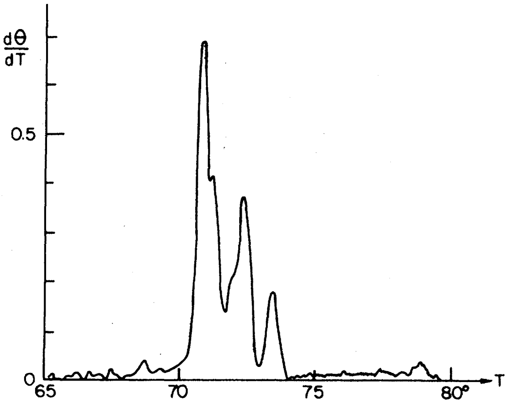

FIG. 6. The experimental differential melting curve (Ref. 21) for the fragment of the DNA of phage ϕX-174 (fragment y_1 of Ref. 21). The quantity θ is the relative number of melted sites; $d\theta/dT=N^{-1}dM/dT$, where N is the total number of sites.

$$\omega_{lL}=\exp\{[f(l+L)-(\tfrac{1}{2}lh-\tfrac{1}{2}Lh+bkT\ln L+J)]/T\}\,, \tag{28}$$

where we (i) do not need the subscript "r" in h_r in Eq. (1) for a homopolymer, (ii) consider, for simplicity, $\chi=1$, and (iii) denote by f the free energy per site.

Obviously,

$$\sum_{l,L=1}^{\infty}\omega_{lL}=1\,, \tag{29}$$

and thus, by Eqs. (28) and (29),

$$\exp[(h-\phi)/T]-1$$
$$=\exp(-J/T)\sum_{L=1}^{\infty}L^{-b}\exp(L\phi/T)\,, \tag{30}$$

$$\phi\equiv f+\tfrac{1}{2}h\,, \tag{31}$$

where the convergence of the right-hand side in Eq. (30) implies $\phi\leq 0$; $\phi=0$ is possible only if $b>1$.

When $b>1$, then Eq. (30) implies

$$\begin{aligned}&-\phi\propto\exp(-J/T)\,,\quad h<h_k\,,\\&\phi\equiv 0\,,\quad h>h_k\,,\end{aligned} \tag{32}$$

where, by Eq. (30),

$$\exp(h_k/T_k)-1=\zeta(b)\exp(-J/T_k)\,, \tag{33}$$

and thus

$$h_k/T_k\approx\zeta(b)\exp(-J/T_k)\,, \tag{33a}$$

$\zeta(b)$ being the Riemann ζ function

$$\zeta(b)\equiv\sum_{m=1}^{\infty}m^{-b}\,. \tag{33b}$$

Equation (30) implies the Poland-Scheraga-Applequist phase transition.[24-26]

Now we may apply this reasoning to Eq. (25). At helix ground-state domains, according to Sec. II, $\langle h\rangle < 0 < h_k$, $\langle h\rangle$ being the average h_r in the domain. Therefore, the contribution of high-energy excitations there is exponentially small. In melted ground-state domains $\langle h\rangle > 0$. As $h_k \propto \exp(-J/T)$, these domains (except for their exponentially small fraction and for effects related to their finite length) do not have excitations according to Eq. (32).

Now suppose $b<1$. Then, if $-h \gtrsim T$, the sum in Eq. (30) quickly converges owing to $\phi \approx h$, and Eq. (30) provides

$$h-\phi \approx \exp(-J/T)\sum_{L=1}^{\infty} L^{-b}\exp(Lh/T) \propto \exp(-J/T)\,. \tag{34}$$

If $h/T \gtrsim 0$, then $\phi \approx 0$ and Eq. (30) provides[41]

$$\exp[(h+|\phi|)/T]-1 \sim \exp(-J/T)\,|\phi|^{b-1}\,, \tag{35}$$

and thus

$$\begin{aligned} &\phi \sim \phi_0 \sim \exp[-J/(2-b)T]\,, \quad |h| < \phi_0;\\ &\phi \sim |h|^{-1/(1-b)}\exp[-J/(1-b)T]\,, \quad h > \phi_0\,. \end{aligned} \tag{36}$$

Thus, as should be expected, the high-energy excitations in all cases give exponentially small relative refinements in Eq. (25) which are unimportant, at least until the ground state is a two-phase one.

Now we consider a periodic polymer which consists of identical Λ-length "fusible" domains, which melt first, with λ-length "refractory" domains between them, which melt second, and assume $\chi=1$ and a temperature at which Λ-length domains are already rigid melted ground-state domains. The Gibbs probability ω_m of $m>1$ consequent melted λ-length domains, which follow the helix λ-length domain, is

$$\begin{aligned} \omega_m = \exp(\{&m\Lambda^* f - [\tfrac{1}{2}\lambda\epsilon_\lambda - \tfrac{1}{2}m\Lambda\epsilon_\Lambda - \tfrac{1}{2}(m-1)\lambda\epsilon_\lambda\\ &+ J + bT\ln(m\Lambda^* - \lambda)]\}/T)\,,\\ \Lambda^* = \Lambda&+\lambda\,, \end{aligned} \tag{37}$$

where we denote the free energy per site by f, the helix energy per site of the λ- and Λ-length domains by $\frac{1}{2}\epsilon_\lambda$ and $\frac{1}{2}\epsilon_\Lambda$, while the coiled energies of λ and Λ domains are $-\frac{1}{2}\epsilon_\lambda$ and $-\frac{1}{2}\epsilon_\Lambda$ per site. The equation

$$\sum_{m=1}^{\infty}\omega_m = 1 \tag{38}$$

is analogous to Eq. (29) and provides

$$\exp(\lambda\tilde{h}/T) = \sum_{m=1}^{\infty}\exp(m\Lambda^*\phi/T)(m-\lambda/\Lambda^*)^{-b}\,, \tag{39}$$

where

$$\begin{aligned} &\phi = f + \tfrac{1}{2}\bar{h}\,,\\ &\bar{h} = (h_\lambda\lambda + h_\Lambda\Lambda)/\Lambda^*\,,\\ &\tilde{h} = h_\lambda + (J + bT\ln\Lambda^*)/\lambda\,. \end{aligned} \tag{39a}$$

All further considerations are similar to those for a homopolymer but they no longer involve a large parameter J/T. For instance, $b>1$ implies a phase transition (to $\phi \equiv 0$) of the same nature as in a homopolymer, but at a temperature at which

$$\tilde{h} = \tilde{h}_k = (T/\lambda)\ln\sum_{m=1}^{\infty}(m-\lambda/\Lambda^*)^{-b} \sim T/\lambda\,. \tag{40}$$

[i.e., by Eq. (39a), at a temperature somewhat higher than that at which the ground state becomes a homophase one]. The transition is related to the "many-domain" effect (as the transition singularity is provided by $m \to \infty$), i.e., to the collective effects of interacting adjacent flexible domains. According to previous reasoning [see, e.g., Eqs. (15b) and (15c)] this interaction slightly renormalizes the energy which determines the transition temperature.[42]

Equation (39) determines the free energy. For instance, near the phase transition point, when $1<b<2$,

$$\begin{aligned} &-\phi \sim (T/\Lambda^*)[\lambda(\tilde{h}_k - \tilde{h})/T]^{1/(b-1)}\,,\\ &0 < \tilde{h}_k - \tilde{h} \ll (T/\lambda)\,,\\ &\phi \equiv 0\,, \quad \tilde{h} > \tilde{h}_k\,. \end{aligned} \tag{41}$$

If $b<1$ (so there is no phase transition) and $\lambda\tilde{h}/T > 1$, then

$$-\phi \sim (T/\Lambda^*)\exp[-\lambda\tilde{h}/T(1-b)]\,. \tag{42}$$

In all cases $\phi \propto 1/\Lambda^*$ and therefore, when the ground state is a two-phase one, the adjacency of flexible domains just renormalizes the terms in Eq. (25) [as should be expected from general considerations in Sec. II and Eqs. (15b) and (15c)]. But when the ground state becomes a homophase one this renormalization either implies a phase transition when $b>1$ and Λ and λ are finite, or determines the difference between the free energy and the ground-state energy when $b<1$ and $\lambda\tilde{h}/T > 1$.

The above considerations are readily applied to a general case of adjacent ground-state flexible domains in Eq. (25). The necessity in the consideration of the contribution of the adjacency of flexible domains is related to the fact that in the immediate vicinity of the temperature T_k of the transition to the homophase melted ground state all remaining ground-state helix domains obviously have their melting temperatures very close to T_k and are therefore flexible. Thus every ground-state helix domain has two adjacent flexible domains bordering on each of its neighboring rigid melted ground-state domains.

V. PHASE TRANSITIONS

According to previous reasoning, the phase transition, related to DNA melting, occurs when $b>1$ (and of course $N\to\infty$) and can take place at $T=T_k$ in the following ways:

(i) The ground state is always homophase and just changes from a completely helix to completely melted one. It means that, with respect to the ground state, a heteropolymer is quasihomogeneous, there are no regions long enough and inhomogeneous enough, to provide the inhomogeneous ground state. This case is similar to the case of a homopolymer [related to the average heteropolymer with $\bar{h}$ from Eq. (14a)] and can be treated similarly. In this case the half-width of the melting curve, by Eq. (34), is exponentially small over J/T. The order of the phase transition depends on b. When $\tau\equiv(T_k-T)/T_k>0$, then, similar to Eqs. (30) and (33),

$$\begin{aligned} &\phi\propto\tau^{1/(b-1)}\,,\quad 1<b<2\,,\\ &\phi\propto\tau\ln\tau\,,\quad b=2\,,\\ &\phi\propto\tau+\rho\tau^{b-1}\,,\quad b>2,\ \text{nonintegral}\,,\\ &\phi\propto\tau+\rho\tau^{b-1}\ln\tau\,,\quad b>2,\ \text{integral}\,, \end{aligned}\tag{43}$$

where ρ is a constant. When $\tau<0$, then $\phi\equiv0$ (cf. Ref. 26).

(ii) The ground state changes from a two-phase state to a completely melted state; at the transition point of the ground state, average lengths of ground-state helix and melted domains remain finite. This case is similar to the case of a periodic heteropolymer of Sec. IV and can be treated similarly. The phase transition half-width is inversely proportional, by Eq. (40), to an average length of the ground-state helix domains; the nature of the transition is the same as in Eq. (43). In particular, the singularities of the type described by Eq. (43) are typical for quasiperiodical sequences.

(iii) The polymer reduces to a set of (infinite) polymers; the density of the polymers with the "critical" $h=h'$ and the corresponding free energy per site $\phi(h'-h)\sigma(h'-h)$, [where $\sigma(X)=0$ when $X>0$ and $\sigma(X)=0$ when $X<0$; the singularity of $\phi(X)$ is determined by Eq. (43)] is $c(h')$. Then the free energy ϕ per site of the initial heteropolymer is

$$\phi=\int\phi(h'-h)\sigma(h'-h)c(h')dh'\,.\tag{44}$$

Obviously, the resulting singularity can be arbitrarily smeared.

(iv) The ground state melts with infinitely increasing average domain lengths. Then, by Eqs. (40) and (41), the transition occurs at the melting point of the ground state and is described by Eq. (25). The phase transition singularity is determined by the singularity in the density of the ground-state helix sites at the ground-state transition temperature and can be of a different nature. Let us consider, for instance, the case of a sequence with finite correlation radius r_c. When the characteristic lengths of helix l_c and melted L_c domains become much larger than r_c, the probability density $\omega(l,X)$ of an l-length segment with concentration X of the first component is determined by the Gaussian distribution

$$\omega(l,X)\propto\exp\{-l(X-\bar{X})^2/2\Delta X^2\}\,,\tag{45}$$

$\Delta X/\sqrt{l}$ being the mean quadratic concentration fluctuation and $\bar{X}$ the average X. By Eqs. (16) and (7) the segment (l,X) may be the ground-state helix domain (with $n=-2$, as helix ground-state domains are rare and close to their melting temperature near the transition point), if

$$h^{(1)}X+h^{(2)}(1-X)\approx-[J+bkT\ln(\chi^2L_c)]/l_c\,.\tag{46}$$

The characteristic l_c is determined by

$$\omega_0(L)\equiv\max_l\omega(l,X)\,,\tag{46a}$$

with X from Eq. (46), while

$$L_c\sim1/\omega_0(L_c)\,.\tag{47}$$

Equation (47) determines L_c, and thus $\phi\propto1/L_c$ (ϕ is measured from the completely melted energy), the transition temperature T_k, and the transition singularity. The latter is of the essential singularity type: $\ln|\phi|\propto(T_k-T)^{-1}$ (cf. Ref. 26). Obviously, these considerations are applicable whenever $l_c, L_c\gg r_c$. (Thus the essential singularity is characteristic for quasirandom sequences.) According to Sec. II, l_c, $L_c\propto J/T\gg1$, so for r_c, which is not too large, they may be applicable in the whole melting interval. In the leading approximation over T/J the exact knowledge of L_c is immaterial, as L_c enters the formulas only in the form of $J+bT\ln L$. The quantity $bT\ln L$ is not negligible compared to J only when L is exponentially large, most of the DNA is melted, helix domains are rare, and $n=-2$. Therefore, to account for $b\neq0$, we should just replace J (in the equation for $b=0$) by $J^*\equiv J+bT\ln L_c$, with L_c determined by Eqs. (45)–(47). Simple evaluations (providing the same accuracy and accounting for $L_c>1$ at any temperature) result in (cf. Ref. 26)

$$J^* = J/[1 - 2bT\bar{h}/(\Delta h \Delta X)^2] ,$$
$$\Delta h = h^{(1)} - h^{(2)} , \qquad (48)$$
$$\bar{h} = h^{(1)}\bar{X} + h^{(2)}(1 - \bar{X}) .$$

When $b = 0$ the leading (over T/J) approximation for the free energy is provided by the Vedenov-Dykhne-Lifshitz formula (14c); a more accurate Vilenin formula for the ground-state energy E of a random sequence [with $\Delta X^2 = \bar{X}(1 - \bar{X})$] is provided by Eqs. (14), (14a), and (14b).

To summarize, the nature of the phase transition essentially depends on the value of b and on the nature of the "randomness" of the sequence.

VI. INVERSE PROBLEM

The differential melting curve $dM/dT \equiv dN_c/dT$ is described by Eq. (26). Each term in the sum in this equation has a sharp maximum at $T = T_m$. The relative half-width of the peak is $\delta T/T \sim 1/sl \ll 1$. The maxima provide the oscillatory picture observed in numerous experiments. When the DNA total number of sites is not too large[43] each peak is related just to one or several melting domains. That is why the fitting of the experimental dN_c/dT to the theoretical formula reduces mainly to the fitting of each separate peak to its domains. This allows us to determine l_m and T_m for these domains.

The light absorption A changes with the light wavelength somewhat differently for different components,[51,44,45,52,18,11] so

$$\frac{dA}{dT} = \mu_1 \frac{dN_c^{(1)}}{dT} + \mu_2 \frac{dN_c^{(2)}}{dT} , \qquad (49)$$

where μ_1 and μ_2 are independent of temperature and depend only on the wavelength, while the numbers of the first $N_c^{(1)}$ and the second $N_c^{(2)}$ component melted sites are independent of wavelength and depend only on temperature. Thus the use of different wavelengths allows us to determine separately $dN_c^{(1,2)}/dT$. The latter quantities are derived from Eq. (25) and are described by equations which differ from Eq. (26) only in the replacement of l_m^2 by $i_m^{(1,2)} l_m$:

$$\frac{dN_c^{(1,2)}}{dT} = \sum_m \tfrac{1}{2} \beta i_m^{(1,2)} \frac{l_m}{\cosh^2[\beta l_m (T - T_m)]} . \qquad (50)$$

The knowledge of $dN_c^{(1,2)}/dT$ allows us to determine $i_m^{(1)}$ and $i_m^{(2)}$, which (together with T_m) determine the change n_m in the number of phase boundaries.

An example of the solution of the inverse problem and its accuracy (compared to the known DNA sequence[4]) is demonstrated in Table I.

TABLE I. Domain melting temperatures T_m, lengths l_m, and first-component concentrations X_m, as according to the present analysis of experimental data of Wada *et al.* (Ref. 21) for the DNA fragment (the fragment y_1 containing 2745 sites) of phage ϕX-174. Superscript "E" indicates corresponding data for the known (Ref. 4) phage sequence (with unknown sites in it, replaced by the first component). The last column gives the ordinal number of the domain in the fragment. The adjustable parameters used are $\beta = 0.0106\,°\mathrm{C}^{-1}$ and $T_b = 100\,°\mathrm{C}$. Parameters $T^{(1)} = 52.5\,°\mathrm{C}$ and $T^{(2)} = 92.9\,°\mathrm{C}$ are taken from Ref. 50. The experimental plot, which is analyzed in Table I, is presented in Fig. 6.

T_m	l_m	l_m^E	X_m	X_m^E	Number of domains
70.5	233	239	0.585	$0.582\left\{\begin{smallmatrix}+0.01\\-0.1\end{smallmatrix}\right.$	I
70.87	641	671	0.570	0.566 ± 0.002	VII
71.2	468	457	0.559	0.558 ± 0.002	VI
71.5	215	220	0.574	0.591 ± 0.005	III
72.0	333	333	0.526	$0.571\left\{\begin{smallmatrix}+0.01\\-0.05\end{smallmatrix}\right.$	II
72.44	481	490	0.530	0.559 ± 0.015	IV
73.5	374	335	0.497	0.501 ± 0.003	V

When DNA is very long (i.e., contains 10^5 or more sites) the peaks may overlap and the plot of dN_c/dT may have only one or few[38] maxima.[46] Then Eqs. (26) and (50) become integral equations with respect to distribution functions g_n [from Eq. (13)] and allow us to determine these functions.[49]

VII. SUMMARY

(i) DNA is a double-stranded two-component molecule. When the temperature of its solvent increases it exhibits the transition to separate unbound ("melted") strands. This transition is described by a one-dimensional Ising Hamiltonian (1) in an external temperature-dependent inhomogeneous "magnetic field" (the sequence of magnetic fields is related to the DNA component sequence) with a long-range polyspin interaction. Spin "up" ($S = +\frac{1}{2}$) describes an unbound ("melted" or "coiled") site, spin "down" ($S = -\frac{1}{2}$) refers to a bound ("helix") site.

(ii) An explicit analytical formula for the thermodynamics of the Hamiltonian (1) is presented. No assumptions are made about the component sequence and therefore the solution is not related to usual approaches (such as perturbation or mean-field theory, scaling, renormalization group, etc.). In the special case of an infinite random component sequence the free energy f per site in the leading approximation (which becomes accurate when we approach the phase transition point) equals

$$f=\tfrac{1}{2}\bar{h}-\tfrac{1}{4}J^{*}-\bar{h}\frac{\bar{h}+\lambda h^{(1)}h^{(2)}}{\bar{h}[1-\exp(-\lambda J^{*})]+\lambda h^{(1)}h^{(2)}}, \quad (51)$$

where

$$\bar{h}=h^{(1)}\bar{X}^{(1)}+h^{(2)}\bar{X}^{(2)},$$

$$J^{*}=J/[1-2b\chi^{2}T\bar{h}/\bar{X}^{(1)}\bar{X}^{(2)}(\Delta h)^{2}], \quad (52)$$

$$\Delta h=h^{(1)}-h^{(2)};$$

$\bar{X}^{(1)}$ and $\bar{X}^{(2)}$ are the concentrations of the components ($\bar{X}^{(1)}+\bar{X}^{(2)}=1$), χ is a characteristic winding angle related to the elasticity of the DNA strands, and λ is the root of the equation

$$\bar{X}^{(1)}\exp(-\lambda h^{(1)})+\bar{X}^{(2)}\exp(-\lambda h^{(2)})=1. \quad (53)$$

(iii) DNA melting is not continuous but local. It occurs "step by step": each time a certain specific domain (containing typically several dozen or several hundred sites) melts. That is why the characteristic feature of the differential melting curve $dM/dT=dN_c/dT$ [M being the magnetic moment of the Ising Hamiltonian (1), N_c the number of melted sites, and dN_c/dT the experimentally measured quantity], in the general case of a finite arbitrary sequence is multiple sharp peaks (originated by the "quasijumps" of N_c due to local "melting"). These oscillations were observed in numerous experiments and are demonstrated in Fig. 2. In the cases when both the DNA sequence and the DNA melting curve are known the theoretical formula for dN_c/dT, applied to the given sequence, demonstrates an agreement with the experimental data within their accuracy.

(iv) If one fits the experimental oscillatory plot dN_c/dT to the theoretical formula, one can determine the length l, the component concentration X, and the number n of phase boundaries of melting domains. This is verified by the example of phage ϕX-174 DNA, where l, X, and n for all melting domains are in surprisingly good agreement with those for the known ϕX-174 component sequence (see Table I).

(v) When $b>1$ in the Hamiltonian (1), then DNA melting is a phase transition. The nature of this phase transition depends crucially on the quantity of b and on the component sequence. Correspondingly, the singularity at the transition point may be of any nature, from the first-order phase transition to the essential-singularity type.

Note added in proof. Recently J. Gabarro-Arpa, P. Tougart, and C. Reiss [Nature (to be published)] used Eqs. (26) and (27), first presented in Ref. 31, and proved their coincidence with their experiments on the virus SV-40 and with those of D. Vizard, R. White, and A. Ansevin [Nature 275, 250 (1978)] on the phage ϕX-174.

*This work was initiated during a stay at the Dept. of Physics of the Univ. of Pennsylvania.

[1] See, for example, e.g., A. A. Vedenov *et al.*, Sov. Phys. Usp. 14, 715 (1972).

[2] W. Fiers *et al.*, Nature (London) 260, 500 (1976).

[3] F. Sanger *et al.*, Proc. Nat. Acad. Sci. U.S.A. 74, 5463 (1977).

[4] F. Sanger *et al.*, Nature (London) 265, 687 (1977).

[5] V. B. Reddy *et al.*, Science 200, 494 (1978).

[6] W. Fiers *et al.*, Nature (London) 273, 113 (1978).

[7] Beck *et al.* (unpublished).

[8] Nucleotides along the strand are bound by the chemical bonds whose energy is about 30 000°. Thus their order can be considered fixed at room temperatures.

[9] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transition in Biopolymers* (Academic, New York, 1970).

[10] R. W. Wartell and E. W. Montroll, Adv. Chem. Phys. 22, 129 (1972).

[11] See, for instance, R. D. Blake and S. G. Lefoley, Biochim. Biophys. Acta 578, 233 (1978). The derivative dN_c/dT is obtained in the measurement of the difference in light absorption for solvents slightly different in temperature or in ionic strength.

[12] M. Ya. Azbel, Phys. Rev. Lett. 31, 589 (1973).

[13] M. Steinert and S. Van Assel, Biochem. Biophys. Res. Commun. 61, 1249 (1974).

[14] S. Yubuki *et al.*, Biochim. Biophys. Acta 395, 258 (1975).

[15] D. L. Vizard and A. T. Ansevin, Biochemistry 15, 741 (1976).

[16] A. T. Ansevin *et al.*, Biopolymers 15, 153 (1976).

[17] Gotoh *et al.*, Biopolymers 15, 655 (1976).

[18] C. Akiyama, D. Gotoh, and A. Wada, Biopolymers 16, 427 (1977).

[18(a)] H. Tachibana, A. Wada, D. Gotoh, and M. Takanami, Biochim. Biophys. Acta 517, 319 (1978).

[19] A. Wada *et al.*, Nature (London) 263, 439 (1976).

[20] C. Reiss and T. Arpa-Gabarro, Progr. Mol. Subcell. Biol. 5, 1 (1977).

[21] A. Wada *et al.*, Nature (London) 269, 352 (1977).

[22] Y. L. Lyubchenko *et al.*, Nature (London) 271, 28 (1977).

[23] H. Jacobson and W. H. Stockmayer, J. Chem. Phys. 18, 1600 (1950). We obviously assume $ACB=BDA$, which is practically always true in the case of DNA, as a DNA sequence is highly inhomogeneous and an accidental complementarity of nucleotide pairs is highly improbable.

[24] D. Poland and H. A. Scheraga, J. Chem. Phys. 45, 1456 (1966).

[25] J. Applequist, J. Chem. Phys. 45, 3459 (1966); 50, 600 (1969).

[26] M. Ya. Azbel, J. Chem. Phys. 62, 3635 (1975).

[27] The phase transition in a one-dimensional system seems to be forbidden by the Landau-Lifshitz theorem.[28] According to this theorem, a phase fluctuation may appear at any of $N\to\infty$ sites, thus providing

the entropy contribution $-T\ln N \rightarrow -\infty$, which exceeds any finite increase in the energy. However, if the effective long-range interaction increases with distance faster than $bT\ln L$ and $b > 1$, then it provides an increase in energy (due to a single fluctuation) $bT\ln N$ which exceeds the entropy contribution $T\ln N$ and makes the theorem invalid.

[28] L. D. Landau and I. M. Lifshitz, *Statistical Physics*, 2nd ed. (Addison-Wesley, Reading, Mass. 1969).

[29] I. M. Lifshitz, Sov. Phys. JETP 38, 545 (1974).

[30] M. Ya. Azbel, Ann. Isr. Phys. Soc. 2, 25 (1977).

[31] M. Ya. Azbel, J. Phys. A 12, L29 (1979).

[32] This is related to the fact that any thermodynamical quantity ϕ is determined from the partition function and therefore is an integral of a certain system characteristic $y(X)$ over some variable X: $\phi(X) = \int K(X,X')y(X')dX'$. A change of y by δy, where δy oscillates very quickly with X, may result in a very small $\int K(X,X')\delta y(X')dX'$, i.e., in a very small (on the order of an experimental error) change in $\phi(X)$, though the average of $(\delta y)^2$ may be on the order of y^2.

[33] The same value of s for both the first and the second components follows from the comparison of theoretical[34] and empirical[35] formulas for DNA melting temperatures (see also later).

[34] A. A. Vedenov and A. M. Dykhne, Zh. Eksp. Teor. Fiz. 55, 357 (1968) [Sov. Phys. JETP].

[35] J. Marmur and P. Doty, J. Mol. Biol. 5, 109 (1962).

[36] In fact, b and J also depend slightly on L because of elasticity,[26,30] knots[53-60] and self-avoidance[61-64] of strands. When $L \lesssim \chi^{-2/3}$, strands are separated everywhere by a distance of less than one site and the loop entropy is then negligible and $b \simeq 0$. When $\chi^{-2/3} \lesssim L \lesssim \chi^{-1}$, the complementary sites of strands may be essentially separated, but they are always opposite each other; the probability W of their meeting is on the order of $\chi^2 L^3$, the "loop entropy" is $\ln(\chi^2 L^3)$, so $b \approx 3$, while J is slightly renormalized. When $\chi^{-1} \lesssim L \lesssim \chi^{-2}$, the complementary sites are separated in all directions, $W \sim \chi^4 L^5$ and $b \approx 5$. When $L \gtrsim \chi^{-2}$, the motion of the strands is completely randomized and the "step" of random Brownian motion is $\sim L^{-2}$, so $W \sim \chi^{-3} L^{3/2}$ and $b \approx 1.5$. The quantity of χ is determined by $E_e \alpha \chi^2 \sim kT$, E_e being the elastic energy for the winding angle χ. The melting interval is typically a few degrees, while $T \sim 400\,°\mathrm{K}$; therefore χ is practically independent of temperature.

[36(a)] When l is very large, $T_m \approx T^{(1)}X^{(1)} + T^{(2)}X^{(2)}$, which coincides with the empirical Marmur-Doty formula.[35] This proves to have the same value of s for $h^{(1)}$ and $h^{(2)}$, as $h^{(1,2)} = s^{(1,2)}(T - T^{(1,2)})$ with different $s^{(1)}$ and $s^{(2)}$ would not imply this formula.

[37] Dividing Eq. (5) by $\frac{1}{2}\Delta h$, we can rewrite it in the form $i_1 - i_2 \equiv \Delta i = \frac{1}{2}nw - pl$. Thus parameters n, p, w, l determine i_1 and i_2 of the melting domain:

$$i_1 = \tfrac{1}{2}(l + \Delta i) = \tfrac{1}{4}nw + \tfrac{1}{2}(1-p)l\,,$$

$$i_2 = \tfrac{1}{2}(l - \Delta i) = -\tfrac{1}{4}nw + \tfrac{1}{2}(1+p)l\,.$$

[38] A. Vilenkin, Biopolymers 16, 1657 (1977).

[39] Note that T_m determined by Eq. (18) differs from T_m of Eq. (6) owing to the contribution of the loop-entropy term. The shift of the domain boundaries B and E to B_1 and E_1 increases the energy, so $|L(BB_1)|$, $|L(EE_1)| \ll L(BE)$; we neglect terms on the order of $bT[L(B_1B) + L(EE_1)]^2/L^2(BE)$ in $\epsilon_c(BB_1)$ and $\epsilon_c(EE_1)$.

[40] Note that the contribution to ϵ_c of the loop-entropy term in this case is different from that of Eq. (18). According to Eq. (16)

$$\epsilon_c = -h^{(1)}i^{(1)} - h^{(2)}i^{(2)} - J - bT \times \ln[L_1 L_2 \chi^2/(L_1 + L_2 + l)] \equiv s(T_m - T)\,,$$

where L_1 and L_2 refer to the rigid domains and $i^{(1)}$ $i^{(2)}$, and $l = i^{(1)} + i^{(2)}$ to the flexible one. Since T_m depends on the domains adjacent to a flexible domain, two flexible domains, bordering on the same melted rigid domain, do not provide additive free-energy contributions. To account for this fact we shall denote as "adjacent flexible domains" those domains which are adjacent either to each other, or on the same rigid melted domain.

[41] To elucidate the physical meaning of, e.g., Eq. (35), let us consider the case of $|h| \ll T$, when Eq. (35) can be written in the form

$$|\phi| \sim 1/\overline{L} \sim \overline{l}\exp[-(J + bT\ln\overline{L})/T]\,,$$

$$\overline{l} \sim T/(h + |\phi|)\,.$$

This represents the proportionality of $|\phi|$ to the relative number of phase boundaries ($\overline{L}$ being the characteristic length of long melted segments in the basically melted state) which is on the order of the probability of a helix segment. The latter is determined by the excitation energy $J + bT\ln\overline{L}$ and the number of such segments, i.e., the number $\overline{l}$ of their possible lengths [from 1 to $\sim\overline{l}$, $\overline{l}$ being the characteristic length of a helix segment whose excitation energy increase $\overline{l}(h + |\phi|)$ with $\overline{l}$ should be on the order of T].

[42] Obviously, it is no problem to account for the shifts of the helix domain boundaries, they will also slightly renormalize the transition value of $\tilde{h}$.

[43] In the case, for instance, of a random sequence, the characteristic length of a domain is proportional to[29,38] $(J/T)^2 \sim s^2 \sim 100$. The relative half-width of the peak is $\sim 1/sl \sim 1/s^3$, while the melting interval ΔT_m, by Eq. (14c), is $\Delta T_m \sim (T^{(2)} - T^{(1)})/s \sim 1/s^2$ (as, numerically, $T^{(2)} - T^{(1)} \sim T/s$) so the number of peaks is $\sim s$. Thus the total length of DNA with clearly separated peaks is $\sim s^2 s \sim s^3 \sim 10^3$. Therefore DNA is "short" if it contains several thousand sites or less.

[44] S. Z. Hirschman and G. Felsenfeld, J. Mol. Biol. 16, 347 (1966).

[45] S. Z. Hirschman, M. Gellert, S. Falkow, and G. Felsenfeld, J. Mol. Biol. 28, 469 (1967).

[46] This may not be the case for even the longest DNA's, such as mammalian ones. Recent experiments [see, e.g., Refs. 47–48(a)] indicate that a considerable part of these DNA's may consist of repetitive segments which will provide clearly distinguished peaks in dN_c/dT. This may be the origin of multiple oscillations, observed in the quoted experiments for DNA's containing $\sim 10^5$ sites. The proposed analysis of sequence statistical structure may be the quickest and simplest one.

[47] R. J. Britten and D. E. Kohne, Science 161, 529 (1968).

[48] M. Waring and R. J. Britten, Science 154, 791 (1966).

[48(a)] H. Rosenberg, M. Singer, and M. Rosenberg, Science 200, 394 (1978).

[49] The relative accuracy of this determination, indicated in Sec. II, is related to a very fast decrease of the kernel of the integral equation, obtained from Eqs. (26) and (50).

[50] Yu. S. Lazurkin *et al.*, Biopolymers 14, 1551 (1974).

[51] J. R. Fresco, L. C. Klotz, and E. G. Richards, Cold Spring Harbor Symp. Quant. Biol. 28, 83 (1963).

[52] G. Felsenfeld and S. Z. Hirschman, J. Mol. Biol. 13, 407 (1965).

[53] M. Delbrück, in *Mathematical Problems in the Biological Sciences*, edited by R. E. Bellmann, Proc. Symp. Appl. Math. 14, 55 (1962).

[54] H. L. Frisch and S. Prager, J. Chem. Phys. 46, 1475 (1967).

[55] S. F. Edwards, Proc. Phys. Soc. London 91, 513 (1967).

[56] R. Alexander-Katz and S. F. Edwards, J. Phys. A 5, 674 (1972).

[57] N. Saito and Y. Chen, J. Chem. Phys. 59, 3701 (1973).

[58] S. F. Edwards and K. F. Freed, J. Phys. C 3, 739, 750, 760 (1970).

[59] S. F. Edwards, J. Phys. A 1, 15 (1968).

[60] S. F. Edwards and J. W. Kerr, J. Phys. C 5, 2289 (1972).

[61] P. G. de Gennes, P. Pincus, and R. M. Velasco, J. Phys. (Paris) 37, 1461 (1976).

[62] P. G. de Gennes, Riv. Nuovo Cimento 7, 363 (1977).

[63] P. G. de Gennes, Polym. Lett. Ed. 15, 623 (1977).

[64] P. Pincus, Macromolecules 9, 386 (1976).

[65] M. Azbel, Proc. Nat. Acad. Sci. USA 76, 101 (1979).