LETTER TO THE EDITOR

Phase transition in DNA

M Ya Azbel†

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

Abstract. DNA solution is described by a one-dimensional Ising model with a long-range poly-spin interaction in an inhomogeneous 'magnetic field'. The model is used to calculate DNA thermodynamic properties, e.g. heat capacity and optical density. They exhibit oscillations which can be used to obtain important information about the DNA component sequence. This is demonstrated by an example of a particular DNA. The long-range interaction yields a phase transition. Its nature is shown to depend crucially on well defined properties of the DNA sequence.

DNA is a long double-strand molecule which consists of two types of molecule pairs ('components'). When the temperature T is increased, pairs may unbind, thus forming a 'coiled' or 'melted' state (see figure 1). This is known as a helix–coil transition. The light absorption in the region of 2600 Å is different for bound ('helix') and melted pairs. Therefore the number N_c of melted sites is proportional (Vedenov *et al* 1972) to the optical density of DNA solution and may be directly measured (Vedenov *et al* 1972, Poland and Scheraga 1970, Wartell and Montroll 1972, Lazurkin *et al* 1970).
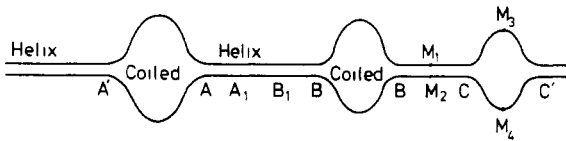
Figure 1. DNA melting: M_1M_2 and M_3M_4 are bound ('helix') and unbound ('melted' or 'coiled') molecule pairs respectively.

The helix–coil transition in DNA is described (Vedenov *et al* 1972, Poland and Scheraga 1970, Wartell and Montroll 1972, Lazurkin *et al* 1970, Azbel 1975, 1978) by a one-dimensional Ising model with long-range poly-spin interactions in an inhomogeneous 'magnetic field'. The variation of this field is related to the sequence of the DNA components, which contains specific genetic information. The sequence cannot be subject to any simplifying assumptions, and this makes DNA a very unusual physical system.

In this letter we use this Hamiltonian, including for the first time *the arbitrary component sequence and the long-range interaction*, to calculate analytically DNA thermodynamic properties, e.g. the number of melted sites. The latter coincides very accurately with experimental data (Wada *et al* 1977, Lyubchenko *et al* 1977) for a

† Visitor at The Institute for Advanced Study, Princeton, New Jersey 08540, USA.

particular DNA (of phage ϕX-174). We also show that a detailed inspection of the DNA melting curve provides important information about DNA, which is easy to obtain and analyse and which may complement the usual (Sanger and Coulson 1975, Maxam and Gilbert 1977, Sanger *et al* 1977a, b) biochemical methods of DNA sequencing. (Up to now only four sequences have been completely sequenced (Fiers *et al* 1976, 1978, Sanger *et al* 1977a, b, Reddy *et al* 1978, Beck *et al* 1979).) The proposed procedure is demonstrated for an example of DNA for which both the melting curves (Wada *et al* 1977, Lyubchenko *et al* 1977) and the sequence (Sanger *et al* 1977a, b) are known.

The long-range interaction leads to a specific phase transition (we shall discuss its specificity later) in such one-dimensional systems. The nature of the phase transition is analysed and is found to depend crucially on certain well defined properties of the sequence of 'magnetic field' variables. It may exhibit various kinds of singularities, ranging from an essential singularity for quasi-random sequences to a first-order transition for quasi-periodic ones. Accurate experiments on these phase transitions are practically absent, and it is hoped that the theory may stimulate them.

The DNA effective Hamiltonian H, measured from the completely bound state, is given by (Vedenov *et al* 1972, Poland and Scheraga 1970, Wartell and Montroll 1972, Lazurkin *et al* 1970, Azbel 1975, 1978)

$$H = -\sum_r h_r(\tfrac{1}{2}+s_r) + J \sum_r (\tfrac{1}{4}-s_r s_{r+1})$$
$$+ \sum_{r,L} (\tfrac{1}{2}-s_r)(\tfrac{1}{2}+s_{r+1})(\tfrac{1}{2}+s_{r+2})\ldots(\tfrac{1}{2}+s_{r+L})(\tfrac{1}{2}-s_{r+L+1})bkT \ln (L\chi^2). \quad (1)$$

Spins 'down' ($s_r = -\tfrac{1}{2}$) and 'up' ($s_r = +\tfrac{1}{2}$) label unbound and bound states respectively.

The magnetic field h_r is related to the component j_r at the rth site ($j_r = 1$ or 2) by $h_r = S(T - T^{(j_r)})$, and depends on temperature.† Typically (Vedenov *et al* 1972) $T^{(2)} - T^{(1)} \sim 40$ K, $T^{(2)} \sim 400$ K, $S \sim 10k$, k being the Boltzmann constant. The effective Hamiltonian depends on temperature, for it is in fact the free energy for a given set $\{s_r\}$ of spins; the summation over all other degrees of freedom has already been performed. These degrees of freedom include, in particular, loop degrees of freedom of unbound portions (see figure 1) which are responsible for the last sum in the Hamiltonian (1); χ accounts (Azbel 1978) for the elasticity of the strands, and is a characteristic winding angle between adjacent sites. According to Vedenov *et al* (1972) $\chi \sim 0.1$, $b \sim 1$.

The coupling energy (Vedenov *et al* 1972) $J \sim ST \gg kT$. Therefore the effective temperature is low, and the free energy F differs only slightly from the ground-state energy $E = \min H$ owing to rare low-energy excitations. We start with the ground state and with $b = 0$. In the temperature interval $T^{(1)} < T < T^{(2)}$, the magnetic field has opposite signs for the two components: $h_r > 0$ for $j_r = 1$; $h_r < 0$ for $j_r = 2$. Thus the ground state consists of 'spin up' and 'spin down' domains. These domains were precisely determined by Azbel (1973a, b) and Lifshiftz (1974). When the temperature increases, h_1 increases, $|h_2|$ decreases and the average field h increases. Therefore certain domains (Azbel 1973a, b) change their magnetisation from 'spin down' to 'spin up' (we shall denote this process as 'melting'). Suppose, for instance, that in equation (1) $b = 0$, $h_1 = 1$, $h_2 = -1$, $J = 4$ and the component sequence is 111111222221111111. By equation (1), when all spins are 'up', the five 2's contribute an energy $-5h_2 = 5$; when

---

† The same value of S for any h_r follows from the comparison of theoretical (Vedenov and Dykhne 1968) and experimental (Marmur and Doty 1962) formulae for the DNA melting temperature.

the 2's have spins 'down', two domain boundaries are created, contributing $J = 4$ to the energy. Thus minimal H is achieved by three domains: spins 'up' at 1's, spins 'down' at 2's. However, when $h_1 = \frac{3}{2}$ and $h_2 = -\frac{1}{2}$, then $-5h_2 < J$, and all the spins are 'up' in the ground state.

Obviously, all the 2's melt when $h_2 = -\frac{1}{5}J = -\frac{4}{5}$. In a general case, a domain which contains $i_m^{(1)}$ 'unities' and $i_m^{(2)}$ 'twos', and which creates while melting n_m new domain boundaries (in our example $i_m^{(1)} = 0$, $i_m^{(2)} = 5$, $n_m = -2$), melts when $-h_1 i_m^{(1)} - h_2 i_m^{(2)} + \frac{1}{2}Jn_m = 0$. Thus the melting temperature T_m is

$$T_m = T^{(1)}x_m^{(1)} + T^{(2)}x_m^{(2)} + T_b n_m / l_m$$

$$x_m^{(1,2)} = i_m^{(1,2)}/l_m; \qquad l_m = i_m^{(1)} + i_m^{(2)}; \qquad T_b = \tfrac{1}{2}J/S. \qquad (2)$$

When $b \neq 0$, melted domain energies may be non-additive† (domains 'interact'). However, the non-additivity is always relatively small compared with J: when $L \sim 1$, this is due to $kT \ll J$, while when L is large it is due to the slow change† of $\ln L$ with L. Such a non-additivity can be accounted for as a perturbation. Thus the ground state for $b \neq 0$ can be determined, according to Azbel (1973a, b) and Lifshitz (1974). A ground-state domain melts when its melting energy ϵ_c equals zero. The melting energy may depend on the adjacent domains. For instance, if a domain has $l_m = i_m^{(1)} + i_m^{(2)}$ sites, and melts between 'spin up' domains containing L_1 and L_2 sites (which are not at the edges of the DNA), then

$$\epsilon_c = -h_1 i_m^{(1)} - h_2 i_m^{(2)} - J - bkT \ln[\chi^2 L_1 L_2 /(L_1 + L_2 + l_m)]. \qquad (2a)$$

Now we consider low-temperature excitations. There are two types of low-energy excitations: either a boundary between two domains moves, or a full domain, with small $|\epsilon_c|$ (the domain is close to melting already in the ground state), undergoes melting.

Suppose we monotonically increase the temperature, determine as by Azbel (1973a, b) successive melting domains, and denote them as 'flexible' in the vicinity of their melting (where $|\epsilon_c| \lesssim kT$) and as 'rigid' outside it (where $|\epsilon_c| \gg kT$).

A flexible domain, situated between rigid 'spin down' domains, contributes, by equation (1), zero to the energy when it is 'spin down', ϵ_c when it is 'spin up', and inputs additional energies ϵ_1, ϵ_2 when its first and second boundaries are shifted. The corresponding contribution ΔF^+ to the free energy F equals

$$\Delta F^+ = -kT \ln (1 + Q_1 Q_2 t); \qquad Q_{1,2} = \sum \exp (-\epsilon_{1,2}/kT); \qquad t = \exp (-\epsilon_c/kT). \qquad (3a)$$

The summation refers to possible boundary shifts; superscript ' + ' indicates that $n_m > 0$ (here $n_m = 2$).

Suppose a flexible domain is situated between rigid 'spin up' domains. Then the reasoning is similar, but we have to take into account that rigid 'spin up' domains are those domains which melted (and therefore were flexible) at lower temperatures. Thus their boundary shift contribution $-kT \ln (Q_1 Q_2)$ (see equation (3a) with $-\epsilon_c \gg kT$) has already been taken into account and should not be considered twice. The contribution ΔF^- to F is thus

$$\Delta F^- = -kT \ln (Q_1 Q_2 + t) + kT \ln (Q_1 Q_2). \qquad (3b)$$

† For instance, suppose in figure 1 consequent domains A'A, AB, BB', B'C, CC' contain respectively L, l, L, l, L sites, $l \ll L$. Then, by equation (1), the binding of a single AB or B'C at the completely melted A'C changes the winding energy by $\epsilon_1 \simeq |b_1 \ln L + b_1 \ln (2L)| - b_1 \ln (3L) = b_1 \ln (\frac{2}{3}L)$, $b_1 = bkT$. The simultaneous binding of AB and B'C at the completely melted A'C changes the winding energy by $\epsilon_2 \simeq 3b_1 \ln L - b_1 \ln (3L) = b_1 \ln (\frac{1}{3}L^2)$. The 'interaction energy' is $\delta\epsilon = 2\epsilon_1 - \epsilon_2 \simeq bkT \ln \frac{4}{3} \ll J$.

Finally, if a flexible domain borders rigid 'spin up' and 'spin down' domains, then

$$\Delta F^0 = -kT \ln (Q_1 + Q_2 t) + kT \ln Q_1 \tag{3c}$$

where Q_1 relates to the boundary with the 'spin up' domain.

A flexible domain, which borders the DNA edge, is accounted for in the same way; since it has only one 'spin up'–'spin down' boundary, only one Q enters the corresponding formula; the other Q should be replaced by 1.

In the approximation of non-adjacent flexible domains, F equals the sum of contributions (3a, b, c) from all consequently melting domains. The higher approximations, which account for, e.g., the adjacency of flexible domains, are given analogous consideration.

Note that the calculation of F is not based on perturbation theory. The ground state and its excitations (in particular, their locations) depend on the $\{h_m\}$ sequence. The excitations are separated by distances large enough to prevent their interaction. When $b > 1$, F exhibits a phase transition† to 'all spins up'. The transition singularity is obviously determined by the singularity in the concentration c of ground-state 'spins down' at the temperature T_g, where the ground state becomes one 'spin up' domain. The analysis of the ground state shows that this singularity may vary from an essential singularity for a random or quasi-random $\{h_m\}$ sequence, $-\ln|F| \propto (T_g - T)^{-1}$, to a first-order phase transition for a periodic or quasi-periodic $\{h_m\}$ sequence with $b > 2$ (see also Poland and Scheraga 1966, Applequist 1969, Azbel 1975). The nature of the phase transition is rather specific. It is related to the change in the temperature-dependent ground state rather than to fluctuations. For instance, when $c \to 0$, and thus the characteristic length L^* of melted domains goes to infinity together with the 'binding energy' $(-\epsilon_c)$ from equation (2a), the relative contribution of fluctuations from F approaches *zero*. Such specific phase transitions may be typical in systems where the higher energy level has higher degeneracy‡, so the 'magnetic field' in the effective Hamiltonian depends on the temperature, and the transition is related to the 'remagnetisation' of the ground state.

Outside the immediate vicinity of the phase transition, where $kT \ln (L^* \chi^2) \ll J$, the formula for F provides the following equation§ for dM/dT, M being the magnetisation:

$$dM/dT = \sum_m \tfrac{1}{2}\beta l_m^2 / \cosh^2[\beta l_m (T - T_m)], \qquad \beta = \tfrac{1}{2}S/kT \tag{4}$$

with T_m from equation (2) for the ground-state domain, which is the mth one among successively melting (when temperature monotonically increases) domains.

By equation (4), dM/dT exhibits oscillations related to the successive melting of ground-state domains (see also Azbel 1972, 1973a, b). These oscillations have been observed in numerous experiments (Steinert and van Assel 1974, Poland 1974, Yubuki *et al* 1975, Vizard and Ansevin 1976, Lyubchenko *et al* 1976, Ansevin *et al* 1976, Gotoh *et al* 1976, Reiss and Arpa-Gabarro 1977, Wada *et al* 1976, Akiyama *et al* 1977). Fitting equations (4, 2) to experimental data, we can determine l_m, x_m and n_m for the corresponding domains. An example of such a determination and its accuracy is demonstrated in table 1. If a DNA molecule (and therefore certain domains in it) is fragmentised, the identification of the fragmentised domains (from DNA melting curves

---

† Of course, only in the case of an infinite system.

‡ And therefore $(\epsilon_1 - \epsilon_2) - kT \ln (v_1/v_2)$ changes its sign when the temperature increases (ϵ_1 and ϵ_2 are energy levels; v_1 and v_2 are their degeneracies).

§ $dM/dT = dN_c/dT$ is the experimentally measured quantity (N_c is the number of melted sites).

Table 1. Domain melting temperatures T_m, lengths l_m and first-component concentrations x_m, according to the present analysis of the experimental data of Wada *et al* (1977) for the DNA fragment (the fragment Y_1 there, containing 2745 sites) of phage ϕX-174. Superscript 'E' indicates corresponding data for the known (Sanger *et al* 1977a, b) 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 $\beta = 0 \cdot 0106$ deg^{-1} and $T_b = 100$ K; they are the same for all three experimental plots of Wada *et al* (1977) and Lyubchenko *et al* (1977). Parameters $T^{(1)} = 52 \cdot 5$ K, $T^{(2)} = 94 \cdot 9$ K are taken from Lazurkin *et al* (1974). The experimental plot, which is analysed in the table, is presented in figure 2.

T_m	l_m	l_m^E	x_m	x_m^E	Number of domain
70·5	233	239	0·585	$0 \cdot 582^{+0 \cdot 01}_{-0 \cdot 1}$	I
70·87	641	671	0·570	$0 \cdot 566 \pm 0 \cdot 002$	VII
71·2	468	457	0·559	$0 \cdot 558 \pm 0 \cdot 002$	VI
71·5	215	220	0·574	$0 \cdot 591 \pm 0 \cdot 005$	III
72·0	333	333	0·526	$0 \cdot 571^{+0 \cdot 01}_{-0 \cdot 05}$	II
72·44	481	490	0·530	$0 \cdot 559 \pm 0 \cdot 015$	IV
73·5	374	335	0·497	$0 \cdot 501 \pm 0 \cdot 003$	V
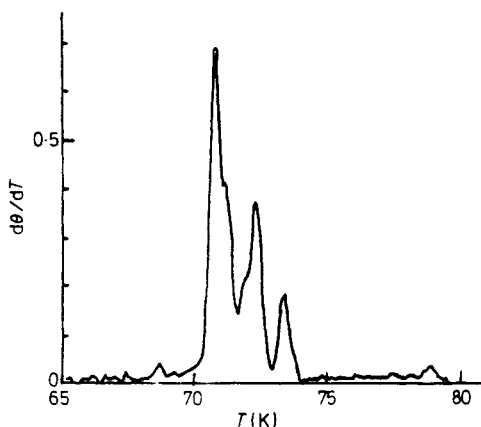
Figure 2. The experimental differential melting curve (Wada *et al* 1977) for the fragment of the DNA of phage ϕX-174 (the fragment Y_1). The quantity θ is the relative number of melted sites. For Hamiltonian (1), $d\theta/dT = N^{-1} dM/dT$, where M is the magnetisation and N is the total number of sites.

before and after the fragmentation) allows us to find out the order of DNA fragments, which is an important and troublesome part of the usual DNA sequencing (Sanger and Coulson 1975, Maxam and Gilbert 1977, Fiers *et al* 1976, 1978, Sanger *et al* 1977a, b, Reddy *et al* 1978, Beck *et al* 1979). All our considerations are based on a large value of *J*. They are readily generalised to any refinement of the Hamiltonian (1) (e.g. to that accounting for the interaction of DNA adjacent sites), and have a rather general character.

In conclusion, the study of the transitions in the Ising model studied above has important consequences both with respect to the solution of the 'inverse' thermodynamic problem (learning about the sequence of the 'fields' from the experiments)

and to our understanding of random systems (the nature of the transition depends on the nature of the randomness).

References

Akiyama C *et al* 1977 *Biopolymers* **16** 427–35
Ansevin A T *et al* 1976 *Biopolymers* **15** 155–74
Applequist J 1966 *J. Chem. Phys.* **45** 3459–61
—— 1969 *J. Chem. Phys.* **50** 600–12
Azbel M Ya 1972 *JETP Lett.* **16** 128–31
—— 1973a *Biopolymers* **12** 1592–609
—— 1973b *Phys. Rev. Lett.* **31** 589–93
—— 1975 *J. Chem. Phys.* **62** 3635–47
—— 1978 *Ann. Isr. Phys. Soc.* Vol. 2(1) *Statist. Phys.–STATPHYS.* 13 (Haifa and Bristol: The Israel Physical Society and Adam Hilger) pp 25–36
Beck E *et al* 1979 *Preprint*
Fiers W *et al* 1976 *Nature* **260** 500–7
—— 1978 *Nature* **273** 113–20
Gotoh O *et al* 1976 *Biopolymers* **15** 655–70
Lazurkin Yu S *et al* 1970 *Biopolymers* **9** 1255–306
—— 1974 *Biopolymers* **14** 1551–62
Lifshitz I M 1974 *Sov. Phys.–JETP* **38** 545–9
Lyubchenko Yu L *et al* 1976 *Biopolymers* **15** 1019–36
—— 1977 *Nature* **271** 28–31
Marmur J and Doty P 1962 *J. Molec. Biol.* **5** 109–15
Maxam A and Gilbert W 1977 *Proc. Natl Acad. Sci. USA* **74** 560–4
Poland D 1974 *Biopolymers* **13** 1859–71
Poland D and Scheraga H A 1966 *J. Chem. Phys.* **45** 1456–69
—— 1970 *Theory of Helix–Coil Transition in Biopolymers* (New York: Academic)
Reddy V B *et al* 1978 *Science, NY* **200** 494–502
Reiss C and Arpa-Gabarro T 1977 *Prog. Molec. Subcoll. Biol.* **5** 1–30
Sanger F and Coulson A R 1975 *J. Molec. Biol.* **94** 441–8
Sanger F *et al* 1977a *Nature* **265** 687–95
—— 1977b *Proc. Natl Acad. Sci. USA* **74** 5463–7
Steinert M and van Assel S 1974 *Biochem. Biophys. Res. Commun.* **61** 1249–55
Vedenov A A and Dykhne A M 1968 *Zh. Eksp. Teor. Fiz.* **55** 357–62
Vedenov A A *et al* 1972 *Sov. Phys.–Usp.* **14** 715–36
Vizard D L and Ansevin A T 1976 *Biochemistry* **15** 741–50
Wada A *et al* 1976 *Nature* **263** 439–40
—— 1977 *Nature* **269** 353–5
Wartell R W and Montroll E W 1972 *Adv. Chem. Phys.* **22** 129–203
Yubuki S *et al* 1975 *Biochem. Biophys. Acta* **395** 258–73