Multiple Helix Model = 1d Ising Model

Zimm - Bragg

c c c h h c c c h h h c

length N

2^N configs

h_0 ↑ ↓ ↓ ↓ ↑ ↑ ↓ ↓ ↓ ↑ ↑ ↑ ↓

helix model $$Z_N = \sum_{\text{configs}} \sigma^g s^n$$

where n = number of helical residues

g = number of helices

Ising model $$\mathcal{H} = -h_0 \sum_i S_i - J \sum_i S_i S_{i+1}$$

h_0 ↑ applied field

J ↑ coupling strength

$$s \equiv e^{2h_0/kT}$$

$$\sigma = e^{-4J/kT}$$

Let $Z_N = Z_N^c + Z_N^h$

Z_N ↗ P.f. for chain of length N

Z_N^c ↑ P.f. constrained so that last unit is a c

Z_N^h ↖ p.f. constrained so that last unit is h

recursion relations

$$xxxxc = xxxcc + xxxhc$$

$$Z_N^c = Z_{N-1}^c + Z_{N-1}^h$$

$$xxxxh = xxxch + xxxhh$$

$$Z_N^h = Z_{N-1}^c \sigma s + Z_{N-1}^h s$$

$$\begin{pmatrix} Z_N^c \\ Z_N^h \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \sigma s & s \end{pmatrix} \begin{pmatrix} Z_{N-1}^c \\ Z_{N-1}^h \end{pmatrix}$$

$$\begin{pmatrix} Z_N^c \\ Z_N^h \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \sigma s & s \end{pmatrix} \begin{pmatrix} Z_{N-1}^c \\ Z_{N-1}^h \end{pmatrix}$$

begin with $\begin{pmatrix} Z_0^c \\ Z_0^h \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ initial conditions

$$\therefore \begin{pmatrix} Z_1^c \\ Z_1^h \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \sigma s & s \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ \sigma s \end{pmatrix}$$

$Z_1 = 1 + \sigma s$ ✓
$\quad\;\; c \quad h$

$$\begin{pmatrix} Z_2^c \\ Z_2^h \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \sigma s & s \end{pmatrix} \begin{pmatrix} 1 \\ \sigma s \end{pmatrix} = \begin{pmatrix} 1+\sigma s \\ \sigma s + \sigma s^2 \end{pmatrix}$$

$Z_2 = 1 + 2\sigma s + \sigma s^2$ ✓
$\quad\;\; cc \quad ch \quad hh$
$\qquad\qquad\;\; hc$

$$\begin{pmatrix} Z_N^c \\ Z_N^h \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \sigma s & s \end{pmatrix}^N \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\underset{\sim}{Z}_N = M^N \underset{\sim}{Z}_0$$

Eigenvectors / Eigenvalues

$$\lambda_1 \underset{\sim}{v}_1 = M \underset{\sim}{v}_1 \qquad \lambda_2 \underset{\sim}{v}_2 = M \underset{\sim}{v}_2$$

$$\underset{\sim}{Z}_0 = a \underset{\sim}{v}_1 + b \underset{\sim}{v}_2$$

$$\underset{\sim}{Z}_N = a \lambda_1^N \underset{\sim}{v}_1 + b \lambda_2^N \underset{\sim}{v}_2$$

$$Z_N = A\lambda_1^N + B\lambda_2^N$$

For large N $\quad Z_N \sim \lambda_1^N$ (largest eigenvalue)

Only interested in extensive quantities

(things proportional to N)

e.g. $F = -kT \ln Z_N$

$= -NkT \ln \lambda_1 +$ small terms

Find λ_1 : $\begin{vmatrix} 1-\lambda & 1 \\ \sigma s & s-\lambda \end{vmatrix} = 0.$

$$(1-\lambda)(s-\lambda) - \sigma s = 0.$$

$$\lambda^2 - (1+s)\lambda - \sigma s + s = 0.$$

$$\lambda = \frac{1+s \pm \sqrt{(1+s)^2 - 4s(1-\sigma)}}{2}.$$

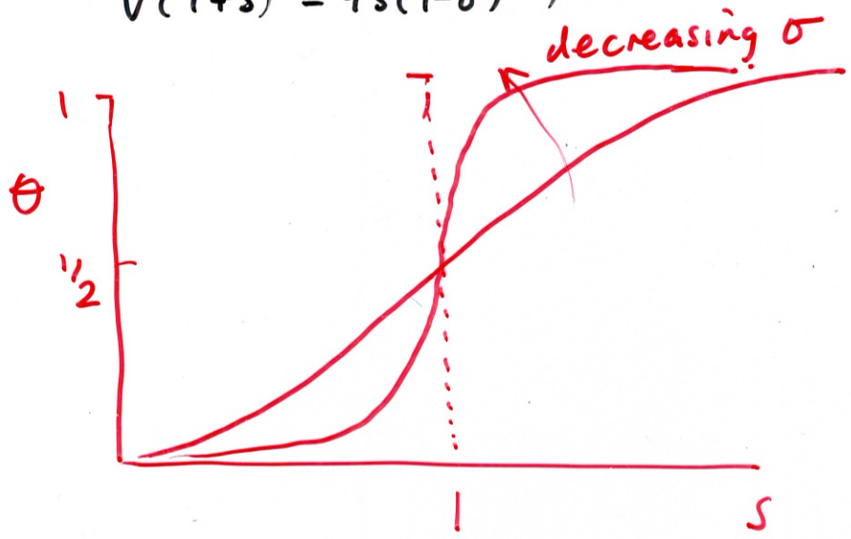
Now $Z_N = \sum_{n=0}^{N} f_n(\sigma)\, s^n$ for some $f_n(\sigma)$

$$\theta = \sum_n \frac{n}{N} \frac{f_n(\sigma)\, s^n}{Z} = \frac{1}{NZ}\, s \frac{\partial Z}{\partial s} = \frac{s}{N} \frac{\partial \ln Z}{\partial s}$$

(as for zipper model)

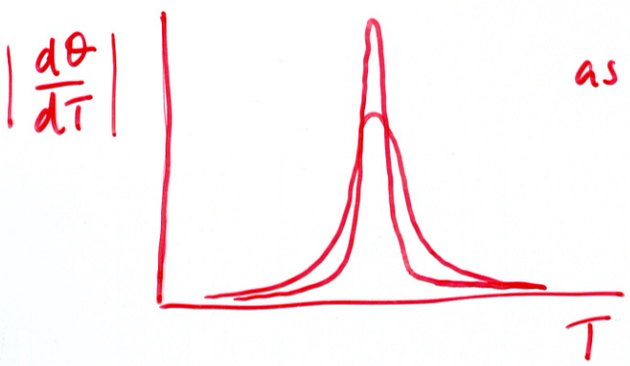
$$= s \frac{\partial \ln \lambda_1}{\partial s} = \boxed{\frac{s}{\lambda_1} \frac{\partial \lambda_1}{\partial s}}$$

$$\theta = \frac{s}{2\lambda_1}\left(1 + \frac{(s-1)+2\sigma s}{\sqrt{(1+s)^2 - 4s(1-\sigma)}}\right)$$



σ controls sharpness of "transition"

Not a real phase transition because curve is smooth even for $N \to \infty$

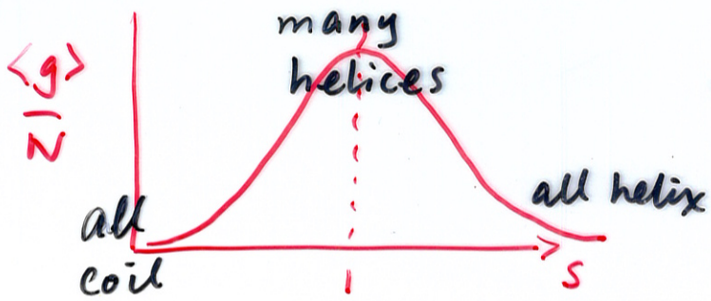


as usual

Can also write $Z_N = \sum_{g=0}^{N} f_g(s)\,\sigma^g$

$\therefore$ mean number of helices per unit length $\therefore$ $= \sum_g \frac{g}{N} \frac{f_g(s)\,\sigma^g}{Z}$

$= \frac{\sigma}{\lambda_1}\frac{\partial \lambda_1}{\partial \sigma}$



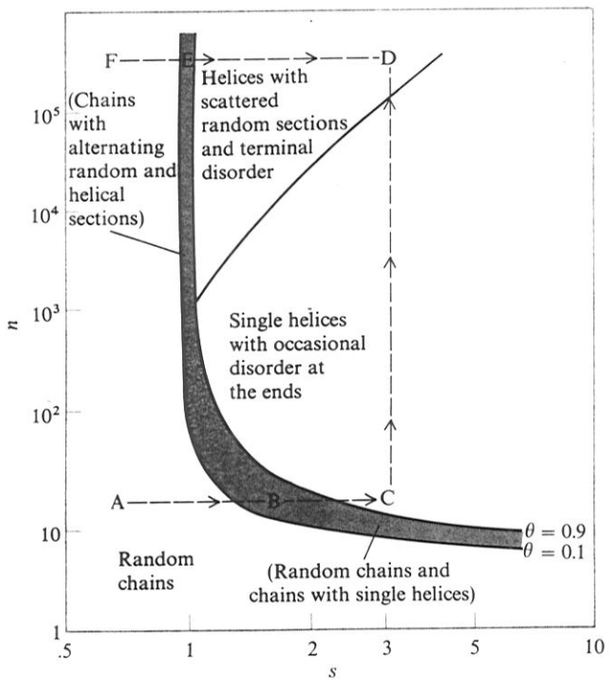


Figure 20-12

The n–s plane for $\sigma = 10^{-4}$. The contours of $\theta = 0.9$ and $\theta = 0.1$ have been chosen arbitrarily as the boundaries of the transition. Along the solid diagonal line, half of the chains contain a single unbroken helical section. [After B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* 31:526 (1959).]

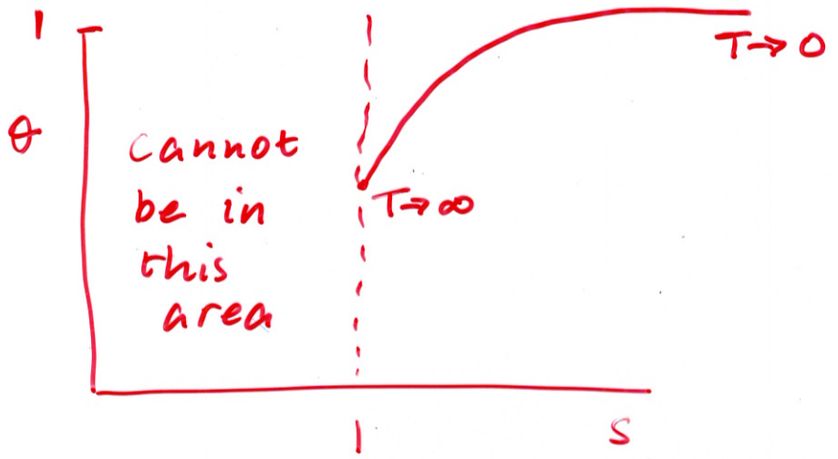
dominate under these conditions. If we now pass from the helical to the coil state at large n (along D → E → F), we move through a transition region in which the chains have alternating random and helical sections. Thus, the microscopic species present in the transition vary considerably with n.

The treatment in Section 20-7 is inaccurate on one point. In actuality, a helix sequence starts by ordering three residues simultaneously, so that the first hydrogen bond can be formed. We ignored this fact, but it can be taken into account by increasing the dimensions of the matrix $\underset{\sim}{\mathbf{M}}$. However, the approximate model used here is quite accurate for large n. Furthermore, the accuracy improves around $\theta = 1/2$. Therefore, in many practical situations, there is no need to consider higher-order matrices.

1 d Ising model

$s = e^{\frac{2h_0}{kT}}$

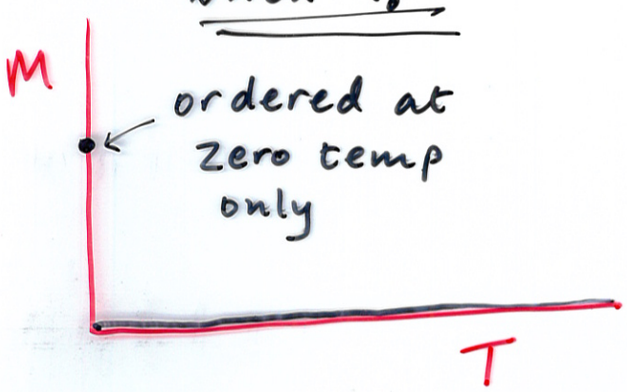
θ = fraction of up-spins

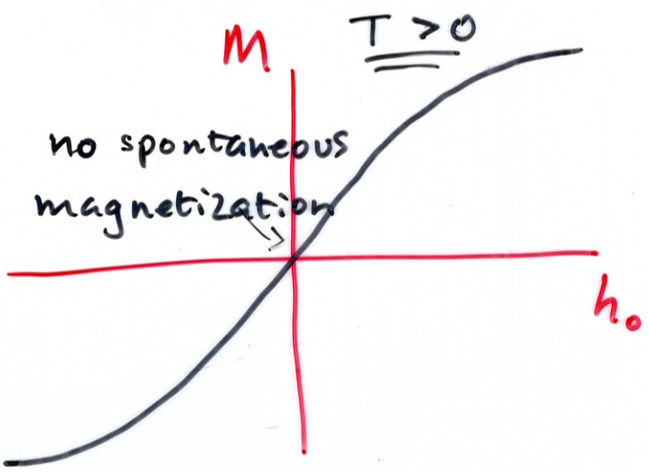


For helix model, $s = e^{\frac{\Delta H - T\Delta S}{kT}}$ ← This changes sign

$s < 1$ for $T < T_m$

Magnetization $M = \theta - (1-\theta) = 2\theta - 1$



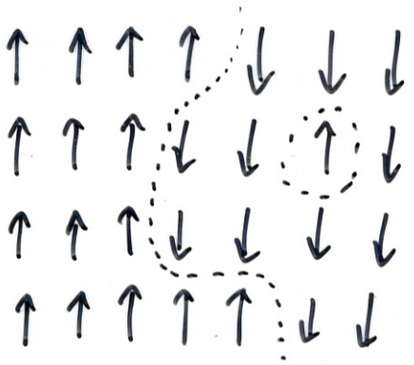


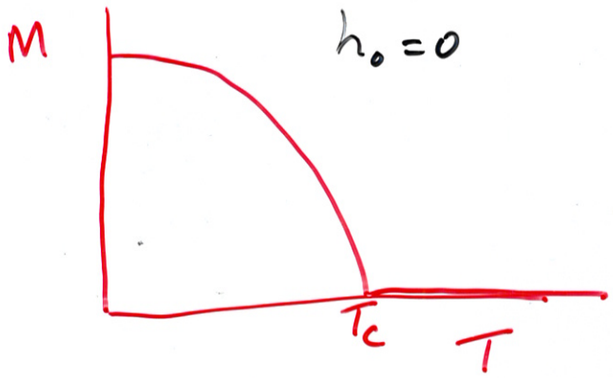
↑ ↑ ↑ ↑ ↑ ↑ ↓ ↓ ↓ ↓ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↓

ξ

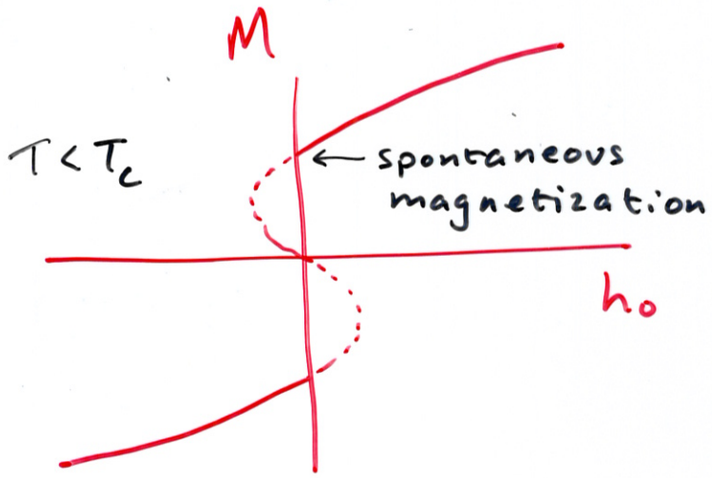
correlation length is finite for $T > 0$

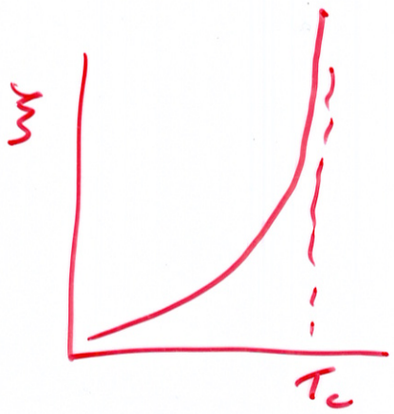
2d Ising model





ordered for $T < T_c$ even when $h_0 = 0$





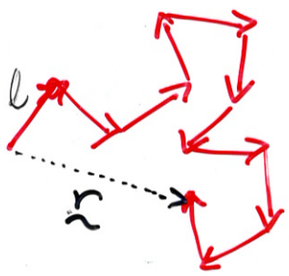
for T just below T_c

Exponents $M \sim (T_c - T)^{\beta}$

$$\xi \sim \frac{1}{(T_c - T)^{\nu}}$$

Rule : For 1d models with short-range interactions you cannot get phase transitions

Loop Entropy



random walk.

$$p(\underset{\sim}{r}) = \left(\frac{3}{2\pi n l^2}\right)^{3/2} e^{-\frac{3r^2}{2nl^2}}$$

Prob. of returning to the origin $p(0)l^3 = \left(\frac{3}{2\pi n}\right)^{3/2}$

Entropic penalty for closing loop

Total configurations $\sim z^n$

configs that close loop $\sim \frac{z^n}{n^{3/2}}$

$$\frac{\Delta S}{k} = n\ln z - \left(n\ln z - \frac{3}{2}\ln n\right)$$

$$= \frac{3}{2}\ln n$$

Self Avoiding Walk

$p(0) \sim \frac{1}{n^b}$

with $b = \frac{7}{4}$

$$\frac{\Delta S}{k} = \frac{7}{4}\ln n$$

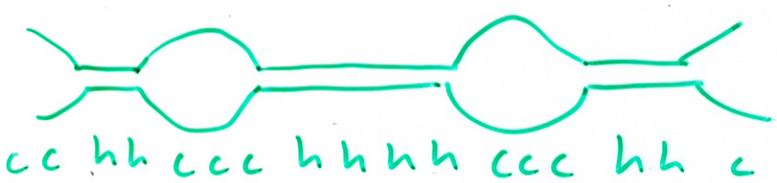
DNA melting model

Poland + Scheraga

Before : σ was penalty for ~~closing loop~~ starting new helix

$$\times\times\times\times h = \times\times\times h h + \times\times\times c h$$

$$Z_N^h = s Z_{N-1}^h + \sigma s Z_{N-1}^c$$



Now : $\frac{\sigma}{n^b}$ = penalty for starting helix and closing loop

$$\times\times\times\times\times\times h = \times\times\times\times\times hh + \times\times\times\times hch + \times\times\times hcch \ldots$$

$$Z_N^h = s Z_{N-1}^h + \frac{\sigma s}{1^b} Z_{N-2}^h + \frac{\sigma s}{2^b} Z_{N-3}^h \cdots$$

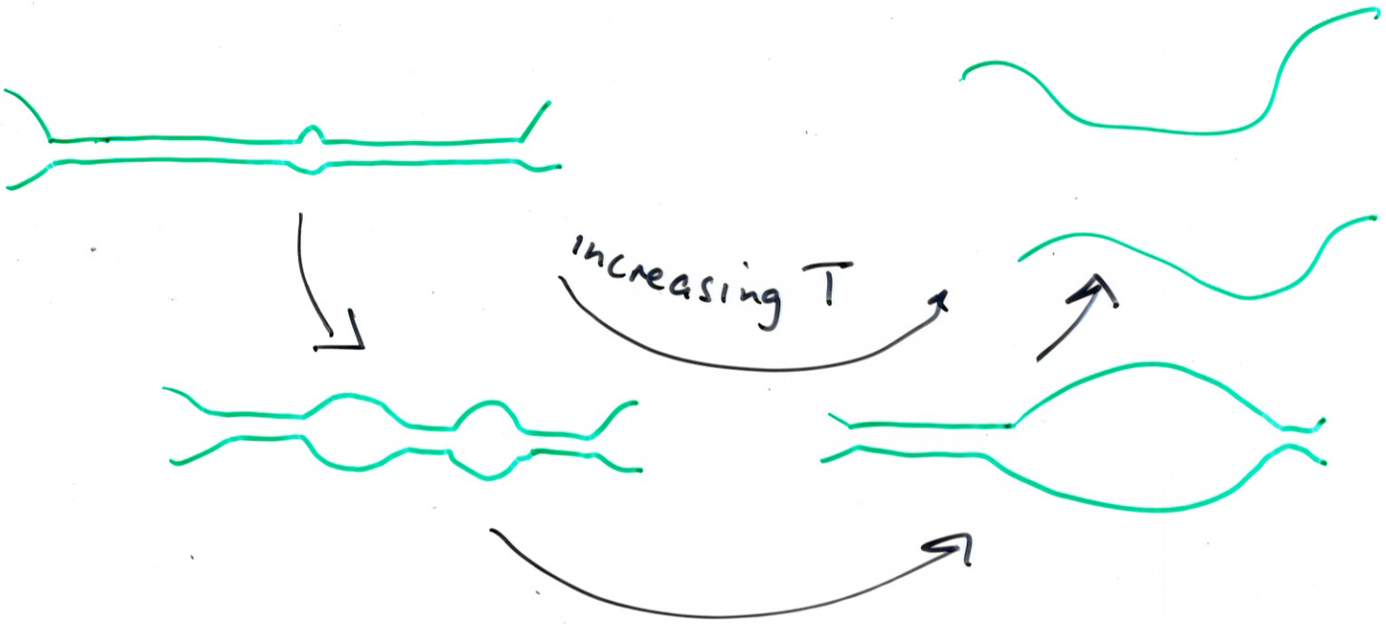
$$Z_N^h = s Z_{N-1}^h + \sigma s \sum_{\ell=2}^{N-1} \frac{Z_{N-\ell}^h}{(\ell-1)^b}$$

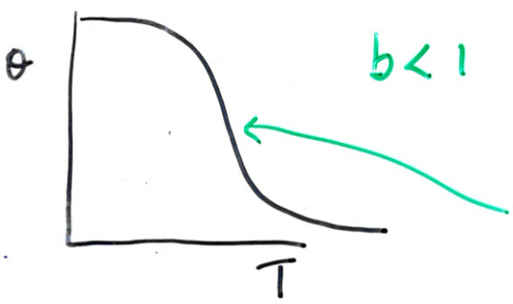
for $N >> 1$, assume $Z_N^h \sim \lambda^N$

$$\lambda^N = s \lambda^{N-1} + \sigma s \sum_{k=1}^{\infty} \frac{\lambda^{N-1-k}}{k^b}$$

$$\lambda = s + \sigma s \sum_{k=1}^{\infty} \frac{1}{\lambda^k k^b}$$

Show that for $b > 1$ there is a phase transition at finite T

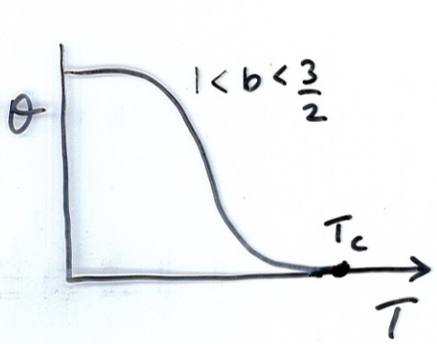


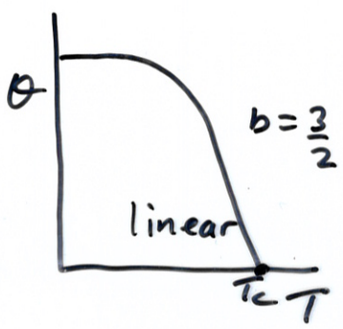


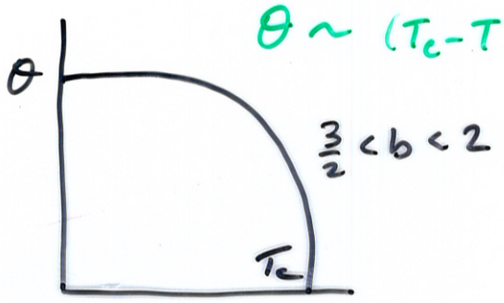
$b < 1$ no phase transition

Smooth change is what we called helix-coil "transition" so far.

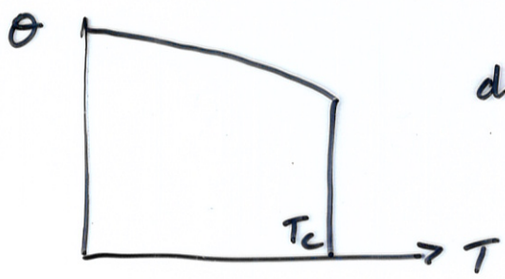
$1 < b < 2$ 2nd order phase transition $\theta \sim (T_c - T)^{\frac{2-b}{b-1}}$







$b > 2$: 1st order transition.



discontinuous change at T_c

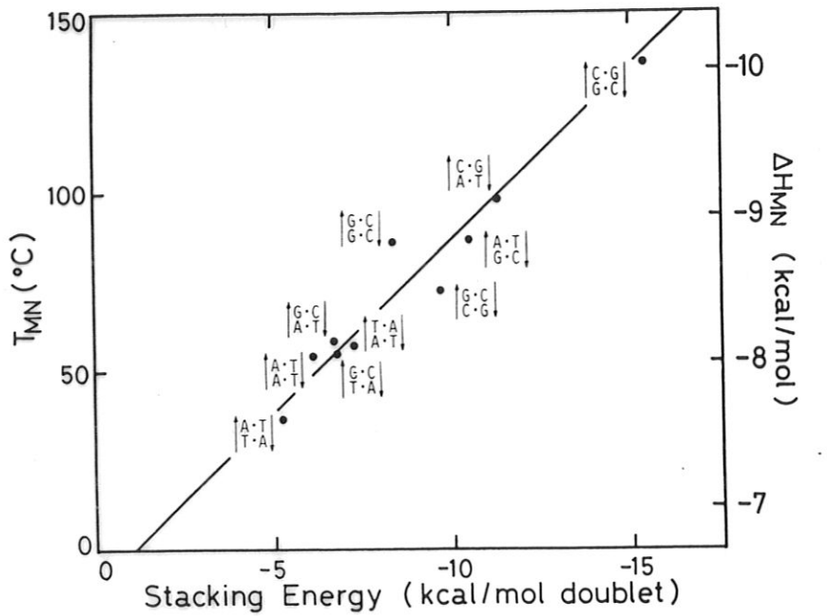


Fig. 3. Correlation diagram between T_{MN} values and stacking energies. The values for stacking energies are taken from Ref. 34. The regression line is also shown.

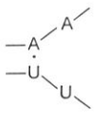
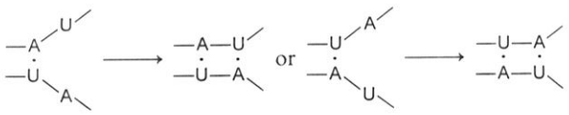
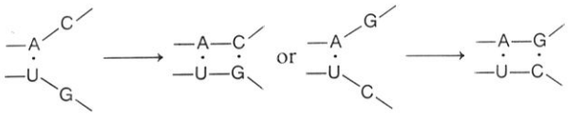
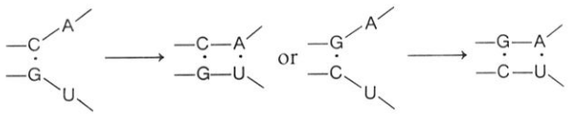
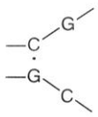
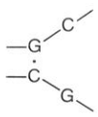
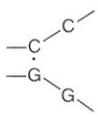
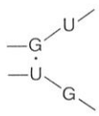
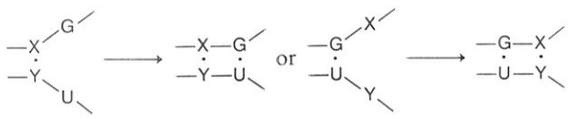
sequence has the lowest stability and that by UAG the second lowest among triplet codons if the order of stability of doublets in RNA is similar to that of the negative stacking energies obtained by Rein and coworkers.[33,34] This adds credibility to the hypothesis that "termination sequences have been selected because of their low abilities to form complementary complexes" proposed by Grosjean et al.[38] on the basis of their observation that the anticodon–anticodon interaction between tRNAs involving the UAG termination sequence is anomalously weak.

There has recently been much interest in detecting local deformation in the DNA double helix.[29,39–41] The wide range of vertical stacking forces may be involved in these local structural variations in DNA. It is also likely that the weak pyrimidine–purine links are the sites of kinks in helix,[42] strand breakage, and/or acceptor sites of chemical drugs,[43,44] mutagens, and carcinogens.[45] Furthermore, stacking heterogeneity may be an important factor that determines stabilities of short duplex stems in tRNAs, mRNAs, and hairpinlike structures near functional loci in DNA (e.g., Ref. 46) predicted from nucleotide sequences. Our present results afford a new insight into the correlation between the sequence, stability, structure, and function of regions of DNA.

The authors are grateful to Professor Akiyoshi Wada, Dr. Hideki Tachibana, and Dr. Yuzuru

2. Vologodskii, A. V. & Frank-Kamenetskii, M. D. (1978) *Nucleic Acids Res.* **5**, 2547–2556.
3. Ueno, S., Tachibana, H., Husimi, Y. & Wada, A. (1978) *J. Biochem.* **84**, 917–924.
4. Wada, A., Ueno, S., Tachibana, H. & Husimi, Y. (1979) *J. Biochem.* **85**, 827–832.
5. Vizard, D. L., White, R. A. & Ansevin, A. T. (1978) *Nature* **275**, 250–251.
6. Tong, B. Y. & Battersby, S. J. (1978) *Biopolymers* **17**, 2933–2937.
7. Tong, B. Y. & Battersby, S. J. (1979) *Biopolymers* **18**, 1917–1936.
8. Wada, A., Yabuki, S. & Husimi, Y. (1980) *CRC Crit. Rev. Biochem.* **9**, 87–144.
9. Gabbarro-Appa, J., Tougard, P. & Reiss, C. (1979) *Nature* **280**, 515–517.
10. Azbel, M. Ya. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 101–105.
11. Azbel, M. Ya. (1980) *Biopolymers* **19**, 61–80.
12. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., III, Slocombe, P. M. & Smith, M. (1978) *J. Mol. Biol.* **125**, 225–246.
13. Beck, E., Sommer, R., Auerswald, E. A., Kurz, Ch., Zink, B., Osterburg, G., Schaller, H., Sugimoto, K., Sugisaki, H., Okamoto, T. & Takanami, M. (1978) *Nucleic Acids Res.* **5**, 4495–4503.
14. Wada, A., Tachibana, H., Gotoh, O. & Takanami, M. (1976) *Nature* **263**, 439–440.
15. Tachibana, H., Wada, A., Gotoh, O. & Takanami, M. (1978) *Biochim. Biophys. Acta* **517**, 319–328.
16. Wada, A., Tachibana, H., Ueno, S., Husimi, Y. & Machida, Y. (1977) *Nature* **269**, 352–353.
17. Wells, R. D., Larson, J. E., Grant, R. C., Shortle, B. E. & Cantor, C. R. (1970) *J. Mol. Biol.* **54**, 465–497.
18. Belintsev, B. N., Vologodskii, A. V. & Frank-Kamenetskii, M. D. (1976) *Mol. Biol.* **10**, 764–769.
19. Ratliff, R. L., Williams, D. L., Hayes, F. N., Martinez, E. L., Jr. & Smith, D. A. (1973) *Biochemistry* **12**, 5005–5012.
20. Morgan, A. R., Coulter, M. B., Flintoff, W. F. & Paetkau, V. H. (1974) *Biochemistry* **13**, 1596–1603.
21. Gotoh, O., Wada, A. & Yabuki, S. (1979) *Biopolymers* **18**, 805–824.
22. Gotoh, O., Husimi, Y., Yabuki, S. & Wada, A. (1976) *Biopolymers* **15**, 655–670.
23. Fixman, M. & Freire, J. J. (1977) *Biopolymers* **16**, 2693–2704.
24. Schildkraut, C. & Lifson, S. (1965) *Biopolymers* **3**, 195–208.
25. Gruenwedel, D. W., Hsu, C.-H. & Lu, D. S. (1971) *Biopolymers* **10**, 47–68.
26. Lazurkin, Yu. S., Lyubchenko, Yu. L., Pavlov, V. M., Frank-Kamenetskii, M. D. & Berestetskaya, I. V. (1975) *Biopolymers* **14**, 1551–1552.
27. Mandel, M., Igambi, L., Bergendahl, J., Dodson, M. L., Jr. & Scheltgen, E. (1970) *J. Bacteriol.* **101**, 333–338.
28. Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C. & Watts, L. (1974) *J. Mol. Biol.* **88**, 523–533.
29. Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z. & Steitz, T. A. (1979) *J. Mol. Biol.* **131**, 669–680.
30. DeVoe, H. & Tinoco, I., Jr. (1962) *J. Mol. Biol.* **4**, 500–517.
31. Pullman, B., Ed. (1968) *Molecular Associations in Biology*, Academic Press, New York.
32. Fujita, H., Imamura, A. & Nagata, C. (1974) *J. Theor. Biol.* **45**, 411–433.
33. Rein, R. (1973) *Adv. Quant. Chem.* **7**, 335–396.
34. Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. (1978) *Biopolymers* **17**, 2341–2360.
35. Gupta, G. & Sasisekharan, V. (1978) *Nucleic Acids Res.* **5**, 1655–1673.
36. Miller, K. J. (1979) *Biopolymers* **18**, 959–980.

Table 23-4
Thermodynamics of adding a base pair to a double-stranded helix

Reaction	ΔH^0 (kcal)	ΔS^0 (kcal deg^{-1})	ΔG^0 (kcal) at 25°C
	−8.2	−0.0235	**−1.2**
	−6.5	−0.0164	**−1.6**
	−5.9	−0.0127	**−2.1**
			
	−13.0	−0.0335	**−3.0**
	−14.7	−0.0349	**−4.3**
	−13.7	−0.0298	**−4.8**
	—	—	**0.3**
	—	—	**0.0**

NOTE: The assumed values for the standard (strand concentration of 1 M) free energy of initiation at 25°C are +6.0 kcal for an A–U base pair and +5.0 kcal for a G–C base pair. In the last reaction, X and Y are any Watson–Crick complementary base pair.

SOURCE: After P. Borer et al., *J. Mol. Biol.* 86:843 (1974).

$$L = \sum_i w_i l_i \{T_{e,i} - T_{t,i}(T_{MN}[\text{old}]) + T_{a,i}(T_{MN}[\text{old}]) - T_{a,i}(T_{MN}[\text{new}]) + u_k D_k I_{i,k}\}^2$$

is minimized, where $I_{i,k} = 1$ if region i is included in DNA fragment k, but otherwise $I_{i,k} = 0$. u_k is a weight in the range $0 \sim 1$, and D_k, which is to be determined, represents the degree of shift of the melting profile of fragment k from the standard profile. Melting profiles with similar D_k values were combined, and the actual number of independent D_k's was reduced to two or three in most iteration series.

All the calculations were carried out on a disc-based YHP 2108A computer. The programs were written in FORTRAN with several ASSEMBLER subroutines which were devoted to manipulation of nucleotide sequences encoded into binary numbers.

RESULTS AND DISCUSSION

We began the least-squares iteration series with $u_k = 0$ for all fragments, $w_i = 1$ for the 24 regions with $\Delta n_i = 0$, and $w_i = 0.1$ for the other regions. Several values of ΔS ($18 \sim 28$ e.u.) and various functional forms of σ_j[21,23] were examined in order to get better fits of the calculated and observed melting profiles. Although coarse outlines of melting profiles were determined by $-\ln(\sigma_j/\Delta S)$ ($j \simeq 300$), the best results were obtained with $\Delta S = 22 \sim 26$ e.u. and σ_j of the form[23] $\sigma_j = \sigma_0(j + d)^{-c}$ ($c = 1 \sim 2$, $d > 200$).

The fits were further improved by using larger w_i and/or u_k values for the regions or fragments for which discrepancies between $T_{t,i}$ and $T_{e,i}$ were prominent. We examined various combinations of w_i's and u_k's. Iterations were repeated a few times for each set of w_i's and u_k's. Further repetitions did not usually improve the fits. Occasionally fitting was attempted on a subgroup of fragments (e.g., fd fragments only). About 200 different sets of T_{MN}'s were thus obtained. Each time an iteration series was accomplished, several values for σ_0 (and sometimes for c, d, and ΔS) were reexamined. The total number of different sets of parameters used for calculations amounted to nearly 1000.

The T_{MN} values listed in Table I and other parameters specified in the legend to Fig. 1 gave the best-fit profiles among them. T_{MN} values which were largely different from those listed in Table I yielded significant loss of fits for several fragments, even though they gave better fits for one or two fragments. However, uncertainty of several degrees in each T_{MN} value could not be eliminated due to the limited number of sample profiles and to ambiguity in the relative importance of the shift term $u_k D_k I_{i,k}$.

Figures 1 and 2 show the calculated melting profiles superimposed on the observed ones. Good agreement of the two profiles is apparent. We also calculated hundreds of melting profiles for each DNA fragment using two terms of stability parameters for A·T and G·C base pairs with various combinations with other thermodynamic parameters (unpublished results

TABLE I
Stability Matrix for Nearest-Neighbor Stacking Doublets[a]

5′	3′ A	3′ T	3′ G	3′ C
T	36.73	54.50	54.71	86.44
A	54.50	57.02	58.42	97.73
C	54.71	58.42	72.55	85.97
G	86.44	97.73	85.97	136.12

[a] T_{MN} values at $Na^+ = 19.5$ mM are shown in °C.

$-T_{A\cdot T} = 40 \sim 50$°C; chain association factor, $\beta C_0 = 10^{-3} \sim 10^{-10}$; and various functional forms of σ_j[21,23] (e.g., $\sigma_0 = 10^{-3} \sim 10^{-9}$, $c = 0 \sim 2$, and $d = 1 \sim 500$ for σ_j of the form $\sigma_j = \sigma_0(j + d)^{-c}$). In this way, however, we

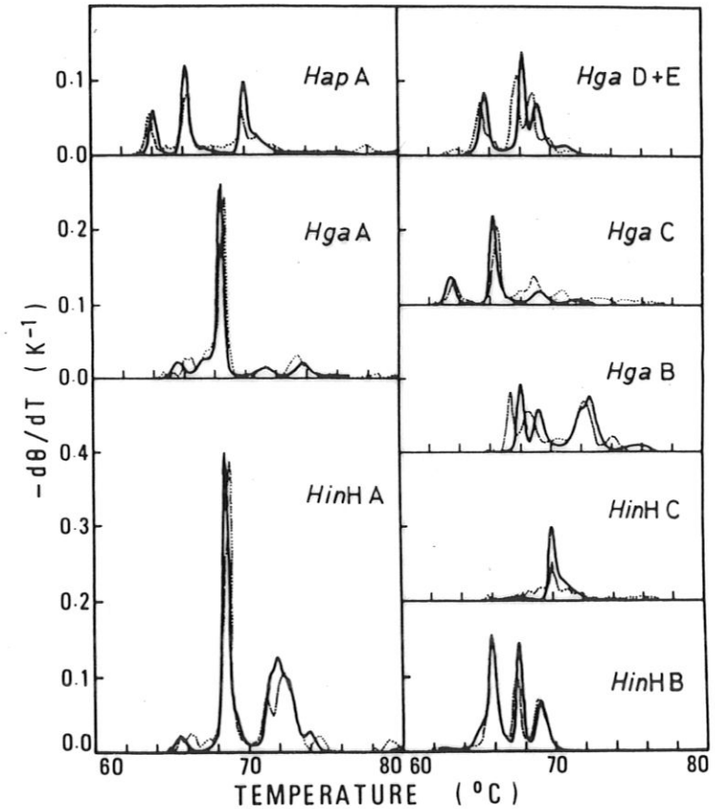


Fig. 1. Comparison of theoretical (solid line) and experimental (dotted line) melting profiles of restriction fragments of fd DNA. Helical contents θ were calculated at intervals of 0.1°C, and thermal derivatives were obtained by six-point linear fitting routine (Ref. 22). The T_{MN} values used for the calculations are listed in Table I. Other parameters are: $\Delta S = 24.54$ e.u.;

The bonding matrix defined in Figs 1(a) and 2(a) exhibits all potentially admissible base-pairings but not their stability ranks. The latter can be exhibited by plotting the probability of each base-pairing in a helix state on the axis of the direction vertical to the two-dimensional plane of the matrix. We refer to this map as the bonding probability map (Suyama *et al.*, 1983). The entire picture of the progressive unfolding transition can be manifested by displaying a series of changing bonding probability maps. The unfolding starts from the map of the ground state. As the transition proceeds, the probabilities in the helix state decrease from the unstable portion and finally all vanish at the complete separation of two chains or the complete denaturation in DNA or RNA, respectively. The problem of stability distribution is solved when all the patterns of these changing bonding probability maps are known.

The series of unfolding patterns in DNA is displayed in a three-dimensional illustration, as shown in Fig. 4. Here, a DNA melting surface in a three-dimensional space consisting of the base-position axis, i, the temperature (or other double helix unfolding perturbation) axis, T, and the axis that indicates the state probability of each base, p, is plotted. In the DNA bonding probability map, only the upper left to lower right diagonal elements have a meaningful value. The unfolding of the DNA secondary structure is thus clearly represented by the melting surface shown in the figure.

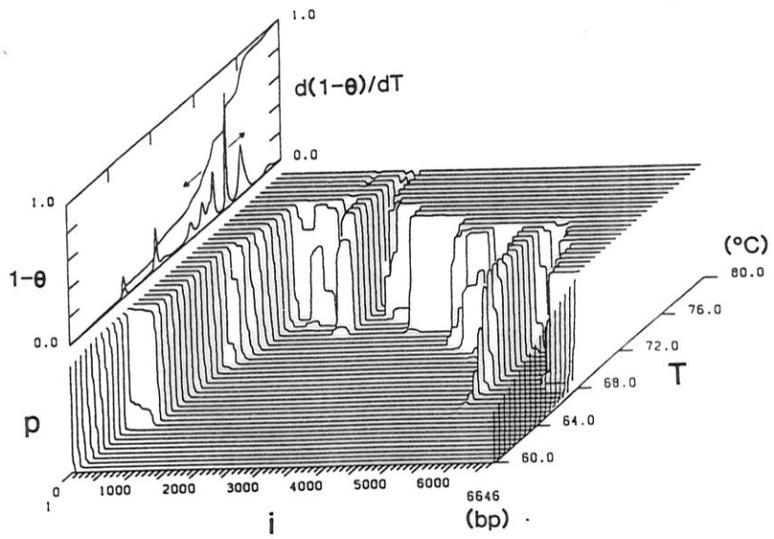


FIG. 4. A melting surface of Col E1 DNA in 0.1 × SSC plotted in a three-dimensional space which consists of the temperature, T, base position, i, and base-pairing probability, p, axes. The melting profile (a T vs $(1-\theta)$ plot, where θ is a helix content), which is a projection of the melting surface onto the T–p plane and is conventionally obtained by spectroscopic measurement, is plotted on the upper left together with its differential curve.

2. *Different Theoretical Achievements Between DNA and RNA*

In comparison with the theoretical achievement in DNA which made it possible to delineate the melting surface shown in Fig. 4 in the above section, the statistical thermodynamics of the RNA chain remains obscure. The reason is based on the structural complexity of RNA secondary structure due to the lack of perfect complementarity in the base sequence and molecular chain topological constraints in counting admissible bondings (see Section II. 4).

By definition, each admissible base-pair can take two states, bonded and non-bonded. Hence, the number of possible states of DNA is $\sim 2^N$ and that of RNA is $\sim N^N$. We must account for many more states in RNA than DNA in statistical thermodynamic calculations.