ACCOUNTS OF CHEMICAL RESEARCH

VOLUME 2 NUMBER 8 AUGUST, 1969

On the Mechanism of Deoxyribonucleic Acid Unwinding

DONALD M. CROTHERS

Department of Chemistry, Yale University, New Haven, Connecticut

Received February 3, 1969

Deoxyribonucleic acid (DNA), like many biological macromolecules, possesses a highly ordered secondary structure. The hypothesis by Watson and Crick[1] of a double-helical form provided one of the first and most important correlations between macromolecular structure and biological function. In their proposed structure, the purine and pyrimidine bases attached to the sugar phosphate backbone strands serve to link the two strands together by complementary hydrogen bonds, adenine to thymine and guanine to cytosine.

In its biological role the molecule, as the repository of genetic information, must be suited for duplication, since each viable progeny cell must receive a copy. The structure with complementary hydrogen bonds is ideally constructed for this, since each strand can serve as the template for synthesis of its complement. It has since been found[2] that, when DNA is duplicated, one parent strand is generally transmitted to each daughter cell, thus requiring that the DNA molecule be unwound and re-formed in each replication cycle.

From studies on DNA samples in solution, it has been known for roughly 15 years that the double-helical structure can be disorganized and unwound by exposure of the molecule to extreme conditions. The mechanism of this is certainly different from the unwinding that occurs in the cell during replication, but its study does provide insight into a number of relevant questions concerning the molecular properties of DNA.

The denaturation of DNA samples can be brought about in a number of ways. Heat, high and low pH, low salt concentration, and nonaqueous solvents all have a destabilizing influence. Since many molecular properties, such as optical absorbance and rotation, viscosity, and sedimentation coefficient, are altered in the process, it is readily accessible to physical measurement. The evidence (which may be found summarized in earlier reviews[3-5]) indicates that when the double-helical molecule is denatured it is converted to two coiling polymer chains. The secondary structure of these, involving stacking of adjacent bases[6-8] to provide some rapidly fluctuating local order, is of a much lower degree than in the parent double helix.

Qualitative Features of the "Melting" Transition

The most common method for detecting denaturation of DNA is the "melting curve"[9] determined by measuring the ultraviolet absorbance of a solution as a function of temperature. Figure 1 shows the result of such an experiment, in this case in the high pH range. The molecule is progressively destabilized by higher temperatures, and the "melting point" is reduced from 90°, as typical at neutral pH, to near room temperature because of ionization of guanine and thymine bases in the coil form. This deprotonation destroys the hydrogen-bonding complementarity of both adenine–thymine (AT) and guanine–cytosine (GC) base pairs.

An important feature of the transition is its cooperativity. As seen in Figure 1, the major part of the change in absorbance occurs within a few degrees centigrade, much less than would be expected on the basis of the heat per base pair if these were completely independent. However, it is also clear that the profile is not a true phase transition, with an abrupt change from the double-helical state to separated single strands. Rather, there are intermediates in which part of a molecule is disorganized and the rest still helical, for which the most direct evidence is visualization of partially denatured molecules in the electron microscope.[10]

Considerable use of the "melting" transition and its

(1) J. D. Watson and F. H. C. Crick, *Nature*, **171**, 737 (1953).
(2) M. Messelson and F. W. Stahl, *Proc. Natl. Acad. Sci., U. S.*, **44**, 679 (1958).
(3) B. H. Zimm and N. R. Kallenbach, *Ann. Rev. Phys. Chem.*, **13**, 171 (1962).
(4) J. Marmur, R. Rownd, and C. L. Schildkraut, *Progr. Nucleic Acid Res.*, **1**, 231 (1963).
(5) G. Felsenfeld and H. T. Miles, *Ann. Rev. Biochem.*, **36**, 407 (1967).
(6) J. Brahms, A. M. Michelson, and K. E. Van Holde, *J. Mol. Biol.*, **15**, 467 (1966).
(7) M. Leng and G. Felsenfeld, *ibid.*, **15**, 455 (1966).
(8) D. Poland, J. N. Vournakis, and H. A. Scheraga, *Biopolymers*, **4**, 423 (1966).
(9) R. Thomas, *Biochim. Biophys. Acta*, **14**, 231 (1954).
(10) R. Inman, *J. Mol. Biol.*, **18**, 464 (1966).

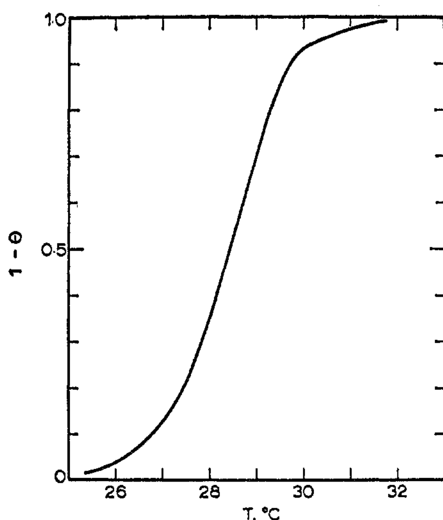


Figure 1. A "melting curve" for bacteriophage T7 DNA, showing $1 - \theta$, the fraction of bases unpaired, as a function of temperature. $1 - \theta$ is calculated by assuming that the fraction of the total change in optical density is proportional to the fraction of bases unpaired; pH 11.5, $[Na^+]$ 0.42 M.

reverse has been made in application to biological and biochemical problems. It is not the purpose of this Account to review these, but instead to assess the present state of our knowledge of the physical mechanisms of the process itself. The approach is largely physicochemical, utilizing statistical mechanics to formulate a theoretical description of the transition equilibrium and kinetic arguments to make deductions about possible mechanisms.

Equilibrium Theory of the Melting of Polynucleotides

The main function of theoretical description of the melting transition is to relate such experimental quantities as shape and breadth of the transition curves to molecular parameters. Because of the great complexity of the system, it is not yet possible to evaluate all of its energy levels, or even to enumerate them. Therefore, theories of the melting transition are, of necessity, thermodynamic in character and rely heavily on comparison with experiment.

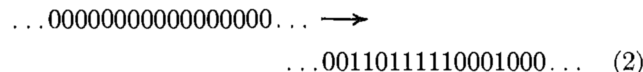
There are many partially melted states intermediate between fully helical and completely denatured. Let the probability that the molecule is in state i be p_i. According to thermodynamics (and analogous to the Boltzmann distribution over discrete energy levels)

$$p_i = \exp(-\Delta G_i/RT)/\sum_i \exp(-\Delta G_i/RT) \quad (1)$$

where ΔG_i is the free energy of forming state i from the reference state, usually taken by convention to be the fully denatured form. Thus in order to calculate the probability of each state, one must know the free energy of all states.

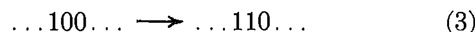
It is common practice to define the states of the molecule according to whether or not each particular base pair is hydrogen bonded. Thus, the model adopted is that of a one-dimensional lattice of length N, the sites of which correspond to the base pairs, giving 2^N possible states for the whole molecule. A typical state can be depicted by a sequence of 0's and 1's, standing for unbonded and bonded pairs, respectively.

Since little can be calculated concerning the detailed interactions, it is convenient to introduce the free energies in eq 1 in parametric form,[11] leaving the question of their evaluation to later. Several classes of reactions are involved in producing a typical state from the reference state (all 0's), as illustrated in reaction 2. One re-



$$\ldots 00000000000000000 \ldots \longrightarrow \ldots 00110111110001000 \ldots \quad (2)$$

action is the bonding of a base pair adjacent to a bonded pair, in what might be called a "growth" reaction (eq 3). Wet let the equilibrium constant for reaction 3 be s,

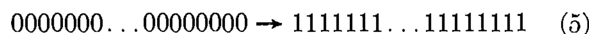


$$\ldots 100 \ldots \longrightarrow \ldots 110 \ldots \quad (3)$$

and by a familiar thermodynamic relation, set

$$s = \exp(-\Delta G_g/RT) \quad (4)$$

where ΔG_g is the standard free energy change for reaction 3. Furthermore, reaction 5, representing the



$$0000000 \ldots 00000000 \rightarrow 1111111 \ldots 11111111 \quad (5)$$

reversal of complete melting, has equilibrium constant s^N (neglecting a single nucleation step; see below). The heat of this reaction is $N\Delta H$; ΔH is the heat of helix formation per base pair. By the van't Hoff relation

$$\partial \ln s^N/\partial T = N\Delta H/RT^2 \quad (6)$$

or

$$\partial \ln s/\partial T = \Delta H/RT^2 \quad (7)$$

Consequently the variation of s with temperature may be obtained from the calorimetric heat of the reaction.[12,13] In general, $s = 1$ very near the midpoint of the transition, so this parameter is known at all temperatures.

If a helical region is nucleated in the middle of a denatured region (eq 8), a process rather different from

$$\ldots 00000000 \ldots \longrightarrow \ldots 00010000 \ldots \quad (8)$$

the growth reaction is involved. First, the newly formed base pair does not have an adjacent pair to interact with by "stacking" of the neighboring unsaturated ring systems, so it is much less stable. Second, the newly paired bases had, in the unpaired state, a much larger volume to move about in than was the case in reaction 3, where they were closely restrained by being at the end of a helical section. This feature also makes the nucleation reaction 8 less likely than the growth reaction 3.

Letting ΔG_n be the free energy of nucleation reaction 8, this can be broken up into

$$\Delta G_n = \Delta G_g - \Delta G_{st} + \Delta G_r \quad (9)$$

where ΔG_{st} is the correction for the missing (negative) free energy of stacking the new pairs on the existing

(11) B. H. Zimm and J. K. Bragg, *J. Chem. Phys.*, **31**, 526 (1959).
(12) E. Neumann and T. Ackermann, *J. Phys. Chem.*, **71**, 2377 (1967).
(13) H. Krakauer and J. M. Sturtevant, *Biopolymers*, **6**, 491 (1968).

helix, ΔG_r is the (positive) free energy correction for closing a larger ring of the two backbones than in reaction 3, and ΔG_g is defined for reaction 3. The magnitude of ΔG_r will depend on the size of the ring which is closed in the nucleation reaction. For generality, the "ring-weighting factor" is expressed as a function $\omega(j)$,

$$\omega(j) = \exp(-\Delta G_r(j)/RT) \tag{10}$$

where j is one-half the number of internucleotide (sugar phosphate) links in the ring. Furthermore, we let

$$\sigma = \exp(\Delta G_{st}/RT) \tag{11}$$

Examination of eq 4, 10, and 11 reveals that the term $\exp(-\Delta G_i/RT)$ for the free energy of molecular state i can be generated by multiplying together appropriate factors of s, σ, and $\omega(j)$ for each interaction in the system. This process involves the methods of lattice statistics and has been done in many different ways;[14-17] these will not be reviewed here.

The qualitative aspects of the transition can be understood from simple free-energy arguments. Since the nucleation process is less favorable than growth, and since a nucleation must be performed to make each separate helical section, the free energy will be minimized by sharply restricting the number of helical sections, with increased average length for those that are present. In the limit of an infinitely difficult nucleation process (corresponding to $\sigma = 0$), boundaries between helix and coil would be suppressed entirely, and the system would show a phase transition from helix to coil at a discrete temperature.[18] When σ is small but nonzero, a transition that is sharp but of finite breadth results, with occasional alternations between helix and coil.

Thus two physical features contribute to producing a sharp melting transition. One is the strong stabilization of base pairs which results from stacking them together, an interaction measured by σ. Melting of DNA requires roughly 8 kcal per base pair;[12,13] it is clearly the enthalpy of stacking the bases[6-8] that is responsible for most of this. From the effects of nonaqueous solvents,[19,20] one may argue that hydrophobic bonding is important for stabilizing the helix, but the expected thermodynamic consequences of this (negative entropy and enthalpy changes of the solvent on melting the helix) are masked by the strong energy of interaction between the bases and the large entropy increase when the double helix is disorganized. This is probably not an unusual situation for complexes involving interaction of unsaturated ring systems.[21]

The second feature that makes nucleation difficult is

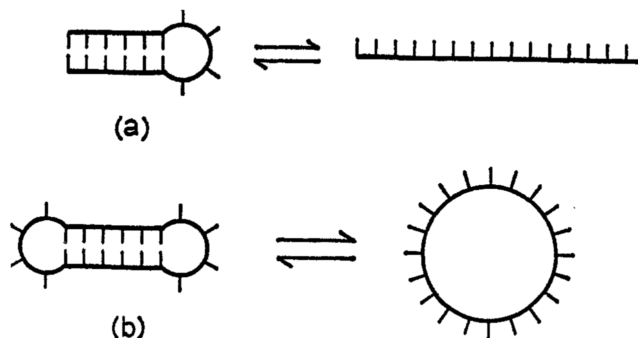


Figure 2. Diagram of the melting of a hairpin helix (a) and a closed single-strand circle (b). The increased T_m of (b) can be related to the free energy of ring closure.[25]

that in a reaction like (8) the two bases must find each other within the volume determined by the size of the coil region. It has been customary[15,16] to take account of this by using the formula of Jacobson and Stockmayer,[22] which gives the relative probability of closing a ring in a random flights polymer chain. Hence

$$\omega(j) = aj^{-\alpha} \tag{12}$$

where $j - 1$ is the number of unbonded base pairs, j is proportional to the number of links in the random flights chain, and α is $^3/_2$ for ideal conditions. The proportionality constant a is not easily evaluated, and is consequently absorbed in the parameter σ, which is possible since every factor of σ has an associated factor of $\omega(j)$ in the product $\exp(-\Delta G_i/RT)$. We have previously estimated that a is not greatly different from one,[23] so σ remains a rough measure of the stacking interaction as shown in eq 11.

Since DNA is not found in ideal or Θ solvent conditions, the actual value of α cannot be $^3/_2$. Excluded volume effects in a polymer chain cause its dimensions to increase more rapidly with chain length than would be the case under ideal conditions. It is therefore reasonable that ring closure should increase in improbability more rapidly with chain length than would be the case under ideal conditions. Theoretical estimates place the value of α between $^3/_2$ and 2.[24] The most promising experimental method for obtaining an estimate of α for small- and medium-sized rings is based on comparison of the melting of hairpin and circular polynucleotides (Figure 2), as described by Baldwin and his collaborators.[25] Since form b must melt into the unfavorable single-strand circle, its melting temperature is higher than that of form a and the magnitude of the shift can be related to the free energy of ring closure.

As indicated earlier, the degree of difficulty of nucleation controls the breadth of the melting transition. The nucleation free energy is in turn influenced by the extent of transition, since nucleation in large coil regions is less likely than in small ones. This implies that transition curves should be asymmetric, being sharper at

(14) S. A. Rice and A. Wada, *J. Chem. Phys.*, **29**, 233 (1958).
(15) J. H. Gibbs and E. A. DiMarzio, *ibid.*, **30**, 271 (1959).
(16) B. H. Zimm, *ibid.*, **33**, 1349 (1960).
(17) S. Lifson and B. H. Zimm, *Biopolymers*, **1**, 15 (1963).
(18) M. E. Baur and L. H. Nosanow, *J. Chem. Phys.*, **37**, 153 (1962).
(19) G. K. Helmkamp and P. O. P. Ts'o, *J. Am. Chem. Soc.*, **83**, 138 (1961).
(20) T. T. Herskovits, S. J. Singer, and E. P. Geiduschek, *Arch. Biochem. Biophys.*, **94**, 99 (1961).
(21) D. M. Crothers and D. I. Ratner, *Biochemistry*, **7**, 1823 (1968).
(22) H. Jacobson and W. H. Stockmayer, *J. Chem. Phys.*, **18**, 1600 (1950).
(23) N. R. Kallenbach and D. M. Crothers, *Proc. Natl. Acad. Sci. U. S.*, **56**, 1018 (1966).
(24) M. E. Fisher, *J. Chem. Phys.*, **45**, 1469 (1966).
(25) I. Scheffler, Ph.D. Dissertation, Stanford University, 1968.

their conclusion than at the beginning. Figure 3 shows that this is so for a homogeneous polynucleotide (the acid double helix of polyadenylic acid) of high molecular weight,[26] with low polymers removed by fractionation.[27]

From curves like Figure 3, σ and α can, in principle, be determined, since both parameters influence the transition breadth, but only α determines the asymmetry or shape of the curve. Earlier comparison of theory[28] with experiment[29] on the melting of polynucleotides used the assumption that $\alpha = {}^3/_2$, with the result that $\sigma = 10^{-3}$ to 10^{-4}. A larger revised estimate for α will mean larger values of σ. Recently a minimum value of $\sigma\omega(^5/_2) = 10^{-3}$ has been estimated from experiments on the melting of hairpin oligonucleotides.[25] Refinement of these parameters is a matter of current concern,[30,31] and reasonable agreement between the various approaches to the problem can be expected soon.

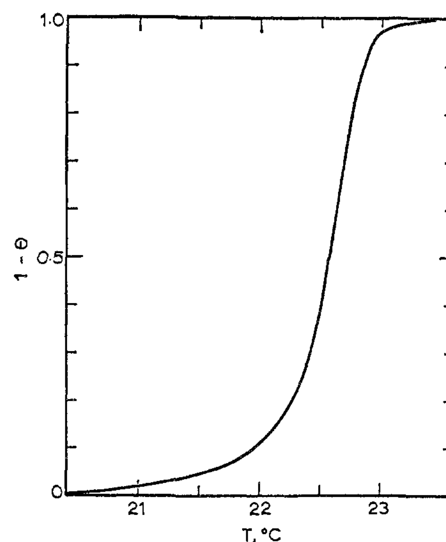


Figure 3. Melting curve of the acid double helix of polyadenylic acid;[26] pH 6.0, [Na+] 0.1 *M*.

Theory of the Melting of DNA

Calculation of melting curves for DNA poses problems of a new dimension in comparison with synthetic polynucleotides. There are several important physical differences between the two materials. The strands in a synthetic polymer can slide with respect to each other and still maintain complementary register, whereas they match in one unique way in DNA. This by itself should mean broader melting curves for the synthetic polymers since they have many more partially bonded states than does DNA. However, the experimental fact is just the opposite, as may be seen by comparing Figures 1 and 3. The source of the broadened transition of DNA is the heterogeneous composition. It is known that the stability of a DNA sample depends on its percentage composition of GC base pairs: the melting temperature increases nearly linearly with per cent GC content.[32] Within a particular molecule some regions will be richer in GC than others because of sequence fluctuations; these will melt at higher temperatures, giving a kind of intramolecular dispersion of the melting curve. The result is to broaden the whole melting curve well beyond that predicted if the base pairs had equal stability.

The main problem in the theory of the melting of natural DNA is to take this heterogeneity of base pair stability into account. The problem is not difficult if it is permitted to assume that the base sequence is periodic,[30,33] but this is not the case for DNA. If one leaves out the dependence of the nucleation free energy, ΔG_n, on the size of coil regions, only nearest-neighbor interactions remain to be taken into account, a considerable mathematical simplification. With the assumption of a random (or other statistical) sequence of bases, the problem has been solved at this level by at least two methods, one analytical[34] and the other Monte Carlo,[35] with quantitative agreement.

Only approximate solutions have yet been obtained for the heterogeneous DNA problem with inclusion of ring weighting. These do, however, give considerable insight into the details of the transition mechanism. An approximation that has been useful is to group together a number of similar molecular states i and assign them identical free energies, an approach we refer to as "coarse graining."[36] In this way it is possible to reduce enormously the number of states which must be counted, thus bringing the problem within the range of present-day computers. The computational details are unimportant here; the theory is, as before, based on eq 1.

One of the important points in this calculation is to establish the relation between base sequence and transition behavior. Figure 4 shows sequence information for a hypothetical DNA molecule, generated by random selection rules on a computer.[36] The sequence contains 5000 base pairs broken up into 100 groups of 50 each. Composition fluctuations are shown in the figure, in the form of the excess of GC pairs in a group of 50 above the average expectation value of 25. According to the coarse-graining approximation each of these groups acts as a unit, being either all helix or all coil. (The practical consequence in the computation is that one has to deal with a 100 × 100 matrix instead of one 5000 × 5000.) A condition on the approximation is that the average length of a helix or coil section in the calculated transition must be considerably larger than the 50 base-pair

(26) D. W. Hennage and D. M. Crothers, to be published.
(27) H. Eisenberg and G. Felsenfeld, *J. Mol. Biol.*, **30**, 17 (1967).
(28) D. M. Crothers and B. H. Zimm, *ibid.*, **9**, 1 (1964).
(29) R. B. Inman and R. L. Baldwin, *ibid.*, **8**, 452 (1964).
(30) N. S. Goel and E. W. Montroll, *Biopolymers*, **6**, 731 (1968).
(31) N. S. Goel, N. Fukuda, and R. Rein, *J. Theoret. Biol.*, **18**, 350 (1968).
(32) J. Marmur and P. Doty, *J. Mol. Biol.*, **5**, 109 (1962).
(33) D. M. Crothers and N. R. Kallenbach, *J. Chem. Phys.*, **45**, 917 (1966).
(34) G. W. Lehman in "Proceedings of IUPAP Conference on Statistical Mechanics, Copenhagen, July 11-16, 1966," T. Bak, Ed., W. A. Benjamin, New York, N. Y., 1967.
(35) T. R. Fink and D. M. Crothers, *Biopolymers*, **6**, 863 (1968).
(36) D. M. Crothers, *ibid.*, **6**, 1931 (1968).

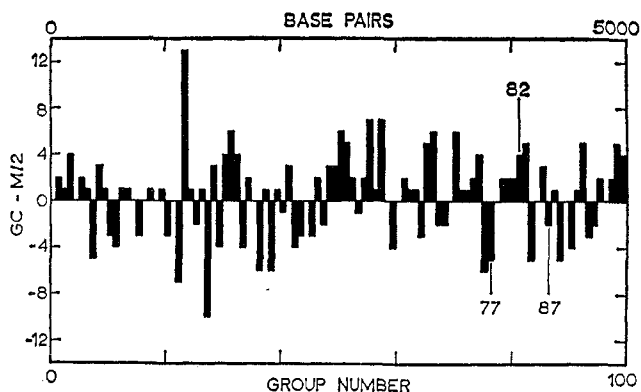


Figure 4. Composition fluctuations in a hypothetical DNA molecule,[36] showing the excess of GC pairs in a group of 50 over the random expectation value of 25.

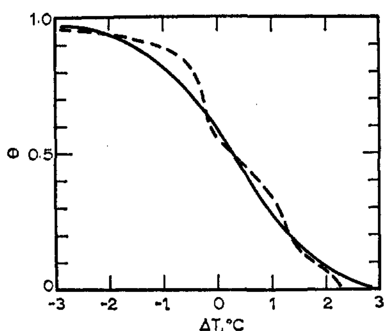


Figure 5. Calculated melting curve (- - - -) for the molecule diagrammed in Figure 4 (with periodic repetition of the indicated sequence); an experimental melting curve for T2 DNA (——) is shown for comparison.

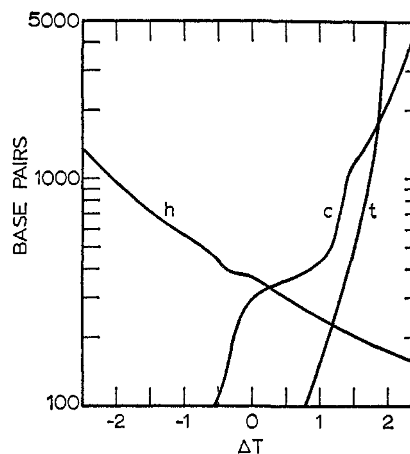


Figure 6. Average length of helix (h) and coil (c) regions, and the length of the coil section at the left end of the molecule (t) calculated for the sequence shown in Figure 4.

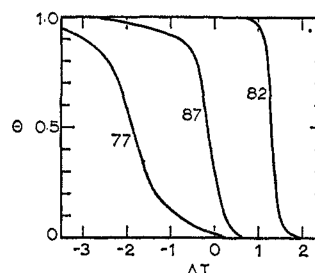


Figure 7. "Melting curve" for individual groups numbered in the sequence diagram in Figure 4. θ is the probability that the base pairs in the particular group are in the helical state.

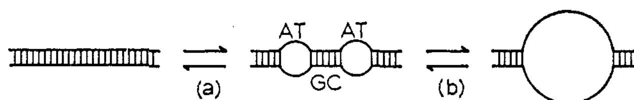


Figure 8. Diagram of two distinguishable effects in the melting of DNA: (a) nucleation and growth of denatured regions at AT-rich loci, and (b) disappearance of the helix section separating the two coils, to produce one large loop.

group, so that the probability of helix–coil alterations within a group of 50 is small and can be ignored. The formalism used for this particular calculation[36] produces the melting behavior of an infinitely long polymer made by periodic repetition of the random-sequence unit containing 5000 base pairs. This again is only an approximation to a molecule of truly random sequence, but in the limit as the repeated unit becomes infinitely long, the fact of its repetition becomes insignificant. For the approximation to be reasonably good, the length of the repeated sequence must be many times the average length of helix and coil sections of the molecule. These conditions are only moderately well met by the sequence and its division illustrated here, but they approach the practical limit of present day computers.

In spite of the approximations, these calculations improve our qualitative understanding of the melting process. Figure 5 shows the calculated melting curve for the sequence illustrated, with $\sigma = 10^{-4}$ and $\alpha = {}^3/_2$, values chosen because they are consistent with the melting breadth of synthetic polynucleotides.[28] An experimental melting curve is shown for comparison. The calculated curve has approximately the correct breadth, although it shows multiphasic character because of the limited size of the sequence period. Figure 6 shows the calculated average lengths of helix and coil regions and the amount of melting from the left end of the molecule.

It is possible to obtain information in considerably greater detail about the melting of the hypothetical molecule shown in Figure 4. The theory allows one to calculate the probability that any particular base pair is bonded or unbonded as a function of temperature. Figure 7 shows examples of this sort of calculation, chosen from representative regions of the molecule. The melting curves for particular bases are labeled with the appropriate group number corresponding to the sequence diagram, Figure 4. As expected, the AT-rich regions of the molecule melt at lower temperatures, but there is also a large difference in the breadth of the transition curves for the two kinds of regions. The process may be visualized as shown in Figure 8, where (a) is the nucleation of denaturation in AT-rich sections, and (b) is the disappearance of GC-rich helical sections to form a single large coil. A sharper transition for (b) than (a) means that the intermediates are represented to a greater extent in process a than b. The apparent heat (van't Hoff) of the transition of group 82 is found to be

about 250 times the heat of melting per base pair; since according to the sequence diagram there are only about 250 base pairs in the GC-rich region around group 82, the van't Hoff heat and the total heat available seem to be approximately the same. Consequently, the representation of intermediates is negligible, and the melting transition of the GC-rich region around group 82 is essentially an all-or-none or two-state transformation.

Thus the picture developed is of relatively gradual nucleation of melting at AT-rich loci as the temperature is raised. When a critical temperature zone is reached, helical regions separating two coil sections melt in a very sharp transition to produce a single larger coil section. The characteristic temperature of this second type of transition will depend on the local GC composition. The physical basis for the difference in the melting curves of AT- and GC-rich regions is the variation of the nucleation free energy with the size of coil regions. It is much more favorable to open a single base pair in the middle of a long helix than to form a single pair in the middle of a long coil region. The former is the nucleation event in transition a, the latter in transition b; the transition breadths are reflecting the extent to which the nucleation process is thermodynamically unfavorable.

Kinetics of DNA Melting

Some of these conclusions from the equilibrium theory of the melting transition are supported by studies of the rate of DNA denaturation, and these latter also provide insight into further aspects of the mechanism of unwinding. One feature to be expected in the unwinding of a double helix has been apparent for many years: when unwinding is to be accomplished in the middle of the molecule, the two ends must rotate relative to each other. If the molecule is long enough, the rate can become limited by this factor. Several theories describing this process have been published;[37-40] all predict that the time required for unwinding should increase with at least the molecular weight squared. One way to rationalize this is to view unwinding as a Brownian process in which the twist of the two strands about each other "diffuses" from the interior of the molecule to the ends. Since the time required for a diffusion process depends on the square of the path length, a square dependence on molecular weight is predicted.

Another feature one should anticipate in considering the unwinding kinetics is the possible entry of nucleation processes into rate limitation of the transition. The elementary step of base-pair formation is thought to be very fast,[41] but the combination of many of these into a cooperative process involving a nucleation event could lead to very slow kinetics. In particular, one might expect transition b in Figure 8 to be slow because of the very large nucleation free energy.

One more expected feature of the unwinding kinetics should be noted. If the ends of a double helix are prevented from turning, for example, by forming a closed double circular molecule, the helix is stabilized against denaturation. However, sufficiently extreme conditions of temperature or pH can cause disorganization of the molecule, as evidenced by increase of the optical density in the ultraviolet,[42] even though unwinding is prohibited. Thus, for sufficiently extreme perturbation conditions the optical changes can become completely decoupled from actual unwinding, and the kinetics of the former can be very different from the kinetics of the latter.

Perhaps the simplest experiments to understand on the kinetics of DNA unwinding are those of Davison[43,44] concerning the time required at very high pH for the molecule to unwind to the point where separation of the strands is preferred over rapid rewinding when neutral conditions are restored. Under the high pH perturbation the optical changes are over in less than 10 msec,[45] but "strand separation" takes orders of magnitude longer than this, illustrating the decoupling between optical changes and unwinding. Davison[43,44] showed that the time for "strand separation" is proportional to the square of molecular size and linear in solvent viscosity, supporting the view that hydrodynamic resistance to unwinding is rate limiting.

On the other hand, when the rate of the optical changes is measured following a much smaller perturbation in the transition region, complicated kinetic curves containing very slow effects are seen. In addition, one part of the curve may have a different molecular weight dependence than another. Figure 9 shows logarithmic plots of the kinetic data following a temperature jump of 1°[46] (refer to Figure 1 for the equilibrium transition curve), including the DNAs from T2 and T7 bacteriophage, which have a molecular weight ratio of about 5:1. As may be seen, the initial slopes of the kinetic plots differ sharply, but the slowest effects are of comparable rates for the two DNAs.

One must conclude that more than one effect is appearing in the kinetic curves, since the dependence on molecular weight is not uniform. This conclusion is further strengthened by the large difference in the temperature coefficient of different parts of the kinetic curve. Figure 10 shows how two different measures of the time required for the optical changes vary with perturbation size. Shown are τ_{max}, the slowest exponential decay time, and $\bar{t}$, an average decay time obtained by dividing the area under the measured kinetic curve by its amplitude, for the two DNAs, T2 and T7. Especially

(37) W. Kuhn, *Experientia*, **13**, 301 (1957).
(38) H. C. Longuet-Higgins and B. H. Zimm, *J. Mol. Biol.*, **2**, 1 (1960).
(39) M. Fixman, *ibid.*, **6**, 39 (1963).
(40) D. M. Crothers, *ibid.*, **9**, 712 (1964).
(41) M. Eigen in "Nobel Symposium No. 5," S. Claessan, Ed., Almquist and Wisbell, Publishers, Stockholm, 1967, pp 354–363.
(42) J. Vinograd, J. Lebowitz, and R. Watson, *J. Mol. Biol.*, **33**, 173 (1968).
(43) P. F. Davison, *ibid.*, **22**, 97 (1966).
(44) P. F. Davison, *Biopolymers*, **5**, 715 (1967).
(45) M. Ageno, E. Dore, and C. Frontali, *Rend. Accad. Naz. Lincei*, **40**, 346 (1966).
(46) H.-CH. Spatz and D. M. Crothers, *J. Mol. Biol.*, in press.

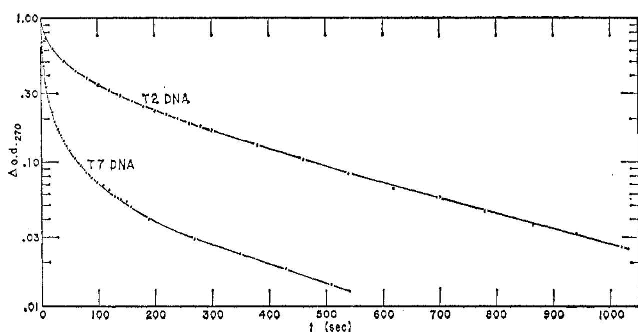


Figure 9. Kinetic curves for melting of T2 and T7 DNAs,[46] showing the difference between the final optical density and that at time t on a logarithmic scale. The total change is normalized to 1 at $t = 0$. The temperature perturbation is 1° in the middle of the melting curve shown in Figure 1.

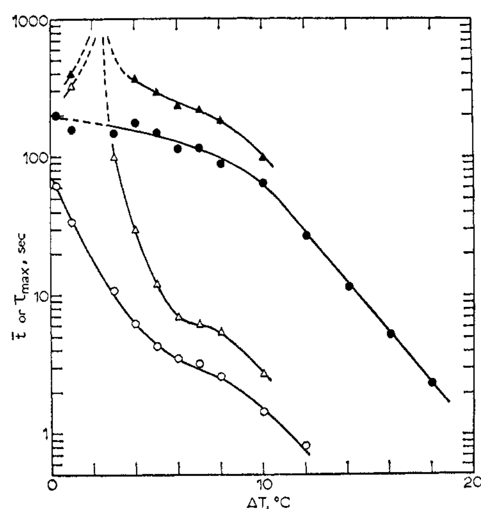


Figure 10. Characteristic times for DNA denaturation as a function of perturbation size, showing $\bar{t}$ (circles) and τ_{max} (triangles) for T7 DNA (open symbols) and T2 DNA (filled symbols). The complex character of the kinetics is revealed by the varying temperature coefficient of the rate.

notable for T7 is the strong variation with ΔT of t and τ_{max} when the perturbation size is below $\sim 5°$. There follows a sudden plateau, in which region the rates for T2 and T7 are more nearly constant and change with ΔT in a parallel manner. In this region the ratio of the two rates is roughly the square of the molecular weights, and one can conclude that viscous resistance to rotation is the likely rate-limiting factor.

The rapid decline in $\bar{t}$ when ΔT exceeds $\sim 10°$ is symptomatic of the decoupling between unwinding and optical changes. When the perturbation becomes larger the helical structure can be disorganized with less and less untwisting, so the optical changes, which reflect the former, become faster and faster.

The large change in $\bar{t}$ and τ_{max} when ΔT is small is not so easy to understand. The main question is the origin of this very slow effect—much slower than the hydrodynamically limited unwinding time for T7 DNA—with its large temperature coefficient. The present working hypothesis[43] is that this effect arises from equilibration of the type shown in Figure 8b. All of the kinetic observations can be explained on this basis. For example, the transition 8b in the forward direction should have a very large activation energy because of the heat required to melt all of the base pairs. The average helix length is several hundred pairs (see Figure 6), and the heat is approximately 8 kcal per base pair. One would expect the activation energy in the forward direction to be roughly the equilibrium enthalpy, which amounts to thousands of kilocalories. Consequently, only GC-rich helical sections which are very near their critical temperatures will melt at the slow relaxation rate of transition 8b, accounting for the small amplitude of the effect associated with τ_{max} for T7 DNA in Figure 9. Most of the melting occurs at a rate faster than τ_{max}, but limited by the time set by viscous resistance to rotation. Since this is much slower for T2 DNA, the initial part of the curve in Figure 9 is much slower in that case.

A concrete basis for these suggestions is provided by recent work on the rate of melting of oligonucleotides,[41] which undergo an all-or-none transition between separated strands and double helix that is closely analogous to that diagrammed in Figure 8b in the interior of a long molecule. A large temperatue coefficient for the rate of the melting reaction is indeed found for oligonucleotides.

When close to equilibrium, transformation 8b requires fluctuations in the location of twist in order to form the necessary partly melted transient intermediates. The twist in the helical section must be disposed of somewhere, and if this is done by transferring it to neighboring coil regions, the rate should be independent of molecular size. This is presumably the case for the data for T2 and T7 DNAs in Figure 9. However, two conditions could cause τ_{max} in the transition region to depend on molecular weight: when the molecule is short, the fluctuations could involve rotation of the chain out to its end, at a rate dependent on molecular weight, or, if the molecule is very large, equilibration 8b could become fast relative to friction-limited unwinding, and one would measure the slower rate of the latter. These considerations may help explain the disparate reports on the molecular weight dependence of the relaxational rate of unwinding in different solvent systems.[47,48]

Finally, in connection with process 8b it is important to recognize the distinction between the kinetic barrier to the fluctuations in twist which are necessary for internal equilibration and the kinetic barrier to the unwinding that follows (or accompanies, depending on relative rates) the internal equilibration. Fluctuations occur against a local free-energy gradient and are relatively much slower than the general unwinding, which is driven by the net decrease in the free energy of the molecule as it unwinds. Thus, the hydrodynamically limited time, measured from the rate following a large perturbation (where internal equilibration is rapid), is much

(47) M. T. Record, Ph.D. Dissertation, University of California, San Diego, 1967.
(48) H. Massie, Ph.D. Dissertation, University of California, San Diego, 1967.

faster than τ_{max} in the transition, even though the latter is in a sense also limited by a frictional term, the resistance to fluctuations in twist.

Conclusion

The equilibrium and kinetic properties of the helix-coil transition in nucleic acids clearly reflect a number of detailed properties of the molecule, such as average dimensions and interactions in coil and helix regions and fluctuations in the local base composition. There is reason to expect similar dependences to be found for other nucleic acids such as the naturally occurring transfer and ribosomal RNAs, whose secondary structure is of considerable current interest. Studies of the details of the melting transition of these should give valuable insight into their specific structure. Nucleic acids differ from proteins in having much simpler, more predictable mechanisms of denaturation. It is for this reason that studies of unwinding can be interpreted in terms of molecular parameters for nucleic acids and that work of this kind can be expected to yield specific information more easily than is the case with proteins.

A Quantitative Approach to Biochemical Structure–Activity Relationships

CORWIN HANSCH

Department of Chemistry, Pomona College, Claremont, California 91711

Received January 9, 1969

This Account dates from 1946 when, under less affluent conditions, Robert Muir of the Pomona botany department was housed in our chemistry building. His interest in plant growth regulators and my interest in indole derivatives soon led to a joint effort to correlate chemical structure with biological activity of the indoleacetic acid-like synthetic hormones.[1] Although we at once began to obtain interesting qualitative results, attempts to formulate these in quantitative terms were frustrated by our conceptual training. As a plant physiologist, Muir was well aware of "lock and key" theory of enzyme–substrate reactions and, as a chemist, I was conditioned to explain substituent effects in the electronic terms of the Hammett equation. Neither of us having had training in pharmacological work, the arguments of Veldstra[2] that relative lipophilic character of substituents was highly important fell on deaf ears.

It was during the reflective period of a sabbatical leave in 1960 in Professor Huisgen's laboratory in Munich that it finally became apparent to me that a study of partition coefficients of the phenoxyacetic acid growth regulators might be helpful in assigning relative lipophilic character to the different members of the series. Soon after my return to Claremont, such a study with Maloney[3] showed the importance of this parameter in explaining the relative activity of the many phenoxyacetic acids which had been tested on *Avena* coleoptiles by Muir, now at the University of Iowa.

It was clear from different ways of mathematically combining Hammett σ constants and log P values (P is the octanol–water partition coefficient of the un-ionized molecule) that a considerable reduction in the variance in the data was possible. At this point Professor Toshio Fujita from Kyoto University joined our group and made the important suggestion that we follow an approach used by Taft[4] and linearly combine the two constants as in eq 1. In eq 1, C is the molar

$$\log (1/C) = k_1\pi + k_2\sigma + k_3 \quad (1)$$

concentration of compound producing a standard response in a constant time interval. The constants k_1, k_2, and k_3 are obtained *via* the method of least squares. In eq 1, π is defined analogously to σ; *i.e.*, $\pi = \log P_X - \log P_H$. P_X is the octanol–water partition coefficient of a derivative and P_H that of a parent molecule, in this case phenoxyacetic acid.

Although eq 1 gave much better correlations than the simple Hammett equation ($k_1 = 0$), or what might be called a Meyer–Overton[5] equation ($k_2 = 0$), it still left a good deal to be desired. It was not until I could bring myself to postulate that log $(1/C)$ was not linearly but *parabolically* dependent on log P that a generally useful equation (eq 2) was obtained. The idea

$$\log (1/C) = -k_1(\log P)^2 + k_2(\log P) + k_3\sigma + k_4 \quad (2)$$

behind eq 2 is that molecules which are highly hydrophilic will not penetrate lipophilic barriers readily and hence will have a low probability of reaching the biolog-

(1) (a) R. M. Muir and C. Hansch, *Ann. Rev. Plant Physiol.*, **6**, 157 (1955); (b) R. M. Muir, T. Fujita, and C. Hansch, *Plant Physiol.*, **42**, 1519 (1967).

(2) H. Veldstra, *Ann. Rev. Plant Physiol.*, **4**, 151 (1953).

(3) C. Hansch, P. P. Maloney, T. Fujita, and R. Muir, *Nature*, **194**, 178 (1962).

(4) R. W. Taft in "Steric Effects in Organic Chemistry," M. S. Newman, Ed., John Wiley & Sons, Inc., New York, N. Y., 1956, p 556.

(5) K. H. Meyer and H. Hemmi, *Biochem. Z.*, **277**, 39 (1935).