ACCOUNTS OF CHEMICAL RESEARCH
VOLUME 4 NUMBER 8 AUGUST, 1971

Experimental Tests of the Theory of Deoxyribonucleic Acid Melting with d(T-A) Oligomers

ROBERT L. BALDWIN†

Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

Received December 17, 1970

Hairpin Helices Formed by d(T-A) Oligomers and the Theory of DNA Melting

A few years ago, Elson, Scheffler, and I began a study of loops in hairpin helices formed by d(T-A)[1] oligomers of defined chain lengths. The properties of loops in nucleic acid helices are important for several reasons. They are found in all transfer RNAs, and the loops as well as the base pairs must be taken into account in estimating the relative stabilities of different cloverleaf models for tRNA structure. Loops also play an important role in determining the melting behavior of macromolecular DNAs.

A basic aim of our work, one which provides a focus for this Account, was to test some fundamental assumptions in the theory of DNA "melting," a term used by DNA chemists for the transition from a helix to a non-base-paired chain. Oligomers are ideal materials for such a study because helix stability depends strongly on chain length in this size range, and a detailed study of this size dependence can be used to test the theory.

Compared to other biological macromolecules, double-helical DNA has a very simple structure, so that one can hope to predict the position of equilibrium for any specified reaction, such as the opening of a particular sequence in a long DNA helix, by measuring only a few basic equilibrium constants as a function of temperature and monovalent counterion concentration (M). A theoretical framework for doing this has been worked out, with minor variations, by several authors. In particular, the theory of Zimm[2] has been widely used.

In its present form, the theory of DNA melting contains the following elements. (1) There are two physical types of base pairs: the *isolated* base pair, which nucleates a new helical segment, and the *stacked* base pair, which is added at the end of an existing helix and gains stability *via* a *stacking interaction* with the adjacent base pair. (As customarily defined, there are $n - 1$ stacking interactions in a helical stack of n base pairs.) (2) There are two chemical types of base pairs, A·T and G·C, which differ in stability. (3) There are three types of loops of unpaired bases. In natural DNAs a *symmetric two-chain loop* is formed when a new helical segment is nucleated in a partly helical, partly unpaired, molecule (see Figure 1). Double-stranded synthetic DNAs with repeating sequences can form *asymmetric* two-chain loops, with different numbers of unpaired bases on opposite strands. Finally there are *hairpin loops*, formed by the looping back of one chain on itself, as in tRNA molecules and in d(T-A) hairpins.

In our work with d(T-A) oligomers, we bypass the problem of two chemical types of base pairs and, since every T·A base pair in the interior of a helix has the same nearest neighbors, we suppose that the stacking interaction is the same for each stacked base pair. Then we are free to focus on three basic questions. (1) How stable is an isolated T·A base pair in water, when it closes the minimum-size hairpin loop? (2) Does the "stability constant"[2] s for a stacked base pair depend on helix length and on the size of an oligomer? (3) How does the equilibrium constant for hairpin loop formation vary with loop size? For sufficiently large loops, the difference in geometries between hairpin and two-chain loops should become unimportant, so that a study of hairpin loops should also be useful in understanding the properties of two-chain loops.

Nucleation of the DNA helix has a low probability compared to helix propagation, for two reasons. (1) The isolated base pair lacks a stacking interaction. Studies of base stacking in systems where base pairing does not occur (*e.g.*, in dinucleoside phosphates)[3] show that base stacking in water is driven by a favorable standard enthalpy ($\Delta H°$) and is opposed by the standard entropy ($\Delta S°$). Probably this is true also of the

† Supported by research grants from the National Science Foundation (GB 8016) and the National Institutes of Health (AM 04763).

(1) Abbreviations: DNA, deoxyribonucleic acid; RNA, ribonucleic acid; d, DNA chain; r, RNA chain; tRNA, transfer RNA; A, G, C, T, U, adenine, guanine, cytosine, thymine, and uracil, respectively; (T-A), chain with a repeating T-A-T-A ... sequence; $d(T\text{-}A)_N$, chain with the T-A sequence repeated N times and with T at the 5′-phosphate end of the chain; poly[d(A-T)], polymeric chain with a repeating A-T sequence; T·A, the base pair of T hydrogen bonded to A. See W. E. Cohn, *Eur. J. Biochem.*, **15**, 203 (1970).

(2) B. H. Zimm, *J. Chem. Phys.*, **33**, 1349 (1960). For a recent review of the theory of DNA melting, see Y. S. Lazurkin, M. D, Frank-Kamenetskii, and E. N. Trifonov, *Biopolymers*, **9**, 1253 (1970).

(3) R. C. Davis and I. Tinoco, Jr., *ibid.*, **6**, 223 (1968).

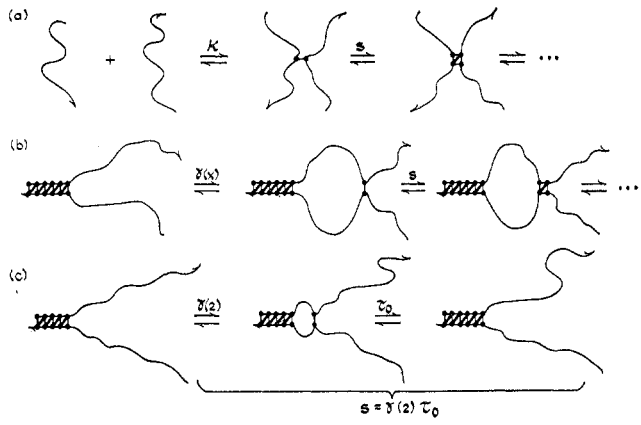

Figure 1. Diagrams of helix nucleation *via* the formation of an isolated base pair: (a) by joining the two complementary DNA chains—the bimolecular equilibrium constant is κ and the helix propagation constant (or stability constant) is s; (b) by closing a two-chain loop in a partly helical DNA molecule—the loop has x internucleotide links and $x - 2$ unpaired bases. In (c) the formation of a stacked base pair is divided into two processes for the purpose of introducing the stacking interaction constant τ_0: first, an isolated base pair is formed with equilibrium constant $\gamma(2)$; second, the isolated base pair is added on to the helix with equilibrium constant τ_0. (For a hairpin loop (see Figure 2) with $x - 1$ unpaired bases there are x internucleotide links counted as phosphate residues; $x - 1$ are backbone links with six rotatable bonds per link and one is a base-pair link with 11 rotatable bonds, treating the base pair as rigid unit.)

stacking interaction between base pairs in the DNA helix. Since the values of $\Delta H°$ for isolated and stacked base pairs are quite different, the equilibrium constants for isolated and stacked base pairs will have quite different dependences on temperature. (2) The DNA helix may be nucleated in a variety of ways (by joining two separate chains or by closing loops of different types), all of which have unfavorable $\Delta S°$'s. To form a base pair the two bases must be brought within a critical volume element δv, whose radius is of the order of the H-bond length.[4]

The probability of closing a loop consequently decreases with loop size. This is expressed quantitatively by the loop-weighting function (or "lwf") which gives the time-average effective concentration of one base in the vicinity of its partner-to-be. By use of the lwf $\rho(x)$, the equilibrium constant for helix nucleation *via* loop closure, $\gamma(x)$,[5] may be correlated with the nucleation constant κ for joining two chains (see Figure 1).[6] The units of κ and $\rho(x)$ are reciprocal; if the

$$\gamma(x) = \kappa\rho(x) \qquad (1)$$

effective concentration $\rho(x)$ has units of molecules/Å^3, then the bimolecular constant κ has units of Å^3/molecule. The term "effective" concentration is used to describe $\rho(x)$ because the probability of base-pair formation depends both on the concentration of the complementary base and on the mutual orientation of the bases, which may be correlated with the concentration if the loop is small. Equation 1 is analogous to the correlation by Jacobson and Stockmayer[4] of cyclization with the synthesis of linear polymers in condensation polymerization.

Helix nucleation and propagation have been studied for rA·rA oligomeric helices, formed at acid pH's, by Applequist and Damle[7] and recently by Eigen and Pörschke.[8] This helix, which is formed when rA chains are protonated, has two parallel strands and the bonding between strands includes H bonding between adenine and phosphate.[9] Because its electrostatic properties are quite different from those of the DNA helix, and also because the structure and bonding between strands are unlike those of DNA, the rA·rA acid helix is not a good model for DNA. However, these studies have been very helpful in understanding helix formation by oligonucleotides.

Some Properties of Poly[d(A-T)]

The alternating DNA copolymer poly[d(A-T)] was first discovered by accident,[10] in a control for an experiment studying DNA synthesis *in vitro* by the *E. coli* DNA polymerase. "Nearest-neighbor frequency" experiments showed that the base sequence of poly[d(A-T)] is entirely the alternating, self-complementary sequence A-T-A-T-... within the error of measurement (0.5%). The mechanism of its *de novo* synthesis remains unknown, but a clue has been provided by the finding that short T-A-T-A... oligomers will serve as templates for the enzymatic synthesis of the polymer.[11] The helical structure of poly[d(A-T)] is like that of natural DNAs (the lithium salt gives the B form helix), although a strange form (the "D" form) has been found in fibers of the ammonium and sodium salts.[12]

Several physical experiments indicate that poly[d(A-T)] forms hairpin helix branches[13-15] before and during melting. The ability to branch endows poly-

(4) H. Jacobson and W. H. Stockmayer, *J. Chem. Phys.*, **18**, 1600 (1950). Compare also the correlation of "Hershey circle" formation with the joining of half-molecules of λ DNA: J. C. Wang and N. Davidson, *J. Mol. Biol.*, **15**, 111 (1966); **19**, 469 (1966).

(5) In the case of the α helix, the equilibrium constant for nucleation is written as σs, not γ; σ is then the ratio of the equilibrium constants for nucleation and for propagation. Since $\Delta H°$ for the first H bond formed (between residues 1 and 5 of a new helical segment) is probably the same as for successive H bonds (between residues 2–6, 3–7, etc.), σ is purely entropic and may be assumed to be independent of temperature. A similar notation has generally been used for nucleation of the DNA helix, *i.e.*, writing $\sigma(x)s$ rather than $\gamma(x)$. In our view, this proves awkward. Including s as one factor of the nucleation constant implies a similarity in the thermodynamic properties of isolated and stacked base pairs that does not exist. Also, $\Delta H°$ for $\gamma(x)$ is probably close to 0 while s is strongly temperature dependent; if $\gamma(x)$ is independent of temperature, then $\sigma(x)$ has the inverse temperature dependence of s.

(6) I. E. Scheffler, E. L. Elson, and R. L. Baldwin, *J. Mol. Biol.*, **48**, 145 (1970).

(7) J. Applequist and V. Damle, *J. Amer. Chem. Soc.*, **87**, 1450 (1965).

(8) M. Eigen and D. Pörschke, *J. Mol. Biol.*, **53**, 123 (1970).

(9) A. Rich, D. R. Davies, F. H. C. Crick, and J. D. Watson, *ibid.*, **3**, 71 (1961).

(10) H. K. Schachman, J. Adler, C. M. Radding, I. R. Lehman, and A. Kornberg, *J. Biol. Chem.*, **235**, 3242 (1960).

(11) A. Kornberg, L. L. Bertsch, J. F. Jackson, and H. G. Khorana, *Proc. Nat. Acad. Sci. U. S.*, **51**, 315 (1964).

(12) D. R. Davies and R. L. Baldwin, *J. Mol. Biol.*, **6**, 251 (1963).

(13) R. B. Inman and R. L. Baldwin, *ibid.*, **5**, 172 (1962); R. B. Inman and R. L. Baldwin, *ibid.*, **5**, 185 (1962).

(14) H. Ch. Spatz and R. L. Baldwin, *ibid.*, **11**, 213 (1965).

(15) For a review, see R. L. Baldwin in "Molecular Associations in Biology," B. Pullman, Ed., Academic Press, New York, N. Y., 1968, p 145.

[d(A-T)] with an unusual conformational mobility, that may be detected by viscosity measurements[13] at temperatures even 30–40° below the T_m (the temperature midpoint of the melting transition). This conformational mobility probably explains many of the unusual responses of poly[d(A-T)] to enzymes acting on nucleic acids, such as *E. coli* exonuclease I.[16] Its mobility appears important for the interferon-inducing ability of poly[d(A-T)] and of its RNA analog poly[r(A-U)]; the effectiveness of poly[r(A-U)] for interferon induction is in fact increased 10^6-fold by preincubation for 2 hr at 37°.[17]

Preparation and Characterization of d(T-A)$_N$ Oligomers

Our first problem was, of course, to prepare the oligomers and to fractionate them in a size range where complete separation has been difficult to achieve (chain lengths of 10–50 residues). Cleavage of poly[d(A-T)] with an endonuclease from bovine pancrease (pancreatic DNase) revealed a fortunate specificity of this enzyme for the ApT bond in poly[d(A-T)].[18] The resulting oligomers have equal numbers of A's and T's, with T at the 5′-phosphate end of the chain and A at the 3′-hydroxyl end, and therefore are described by the generic formula d(T-A)$_N$. This fourfold reduction in the number of possible oligomers greatly simplified the problem of fractionation. Electrophoresis in concentrated acrylamide gels was found capable of resolving some 20 d(T-A)$_N$ oligomer bands,[19] both on an analytical and on a preparative scale, with the size range of optimum resolution depending on the acrylamide concentration.

Oligomers in the size range $5 \leq N \leq 25$ were prepared and studied.[18] Their molecular weights were determined by equilibrium centrifugation at an alkaline pH sufficient to melt the helix by titrating the thymidylate residues, thus ensuring that no dimer helices were present. A gift of a chemically synthesized decanucleotide, d(T-A)$_5$, from Dr. H. G. Khorana allowed us to identify the d(T-A)$_5$ band in acrylamide gels and to calibrate the molecular weight measurements.

The first melting curves measured for purified d(T-A)$_N$ oligomers showed the presence of two helical forms with different molar absorbancies. The more hypochromic form was found only at low temperatures and underwent a transition to the normal form between 15 and 35°, depending on N and M. Molecular weight measurements as a function of temperature[18] identified the low-temperature form as a dimer helix (which may aggregate to form longer helices) and identified the normal form as an intramolecular hairpin helix. Conditions were found in which the hairpin helices could be studied in the absence of dimers.

Measurement of the Parameters Needed to Predict Hairpin Melting Curves

The procedure used to test the theory of DNA melting with d(T-A)$_N$ oligomers is reasonably straightforward. Only two equilibrium constants are needed for a satisfactory analysis of the melting curves for open hairpin helices, formed by linear oligomers, and one of these, the stability constant s, can be measured independently of the oligomer melting curves. To describe nucleation, only γ—the equilibrium constant for the minimum-size hairpin loop—is needed for a first analysis, since the minimum-size loop predominates during melting.[6] Because ΔH°_{nucl} appears to be small, γ is assumed to be independent of temperature. A statistical weight is assigned to each hairpin species (Figure 2) which, if the theory is valid, equals the equilibrium constant for the formation of this species from the unpaired chain. Neglecting electrostatic effects, the statistical weight of a partly helical molecule with one minimum-size loop and k base pairs is γs^{k-1}, and the statistical weight of the unpaired chain is 1. Before the statistical weights of all species are summed on a computer to give a predicted melting curve,[6] the weight of each species is multiplied by its average molar absorbance per pair of T and A residues.

Studies on poly[d(A-T)] have shown that the helix hypochromicity (the relative change in absorbance on forming the helix from the unpaired chain) is reasonably independent both of temperature (T) and monovalent counterion concentration (M). When this is assumed to be true also for oligomers, two problems remain. One is to know how the hypochromicity of a helical base pair depends on helix length.[20] In our work Applequist's approximation[21] was used (eq 2), according

$$\mathrm{Hy}(k) = \mathrm{Hy}(\infty)\left(\frac{k-1}{k}\right) \qquad (2)$$

to which the hypochromicity varies with helix length as the ratio of stacking interactions $(k-1)$ to total base pairs (k).

The second problem was to find g, the number of bases in the minimum-size loop, since the hypochromicity of a hairpin helix is a weighted average of contributions from bases in the loop and in the helix. Because g can have only the even values 2, 4, 6... (there must be equal numbers of A's and T's in the loop), we compared the experimental plot of helix hypochromicity *vs.* N with predicted curves for $g = 2, 4, 6, \ldots$[6] The value of $g = 4$, which gave good agreement, is reasonable according to models of the loop built with space-filling models; however, it must be considered tentative because of the assumptions involved in finding g in this way.

The stability constant s, as defined in the statistical

(16) I. R. Lehman and A. L. Nussbaum, *J. Biol. Chem.*, **239**, 2628 (1964).
(17) E. DeClercq, R. D. Wells, and T. V. Merigan, *Nature*, **226**, 364 (1970); E. DeClercq, R. D. Wells, R. C. Grant, and T. C. Merigan, *J. Mol. Biol.*, **56**, 83 (1971).
(18) I. E. Scheffler, E. L. Elson, and R. L. Baldwin, *ibid.*, **36**, 291 (1968).
(19) E. L. Elson, Ph.D. Thesis, Stanford University, 1966; E. L. Elson and T. Jovin, *Anal. Biochem.*, **27**, 193 (1969).
(20) H. K. DeVoe and I. Tinoco, Jr., *J. Mol. Biol.*, **4**, 518 (1962).
(21) J. Applequist in "Conformations of Biopolymers," G. N. Ramachandran, Ed., Academic Press, New York, N. Y., 1967, p 403.

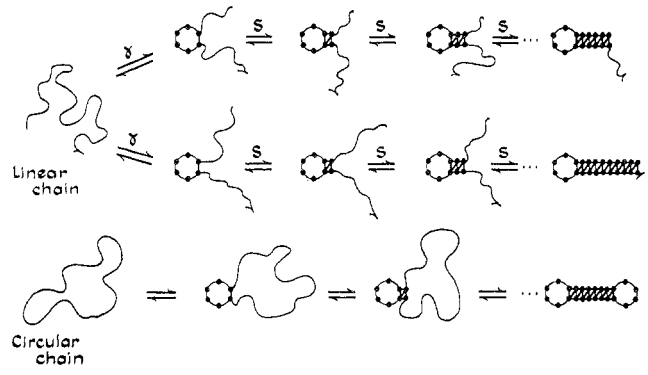

Figure 2. Diagrams of hairpin helix formation by linear and circular oligomers with a T-A-T-A...base sequence. Only the minimium-size hairpin loop is shown. If the loop is central, the linear oligomer can form a complete hairpin helix; if the loop is off-center, part of the chain remains nonbonded and forms a dangling end terminated either by a 5′-P or a 3′-OH group. The entire reaction pathway shown for the linear oligomer may be considered as a loop migration reaction; many such reactions are possible. For the circular oligomer, which forms a closed hairpin helix with two end loops, the complete helix has the same geometry regardless of where the first loop is formed, except as the base sequence in the end loops reads A-T-A-T or T-A-T-A. The equilibrium constants for base-pair formation in the closed helix are complex, since they depend on the sizes of the two loops. The simple scheme of equilibrium constants shown for the open helix has been found to be valid only in special circumstances (at high counterion concentrations).[22]

theory of DNA melting, is required to have certain properties. One requirement is that, for a homogeneous DNA, s should be independent of helix length and not be dependent on oligomer size. Our results showed that the contrary is true in many conditions.[22] Therefore we begin here by describing the apparent stability constant, s^{app}. It is the equilibrium constant for the formation of a stacked base pair in an infinite DNA helix, and it is obtained from measurements made on poly[d(A-T)]. The T_m of poly[d(A-T)] (or T_m^∞) is assumed not to be influenced by nucleation processes, so that $s^{app} = 1$ at T_m^∞. According to a theoretical study by Hijsmans,[23] s does equal 1 at T_m^∞ for poly[d(A-T)] despite the effects of branching preceding melting. The temperature dependence of s^{app} is given by eq 3, and, since $s^{app} = 1$ at T_m^∞, eq 4 follows.

$$-RT \ln s^{app} = \Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad (3)$$

$$T_m^\infty = \Delta H^\circ / \Delta S^\circ \quad (4)$$

Calorimetric measurements by Scheffler and Sturtevant[24] of the poly[d(A-T)] melting curve gave $\Delta H^\circ = -7.9$ kcal/mole base pairs at $M = 0.01$ and showed that ΔH° is reasonably independent of temperature, *i.e.*, that ΔC_p is small compared to ΔH°. We have assumed that both ΔS° and ΔH° are independent of temperature. Therefore eq 3 and 4, together with measurements of ΔH° and T_m^∞, are sufficient to give s^{app} as a function of temperature at $M = 0.01$.

Since s^{app} depends strongly on the counterion concentration, either ΔH° or ΔS°, or both, must vary with M. The dependence of s^{app} on M results from a difference in electrostatic free energy, ΔG°_{el}, per stacked base pair, between the helix and the unpaired chain(s). If we divide ΔG° into a "chemical" and an electrostatic part

$$\Delta G^\circ = \Delta G^\circ_{ch} + \Delta G^\circ_{el} = \Delta H^\circ_{ch} - T\Delta S^\circ_{ch} + \Delta G^\circ_{el} \quad (5)$$

then, since $\Delta G^\circ = 0$ at $T_m^\infty(M)$, we have

$$T_m^\infty(M) = (\Delta G^\circ_{el} + \Delta H^\circ_{ch})/\Delta S^\circ_{ch} \quad (6)$$

Both experiments and some theories[25,26] indicate that a suitable approximate relation between ΔG°_{el} and M, for values of M less than M_{hi}, is

$$\Delta G^\circ_{el} = \Delta G^\circ_{el}(M_{hi}) + B \log (M/M_{hi}) \quad (M \leq M_{hi}) \quad (7)$$

where $\Delta G^\circ_{el}(M_{hi})$ is quite small compared to ΔG°_{el} at low values of M. Thus the relation between T_m^∞ and M, when ΔH°_{ch} and ΔS°_{ch} are taken to be constants, is given by eq 8. The value of M_{hi}, at which T_m^∞ reaches

$$T_m^\infty(M) = T_m^\infty(M_{hi}) + \left(\frac{B}{\Delta S^\circ_{ch}}\right) \log (M/M_{hi}) \quad (8)$$

a plateau, has been found to be about 0.5 for poly[d(A-T)][27] and $M_{hi} = 1$–2 for natural DNAs.[25,26] Typical values of $dT_m/d \log M$ for natural DNAs are about 18°, and a value of 22° has been reported for poly[d(A-T)].[27] When a detailed analysis has been made, the variation of ΔH° and ΔS° with M may well prove to be complex. This problem has been discussed by Record[26] for the case of natural DNAs. He concludes that, to a first approximation, ΔS° is independent of M and ΔH° varies with T_m according to eq 4. We have used this assumption in analyzing our results, but probably the conclusions would not be very different if instead ΔH° were treated as independent of M.

Agreement with Theory at a High Counterion Concentration

To measure γ it is necessary, in principle, to use only the breadth or the T_m of one oligomer melting curve when the other parameters described above have been measured. In practice, γ is found by comparing the entire experimental curve of $T_m(N)$ as a function of N (or $1/(N - 2)$) with curves predicted for different values of γ. The first test of the theory is then whether one value of γ will reproduce the T_m's of all oligomers. At $M = 0.5$, $\gamma = 0.003 \pm 0.001$ was found to do this satisfactorily.[6] The second test of the theory is to use this value of γ to predict the breadths and shapes of the different oligomer melting curves. Again at $M = 0.5$ the agreement was satisfactory with $\gamma = 0.003$.

(22) E. L. Elson, I. E. Scheffler, and R. L. Baldwin, *J. Mol. Biol.*, **54**, 401 (1970).
(23) J. Hijsmans, *J. Chem. Phys.*, **47**, 5116 (1967).
(24) I. E. Scheffler and J. M. Sturtevant, *J. Mol. Biol.*, **42**, 577 (1969).
(25) C. Schildkraut and S. Lifson, *Biopolymers*, **3**, 195 (1965).
(26) M. T. Record, Jr., *ibid.*, **5**, 975 (1967).
(27) R. B. Inman and R. L. Baldwin, *J. Mol. Biol.*, **8**, 452 (1964).

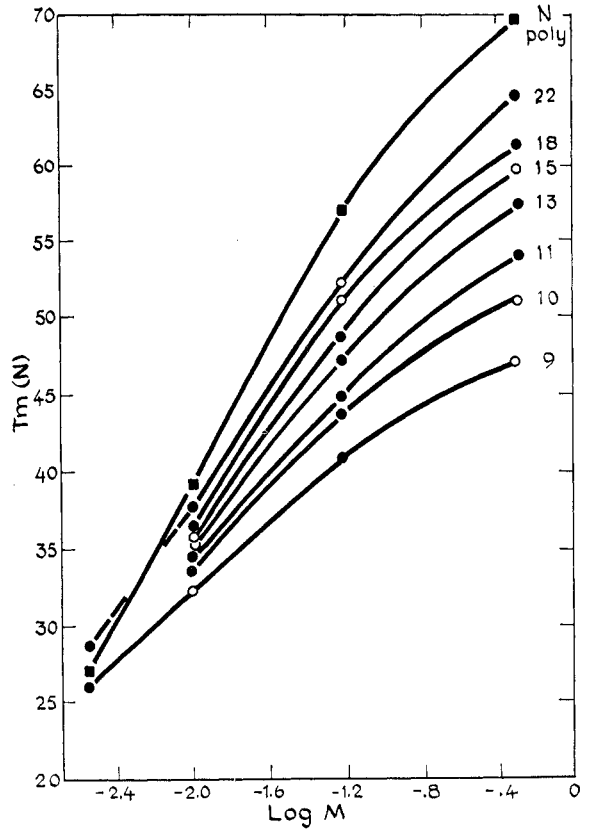

Figure 3. For the linear oligomers d(T-A)$_N$, this graph shows the dependence of the melting temperature $T_m(N)$ of the hairpin helix on N and on the monovalent counterion concentration, M.[22] These results showed that the usual approximation s^{app} = constant, independent of helix length or oligomer size, would have to be modified for DNA oligomers (reprinted with permission from the *Journal of Molecular Biology*).

Dependence of s^{app} on Oligomer Size for $M < M_{hi}$

At $M = 0.06$ and 0.01 it was not possible to reproduce both the breadth and the T_m of even one oligomer melting curve with one value of γ,[22] using the procedures and assumptions which were successful at $M = 0.5$. The reason for this behavior turned out to be very interesting and, in hindsight, not surprising. It is best explained by the data in Figure 3, which show $T_m(N)$ *vs.* log M for several oligomers. It is apparent that $dT_m/d \log M$ depends on oligomer size. Since $dT_m/d \log M$ is a straightforward function of ΔG°_{el} for the infinite helix, the simplest explanation of these data is that ΔG°_{el} depends on helix length and on oligomer size. In turn s^{app} would be a function of oligomer size, and this would explain why the procedure used successfully to fit the oligomer melting curves at $M = 0.5$ (where ΔG°_{el} is small) fails at $M = 0.06$ and 0.01.

To study this problem further, ΔG°_{el} was computed[22] by the method of Schildkraut and Lifson,[25] who had shown that T_m^∞ for the DNA helix is expected to be a linear function of log M at low and moderate M. Pairwise interactions between charges on opposite strands of the helix are summed numerically, and the unknown parameters can be included in one adjustable parameter[26] which may be found from $dT_m^\infty/d \log M$. The dependence of ΔG°_{el} on helix length predicted in this way was found to agree surprisingly well with the experimental values. The calculations also indicated that charge interactions involving unpaired chain segments cannot be neglected. At $M = 0.5$, the computed values of ΔG°_{el} were quite small.

These calculations show the nature of the problem but they leave open the question of how to compute equilibria between different, partly helical, species at low M. *A priori* calculations of these electrostatic effects are semiquantitative at best. In the statistical theory of DNA melting, s^{app} has been taken to be independent of oligomer size and helix length. This assumption, which we find is satisfactory only for $M \geq M_{hi}$, may be regarded as a zero-order approximation, valid only when ΔG°_{el} is quite small. A reasonable first-order approximation for $M < M_{hi}$ is to allow s^{app} to be a function of oligomer size but to require that it be independent of conformation, not dependent on the fraction helix or on the disposition of unpaired chain segments. This approximation was used[22] to predict oligomer melting curves at $M = 0.01$ and 0.06 with reasonable success. To use the approximation, only one new parameter is needed: the experimental T_m of the oligomer. The nucleation constant γ is taken to be independent of both oligomer size and conformation, although γ is allowed to vary with M. The approximation allows rapid calculation of the equilibria and reproduces quite well the breadths and shapes of the oligomer melting curves, but further work is needed to define the limits of its validity.

Helix Formation by "Minicircles" and the Hairpin Loop-Weighting Function

The discovery by Olivera, Scheffler, and Lehman[28] that the DNA-joining enzyme closes d(T-A)$_N$ oligomers into single-strand circles for $N \geq 16$ made it possible to study loops of different sizes. These "minicircles" form closed hairpin helices with a loop at each end (Figure 2) and they can melt only by enlarging loops. In predicting melting curves for circular oligomers, we began by using the Jacobson–Stockmayer lwf,[4] which is valid for long chains whose end-to-end conformations obey a gaussian distribution. In eq 9 J is a constant

$$p(x) = J/x^{3/2} \qquad (9)$$

dependent on chain stiffness. The melting curves for d(A-T)$_c$ minicircles at $M = 0.5$[6] showed that something was wrong: the predicted melting curves gave T_m's lower than the T_m of poly[d(A-T)] and the experimental T_m's were 4–5° higher for $c = 20$ and 31 (Figure 4). No new parameters were involved because the constant J in eq 9 cancels from the expression for statistical weights. Some unexpected feature of the actual lwf for DNA loops evidently was being expressed. Two possibilities were explored: (a) the exponent of $^3/_2$ might be too low, since other studies of polymer chains have suggested an exponent near 1.75[29] or even 2;[30] (b) it

(28) B. M. Olivera, I. E. Scheffler, and I. R. Lehman, *J. Mol. Biol.*, **36**, 291 (1968).
(29) M. E. Fisher, *J. Chem. Phys.*, **45**, 1469 (1969).

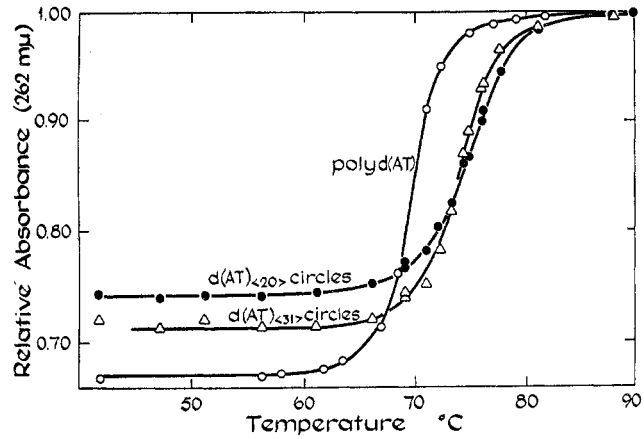

Figure 4. Melting curves of the closed hairpin helices formed by two circular oligomers, relative to the melting curve of poly-[d(A-T)].[6] These results showed that the loop-weighting function which had been used in the theory of DNA melting would have to be modified for small loops, since it predicts that helices formed by circular oligomers should be slightly less stable than poly[d(A-T)] (reprinted with permission from the *Journal of Molecular Biology*).

might be necessary to modify the lwf for small loops,[31] since the Jacobson–Stockmayer lwf is a limiting expression, valid for long chains.

The predicted melting curves proved to be insensitive to the choice of exponent between $^3/_2$ and 2, but sensitive to the form of the lwf used for small loops and especially to the value for $\rho_h(5)$, the lwf for the minimum-size loop. According to Flory and Semlyen[31] one may expect that $\rho_h(x)$ for small hairpin loops will be of the form

$$\rho_h(x) = \frac{J}{[(C_x/C_\infty)x]^{3/2}} \qquad (10)$$

when the distribution of chain conformations is gaussian and in the absence of special orientation effects. The quantity C_x is Flory's "characteristic ratio" which is always less than C_∞; C_x/C_∞ asymptotically approaches 1 for large x. If C_∞ is large (compared to 1), one can expect a large drop in C_x/C_∞ at small x. The constant J also depends on C_∞, which is a measure of the chain stiffness. For the unstacked poly(rU) chain, Inners and Felsenfeld[32] have found $C_\infty = 18$; this is similar to their values for unstacked poly(rA) and depurinated DNA. We tried different empirical representations of C_x/C_∞ in predicting the minicircle melting curves;[6] each approximation tried had only one new parameter. The one which was most successful in predicting the minicircle melting curves was $C_x/C_\infty = 1 - (K/x)$, but all three representations that were tried indicated that $\rho_h(5)$ was about 15 times larger than predicted by the Jacobson–Stockmayer lwf. Consequently there is a simple explanation for why helices formed by minicircles have unexpectedly high T_m's: small loops are more stable relative to large ones than predicted by the Jacobson–Stockmayer lwf.

(30) F. T. Wall, L. A. Hiller, Jr., and W. E. Atchison, *J. Chem. Phys.*, **29**, 2314 (1955).
(31) P. J. Flory and J. A. Semlyen, *J. Amer. Chem. Soc.*, **88**, 3209 (1966).
(32) L. D. Inners and G. Felsenfeld, *J. Mol. Biol.*, **50**, 373 (1970).

Evaluation of the Stacking Interaction

For the purpose of defining the stacking interaction, suppose that a stacked base pair is formed in two steps (see Figure 1). In the first step an isolated base pair is formed with an equilibrium constant $\gamma(2)$. In the second step this isolated base pair is added on to the helix with an equilibrium constant τ_0. The overall equilibrium constant is $s = \gamma(2)\tau_0$. From the results with d(T-A) hairpin helices $\gamma_h(5) = 0.003$, and therefore $\gamma(2) = 0.003\rho(2)/\rho_h(5)$. Since the effective concentration $\rho(2)$ is probably at least as large as $\rho_h(5)$, it is probable that $\tau_0 < s/0.003$. At $M = 0.5$ s varies from 1 at T_m^∞ to about 8 at 20°; thus $\tau_0 < 2 \times 10^3$ at room temperature.

A defined stacking interaction τ is a basic parameter in Zimm's 1960 theory.[2] It was evaluated by Crothers and Zimm[33] from the widths of the melting curves of some DNA homopolymer pairs[27] and was found to be 10^3–10^4. In order to make τ a measurable quantity, Zimm[2] gave a defined value to $\gamma(2)$ based on the Jacobson–Stockmayer expression for $\rho(2)$. The estimate of $\tau = 10^3$–10^4 is quite consistent[34] with $\tau_0 < 2 \times 10^3$ since the Jacobson–Stockmayer expression underestimates $\rho_h(5)$ by a factor of 15 and since this behavior is now expected for small DNA loops, in the absence of special orientation effects.

Prediction of γ and κ

Since γ depends only on κ and on $\rho_h(5)$ (eq 11) it is

$$\gamma = \gamma_h(5) = \kappa\rho_h(5) \qquad (11)$$

possible to predict γ from a measurement of κ (this requires study of dimer helices) and from an analysis of the chain conformation of unpaired d(T-A) chains. Considerable progress has been made in understanding polynucleotide chain conformation, both from experimental[32] and from theoretical studies,[35] and it is possible to compute $\rho_h(5)$ for the different models of chain conformation under current consideration. A value for κ has not yet been measured for the d(T-A)$_N$ oligomers. However, if the assumptions discussed below are correct, similar values of κ will be found for the different RNA and DNA double helices, and a value of $\kappa = 22$ Å^3/molecule can be computed from data given by Applequist and Damle[7] for the rA·rA helix.

Let the loop closure reaction be divided into two partial reactions:[4] (a) bringing the two complementary bases together within a volume element δv needed for H-bond formation and (b) base pair formation (eq 12).

$$\Delta G^\circ_{nucl} = \Delta G^\circ_{loop} + \Delta G^\circ_{bp} \qquad (12)$$

Bringing the bases together is assumed to be purely an entropic reaction, so that $\Delta H^\circ_{loop} = 0$. Also ΔS°_{bp} is

(33) D. M. Crothers and B. H. Zimm, *ibid.*, **9**, 1 (1964).
(34) More specifically, one can compute the product κJ (see eq 10 and 11) from both τ and γ; the same value of κJ is found, within experimental error, from $\kappa J = 2^{3/2}s\tau^{-1}$ and from $\gamma_h(5) = 0.003$, with $C_x/C_\infty = 0.15$ for $x = 5$ in eq 10.
(35) W. Olson, Ph.D. Thesis (with P. J. Flory), Stanford University, 1970; C. DeLisi and D. M. Crothers, *Biopolymers*, in press.

assumed to be small and is neglected. Then

$$\Delta G^\circ_{nucl} = -RT \ln \gamma = \Delta H^\circ_{bp} - T\Delta S^\circ_{loop} \quad (13)$$

with

$$\Delta S^\circ_{loop} = R \ln \rho_h(5)\, \delta v \quad (14)$$

Next, consider the case in which $\Delta H^\circ_{bp} = 0$. (Since H bonds between the bases and water are broken to form a base pair, ΔH°_{bp} should be small.) Then

$$\gamma = \rho_h(5)\delta v \quad (15)$$

and therefore (compare eq 11 and 15)

$$\kappa = \delta v \quad (16)$$

These assumptions lead to an interpretation of κ as the volume element of contact between two bases needed for base-pair formation. If they are correct, as a first approximation, then the radius r of δv should be comparable to the H-bond length (2.8 Å). For the rA·rA helix $r = 1.7$ Å if $\delta v = \kappa = 22$ Å^3/molecule. This is reasonable agreement (within one order of magnitude for δv).

For the four Inners–Felsenfeld models of the polyribonucleotide chain, $\rho_h(5)$ varies between 1.7×10^{-5} and 2.4×10^{-5} molecule/Å^3 if special orientation effects are neglected (E. L. Elson, private communication, 1970). The corresponding values of r, computed from eq 15 with $\gamma = 0.003$ and $\delta v = (4/3)\pi r^3$, are 3.45–3.06 Å. For a free rotation model, $\rho_h(5) = 9.6 \times 10^{-5}$ molecule/Å^3 and $r = 1.96$ Å. The agreement with the H-bond length is as close as can be expected. We conclude that, on the one hand, the assumptions listed above are plausible and that, on the other hand, this comparison is not very sensitive to the particular model used for the polynucleotide chain.

However the comparison of γ with a value predicted from $\rho_h(5)$ is sensitive to the value assigned to ΔH°_{bp}. If ΔH°_{bp} is given a definite value of only -2 kcal, then ΔS°_{loop} is changed from -11.5 eu for $\Delta H^\circ_{bp} = 0$ to -18.2 eu for $\Delta H^\circ_{bp} = -2$ kcal (see eq 13). If $\rho_h(5) = 2 \times 10^{-5}$ molecule/Å^3, r is then only 1.1 Å (see eq 13 and 14). Consequently it is likely that ΔH°_{bp} is small and that γ is reasonably independent of temperature (compare the discussion by Eigen and Pörschke[8]).

Other Conclusions about Equilibria in Hairpin Helices

Does a hairpin helix melt from the open end or by enlarging the hairpin loop? A study of this question,[6] based on the results obtained with circular oligomers, shows that the hairpin melts chiefly from its open end and that the minimum loop is the predominant loop size throughout melting.

What factors control the transition from monomolecular hairpin helices to dimer helices? Obviously higher concentrations will favor dimers and aggregates of dimers. The finding that dimer helices always undergo a transition to hairpins on raising the temperature was unexpected, but is easily understood.[18] The statistical weights, or concentrations relative to the nonbonded chain, of hairpin and dimer helices are

$$\frac{c_n}{c_0} = \gamma s^{n-1} \quad \text{(hairpin, } n \text{ base pairs)} \quad (17)$$

$$\frac{c_N}{c_0} = \frac{\kappa}{2} c_0 s^{N-1} \quad \text{(dimer, } N = 2n + g \text{ base pairs)} \quad (18)$$

where g is the number of unpaired bases in the minimum hairpin loop. The hairpin is more stable than the dimer at high temperatures because it is formed intramolecularly and s becomes small at high temperatures; at T_m^∞, $s = 1$ and the ratio of hairpin to dimer is about 10^5 if $\gamma = 10^{-3}$, $\kappa = 10^{-3}\ M^{-1}$, and $c_0 = 10^{-5}\ M$.

Kinetic Behavior of Hairpin Helices

In principle, the rate constants for opening and closing a T·A base pair can be found by analyzing the kinetics of melting of hairpin helices of different sizes.[36] Since a complete understanding of the equilibrium constants is required for such an analysis, it seems best to analyze first the kinetic results for $M \geq 0.5$, where one can hope to bypass an explicit consideration of electrostatic effects. Unfortunately the d(T-A)$_N$ hairpin helices melt so rapidly at $M = 0.5$ that most of the reaction is complete[37] within the dead time of our temperature-jump instrument[38] (a few microseconds), and it appears that a different method will be required to complete the study.

Analysis of the results is a complex problem for these hairpin helices because a large number of species, intermediates between the complete helix and the unpaired chain (see Figure 2), contribute significantly to the signal. The dimer helices studied by Eigen and Pörschke[36] have been analyzed by a simple steady-state equation, because only two physical species—the complete helix and the unpaired chains—contribute significantly to the signal, to a first approximation. For the d(T-A)$_N$ hairpins, it is necessary to take account of partly melted helices by making a complete analysis of the eigenvectors and eigenvalues of the rate coefficient matrix.[38]

Recently E. L. Elson (private communication, 1970) has obtained an analytical solution to the problem for the simplest case, in which there are only four rate constants: forward and backward rate constants for helix nucleation and propagation. With his solution it is possible to generate predicted kinetic melting curves with reasonable demands on computer time, even for large oligomers, and to compare these with the measured ones. An important feature of these

(36) D. M. Pörschke, Ph.D. Thesis, Göttingen University, 1968; M. Eigen in "Fast Reactions and Primary Processes in Chemical Kinetics," S. Claesson, Ed., Interscience, New York, N. Y., 1967, p 358.

(37) T. Y. Tsong, unpublished; I. E. Scheffler, Ph.D. Thesis, Stanford University, 1968.

(38) M. Eigen and L. C. deMaeyer in "Techniques of Organic Chemistry," S. L. Freiss, E. S. Lewis, and A. Weissberger, Ed., Vol. 8, Part II, Interscience, New York, N. Y., 1963.

predicted curves is that they depend almost entirely on the rate of helix propagation for the longer oligomers, when the temperature jumps are limited to the melting zone, and are relatively insensitive to the rate of helix nucleation. We expect that it will be possible to use this approach to measure the rate constant for propagation and eventually to work out the kinetics of a defined loop migration reaction (see Figure 2).

The studies of oligomers which are described here have been a collaborative effort between Drs. E. L. Elson, I. E. Scheffler, and myself, and the work is being carried forward by Dr. Elson at Cornell. In writing this summary I have made use of his current analyses and unpublished work. While doing this research, we benefitted considerably from discussions with other scientists, in particular: Drs. B. H. Zimm, J. A. Schellman, D. M. Crothers, P. J. Flory, and M. T. Record, Jr. I also thank Barry Nall for his comments on this Account.

The Möbius–Hückel Concept in Organic Chemistry. Application to Organic Molecules and Reactions

HOWARD E. ZIMMERMAN

Chemistry Department, University of Wisconsin, Madison, Wisconsin 53706

Received January 11, 1971

In 1966[1] we presented an approach to determine the forbiddenness or allowedness of organic reactions which is a suitable alternative to the thoroughly documented methods of Woodward and Hoffmann.[2] The present Account summarizes the original approach and applications presented in various subsequent publications and describes some further examples of interest. These include application of the concept not only to prediction of allowedness of reactions but also to description of some ground-state organic systems.

The Concept

One concept which has been of special value in organic chemistry is the Hückel rule[3] which says that, for ground-state molecules with a cyclic array of orbitals, $4N + 2$ electrons lead to aromaticity and special stability deriving from the presence of a closed shell. Systems which have $4N$ electrons are said to be antiaromatic.[4] This rule holds not only for cyclic polyenes but also for cyclic transition states.

However, the $4N + 2$ rule can be shown to apply only to cyclic systems composed of orbital arrays in which there are zero or an even number of sign inversions resulting from plus–minus overlaps.[1] Ordinary cyclic polyenes and other cyclic π systems such as cyclopropenyl, cyclobutadiene, cyclopentadienyl, benzene, cycloheptadienyl, etc., fit this requirement. These are conveniently termed Hückel systems since they fit the Hückel rule and have molecular orbital solutions of the normal Hückel type.

However, many systems in organic chemistry consist of monocyclic arrays of orbitals in which there is one or, alternatively, an odd number (*vide infra*) of overlaps between adjacent orbitals of different sign. These molecules do not have a closed shell with $4N + 2$ electrons but rather need $4N$ electrons for stability. With $4N + 2$ electrons they are antiaromatic.[1] We have termed such molecular species Möbius because the molecular orbital situation is quite like that of Heilbronner's[5] Möbius cyclic polyenes.[1]

Hence the first problem is to learn to recognize which orbital arrays are of the Hückel type and which are Möbius. Figure 1 depicts two unlikely but instructive arrays of arbitrarily chosen orbitals; **1a** is a Hückel system while **1b** is a Möbius one. Note that these arrays may consist of p orbitals, hydrogen 1s orbitals, carbon 2s orbitals, etc., and will roughly approximate situations where all are of the same type.

In Figure 1 it should be clear that the orbitals shown are "basis set orbitals," namely, the assortment of orbitals present *prior to* a molecular orbital calculation; such a set can be chosen with the orientations and assignment of sign selected for convenience. However the molecular orbital calculation results prove independent of the orientations selected. Any concern about adjacent plus–minus overlap as unfavorable is premature. Hence the categorization of a system as Hückel *vs.* Möbius does not require an explicit MO calculation but is made from inspection of the basis orbitals available to the system.

(1) (a) H. E. Zimmerman, *J. Amer. Chem. Soc.*, **88**, 1564 (1966); (b) *ibid.*, **88**, 1566 (1966); (c) *Science*, **153**, 837 (1966).

(2) (a) R. B. Woodward and R. Hoffmann, *J. Amer. Chem. Soc.*, **87**, 395 (1965); (b) *ibid.*, **87**, 2511 (1965); (c) R. Hoffmann and R. B. Woodward, *ibid.*, **87**, 2046 (1965); (d) *ibid.*, **87**, 4389 (1965); (e) R. B. Woodward and R. Hoffmann, *Accounts Chem. Res.*, **1**, 17 (1968); (f) *Angew. Chem., Int. Ed. Engl.*, **8**, 781 (1969).

(3) E. Hückel, *Z. Phys.*, **70**, 204 (1931); **76**, 628 (1932); **83**, 632 (1933).

(4) R. Breslow, J. Brown, and J. J. Gajewski, *J. Amer. Chem. Soc.*, **89**, 4383 (1967).

(5) (a) E. Heilbronner, *Tetrahedron Lett.*, 1923 (1964). (b) This is equivalent to defining X as $X = (\alpha - E)/\beta = [E - \alpha]/-\beta$. (c) Note also that where parallel p orbitals with opposed signs are present, this counts as only one node although there are two sets of lobes; this is because a single node is the occurrence of two adjacent oppositely signed orbitals (not lobes). Also, there is no node between lobes of a single p orbital.