

Thermodynamics of the Helix-Coil Equilibrium in Oligoadenylic Acid from Hypochromicity Studies

Jon Applequist and Vinayak Damle

Contribution from the Department of Chemistry, Columbia University, New York, New York 10027. Received September 10, 1964

The theory of helix-coil equilibria has been applied to hypochromicity studies of oligoadenylic acids. Computer methods were used to solve equations relating the two fundamental equilibrium constants needed to describe the process to the measured hypochromicities. From the temperature dependence of the stability constant for base pairing, the enthalpy of the base pairing reaction was found to be -8.0 kcal./mole of base pairs, and the entropy -21 cal./deg./mole of base pairs, both quantities being constant to within 10% for oligomers of eight to eleven residues. The association factor, which is the additional factor in the equilibrium constant for formation of the first base pair between two strands, was found to be 2.2×10^{-3} l./mole. Good agreement between theoretical and experimental hypochromic "melting" curves was obtained for oligomers of two to eleven residues only when the following contributions to hypochromicity were incorporated in the theory: (i) the dependence of hypochromicity on helix length predicted by Rich and Tinoco and (ii) a temperature dependence of hypochromicity of nonbonded bases proposed to account for anomalous behavior of the melting curves toward the high temperature end. The observed agreement is taken to be support for the existence of both effects and particularly for the hypothesis that interactions between neighboring base pairs are responsible for hypochromism in the helix.

I. Introduction

It has been the object of much work in recent years to understand the phenomena associated with helix-coil equilibria in macromolecules in terms of fundamental molecular processes. Most of the attention so far has been given to the problems of developing suitable theoretical models and mathematical methods for treating them. The results of this work have provided an enlightenment that has been unfortunately limited by the absence of experimental data that offer any more than a qualitative test of the theory, particularly in the case of nucleic acids. Furthermore, the potential value of the theory as it has been developed thus far lies in its means of providing fundamental thermodynamic information from observed properties for a given system; clearly an exploitation of this potential is greatly needed.

Fresco, Blake, and Doty¹ have studied the temperature dependence of ultraviolet absorption in a series of well-resolved fractions of oligoadenylic acid (oligo-A). The purpose of this paper is to present a detailed analysis of their data in terms of some alternative models for the helix-coil equilibrium.

It has been established that polyadenylic acid (poly-A) in aqueous solution at low pH forms a two-

stranded helical molecule.² A detailed structure of the helix consistent with X-ray diffraction patterns for poly-A fibers has been described.³ The studies of the hypochromic effect in the near-ultraviolet suggest that an equilibrium between the helical and randomly coiled⁴ states of the polymer exists and is sensitive to temperature in a way that can be quantitatively described.

The oligo-A systems are of considerable interest in two respects. (i) A substantial body of speculation has appeared regarding the molecular states present in a helix-coil equilibrium of this type.^{5,6} For these oligomers the theoretical treatment is relatively straightforward, and the system therefore offers an unusually good test of the predictions. It is also possible to use the theory to determine from the experimental data fundamental thermodynamic quantities for the equilibrium, and from these to predict features of the system not yet experimentally verified. (ii) The electronic origin of the hypochromic effect has itself aroused much theoretical speculation,⁷⁻¹⁰ and one must accept one or the other of divergent views in order to interpret measured hypochromicities. The view developed by Tinoco,⁷ and subsequently by Rhodes,⁸ is that hypochromism¹¹ in these systems arises from interactions between transition dipole moments of nondegenerate electronic transitions of all neighboring bases along the chains. An important consequence of this view is that the hypochromicity should diminish with decreasing chain length, as calculated by Rich and Tinoco.¹² Similarly, any hypothesis involving interactions of neighboring base pairs should result in a predicted chain-length dependence, such as the interaction due to π -orbital overlap, whose importance was suggested by the calculations of Ladik.¹⁰ On the other hand, Nesbet⁹ has considered the dipole interactions of degenerate transitions and has concluded that these produce significant contributions to the hypochromicity. His calculations suggest that the hypochromicity due to interactions within a single base pair could be large enough to account for most of the effect observed in

(2) J. R. Fresco and P. Doty, *J. Am. Chem. Soc.*, **79**, 3928 (1957); J. R. Fresco and E. Klemperer, *Ann. N. Y. Acad. Sci.*, **81**, 730 (1959).

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

(4) We use the term "randomly coiled" interchangeably with "one-stranded," although the degree of randomness of the single strand remains to be established.

(5) B. H. Zimm, *J. Chem. Phys.*, **33**, 1349 (1960), and references cited therein.

(6) J. Applequist and V. Damle, *ibid.*, **39**, 2719 (1963).

(7) I. Tinoco, Jr., *J. Am. Chem. Soc.*, **82**, 4785 (1960).

(8) W. Rhodes, *ibid.*, **83**, 3609 (1961).

(9) R. K. Nesbet, *Biopolymers Symp.*, **1**, 129 (1964).

(10) J. Ladik, *Acta Sci. Hung.*, **11**, 239 (1960).

(11) In order to avoid ambiguity, we will regard the term "hypochromism" as synonymous with "hypochromic effect," in the qualitative sense, while "hypochromicity" is the quantitative measure of the effect.

(12) A. Rich and I. Tinoco, Jr., *J. Am. Chem. Soc.*, **82**, 6409 (1960).

(1) J. R. Fresco, R. Blake, and P. Doty, manuscript in preparation.

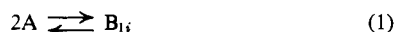
polymers. If this were true, then the hypochromicity should be almost independent of chain length. The hypochromicities of oligomers, interpreted with the aid of additional knowledge regarding the distribution of molecular states, offer an experimental means of determining whether the effect results primarily from *intra-* or *inter-base-pair* interactions. It will be seen that the data behave approximately according to the predictions of Rich and Tinoco, and therefore confirm at least that a nearest-neighbor interaction of some sort is of predominant importance.

II. The Model

The calculation of theoretical hypochromicity will proceed by averaging the hypochromicities of all of the species assumed to be present in the system, each weighted by its relative population. The object of the model, then, is to provide the basis for calculating the populations of the various species. Three different models are compared in this study: (a) the staggering zipper model, which is an extension of (b) the non-staggering zipper model, for which most of the equations have been derived elsewhere,⁶ and (c) the all-or-none model, also a well-known hypothetical case.¹³

The Staggering Zipper Model. The molecular species allowed in this model are illustrated in Figure 1 for chains with a degree of polymerization $N = 3$. The features of the model that are particularly appropriate to poly-A are (i) that both strands in the helical species are identical and parallel³ (as opposed to the anti-parallel arrangement in some nucleic acids), and (ii) that the chains may be staggered, *i.e.*, the ends of the strands need not be in register in partially bonded species, due to the fact that all of the nucleotides are identical (except for minor structural differences in the end groups, which will not be taken into account in this treatment). The term "zipper" is used here to imply that not more than one uninterrupted bonded sequence may exist in any species. This is justified on the ground that the formation of an interruption requires that nonbonded residues be placed in a closed loop at the expense of such residues in "free" chains. Since the free energy for this process is unfavorable, the populations of interrupted species are assumed to be negligible for short-chain polymers. (The term "zipper" may be intended by some authors to imply a *mechanism* by which one species is transformed to another. The treatment here deals only with the equilibrium distribution of species and therefore makes no assumptions about such mechanisms.)

Let A represent single-stranded poly-A whose degree of polymerization is N , and let B be the generic symbol for the two-stranded complexes. More specifically let B_{ni} be the i th species [according to any suitable numbering system such that $1 \leq i \leq (N - n + 1)^2$] among those molecules with n bonded base pairs. To perform the desired calculations we need only the equilibrium constants for the following two reactions



for which the equilibrium constant is defined as β_s , and



(13) J. Schellman, *Compt. rend. trav. lab. Carlsberg, Ser. chim.*, 29, No. 15 (1955).

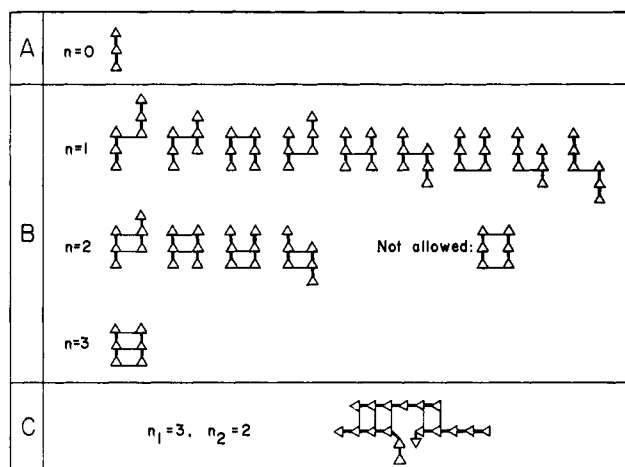


Figure 1. The staggering zipper model. Each triangle represents a nucleotide residue, and lines joining strands represent hydrogen bonds. All species enumerated are shown for $N = 3$, those with equal n having the same population. A typical example of species C is shown for $N = 6$.

for which the equilibrium constant is s . Strictly speaking, the equilibrium constant for formation of staggered species of B_{ni} should be twice that for non-staggered species, since the symmetry number is two for the latter and one for the former. We have accounted for this fact by including in the enumeration of staggered species those which differ only by the interchanging of strands, and which are therefore in reality indistinguishable. (See Figure 1.) Each distinguishable staggered species is thereby doubled in population, in accordance with the symmetry requirement, and the effective number of species with n bonded base pairs is seen to be simply $(N - n + 1)^2$. The quantity β , which we shall call the "association factor," is to be understood as the value appropriate for the formation of species with a symmetry number of two. The quantity s was designated the "stability constant" by Zimm.⁵ It is implied that β and s are constant for all i, j , and n ($1 \leq n < N$) at a given temperature and pressure. It follows from the above that for the reaction



the equilibrium constant K is

$$K = \beta \sum_{n=1}^N (N - n + 1)^2 s^n \equiv \beta L(s) \quad (3b)$$

If K is in terms of molar concentrations, β has units of volume per mole and is equal to $N_0\alpha/2$ in the notation of our previous treatment,⁶ taking the symmetry number into account. As shown before,⁶ the fraction of nucleotides paired f is given by

$$f = \frac{s\{1 + 4\gamma L(s) - [1 + 8\gamma L(s)]^{1/2}\} L'(s)}{4\gamma N L(s)^2} \quad (4)$$

where, from the definition of $L(s)$ in eq. 3b

$$L(s) = s[N^2 - (2N^2 + 2N - 1)s + (N + 1)^2 s^2 - s^{N+1} - s^{N+2}]/(1 - s)^3 \quad (5)$$

and the derivative is

$$L'(s) = [N^2 - 2(N^2 + 2N - 1)s + (N + 2)^2 s^2 - (N + 2)s^{N+1} - 4s^{N+2} + Ns^{N+3}]/(1 - s)^4 \quad (6)$$

Equations 5 and 6 hold for $s \neq 1$. The concentration parameter γ in eq. 4 is

$$\gamma = \beta c \quad (7)$$

where c is the total molar concentration of strands, bound or unbound.

The quantity obtained directly from experiment is the hypochromicity \bar{G} , rather than f . Since the contribution of a nucleotide to \bar{G} depends on the length of the bonded sequence in which it is located, we write

$$\bar{G} = \sum_{n=0}^N G_n w_n \quad (8)$$

where

$$G_n = (\epsilon' - \epsilon_n)/\epsilon' \quad (9)$$

ϵ' being the molar extinction coefficient for isolated nucleotides, and ϵ_n being that for nucleotides in sequences of n base pairs. The weighting factor w_n is the fraction of nucleotides in bonded sequences of length n . The term in G_0 in eq. 8 was assumed to be zero in our earlier calculation⁶; *i.e.*, nonbonded nucleotides were assumed to make no contribution to the hypochromicity. This assumption will be removed below. First we recall that omission of the term for $n = 0$ in eq. 8 leads to the following expression for the hypochromicity,⁶ denoted \bar{G}^*

$$\frac{\bar{G}^*}{G_\infty} = s \left\{ 1 + 4\gamma L(s) - [1 + 8\gamma L(s)]^{1/2} \right\} [L'(s) - AL'(se^{-a}) - BL'(se^{-b})] / 4\gamma NL(s)^2 \quad (10)$$

where the following numerical constants were obtained from Rich and Tinoco's¹² predicted values of G_n/G_∞ : $A = 0.18$, $a = 0.0345$, $B = 0.82$, $b = 0.392$. Equation 10 is not adequate to describe the behavior of poly-A, however, as will be shown below. We therefore write for the total hypochromicity \bar{G}

$$\bar{G} = \bar{G}^* + (1 - f)G_0 \quad (11)$$

which follows from eq. 8 since $w_0 = 1 - f$. It will be seen below that G_0 is a function of temperature for poly-A, a fact which adds to the complicated temperature dependence of \bar{G} through s . Also it will be seen that eq. 11 adequately describes the observations of Fresco, *et al.*¹

The possibility that aggregates of more than two strands occur in poly-A has been ignored in the above treatment. This turns out to be justified for the monodisperse oligomer case, as will be shown in section IV.

The Nonstaggering Zipper Model. This model allows only those species in Figure 1 in which the chain ends are opposite each other. There is no *a priori* justification for this model for poly-A, but we include it in this study in order to test the sensitivity of our conclusions to assumptions made about the distribution of species. The equations for f and \bar{G}^*/G_∞ were derived previously⁶ for this case and are identical with eq. 4 and 10, respectively, except that $J(s)$ is substituted for $L(s)$, where

$$J(s) = s[N - (N + 1)s + s^{N+1}] / (1 - s)^2 \quad (12)$$

and

$$J'(s) = [N - (N + 2)s + (N + 2)s^{N+1} - Ns^{N+2}] / (1 - s)^3 \quad (13)$$

The All-or-None Model. This model allows only the species A and B_N where B_N represents the completely bonded species. This model is expected *a priori* grounds to be less satisfactory than either of the zipper models, but is considered in this study because the degree to which it approximates the true state of affairs is of some interest. A single equilibrium constant k determines the behavior of the system

$$k = [B_N]/[A]^2 \quad (14)$$

We then have

$$f = [1 + 4kc - (1 + 8kc)^{1/2}] / 4kc \quad (15)$$

and, from eq. 8

$$\bar{G}/G_\infty = f \{ 1 - A \exp[-a(N - 1)] - B \exp[-b(N - 1)] \} + (1 - f)G_0/G_\infty \quad (16)$$

Experimental data on hypochromicity vs. temperature therefore allow k to be determined as a function of temperature. The enthalpy ΔH_N and entropy ΔS_N for the process $2A \rightleftharpoons B_N$ may then be determined from the relationship

$$\ln k = \frac{\Delta S_N}{R} - \frac{\Delta H_N}{RT} \quad (17)$$

III. Method of Calculation

We now show how the preceding considerations may be applied to the data of Fresco, *et al.*,¹ on the temperature dependence of hypochromicity in oligo-A samples. The following two temperature-dependent factors in the theory will be considered.

(i) The stability constant s will have the temperature dependence given by

$$\ln s = \frac{\Delta S_1}{R} - \frac{\Delta H_1}{RT} \quad (18)$$

where ΔS_1 and ΔH_1 are the entropy and enthalpy, respectively, in units per mole of nucleotide pairs, for reaction 2. It will be assumed that the enthalpy for reaction 1 is also equal to ΔH_1 , so that β may be regarded as independent of temperature. This assumption would be correct, for example, if the enthalpy change were entirely due to hydrogen bonding between paired bases. If there is a significant contribution to ΔH_1 from "stacking" interactions between adjacent bases, then β would be dependent on temperature. The error in the analysis resulting from possible stacking interactions is likely to be small, as will be shown in section IV.

(ii) Our original attempts to compare theory with experiment for the oligo-A system showed poor agreement, particularly for the shorter chains. The nature of the discrepancy was such as to lead us to propose that the extinction coefficient of nonbonded bases ϵ_0 is a slowly varying function of temperature, $\epsilon_0(T)$. The empirical form in eq. 19 suggested itself. By

$$\epsilon_0(T) = \epsilon_0(\infty)e^{-\tau/T} \quad (19)$$

plotting the experimental data in the form $\log \epsilon_0(T)$ vs. $1/T$, the limiting slope at high temperature gave the parameter τ . It was found that a value of $\tau = 130^\circ\text{K}$. fit the limiting slopes for all chain lengths within an average deviation of 15%; hence this value has been used throughout the following calculations.

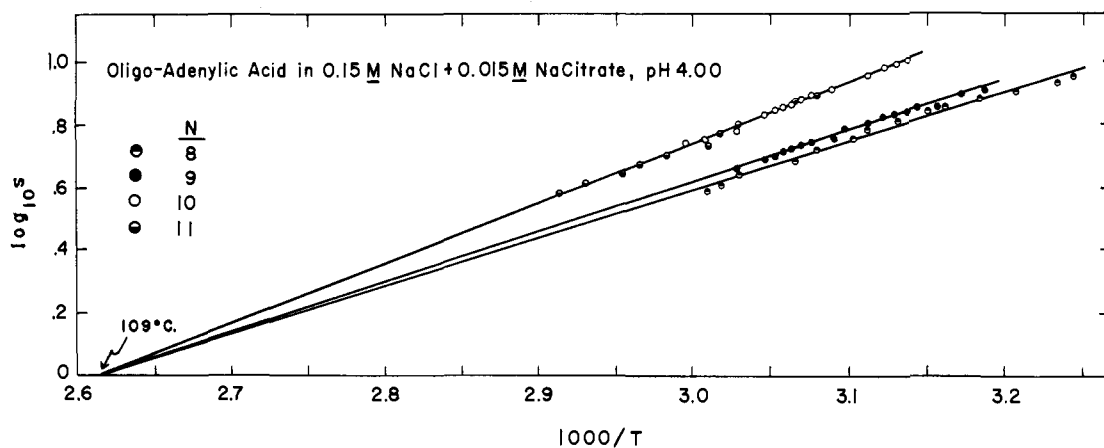


Figure 2. Temperature dependence of stability constant s for various degrees of polymerization N . The concentration parameter γ for each sample was chosen to place the horizontal intercept at $T_c = 109^\circ$. Straight lines were fitted to the data by the method of least squares (see Table I for thermodynamic quantities).

Since the hypochromicity appears to be very nearly zero at 90° (363°K.) for oligo-A, \bar{G} is obtained experimentally as

$$\bar{G} = 1 - \epsilon(T)/\epsilon(363) = 1 - A(T)/A(363) \quad (20)$$

where $A(T)$ is the absorbance of the sample at temperature T . The contribution from nonbonded nucleotides is therefore calculated as

$$G_0 = 1 - \epsilon_0(T)/\epsilon_0(363) = 1 - \exp[\tau/363 - \tau/T] \quad (21)$$

The primary task is to determine values of ΔH_1 , ΔS_1 , and β that give the best agreement between eq. 11 and 20 for \bar{G} . To simplify the task, we recognize that $s = 1$ at some temperature T_c , and hence, from eq. 18

$$\Delta S_1 = \Delta H_1/T_c \quad (22)$$

According to the simple Ising model¹⁴ for an equilibrium of this type in very long chains, T_c is the temperature at which $f = 0.5$. We therefore take $T_c = 382^\circ\text{K.}$ (109°C.), since this is the midpoint of the transition for high molecular weight poly-A, as found¹ by extrapolating T_c values obtained between pH 4.18 and 5.10 to pH 4.00. (This choice of T_c may be subject to some question because the degree to which poly-A behaves like the Ising model is not known. The treatment by Zimm⁵ for very long two-stranded molecules shows that $f > 0.5$ at T_c . On the other hand, our calculations¹⁵ for very long molecules with random breaks in the strands show that $f < 0.5$ at T_c . Since some features of both models are likely to be applicable to poly-A, the compromise adopted above seems the most reasonable approach.) In what follows, we therefore regard ΔH_1 and β as the only adjustable parameters.

The procedure by which ΔH_1 and β are obtained from the experimental values of \bar{G}/G_∞ for a particular N is, briefly, as follows. The calculations for this work were carried out on the IBM 7090 and 7094 digital computers at the Columbia University Computer Center.

The quantity \bar{G}/G_∞ is a function of s and γ , according

(14) J. Applequist, *J. Chem. Phys.*, **38**, 934 (1963).

(15) V. Damle and J. Applequist, Abstracts, 145th National Meeting of the American Chemical Society, New York, N. Y., Sept. 1963, p. 28C.

to eq. 10 and 11. For trial values of γ , then, s is obtained at each experimental temperature by computer solution of eq. 10 and 11. ΔH_1 is then obtained by a least-squares fit of the data to eq. 18. The correct value of γ is selected as the one for which $s = 1$ at $T = 382^\circ\text{K.}$, by extrapolation from the experimental temperature range. A rapidly converging trial and error technique was used to find this value of γ .

Since G_0 is an empirical function of temperature, T must be eliminated from eq. 11 to solve for s . To accomplish this we have used a method of successive approximations, whereby a trial value of ΔH_1 was assumed, so that T could be eliminated by eq. 18. The value of ΔH_1 subsequently obtained by solution of the equations was then used as a second trial value, and the entire calculation was repeated. Repetition was continued until successive values of ΔH_1 were found to differ by less than 0.1 kcal. Depending on the selection of trial values, the final results were achieved in most cases with two to five successive approximations.

For the all-or-none model, eq. 15 and 16 were solved for k at each experimental temperature, using the successive approximation method to eliminate T from the G_0 term. ΔH_N and ΔS_N were then obtained by a least-squares fit of the data to eq. 17.

IV. Results

The data of Fresco, *et al.*,¹ on oligo-A solutions were obtained at pH 4.00 in 0.15 M NaCl + 0.015 M sodium citrate. \bar{G} was measured at 257 m μ in all cases. From high molecular weight poly-A, G_∞ was found to be 0.335.

Thermodynamic Quantities. In Figure 2, $\log s$ obtained for the staggering zipper model for oligo-A is plotted against $1/T$ from the experimental data on oligomers of $N = 8, 9, 10$, and 11. Only those data from the central region of the transition were used in each case, as the errors in $\ln s$ become large where $d\bar{G}/d \ln s$ is small. The data shown were obtained for the values of γ giving $T_c = 382^\circ\text{K.}$ To illustrate the sensitivity of the results to variations in γ , the data for $N = 10$ obtained from three trial values of γ are shown in Figure 3. Since the slope of these lines gives ΔH_1 , it is seen that this quantity is extremely insensitive to errors in γ . The data from the nonstaggering and

Table I. Thermodynamic Parameters^a for the Helix-Coil Equilibrium in Oligoadenylic Acid in 0.15 M NaCl + 0.015 M Sodium Citrate (pH 4.00)

N	c, μM	Staggering zipper			Nonstaggering zipper			All-or-none	
		ΔH ₁ , kcal.	ΔS ₁ , cal./deg.	β, l./mole	ΔH ₁ , kcal.	ΔS ₁ , cal./deg.	β, l./mole	ΔH _N /N, kcal.	ΔS _N /N, cal./deg.
8	5.08	-7.01	-18.3	2.8 × 10 ⁻²	-6.57	-17.2	8.3 × 10 ⁻²	-5.89	-15.6
9	3.51	-7.33	-19.2	8.6 × 10 ⁻³	-6.86	-17.9	2.8 × 10 ⁻²	-6.20	-16.6
10	4.47	-8.80	-23.0	2.2 × 10 ⁻⁴	-8.33	-21.8	7.8 × 10 ⁻⁴	-7.68	-21.2
11	5.84	-8.75	-22.9	4.3 × 10 ⁻⁴	-8.19	-21.4	1.7 × 10 ⁻³	-7.39	-20.1
Average ^b		-7.98	-20.9	2.2 × 10 ⁻³	-7.48	-19.5	7.5 × 10 ⁻³	-6.59	-18.3

^a All thermodynamic quantities are in units per mole of base pairs. ^b The averages for β are taken from the arithmetic mean of log β. Simple arithmetic means are given for the other quantities.

all-or-none models are not shown, but it may be of interest to note that this type of plot gave equally good straight lines for these models.

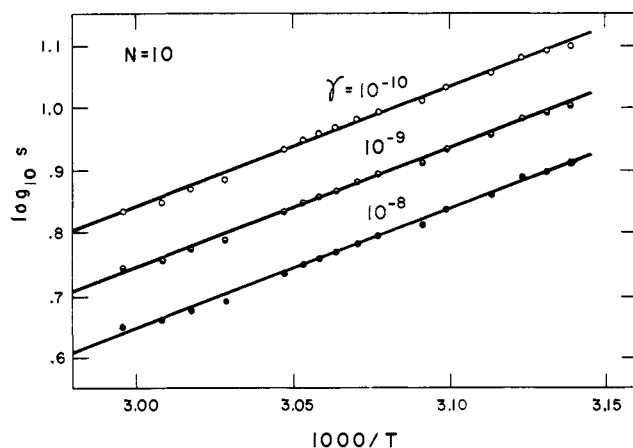


Figure 3. Temperature dependence of stability constant *s* for decaadenylic acid, using trial values indicated for the concentration parameter γ .

The experimental data for $N < 8$ did not yield reliable results by our computation procedure, primarily because the hypochromicity has a large contribution from G_0 , for which we have only approximate empirical values.

The thermodynamic quantities obtained for the longer oligomers are shown in Table I. The results are given for all three of the models discussed above. For the all-or-none model, the quantities $\Delta H_N/N$ and $\Delta S_N/N$ are in units per mole of base pairs and therefore correspond to ΔH_1 and ΔS_1 for the zipper models. The variation of the various parameters with chain length for any one assumed model is probably the result of systematic errors in the data for each sample. Therefore the averages in Table I are the best values for each model. As mentioned in section II, the staggering zipper model is expected to be the most accurate representation of oligo-A, and therefore the thermodynamic quantities obtained for this model are to be regarded as the most nearly correct ones. The other values are listed because they show that the interpretation of the experimental data is only moderately sensitive to assumptions made about the model. The reasons for this will become clear when the population analysis is discussed in section IV.

As a test of the validity of the theory and of the thermodynamic quantities given in Table I, all of the

experimental data are shown in Figure 4 superimposed on the theoretical curves for the staggering zipper model, given by eq. 11. The values of ΔH_1 , ΔS_1 , and β for these curves are the averages for the staggering model given in Table I. The only parameters that are varied among the curves are the known values of N and c . Figure 4 illustrates the following successes of the theory. (i) The curves for $N \leq 6$ are satisfactorily predicted from parameters obtained from the data for the longer chains. (ii) The plateaus in the theoretical

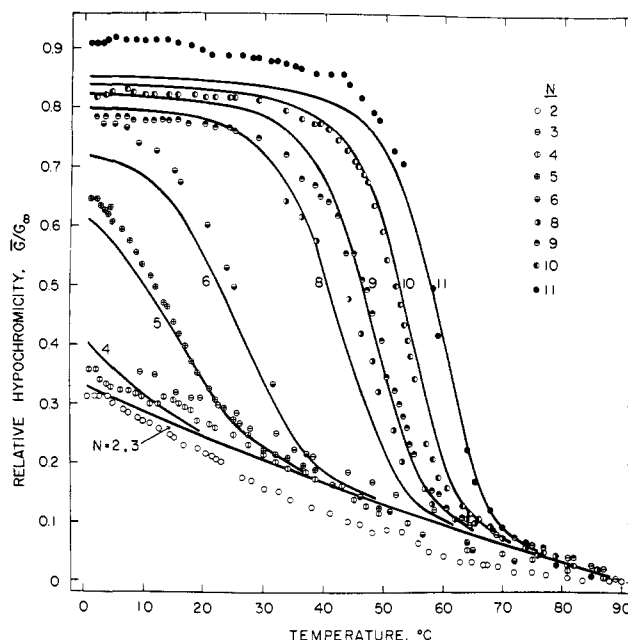


Figure 4. Relative hypochromicity curves at 257 mμ for oligoadenylic acids in 0.15 M NaCl + 0.015 M sodium citrate, pH 4.00. Solid curves are calculated for the staggering zipper model using $\Delta H_1 = -7.98$ kcal., $\Delta S_1 = -20.9$ cal./deg., $\beta = 2.2 \times 10^{-3}$ l., all per mole of base pairs. The only parameters that differ for the various curves are the chain lengths N and the concentrations, which are listed in Table II.

curves at low temperatures are within about 5% of the plateaus in the experimental data, where these are clearly observable, *i.e.*, for $N = 9, 10,$ and 11 . This suggests that the chain-length dependence of hypochromicity predicted by Rich and Tinoco¹² is substantially correct, and justifies our use of their predictions (in approximate form) in calculating average hypochromicities. We have not attempted to modify the numerical constants describing this effect (eq. 10) to give a better fit, as the observed discrepancies are

probably not beyond the limits of systematic errors. The improvement in thermodynamic quantities in any case is not likely to be significant. (iii) The gradual slope of the curves at the high temperature end, as opposed to the low temperature plateaus, is seen in both theory and experiment. This effect is attributed to the temperature-dependent hypochromicity of non-bonded nucleotides, and the observed agreement indicates that the empirical description of the effect by eq. 19 is satisfactory. The significance of the effect is emphasized by the fact that for $N = 2$ no appreciable degree of base pairing is predicted over the observed temperature range, yet the hypochromicity is large.

It is of some interest to note that theoretical curves for the nonstaggering model, using the average thermodynamic quantities for this model in Table I, coincide almost exactly with the solid curves shown in Figure 4. The agreement with experiment, therefore, does not serve as a criterion for distinguishing between the two zipper models. The argument for the staggering model remains an *a priori* one. On the other hand, the curves for the all-or-none model, calculated using the average parameters in Table I, showed significant disagreement with experiment for $N = 4, 5,$ and $6,$ being displaced toward higher temperature several degrees in each case. This can be attributed to errors inherent in the all-or-none model.

Variability in Thermodynamic Quantities. The variations in the quantities listed in Table I for various N raise questions as to whether these should be regarded as constants. However, we will show that the variations of $\Delta H_1,$ $\Delta S_1,$ and β stem from the same source, and that these are not, therefore, a reflection of a real dependence upon $N.$ If the data fit the assumptions of this treatment ideally, all of the points in Figure 2 would fall on a single straight line. A closer coincidence of the observed lines can, in fact, be brought about by using the same β for all $N,$ although there remain variations in the slopes. Systematic errors in the slopes would therefore account for the variability of all of the parameters as determined by our procedure. By a straightforward derivation, it can be shown that the error $\delta \log \beta$ is related approximately to the error $\delta \Delta H_1$ by

$$\delta \log \beta \cong -N \log s' \delta \Delta H_1 / \Delta H_1$$

where s' is the value of s at the one point on the experimental line where there is no systematic error. A reasonable value of $\log s'$ would be around unity, and since ΔH_1 falls within the range -8.0 ± 0.8 kcal./mole of base pairs, the limits of $\delta \log \beta$ should be near $\pm 1.$ Therefore, the observed variations in β by a factor of ten from the average can be attributed to the same systematic errors as are present in $\Delta H_1.$ The variation of ΔS_1 in the range -21 ± 2 cal./deg./mole of base pairs obviously has the same origin.

Two sources of error can be mentioned to account for these variations: (i) discrepancies between theoretical and experimental values of $G_n/G_\infty,$ leading to errors in the slopes of the order of 5%, and (ii) contamination of the sample by other ultraviolet-absorbing materials, which could not be ruled out in all cases,¹ and which could have led to spurious broadening of the melting curves. Any true dependence of the parameters on N would therefore not be distinguishable in these data,

and we conclude that our model correctly represents the behavior of the system to within reasonable limits of experimental error.

We have assumed throughout that the parameters are independent of temperature, although this requires closer scrutiny. It was noted in section III that if ΔH_1 has a significant contribution from stacking interactions, then β is temperature dependent. We may assume, for example, that all of ΔH_1 is due to stacking, in which case $RT^2 d \ln \beta / dT = +8.0$ kcal./mole of base pairs). We find on recalculating the curves of Figure 4 for $N = 9, 10,$ and $11,$ taking β to be 0.0022 l. at $50^\circ,$ but varying with temperature according to this assumption, that the effect is to reduce the slopes of the theoretical curves by 10% at the midpoints. Since this slope is approximately proportional to ΔH_1 (see Addendum), then the "true" values of ΔH_1 would be 10% larger (in magnitude) than was found assuming β is constant. This is a relatively small error, and it is unlikely that the stacking effect would greatly increase the over-all limits of uncertainty.

The variation of G_0 with temperature may result from a temperature dependence of the structure of the single strand, in which case ΔH_1 and ΔS_1 would also be expected to be temperature dependent. For example, if we assume that the single strand at very low temperature forms a secondary structure whose hypochromicity G_0/G_∞ is 0.4 (a reasonable lower limit), then we can estimate the enthalpy of formation of this structure from the random coil ΔH_s from $dG_0/dT,$ assuming in addition that the structure melts noncooperatively. The result is $\Delta H_s = -2.0$ kcal./residue-mole. This is an upper limit (in magnitude), since cooperative effects would act to lower the estimate, as would also a higher value of G_0/G_∞ at the low temperature limit. The data used to determine ΔH_1 for a given sample cover a 20° temperature interval; from the change in G_0 we estimate that the variation in enthalpy of the single strand over this interval due to its structural change is 0.35 kcal./residue-mole (upper limit), which would then appear as a variation of 0.7 kcal./mole of base pairs in ΔH_1 over $20^\circ.$ One would then expect to see distinguishable curvature in Figure 2. Since none is obvious, we conclude that the temperature dependence is probably smaller than this, although not necessarily too small to create discrepancies between more precise experimental data and the theoretical melting curves of Figure 4.

Calculation of $f.$ The fraction of bases paired, $f,$ is the quantity that has generally been sought in studies of hypochromicity in nucleic acids, and, in fact, it is often assumed that $f = \bar{G}/G_\infty.$ For the oligo-A systems, however, we have seen that f is not simply related to $\bar{G}/G_\infty.$ It is therefore of interest to calculate f for the samples studied, using eq. 4 and the average thermodynamic parameters for the staggering model in Table I. The results are shown in Figure 5. Comparison with Figure 4 shows that f is zero over considerable temperature ranges where \bar{G}/G_∞ is significantly greater than zero, particularly for the shorter chains. As noted above, this results from the temperature dependence of $G_0.$ The temperatures T_m at which $f = 0.5$ are listed in Table II. These differ from T_m values estimated directly from hypochromicity curves by $1-3^\circ,$ depending on the approximations made in

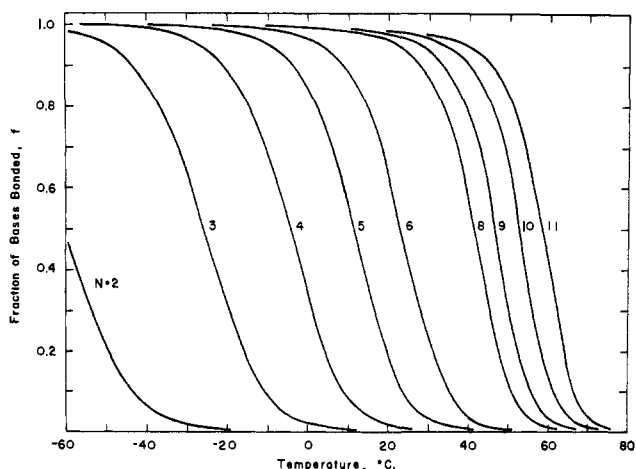


Figure 5. Fraction of bases paired, f , for the staggering zipper model, using the same parameters as for Figure 4.

the latter case. (See Addendum for a discussion of the significance of T_m .)

Population Analysis. It is possible now to calculate the relative populations of the various species throughout the helix-coil transition range. We consider here

Table II. Temperature of Midpoint of Transition for Oligoadenylic Acids in 0.15 M NaCl + 0.015 M Sodium Citrate (pH 4.00)

N	$c,^a$ μM	$T_m,$ $^{\circ}C.$
2	23.8	-60.5
3	16.0	-25.8
4	8.75	-4.5
5	7.24	11.8
6	5.34	23.4
8	5.08	41.4
9	3.51	46.9
10	4.47	53.2
11	5.84	58.5

^a Determined by spectrophotometric analysis of stock solution hydrolyzed to monomer by alkali.¹

the fraction F_n of strands in species with n bonded base pairs. From the formulas of section II, this is readily found to be, for $n = 0$

$$F_0 = \{[1 + 8\gamma L(s)]^{1/2} - 1\} / 4\gamma L(s) \quad (23a)$$

and for $1 \leq n \leq N$

$$F_n = 2\gamma(N - n + 1)^2 s^n F_0^2 \quad (23b)$$

The results of this calculation for $N = 11$ are shown in Figure 6. Here β was taken as 2.2×10^{-3} l./mole and $c = 5.84 \mu\text{moles/l.}$ It is seen that the predominant species throughout the transition are the single strands and the almost completely bonded helices. The existence of significant numbers of helical species with one or two base pairs nonbonded is the only reason for the differences in results obtained for the zipper models and the all-or-none model (Table I) and accounts for the fact that these differences are small (though significant). The results for other N are qualitatively similar to those in Figure 6. For $N = 2$, $F_1 \cong 0$ throughout the transition when $c = 23.8 \mu\text{moles/l.}$; *i.e.*, the reaction is essentially all-or-none for this case. These results sug-

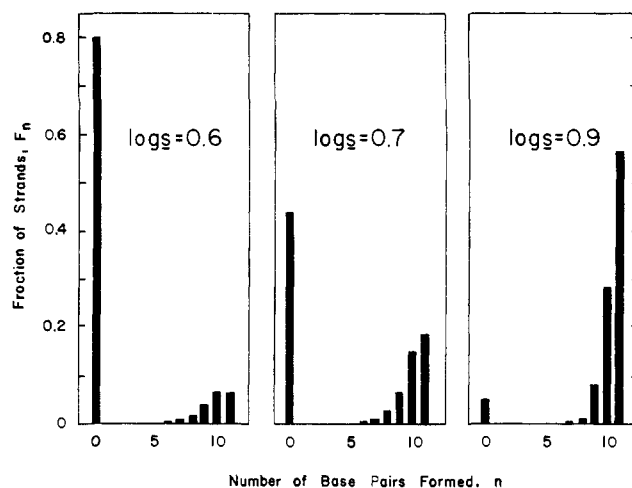


Figure 6. Population analysis for the staggering zipper model with $N = 11$, $\gamma = 1.3 \times 10^{-9}$, corresponding to undecaadenylic acid under the conditions of this study.

gest that any experiment designed to determine directly the species present in the equilibrium is likely to lead to the conclusion that an all-or-none process is taking place. Such evidence would now be understandable in terms of a reasonable molecular model.

It is worth noting that the population analysis in Figure 6 is qualitatively similar to that calculated by Schellman¹⁶ for a hypothetical polypeptide case, also using the zipper model. The underlying causes are different in the two cases, however; *e.g.*, there is no concentration effect in the polypeptide case, while the distribution found here results, at least in part, from the extreme dilution of the system.

Multiple-Stranded Species. It has been assumed throughout the above treatment that the only species present are those consisting of one or two strands. It is known, however, that higher aggregates of poly-A form under some conditions.² We are now in a position to make some predictions about the amounts of such aggregates in the oligo-A systems. Let us consider only species formed from three strands, such as that illustrated at the bottom of Figure 1. Let C be the generic symbol for such species, and let n_1 and n_2 be the numbers of base pairs in the two helical sections in a given species. The equilibrium constant K_3 for the reaction $3A \rightleftharpoons C$ is

$$K_3 = \frac{\beta^2}{2} \sum_{n_1=1}^{N-1} \sum_{n_2=1}^{N-n_1} (N - n_1 + 1)(N - n_2 + 1)(N - n_1 - n_2 + 2)(N - n_1 - n_2 + 1) s^{n_1+n_2} \quad (24)$$

We have evaluated K_3 by direct summation on the computer, and by straightforward calculation have found the fraction F' of strands in C for all of the oligo-A systems studied.¹⁷ F' is plotted against f (from eq. 4) in Figure 7. It is seen that F' is never greater than about 10^{-6} and is therefore negligible. It seems likely, then, that the total amount of higher aggregates is

(16) J. A. Schellman, *J. Phys. Chem.*, **62**, 1485 (1958).

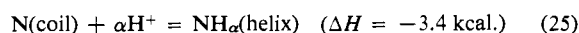
(17) Since completion of this work it was discovered that eq. 24 underestimates K_3 , and likewise F' , by a factor between four and eight, through failure to consider that the factor of $1/2$ in the sum should appear only in those terms for which $n_1 = n_2$ and that all species C have a symmetry number of unity, the latter giving rise to an additional factor of four in K_3 , according to our convention regarding β . This error does not affect our qualitative conclusions.

negligible at the concentrations in the micromolar range. The fact that F' vanishes at both extremes of the transition, as seen in Figure 7, is consistent with general considerations. As f approaches unity, C could be imagined to disappear in favor of higher aggregates, were it not for the fact that all of the bases can be paired at a lower free energy if only two-stranded species are present at $f = 1$. This argument depends on two additional assumptions: (i) the single strands are all identical in length, and (ii) a base pair is bonded primarily by interactions within the pair, rather than by "stacking" interactions with neighboring base pairs. Precautions were taken by Fresco, *et al.*,¹ to ensure (i) in their experiments. Assumption (ii) seems reasonable in view of the extensive hydrogen bonding (four bonds per base pair) and electrostatic linkages (two) proposed for the poly-A helix.³ We conclude, therefore, that the two-stranded model chosen for the oligo-A system is a realistic one.

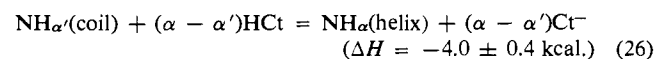
Effects of Concentration. One of the predictions of the theory that has not yet been adequately tested by experiment is the effect of concentration on the "melting" curve. Some numerical calculations of this effect for the nonstaggering zipper model showed that it is likely to be very significant in the range of N studied here.⁶ For the low concentrations generally of interest, the quantity $(\partial \ln s / \partial \ln c)_T$ is roughly constant over a wide range of f and c for a given N , and is -0.3 for $N = 3$ and -0.1 for $N = 10$, on the basis of theoretical curves for the staggering zipper model. Using the data in Table I one calculates, then, that an increase in c by a factor of ten would shift the melting curves to higher T by 10° for $N = 3$ and 7° for $N = 10$ (see Addendum for further discussion of this effect).

V. Discussion

Comparison with Calorimetric Data. The calorimetric enthalpy of helix formation in poly-A in 0.1 M KCl has been reported by Rawitscher, Ross, and Sturtevant.¹⁸ They measured the heat evolved as HCl was added to a solution of poly-A initially at pH 7, and thereby obtained the enthalpy of the reaction



where N represents a nonprotonated nucleotide residue and α is the fraction of adenine groups protonated at pH 4.00 in the helical form. In the buffered system studied by Fresco, *et al.*,¹ a somewhat different process occurs, namely



where α' is the fraction of adenine groups protonated at pH 4.00 in the randomly coiled form, and HCl represents citric acid. (For simplicity we omit explicit notation of the fact that citric acid has three acidic protons, all of which play some role in this reaction.) The ΔH for eq. 26 is in units per residue-mole and is therefore one-half of ΔH_1 . It is probably a coincidence that the enthalpies for (25) and (26) are so nearly the same, since different ionization processes are involved and different added salts were present in the experiments performed. To remove the effect of

(18) M. A. Rawitscher, P. D. Ross, and J. M. Sturtevant, *J. Am. Chem. Soc.*, **85**, 1915 (1963).

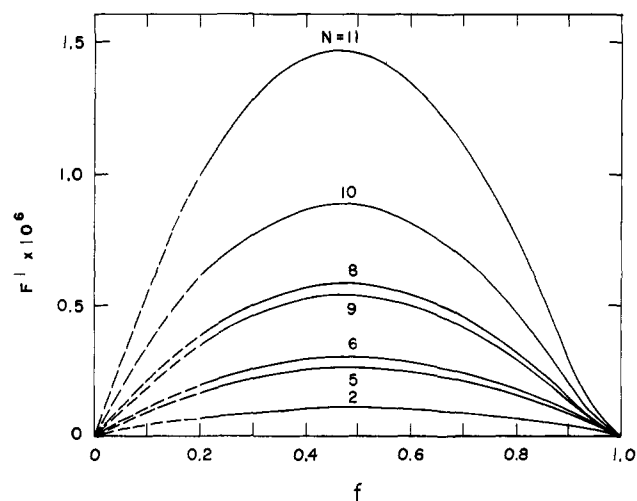


Figure 7. Fraction F' of strands associated in three-stranded species for various chain lengths at the concentrations listed in Table II. Numerical calculations were carried out only in the regions indicated by solid curves.

ionization, Rawitscher, *et al.*,¹⁸ took into account the heat of ionization of the helix, giving a corrected value of $\Delta H = -1.55$ kcal./residue-mole. It is to be noted that this value is for the enthalpy of base pairing at $\alpha = 0$, *i.e.*, for a situation where the base pair is unstable. The same quantity would be obtainable from reaction 26 if the enthalpies of ionization were known for adenine in both the helix and random coil and for citric acid. It is doubtful that a meaningful comparison could be made from available data, however, and we will not attempt it here. Suffice it to note that the extent of proton transfer in eq. 26 is considerable, since α is about 0.9, judging from the titration curve for poly-A,² while α' is about 0.4, assuming a pK_a of 3.8, as found for a number of adenosine phosphates.¹⁹ Therefore, the ΔH given for eq. 26 does not correspond to a fixed degree of protonation of adenine, and, while it does not measure the bonding forces in the helix, it does correctly measure the total energetic contribution to the stability of the helix under the conditions of the experiment.

Hypochromism. The dependence of hypochromicity on chain length was shown most clearly in Figure 4 for the 9-, 10-, and 11-mers. Although the maximum hypochromicity was not observed for shorter chains, the agreement with the theoretical curves shows that chains at least as short as the 5-mer have the approximate relative hypochromicities predicted by Rich and Tinoco.¹² This conclusion is not likely to be altered by further refinements of the model for the helix-coil equilibrium, since it has been shown that the curves are insensitive to details of the model. The feature of Tinoco's theory⁷ which leads to this chain-length dependence is the interaction between neighboring bases which falls off approximately as the inverse cube of the distance between bases, *i.e.*, in the manner of a dipole-dipole interaction. However, for the range of chain lengths studied here, it is unlikely that the dipole-dipole interaction could be distinguished from any other interaction which is primarily a nearest-neighbor

(19) D. O. Jordan, "The Chemistry of Nucleic Acids," Butterworths, Washington, D. C., 1960, p. 137.

effect, such as π -orbital overlap. The results are nevertheless inconsistent with the notion that the hypochromicity could arise largely from interactions within single base pairs. Since Nesbet's recent revival of this idea was based on calculations for only adenine-thymine and guanine-cytosine pairs,⁹ our observations do not directly conflict with his theory, although to reconcile them it would be necessary to demonstrate that adenine-adenine pairs interact in a significantly different way. Our conclusions are consistent with those reached by Michelson²⁰ from studies of hypochromicities of various oligonucleotides.

Acknowledgments. We are indebted to Dr. Jacques R. Fresco for his cooperation and encouragement during the course of this study, both by making data available to us in advance of publication and through numerous stimulating and informative discussions. This investigation was supported by a Public Health Service research grant (GM-10882) from the Division of General Medical Sciences, Public Health Service.

Addendum

A number of useful relationships which are valid at infinite dilution can be derived for the models discussed here. Since $s \gg 1$ in the transition region at high dilution, the functions $L(s)$ and $J(s)$ both approach s^N in this limit, and the system behaves in an all-or-none fashion. Equation 4 then reduces to

$$f = \frac{1 + 4\gamma s^N - (1 + 8\gamma s^N)^{1/2}}{4\gamma s^N} \quad (27)$$

The midpoint of the transition, $f = 1/2$, occurs at $\gamma s^N = 1$. At this point the sharpness of the transition is

$$(df/d \ln s)_m = N/6 \quad (28)$$

where the subscript m refers to the midpoint. From eq. 18 this becomes

$$(df/dT)_m = N\Delta H_1/6RT_m^2 \quad (29)$$

Thus ΔH_1 can be evaluated directly from the slope of the melting curve.

The hypochromicity is simply interpreted in this limit, since $\tilde{G}^* = fG_N$; thus, eq. 11 can be solved for f from experimental quantities.

(20) A. M. Michelson, *Biochim. Biophys. Acta*, **55**, 841 (1962).

The rate at which the melting curve shifts with concentration is measured by the following derivative, which is simply evaluated from the condition at constant f , $\gamma s^N = \text{constant}$.

$$(\partial \ln s / \partial \ln \gamma)_{f,N} = -1/N \quad (30)$$

Numerical calculations for the staggering zipper model show agreement with eq. 30 to within 5–10% for γ in the range 10^{-4} – 10^{-9} . The temperature shift is given by

$$(\partial T / \partial \ln \gamma)_{f,N} = -RT^2/N\Delta H_1 \quad (31)$$

Integrating eq. 31 between concentrations c and c' , assuming constant β , the relationship between the corresponding temperatures T and T' at some fixed f is

$$\frac{1}{T'} = \frac{1}{T} + \frac{R}{N\Delta H_1} \ln \frac{c'}{c} \quad (32)$$

At the midpoint, the following relationship should prove useful.

$$(\partial T_m / \partial \ln \gamma)_{f,N} = -1/6(df/dT)_m \quad (33)$$

The dependence of T_m on chain length follows directly from the condition $\gamma s^N = 1$, incorporating eq. 18 and 22.

$$\frac{1}{T_m} = \frac{R \ln \gamma}{N\Delta H_1} + \frac{1}{T_c} \quad (34)$$

If ΔH_1 and γ are constant, a plot of $1/N$ vs. $1/T_m$ should be linear, and the slope and intercept determine γ and T_c .

The data on oligo-A can be expected to follow eq. 27–34 to within the accuracy of the all-or-none model. By applying eq. 29 to the data for $N = 6$ through 11, we find $\Delta H_1 = -7.53 \pm 0.3$ kcal., which is within the range of $\Delta H_N/N$ in Table I, but differs because we have used experimental values of G_N in the present case. The spread of values is seen to be somewhat reduced by this procedure. Equation 34 was applied to the same samples, using eq. 32 to correct experimental T_m 's to values at a fixed concentration of $5.00 \mu M$. By least-squares analysis, $\beta = 24 \times 10^{-3 \pm 0.5}$ l./mole, and $T_c = 105.1 \pm 4.4^\circ$, in satisfactory agreement with the value of 109° found for the high polymer. At higher dilutions the results would be expected to approach those obtained by the exact numerical treatment of the staggering zipper model.