dipole moment, again because the total change of ϵ' through the secondary dispersion region is small. From either relation, and for each of the two branched polymers, it is found that the root-mean-square type A moment per repeat unit is 0.18 D. This numerical value is exactly the same as that found previously[2] for linear poly(propylene oxides), and compares well with the observed[16] moment of 0.22 D along the C–C bond of the ring in monomeric propylene oxide.

(16) J. D. Swalen and D. R. Herschbach, *J. Chem. Phys.*, 27, 100 (1957).

Acknowledgments. Most of this work was supported at M.I.T. (1959–1961) by the U. S. Army Research Office (Durham). Later support at Dartmouth College was from the NSF. The apparatus had been constructed earlier with the aid of grants from the American Chicle Co. and Dunlop of Canada, Ltd. Fellowship support of J. J. B. by Phi Kappa Pi, the American Cyanamid Co., and the NSF is gratefully recalled.

Toward a Statistical Theory of Superhelices. The Configurational Entropy of Cyclic Molecules with Random Loops

Homer Jacobson

Department of Chemistry, Brooklyn College of the City University of New York, Brooklyn, New York 11210. *Received June 25, 1969*

ABSTRACT: Expressions for the configurational entropy of superhelix-like molecules have been developed, using a random-walk cyclic model containing N rigid links of length b, with a minimum loop size of q links. Criterion of the configurational restriction into $t + 1$ loops is chain contact, within a volume $\pi r_e^2 b$, at t points, where r_e is an averaged closure radius. Application of ring-closing theory gives approximately $S(N,q,t) = k \ln [(2t!)^{-1} N^{t+5/2} (\mathrm{LCF})^t] + k\overline{I \ln I/\bar{I}}$, where $\mathrm{LCF} = (3/2)^{3/2}\pi^{-1/2}(r_e/b)^2$, and I is Π (from $i = 1$ to $t + 1$)$(m_i)^{-1/2}/m_1 m_{t+1}$, m_i being the size of the ith loop in the molecule. The indicated averaging was performed by Monte Carlo summation over a large sample of molecules with random m_i's, yielding a function approximately equal, for $\overline{I \ln I/\bar{I}}$, to $A - Bt$, for t above 2. Computations for a λ-phage–DNA-like model with $N = 200$, $q = 2$, and $\mathrm{LCF} = 10^{-4}$ at varied t show a monotonic decrease of S with t of an order slightly higher than linear, reaching $-965k$ eu for the maximally looped structure with $t = 99$. Expressions for distributions of t in ensembles with fixed net twist are derived; it is shown that reverse twists are unlikely in many loop molecules. Loop size distributions are also generable from expressions given by the theory; hydrodynamic variables are more properly calculated therefrom than from earlier simplified superhelix models.

Recent interest in superhelices stems from their discovery in viral and mammalian nucleic acids.[1–5] Mechanistic models have been produced[6,7] to account for their sedimentation and viscosity properties. This paper describes the start of a general configurational theory of the superhelix structure ensemble, as distributions of such configurations may be expected in response to the parameters affecting the ensemble. Its main thrust is the calculation of probabilities for the possible configurations in looped molecules, and by direct inference the distributions of loop numbers and sizes, and the entropy of specific collections of loops resulting from superhelix-generating chain twists.

The presence of multiple closed-end loops and contacts of crossed sections of the primary helix chain distinguishes a superhelix from an ordinary cyclic molecule. The latter may develop a small number of loops in random fashion, but the superhelix has an average excess of loops twisted in one direction. Figure 1 shows two possible models of superhelical chains, representing the twisted loop, or "interwound" model[2,8] considered here. We reject the "toroidal" model[2,8] as a physical reality, on the basis of experience with physical models of superhelix twist, which promptly close the opened loops, to assume the interwound configuration.

We define a loop index t to be the total number of chain contact crossings in either twist direction, leading to $t + 1$ loops. In Figures 1a and 1b, the twist directions are represented to be in the same direction. In fact t is a continuous variable, but here it will be considered discrete, as a convenience of theoretical analysis. We will also regard t as a *de facto* number holding for a particular configuration at a particular time, rather than as a time average, or as the quantity of potential chain twist which would just relieve the uncoiling of the primary helix. We are then free to consider the ensemble of possible looped molecules with any specific integral t, and to compute the entropy of the ensemble. Then any needed study of collections of configuration with various t's lumped together, or integrations over varied t, can be performed.

(1) J. Vinograd, J. Lebowitz, R. Radloff, R. Watson, and P. Laipis, *Proc. Nat. Acad. Sci. U. S.*, **53**, 1104 (1965).
(2) W. Bauer and J. Vinograd, *J. Mol. Biol.*, **33**, 141 (1968).
(3) H. S. Jansz and P. H. Pouwels, *Biochem. Biophys. Res. Commun.*, **18**, 589 (1965).
(4) L. V. Crawford, *J. Mol. Biol.*, **13**, 362 (1965).
(5) A. M. Kroon, P. Borst, E. F. J. Van Bruggen, and G. J. C. M. Ruttenberg, *Proc. Nat. Acad. Sci. U. S.*, **56**, 1836 (1966).
(6) V. A. Bloomfield, *ibid.*, **55**, 717 (1966).
(7) H. B. Gray, Jr., V. A. Bloomfield, and J. E. Hearst, *J. Chem. Phys.*, **46**, 1493 (1967).
(8) J. Vinograd, J. Lebowitz, and R. Watson, *J. Mol. Biol.*, **33**, 173 (1968).

A Model for Calculation of Configurational Probabilities. We take as our model for configurational calculations an artificial but familiar one, consisting of N freely rotating rigid links of length b each, looped by t crossings representing internal points of chain contacts, and giving $t + 1$ loops in a closed cyclic molecule. We specify the loop sizes as $m_1, m_2 \ldots m_i \ldots m_{t+1}$, with $\sum_{i=1}^{t+1} m_i = N$. We consider only an unbranched loop structure such as that in Figure 1a for this simple model, leaving branched structures for later refinement. The loop sizes m_i are consequently indexed from one end of the model to the other, and there is a twofold ambiguity of indexing, depending on the end used to start counting configurations. In work to follow the m_i are considered as discrete integers, although they, too, are actually continuous. We further restrict the loop size to some minimum value q. For independent segments with zero volume, q would be 2; considering excluded volume requires a q of at least 3. We finally assume that there are no directional restrictions on the segments arising from the loop-generating chain contacts, but only the restriction of configuration due to the termination of ends of the segments. We are now prepared, with the aid of previous theory on entropy of ring closure[9] and assumptions about the closeness of chain contact, to calculate probabilities of specific configurations, and entropy of a given ensemble of molecules fitting these model parameters.

The Probability of Two-Loop Structures. Let us consider the formation of a single chain crossing, leading to a superhelix-like configuration with $t = 1$. The series of steps used is depicted in Figure 2. The single configuration and position of the contact must be specified, using the following choices, corresponding to the steps in Figure 2.

a. Choose a starting point, indicated by the large dot. Such points are treated as distinct only if they are one link apart. The quantizing of the chain in units of single links gives an absolute value of entropy dependent on the link size, but an entropy change between configurations which is independent of the link size, except for ring closures. For the latter, theory is used[9] which stems from quantization of the chain into units of this size. There are, then, N different places at which the starting point can be chosen. However, either of two starting places will give the same configuration, with m_1 and m_2 interchanged, namely the notch and dot. The choice factor is thus $N/2$.

b. Choose the size of loops 1 and 2, *i.e.*, m_1 and m_2. The notch and dot in Figure 2b fix these. This can be done in $N - 2q + 1$ ways, by choosing a point among the remaining link joints at least q units from the starting point.

c. Break a bond at the starting point. The probability of this operation is the inverse of the probability of ring closure, as given by previous theory.[9] This equals $N^{3/2}/\mathrm{RCF}$, where RCF is a geometrical ring closing factor, of value $(6/\pi)^{1/2}(r_0/b)^3$, with r_0 the average distance over which the bond can form. Since this factor cancels out in these derivations, its exact value is unimportant.

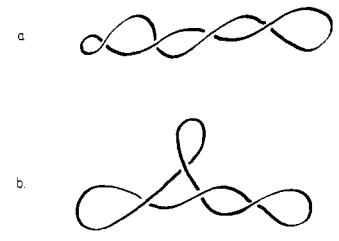

Figure 1. Two simple superhelix configurations, $t = 4$.

d. Close the loop of size m_1. This is the crucial operation in determining the major portion of the change in entropy. For what position can the broken end of the first loop be considered closed? From the chain quantization, we consider that over a length b, the loop size will be a constant, *i.e.*, m_1. And there is some effective cross-sectional area about this chain link within which the loop end must find itself. Using, as simplest expression, a constant effective cross-sectional area defined as πr_c^2 as criterion of loop closure, the required effective volume of loop closure becomes $\pi r_c^2 b$. This volume defines a loop-closing factor, or LCF, which uses the above volume in place of the $4/3\pi r_0^3$ used for RCF. Taking the ratio of these volumes, and multiplying the above expression for RCF, we calculate that LCF is $(3/2)^{3/2}\pi^{-1/2}(r_c/b)^2$. LCF does not cancel out of the entropy expression, whose major term will be seen to be proportional to ln LCF. In actual superhelices, chain torsional forces push the chain sides into contact in t places, until they meet with electronic charge cloud repulsions, with vacillation about equilibrium distances caused by thermal torsional oscillations. The quantity r_c is therefore some geometric average over the available distances, from the deepest possible penetration of squeezed-together chains to the farthest excursion from contact that thermal oscillation will occasionally allow. With understanding that the LCF factor may show some dependency on t or m_1 in a more detailed treatment, we continue with a constant one. The probability factor for the closing of the loop of size m_1 is thus taken as $\mathrm{LCF} \cdot (m_1)^{-3/2}$.

e. Close the loop of size m_2. Since the end of this free loop must make contact with the end of the broken chain, within one bond length, *i.e.*, r_0, the closing probability factor is given by ring-closing theory as $\mathrm{RCF} \cdot (m_2)^{-3/2}$.

We now define a configuration probability factor

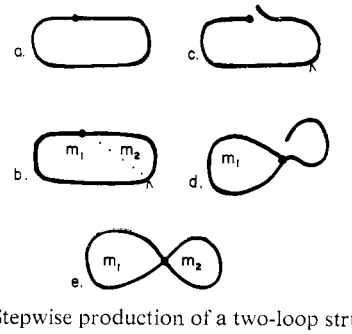

Figure 2. Stepwise production of a two-loop structure from a cyclic molecule.

(9) H. Jacobson and W. H. Stockmayer, *J. Chem. Phys.*, **18**, 1600 (1950).

for any given assignment of the quantities N, q, t, and the m_i's, in this case m_1 and m_2, to be[9] $\Omega(N,q,t,m_i)$. As given by multiplying the results of the above sequence of five operations

$$\Omega(N,q,1,m_1,m_2) = \{N(N - 2q + 1)/2\}\{\text{LCF}\}\{N^{3/2}(m_1m_2)^{-3/2}\} \quad (1)$$

The division of bracketed terms in (1) represents the three sources of entropy change in ring formation: first, the number of ways of generating loops, a large number; second, the likelihood of closing the loop with a chain contact, a small number; and last the random-walk replacement of a single large ring, size N, by a number of small rings of size m_i, a small number. Both of the latter two factors are interrelated and stem from ring-closing theory, but the first is the result of end-of-loop position geometry, and the second result of replacing a long random walk which returns by a series of shorter ones. In analogy with subsequent work, we will define $[1 - (2q - 1)/N](m_1m_2)^{-3/2}$ as I, leaving the remaining terms, $N^{7/2}$ LCF/2, as $F(N,q,1)$; we define I and F as

$$\Omega(N,q,t,m_i) \equiv F(N,q,t) \cdot I(N,q,t,m_i) \quad (2)$$

F is that part of Ω which is independent of loop sizes.

If we now ask whether Ω must be always smaller than unity, a paradox develops. Examination of (1) convinces that for large enough N, Ω can exceed unity. This seems to imply that an entropy *increase* occurs on going from the unrestricted cyclic configuration into a restricted, looped one. Reflection, however, interprets the probability factor as something other than the quantity whose logarithm gives the entropy. In the case of very large N, a molecule containing a loop is certainly more likely than one without one. A set of such large molecules not subject to chain torsions will then possess more looped structures, the loops twisting in random directions, than unlooped ones. Consideration of entropy must then involve weighting and integrating the logarithms of the probability factors over the entire ensemble of loop sizes and values of t, using the appropriate distribution functions.

We now describe the necessary summation over the ensemble of varied loop sizes leading to calculation of configurational entropy change to give the loops. We term this entropy change $S(N,q,t)$, and define it as k times the mean value of the logarithm of $\Omega(N,q,t,m_i)$ averaged over all accessible values of m_i, with Ω, the probability of each configuration, as weighting factor

$$S(N,q,t) = k\frac{\sum_{m_i}\Omega(N,q,t,m_i) \ln \Omega(N,q,t,m_i)}{\sum_{m_i}\Omega(N,q,t,m_i)} \quad (3)$$

where the summation extends over all possible values of m_i. For $t = 1$, this is a single summation over m_1 from q through $N - q$, with m_2, which equals $N - m_1$, taking on complementary values. This summation has no simple answer in closed form, although readily performed, as below, on computer.

The process of eq 3 may appear to be a duplication of the summing over all possible combinations of m_i, here m_1 and m_2, of which there are $N - 2q + 1$ possible, as already performed in the computation of Ω for eq 1, particularly in step b. Had we omitted the term, which we define as $Z(N,q,t) = \sum_{m_i}(1)$, from Ω, it would be necessary to insert $k \ln Z$ into (3), which simply averages the entropy per configuration of m_1 and m_2. It is more convenient to have put Z into Ω, as we have done here and do later. A similar problem arises with summing over t, with which we have not done likewise, as the distribution over t is in general quite sharp, and t can be taken as a system constant or near constant.

For an approximate answer in closed form, we may replace the summations in (3) by integrals. From the definitions of (2) and (3), and within the approximation of integral for sum

$$S(N,q,1) = k\int_q^{N-q} \Omega \ln \Omega \, dm / \int \Omega \, dm = k \ln F(N,q,1) + k\overline{I \ln I/I} \quad (4)$$

where the bars signify averaging I over the range of m_1 and m_2, by summing or integrating.

The averaging of $\ln \Omega$ weighted by the state probability in (3), (4), (19), (20), and (21) is correct for the usual concept of thermodynamic entropy change, particularly where we shall later roughly equate $T\Delta S$ to ΔH, the driving energy causing superhelix formation. For simple calculation of the total fraction of superhelical configurations, however, a plain summation of Ω over m_i will suffice. This sum, if performed with Ω from (1), would indeed include Z twice, and so must either be divided by $\Sigma_{m_i}(1)$, or the Z factor eliminated from (1).

Now substituting the expression for Ω and its parts from (1)

$$S(N,q,1) = k \ln \{N^{7/2}[1 - (2q - 1)/N]\text{LCF}/2\} + k\frac{\int_q^{N-q}(m_1m_2)^{-3/2} \ln (m_1m_2)^{-3/2} \, dm}{\int_q^{N-q}(m_1m_2)^{-3/2} \, dm} \quad (5)$$

From (4), the numerator of the integral-containing term is $\overline{I \ln I}$, and the denominator is $\bar{I}$, except that the m-independent correction term $[1 - (2q - 1)/N]$ is shifted back with F here.

By elementary methods of integration, replacing m_1 by m, and m_2 by $N - m$

$$\bar{I} = (N - 2q)(2/N)^2[q(N - q)]^{-1/2} \quad (6)$$

$$\overline{I \ln I} = -3/2\{[\ln (N - q) + \ln q](N - 2q)(2/N) \times [q(N - q)]^{-1/2} + 2(2/N)^2(N - 2q) \times [q(N - q)]^{-1/2} + 4(2/N)^2 [\sin^{-1} (q/N)^{1/2} - \sin^{-1} (1 - q/N)^{1/2}]\} \quad (7)$$

Whence

$$\overline{I \ln I/I} = -3/2 \{\ln (N - q) + \ln q + 2 + 4[\sin^{-1} (q/N)^{1/2} - \sin^{-1} (1 - q/N)^{1/2}][q(N - q)]^{1/2}/(N - 2q)\} \quad (8)$$

We will not reproduce the integration, as performed with parts and reference to standard tables, but the two indefinite integrals corresponding to $\bar{I}$ and $\overline{I \ln I}$,

respectively, are $(m - N/2)(2/N)^2[m(N - m)]^{-1/2}$, and $-3/2\{[\ln m + \ln (N - m)](m - N/2)(2/N)^2[m(N - m)]^{-1/2} - (2/N)^2(N - 2m)[m(N - m)]^{-1/2} - 4(2/N)^2 \sin^{-1}(m/N)^{1/2}\}$; differentiation will give back the integrands of (5); substitution of the limits gives the results cited in (6) and (7).

To achieve physical feel for results of this model, we proceed with calculation of the entropy of a two-loop ring, using (5) and (8), and representative parameters. We take $N = 200$, corresponding approximately to estimates of the link length of λ-phage DNA,[10] use $q = 2$, and a guess of 10^{-4} for LCF, corresponding to r_e of about 7 Å. Substituting these values in (5) and (8) gives $S(200,2,1) = -2.46k$ eu. The use of the integral instead of the sum gives a value for $\overline{I \ln I/I}$ (without the minute correction term) of $-11.08k$ eu, a figure within 0.1 (as seen below) of the one computed with the summation formula. The other terms include $-9.21k$ eu for ln LCF, and $17.83k$ eu for the logarithm of the balance of the quantity $F(200,2,1)$ (with the correction term). This figure would rise to $S = 0$ if r_e were as large as 25 Å; r_e could be more constrictive and give yet more negative entropy values. Fairly large numbers of two-looped structures may thus be expected in untwisted helical structures as large as λ DNA, if the other parameters are reasonably good estimates. Smaller molecules without chain torsions would have little tendency to show closed loops.

Direct addition of configuration probabilities, as mentioned above, give entropy results slightly more negative than the correct weighted-logarithm ones. The differences are only in ln I, and change the over-all entropy by about k eu for $t = 1$, rising to $3k$ eu for $t = 3$, and dropping thereafter to an insignificant portion of the total ΔS. The actual fraction of single-looped structures in the example above is about 0.03, rather than the 0.085 which would be calculated from an entropy of $-2.46k$ eu.

The Probability of Many-Looped Structures. We now proceed to the calculation of the probabilities of configurations with more than two loops. Figure 3 shows the steps in a process similar to that of Figure 2, leading from a cyclic molecule to a looped but unbranched structure containing $t + 1$ loops. We enumerate the depicted steps and the resultant probability factors.

a. Choose a starting point (large dot). As before, the probability factor is $N/2$.

b. Choose all $t + 1$ values of m_i, *i.e.*, the sizes of the loops, in order. The probability factor associated with this step is

$$\{N - (t + 1)q + t\}!/\{N - (t + 1)q\}!t! = \{N^t/t!\} \prod_{i=1}^{t} [1 - \{(t + 1)(q - 1) + i\}/N]$$

This factor, the number of ways of arranging N-numbered objects into $t + 1$ piles of contiguous numbering, none smaller than q, is a known combinatorial quantity. The author will provide a derivation, on request.

(10) J. E. Hearst and W. H. Stockmayer, *J. Chem. Phys.*, 37, 1425 (1962).

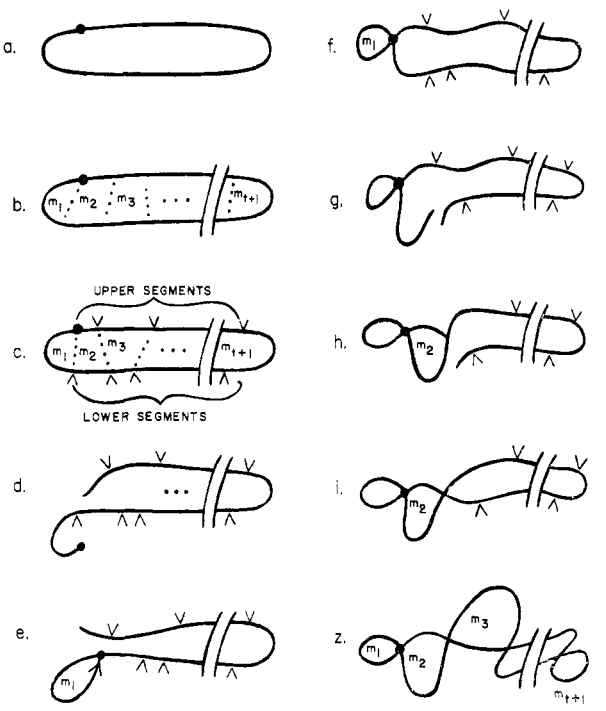

Figure 3. Stepwise production of a multi-loop structure from a cyclic molecule.

c. Set the crossing points within the loops. To any nonend loop of m_i links, there correspond $m_i + 1$ ways of forming the crossing points, with more or less links partitioned between the upper and lower segments. The notches in Figure 3c delineate the exact position of the crossings, and consequently the segment sizes for both upper and lower ones. The configurations containing zero links (none are depicted in Figure 3) for upper or lower segments are seen to be branch points containing a single loop. These are reasonable possible configurations in the model here, and have been retained in further calculations.

The crossing position on the end loops is fixed by the choice of starting point, the size of m_1, and the positions chosen for the middle-loop segments. Therefore the probability factor for the crossing point selection is the product, for the middle loops only, $\prod_{i=2}^{t} (m_i + 1)$.

d. Break the bond at the initial point. As previously, this has the probability factor of $N^{3/2}$/RCF.

e. Close the first loop, with m_1 links, at the selected crossing point. As previously, the probability factor is $\mathrm{LCF}(m_1)^{-3/2}$.

f. Close the remainder of the molecule at the selected point, and form the bond. Here the probability factor is $\mathrm{RCF}(N - m_1)^{-3/2}$.

g–i. Repeat steps d–f to form the second loop of m_2 links. The probability factors, in order, are $(N - m_1)^{3/2}$/RCF, $\mathrm{LCF}(m_2)^{-3/2}$, and $\mathrm{RCF}(N - m_1 - m_2)$.

j–z. Repeat steps d–f $t - 2$ more times, to form the remaining loops, including the final configuration with $t + 1$ loops. As a result, the probability factors multiplied together from the $3t$ steps d–z will include the following.

1. No RCF factors; they, together with any energy terms from bond breaks and reforming, will cancel out.

2. $(\mathrm{LCF})^t$. Each of the t closures of loops will bring in one such geometric factor.

3. $N^{3/2}\prod_{i=1}^{t+1} m_i^{-3/2}$. All of the factors containing an intermediate ring size, *e.g.* $(N - m_1 - m_2 - \ldots - m_i)^{3/2}$, will cancel, except the last one, $m_{t+1}^{-3/2}$, which is the last term in the $-3/2$ power product. The other t terms in the product arise in loop-closing steps such as e and h.

The over-all probability factor for the structure with loops of size m_i, *i.e.*, $m_1, m_2, \ldots m_i, \ldots m_{t+1}$, is therefore

$$\Omega(N,q,t,m_i) = [N^{t+5/2}/(2t!)](\mathrm{LCF})^t \prod_{i=2}^{t} (m_i + 1) \times \prod_{t=1}^{t} [1 - \{(t+1)(q-1) + i\}/N] \prod_{t=1}^{t+1} (m_i)^{-3/2} \quad (9)$$

The first two of the product expressions in (9) cause difficulties in visualization of effect of q limitation on loop sizes (second product expression), and of the presence of single-loop branches (first product expression) on the entropy. These can be overcome by making reasonable approximations to the expressions.

The second product expression in (9) can be closely approximated by using the average value of i, namely $(t + 1)/2$, therein. This gives $N^t\{1 - (t+1)(q - {}^1/_2)/N\}^t$ for the value of this term. The bracketed term represents the q correction for loop size on the number of ways of partitioning into loops; it is small for small t, and is small when $q \ll$ average m_i.

The first product expression in (9) can be broken up similarly to give $\prod_{i=2}^{t} (m_i) \prod_{i=2}^{t} (1 + 1/m_i)$. The first part of this can be included with the third product expression of (9), as below. Evaluation of the second part, essentially a correction term due to single-loop branches, cannot be made quite as simply as was the correction term above. Simple replacement of $1/m_i$ by its average value $(t + 1)/N$ leads to a poor estimate of the average value of $\prod_i (1 + 1/m_i)$, or to the desired weighted average of its logarithm, $\overline{\ln (1 + 1/m_i)}$, since small values of m_i have a disproportionate effect on this correction term. The averaging, as done in Appendix A, leads to a fairly complicated but closed transcendental function of N, q, and t, which we simply represent as $\overline{\ln (1 + 1/m_i)}$ in what follows; its value is tabulated below.

The over-all result of these approximations and shifts is

$$\Omega(N,q,t,m_i) \cong (2t!)^{-1}N^{t+5/2}(\mathrm{LCF})^t\{1 - (t+1)(q - {}^1/_2)/N\}^t [(1 + 1/m_i)^{t-1}] \left[\prod_{i=1}^{t+1} (m_i^{-1/2})\right] \Big/ m_1 m_{t+1} \quad (10)$$

Again we can break up Ω into terms (I) which depend on m_i, and those (F) which do not. (Although the bracketed approximation correction terms do not seem to involve m_i, we include them with the product terms at the end of the expression, since their accurate value, as in (9), does indeed depend on the values of the m_i. In either case, we incorporate the N^t term into $F(N,q,t)$ leaving the balance of the terms in either (9) or (10) for I.)

$$F(N,q,t) = (2t!)^{-1}N^{t+5/2}(\mathrm{LCF})^t \quad (11)$$

Collections of Configurations with Constant t. The set of all configurations with fixed t, as in the case integrated above for $t = 1$, must now be considered in order to estimate the entropy as a function of t. The F portion of Ω does not depend on the configuration, but the I portion must be appropriately summed. We start by finding limits for maximum and minimum values for Ω. We need only do this for I, and for this purpose we will ignore the approximation terms in (10), using the last term only. Writing this m_i-dependent term as

$$U(N,q,t,m_i) = \left[\prod_{i=1}^{t+1} (m_i^{-1/2})\right] \Big/ m_1 m_{t+1} \quad (12)$$

the method of Lagrangian multipliers immediately suggests itself. U is subject to the restriction that

$$\sum_{i=1}^{t+1} m_i = N \quad (13)$$

Then, by the customary "maximization" of such expressions

$$\partial \ln U/\partial m_i + \alpha \partial N/\partial m_i = \partial/\partial m_i \left[-{}^1/_2 \sum_{i=1}^{t+1} (\ln m_i) - \ln m_1 - \ln m_{t+1} \right] + \alpha \quad (14)$$

$$= \alpha - {}^1/_2 m_i \qquad (2 \le i \le t), \text{ or}$$

$$= \alpha - {}^3/_2 m_i \qquad (i = 1 \text{ or } i = t+1) \quad (15)$$

whence, $m_i = 3/(2\alpha)$, when $i = 1$ or $i = t + 1$; and $m_i = 1/(2\alpha)$, otherwise. Substituting these values for m_i into condition 13, $\alpha = (t + 5)/2N$. Whence, m_1 and $m_{t+1} = 3N/(t + 5)$, and all other $m_i = N/(t + 5)$. Further

$$U(N,q,t,m_i') = 3^{-3}(t + 5)/N^{(t+5)/2} \quad (16)$$

the use of m_i' indicating this extremum of choice for the loop sizes of the configuration. This extreme-entropy distribution consists of molecules with uniform center loops, but with end loops which are three times as large. While this is an esthetically satisfying result, a little contemplation, such as taking the second derivative of $\ln U + \alpha N$ and finding it negative, results in the conclusion that this is not the situation of maximum, but of minimum entropy, *i.e.*, the *least* probable loop size distribution. This reversal of the ordinary result of the Lagrangian method is due to U being a decreasing, rather than the customary increasing, function of the m_i's.

We can now infer the most probable configurations of loop size distribution for this model. The equal-sized loops being those of lowest probability, the highest probability configuration must have all loops but one of size q, and that one bearing the remainder of the links. In chains of three or more loops, the large loop, of size $N - tq$, should be an inner loop to take advantage of the $-{}^1/_2$ power. With $t = 1$, the sizes are q and $N - q$, and the m_i-dependent probability factor becomes

$$U(N,q,1,N - q,q) = q(N - q)^{-3/2} \quad (17)$$

For a molecule of three or more loops

$$U(N,q,t,m_i'') = q^{-2-t/2}(N - tq)^{-1/2} \qquad (18)$$

m_i'' here indicating the distribution of highest probability. For a distribution of small numbers of configurations, as epitomized by selection of m_i among a small number of choices, the most probable distribution as given by (17) and (18), even including its neighbors in phase space, does not represent a good approximation to the entropy. Many configurations of a nonmaxiprobable sort represent the state of the system often enough to make it necessary to sum the weighted logarithm of the probability factors over phase space. In analogy with (3) and (5), but using summation instead of integration, and indicating the nature of the summation over phase space more explicitly, we have

$$S(N,q,t) = k\frac{\sum_{m_1=q}^{N-q}\sum_{m_2=q}^{N-q-m_1}\cdots\sum_{m_i=q}^{N-q-\Sigma m_i}\cdots\sum_{m_{t-1}=q}^{N-q-\Sigma m_i}[\Omega(N,q,t,m_i)\ln\Omega]}{\sum_{m_i}(\Omega)} \qquad (19)$$

where the summations which appear as limits of summations go from 1 to $i - 1$; the summation limits of the denominator are identical with those in the numerator. For constant t, this reduces to (4), where the average here is a summation rather than an integration calculation. Substituting in (10), for constant t

$$S(N,q,t) = k\ln[(2t!)^{-1}N^{t+5/2}(\mathrm{LCF})^t] + k(t-1)\overline{\ln(1+1/m_i)} + kt\ln\left(1-(t+1)(q-1/2)/N\right) + k\frac{\sum_{m_i}\left[\left(\left\{\prod_{i=1}^{t}(m_i^{-1/2})\right\}/m_1m_{t+1}\right)\left(-1/2\sum_{i=1}^{t+1}\ln m_i - \ln m_1 - \ln m_{t+1}\right)\right]}{\sum_{m_i}\left(\left\{\prod_{i=1}^{t+1}(m_i^{-1/2})\right\}/m_1m_{t+1}\right)} \qquad (20)$$

Substituting (19) into (9),and making the same shifts of terms made in (20)

$$S(N,q,t) = k\ln[(2t!)^{-1}N^{t+5/2}(\mathrm{LCF})^t] + k\sum_{m_i}\left[\left(\prod_{i=2}^{t}(1+1/m_i)\right)\left(\prod_{i=1}^{t}\{1-[(t+1)(q+1)+i]/N\}\right)\left(\left\{\prod_{i=1}^{t+1}m_i^{-1/2}\right\}/m_1m_{t+1}\right)\times\left(\sum_{i=2}^{t}\ln(1+1/m_i)+\sum_{i=1}^{t}\ln\{1-[(t+1)(q+1)+i]/N\}-1/2\sum_{i=1}^{t+1}\ln m_i-\ln m_1-\ln m_{t+1}\right)\right] \Big/ \sum_{m_i}\left[\left(\prod_{i=2}^{t}(1+1/m_i)\right)\left(\prod_{i=1}^{t}\{1-[(t+1)(q+1)+i]/N\}\right)\left(\left\{\prod_{i=1}^{t+1}m_i^{-1/2}\right\}/m_1m_{t+1}\right)\right] \qquad (21)$$

The first two product terms in (21) and their logarithms, and the second and third additive terms in (20) represent, as noted, the correction terms for single-loop branches, and q-limited loop sizes.

Calculations and Results

Exact summation of (20) or (21) is not possible in closed form. Even if the correction terms are ignored, and the integration approximation as performed in (4) to (8) is tried, the expressions of $\overline{I\ln I}$ and $\bar{I}$ become extremely cumbersome, as soon as t increases beyond 1. With a computer, explicit summation of (20) and (21) is possible for small values of t. As t becomes reasonably large, however, even this overloads the computer program with impractical nests of DO loops. The summation was carried out, consequently, by using representative parameters, and the Monte Carlo method. The same λ-phage-like parameters were chosen as above, namely $N = 200$, $q = 2$, and LCF = 10^{-4}. Integral values of t from 1 to 15, and by fives through 50, were assigned for specific calculations and in general, random samples of 1000 were selected from the m_i phase space, as described below, with duplicate summations carried out for a number of values of t. For $t = 40$, 45, and 50, the sample size was reduced to 500. The sample m_i's were obtained by setting the computer to make the t random cuts described in step b of the process described in Figure 3, with all cuts coming within less than q (here 2) links of an already selected cut discarded; the m_i become the distances between the cuts. For each sample of a group of m_i, the value of $I(N,q,t,m_i)$, as given by the m_i-dependent product terms in (9), was computed, together with its logarithm and the product of itself by its logarithm. Summation of a large-sample $I\ln I$, divided by the summation of the corresponding I's, is the sampling equivalent of the summations specified in (21), and is accurate to the extent that the sampling process covers phase space uniformly and without bias. The requisite entropy figures were then computed simply by adding $k\ln F$ to the estimated $k\overline{I\ln I}/\bar{I}$.

The q limitation cut correction term, the second product term in (9), was computed in the rigorous form, instead of the first approximation as shown in (20). Uncorrected values of $\overline{I\ln I}/\bar{I}$, computed without inclusion of either of the first two product terms in (9), were simultaneously calculated so as to estimate the importance of such terms, and the accuracy of the approximations made in (10) for them.

Table I presents results of these calculations for a selection of the values of t. The first row gives $\ln F(N,q,t)$ with the parameters as stated. The next three rows give maximum, minimum, and Monte Carlo estimates of $\ln I(N,q,t,m_i)$. Maximum estimate comes from (17) or (18), minimum from (16), and the Monte Carlo averages as described above. Where several samples were taken cited figures represent the averages.

TABLE I
TOTAL AND CONTRIBUTORY ENTROPIES OF SUPERHELIX FORMATION FOR VARIOUS t's
($N = 200$, $q = 2$, LCF = 10^{-4})

	t							
	1	2	3	4	5	10	20	50
ln $F(N,q,t)$	8.64	4.04	−0.98	−6.27	−11.79	−41.67	−108.0	−331.5
ln I_{max}	−8.97	−4.72	−5.06	−5.40	−5.74	−7.45	−10.86	−21.02
ln I_{min}	−13.81	−15.03	−16.17	−17.25	−18.27	−22.72	−29.29	−38.79
$\overline{I \ln I/I}$	−11.00	−9.02	−9.90	−11.16	−12.01	−16.01	−22.53	−44.06
$S(N,q,t)$	−2.36	−4.98	−10.88	−17.44	−23.80	−57.68	−130.5	−375.6
$S(N,q,t)$	−3.05	−6.37	−12.96	−20.21	−27.27	−64.61	−144.4	−410.2
Calcd cor (q limitation)	−0.015	−0.05	−0.09	−0.15	−0.23	−0.85	−3.44	−24.4
Calcd cor (one-loop branch)	0.000	0.04	0.09	0.17	0.26	0.90	2.94	12.8
Calcd cor (total)	−0.015	−0.01	0.00	0.01	0.03	0.06	−0.50	−11.6
Obsd cor	−0.015	−0.06	0.00	0.03	0.12	0.21	0.00	−11.7

The following row gives the total corrected estimate of the entropy $S(N,q,t)$ obtained by adding ln F to the corrected Monte Carlo $\overline{I \ln I/I}$. The following row gives an entropy estimate for a superhelix structure, with all loops twisted in the same direction; this is estimated, as described below, by subtracting t ln 2 from the entropy of the randomly looped structure with t twists. The last four rows present the sum of the two correction factors represented by the first two product terms in (9). In the first three rows, they are calculated from the two bracketed approximation terms of (10), and (5) of Appendix A, and in the last row as the difference between the corrected and uncorrected value of $\overline{I \ln I/I}$ in the Monte Carlo calculation. All entropy values are given in multiples of k eu, *i.e.*, the table is a representation of S/k.

Figure 4 is a plot of $\overline{I \ln I/I}$, as given by (10), and calculated by the Monte Carlo method, for the cited values of t. Duplicate sample calculations, where performed, are here presented as separate points. Scatter in the Monte Carlo points from $t = 2$ on up is visible, generally decreasing with increasing t. The uncertainty of an individual sample of 1000 is about half of k eu in the value of $\overline{I \ln I/I}$; in the value of the correction terms, it is one- or two-tenths of k eu. The solid line represents the corrected values of $\overline{I \ln I/I}$, and the dashed line the uncorrected ones.

Figure 5 is a plot of S/k, using corrected calculations with $\overline{I \ln I/I}$ and mean values of all samples, for the same values of t.

A picture of the entropy changes in the twisting of such models into randomly looped or superhelically looped molecules emerges from the calculations. The major terms in the entropy expression are those appearing in F, the loop-size-independent part of S. F is given explicitly by (11), and its logarithm is of the form

$$\ln F = \ln A - Bt - \ln (t!) \qquad (22)$$

The weighted logarithm of I is a complex function, with the following features.

1. A local peak appears at $t = 1$. This is due to

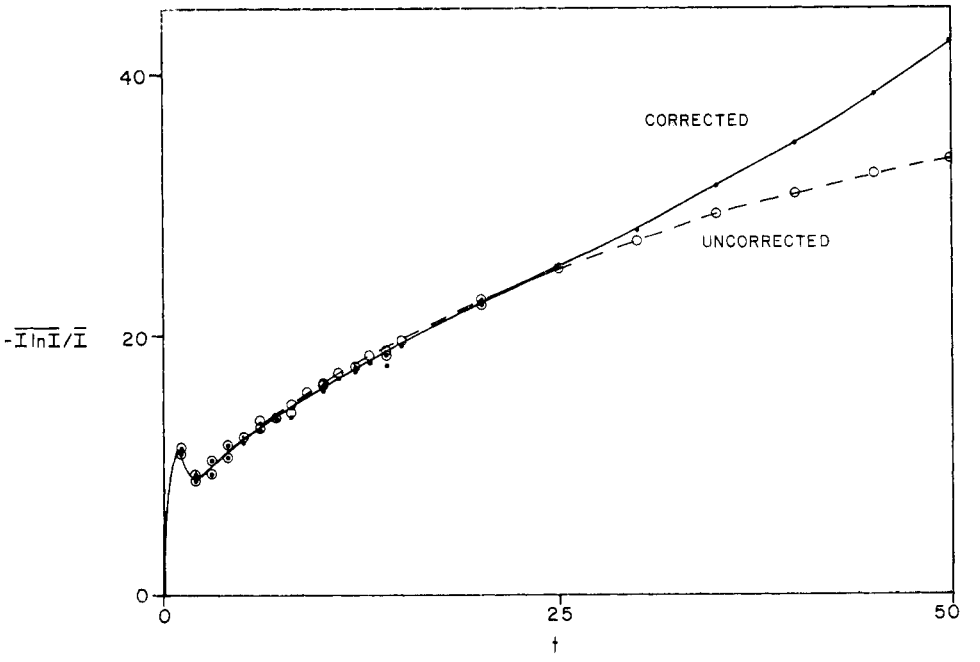

Figure 4. Mean entropy contribution of loop-size distribution dependent factor: $N = 200$, $q = 2$, LCF = 10^{-4}.

the lower probability factor associated with end loops (*i.e.*, m_1 and m_{t+1} to the $-3/2$ power), and the greater ease of formation of middle loops, for which the $-1/2$ power appears in the probability factors.

2. There is a sharp drop toward zero entropy, by definition, for $t = 0$.

3. There is a monotonic rise toward more negative entropy, for $t > 2$. Without the correction factors, the rise is somewhat slower at the higher t's. The correction factors are in opposite directions, and nearly cancel one another out, up to $t = 25$. They are tiny and negative at small t's, slightly positive at intermediate t's, and negative at t's above 20, having the effect of allowing a straight-line approximation from $t = 2$ to $t = 50$, valid within k eu, of the general form

$$\overline{I \ln I}/\bar{I} = \ln A' - B't \tag{23}$$

where B' is approximately $2/3$. Above $t = 50$, the reduction in entropy due to the q limitation predominates strongly enough to cause an upward curve to the negative entropy in Figure 4, ultimately reaching the high level calculated for $t = 99$ below. Absolute differences between the last two rows of Table I are small, and represent Monte Carlo scatter, rather than failure of the approximations used. We further note that these correction terms should be added to the maximum and minimum I terms calculated for the second and third rows of Table I; otherwise the corrected value of $\overline{I \ln I}/\bar{I}$ for high t's will appear to be less than the estimate of the minimum.

The interesting and wiggly ln I part of the entropy is nearly lost in the larger ln F, and the peak is absorbed in the general rise of negative entropy, except as a slope flex. For the value $N = 200$, B is 5.3, and a $3k$ eu drop in $-\ln I$ between $t = 1$ and $t = 2$ is more than compensated by the $5.3k$ eu rise from $-\ln F$. For a much smaller molecule, however, the effect of the two-loop entropy disadvantage may become observable.

The theory does not describe entropies at nonintegral values of t; for such values, some reasonable interpolation could be chosen, based upon empirical approximations to the curves drawn through integral t's. A total approximation based on addition of (22) and (23) is roughly valid from $t = 2$ to $t = 50$, and rather accurately for t's starting a bit higher. At low t, taking the peak in ln I and the sharp drop at $t = 0$ into account, a much more difficult function fitting is needed to replace the discrete results with a closed empirical function.

Calculations beyond $t = 50$ were not made, not merely because of their laboriousness, but because of the breakdown of approximation assumptions, and questions of the physically valid meaning in an area where most chains approach the minimum loop size q. In theory as it stands, t can take the maximum value of $N/q - 1$, here 99. It is possible to calculate the predictions of the model, howsoever unreal, at or near this value, with a different approach from that of the correction terms. At maximum t, there are only N/q possible loops, and all of the $m_i = q$. Each such loop, not including the end ones, can be formed by $q + 1$ combinations of upper and lower segments, for a total of $(q + 1)^{t-1}$ configurations of segments, leading to $N(q + 1)^{t-1}/2$ possible loop configurations. Together with other terms in Ω, *i.e.*, $(\mathrm{LCF})^t N^{3/2}/(q^{3/2})^{t+1}$, we can write

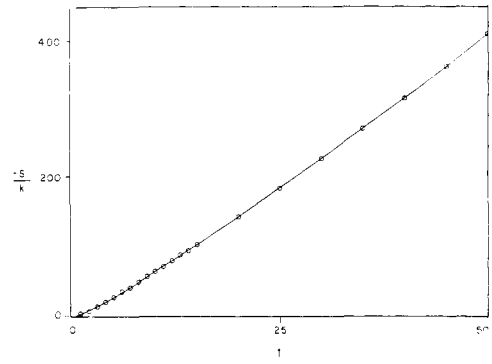

Figure 5. Total entropy of superhelix formation: $N = 200$, $q = 2$, LCF $= 10^{-4}$.

$$\Omega(N,q,t_{\max}, m_i = q) = \frac{[(\mathrm{LCF})(q + 1)q^{-3/2}]^t{}_{\max} N^{5/2}}{2(q + 1)q^{3/2}} \tag{24}$$

where $t_{\max} = N/q - 1$. As there is no m_i-dependent distribution of probability of individual configurations, the entropy is simply given by k ln Ω, and for the example of the calculations above $S(200,2,99)k = -896$ without twist entropy, as in (24), and -965 (99 ln 2 has been subtracted) for the superhelix itself. An expansion about this value of t for near-maximal values would be possible for configuration counting, but has not been done here. We presume that Figure 5 would show a fairly sharp rise near this value, although it is not much higher than one might guess from extrapolating the calculated curve (which includes the t ln 2 entropy of twist) up to $t = 99$.

Effect of Twist Direction. The calculations above apply to configurations of fixed t, without regard to the direction of twist for each loop (except as noted for Figure 5 and the sixth line of Table I). An actual superhelix is perforce twisted predominantly in one direction, resulting in the over–under pattern visible in Figure 1. Such choice of twist direction reduces the number of possible chain configurations by the factor 2^t, if twists in one direction only are allowed, and decreases the entropy by kt ln 2 eu. The sixth row of Table I includes this quantity. This entropy of twist choice is not large, and could be hidden within the uncertainty of the LCF, of which it is in fact a part.

In actual superhelices, the chain twists might not be all in the same direction. Any twisted chain observed to have t twists may belong to any of a number of families with *net* chain twist in one direction defined as the one-dimensional vector quantity **t**. Given a single value of **t**, representing a family of t's each made up of l left-hand and r right-hand twists opposing each other, we see, using a left-hand convention for net twist, that $\mathbf{t} = l - r$. It follows

$$\mathbf{t} = t - 2r \tag{25}$$

where r can vary from 0 up, and is a continuous variable, just as are t and **t**. The family defined by a single value of **t** will thus be enumerated by the sum of all

configurations satisfying '25). Defining this inclusive number of configurations as $\Omega'(\mathbf{t})$, we have

$$\Omega'(\mathbf{t}) = \sum_{r=0}^{\infty} \Omega(t)g(t) \tag{26}$$

Ω is the function defined by (10) or (9), and $g(t)$ is a weighting factor giving the fraction of configurations within the family with the particular value of r in that term of the sum. From (25)

$$\Omega'(\mathbf{t}) = \sum_{r=0}^{\infty} \Omega(\mathbf{t} + 2r)g(\mathbf{t} + 2r) \tag{27}$$

The value of $g(\mathbf{t} + 2r)$ is the fraction of configurations with $\mathbf{t}$ net left-hand twists, and $\mathbf{t} + 2r$ total twists. There are $2^{\mathbf{t}+2r}$ equivalent configurations with $\mathbf{t} + 2r$ total twists included in $\Omega(\mathbf{t} + 2r)$. Of these, $(\mathbf{t} + 2r)!/(\mathbf{t} + r)!r!$, *i.e.*, the binomial coefficient $\binom{\mathbf{t} + 2r}{r}$, have the necessary $\mathbf{t} + r$ left-hand, and r right-hand, twists. Therefore $g(\mathbf{t} + 2r)$ is $\binom{\mathbf{t} + 2r}{r} \Big/ 2^{\mathbf{t}+2r}$, and

$$\Omega'(\mathbf{t}) = \sum_{r=0}^{\infty} \binom{\mathbf{t} + 2r}{r} \Omega(\mathbf{t} + 2r)/2^{\mathbf{t}+2r} \tag{28}$$

Equation (28) gives $\Omega'(\mathbf{t})$ as a summation; legitimately it should be an integral, reflecting the continuous nature of $\mathbf{t}$, t, and r. We will continue to treat it as a summation, since first the intermediate values of the binomial coefficients are not meaningful, even with extension to γ functions, and second the correctness of the theory for nonintegral numbers of loops has not been shown for the model used. Moreover, integration of the transcendental functions in closed form is not feasible. Seen as a summation, however, (28) is approximately summable in the interesting region of intermediate $\mathbf{t}$. Here, as given by (22) and (23)

$$\Omega(\mathbf{t}) = AA'e^{-(B+B')\mathbf{t}}/\mathbf{t}! \tag{29}$$

Whence

$$\Omega(N,q,\mathbf{t} + 2r) = AA'e^{-(B+B')(\mathbf{t}+2r)}/(\mathbf{t} + 2r)! \tag{30}$$

and, from (28)

$$\Omega'(\mathbf{t}) = \Omega(\mathbf{t})2^{-\mathbf{t}} \sum_{r=0}^{\infty} 2^{-2r}e^{-2(B+B')r}(\mathbf{t}!)/(\mathbf{t} + r)!r! \tag{31}$$

Making one further assumption, valid for $\mathbf{t} \gg r$, *i.e.*, large $\mathbf{t}$, or little contribution from terms with an appreciable fraction of right-handed twists, the quotient $\mathbf{t}!/(\mathbf{t} + r)!$ becomes approximately $\mathbf{t}^{-r}$, and

$$\Omega'(\mathbf{t}) = \Omega(\mathbf{t})2^{-\mathbf{t}} \sum_{r=0}^{\infty} [e^{-2(B+B')}/4\mathbf{t}]^r/r! \tag{32}$$

In (32), we recognize the series $\sum_{r=0}^{\infty} (e^{ar}/r!) = e^{e^a}$, where a here is $-2(B + B') - \ln 4\mathbf{t}$. Thus, for the number of configurations $\Omega'(\mathbf{t})$ with a fixed net helix twist $\mathbf{t}$, relative to the number $\Omega(\mathbf{t})$ with $\mathbf{t}$ total twists

$$\Omega'(\mathbf{t})/\Omega(\mathbf{t}) = 2^{-\mathbf{t}} \exp [e^{-2(B+B')}/4\mathbf{t}] \tag{33}$$

The $2^{-\mathbf{t}}$ term in (33) reflects the selection of only one of the $2^{\mathbf{t}}$ possible configurations for the first member of the family. If only this member contributes significantly, the series terminates with the first term, and $\Omega'(\mathbf{t})/\Omega(\mathbf{t}) = 2^{-\mathbf{t}}$. In this case, the entropy given in the sixth row of Table I corresponds to $\Omega'(\mathbf{t})$, the number of configurations in the family. Where the quantity a is very negative (in the example used it is about $-6 - \ln \mathbf{t}$, and e^a is about $(200\mathbf{t})^{-1}$), the probabilities for configurations with $r > 0$ in the family become exceedingly small. Thus an ordinary superhelix of viral DNA dimensions should show little tendency for any twists to form in the counterdirection. At higher $\mathbf{t}$, the decrease of entropy with $\mathbf{t}$ becomes even more pronounced, and the likelihood of reverse twists here is even fainter. Only when the chain is exceedingly long, as $N \rightarrow (\mathrm{LCF})^{-1}$, does the chance of reverse twists at moderate or high $\mathbf{t}$'s emerge. For low $\mathbf{t}$, especially the interesting case of $\mathbf{t} = 0$, twists in both directions become possible. Here the approximations of our present treatment break down, and a more sophisticated approach to the configuration fraction would be needed to predict twist distributions. At zero $\mathbf{t}$, some symmetric distribution is present, dependent on chain stiffness and energy fluctuations.

Entropy computations for the collections with a single value of $\mathbf{t}$ are identical with those given in (20) and (21), corrected by $k\mathbf{t} \ln 2$, for the case in which reverse twists are not frequent. For low $\mathbf{t}$, or very high N, the assumption of constant t, used in formulating (20) and (21), is incorrect. In this instance, we must substitute for (20)

$$S(N,q,\mathbf{t}) = k \frac{\int_{\mathbf{t}} \sum_{m_i} \Omega' \ln \Omega' \, d\mathbf{t}}{\int_{\mathbf{t}} \sum_{m_i} \Omega' \, d\mathbf{t}} \tag{34}$$

Once again replacing Ω' by a pair of factors $F'(N,q,\mathbf{t})$ and $I'(N,q,\mathbf{t},m_i)$ respectively independent of and dependent on the loop size distribution, we now have

$$S(N,q,\mathbf{t}) = \frac{\int_{\mathbf{t}} F' \left[\ln F' \sum_{m_i} I' + \sum_{m_i} (I' \ln I') \right] d\mathbf{t}}{\int_{\mathbf{t}} F' \left[\sum_{m_i} I' \right] d\mathbf{t}} \tag{35}$$

The integration over $\mathbf{t}$ in (35) is similar to that replaced by a summation and performed in (28)–(33). For evaluation in closed form, integrable or summable functions $F'(\mathbf{t})$ and $I'(\mathbf{t})$ are needed. If the exponential approximations for $F'(\mathbf{t})$ and $I' \ln I'/I'$ of (22) are used, (35) becomes integrable. In the region of applicability of (22) and (23), however, the lowest value of $\mathbf{t}$ alone, *i.e.*, t, gives almost all the contributions to Ω. And in this case, there is no need to evaluate (35), since $\mathbf{t}$ is sensibly constant. At low values of $\mathbf{t}$, a correct function meeting the known integral values of $\mathbf{t}$ is hard to find, and has not been considered.

Applications of Theory. Consideration of the entropy of such multiply looped structures as faced here has been the major stumbling block to formulation of a complete statistical model of the superhelix molecule.

The following problems are now amenable to consideration, and in some cases comparison with experiment.

1. Distributions of **t**. Given a fixed torsional energy, resulting from uniform untwisting of the primary helix with a known torsional modulus, the twisting of the chain into a superhelix is driven with a finite available free energy. Opposed to this energy is the entropy of loop formation and twist, as here treated. Neglecting fluctuations, the equilibrium value of **t** will be reached which corresponds to zero free energy, a number not necessarily integral. And given this value of **t**, a family of different t's results, with possibility of some reverse twists which is generally small, but given by some expression like (28) or its integral equivalent, viewed as a distribution over r. Besides the distribution of apparent twists, a further distribution of **t** is possible from fluctuations in the energy, a function which must be calculated from additional parameters involving torsional degrees of freedom and rotational thermal random oscillations.

2. Superhelix loop distributions for fixed t. Given a weighting function for loop closing probabilities, such as that chosen in the present model (*i.e.*, sharp cutoff at minimum loop size = q, loop closure probability proportional to $m_i^{-3/2}$), or an improved model with a more realistic loop closure probability taking excluded volume and chain stiffness into account, loop size distributions are generable. Using the random-number approach of the programs used to calculate entropy with the present model, assemblies of loops into representative samples can be generated pictorially, or semiempirical analytical distributions can be prepared. It is only necessary to generate a large sample of groups of m_i, and to decide whether to include a given set in the representative sample by a chance decision maker whose probability of acceptance is proportional to $I(N,q,t,m_i)$ for this set.

3. Hydrodynamic variables. These depend on the expected values of t, and on the loop size distribution. As given by this or an improved theory, such variables can now be calculated more realistically, with more correct dependency on t.

Two lines of observation show the presence of actual superhelical forms in cyclic double-stranded DNA, the direct and the indirect. The direct electron microscopic observations suffer from uncertainty in artifactual effects. Yet one can hardly doubt the existence of substantial supercoiling from available electron microscope pictures,[11] even though there might be quantitative doubts about artifacts in the values of t apparently observed. The indirect observations[12–15] of changes in sedimentation velocity, titration, or binding of intercalative dyes, are subject to uncertainties of molecular shapes, and estimates of such provided by this theory may be useful in making them more quantitative. Two sizes of DNA in superhelical form have been studied, the 3×10^6 Dalton rings of SV 40 and polyoma DNA, and the 3.2×10^7 rings of λ, corresponding to the calculations performed here. If we use the Hearst-Stockmayer estimate[10] of 717 Å for the independent link length of the DNA, and a molecule length as estimated at 200 Daltons/Å for the linear density,[11] we arrive at $N = 220$ for λ-DNA, and $N = 21$ for the shorter species. Bode and MacHattie[11] observed variable **t**'s up to about 140 by electron microscope measurement, somewhat above what a sharp cutoff at $q = 2$ would allow. Obviously the link is not a stiff volumeless rod, and by dint of bending beyond that acknowledged in this theory, more than N/q loops can be formed. In the case of the smaller DNA's, the estimated **t**'s run[2] 12–15, also slightly exceeding the N/q estimate. These observations suggest that these observed superhelices have twist forces available which substantially exceed the resistance due to entropy, and which push up against the steeper forces of chain bending or sharp configurational limitation not predicted by our model. With λ-DNA prepared at medium high salt concentrations, the value of **t** is decreased to the intermediate range, where relevancy of the theory is high. The DNA produced at low salt concentrations shows substantial branching, while that from medium high concentrations has little or none. This is a point not directly treated by the present theory. However, the defining of a branch as a point where one of the segments in a given loop contains zero of the m_i links assigned to that loop, and the other segment contains all m_i of them, results in a similar prediction, for the single-loop branches allowed. The correction to the entropy rises approximately as t^2 for low t, and becomes very large at high t's. At maximum t, two out of $q + 1$ loops are branches. This type of branch will be decreased in quantity by directional restrictions which have not been considered here for high values of t requiring a predominance of small loops.

The loop size distributions are in good harmony with observed electron microscope pictures, at least qualitatively. The hydrodynamic model used by Bloomfield[6] and the "constant superhelix density" used by Vinograd, *et al.*,[8] do not agree with predictable distributions for low or moderate t's. At or near maximum t, the loop sizes approach the equality of these simple models. However, both theory and the electron microscope pictures suggest that branching should be taken into account in hydrodynamic calculation when the molecule is heavily looped.

As for the energies involved in the helix twist, estimates cited by Bauer and Vinograd[2,16] of 200- and 350-kT per molecule for ΔG have been made for superhelical polyoma DNA, with a hydrodynamic estimate of 16 loops, or $t = 15$. Using a completely twisted model, and allowing $N = 32$ (the estimate above of 21 would allow only ten loops or less), and $q = 2$, we calculate a value of $130kT$ eu for the configurational entropy loss of the totally looped molecule. Though quite unreliable for theoretical reasons, as well as due to the uncertain guess of 10^{-4} for LCF, this is of comparable magnitude to the free energy estimate. The

(11) V. C. Bode and L. MacHattie, *J. Mol. Biol.*, **32**, 673 (1968).
(12) J. Vinograd and J. Lebowitz, *J. Gen. Physiol.*, **49**, 103 (1966).
(13) J. P. Le Pecq, Thesis, Faculte des Sciences, Paris, 1965.
(14) M. J. Waring, *J. Mol. Biol.*, **13**, 269 (1965).
(15) L. V. Crawford and M. J. Waring, *ibid.*, **25**, 23 (1967).
(16) W. Bauer and J. Vinograd, Abstracts, 1969 Meeting of the Biophysical Society, Los Angeles, Calif.

presence of ΔH terms in the free energy loss, coming from solvation changes or intramolecular heat changes as the strain is relieved, cannot be picked up by this comparison. If ΔH is zero, the equality between ΔG, as stored in the superhelix, and $-T\Delta S$, as observed experimentally on such molecules, together with an accurate theory of the configurational entropy, should allow verification of dependence of superhelix free energy on chain twist.

We question whether the dependence of ΔG on t is quadratic, as suggested by Bauer and Vinograd.[2] The curve of Figure 5, which gives the free energy, if ΔH terms are small, is more nearly linear along most of its length than quadratic, and only starts to rise in curvature at high t. Dependence of r_e upon t, and consequently a nonconstant LCF, would introduce higher power dependence of S upon t. Our first intuitive scrutiny of the question leads us to believe that for a reasonably torsionally stiff chain, increase in t would simply spread the increased force driving the chains into contact over a larger number of contacts, leaving the value of r_e affected to only a small degree. At and near maximum twist, of course, we would agree that r_e should become substantially smaller. And since the resulting entropy change has a logarithmic dependence on r_e, we would expect the form of the curve of Figure 5 to reach quadraticity only by happenstance.

Acknowledgments. The author thanks Dr. K. Solc of Dartmouth and Dr. T. Ishida of Brooklyn College for programming the calculation of $\overline{I \ln I}/\bar{I}$. He also thanks the Chemistry Department of Dartmouth College for hospitality and computer facilities during the formulation of this work. He further thanks the National Institutes of Health for support under Grant AI-08038.

Appendix A

Contribution of Single-Loop Branches to Superhelix Entropy. In eq 9 of the main text, we seek the product $\prod_{i=2}^{t}(m_i + 1) = \prod_{i=2}^{t} m_i \prod_{i=2}^{t}(1 + 1/m_i)$, finding suitable approximation for the latter product term, and especially for its logarithm $\sum_{i=2}^{t} \ln(1 + 1/m_i)$, where m_i are all possible sizes of the $t + 1$ loops from q through $N - tq$, such that $\sum_{i=2}^{t} m_i = N$. The second product term is a correction term which would be expressed as $\prod_{i=2}^{t}(1 - 1/m_i)$, if we had not included loops with zero upper or lower segment links, *i.e.*, single-loop branches, in our count. We consequently refer to it as a single-loop branch contribution correction. Strictly, it is about half the single-loop branch contribution, the other half being hidden in the product $\prod_{i=2}^{t} m_i$.

An averaging approximation, in which the mean value of $\ln(1 + 1/m_i)$, represented as $\overline{\ln(1 + m_i)}$, is computed, gives the desired result as $(t - 1)\,\overline{\ln(1 + 1/m_i)}$. The random nature of the t cuts in the N-link chain made in step b of the process shown in Figure 3 determines the m_i in a loop size distribution which is approximately exponential, with minimum size q, and maximum size $N - tq$. We therefore can consider such an exponential distribution, giving rise to a population of m_i, and can calculate the desired average correction therefrom.

Consider each loop size m_i to arise from a cut made in the N-link chain, m_i units from the previous cut, with a minimum of q. The likelihood of such an eventuality is $p(1 - p)^{m_i - q}$, where p is the probability of a cut at one point, since the likelihood of no cuts at the m_i points from q through $m_i - 1$ is $1 - p$ each. The quantity p is determined by considering the cut density. There are t cuts at $N - 1 - 2(t + 1)(q - 1)$ possible points, to first approximation. This figure arises from consideration of the obliteration of $2(q - 1)$ points from cut eligibility for each of t cuts because of the q minimum restriction, and from the ineligibility of the $2q$ points at each end of the $N + 1$-point, N-link chain, for essentially the same reason. The cut density is thus approximately

$$p = t/[N - 1 - 2(t + 1)(q - 1)] \qquad \text{(A1)}$$

This value is only strictly correct for $t = 1$, *i.e.*, when no overlap of the obliterated points occurs. At larger values of t, overlap will reduce the obliterated points from $2(q - 1)$ points, for $t = 1$, to an average of $(q - 1)$ points per cut for $t = N/q - 1$. We linearize the overlap to give $2(q - 1) - (t - 1)q(q - 1)/(N - 2q)$ as the average number of obliterated points per cut. The same fractional decrease applies to the points at the chain ends. We consequently replace the $2(q - 1)$ in (A1) by the corrected approximation above for average number of obliterated points per cut, in essence considering the origin as the $(t + 1)$th cut, and replace (A1) by the improved approximation

$$p = t/\{N - 1 - (q - 1)(t + 1) \times [2 - q(t - 1)/(N - 2q)]\} \qquad \text{(A2)}$$

Equation A2 reduces to the simpler (A1) for $t = 1$, and nearly so for small t, in which overlap is infrequent. When $t = N/q - 1$, at which point all possible cuts *must* be made to divide the chain into N/q loops of q links each, $p = 1$, from (A2). Equation A1 has long since ceased to hold.

We now calculate the mean values of $\ln(1 + 1/m_i)$ in the assembly of m_i's resulting from the exponential distribution given by $p(1 - p)^{m_i - q}$. The average of any function of m_i can be computed from the quotient

$$\overline{f(m_i)} = \sum_{m_i = q}^{N - tq} f(m_i)p(1 - p)^{m_i - q} \Big/ \sum_{m_i = q}^{N - tq} p(1 - p)^{m_i - q} = \sum_{m_i} f(m_i)(1 - p)^{m_i} \Big/ \sum_{m_i} (1 - p)^{m_i} \qquad \text{(A3)}$$

The summations over m_i can be approximately replaced by integrals, using the same limits; the function of interest is $\ln(1 + 1/m_i)$. We get its approximate average from

$$\overline{\ln(1 + 1/m_i)} = \int_q^{N - 2q} \ln(1 + 1/m) \times (1 - p)^m \, dm \Big/ \int_q^{N - 2q} (1 - p)^m \, dm \qquad \text{(A4)}$$

where m_i has been replaced by m in the integrals.

The upper limit of the integral is not very critical, as $(1 - p)^{m_i}$ becomes very small for values of t or of m_i beyond small integers; moreover the contribution of the infrequent large m_i's to the value of the numerator (A3) is small. The upper limit could be taken as infinity, without great error, or as an arbitrary $N/2$, reflecting an average point of impossibility of loop continuance. A more accurate approximation would introduce a weighting factor of approximately triangular form to represent the lower likelihood of the longer loops. None of these changes would affect the entropy contribution by sensible amounts, and we proceed as indicated in (A4).

Performing the indicated integrations, we find

$$\overline{\ln (1 + 1/m_i)} = \{\ln (q + 1) - \ln q - (1 - p)^{N-(t+1)q} \ln [1 + 1/(N - tq)] - \mathrm{Ei}[(q + 1) \times \ln (1 - p)]/(1 - p)^{q+1} + \mathrm{Ei}[q \ln (1 - p)]/(1 - p)^q + \mathrm{Ei}[(N - 2q + 1) \ln (1 - p)]/(1 - p)^{q+1} - \mathrm{Ei}[(N - 2q) \ln (1 - p)]/(1 - p)^q\} / [1 - (1 - p)^{N-(t+1)q}] \quad (A5)$$

where Ei is the exponential integral $\int_x^\infty e^x \, dx/x$, and is read from tables, or computed in subroutine series, as the negative of the tabulated values of $-\mathrm{Ei}(-x)$, since all arguments thereof are negative.

We omit description of the integration, performed with standard substitutive and parts methods, but cite the results of the indefinite integration of the numerator Nu, and the denominator, De, of (A4)

$$\mathrm{Nu} = (1 - p)^m \ln (1 + 1/m) + \mathrm{Ei}[m \ln (1 - p)] - \{\mathrm{Ei}[(m + 1) \ln (1 - p)]\}/(1 - p) \Big/ \ln (1 - p) \quad (A6)$$

$$\mathrm{De} = (1 - p)^m/\ln (1 - p) \quad (A7)$$

where Ei is here the indefinite form of the exponential integral. Substitution of limits and division of (A6) by (A7) gives (A5), and the correctness of the integration can be verified by differentiating to give (A4).

The quantity $\overline{\ln (1 + 1/m_i)}$, as given by (A5) and (A2), is the correction factor given in the main text in eq 9 and Table I. Computation thereof was programmed on Wang 380 calculator, for the same values of N, q, and t as given for the other entropy parameters in Table I. The two terms in the exponential integral of the upper limit, as well as the logarithmic term in the numerator containing the upper limit, in eq A5, were negligible, except in the case of $t = 2$, for which they were included in the calculation. The denominator of (A5) differs appreciably from unity only for t from 2 to 5, but was correctly included in the program for all values of t. Thus above $t = 5$, the existence of an upper limit to the loop size (with $N = 200$ and $q = 2$) could be totally ignored. Above $t = 2$, such limit only contributes a small factor to the denominator of (A5). Since the terms which are appreciably affected by an uncertainty in the upper limit of loop size only occur at a time when the whole correction is small, we feel the computed correction is a satisfactory representation of the contribution of single-loop branches to the entropy, and certainly more accurate than the uncertainty in the Monte Carlo method used would necessitate.

Thermodynamics of Equilibrium Polymerization in Solution. Effect of Polymer Concentration on the Equilibrium Monomer Concentration

J. Leonard

Département de Chimie, Université Laval, Québec 10, Canada.
Received May 27, 1969

ABSTRACT: Recently, it has been shown that for an equilibrium polymerization in solution, the equilibrium monomer volume fraction ϕ_m varies with the polymer volume fraction ϕ_p, in accordance with the linear relation $\phi_m = A + B\phi_p$. In this paper the relation between the constants A, B, ΔG_{le}, the free-energy change upon the conversion of 1 mol of liquid monomer to 1 base-mol of long chain amorphous polymer, and the interaction parameters χ_{sp}, χ_{mp}, and χ_{ms} is emphasized, the subscripts m, p, and s referring to monomer, polymer, and solvent, respectively. Calculations show that, for a given monomer, the equilibrium position is predominantly determined by χ_{ms}, the monomer–solvent interaction parameter. Expressions for the variation of A with temperature and the effect of short chains on the value of ϕ_m are deduced. Results calculated from these expressions are in good agreement with those of the literature.

A state of equilibrium between monomer and "living" polymer may be attained in the course of an anionic or cationic polymerization. Ivin and Leonard[1] have recently observed that in the case of the equilibrium anionic polymerization of α-methylstyrene in tetrahydrofuran (THF), at certain temperatures, the equilibrium monomer concentration decreased linearly with increasing concentration of polymer.

Taking into consideration the effect of polymer concentration on the equilibrium monomer concen-

(1) K. J. Ivin and J. Leonard, *Eur. Polym. J.*, in press.