

# The equilibrium statistical mechanics of the helix-coil transition in torsionally stressed DNA

Craig J. Benham<sup>a)</sup>

The Institute for Cancer Research, The Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111  
(Received 30 August 1979; accepted 16 November 1979)

A theory of helix-coil transitions in torsionally stressed heteropolymeric DNA is developed. Specifically, the polymer is constrained to have a fixed value of total twist in all accessible states, which may differ from its unstressed total twist. This imposed deformation is partitioned between twisting of the helical portions and melting with subsequent twisting in the coil regions. Quadratic free energy densities are associated to torsion in both types of regions. In this theory two parameters influence helix stability—temperature and the magnitude of the imposed twist. Consequently the predicted transition behavior is considerably more intricate than that arising from thermal transitions in unconstrained molecules. In particular, sufficient untwisting can significantly destabilize the helix, even at low temperatures. In this case melting occurs predominantly in AT-rich regions, although the transition behavior of such regions is complicated. For example, if the imposed twist is altered to decrease helix stability at fixed temperature, the probability of melting of a given monomer unit need not increase monotonically. The predictions of this theory compare favorably with a previous thermodynamic analysis of the same problem.

## I. INTRODUCTION

Recent developments have brought a recognition of the importance of the mechanical properties of DNA to questions of molecular structure and function.<sup>1-5</sup> In this view elastic properties are ascribed to the duplex. That is, in its unstressed, elastic equilibrium state the helix is straight with twist rate of  $A = 10.4$  base pairs per turn.<sup>3,6,7</sup> Deviations from this conformation are opposed by elastic restoring stresses, hence require energy. A linear segment of DNA with free ends can deform so as to relieve these stresses. However, DNA can be constrained so as to preclude unrestricted deformation. One such constraint which occurs commonly in natural DNA is covalent closure of the molecule to form a ring. Alternatively, a molecule of DNA which is held at points along its length can be stressed (or relaxed) by molecular activity in the intervening regions. Commonly these imposed stresses involve torsion. Examples include the unwinding of the *lac* operator by binding of *lac* repressor,<sup>8</sup> and the activities of DNA gyrases<sup>9,10</sup> and topoisomerases.<sup>11,12</sup>

It is important to consider possible effects of imposed stresses on the structure and function of DNA. To this end, the present contribution extends the theory of helix-coil transitions in heteropolymers to treat torsionally stressed molecules. Specifically, the duplex is regarded as being constrained to be either underwound or overwound relative to the elastic equilibrium state. This deformation is partitioned between helical and coil regions. Changes  $\tau_h$  in the twist rate of the helical regions from its unstressed value require quadratic energy density  $C_h \tau_h^2/2$ , where  $C_h$  is the torsional stiffness of the duplex.<sup>4</sup> Similarly, the two denatured strands in a coil region may twist around each other with rate  $\tau_c$ . This deformation also requires a free energy density  $C_c \tau_c^2/2$ , where  $C_c$  is the effective torsional stiffness of the coil. Because twisting occurs in the coil regions of a stressed molecule, further considerations of loop entropy do not

arise. Indeed, the free energy associated with this twisting is itself predominantly entropic in nature.<sup>13</sup>

A thermodynamic treatment of denaturation in torsionally stressed DNA has been developed previously.<sup>14</sup> The present statistical mechanical reformulation of this problem extends the thermodynamic results by rendering computable (at least in principle) both the melting profiles of specific base sequences and the magnitudes of fluctuation of important parameters.

An alternative statistical mechanical approach to the melting of DNA subject to the single specific constraint of ring closure has been presented.<sup>15</sup> However, in that work it was assumed that coil regions always have zero twist while helical regions always remain at their unstressed twist rate. This presumes that duplex DNA is torsionally undeformable, which is clearly unrealistic. In that treatment energy is associated to the states of the system in a completely nonspecific way. Indeed, the possible accessible states (which would involve twisting, bending, and melting<sup>3</sup>) are never explicitly described at all.

In the following sections the helix-coil transition theory for torsionally stressed heteropolymers is developed. Theoretical predictions are made regarding the melting profiles of specific DNA sequences. It is shown that stress-induced melting has properties that are somewhat different from those of the strictly thermal transition in unstressed molecules. Present results agree well with the predictions of the thermodynamic theory. A detailed treatment of possible biological roles for stress-induced strand separation will be presented elsewhere. Applications of this approach to problems involving structure and function of DNA are briefly described in Ref. 14.

## II. HELIX-COIL TRANSITION THEORY FOR TORSIONALLY STRESSED DNA

Consider a heteropolymer sequence of  $N$  base pairs whose total twist  $Tw$  deviates from its unstressed value  $Tw_0$  by  $q = Tw - Tw_0$ . (Here  $Tw_0 = N/A$ , where  $A$  is the

<sup>a)</sup>Present address: Mathematics Department, University of Kentucky, Lexington, Kentucky 40506.

number of base pairs per turn of the unstressed helix,  $A=10.4$ .) In consequence the twist rates of helical and coil regions may differ by amounts  $\tau_h$  and  $\tau_c$  from their respective unstressed values. (Note that the unstressed twist rate of a coil region is zero, while that of a helix is  $2\pi/A$  when measured in radians per base pair.) In the present, initial formulation we assume that  $\tau_h$  (respectively  $\tau_c$ ) has the same value for all bases in all helical (respectively coil) regions, as is true at (stressed or unstressed) elastic equilibrium.<sup>3</sup> If a total of  $n$  base pairs are in the coil state, then

$$\tau_h = \begin{cases} 2\pi q/N & n=0 \\ \left[2\pi\left(q + \frac{n}{A}\right) - n\tau_c\right] / (N-n) & 0 < n < N. \end{cases} \quad (1)$$

It must be emphasized that the present approach differs from the standard theory of helix-coil transitions in heteropolymers by imposing the constraint of constancy of twist in all accessible states. This deformation is partitioned between the twisting of helical regions and the melting and subsequent twisting of the coil regions. In the standard theory the molecule is unconstrained, permitting it to untwist as it melts so that helical and coil regions remain at their unstressed twist rates regardless of how much denaturation occurs. For this reason the present theory does not reduce to the traditional one, even when  $q=0$ . [To see this, note from Eq. (1) that when  $q=0$  and  $n>0$ , either  $\tau_c$ , or  $\tau_h$  must be nonzero.]

A system whose states are discrete (such as is found in the traditional helix-coil theory) has associated partition function

$$Z = \sum_r \exp(-\beta E_r),$$

where the index  $r$  enumerates all accessible states. If a continuum of accessible states exists, the above summation is replaced by an integral. The present problem amalgamates both of these cases. For the free energy of a given state of a stressed heteropolymer is the sum of a discrete term  $E_{bp}$  arising from the state of base pairing along the molecule, and a continuous term  $E_n$  contributed by the twist distribution:

$$E = E_{bp} + E_n. \quad (2)$$

The term  $E_{bp}$  is determined by the sequence of helix and coil bases along the molecule. It is identical to the free energy involved in the standard helix-coil transition theory for unconstrained DNA. Once the sequence of helix and coil bases is specified, the number  $n$  of denatured bases is determined. If the molecule is either totally coil ( $n=N$ ) or totally helix ( $n=0$ ) the energy of twist is determined by  $q$  to be

$$E_0 = 2\pi^2 q^2 C_h/N, \\ E_N = 2\pi^2 \left(q + \frac{N}{A}\right)^2 C_c/N.$$

For all intermediate states,  $0 < n < N$ , the free energy associated to twist is  $E_n = [(N-n)C_h\tau_h^2 + nC_c\tau_c^2]/2$ . Then substitution from Eq. (1) expresses  $E_n$  as the following function of  $\tau_c$  alone,

$$E_n(\tau_c) = \frac{nC_c\tau_c^2}{2} + \frac{C_h}{2(N-n)} \left[2\pi\left(q + \frac{n}{A}\right) - n\tau_c\right]^2.$$

Therefore substitution of the sum in Eq. (2) into the partition function for a torsionally stressed heteropolymer permits factorization into continuous and discrete terms

$$Z = \sum_{n=0}^N \left[ \left( \sum_{\tau_n} \exp(-\beta E_{bp}) \right) \int_{-\infty}^{\infty} \exp[-\beta E_n(\tau_c)] d\tau_c \right] \\ + \exp[-\beta(E_{bp(N)} + E_N)] + \exp[-\beta(E_{bp(0)} + E_0)]. \quad (3)$$

The index  $\tau_n$  enumerates all states of base pairing with  $n$  coil bases. The last two terms arise from the totally coil and the totally helical cases.

The integral in Eq. (3) may be evaluated to be

$$I_n = \int_{-\infty}^{\infty} \exp[-\beta E_n(\tau_c)] d\tau_c \\ = \left( \frac{2\pi(N-n)}{n\beta[C_h n + C_c(N-n)]} \right)^{1/2} \\ \times \exp\left( \frac{-2\pi^2 C_c C_h \beta (q + n/A)^2}{C_h n + C_c(N-n)} \right) \quad 0 < n < N. \quad (4a)$$

Letting

$$I_0 = \exp(-2\pi^2 q^2 C_h \beta/N), \quad (4b)$$

$$I_N = \exp\left[-2\pi^2 \left(q + \frac{N}{A}\right)^2 C_c \beta/N\right], \quad (4c)$$

the partition function for a torsionally stressed heteropolymer becomes

$$Z = \sum_{n=0}^N \left( I_n \sum_{\tau_n} \exp(-\beta E_{bp}) \right). \quad (5)$$

A parameter  $f(\tau_c)$  dependent on the twist rate of the molecule will have mean value given by

$$\bar{f} = \left[ \sum_{n=0}^N J_n \left( \sum_{\tau_n} \exp(-\beta E_{bp}) \right) \right] (Z)^{-1}, \quad (6)$$

where

$$J_n = \int_{-\infty}^{\infty} f(\tau_c) \exp[-\beta E_n(\tau_c)] d\tau_c.$$

In particular, the mean value of  $\tau_c$  is given by Eq. (6) with  $f(\tau_c) = \tau_c$  so that

$$J_n = \int_{-\infty}^{\infty} \tau_c \exp[-\beta E_n(\tau_c)] d\tau_c \\ = 0 \quad n=0 \\ = 2\pi I_n C_h (q + n/A) / [C_h n + C_c(N-n)] \quad 0 < n < N. \quad (7)$$

The variance of  $\tau_c$ , a measure of the magnitude of fluctuations, is given by

$$\text{Var}(\tau_c) = (\overline{\tau_c^2}) - (\overline{\tau_c})^2.$$

The value of  $(\overline{\tau_c^2})$  is determined from Eq. (6) with  $f(\tau_c) = \tau_c^2$ , so that

$$\int_{-\infty}^{\infty} \tau_c^2 \exp[-\beta E_n(\tau_c)] d\tau_c = 0 \quad n=0 \\ = \left( \frac{4\pi^2 C_h^2 (q + n/A)^2}{[C_h n + C_c(N-n)]^2} + \frac{N-n}{n\beta[C_h n + C_c(N-n)]} \right) I_n \quad 0 < n \leq N.$$

The mean value of a quantity  $\mu$  which is not an explicit function of twist is given by

$$\bar{\mu} = \sum_{n=0}^N \left( I_n \sum_{r_n} \mu \exp(-\beta E_{bp}) \right) / Z .$$

In this way expressions for the average number of helical bases, the average number of helical regions, and measures of the magnitude of their fluctuations can be found.

The Boltzmann factors  $\exp(-\beta E_{bp})$  appearing in the partition function of Eq. (5) are identical to those found in the standard theory of helix-coil transitions in unstressed heteropolymers.<sup>16</sup> Briefly, let the zero point of free energy for each base pair occur in the melted state. Then the free energy associated to the  $i$ th pair when native is

$$F^i = [F(\text{nat}) - F(\text{den})]^i .$$

In a heteropolymer,  $F^i$  depends upon the type of the base pair involved, AT or GC. A cooperativity free energy  $F_s$  is associated to each pair which is helical, but whose next neighbor is in the coil state. This corresponds to the fact that an extra amount of free energy is needed to denature a first base pair in a region.  $F_s$  is assumed to have the same value for all base pairs. The equilibrium constant for the  $i$ th pair is  $s_i = \exp(-\beta F^i)$  and the cooperativity factor is  $\sigma = \exp(-\beta F_s)$ . The Boltzmann factor for each specific sequence of helical and coil bases is of the form

$$\exp(-\beta E_{bp}) = \left( \prod_{\alpha} s_i \right) \sigma^{\gamma} .$$

Here  $\alpha$  indexes the helical base pairs and  $\gamma$  is half the number of pairs in the helix state which have a coil neighbor. (That is,  $\gamma$  is the number of beginnings of sequences of coil bases.)

With these variables the average values of the most important parameters can often be expressed in an easily calculable form. For example, the probability  $p_i$  that the  $i$ th pair is helical is

$$p_i = \partial \ln Z / \partial \ln s_i . \tag{8a}$$

The expected number of coil bases is then

$$\bar{\theta} = \sum_{i=1}^N (1 - p_i) . \tag{8b}$$

The mean value of  $\nu$ , the number of helical regions, is given by<sup>16</sup>

$$\bar{\nu} = \partial \ln Z / \partial \ln \sigma . \tag{9}$$

Similarly, the magnitude of fluctuations can be found from appropriate second derivatives.

### III. CALCULATIONS

In this section methods are developed for calculating several equilibrium properties of the helix-coil transition in torsionally stressed heteropolymers. Specifically, algorithms are given for evaluating the expected number  $\bar{\nu}$  of helical regions, the probability  $p_i$  of the  $i$ th pair being native, and the expected value  $\bar{\tau}_c$  of the twist rate in the coil regions. The technique used may be easily

modified to calculate other quantities of interest. The numerical method used here is an adaptation of a recursion relation technique developed by Anshelevich *et al.*<sup>17</sup>

Define  $g_n(m)$  to be the conditional partition function of the first  $n$  base pairs when the  $n+1$ st pair is coil and beyond the  $n+1$ st there are exactly  $m$  coil pairs. Let  $g_0(m) = I_m$ . One can show that the following recursion relationship holds<sup>17</sup>:

$$g_1(m) = I_{m+1} + \sigma s_1 I_m , \tag{10a}$$

$$g_n(m) = g_{n-1}(m+1) + s_n g_{n-1}(m) + (\sigma - 1) s_n g_{n-2}(m+1) \quad n \geq 2 . \tag{10b}$$

Finally, the partition function for the sequence of  $N$  base pairs is  $Z = g_N(0)$ .<sup>17</sup>

Let  $G_0(m) = F_0(m) = I_m$ ,  $G_i(m) = g_i(m)/g_{i-1}(m)$ , and  $F_i(m) = g_i(m)/g_{i-1}(m+1)$ , when  $i \geq 1$ . Then, from Eqs. (10a) and (10b),

$$G_1(m) = \sigma s_1 + (I_{m+1}/I_m) , \tag{11a}$$

$$F_1(m) = 1 + (\sigma s_1 I_m / I_{m+1}) , \tag{11b}$$

$$G_n(m) = s_n + [G_{n-1}(m+1) + s_n(\sigma - 1)] / F_{n-1}(m) \quad n \geq 2 , \tag{11c}$$

$$F_n(m) = 1 + [s_n F_{n-1}(m) + s_n(\sigma - 1)] / G_{n-1}(m+1) \quad n \geq 2 . \tag{11d}$$

Now the partition function  $Z$  is given by  $Z = \prod_{i=0}^N G_i(0)$ .<sup>17</sup>

This procedure may be repeated with  $J_n$  replacing  $I_n$  throughout to find analogous recursion relations expressing any sum of the form

$$Z_J = \sum_{n=0}^N \left( J_n \sum_{r_n} \exp(-\beta E_{bp}) \right) .$$

If the recursion functions in this case are denoted  $\bar{G}_n(m)$ ,  $\bar{F}_n(m)$ , then the recursion relations found are those of Eqs. (11a)–(11d) with  $\bar{G}$ ,  $\bar{F}$ ,  $J$  in place of  $G$ ,  $F$ ,  $I$ . Therefore the average value of any function  $f(\tau_c)$  may be evaluated from Eq. (6) to be

$$\bar{f} = \prod_{i=0}^N [\bar{G}_i(0)/G_i(0)] .$$

In particular the average twist rate  $\bar{\tau}_c$  of the coil region is computed this way using the values of  $J_n$  given by Eq. (7) and defining  $\bar{G}_0(0) = 1$ .

Formulae for calculating other equilibrium properties of this system may be derived from Eqs. (11a)–(11d). For example, the probability  $p_j$  of the  $j$ th pair being helical is found [using Eq. (8a)] to be

$$p_j = \sum_{i=0}^N \frac{\partial \ln G_i(0)}{\partial \ln s_j} .$$

Letting  $W_n(m) = \partial \ln G_n(m) / \partial \ln s_j$  and  $V_n(m) = \partial \ln F_n(m) / \partial \ln s_j$ , then

$$p_j = \sum_{i=0}^N W_i(0) .$$

Recursion relations for  $W_n(m)$ ,  $V_n(m)$  are found by direct differentiation of Eqs. (11a)–(11d). These are  $W_0(m) = V_0(m) = 0$ ,

$$\begin{aligned}
 V_1(m) &= W_1(m) = s_1 \sigma I_m / (I_{m+1} + \sigma s_1 I_m) \quad j=1 \\
 &= 0 \quad j \neq 1, \\
 W_n(m) &= \frac{S_j}{G_n(m)} \left( 1 + \frac{\sigma - 1}{F_{n-1}(m)} [1 - V_{n-1}(m)] + \frac{G_{n-1}(m+1)}{S_n F_{n-1}(m)} [W_{n-1}(m+1) - V_{n-1}(m)] \right) \quad j=n \geq 2 \\
 &= \frac{1}{G_n(m)} \left( \frac{s_n(1-\sigma)V_{n-1}(m)}{F_{n-1}(m)} + \frac{G_{n-1}(m+1)}{F_{n-1}(m)} [W_{n-1}(m+1) - V_{n-1}(m)] \right) \quad j \neq n \geq 2, \\
 V_n(m) &= \frac{S_j}{F_n(m) G_{n-1}(m+1)} ([F_{n-1}(m) + \sigma - 1][1 - W_{n-1}(m+1)] + F_{n-1}(m) V_{n-1}(m)) \quad j=n \geq 2 \\
 &= \frac{S_n}{F_n(m) G_{n-1}(m+1)} \{ [V_{n-1}(m) F_{n-1}(m)] - [F_{n-1}(m) + \sigma - 1] W_{n-1}(m+1) \} \quad j \neq n \geq 2.
 \end{aligned}$$

To calculate  $\bar{v}$  from Eq. (9) one proceeds similarly. That is,

$$\bar{v} = \sum_{i=0}^N R_i(0),$$

where  $R_n(m) = \partial \ln G_n(m) / \partial \ln \sigma$ ,  $T_n(m) = \partial \ln F_n(m) / \partial \ln \sigma$ .

The recursion relations in this case are

$$\begin{aligned}
 T_0(m) &= R_0(m) = 0, \\
 R_1(m) &= s_1 \sigma / G_1(m), \\
 T_1(m) &= \sigma s_1 I_m / (I_{m+1} F_1(m)), \\
 R_n(m) &= \frac{1}{F_{n-1}(m) G_n(m)} \{ s_n \sigma + G_{n-1}(m+1) [R_{n-1}(m+1) - T_{n-1}(m)] - s_n(\sigma - 1) T_{n-1}(m) \} \quad n \geq 2, \\
 T_n(m) &= \frac{S_n}{F_n(m) G_{n-1}(m+1)} \{ T_{n-1}(m) F_{n-1}(m) + \sigma - R_{n-1}(m+1) [F_{n-1}(m) + \sigma - 1] \} \quad n \geq 2.
 \end{aligned}$$

This procedure may be adapted to produce algorithms for calculating other equilibrium properties.

#### IV. RESULTS

In this section are shown the results of sample calculations illustrating several important properties of the helix-coil transition in torsionally stressed heteropolymers. Because the computer time necessary to perform the above algorithms increases quadratically with the number of base pairs, this method of evaluation of  $p_i$ ,  $\bar{v}$ , and  $\bar{v}_c$  can only be applied to relatively short sequences. On the small computer available to the author (PDP 11/70) the longest sequence it was possible to treat contained 45 base pairs. This was found to be sufficient to investigate the important properties of the transition.

Approximation techniques must be developed for handling calculations involving thousands of base pairs. This problem is more difficult for torsionally stressed molecules than it is for unconstrained heteropolymers. No coarse-graining approach seems feasible for estimating melting probabilities because the twist in a re-

gion depends upon the number of coil bases found there. This important question will be treated carefully elsewhere.

Because only short sequences are considered here, end effects are quite important. In the present calculation two types of boundary conditions are used, both of which involve adding an extra base pair to each end. In the first case these extra pairs are regarded as free to denature with equilibrium constant equal to that of a GC pair. This assumption of meltable ends is not physically plausible for stressed DNA because the imposition of our torsional constraint ( $q = \text{const}$ ) requires that the ends of the segment be held. It is more reasonable to assume that the added base pairs at the ends are held in the helical state. This is accomplished by endowing them with a high equilibrium constant (here chosen as  $s = 10^7$ ).

The base sequence used in most of the following calculations (excluding the extra pairs added to the ends) is

TTCGT/ACAGC/ATTAT/TCGGA/GTGAC/GCTAA/GGATC/GTAAA/TATCA,

where the sequence of only one strand is given. There are 27 AT and 18GC pairs in all. Two AT-rich regions are found—one occupying positions 11–16 and the other at 37–43. The rest of the sequence has 56% GC content. The first AT-rich region is bracketed by late melting GC-rich sections (three bases out of the nearest four

on each side are GC). The sequences bounding the second AT-rich segment are evenly divided between AT and GC. In these calculations the melting temperature  $T_m$  for an AT and a GC pair are taken to be 340 and 380 °K, respectively. The torsional constraint is expressed in terms of the imposed twist density  $\delta$ , defined as

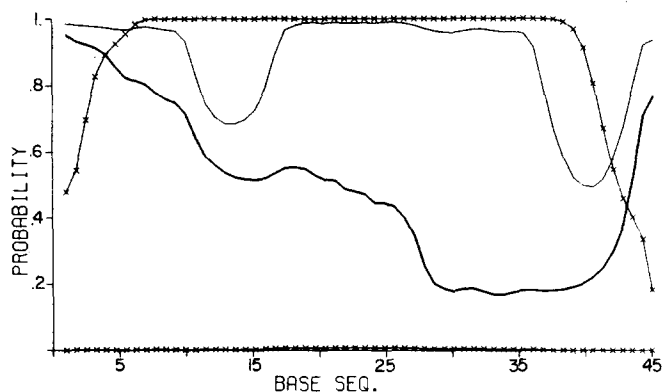


FIG. 1. The probability of being in the helix state is plotted for the sequence of 45 bases described in the text. Here  $\delta = -0.2$  in all cases, while temperature is either 300 °K (favoring helix) or 390 °K (corresponding to denaturing conditions). The cross-hatched lines show the transition profiles for this sequence with meltable ends, while the continuous lines give the corresponding profiles under identical conditions with fixed ends.

$$\delta = (Tw - Tw_0)/Tw_0 = qA/N.$$

Figure 1 displays transition profiles for the above sequence when endowed with meltable and with fixed ends. Profiles are plotted in each case at a high temperature ( $T = 390$  °K) and a low temperature ( $T = 300$  °K). In all cases a twist density  $\delta = -0.2$  was used. It is seen that the negative imposed twist markedly destabilizes the double helix, producing significant melting probabilities even at a temperature substantially below the sequence average  $T_m$ . For fixed ends at 300 °K this melting is largely confined to the AT-rich regions, whereas when the ends are meltable denaturation occurs from the ends. Above  $T_m$  the meltable sequence is essentially completely coil. When the ends are fixed, however, their influence propagates along the sequence, holding it in a partially helical conformation even at high temperature.

The torsional constraint  $q$  (or  $\delta$ ) influences the predicted helix-coil transition in several ways. Increasing  $\delta$  at constant temperature tends to stabilize the helix.

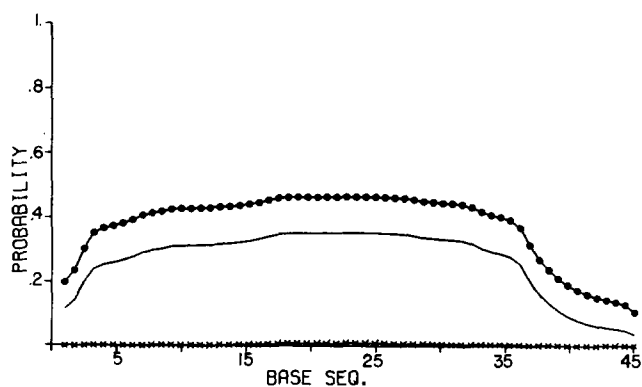


FIG. 2. The transition profile of the sample sequence with free ends is shown at temperature of 390 °K and three different densities. The cross-hatched profile corresponds to  $\delta = -0.2$ ; the continuous line has  $\delta = 0$  and the line with circles has  $\delta = +0.05$ . An increase in  $\delta$  acts to stabilize the helix.

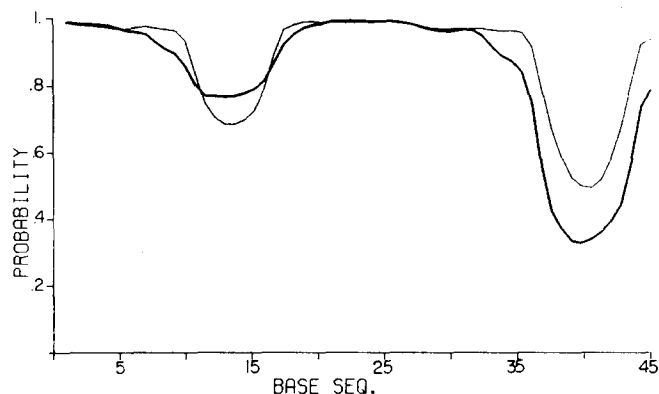


FIG. 3. The transition profiles are shown for the sample sequence with fixed ends at  $T = 300$  °K and two different twist densities. The light line corresponds to  $\delta = -0.2$ , while the heavier line is for  $\delta = -0.28$ . (When  $\delta = 0$  no denaturation occurs under these conditions.) In both cases melting is substantially confined to the two AT-rich portions of the sequence. Although higher (negative) torsional stress decreases overall helix stability, the AT-rich region about base 13 is actually *less* likely to melt at  $\delta = -0.28$  than at  $\delta = -0.20$ ! The opposite conclusion holds for the AT-rich region near base 40.

This is illustrated in Fig. 2, where transition profiles are plotted for our sequence with free ends at  $T = 390$  °K and three values of twist density,  $\delta = -0.2, 0, +0.05$ . (Thermodynamic analysis suggests that extremely high values of  $\delta$  will destabilize the helix.<sup>14</sup> For technical reasons this prediction could not be tested here as problems were encountered in the algorithms with underflow and thereafter with attempts to divide by a spurious zero.)

In the standard theory of helix-coil transitions for unconstrained DNA the probability of melting of an individual base increases monotonically with temperature. When the factor affecting helix stability is the imposition of a torsional deformation this is not always true. Figure 3 displays the melting profiles for our sequence at

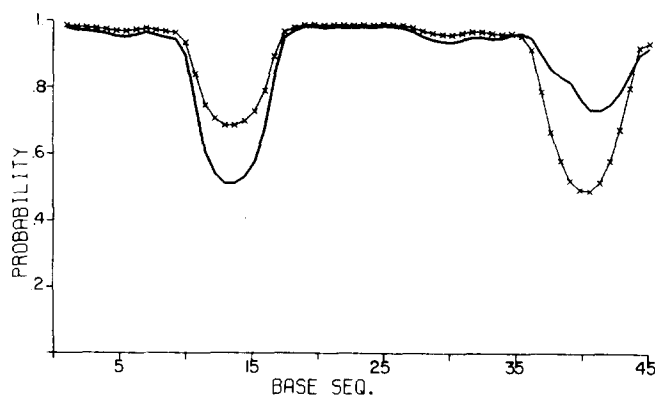


FIG. 4. The transition profiles are shown for two sequences, both with fixed ends at  $T = 300$  °K and  $\delta = -0.2$ . The lighter line shows the helix probabilities for the standard sample sequence described in the text. The heavy line gives the transition profile for a sequence which varies from the standard one by the single replacement of an AT pair with a GC one at position 39. It is seen that when temperature favors base pairing, the transition profile for stressed DNA is extremely sensitive to base sequence in the AT-rich regions.

TABLE I. The stress-induced melting of DNA may be analyzed using either the statistical mechanical approach described here or the thermodynamic theory developed in Ref. 14. This table compares the predictions of these two approaches when applied to the standard sequence of the text under various conditions of temperature and torsional deformation. Here  $\bar{\theta}$  is the expected number of melted bases and  $\bar{\tau}_c$  is the expected twist rate of the coil regions, both calculated from the statistical mechanical theory. Thermodynamic analysis yields the number of denatured bases  $n$  and the coil twist rate  $\tau_c$  which locally minimize the free energy under the given conditions. (In this approach the number of bases which melt must be integral, so  $n$  has been rounded off to the nearest integer.) The predictions of these two approaches to melting of torsionally stressed DNA are seen to agree quite well.

$\delta$	$T(^{\circ}\text{K})$	$\bar{\theta}$	$n$	$\bar{\tau}_c$	$\tau_c$
-0.12	300	3.03	3	-0.35	-0.28
-0.16	300	4.30	4	-0.33	-0.28
-0.2	300	5.49	6	-0.34	-0.28
-0.24	300	6.55	7	-0.32	-0.28
-0.28	300	7.64	8	-0.31	-0.28
-0.2	315	6.05	6	-0.27	-0.25
-0.2	330	6.85	7	-0.19	-0.11
-0.2	345	7.96	8	-0.11	-0.08
-0.2	360	9.81	9	0.00	0.00
-0.2	375	13.18	12	0.14	0.13
-0.2	390	23.25	21	0.32	0.32

two twist densities,  $\delta = -0.2$  and  $\delta = -0.28$ . Here the temperature chosen is 300 °K and fixed ends are assumed. In both cases melting is largely confined to the neighborhoods of the two AT-rich regions. However, as the imposed twist becomes more extreme the melting becomes increasingly concentrated about the most AT-rich region while the probability of melting at the shorter AT-rich segment actually decreases!

The melting profile of a stressed heteropolymer below its transition temperature is extremely sensitive to base sequence in the AT-rich regions. Figure 4 shows the transition profiles for two sequences which differ by a single substitution. In both cases  $T = 300$  °K,  $\delta = -0.2$  and fixed ends are assumed. The base at position 39, an A in our original sequence (light line), is changed to a G in the new sequence. Although the average number of coil bases remains essentially unchanged, this single substitution radically alters the distribution of melting probabilities between the two AT-rich regions.

These results compare favorably with the predictions of the thermodynamic theory. Table I compares the expected number  $\bar{\theta}$  of coil bases calculated from Eq. (8b) (assuming fixed ends) with the number  $n$  of denatured bases predicted by the thermodynamic theory.<sup>14</sup> In addition the expected value of the coil twist rate  $\tau_c$  is shown as computed using both theories. These calculations are performed at several different twist densities and temperatures. It is seen that the predictions of these two theories are in reasonably close agreement.

## V. DISCUSSION

The theoretical helix-coil transition in torsionally stressed DNA is seen to be more complicated than the corresponding transition in unconstrained molecules.

Even substantially below the transition temperature  $T_m$  significant melting can result from the imposition of a negative torsional deformation (i. e.,  $q = Tw - Tw_0 < 0$ ). Such melting is partitioned among the most AT-rich regions in a manner highly dependent upon details of local sequence and the magnitude of the torsional constraint.

Stress-controlled local strand separation could be involved in several important physiological regulatory or structural functions. Certainly it is highly suggestive that many DNAs occur naturally in a (negatively) torsionally stressed state. Further, enzymes have been found which alter or impose torsional stresses in DNA. Also, the interaction of DNA with histones could produce torsional stresses. A detailed treatment of possible biological implications for stress-induced strand separation will be presented elsewhere. A brief outline of evidence suggesting possible roles in replication, recombination, transcription, and repair is given in Ref. 14.

The present treatment of the helix-coil transition in torsionally stressed heteropolymers contains two important simplifications. First, it is assumed that the twist rates  $\tau_h$  (respectively  $\tau_c$ ) are constant over all base pairs in the helix (respectively coil) state. In reality, these rates may fluctuate about some mean values. However, these mean values will be the same for all pairs in a given state.<sup>3</sup> For this reason our approximation is physically reasonable. It is conceptually possible to extend the present theory to include variations in the twist rates of individual bases. However, this would vastly complicate the theory, certainly beyond what seems reasonable for an initial formulation and possibly beyond the performance of important calculations. The inclusion of these variations would not be expected to alter the predictions of the theory significantly.

The second simplification implicit in the work presented here is that hairpin structures at inverted repeat sequences are not considered. A complete theory of torsionally stressed DNA must include the possibility of formation of these structures. Below the transition temperature base pairing is energetically favored, suggesting that hairpin formation might be expected. However, such structures have zero twist about the central axis, whereas coil regions may twist in the opposite direction to the double helix. Therefore untwisting at coil regions in response to negative twist density  $\delta$  at physiological temperatures would relieve more of the torsional stress on the remaining duplex than would hairpins. For this reason it is not clear whether hairpin formation will be important. Certainly this possibility need only be considered when treating relatively long sequences, as an absolute prerequisite is the existence of an inverted repeat sequence. This important question requires further study.

## ACKNOWLEDGMENTS

We acknowledge the support of grants from the National Institutes of Health; CA-22780, CA-06927, RR-05539, and an appropriation from the Commonwealth of Pennsylvania.

- <sup>1</sup>F. B. Fuller, Proc. Natl. Acad. Sci. US 68, 815 (1971).  
<sup>2</sup>C. J. Benham, Proc. Natl. Acad. Sci. US 74, 2397 (1977).  
<sup>3</sup>C. J. Benham, Biopolymers 18, 609 (1979).  
<sup>4</sup>M. Barkley and B. Zimm, J. Chem. Phys. 70, 2991 (1979).  
<sup>5</sup>M. LeBret, Biopolymers 18, 1709 (1979).  
<sup>6</sup>M. Levitt, Proc. Natl. Acad. Sci. US 75, 640 (1978).  
<sup>7</sup>J. Wang, Proc. Natl. Acad. Sci. US 76, 200 (1979).  
<sup>8</sup>J. Wang, M. Barkley, and S. Bourgeois, Nature (London) 251, 247 (1974).  
<sup>9</sup>M. Gellert, K. Mizuuchi, M. O'Dea, and H. Nash, Proc. Natl. Acad. Sci. US 73, 3872 (1976).  
<sup>10</sup>K. Mariani, J. Ikeda, S. Schlagman, and J. Hurwitz, Proc. Natl. Acad. Sci. US 74, 1965 (1977).  
<sup>11</sup>T. Jovin, Ann. Rev. Biochem. 45, 889 (1977).  
<sup>12</sup>J. Champoux, Proc. Natl. Acad. Sci. US 74, 3800 (1977).  
<sup>13</sup>D. Crothers and H. Spatz, Biopolymers 10, 1949 (1971).  
<sup>14</sup>C. J. Benham, Proc. Natl. Acad. Sci. US 76, 3870 (1979).  
<sup>15</sup>A. Vologodskii, A. Lukashin, V. Anshevich, and M. Frank-Kamenetskii, Nuc. Acids Res. 6, 967 (1979).  
<sup>16</sup>A. Vedenov, A. Dykhe, A. Frank-Kamenetskii, and M. Frank-Kamenetskii, Mol. Biol. 1, 272 (1967).  
<sup>17</sup>V. Anshevich, A. Vologodskii, A. Lukashin, and M. Frank-Kamenetskii, Biopolymers 18, 2733 (1979).