Stable Cruciform Formation at Inverted Repeat Sequences in Supercoiled DNA

CRAIG J. BENHAM, *Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506*

Synopsis

This paper analyzes equilibrium superhelical cruciform formation in a topologial domain of DNA containing inverted repeat sequences. The cruciform conformation is shown to be stable when the molecule is sufficiently negatively supercoiled but not when it is positively supercoiled. For a particular sequence containing a single inverted repeat, onset of stability occurs at a degree of negative superhelicity that depends critically on the number of base pairs separating the repeat copies. The free energy associated with the stable cruciform state is constant, independent of the degree of superhelicity, up to the point where the complete inverted repeat participates in the cruciform. In contrast, the free energy of the alternative, unextruded state grows approximately quadratically with the superhelical deformation. Therefore, a degree of negative supercoiling occurs at which the cruciform state becomes energetically favored. The equilibrium properties of this cruciform extrusion transition vary with segment length, the positions and sizes of the inverted repeats involved, environmental conditions, and the degree of superhelicity imposed. In a segment containing multiple repeats, both tandem and inverted, of one or more sequences, the pattern of cruciforms that form is dependent on superhelicity in a complex way. Specific cruciforms may occur at equilibrium only in narrow ranges of superhelicity, their reabsorption being coupled to the extrusion of others.

INTRODUCTION

Two copies of a particular sequence of base pairs can occur in a segment of DNA either with identical orientations (tandem repeat) or with opposite orientations (inverted repeat). When inverted, Watson-Crick complementarity exists between the repeat bases found on the same strand. Therefore, partial *intra*strand base pairing can occur at an inverted repeat sequence, as shown in Fig. 1. The resulting structure, called a cruciform or hairpin, consists of two arms. Each arm contains a deformed region joining it to the rest of the segment, a stem of intrastrand base pairs, and a single-stranded loop at its end that must include all bases intervening between the repeat copies.

Kinetic and thermodynamic arguments suggest that cruciforms are unlikely to occur in linear molecules in solution.[1] However, experiments show that cruciforms can form at susceptible sites when the DNA involved is negatively supercoiled. This conclusion is suggested by the action of chemical probes for single-stranded regions. Probes used include small molecules that bind to unpaired bases[2] and single-strand-specific endonucleases whose nicking sites are revealed by subsequent sequence

Biopolymers, Vol. 21, 679–696 (1982)

CCC 0006-3525/82/030679-18$01.80

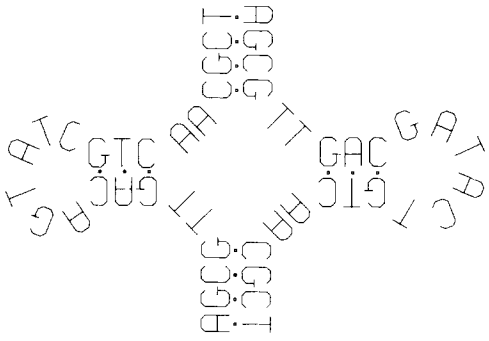

Fig. 1. Diagram of the structure of a cruciform at an inverted repeat sequence. Note that the stem region contains only part of the inverted repeat.

analysis.[3–5] When closed circular DNA containing an inverted repeat was sufficiently negatively supercoiled, the first nick was observed to occur selectively within the putative loop region separating the repeat copies. Superhelical cruciform extrusion has also been observed in artificially synthesized, closed circular, inverted dimer molecules.[6]

A negatively supercoiled molecule is topologically constrained by the deficiency in its linking number Lk.[7–9] If the molecule contains N base pairs and its strands twist at the rate of A base pairs per duplex turn when unstressed, then its superhelicity is $\lambda = Lk - (N/A)$. Superhelicity forces the molecule to twist and/or bend away from its unstressed shape, either smoothly throughout or in some localized fashion. The mechanical properties of DNA dictate that such deformations require energy, so the stable structures of the supercoiled molecule are those that minimize the conformational free energy consistent with constraints.[10,11] In practice, DNA superhelicity may be partially stabilized through interactions with other molecules (nucleosomal winding, intercalative binding, etc.). The considerations developed in this paper pertain to that portion of the total superhelicity that is unstabilized.

Because the cruciform structure has separated strands, it makes no contribution to either the total twist or the linking number of the molecule in which it occurs. Therefore, cruciform extrusion localizes linking deficiency, permitting the rest of the (negatively) supercoiled molecule to approximate its unstressed conformation more closely. When sufficiently negatively supercoiled, the deformation strain energy that is relieved elsewhere in the molecule by extruding a cruciform can exceed the free-energy costs of performing the extrusion.

This paper presents two equilibrium analyses of superhelical cruciform extrusion, one mechanical and the other statistical-mechanical. Because certain parameters (torsional stiffness, free energies) are not presently known to good accuracy, the results of computations based on our analyses should be understood to illustrate the qualitative properties of the transition. Our approach differs from previous treatments of superhelical cru-

ciform extrusion[12–14] in being rigorously based on the mechanical properties of DNA. More precise comparisons with these alternative treatments will be presented in context.

This formulation assumes that any unstabilized superhelicity that is not absorbed by cruciform extrusion is manifested as a uniform torsional deformation of the remaining duplex, with no contribution involving bending (i.e., writhe). One can show theoretically that bending occurs only after the superhelically imposed torsional stresses on the duplex surpass a threshold.[10,11] In the present application, superhelicity is absorbed by cruciform extrusion, so any stresses on the remaining duplex will usually remain small. For this reason, bending deformations are unlikely to be important. (We note that bending would alter slightly the free energy associated with highly stressed duplex regions. However, the qualitative conclusions of the analysis are unchanged by this effect.)

CRUCIFORM STRUCTURE AND ENERGETICS

Cruciforms, Superhelicity, and Twist

Consider a segment of DNA containing N base pairs, supercoiled an amount λ, in response to which a total of N_c base pairs are extruded into one or more cruciforms. Because cruciforms make no contribution to the linking number, their extrusion absorbs superhelicity:

$$\Lambda_c = -N_c/A \tag{1}$$

If $\lambda \neq \Lambda_c$, the residual superhelicity is manifested as a torsional deformation τ [rad/base pair (bp)] of the remaining $(N - N_c)$ duplex base pairs away from their unstressed conformation. At elastic equilibrium one can show that τ is the same for all stressed pairs.[11] It follows that

$$\tau = 2\pi(\lambda - \Lambda_c)/(N - N_c), \quad 0 \leq N_c < N \tag{2}$$

The total elastic strain energy associated with this twisting is $F_{Tw} = (N - N_c)C\tau^2/2$. Here C is the torsional stiffness of B-form DNA, a parameter whose value has been estimated from fluorescence depolarization to be $C \simeq 8.5 \times 10^{-12}$ erg/rad^2.[15,16] Because each cruciform stem may rotate freely about its loop end to relax torsional deformations there, F_{Tw} contains no term arising from twisting of the *intra*strand base pairs. Substitution of Eqs. (1) and (2) displays F_{Tw} at fixed superhelicity λ as a function of N_c above:

$$F_{Tw} = 2\pi^2 C(\lambda A + N_c)^2/A^2(N - N_c) \tag{3}$$

Cruciform Geometry

Consider two copies in inverted orientation of a sequence of N_R base pairs, separated by N_L intervening pairs. Suppose this site extrudes a

cruciform, each arm of which contains n_L unpaired bases in its loop, N_B base pairs in its stem, and n_D bases in its join region. Then the total number of base pairs involved in the cruciform must equal the total number of bases in each arm, so

$$N_c = n_L + 2N_B + n_D, \quad N_B > 0 \tag{4}$$

The cruciform states available to this inverted repeat are indexed by N_B, the number of *intra*strand base pairs that form. The stem region may involve only part of the inverted repeat, so any value $1 \leq N_B \leq N_R$ is possible. The unextruded state corresponds to $N_c = N_B = 0$. Here the number of loop bases is regarded as fixed ($n_L = N_L$), as is n_D. [That is, fluctuational opening of the pairs within each cruciform arm is not considered here. Although transient melting *within* the stem is energetically highly disfavored (by $\simeq$8 kcal/mol[17]), fluctuational stem opening from the ends could occur. However, this effect has little influence on most equilibrium properties (such as cruciform probabilities, distributions among states or expected torsional deformation) and may be disregarded in these cases without sacrificing accuracy. In calculations of properties where such fluctuations might be significant (e.g., the expected number of unpaired bases), they may be incorporated simply by noting that n_L, N_B, n_D may vary separately in each arm subject to the constraint of Eq. (4) and enumerating the cruciform states accordingly.] It follows that an increase in the length N_c of the cruciform occurs by increasing the number of arm base pairs. Because n_D and n_L are regarded as fixed, this can only happen if an equal number of base pairs are disrupted in the duplex regions. That is, the cruciform grows by disrupting *inter*strand pairs to form *intra*strand pairs, with no net change in the total number of base pairs in the segment.

The free energy $F_c(N_B)$ associated with a particular state of the cruciform includes contributions from the conformation of the deformed join region and from the entropy of the single-stranded loops, as well as the free energy F_{BP} needed to denature all single-stranded bases in that state. Because n_L, n_D, and hence the total number of paired bases, are constant in all extruded states, the value of F_{BP} varies only with the sequence of bases in the join region. Computation of examples shows that this influence is very small. It will be disregarded in the sample calculations given below. That is, the free energies F_c of the extruded states of this cruciform are essentially constant independent of stem length N_B (>0). However, F_c does depend on both the length and composition of the loop region:

$$F_c = a_D + a_L + bn_L + 2\alpha RT \ln n_L \tag{5}$$

Here $b = \Delta H[1 - (T/T_m)]$ is the sequence-average free energy needed to melt a loop base pair under given conditions, while a_L ($\simeq$8 kcal/mol[17]) is the extra free energy required to initiate this transition. The entropic coefficient is $\alpha = 1.8$ if excluded-volume effects are included; $\alpha = 1.5$ otherwise.[18–20] The free energy a_D is associated with the conformation of the deformed join region.

Although the examples developed below involve perfect inverted repeats exclusively, defects such as mispaired or unpaired bases may be included. One must enumerate the states of the system appropriately and associate free energies to these defects in all states where they occur. Although the values of these free energies are not presently known, sample calculations may be performed using the free energies associated with the corresponding defects in RNA.[21-23]

MECHANICAL EQUILIBRIUM ANALYSIS

Consider a segment of N base pairs containing one inverted repeat, which is supercoiled an amount λ. The state of this system in which a cruciform is not extruded may be seen from Eq. (3) to have free energy

$$F_0 = 2\pi^2 C\lambda^2/N \tag{6a}$$

Expressed in terms of the superhelix density $\sigma = A\lambda/N$, this is

$$F_0 = 2\pi^2 CN\sigma^2/A^2 \tag{6b}$$

If a cruciform is extruded involving N_c base pairs, its free energy F_1 includes contributions both from the cruciform formation itself and from any deformation of the residual duplex:

$$F_1 = F_c + 2\pi^2 C(N_c + \lambda A)^2/A^2(N - N_c) \tag{7}$$

The minimum of F_1 occurs when

$$N_c = -\lambda A \tag{8}$$

The state of the system in which a cruciform is extruded containing this number of base pairs is therefore an alternative mechanical equilibrium conformation whose free energy is $F_1 = F_c$. In this state, superhelicity is exactly balanced by the cruciform, $\lambda = \Lambda_c$, so that the torsional deformation vanishes [i.e., $\tau = 0$, from Eq. (2)]. If the maximum length of the cruciform is not sufficient to balance the imposed superhelicity, then the extruded mechanical equilibrium state involves the complete inverted repeat, with the remaining superhelicity serving to twist the duplex at a rate τ given by Eq. (2).

The smallest possible cruciform has $N_B = 1$ bp in each arm. It follows that the alternative mechanical equilibrium state comes into existence at superhelicities where

$$\lambda \leq -(N_L + n_D + 2)/A \tag{9}$$

In this case a population of identical molecules distributes itself between these states at equilibrium in a manner described by Boltzmann's Theorem. That is, if p_0 (respectively p_1) is the probability of a molecule being in the unextruded (respectively equilibrium cruciform) state, then

$$p_0 = \begin{cases} 1, & \text{if } \lambda > -(N_L + n_D + 2)/A \\ 1/(1 + \exp[(F_0 - F_1)/kT]), & \text{otherwise} \end{cases} \tag{10a}$$

$$p_1 = 1 - p_0 \tag{10b}$$

Because F_0 increases quadratically with λ whereas F_1 is constant (at least until $N_B = N_R$, so the cruciform is complete), sufficient superhelicity favors the cruciform state. The properties of this transition are shown in Fig. 2 below for a specific case.

Equation (6b) shows that the free energy F_0 of the unextruded state depends quadratically on the superhelix density σ *and* linearly on the length. (That is, supercoiling two molecules to the same superhelix density requires free energies proportional to the lengths of the molecules involved.) These dependencies have been experimentally verified.[24] It follows that cruciform extrusion becomes energetically favored at a superhelix density that depends on segment length N, other factors remaining fixed. A particular inverted repeat (with fixed intervening region) will extrude a cruciform at smaller superhelix densities when it is embedded in longer segments.

This mechanical analysis is similar to a calculation presented by Hsieh and Wang.[12] However, they do not consider states of partial cruciform extrusion. Further, their example does not acknowledge the dependence of the free energy F_0 (and hence also of the superhelical extrusion probability) on segment length N. These omissions invalidate their conclusion that cruciforms will not be extruded at physiological superhelix densities, a result that is also inconsistent with experimental observations.[2-6] In contrast, the present analysis suggests that even relatively small superhelix densities can favor cruciform extrusion in segments of intermediate length ($N \gtrsim 2000$ bp) under physiologically reasonable conditions.

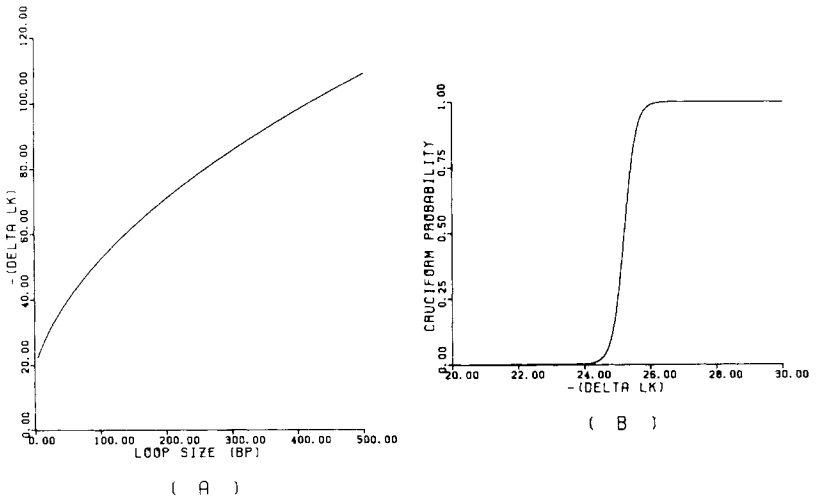

Fig. 2. Plots of important properties of the cruciform transition profiles, as computed from the mechanical theory for a segment of $N = 10000$ bp with $b = 1.5$ kcal/mol bp at $T = 300$ K. (A) Plot of the value of $-\lambda$ ($= -\Delta Lk$) occurring at the midpoint of the transition (i.e., where the free energies of the extruded and unextruded states are equal) as a function of loop size N_L. (B) Sample transition profile (here using $N_L = 10$) showing the abrupt, sigmoidal nature of the transition.

STATISTICAL-MECHANICAL THEORY

The equilibrium properties of a population of identical heteropolymeric molecules may be derived from the governing partition function Z:

$$Z = \sum_r \exp(-F_r/kT) \tag{11}$$

Here the index r enumerates the accessible states (which in this case are discrete) and F_r is the free energy of state r. The equilibrium fractional occupancy of state r is

$$p_r = [\exp(-F_r/kT)/Z] \tag{12}$$

If a given parameter has value q_r in state r, then its expected value for the population is

$$\bar{q} = \sum_r p_r q_r \tag{13}$$

To evaluate the partition function governing superhelical cruciform extrusion one must enumerate the accessible states and evaluate the free energy of each. The free energy in this case is the sum of contributions from duplex twisting and cruciform extrusion:

$$F_r = F_{Tw} + F_c \tag{14}$$

F_{Tw} is given in Eq. (3) above, while the value of F_c for each cruciform may be found from Eq. (5).

To correctly enumerate the possible states of a segment containing multiple copies, both tandem and inverted, of one or several base sequences, one must note that certain collections of cruciforms may be mutually incompatible. To see this, consider a segment containing two distinct inverted repeats. If one copy of each sequence falls within the loop region of the other repeat, then formation of either cruciform precludes extrusion of the other.

To analyze cruciform compatibility in general, suppose that a segment of DNA contains sequences (labeled A, B, C, . . .) which exist in both direct (unbarred) and inverted (barred) copies in an order such as that shown in Fig. 3. Denote intrastrand base pairing between inverted repeats by joining the copies involved with a bracket from below. Then two cruciforms are

Fig. 3. Segment of DNA may contain several repeated sequences, each of which is denoted by a letter in this diagram. Direct repeats are unbarred, whereas inverted repeats are barred. Intervening sequences are not shown. Brackets joining inverted repeat copies denote intrastrand base pairing between the specified sequences. The two brackets given by dashed lines are incompatible with the set of cruciforms denoted by solid brackets. Note that nesting of brackets can occur.

mutually incompatible precisely when their corresponding brackets cross. One can easily construct an algorithm based on this criterion that enumerates all sets of mutually compatible cruciforms for any sequence containing multiple direct and/or inverted repeat sequences.

The pattern of intrastrand base pairing in a mutually compatible set of cruciforms may be quite complex. In particular, nested sets of brackets correspond to situations where one cruciform exists entirely within the loop region of another. When this happens it is the outermost cruciform of the nested set that absorbs superhelicity. That is, extrusion of the outermost cruciform (such as $B\overline{B}$ in the nested set of Fig. 3) disrupts all interstrand base pairs throughout the loop region separating the repeat copies involved. Although this loop may possess internal inverted sequence complementarities, formation of interior cruciforms does not affect the number N_c of interstrand base pairs disrupted by the nested set, the amount $\Lambda_c = -N_c/A$ of superhelicity absorbed thereby, or the torsional deformation τ of the remaining interstrand duplex.

These considerations together with the analysis presented above (Cruciform Structure and Energetics) show that the partition function governing superhelical cruciform extrusion in a segment containing a particular sequence of direct and inverted repeats is

$$Z = \exp\left(\frac{-F_0}{kT}\right) + \sum_l \left[\sum_{m_l} \left(\sum_{n_{m_l}=1}^{N_{Rm}} \exp\left(\frac{-F_{lmn}}{kT}\right)\right)\right] \tag{15}$$

Here, l indexes the collection of all mutually compatible sets of cruciforms for this sequence, m_l enumerates the individual cruciforms within the lth compatible set, and n_{m_l} is the stem length of the m_lth cruciform. The free energy F_{lmn} consists of the elastic strain energy arising from the torsional deformation of the residual interstrand duplex [Eq. (3)] together with the free energy associated with the pattern of single-stranded regions in that state. The free energy $F_0(\lambda)$ associated with the state in which no cruciforms are extruded is found from Eq. (3) to be $F_0(\lambda) = 2\pi^2 C\lambda^2/N$.

In a previous statistical-mechanical approach to superhelical cruciform extrusion, energy was associated with superhelicity in a completely nonspecific way, whereas the duplex was regarded as torsionally undeformable.[13,14] The conclusion reached was that extrusion could occur only at the greatest physiological superhelicities. In contrast, the present analysis, which is founded on more precise mechanical properties of DNA, deduces that cruciform extrusion may occur much more readily.

EQUILIBRIUM TRANSITION PROPERTIES

In this section, important equilibrium properties of superhelical cruciform extrusion are illustrated with sample calculations based on the theory developed above. In all computations reported here we assume that T = 300 K, $n_D = 6$, the loop region is 50% G + C, and all other environmental parameters are set to approximately physiological values so that $b = 1.5$

kcal/mol. The cruciform free energy is given by Eq. (5) with $a_D = 15$ kcal/mol, $a_L = 8$ kcal/mol, and $\alpha = 1.8$. Because the precise values of some parameters are uncertain at present (n_D, a_D), the results reported here should not be regarded as numerically precise predictions. They are intended instead to depict the qualitative aspects of the transition, which will not change with future refinements of parameter values.

One Inverted Repeat

The cruciform states of a system containing one inverted repeat are indexed by N_B, the number of intrastrand base pairs in each arm. The free energies corresponding to the states of this system are

$$F = \begin{cases} 2\pi^2 C\lambda^2/N, & N_B = N_c = 0 \\ F_c + [2\pi^2 C(N_c + \lambda A)^2/A^2(N - N_c)], & 1 \le N_B \le N_R \end{cases} \tag{16}$$

Substitution of this expression into the partition function permits calculation of the fractional occupancy (or probability) $p(N_B)$ of each state from Eq. (12). Then the expected number $\overline{N}_c$ of cruciform base pairs and the expected duplex torsional deformation $\bar{\tau}$ are given by

$$\overline{N}_c = \sum_{N_B=1}^{N_R} (n_D + N_L + 2N_B)p(N_B) \tag{17a}$$

$$\bar{\tau} = \frac{2\pi\lambda}{N} p(0) + \sum_{N_B=1}^{N_R} \frac{2\pi[\lambda + (N_c/A)]p(N_B)}{(N - N_c)} \tag{17b}$$

To understand the basic properties of this transition, recall that cruciform extrusion involving N_c base pairs absorbs superhelicity $\Lambda_c = -N_c/A$. The minimum amount of superhelicity that a cruciform may absorb occurs when each cruciform arm contains a single base pair. If $N_B = 1$, then $N_c = 8 + N_L$; so Λ_c (min) $= -(8 + N_L)/A$. The threshold for the existence of a cruciform is plotted in Fig. 4(A). Although cruciforms may occur in principle whenever $\lambda \lesssim \Lambda_c$(min), in practice significant cruciform extrusion first occurs at more extreme superhelicities, the precise position depending on both N_L and segment length N as shown below. For a given value of λ the state of minimum cruciform free energy was shown to occur when $N_c = -\lambda A$. If base sequence permits extrusion of a cruciform of this size, then one expects a population of identical extruding molecules to be distributed about this most stable state. In other cases, significant torsional deformation will occur.

To demonstrate the effect of loop length N_L on the transition, let the repeat copies, each containing $N_R = 20$ bp, be embedded within a segment of $N = 1000$ bp. The results of these calculations are shown in Fig. 4(B–D) for three different loop lengths: $N_L = 10$ bp (denoted by squares), $N_L = 20$ bp (circles), and $N_L = 40$ bp (crosses). Figure 4(B) shows the probability of cruciform extrusion as a function of imposed (negative) superhelicity, $-\lambda$. Because the free energy F_c of cruciform formation increases with loop length [Eq. (5)], extrusion does not become energetically favored as N_L is

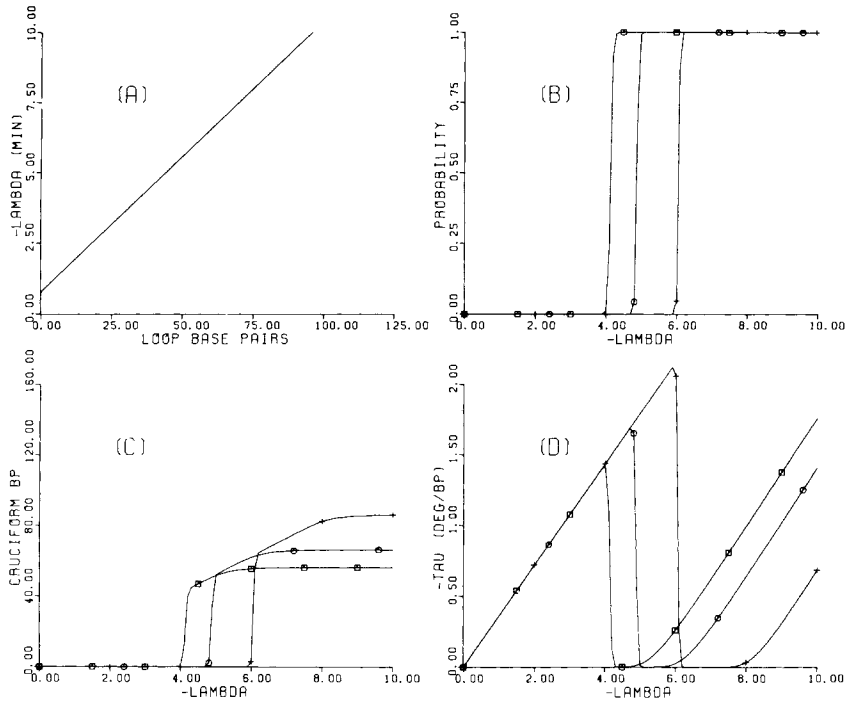

Fig. 4. (A) The smallest amount of superhelicity that can be absorbed by cruciform extrusion involving an inverted repeat sequence whose copies are separated by N_L loop base pairs. (B–D) Plots of the important properties of the superhelical cruciform transition for an inverted repeat whose copies contain N_R = 20 bp and which are separated by loop regions of varying lengths: N_L = 10 bp (denoted by squares), N_L = 20 bp (circles), N_L = 40 bp (crosses). (B) The probability of cruciform extrusion in each case as a function of negative superhelicity, $-\lambda$. (C) The expected cruciform length $\overline{N}_c$ and (D) the expected duplex torsional deformation $\bar{\tau}$, both as functions of $-\lambda$.

increased until correspondingly more extreme superhelicities. (A similar result would occur if the loop length were fixed but loop sequence and/or environmental conditions were altered to increase b or decrease T.) The onset of transition is generally delayed beyond the existence threshold $\lambda = \Lambda_c(\min)$. For example, one has $\Lambda_c(\min) \simeq -1.7$ when $N_L = 10$, whereas transition occurs at $\lambda \simeq -4$.

Figure 4(C) depicts the expected cruciform length $\overline{N}_c$ as a function of (negative) superhelicity, $-\lambda$. Beyond the transition point, $\overline{N}_c$ increases essentially linearly with $-\lambda$ until the complete inverted repeat is involved. (Recall that the cruciform may contain a minimum of $N_L + 8$ bp and a maximum of $N_L + n_D + 2N_R$ bp.) At the transition point the cruciform occurs in a state of intermediate extrusion. For example, when $N_L = 10$ the transition point occurs at $\lambda \simeq -4$, for which $\overline{N}_c \simeq 44$ bp. Therefore, $\overline{N}_B = 14$ bp $[= (\overline{N}_c - n_D - N_L)/2]$ in each arm at this threshold, out of a possible $N_R = 20$ bp. In this case, the transition occurs to a state in which the cruciform is mostly formed. In contrast, when $N_L = 40$, transition

occurs at $\lambda \simeq -6$ to a cruciform with $\overline{N}_c \simeq 64$ bp; this corresponds to $\overline{N}_B = 9$ intrastrand base pairs. The number $\overline{N}_B$ of arm base pairs occurring at the transition point decreases as loop length N_L increases, other factors remaining fixed.

Figure 4(D) displays the expected value of the duplex torsional deformation $(-\bar{\tau})$ as a function of $-\lambda$. Prior to transition all superhelicity is absorbed by twist, so $\bar{\tau}$ increases linearly with λ [see Eq. (2) with $\Lambda_c = N_c = 0$]. In states of intermediate formation ($N_B < N_R$), superhelicity is absorbed by the cruciform, so $\bar{\tau} = 0$. The duplex regions remain unstressed until the superhelicity exceeds that value at which the complete inverted repeat is extruded. Beyond that point the duplex again becomes stressed.

To illustrate the influence of segment length N on the transition, consider an inverted repeat having $N_L = 10$ bp, $N_R = 40$ bp embedded within segments of varying lengths. In this case, the possible cruciform lengths are $18 \leq N_c \leq 96$. Figure 5(A–D) displays the calculated distributions among the states (here indexed by N_c) for segments of various lengths N that are supercoiled an amount λ denoted above the corresponding curve. As segment length increases, transition to cruciforms occurs at more extreme superhelicities and involves a jump to states having longer cruciforms, other

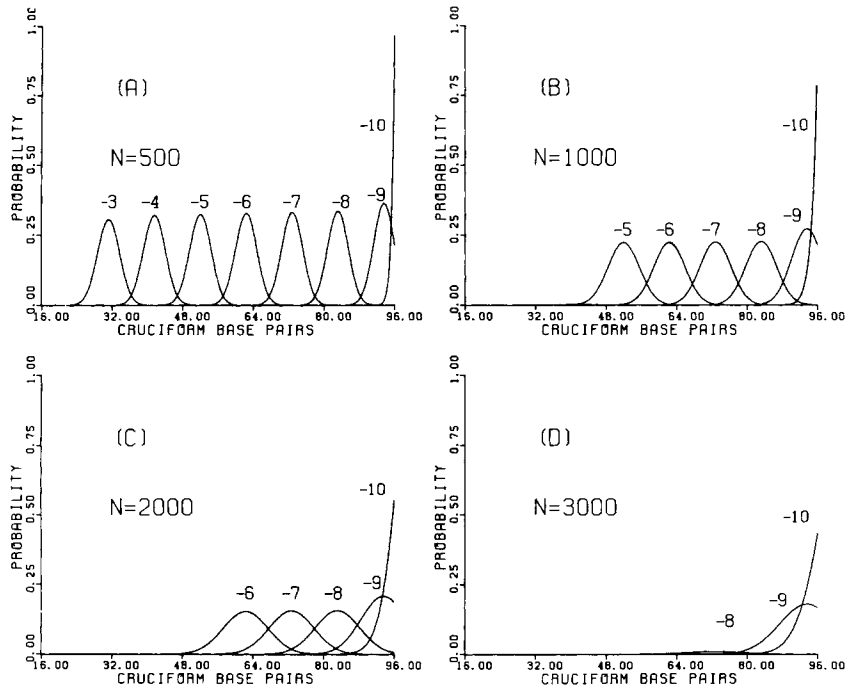

Fig. 5. Graphs of the calculated distributions among the cruciform states of a system consisting of an inverted repeat having copy size $N_R = 40$ bp and loop length $N_L = 10$ bp, embedded in a DNA segment whose total length N is shown. The superhelicity at which each distribution was calculated is given above the corresponding curve.

factors remaining fixed. In sufficiently long segments ($N \gtrsim 4000$ bp in this example), the transition to cruciform states effectively becomes a switching from the unformed state ($N_c = N_B = 0$) directly to the completely formed one ($N_B = N_R$). Only in reasonably short segments (the exact value depending on properties of the inverted repeat) are there superhelicities at which states of intermediate cruciform length are populated. When this happens the resulting probabilities of occupancy are given by Boltzmann distributions. For a fixed value of λ the mean of this distribution does not depend on segment length, although the standard deviation increases as $(N - N_c)^{1/2}$.

Two Inverted Repeats

A more complex repertoire of behaviors is possible in a segment of DNA containing multiple inverted repeats. To illustrate, consider a sequence containing two independent, mutually compatible inverted repeats (that is, ordered $A\bar{A}B\bar{B}$). Denote these repeats as IR1 and IR2, and the cruciforms they extrude as C1 and C2. As the negative superhelicity increases, the first cruciform to extrude will be the one requiring the least free energy of formation, say C1. Extrusion will proceed until superhelicities at which the complete inverted repeat is involved, $\lambda \simeq \Lambda_{C1}(\max)$. Beyond this point two alternatives are possible. If C1 can afford greater (maximal) superhelical relief than C2 (i.e., $\Lambda_{C1}(\max) < \Lambda_{C2}(\max) < 0$), then C1 remains complete, the additional superhelicity serving to twist the duplex until C2 also extrudes. However, if C2 can absorb more superhelicity than C1, then reabsorption of C1 coupled with extrusion of C2 can be the energetically favored response. One sees that the equilibrium pattern of cruciforms in a segment containing multiple inverted repeats can depend on superhelicity in a complex way. If kinetics permits equilibrium to be achieved, some cruciforms may switch off as others switch on.

To illustrate these properties, consider a segment of $N = 1000$ bp containing two inverted repeats, the copies containing $N_{R1} = 20$ bp and $N_{R2} = 30$ bp, respectively. Let the loop regions of each require $b_1 = b_2 = 1.5$ kcal/mol/bp to melt, so that the first cruciform to extrude will be the one with the shortest loop length N_L. Calculations are performed assuming that $N_{L2} = 12$ bp, while N_{L1} may have any of three values, $N_{L1} = 10$, 12, or 14 bp. It follows that C2 can absorb more superhelicity than C1 in all cases, although which cruciform forms first depends on the value of N_{L1}. The results of sample calculations on this system are presented in Fig. 6.

Consider first the top row of figures where $N_{L1} = 14$ bp, $N_{L2} = 12$ bp. In this case, the cruciform that can absorb more superhelicity is also the one that forms first, C2. The extrusion of C2 proceeds essentially as though IR2 were present alone until $\lambda \simeq -12$. At this point, C2 is fully formed and the duplex torsional deformation is so extreme that the simultaneous extrusion of C1 becomes energically favored.

The second row of figures presents the case where $N_{L1} = N_{L2} = 12$ bp,

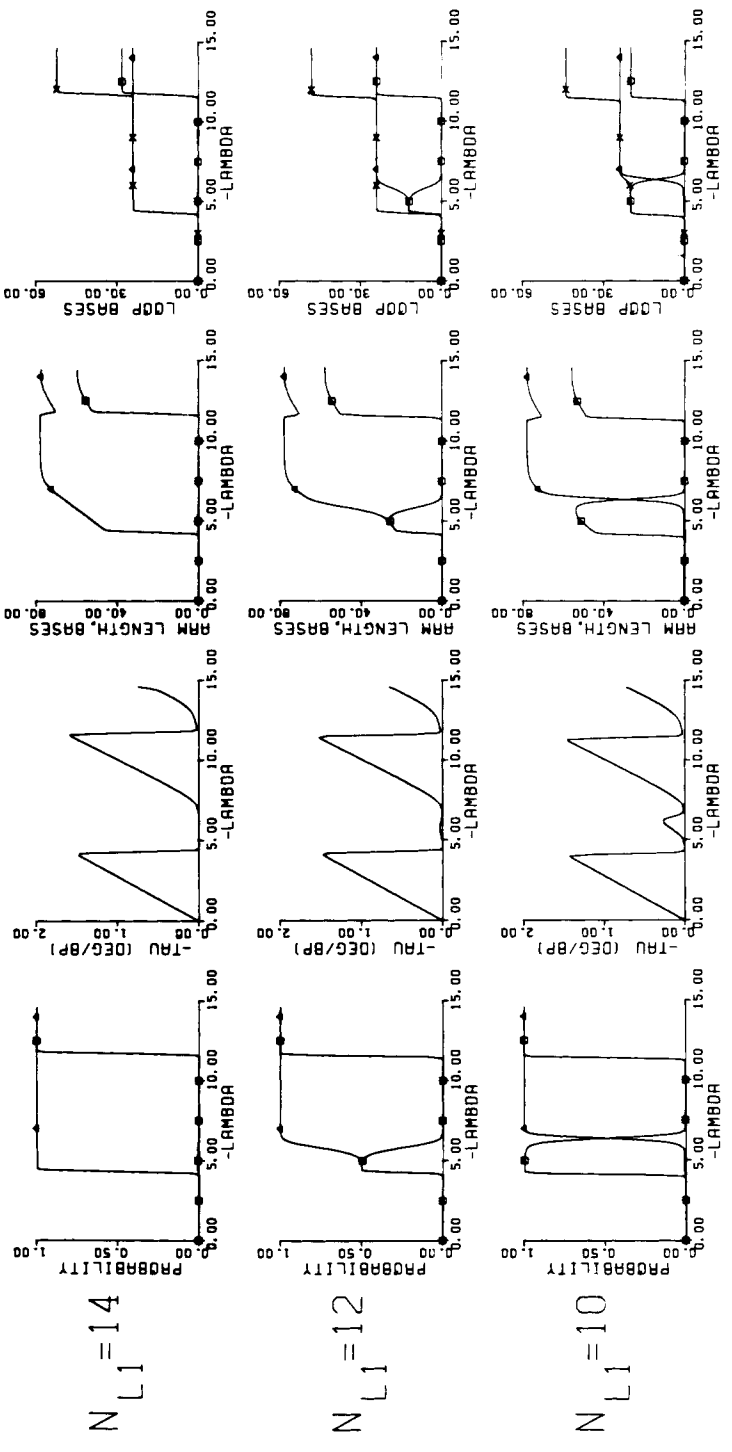

Fig. 6. Calculated equilibrium properties of the superhelical cruciform transition are shown for a segment of $N = 1000$ bp containing two independent inverted repeats. Curves corresponding to cruciform 1 are labeled with squares; those relating to cruciform 2, triangles. The copy lengths of each repeated sequence are chosen to be $N_{R1} = 20$ bp, $N_{R2} = 30$ bp. The loop length N_{L1} separating the copies of repeat 1 is shown to the left of the row of figures to which it pertains, while $N_{L2} = 12$ bp.

so that $F_{C1} = F_{C2}$. Again, the onset of transition occurs near $\lambda \simeq -4.2$, with either cruciform (but *not* both simultaneously) having an equal probability of initial extrusion. If C1 is fully formed, any incremental superhelicity must be partitioned either to twist or to extrusion of C2, both processes requiring energy. The alternative is for C1 to be reabsorbed while C2 is extruded, a process requiring no additional free energy. Therefore, although both cruciforms are equiprobable for small negative superhelicities, C2 comes to dominate as $-\lambda$ increases, whereas the probability of C1 returns to zero.

Consider the bottom row of graphs, in which $N_{L1} = 10$ bp, $N_{L2} = 12$ bp. Now C1 forms first, although C2 can absorb more superhelicity. Once C1 is fully formed ($\lambda \simeq -6$), incremental superhelicity appears as a torsional deformation τ, requiring additional energy. When the total free energy involved exceeds that required to form C2 alone ($\lambda \simeq -6.5$), C1 is reabsorbed while C2 is extruded at equilibrium.

Two properties are common to all these cases. Significant torsional deformation occurs only in situations where existing cruciforms (if any) are of maximal length. Also, the simultaneous extrusion of C1 and C2 occurs only at extreme superhelicities.

The last column of graphs gives the expected number of loop bases in each cruciform and also the total number of loop bases (denoted by crosses) for the three cases. If the frequency of single-strand endonuclease attack in a region depends linearly on the number of unpaired bases there (first-order kinetics), then the probability of initial nicking within the loop of a given cruciform will vary with superhelicity in a manner proportionate to the fraction of single-stranded bases in that loop.

COMPARISON WITH ALTERNATIVE TRANSITIONS

This paper presents an equilibrium, two-state statistical-mechanical analysis of cruciform extrusion at inverted repeats in a supercoiled segment of DNA. As DNA is actually a polymorphic substance, its superhelical behavior is likely to involve competition between several types of local structural transitions. Local denaturations or transitions to Z-form could occur, as well as (or instead of) cruciform extrusions.[25-27] Each of these conformations has its own sequence specificity and conditions for stability. Under these circumstances one expects that a wide variety of superhelical secondary structures are possible. Which actually occurs in a particular molecule would depend in a sensitive way on base sequence, segment length, environmental conditions, and the degree of superhelicity imposed. The theoretical analysis of supercoiled structure in this generality requires a multistate, heteropolymeric statistical-mechanical theory, which the author is developing at present.

Although a rigorous analysis of the competition between cruciform extrusion and alternative modes of transition must await future development, certain qualitative comparisons may be made. First, these three transitions

(to Z-form, to cruciform, or local denaturation) differ in their scale and sequence specificities. Cruciform extrusion requires a degree of inverted repeat homology, a property whose expression involves sequences of some length. In contrast, transitions to Z-form appear to require alternating purine–pyrimidine sequences. Although local denaturation could (in principle) involve any base pair, it is most probable at the most (A + T)-rich regions under physiological conditions. We note that the specific sequences of base pairs that are melting (or possibly switching to Z-DNA) may be quite short. This suggests that cruciform transitions may play a role in certain phenomena involving fairly large superhelical domains, whereas their alternatives could dominate on a smaller scale. However, domains not containing inverted repeats obviously cannot extrude cruciforms, regardless of their size.

The free energy associated with a supercoiled molecule containing regions in one of the alternative stable and stressed conformations (either Z-form or locally melted) has been shown to vary linearly with the imposed torsional deformation.[25-27] As with the cruciform case, these transitions also eventually become favored because the conformational free energy of the untransformed structure is quadratic in the deformation, hence eventually surpasses any line. We note that the analysis of these alternatives presently deals only with that part of the superhelicity that is partitioned to twist. However, once one of these transitions has been initiated, the linearity of the resulting torsional contribution to the total conformational free energy alters the partitioning of incremental superhelical deformation in favor of twist. In this way the free energy of the alternate conformation may become effectively linear in λ. However, this free energy certainly increases with the magnitude of the superhelical deformation $|\lambda|$, whereas that of the perfect cruciform remains constant (at least until the complete inverted repeat is involved). For this reason, as the negative superhelicity increases, the cruciform transition should eventually become preferred in a domain whose sequence permits it. It must be emphasized that the interplay between these types of transitions may be quite complex in practice, involving the possible coexistence of multiple transformed regions within a single stressed domain.

Experimentally, the nicking of superhelical DNA by single-strand-specific endonucleases provides a measure of the presence of denatured regions within a domain. Such regions can occur either as the loops found in the cruciform arms or through stress-induced denaturation. As shown above, cruciforms can form in a stable way only in negatively supercoiled domains, whereas (A + T)-rich sites may denature stably when subjected to either slightly negative or large positive superhelicities.[25] Single-strand-specific endonuclease digestion has been observed in PM2 DNA under both of these conditions, suggesting that stress-induced denaturations do occur under some circumstances.[28] Further, the initial nicking rate of this reaction increases with $|\lambda|$ at both negative and positive superhelicities. This implies that the number of unpaired bases varies with $|\lambda|$, behavior ex-

pected of the stress-induced melting transition.[25] Although some of the single-stranded substrate may occur in cruciform loops, this observation is certainly not consistent with an interpretation that cruciform extrusion is the only transition that occurs in the PM2 genome. In contrast, nicking of superhelical ColE1, pBR322, and ϕX174RF DNA molecules by S1 endonuclease is observed to occur within the loop regions separating inverted repeat sequences.[3-5] This result suggests that superhelical domains containing permissive sequences may, in fact, relieve their intramolecular stresses through cruciform extrusion.

Cruciform extrusion in negatively supercoiled closed circular DNA has been observed directly in artificially produced head-to-head dimer molecules.[6] Here, the high degree of inverted repeat homology and the small size of the unpaired loop region together favor the cruciform transition over its alternatives. Once transition was begun, the portions of the molecule not participating in the cruciform were observed to act in a relaxed fashion. Further, direct electron-microscopic visualization showed that the length of the cruciform arms varied directly with the amount of negative superhelicity imposed. These results are in complete qualitative accord with the conclusions of the present analysis.

DISCUSSION AND CONCLUSIONS

The analysis presented in this paper suggests that cruciform extrusion may be a natural consequence of negative superhelicity. In particular, the cruciforms extruded by a sequence containing multiple repeats have been shown to be intricately interdependent, with some being reabsorbed as others form. It has been shown that a natural DNA sequence possessing a major inverted repeat symmetry is very likely in practice to possess multiple, more minor such homologies.[29] For this reason the interaction of multiple repeats is of practical importance.

Substrate superhelicity is known to influence activities involved in DNA replication, recombination, transcription, and repair.[30-33] This control may be exerted through the creation of local sites of altered secondary structure. If specific enzymic interactions endow these sites with regulatory properties (as experiments suggest in the case of cruciform extrusion in at least one instance[34]), then modulation of superhelicity could control an intricately orchestrated repertoire of activities. The present analysis shows that inverted repeat sequences may be sites of precisely regulated and delicately interacting superhelical cruciform extrusions. In this regard, the fact that inverted repeats occur in transposons, retroviral proviruses, and related structures is especially significant,[35,36] since the insertion or deletion of one inverted repeat can profoundly affect the occurrence of other stress-induced transitions. (As an example, the presence of inverted repeat II drastically alters the circumstances under which cruciform I is extruded in Fig. 6.) Such an insertion could radically change the control of events mediated by the DNA into which it is made, simply by altering the patterns,

types, and/or probabilities of occurrence of other, putatively regulatory stress-induced transitions. For instance, integration of the proviral long terminal repeat from the Moloney sarcoma virus is known to activate the transforming potential of a normal cell sequence (a potential oncogene), presumably by altering the controls on its genetic expression.[37]

An alternative structure has been proposed whereby an inverted repeat sequence assumes a base-paired, four-stranded conformation.[38] Although the stability and energetics of this conformation are not presently known, it is clear that its formation can absorb superhelical deformation. In this structure, the sequence intervening between the inverted repeat copies remains base-paired, in contrast to the denaturation in single-stranded loops required in the cruciform. This strand separation constitutes the dominant contribution to the free energy of cruciform formation, which can be quite large if the intervening sequence is long. For this reason, formation of the Lim-Mazanov four-stranded structure could be energetically favored in negatively supercoiled molecules containing inverted repeats having long intervening sequences, although it has not been observed in synthesized dimers.[6]

This work was supported in part by National Science Foundation Grant PCM-8002814.

References

1. von Heijne, G. & Blomberg, C. (1976) *J. Theor. Biol.* **63,** 347–353.
2. Woodworth-Gutai, M. & Lebowitz, J. (1976) *J. Virol.* **18,** 195–204.
3. Lilley, D. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 6468–6472.
4. Panayotatos, N. & Wells, R. (1981) *Nature* **289,** 466–470.
5. Lilley, D. (1981) *Nucleic Acids Res.* **9,** 1271–1289.
6. Gellert, M., Mizuuchi, K., O'Dea, M., Ohmori, H. & Tomizawa, J. (1978) *Cold Spring Harbor Symp. Quantum Biol.* **43,** 35–40.
7. Fuller, F. B. (1971) *Proc. Natl. Acad. Sci. USA* **68,** 815–819.
8. Crick, F. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 2639–2643.
9. Fuller, F. B. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3557–3561.
10. Benham, C. J. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 2397–2401.
11. Benham, C. J. (1980) *Biopolymers* **18,** 609–623.
12. Hsieh, T. & Wang, J. (1975) *Biochemistry* **14,** 527–535.
13. Vologodskii, A., Lukashin, A., Anshelevich, V. & Frank-Kamenetskii, M. (1979) *Nucleic Acids Res.* **6,** 967–982.
14. Anshelevich, V., Vologodskii, A., Lukashin, A. & Frank-Kamenetskii, M. (1979) *Biopolymers* **18,** 2733–2744.
15. Barkley, M. & Zimm, B. (1979) *J. Chem. Phys.* **70,** 2991–3007.
16. Millar, D., Robbins, R. & Zewail, A. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5593–5597.
17. Crothers, D. & Zimm, B. (1964) *J. Mol. Biol.* **9,** 1–9.
18. Jacobson, H. & Stockmayer, W. (1950) *J. Chem. Phys.* **18,** 1600–1606.
19. Poland, D. & Scheraga, H. (1966) *J. Chem. Phys.* **45,** 1464–1468.
20. Fisher, M. (1966) *J. Chem. Phys.* **45,** 1469–1473.
21. Fink, T. & Crothers, D. (1972) *J. Mol. Biol.* **66,** 1–12.
22. Gralla, J. & Crothers, D. (1973) *J. Mol. Biol.* **78,** 301–319.
23. Tinoco, I., Boer, P., Dengler, B., Levine, M., Uhlenbeck, O., Crothers, D. & Gralla, J. (1973) *Nature [New Biol.]* **246,** 40–41.

24. Pulleyblank, D., Shure, M., Tang, D., Vinograd, J. & Vosberg, H. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 4280–4284.
25. Benham, C. J. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3870–3874.
26. Benham, C. J. (1980) *Nature* **286,** 637–638.
27. Benham, C. J. (1981) *J. Mol. Biol.* **150,** 43–68.
28. Lau, P. & Gray, H. (1979) *Nucleic Acids Res.* **6,** 331–357.
29. Galas, D. (1978) *J. Theor. Biol.* **72,** 57–73.
30. Falco, S., Zivin, R. & Rothman-Denes, L. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 3220–3224.
31. Mizuuchi, K., Gellert, M. & Nash, H. (1978) *J. Mol. Biol.* **121,** 375–392.
32. Hays, J. & Boehner, S. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4125–4129.
33. Liu, L., Liu, C. & Alberts, B. (1979) *Nature* **281,** 456–461.
34. Shlomai, J. & Kornberg, A. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 799–803.
35. Calos, M. & Miller, J. (1980) *Cell* **20,** 579–595.
36. Shimotohno, K., Mizutani, S. & Termin, H. (1980) *Nature* **285,** 550–554.
37. Blair, D., Oskarsson, M., Wood, T., McClements, W., Fischinger, P. & VandeWonde, G. (1981) *Science* **212,** 941–943.
38. Lim, V. & Mazanov, A. (1978) *FEBS Lett.* **88,** 118–123.

Received July 13, 1981
Accepted September 22, 1981