doi:10.1006/jmbi.2001.5361 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2002) **316**, 563–581

JMB

Extrusion of an Imperfect Palindrome to a Cruciform in Superhelical DNA: Complete Determination of Energetics Using a Statistical Mechanical Model

Craig J. Benham[1], Anne G. Savitt[2] and William R. Bauer[2]*

[1]*Department of Biomathematical Sciences Box 1023, Mount Sinai School of Medicine, 1 Gustave Levy Place, New York NY 10029, USA*

[2]*Department of Molecular Genetics and Microbiology Health Sciences Center, State University of New York, Stony Brook, NY 11794-5222, USA*

**Corresponding author*

We present a detailed study of the extrusion of an imperfect palindrome, derived from the terminal regions of vaccinia virus DNA and contained in a superhelical plasmid, into a cruciform containing bulged bases. We monitor the course of extrusion by two-dimensional gel electrophoresis experiments as a function of temperature and linking number. We find that extrusion pauses at partially extruded states as negative superhelicity increases. To understand the course of extrusion with changes in linking number, ΔLk, we present a rigorous semiempirical statistical mechanical analysis that includes complete coupling between ΔLk, cruciform extrusion, formation of extrahelical bases, and temperature-dependent denaturation. The imperfections in the palindrome are sequentially incorporated into the cruciform arms as hairpin loops, single unpaired bases, and complex local regions containing several unpaired bases. We analyze the results to determine the free energies, enthalpies and entropies of formation of all local structures involved in extrusion. We find that, for each unpaired structure, the ΔG, ΔH and ΔS of formation are all approximately proportional to the number of unpaired bases contained therein. This surprising result holds regardless of the arrangement or composition of unpaired bases within a particular structure. Imperfections have major effects on the overall energetics of cruciform extrusion and on the course of this transition. In particular, the extent of the ΔLk change necessary to extrude an imperfect palindrome is considerably greater than that required for extrusion of the underlying perfect palindrome. Our analysis also suggests that, at higher temperatures, significant denaturation at the base of an imperfect cruciform can successfully compete with extension of the cruciform arms.

Keywords: DNA inverted repeat; extrahelical bases; cruciform extrusion; superhelical DNA; DNA transition energetics

Introduction

It has long been clear that structures more complex than the standard DNA double helix are biologically important. These include unpaired or mismatched bases, internal loops, hairpin loops, and Holliday junctions. DNA sequences containing extrahelical bases have been implicated in at least two processes: they appear to serve as intermediates in frameshift mutations,[1,2] especially when present as extrahelical bases in extruded cruciforms;[3] and they are involved in DNA replication in certain eucaryotic viruses and organelles. For example, extrahelical bases have been shown to be required for efficient parvoviral DNA replication. In particular, the 5′ terminal hairpin of the DNA of minute virus of mice (MVM) contains a bubble of four mismatched nucleotides.[4] A closely related virus, identical in DNA sequence except lacking the bubble, has 50-fold lower infectivity. Comparison with a virus variant containing a bubble of different composition at the same site revealed that the location of the bubble, but not its base composition, is the determining factor for infectivity. In congruence with this result, we find that thermodynamic functions of bulge formation depend only upon the number of bases bulged and not upon composition.

E-mail address of the corresponding author: wbauer@ms.cc.sunysb.edu

0022-2836/02/030563–19 $35.00/0

Formation of structures involving bulged and mismatched bases can be accomplished reversibly by extrusion of the appropriate inverted repeat into a cruciform in superhelical DNA. We have developed semiempirical statistical mechanical models to calculate the free energy of formation of such structures, based upon experimental measurements of the dependence of cruciform extrusion upon linking number. By applying this approach at various temperatures, we also evaluate the enthalpies and entropies of formation of these structures. Here we describe our analytical method and apply it to evaluate the naturally occurring bulged bases in the terminal regions of vaccinia virus DNA.

The DNA of vaccinia virus is a large (192 kbp) linear duplex with cross-linked termini, such that a single strand circle forms upon complete denaturation.[5] Beginning at each end and extending inward for about 125 base-pairs, the DNA forms incompletely base-paired hairpins.[6,7] Figure 1 depicts the nucleotide sequences of these terminal hairpins, drawn so as to maximize the number of base-pairs provided that at least two occur contiguously. When drawn in this way, the structures contain four unpaired bases in the terminal loop, six isolated single base bulges (five A/T and one G/C), and two more complex regions, labeled region I and region II. Region I contains two G/C bulges separated by two putative A:T base-pairs. Region II is drawn as an open bulge of six bases, since at most a single isolated A:T pair could be formed within it. As described in detail below, our calculations make no assumptions regarding the presence of any of the nominal base-pairs in either region. The bulged bases at the left terminus are complementary to those of the right terminus at the same locations.

The left and right terminal sequences in native vaccinia virus DNA are related by complementary symmetry,[6,8] as presented in diagrammatic form in Figure 2, upper panel. That is, strand X on the left forms an imperfect hairpin with strand X′, and strand Y′ on the right forms the complementary hairpin with strand Y. The inverted repeat symmetry between X and X′, and between Y and Y′, is

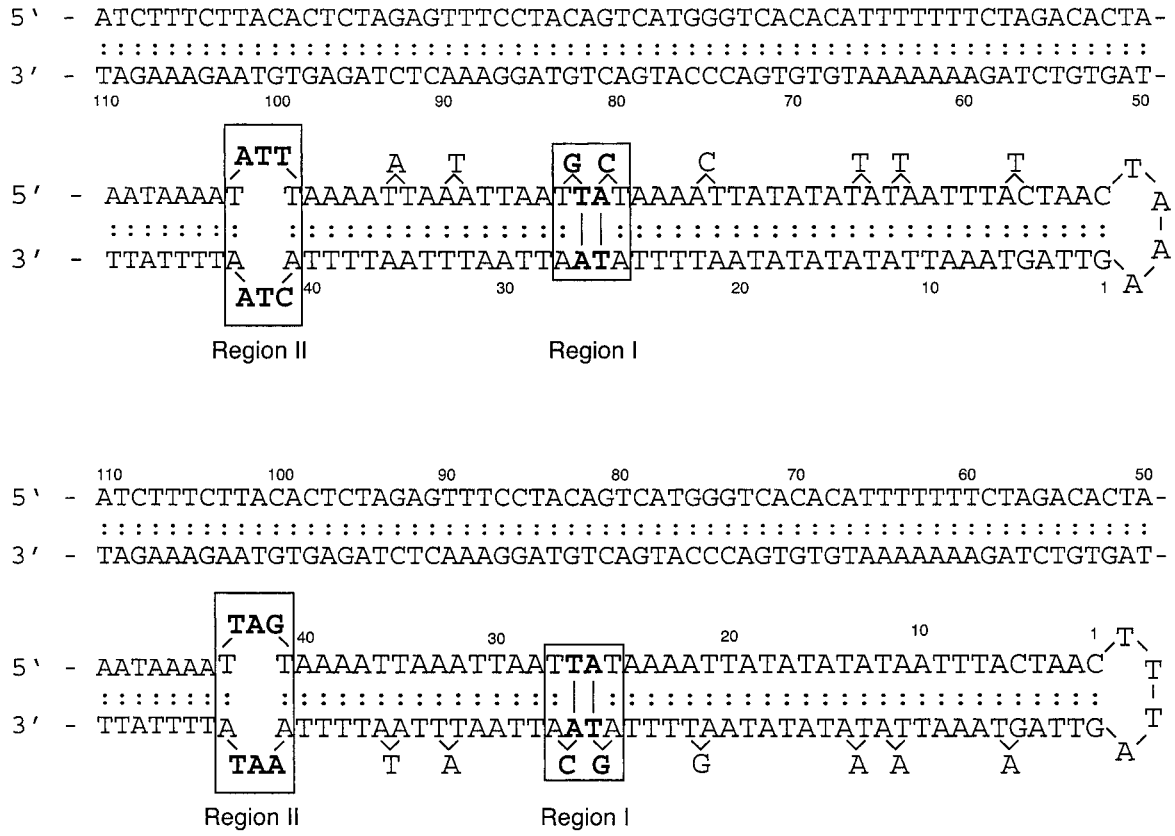

Figure 1. Nucleotide sequence of the hairpin termini of the DNA of vaccinia virus. Except for region II, the maximally base-paired structure is shown. In region II, several isolated interstrand A·T pairings are possible; in the absence of structural information, we show this region as an open loop. The numbering system begins with the first base-pair following the hairpin loops and continues sequentially along the strand with the smaller number of bulged bases. The bulge bases themselves are not counted. Regions I and II are shown in boxes; for clarity, the boxes are drawn one base-pair outside the bases included in each region. The precise nucleotides contained within both regions are shown in bold face. The count does not include the two potential A·T base-pairs at the center of region I (connected by continuous lines), because the model makes no à *priori* assumption about this structure.

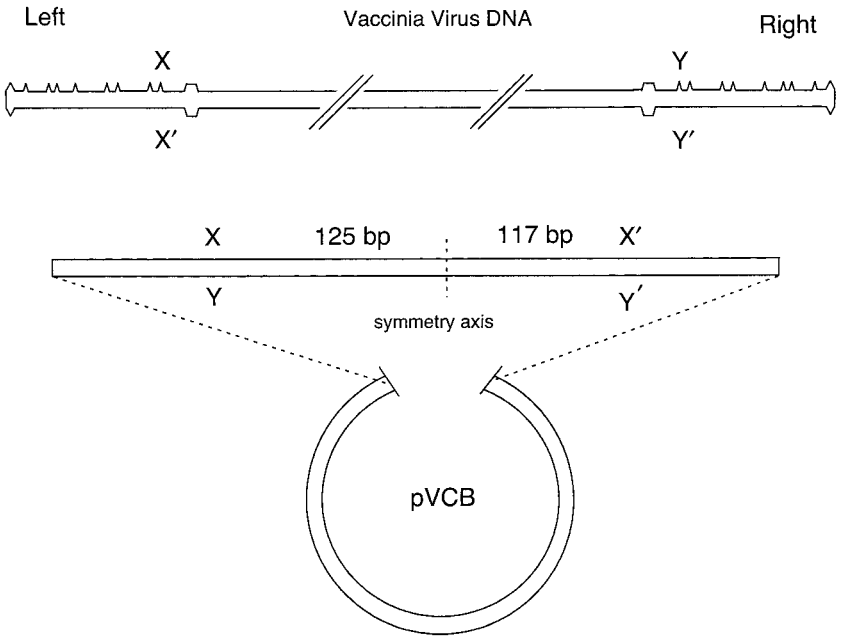

Figure 2. Schematic representation of the construction of plasmid pVCB[40]. The upper panel depicts the left and right terminal hairpin regions of the intact vaccinia virus DNA up to and somewhat beyond the end of region II. Sequence X in the upper left strand is imperfectly paired with sequence X′, and sequence Y in the upper right strand is imperfectly paired with sequence Y′. The lower panel shows the rearranged configuration as it occurs in the native plasmid. The pVCB insert is an imperfect palindrome of 242 base-pairs, 125 to the left and 117 to the right of the center of pseudosymmetry. Here X is perfectly paired with Y and X′ with Y′. This arrangement is similar to that found in the replicative intermediate of vaccinia DNA. The axis of symmetry is indicated.

violated at 12 positions. During DNA replication, the replicative intermediate is a linear concatemer[7,9] of genomic sequences that are joined in a perfectly base-paired imperfect palindrome, in which X is perfectly paired with Y and X′ with Y′. A 242 bp portion of this imperfect repeat representing the natural concatemer junction was cloned into plasmid pVCB,[10] as shown in Figure 2, lower panel. This sequence is perfectly base-paired in the relaxed plasmid, as it is in the viral replicative intermediate DNA. When the plasmid is supercoiled beyond a threshold amount, which depends upon ΔLk and temperature, the imperfect palindrome extrudes into a cruciform containing unpaired bases (Figure 3). Extrusion occurs at the junction between X:Y and X′:Y′, forming a cruciform in which X is paired with X′ in one arm, and Y with Y′ in the other arm. These cruciform arms are the same as the termini of the telomeric regions of vaccinia virus DNA. The nucleotide sequences shown in Figure 1 are those of the completely extruded cruciform arms of pVCB, as well as of the native viral telomeres.

The two-dimensional gel electrophoresis experiments reported here show that cruciform extrusion pauses at positions where the inverted repeat symmetry is interrupted. This is observed at all temperatures from 35 to 60 °C. The driving force for extrusion is a change in the extent of supercoiling. A similar phenomenon was reported previously for the DNA of the terminal regions of the Shope fibroma virus cloned into a small plasmid.[11] Although the analyses in this earlier study lacked the comprehensive statistical mechanical treatment presented here, their general results are entirely consistent with our findings for the plasmid containing vaccinia sequences.

All processes in superhelical DNA that involve changes in duplex winding are strongly coupled to global changes in superhelicity. Calculation of the energetics of formation of both perfect and imperfect cruciforms from the corresponding inverted repeats requires rigorous accounting for this coupling in all states of extrusion. We present this analysis for the first time. Our procedures are formally similar to those previously applied to the denaturation of superhelical pBR322 and pSM1 plasmids.[12,13] We use as known quantities the energy parameters whose values were determined in previous experiments, including the free energy of superhelix formation and the free energy of duplex denaturation. Using the known dependence of the superhelix free energy upon ΔLk,[12–16] we determine the energetics of the initiation event and of the incorporation of the individual imperfections. The unknown energy parameters are initially allowed to vary over predetermined, reasonable ranges. For each experimental temperature, we calculate the extent of transition for each topoisomer by the statistical mechanical procedure described below. We then assess the RMS deviation between these calculated values and the experimental measurements. We repeat the procedure until all combinations of values of these parameters lying within their prescribed ranges have been sampled. Finally, we select the set of free energies giving the best fit at each temperature, as assessed by the RMS criterion. In this way we obtain the temperature dependence of each energy parameter characteristic of the extrusion of the imperfect palindrome.

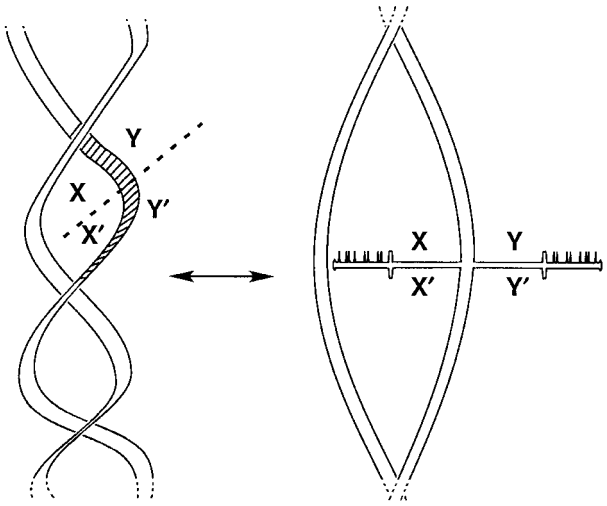

Figure 3. Schematic representation of the formation of an imperfect cruciform from the imperfect inverted repeat in plasmid pVCB. The left structure shows the imperfect inverted repeat in the supercoiled plasmid DNA prior to extrusion. The right structure shows the fully extruded cruciform and the accompanying substantial relaxation of supercoiling in the plasmid DNA.

Theoretical Analysis

Structural transitions of pVCB with changes in supercoiling

Inspection of the plasmid nucleotide sequence shows that the vaccinia virus insert is the only site within the pVCB plasmid at which cruciform extrusion is likely to be driven by negative superhelicity. The resulting extruded structures, whether partial or complete, are all characterized by a terminal loop of four unpaired bases at the end of each arm and a Holliday junction at the intersection of the plasmid duplex and the cruciform arms. In addition, all but the shortest cruciforms contain imperfections. In principle, the extrusion process can continue beyond an imperfection in either of two ways: (a) the cruciform arms extend to incorporate the imperfection either as single base bulges or as more complex unpaired structures, one in each arm; or (b) the short site of denaturation formed by the imperfection at the base of the cruciform extends, forming a larger denatured bubble either prior to or instead of incorporation into the lengthening cruciform arms.

Based upon the experimental results, we construct a model that allows cruciform extrusion to proceed through three types of imperfections, as shown in Figure 1. First, a single base bulge occurs in each arm when a single base-pair violating inverted repeat symmetry is incorporated into the cruciform. There are six of these single base-pair violations found in the vaccinia inverted repeat sequence, five A·T and one G·C. Second, in region I two single base violations of inverted repeat symmetry occur in close proximity. We allow the possibility that in the extrusion process these bases might interact to form a more complex structure than simply two independent single base bulges, and therefore ascribe a separate energy of formation to this region as a whole. The last symmetry violation encountered as extrusion progresses, designated region II, is more complex, involving up to six unpaired bases. We therefore also assign a separate energy to the incorporation of this region into the cruciform arms.

Mathematical model of the transition

We perform an equilibrium statistical mechanical analysis of the cruciform extrusion transition in pVCB. This requires first describing all the states of the system, and then ascribing a free energy to each state.

Enumeration of states

In addition to the unextruded state, there are 110 possible states of extrusion of the cruciform. These are enumerated according to the number n of base-pairs that are formed in one cruciform arm, which takes on all integer values between $n = 0$ (unextruded) and $n = 110$ (fully extruded). We emphasize that the bases within the loop, the bulges and the structures formed by regions I and II are not counted in n (refer to Figure 1).

The imperfections in the inverted repeat symmetry become successively incorporated into the cruciform arms as extrusion proceeds. The number n_d (n) of interstrand base-pairs that must be disrupted when the cruciform arm has extruded to position n is:

$$n_d(n) = 2n + n_u(n) \qquad (1)$$

where $n_u(n)$ is the number of bases within each arm that are not required to be paired. These can be either in the loop region or in the imperfections. To evaluate $n_u(n)$, we note that every extruded state has four unpaired bases in the loop region, and that longer cruciforms include successively more unpaired bases as additional imperfections in the inverted repeat sequence are encountered. As shown in Figure 1, there are six single base bulges, becoming incorporated when $n = 6$, 12, 14, 22, 33, and 36. (The structure we refer to as a ''bulge'' in fact consists of complementary single base bulges, one in each arm. The other structures also occur as complementary pairs.) The bulge next to position $n = 22$ arises from a G·C base-pair, while the other five all come from A·T pairs. Regions I and II are incorporated in all states when $n \geqslant 27$ and $n \geqslant 42$, respectively. Each of these regions is formed from bases that, in the duplex, comprise six interstrand base-pairs. To develop a simple formula for $n_u(n)$ we use the Heaviside step function $H_m(n)$, which has values $H_m(n) = 1$ if $n \geqslant m$, and $H_m(n) = 0$ otherwise. Then $n_u(n)$ may be expressed as:

$$n_u(n) = 4H_1(n) + H_6(n) + H_{12}(n) + H_{14}(n) + H_{22}(n) + 6H_{27}(n) + H_{33}(n) + H_{36}(n) + 6H_{42}(n) \quad (2)$$

The number of turns by which the unstressed interstrand helicity decreases, when a cruciform of length n is extruded, is $n_d(n)/A$, where $A = 10.4$ is the number of base-pairs per turn in *B*-form DNA. It follows that the residual superhelicity in this state is:

$$\Delta Lk_r = \Delta Lk + \frac{n_d(n)}{A} \quad (3)$$

The expression for the residual linking difference becomes more complicated when the state includes both denaturation and cruciform extrusion, as is described below.

We evaluate the residual linking difference, the primary input for our analyses, directly from the migration positions in the first dimension of the gels (see Figure 4). Gel mobility increases monotonically with the magnitude of ΔLk_r, which in topoisomers not experiencing transition is the same as the actual linking difference, ΔLk. In our work, as in all previous analyses of gel migration patterns for superhelical DNA undergoing comparable transitions, we assume that the migration velocity depends only on ΔLk_r. This assumption is supported by the observation that plasmids containing the fully extruded cruciform co-migrate with nicked circular DNA (see Figure 4), showing that the presence of cruciform arms has no effect upon the gel mobility. We measure the ΔLk_r values of topoisomers whose mobilities are reduced by cruciform formation or denaturation by comparing their positions with those of corresponding native, unextruded topoisomers. We described this procedure more fully in a previous publication.[12]

Our experimental results show that, at each temperature, there is a range of ΔLk values over which extrusion pauses at intermediate states; that is, over which ΔLk_r changes much less than ΔLk (see Figures 4 and 5). At higher temperatures, the extrusion pauses longer at the partially extruded state corresponding to the position at which region II becomes incorporated. This situation persists over several topoisomers, the number increasing with temperature. Because region II is the last imperfection encountered, and hence the last possible stopping point, the experiments demonstrate that extrusion does not pass beyond this location for these topoisomers. However, the ΔLk_r values of successive topoisomers in this region increase by less than integer values with increasing $|\Delta Lk|$. We model this by allowing incremental interstrand denaturation to occur at the cruciform base once the interstrand base-pairs of region II become separated. Because this nucleates an unpaired region, the molecule has a choice: it may either extend the region of denaturation or extend the cruciform to incorporate region II into the arms.

In principle, denaturation might be favored because the resulting unpaired strands can twist around each other. This provides more relaxation of superhelicity than does the extrusion of the same number of base-pairs into a cruciform. However, this benefit is offset by the cost of not recovering the free energy associated with base-pairing within the cruciform arms. Because the latter energy decreases as temperature increases, denaturation at the base of a cruciform that is extruded to the region II stopping point is expected to become relatively more favorable as temperature increases.

Because the experiments demonstrate that cruciform extension is not the only event that occurs, our model allows the transition to have the option to sample various denatured states. We enumerate these states according to the number n_o of open (i.e. denatured) base-pairs at each position. For each value of n_o, $1 \leqslant n_o \leqslant n_{max}$, we place the open region in all n_o+1 possible positions around the region II stopping point, and account for the exact energy of opening the specific base-pairs involved in that bubble. This alternative is not needed at region I or at single base bulges because no anomalous relaxation is observed at these points.

Energies associated with states

The free energy associated with each state of a topoisomer is the sum of the free energies required for whatever transitions (extrusion, possibly with denaturation at region II) occur in that state, plus the free energy associated with the residual superhelicity. Thus, the free energy of the unextruded state (i.e. with $n = 0$) is:

$$\Delta G(0) = K(N)\Delta Lk^2 \quad (4)$$

and the free energy associated with residual superhelicity in states where extrusion (but no denaturation) occurs is:

$$\Delta G_r = K(N)\Delta Lk_r^2 \quad (5)$$

The quadratic form of this free energy equation has been repeatedly demonstrated by experiments using a variety of different procedures,[17] and the value of $K(N)$ is known under our experimental buffer and temperature conditions. $K(N)$ varies inversely with the plasmid length in number of base-pairs, N; the effective length of the plasmid at any stage of transition is $N = 2922 - n_d(n)$. This effective shortening occurs because the portions of the molecule extruded in the cruciform are unstressed. The residual linking difference is thus distributed exclusively within the remaining interstrand duplex regions of the molecule.

An additional contribution to the free energy of a state arises from the cost of creating the imperfections in the cruciform arms. The imperfections are of three types: single base bulges, the structure of

region I, and the structure of region II. At present we make no assumptions about the specific structures that form as regions I and II are incorporated into the cruciform arms. In particular, none of the theoretically possible base-pairings that could occur within these regions are included in the cruciform state enumeration parameter n. We simply associate free energies ΔG_I and ΔG_{II} with their formation. These free energies include the free energy needed to denature the interstrand base-pairs from which these structures will be constructed, plus the free energies required to form two of these anomalous structures, one in each arm.

There are six single base violations of inverted repeat symmetry outside of regions I and II. Five of these involve A·T base-pairs and one a G·C pair. The total free energy of forming these structures includes the sequence-specific cost (b_{AT} or b_{GC}) of denaturing the base-pair involved, which is not recovered at the site of an unpaired base, plus the free energy ΔG_b of forming a pair of single base bulges, one on each arm. This bulge free energy ΔG_b is unknown but is modeled as having the same value for A·T pairs as for G·C pairs. Again, we make no assumptions about the precise conformations of these unpaired bases within the cruciform arms. The free energy ΔG_i required for the initiation of cruciform extrusion arises from the cost of creating the two loops and the junction region where the cruciform joins the interstrand duplex.

Analysis of the two-dimensional gel migration patterns shows that extrusion does not stop at any of the first three bulges, all of which are A·T. Presumably this is because the residual superhelicity here is large and readily overcomes the resistance of forming a single bulged base. Since the states in which extrusion stops at any of these first three bulges are not significantly occupied, for the purposes of the fitting procedure we incorporate the total free energies of forming these first three bulges into the effective free energy ΔG_x of cruciform extrusion:

$$\Delta G_x = \Delta G_i + 3(\Delta G_b + b_{AT}) \tag{6}$$

The value we determine in the fitting procedure described below is ΔG_x. Because all other investigators report values of ΔG_i, our values can only be compared to theirs after subtraction of the free energy cost of these first three bulges.

The states in which denaturation occurs at the base of region II are handled separately, as described below. The free energy ΔG_t of transition for all other states of extrusion (i.e. with $n > 0$) is:

$$\begin{aligned}\Delta G_t(n) = {} & \Delta G_x + H_{22}(n)(\Delta G_b + b_{GC}) \\ & + H_{27}(n)\Delta G_I + H_{33}(n)(\Delta G_b + b_{AT}) \\ & + H_{36}(n)(\Delta G_b + b_{AT}) + H_{42}(n)\Delta G_{II}\end{aligned} \tag{7}$$

and the total free energy $\Delta G(n)$ associated with these states is:

$$\Delta G(n) = \Delta G_t(n) + \Delta G_r \tag{8}$$

Finally, we consider the states of denaturation that occur at region II. Because all these states have $n = 41$, the plasmid has effective length $N_d = 2824$ bp. That is, 98 interstrand base-pairs have been disrupted to form the cruciform, whose extrusion has absorbed $98/A$ turns. This leaves a total linking difference of $\Delta Lk_d = \Delta Lk + (98/\mathrm{A})$ to drive denaturation within this 2824 bp domain. Suppose that n_o base-pairs denature at the base of the cruciform, $1 \leqslant n_o \leqslant n_{max}$. (The case where $n_o = 0$ has been handled above.) There are $n_o + 1$ possible locations for this run of denaturation, which may be indexed by the number m of base-pairs that are 5′ to the cruciform, $0 \leqslant m \leqslant n_o$. The other $n_o - m$ denatured base-pairs then will be located on the 3′ side of the cruciform. Then the free energy ΔG_d associated with this state of denaturation is:[18]

$$\Delta G_d = \frac{2\pi^2 CK}{4\pi^2 C + Kn_o}\left(\Delta Lk_d + \frac{n_o}{A}\right)^2 + \sum_{i=1}^{n_o} b_i \tag{9}$$

where $K = K(N_d)$.

The first term in this expression gives the effective free energy associated with residual superhelicity. It differs from the ΔG_r that applies when only cruciform extrusion occurs because in the presence of denaturation the residual linking difference is partitioned between interstrand twisting of the unpaired strands in the denatured region and superhelical deformations of the balance of the domain, each of which requires free energy. This term represents the free energy of the most favorable partitioning between these two choices.[18] Here C is the effective torsional stiffness associated with the interstrand twisting of two denatured DNA strands, whose value we determined previously.[12,13] In the second term of equation (9), the summation is over the particular run of denatured base-pairs being considered, and the values of the denaturation energy b_i are those appropriate to the base sequence of that run.[19,20] Finally, the total free energy of this state is:

$$\Delta G(41, n_o, m) = \Delta G_t(41) + \Delta G_d \tag{10}$$

Statistical mechanical analysis

We perform a statistical mechanical analysis of the cruciform extrusion transition in the pVCB plasmid, on which a linking difference ΔLk has been imposed. This technique calculates the equilibrium distribution of a population of identical molecules among all the possible states, from which equilibrium values of any parameter of interest may be calculated. The partition function for this system is:

$$Z = \sum_{n=0}^{110} e^{-(\Delta G(n)/RT)} + \sum_{n_0=1}^{n_{\max}} \left(\sum_{m=0}^{n_o} e^{-(\Delta G(41,n_o,m)/RT)} \right) \tag{11}$$

The first summation includes all the states in which there is no denaturation, and the second summation includes all states in which denaturation occurs. From this expression we calculate the expected number $\bar{n}$ of base-pairs in the cruciform arm to be:

$$\bar{n} = \frac{\sum_{n=0}^{110} n e^{-(\Delta G(n)/RT)} + \sum_{n_o=1}^{n_{\max}} \left(\sum_{m=0}^{n_o} 41 e^{-(\Delta G(41,n_o,m)/RT)} \right)}{Z} \tag{12}$$

The expected number $\bar{n}_o$ of denatured base-pairs at the position where region II occurs at the base of the cruciform is given by:

$$\bar{n}_o = \frac{\sum_{n_o=1}^{n_{\max}} \left(\sum_{m=0}^{n_o} n_o e^{-(\Delta G(41,n_o,m)/RT)} \right)}{Z} \tag{13}$$

In practice we need not consider states whose populations are negligible. Accordingly, in the calculations reported below we use $n_{\max} = 70$ because the probability of longer denatured runs is insignificant under our conditions.

Finally, we calculate the ensemble average residual linking difference $\overline{\Delta Lk_r}$. This is the amount of superhelicity that remains to retard the progress of the plasmid through the gel, and corresponds to the quantity that is directly measured in our experiments. In states where there is no denaturation the expression for ΔLk_r has been presented in equation (3) above. In states where there is denaturation of n_o base-pairs, the residual linking difference is given by:

$$\Delta Lk_r(n_o) = \frac{4\pi^2 C}{4\pi^2 C + K(N_d) n_o} \left(\Delta Lk_d + \frac{n_o}{A} \right) \tag{14}$$

Then the ensemble average residual linking difference is:

$$\overline{\Delta Lk_r} = \left[\sum_{n=0}^{110} \left(\Delta Lk + \frac{n_d(n)}{A} \right) e^{-(\Delta G(n)/RT)} + \sum_{n_o=1}^{n_{\max}} \left(\sum_{m=0}^{n_o} \Delta Lk_r(n_o) e^{-(\Delta G(41,n_o,m)/RT)} \right) \right] / Z \tag{15}$$

Fitting procedure

We consider a set of topoisomers differing by unity in linking number at a fixed temperature T. Parameters that have been previously evaluated are assigned their known values. This leaves four parameters to be determined: the initiation free energy ΔG_x, the bulge free energy ΔG_b, the region I free energy ΔG_I, and the region II free energy ΔG_{II}. Values for each of these parameters are chosen to lie within reasonable ranges, the boundaries of which may be adjusted as necessary. Then the ensemble average residual linking difference $\overline{\Delta Lk_r}$ is calculated for each topoisomer (i.e. value of ΔLk) using the statistical mechanical procedure described above. These are compared with the experimentally determined values $\Delta Lk_{r,\exp}$. For this purpose we only include topoisomers in the range starting from just before the initiation of transition to the most extreme linking difference examined. The RMS deviation between the calculated and the experimental residual linking differences is evaluated:

$$\delta(\Delta G_x, \Delta G_b, \Delta G_I, \Delta G_{II}) = \sqrt{\frac{\sum_{\Delta Lk} (\overline{\Delta Lk_r} - \Delta Lk_{r,\exp})^2}{M}} \tag{16}$$

where M is the number of topoisomers being considered.

This procedure is repeated for a variety of uniformly spaced values of the unknown parameters, and those values that give the best fit (i.e. smallest RMS deviation) are found. If this optimum occurs on the boundary of the parameter space being considered, the boundaries are extended and the procedure is repeated until the optimum occurs in the interior. Then finer grids can be used as desired, centered on the original optimum, to refine its values. In this way the best fitting values of each of the four unknown energy parameters are determined. In the analyses reported here a unique global optimum was found in every case. This procedure is applied at each temperature where experiments have been performed. In this way the temperature dependences of the free energies ΔG_x, ΔG_b, ΔG_I and ΔG_{II} are determined from the experiments.

Results

Cruciform extrusion as a function of temperature and linking number

We fractionated mixtures of pVCB DNA topoisomers by gel electrophoresis in TPE buffer at various temperatures between 35 and 60 °C. These gels were then analyzed in the second dimension in TAE buffer containing chloroquine phosphate. The results are displayed in Figure 4(a). Similar effects, over a more limited temperature range, were observed in earlier experiments.[21] The diagonal ladder of bands towards the left in each gel represents topoisomers not experiencing extrusion. We use this ladder to determine values of the residual linking difference, ΔLk_r, for topoisomers

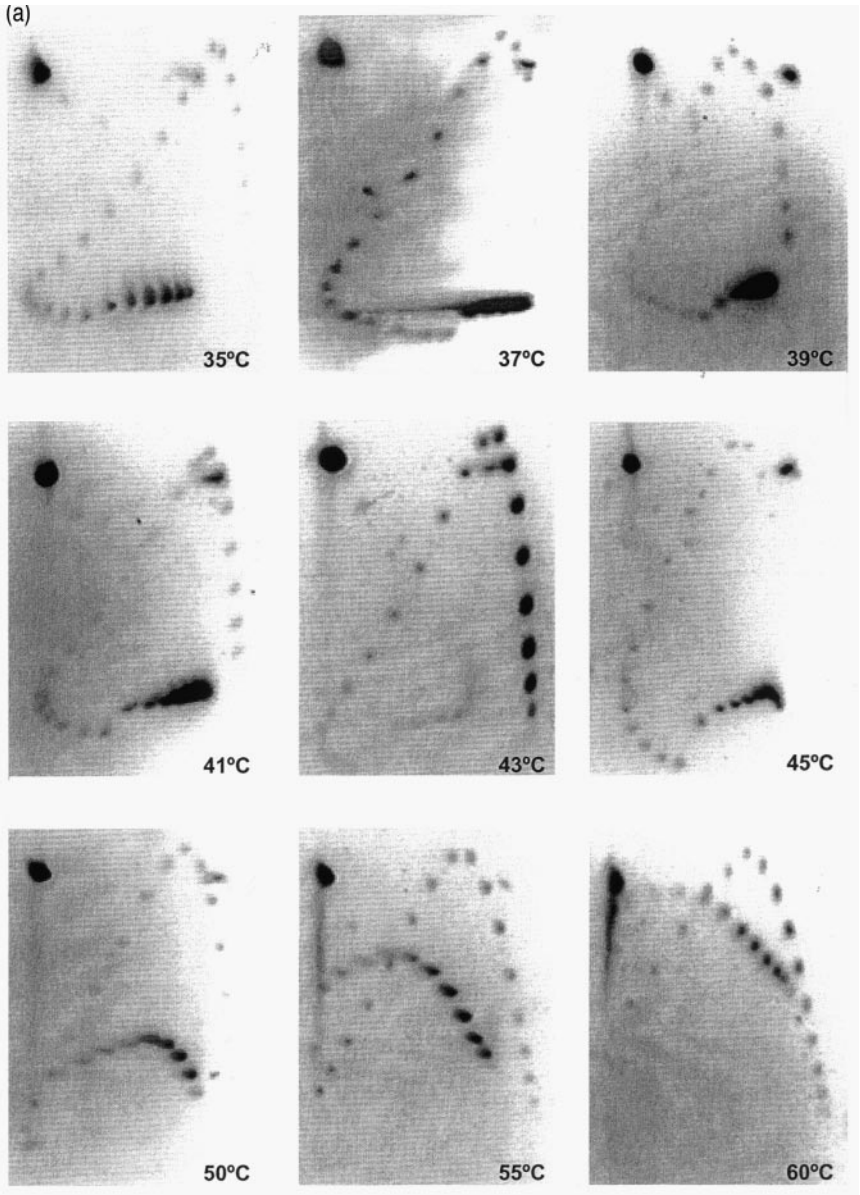

Figure 4. (*Legend opposite*)

with partly or completely extruded cruciforms (see Experimental Methods). The extent of extrusion, n, corresponding to any given value of the residual linking difference is determined using equations (1)-(3). The diagrams in Figure 4(b) show how the locations of the topoisomers at each temperature change with changes in ΔLk.

The intense band at the upper left corner of each gel represents nicked circular DNA. In every case, the nicked circular band migrates slightly faster than the closed, fully relaxed topoisomer. Free rotation about the nicked site allows the nicked species to sample more compact configurations than are easily accessible to the closed fully relaxed topoisomer, thereby increasing its average gel mobility. Because pVCB is relatively small (2922 bp) and the inverted repeat is long (242 bp), complete extrusion is more than enough to relax any level of negative supercoiling within the experimental range. This is shown in the gels by the relatively intense band present at the upper right-hand side at all temperatures except 60 °C. This band co-migrates with the nicked, not the closed relaxed species. This is reasonable, since the Holliday junction should provide a locus of rotational flexibility. We note that the palindrome in this fully relaxed topoisomer is extensively but not completely extruded. This can be seen in the diagrams in Figure 4(b). As pointed out above, the fact that this species co-migrates with the nicked circular plasmid demonstrates that the nearly fully extruded cruciform arms have no significant influence upon the electrophoretic mobility. This strongly supports our use of the ladder of unextruded topoisomers as a reliable measure of ΔLk_r for the partly relaxed topoisomers, each of which

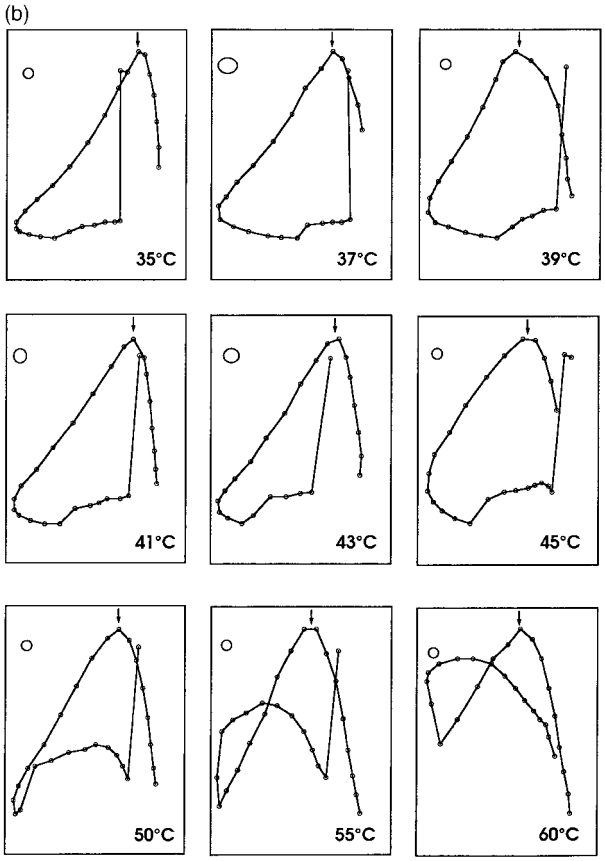

Figure 4. (a) Gel electrophoresis of topoisomer distributions at various temperatures. The first dimension electrophoresis was conducted in 27 cm glass tubes containing 1.5% agarose in TPE buffer, thermostated to the indicated temperature in a Buchler apparatus. Subsequent to the second dimension of electrophoresis, in TAE buffer plus chloroquine phosphate, the resulting slab gels were stained with ethidium bromide and captured with a UVP imaging camera. The isolated band in the upper left-hand corner of each photograph is nicked circular DNA. The direction of electrophoresis is from top to bottom in the first dimension and from left to right in the second dimension. (b) Schematic diagram of all the transitions shown in (a). The location of $\Delta Lk = 0$ is shown by the arrow at the apex of each Figure. The linking numbers of topoisomers to the left of the arrow increase sequentially in the negative sense, and those to the right of the arrow increase sequentially in the positive sense.

has a shorter cruciform than does the fully relaxed species. Since our experiments are conducted at low salt in the absence of a divalent cation, the arms of the four-way junction are expected to be extended and unassociated,[22] hence our results are free of possible complications arising from interactions between the two arms of the cruciform.

The experimental variation of ΔLk_r with ΔLk was used to constrain the statistical mechanical calculations in determining the energy parameters, as described in Theoretical Analysis. We first set up a four-dimensional grid of values of ΔG_x, ΔG_b, ΔG_I and ΔG_{II}, the four parameters to be determined. At each temperature, we determined a unique set of parameter values that gave the best RMS fit, evaluated as described in equation (16). In order to show that these are global and not merely local minima, we also compared the fit value at each minimum with the values at each of its 80 nearest neighbors in this four-dimensional cubic lattice. This procedure was performed at every internal (i.e., non-bounding) lattice point. In every case, no local minimum distinct from the global minimum was detected.

The RMS minimized fitted plots of ΔLk_r *versus* ΔLk at each temperature, including both cruciform extrusion and competing denaturation, are presented in Figure 5 along with the experimental results. Here the open symbols represent the experimental determinations, and the continuous lines indicate the calculated curves of best fit. These results show that at each temperature extrusion occurs progressively over a range of topoisomers. The results of the statistical mechanical calculations agree extraordinarily well with the experimental points, the largest deviations occurring at the highest temperature, 60 °C. This latter temperature is also characterized by failure of the cruciform to extrude completely, due to successful competition from denaturation over the experimental range of superhelicities.

The curves in Figure 5 divide into three regions: (1) an initial rapid rise, indicating the beginning of cruciform extrusion; (2) a central region of greatly reduced slope, in which the imperfections are successively incorporated into the cruciform arms; and (3) a final rapid rise once all imperfections have been incorporated and there is no further impediment to the completion of extrusion. At temperatures below 45 °C, the central region begins with the extruded G/C between locations 21/22 in the cruciform arms and ends with the formation of region II. At temperatures of 45 °C and higher, however, the central region increases in extent and changes in character. In particular, the cruciform persists in a partially extruded state over a greater range of topoisomers, the extent of which increases with temperature. The topoisomers just prior to full extrusion exhibit an increase in $|\Delta Lk_r|$ with increasing $|\Delta Lk|$. This behavior occurs when region II is about to be incorporated into the cruciform arms. Were extrusion simply to stop there, the

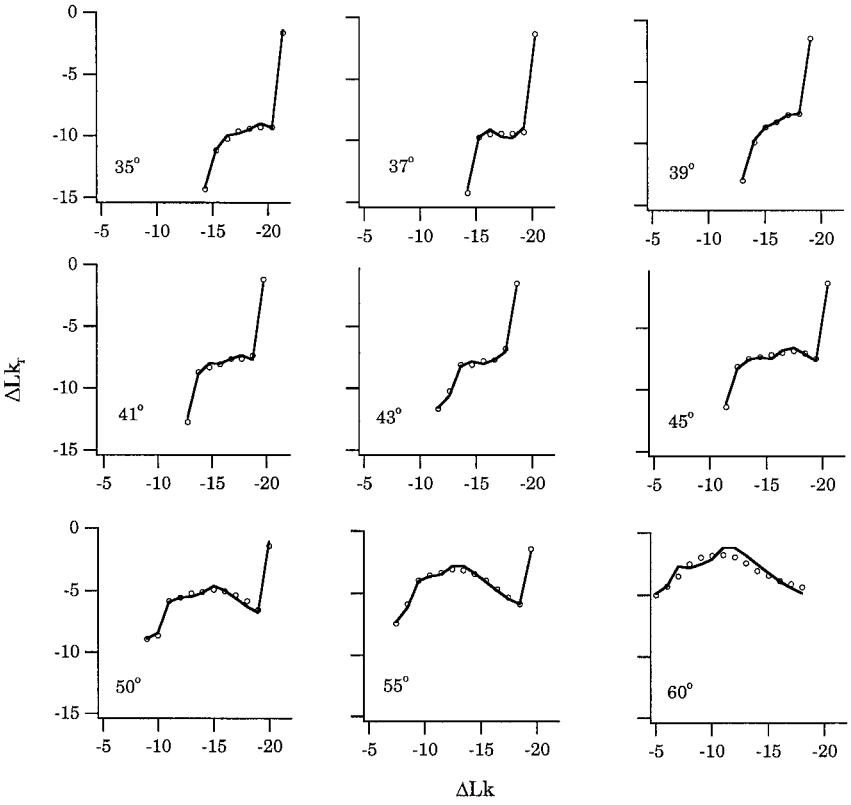

Figure 5. Variation of ΔLk_r with ΔLk as a function of temperature in the first dimension of gel electrophoresis. The symbols in each case represent experimental points determined as described in Experimental Methods, and the continuous lines are the theoretical curves of best fit at each temperature.

residual superhelicity $|\Delta Lk_r|$ would increase by one turn per topoisomer. However, the fact that it increases more slowly (approximately 0.6 turn per topoisomer) shows that some other transition occurs that absorbs about 40% of the ΔLk increment. We ascribe this to local denaturation at the base of the partially extruded cruciform.

The course of the cruciform extrusion is more directly shown by the dependence of $n_d/2$, the number of base-pairs per cruciform arm removed from the original duplex, upon ΔLk. This is presented in Figure 6 for both the experimental (data points) and calculated (continuous line) results, as calculated from equation (3). The curves in (a) are for temperatures of 35 to 43 °C, and those in (b) are for temperatures of 45 to 60 °C. The temperature increases from right to left, and $-\Delta Lk$ increases from left to right. The slopes of all curves in Figure 6(a) initially increase relatively rapidly, passing through the formation of the three A/T bulges at locations 5, 11 and 13 (cf. Figure 1), until the G/C bulge at location 21 is encountered. At this point the slope becomes significantly reduced, reflecting the incremental difficulty of incorporating region I and the two A/T bulges at locations 32 and 35. In the absence of significant denaturation, which is the case at the temperatures plotted in Figure 6(a), the slope increases abruptly following incorporation of region II. The final slope is somewhat greater than the initial slope, showing the effects of the first three A/T bulges in raising the energy required for the beginning of the cruciform extrusion. The curves in Figure 6(b) are similar, except that the influence of denaturation in the vicinity of the base of the Holliday junction can be seen clearly. Here the slope subsequent to encountering region II is greatly reduced, showing the effects of competing denaturation. Finally, at all temperatures except 60 °C, the extrusion is completed at higher values of $|\Delta Lk|$.

Determination of the temperature dependence of the thermodynamic functions for the formation of structures containing unpaired bases

The complete imperfect cruciform is formed from a fully base-paired but imperfect inverted repeat in the original duplex. Each completely extruded arm contains six single base bulges and two more complex bulged structures (regions I and II). By symmetry, the unpaired structures are present in both arms as complements. The total free energy change, ΔG_{tot} associated with formation of the completely extruded imperfect cruciform (in the absence of supercoiling) is then:

$$\Delta G_{tot} = \Delta G_i + 6\Delta G_b + 5b_{AT} + b_{GC} + \Delta G_I + \Delta G_{II} \quad (17)$$

Using the procedure described in the previous section, we have evaluated the free energies associated with each component of this extrusion process. Specifically, we determine ΔG_x, ΔG_b, ΔG_I, and ΔG_{II} by our fitting procedure and obtain ΔG_i from equation (6). In our procedure there are four

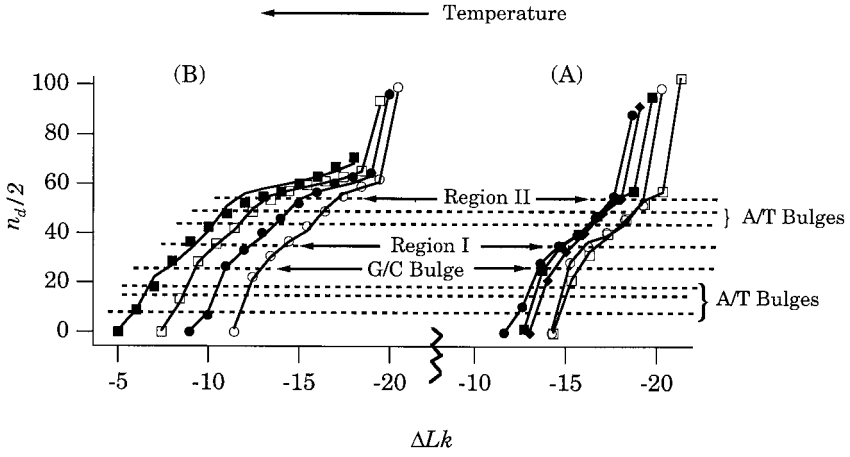

Figure 6. Plot of the number of duplex base-pairs transferred to each cruciform arm as a function of ΔLk. The quantity $n_d/2$ is defined by equation (3). The five curves in (a) are the results for 35 °C (□), 37 °C (○), 39 °C (◆), 41 °C (■) and 43 °C (●). These show little if any evidence of denaturation in the vicinity of the Holliday junction. The four curves in (b) are the results for 45 °C (○), 50 °C (●), 55 °C (□) and 60 °C (■). These show denaturation at the Holliday junction, to an extent increasing with increasing temperature. The three grouped A/T bulges are those following locations 5, 11 and 13, as numbered in Figure 1. The G/C bulge is that following location 21. The two grouped A/T bulges are those following locations 32 and 35. Note that the numbering convention used in Figure 1 counts only intact base-pairs outside of regions I and II, hence is generally smaller than $n_d/2$ at any given value of ΔLk.

unknown quantities to be fitted at each temperature. Since the bulged regions are well separated along the cruciform arms, interaction between them is minimal. For comparison, the determination of the free energy of a single bulged base using the thermal denaturation of oligonucleotides involves six adjustable parameters,[23] and this method is not readily applicable to unpaired structures containing more than a single bulged or mismatched base.

Figure 7 shows the temperature dependence of the components of ΔG_{tot}. The data points in each curve were obtained by applying the statistical mechanical analysis to the experimental gel shifts at each temperature, as described above. All free energies are positive over the entire range of temperatures, in keeping with the expectation that the formation of imperfect structures is relatively disfavored. In each case the variation with temperature is linear within the precision of the experiment. The slopes of the plots in Figure 7 are determined by the entropies, and the intercepts at $T = 0$ K by the enthalpies. The values of ΔH and ΔS for the formation of all structures in both cruciform arms are listed in Table 1. These may be combined to calculate the corresponding free energies at any temperature. Since the enthalpies and entropies are correlated, however, the standard errors of the calculated free energies should be read from the 95% confidence intervals shown in Figure 7 instead of combined from the errors listed in Table 1.

The experimental results were re-analyzed (data not shown) using the integral form of the van't Hoff equation to check for possible temperature dependence of the enthalpies[24]. The heat capacities were very small in all cases except that of region II, where denaturation competes significantly with bulge formation. This region is treated in greater detail in the Discussion.

Effect of the imperfections upon the extrusion of the cruciform

We have shown that the formation of structures containing unpaired bases has a major effect on the cruciform extrusion transition in a superhelical DNA plasmid. The substantial free energy cost of forming unpaired structures when imperfections in the inverted repeat symmetry are encountered has a major impact upon the energetics of extrusion. This is shown in Figure 8, which compares ΔG_{tot} and ΔG_i over the temperature range of the experiments. The bulged bases greatly destabilize the cruciform, as demonstrated by the significantly greater ΔG values at all temperatures. In addition, the entropy contribution of the bulges reduces the difference in free energies as the temperature is raised. Thus, $\Delta G_{tot} - \Delta G_i \approx 70$ kcal/mol at 35 °C, but $\Delta G_{tot} - \Delta G_i \approx 25$ kcal/mol at 60 °C.

Table 1 also presents the thermodynamic functions ΔH and ΔS for the complete extrusion of both arms of the cruciform. This is done in two ways. First we show these state parameters for extrusion of the actual imperfect inverted repeat, including the effects of all unpaired structures formed in the cruciform arms. And second we give the values for a hypothetical perfect cruciform, for

Table 1. Thermodynamic functions for formation of extruded nucleotide structures and of cruciforms

Structure (type)	ΔH (kcal/mol)	ΔS (kcal/mol deg.)
Bulge pair (b)	45.76 ± 4.41	0.132 ± 0.014
Region 1 pair (I)	91.80 ± 4.37	0.275 ± 0.014
Region 2 pair (II)	233.88 ± 43.7	0.667 ± 0.137
Cruciform extrusion, including all imperfections (t)	814.45 ± 49.2	2.363 ± 0.155
Pure cruciform, excluding all imperfections (i)	192.5 ± 10.5	0.565 ± 0.033

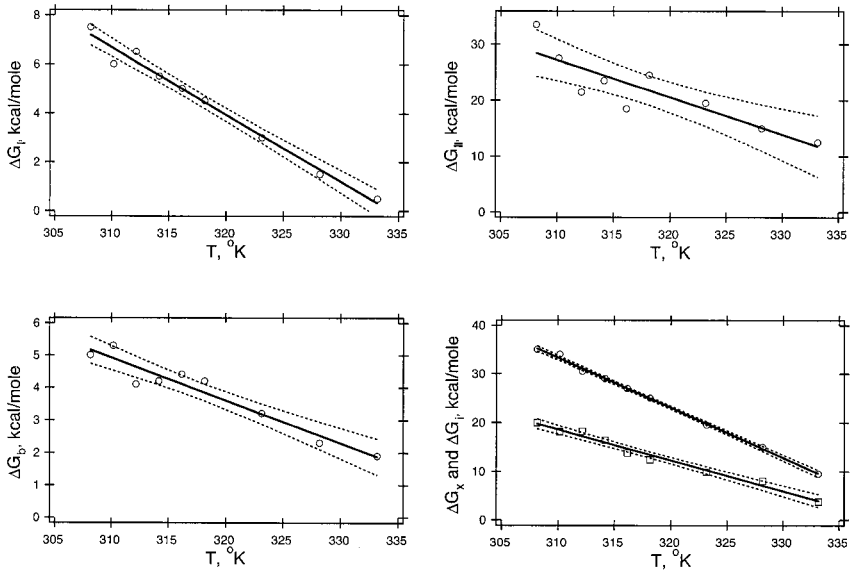

Figure 7. Plots of the free energy of formation, ΔG, for the four different characteristic structures in the imperfect cruciform arms. All free energies depicted here refer to the formation of both cruciform arms, hence are the sum of the formation of the two complementary unpaired structures at corresponding locations. ΔG_I and ΔG_{II} are the free energies of formation of the complementary pairs of regions I and II, respectively. ΔG_b is the free energy of formation of a single base bulge complementary pair. ΔG_i is the free energy of formation of the pair of cruciform arms (including the initiation of denaturation), the unpaired bases at the ends of the arms, and any free energy associated with the junction. ΔG_x is the sum of ΔG_i and the free energies of forming the first three bulges (equation (6)).The bracketing broken lines show the 95% confidence limits.

which the only free energy cost is ΔG_i. As listed in Table 1, both ΔH and ΔS are approximately four times as large for the imperfect cruciform as for the underlying perfect cruciform.

Discussion

The principal objective of this research is to determine the energetics of extrusion of an imperfect palindrome into a cruciform containing bulged bases of various configurations and locations. In the course of this analysis, we have developed a new procedure for evaluating the thermodynamic state functions associated with the formation of a variety of special DNA structures. Our approach is to insert an imperfect inverted repeat into a plasmid, then observe the extent of its extrusion with changes in linking number, as represented by the distribution of topoisomers in two-dimensional gel electrophoresis. Although this inverted repeat is a perfect duplex in the plasmid, it can be induced to extrude to a cruciform containing various local unpaired bases by the appropriate changes in ΔLk and temperature. The theoretical method used to analyze these transitions is a significant extension of the technique previously used to analyze strand separation.[12,13] This method includes the competition between denaturation and cruciform extrusion, and it presents the first statistical mechanical analysis of cruciform extrusion in which all states of extrusion are explicitly examined.

Course of the cruciform extrusion with ΔLk

Our results show that cruciform extrusion can be hindered at sites where imperfections in the inverted repeat symmetry of the participating DNA sequence are encountered. The increments of energy needed to form unpaired structures in the cruciform arms are provided by changes in $|\Delta Lk|$, and hence in the energy of relaxation. This is already evident in the plots in Figures 5 and 6 at temperatures 35-43 °C.

The smallest imperfection, an inserted base-pair that causes a single base bulge in each arm, can delay extrusion. Whether or not extrusion pauses at a specific imperfection depends on two factors: the free energy needed to incorporate the current imperfection into the cruciform arms, and the amount of residual superhelicity remaining when that imperfection is encountered. (Here we regard the initiation of extrusion, in which the loops are incorporated, as an imperfection, the first that is encountered.)

The extrusion pauses at an imperfection when the remaining ΔLk_r provides insufficient energy to drive its incorporation into the cruciform arms. Since the free energy of residual superhelicity is quadratic in ΔLk_r, the energy of relaxation of successive topoisomers increases rapidly until a topoisomer is reached where the blockage is overcome and extrusion proceeds. Extrusion continues beyond this point until the molecule is fully relaxed, the inverted repeat is fully extruded, or

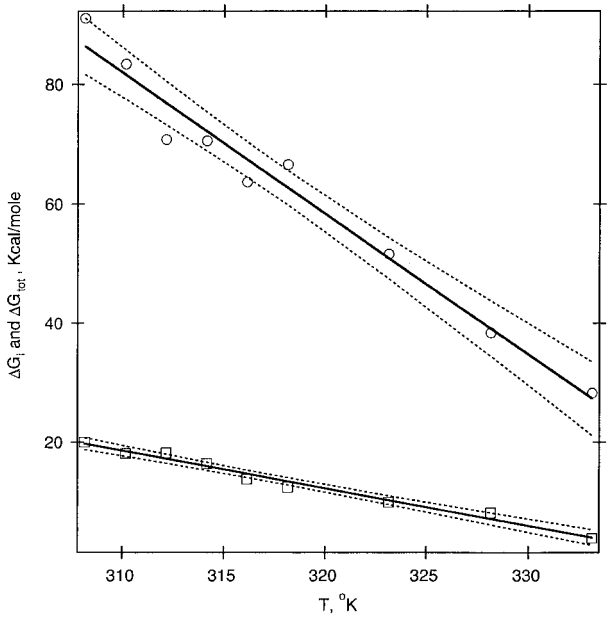

Figure 8. Comparison of the free energies of formation of imperfect (upper curve) and the underlying perfect (lower curve) cruciforms in pVCB DNA. The bracketing broken lines show the 95% confidence limits.

the next imperfection is encountered. In the latter case, whether or not extrusion stops at that imperfection again depends on the relative costs of its incorporation *versus* relaxation for the given topoisomer.

In pVCB, the three single base A/T bulges proximal to the cruciform loop that are encountered early in the extrusion process are ineffective in delaying extrusion within the precision of these experiments. To understand why, we note that, because significant free energy is needed to initiate cruciform extrusion, this process only becomes favored when the DNA is substantially negatively supercoiled. The topoisomers in which extrusion occurs will still have significant residual superhelicity remaining as each of the first three bulges is encountered. Because the free energy of their incorporation is less than the relaxation energy provided by the resulting change in ΔLk_r, they fail to obstruct further extrusion. When the fourth bulge (G/C) is encountered, however, the residual superhelicity has already been significantly reduced by the extrusion process. Consequently, a larger change in ΔLk_r is required to stabilize incorporation of this fourth bulge into the cruciform, a phenomenon that is reflected in local broadening of plots of ΔLk_r *versus* ΔLk, as seen in Figures 5 and 6. A similar mechanism accounts for the substantial broadening of the plots of ΔLk_r *versus* ΔLk during formation of region I.

Six interstrand duplex base-pairs must denature before region II can be incorporated into the cruciform, and this process occurs at a substantially reduced ΔLk_r At this point and beyond, there are two alternative possibilities: either this denatured region can extend, or the cruciform arms can lengthen to incorporate the region II imperfection. When the latter occurs there is no further impediment to full extrusion, so the topoisomers become fully relaxed and comigrate with the nicked species. The competition between these alternatives changes with temperature. At higher temperatures denaturation becomes more favored to persist as $|\Delta Lk|$ increases, primarily because the energy returned by forming intrastrand base-pairs, and hence the competitiveness of continued extrusion, decreases with temperature. At all temperatures except 60 °C a threshold superhelicity is observed where region II becomes incorporated into the cruciform and extrusion proceeds to completion without further competition from denaturation.

Thermodynamics of cruciform formation

Table 2 contains the free energies of cruciform extrusion at 37 °C, both for the hypothetical perfect cruciform and for the actual cruciform containing all bulged bases. The free energy of the perfect cruciform is approximately 7.2 times that of a single base bulge, suggesting that most of the free energy of formation of the cruciform derives from the eight unpaired bases at the two hairpin loops. The presence of imperfections has a drastic effect, causing a more than fourfold increase in free energy of extrusion.

An alternative way to assess the magnitude of the effects of imperfections upon cruciform extrusion is to estimate the added linking number change required to completely extrude the imperfect *versus* the perfect cruciform. As listed in Table 2, the differential free energy at 37 °C is $\Delta\Delta G = 64.3$ kcal/mol. The free energy of superhelix formation is proportional to $(\Delta Lk)^2$, the constant of proportionality at 37 °C being $K = 0.253$ for the unextruded experimental plasmid.[12] If ΔLk_o is the linking difference at which both the underlying perfect and the actual imperfect cruciform begin to extrude, and ΔLk_f is the linking difference at which the imperfect cruciform is fully extruded, then:

$$(\Delta Lk_f)^2 - (\Delta Lk_o)^2 = \frac{\Delta\Delta G}{K} = 254.2 \qquad (18)$$

This simplified equation assumes that ΔLk_r is approximately constant throughout extrusion, which is reasonably accurate at 37 °C (see Figure 5).

Taking $\Delta Lk_o = -14$ from our experimental results at 37 °C, this equation predicts that $\Delta Lk_f = -21$. Reference to the 37 °C panels of Figures 4-6 confirms this prediction. A considerable increment in ΔLk is indeed required to extrude a cruciform containing the array of bulged bases found in the sequence used in the present study. A similar calculation comparing a perfect cruciform with one containing a single bulged pair (resulting from an initial inverted repeat in which one arm contained either a single insertion or del-

Table 2. Free energy at 37 °C and T_e for formation of extruded nucleotide structures and of cruciforms

Structure	ΔG(37 °C) (kcal/mol)	$\Delta G/\Delta G_b$ (37 °C)	T_e (°C)
Single bulge, average	2.42 ± 0.18	1.0	74
Single region 1, average	3.25 ± 0.18	1.34 ± 0.12	61
Single region 2, average	13.6 ± 1.8	5.58 ± 0.84	78
Cruciform, including all imperfections	81.6 ± 4.0	33.7 ± 2.9	71
Cruciform, excluding imperfections	17.3 ± 1.3	7.15 ± 0.56	65

etion) shows that even such a minimal imperfection can increase the final ΔLk by nearly one. This means that even a slight imperfection in a naturally occurring inverted repeat can significantly impede full extrusion under the influence of torsional stress.

The free energy of extrusion of a perfect palindrome into a cruciform was evaluated previously by several groups, using the gel retardation technique under different conditions and applied to different molecules. These results were based upon an application of an equation similar to our equation (18). Gellert's group[25] found a value of 18 kcal/mol at 50 °C. Lilley[26] found a value of 18.4 kcal/mol under unspecified conditions. Courey & Wang[27] estimated that the free energy of cruciform extrusion is about 17 kcal/mol at 65 °C. They argued that this value is about the same at 37 °C. Singleton & Wells[28] found a value of 22 kcal/mol at room temperature. We note that all these results were obtained by equating the differences in superhelical free energy immediately before and after extrusion with the free energy of extrusion. This procedure is reasonable only in the limiting case of the extrusion of a perfect palindrome. A rigorous analysis requires inclusion of all possible states of extrusion, as presented here.

Free energies of formation of single base bulges and of more complex bulged regions

In the course of analyzing cruciform extrusion, we determined the thermodynamics of formation of single bulged bases and of two more complex bulged structures. Table 2 lists the free energies of formation of a single base bulge, and of regions I and II, as calculated from the linear fits of Figure 6 at 37 °C. The free energies relative to that of a single bulge at 37 °C are also calculated, as are the temperature, T_e, at which each $\Delta G = 0$; i. e, at which the structure in question ceases to be destabilizing. The relative values of T_e may be used to compare the destabilizing effects of each structure: region II is the most destabilizing, followed by a single bulge and by region I. All structures remain destabilizing even at 60 °C, the highest temperature attained in these experiments. However, denaturation becomes a significant competitor to cruciform extrusion at 60 °C. Table 2 also lists the ratios of each free energy at 37 °C relative to that of a single base bulge. As is the case with ΔH and ΔS (see next section below), the values of ΔG are approximately proportional to the number of bases involved in forming the structure. The exception to this is region I, in which the ratio is less than the expected 2.0. This is possibly due to stacking of the extrahelical G into the duplex. NMR experiments performed upon oligonucleotides containing region I at $T = 5$ °C (Ulyanov, James and W.R.B., unpublished results) indicate that the G is stacked while the C is truly extrahelical at this temperature, which is much lower than those of our experiments. Other NMR experiments involving a different oligonucleotide[29] found that cytidine bulges change from extrahelical to stacked as the temperature is raised to the vicinity of 40 °C.

The single base bulge free energy found here is 2.45(±0.18) kcal/mol at 37 °C, which is comparable to values reported by others. Woodson and Crothers[30] obtained a range for ΔG of single bulge formation of 2.9-3.2 kcal/mol at 41 °C, using optical melting experiments in 80 mM $NaClO_4$, 20 mM Tris-HCl. Our solvent is approximately 45 mM salt. In order to compare these results with ours, we assume that the dependence of T_e upon salt concentration, C_s, is similar to that of T_m found for denaturation:[20] $T_{m2} - T_{m1} = -16.6\log[C_{s2}/C_{s1}]$. The T_e for DNA under our conditions is then estimated to be about $16.6\log(100/45) = 5.8$ °C lower than under the conditions of Woodson and Crothers. Assuming that an increase in salt concentration has the same effect as the corresponding decrease in temperature, we estimate from the appropriate curve in Figure 6 that our free energy of bulge formation under their conditions would be 2.7 kcal/mol. In experiments performed in 1.0 M salt, LeBlanc and Morden[23] obtained a somewhat higher range of 3.5-4.6 kcal/mol at 37 °C. Again correcting for the difference in salt concentration as above, we estimate from our results that the free energy of bulge formation under their conditions would be 3.9 kcal/mol, in excellent agreement. Ke and Wartell[31] used temperature gradient gel electrophoresis in 70 mM salt to obtain a broader range of 2.5-4.6 kcal/mol at 37 °C. It is clear from the above comparisons that our determinations are in line with those made previously using small oligonucleotides. Our measurements are uninfluenced by the possibly substantial end effects that occur during experiments with oligonucleotides. This might explain why our values are generally somewhat smaller than those reported from experiments on oligonucleotides.

Entropy and enthalpy of the formation of unpaired structures

In single base bulge formation, cruciform extrusion causes a base-pair in the original duplex to become a pair of bulged bases, one in each arm. The average single base bulge free energy is therefore half the value ΔG_b reported here, and the corresponding enthalpy and entropy are likewise half the values given in Table 1. The average free energies, enthalpies and entropies for formation of a single hairpin and a single region I or region II structure also are half the values listed there. The average enthalpy and entropy changes for single bulges, regions I and II, and a single hairpin are listed in Table 3.

We estimate that the average value of ΔH for the formation of a single bulged base is 23 kcal/mol, and that the average value of ΔS is 66 cal/mol deg. At the present level of our analysis, these numbers are independent of base composition. The results in Table 3 show, in addition, that both ΔH and ΔS are approximately proportional to the number of unpaired bases in each structure. Region I nominally involves an asymmetric loop with four nucleotides in one strand and two in the opposing strand. NMR structural studies of this region in TAE buffer (Ulyanov, James and W.R.B., unpublished results), establish that the central A·T base-pairs are indeed formed. Thus the ΔH and ΔS found for this structure, in which only the G and C remain unpaired, are also consistent with this proportionality. No structural information is yet available for region II. If no interstrand base-pairs are formed in the looped out structure, the number of unpaired bases would be six. It is possible, however, for either of two mutually exclusive A·T interstrand base-pairs to form, although perhaps transiently. If this happens, the effective number of unpaired bases in region II would be between four and six. The ΔH and ΔS results that we obtain are consistent with such a structure. Finally, the underlying hairpin contains four unpaired bases in the terminal loop. The ΔS and ΔH ratios that we determine are also proportional to the nominal extent of unpairing.

The ΔH of formation of each unpaired structure is positive, due to the presence of fewer favorable molecular interactions in the vicinity of the defect relative to an intact duplex. The ΔS of formation of each unpaired structure is also positive, reflecting the greater number of degrees of freedom that are associated with a partly disrupted duplex compared to the corresponding intact duplex. The entropy term therefore favors defect formation, while the enthalpy term opposes it. As the temperature increases, the entropy contribution to the free energy becomes relatively more important, until a temperature, T_e, is reached at which the formation of each loop, bulge or region becomes isoenergetic (see Table 2).

Determinations by others of the ΔH and ΔS for single base bulge formation compared the T_m of a bulge-containing oligomer to that of an oligomer lacking the bulged base but otherwise having the same sequence. Woodson and Crothers[30] used a nonamer as control, inserting either an extra A or an extra C at a location between the third and fourth bases from the 3′ terminus of one strand. As pointed out by the authors, this location is close enough to the terminus to promote significant fraying, thereby possibly skewing the results. Nevertheless, analysis of their results yields a ΔH of 11.8 kcal/mol for the C bulge and 16.4 kcal/mol for the A bulge. The values of ΔS were 31 cal/mol deg. for the C bulge and 43 cal/mol deg. for the A bulge. The solvent included 20 mM Tris HCl and 80 mM $NaClO_4$ (pH 7). In a more recent study, LeBlanc and Morden[23] used two different methods to compare the melting of a series of decamers containing an extra bulged base at the center of either strand. Averaging their results for A·T and G·C pairs, the calculated average ΔH depended somewhat on the method, being 17.2 kcal/mol for the NLS method and 21 kcal/mol for the T_m method. The range of ΔH was 12-26 kcal/mol, with most of the values lying between 17 and 26 kcal/mol. Our result of 23 kcal/mol lies near the middle of this range. The reported values for ΔS in the same study ranged from 27 to 70 cal/mol deg., with the average being 46 cal/mol deg. for their NLS method and 55 cal/mol deg. for their T_m method. These experiments were conducted in 1 M NaCl, while ours were in electrophoresis buffer (approximately 45 mM salt). Our estimates for the ΔH and ΔS of single base bulge formation, 23 kcal/mol and 66 cal/mol deg., are similar to these T_m-based estimates.

Most of the reported studies of the thermodynamic functions for bulge base formation have examined oligoribonucleotides, since these relate to the folding of RNA.[32] We hesitate to compare our results with these in a rigorous way, since the interruption of an *A* duplex (in the RNA case) might well be very different from the interruption

Table 3. Thermodynamic functions for formation of single arm defects and of hairpins

Structure	ΔH (kcal/mol)	$\Delta H/\Delta H_b$	ΔS (kcal/mol deg.)	$\Delta S/\Delta S_b$
Bulge, single base	22.9 ± 2.2	1.0	0.066 ± 0.007	1.0
Region 1	45.9 ± 2.2	2.0 ± 0.2	0.137 ± 0.007	2.1 ± 0.3
Region 2	116.9 ± 21.9	5.1 ± 1.1	0.333 ± 0.069	5.0 ± 1.2
Hairpin, excluding all stem defects	96.3 ± 5.2	4.2 ± 0.5	0.283 ± 0.017	4.3 ± 0.6

of a *B* duplex (in the DNA case). Nonetheless, in at least one study[33], the reported values of ΔH were only slightly lower than ours. Using thermal denaturation methods similar to those described immediately above for oligodeoxyribonucleotides, the results for ΔH for those oligonucleotides satisfying the two-state criterion ranged from 13.9 to 20.2 kcal/mol for a single bulged base. This is slightly lower than, but comparable to, the result that we have obtained.

Biological importance of unpaired bases

We determined the energies of formation of particular local DNA structures that form when an imperfect inverted repeat extrudes into a cruciform containing extrahelical bases. Structures of this type are of biological significance in at least two different contexts: they are present in numerous eucaryotic viruses and organelles and are important for DNA replication in these systems; and they appear to serve as intermediates in frameshift mutations,[1,2] especially when present as extrahelical bases in extruded cruciforms.[3]

The presence of terminal hairpin loop structures is characteristic of the DNA of several eucaryotic viruses and organelles. Examples include the viral genomes of iridoviruses,[34] parvoviruses,[35–37] phycodnaviruses[38] and poxviruses,[6,39,40] as well as the rDNA from *Tetrahymena*,[41] the mitochondrial DNA of *Paramecium*[42] and linear mitochondrial DNA of yeast.[43] Many of these terminal sequences contain extrahelical bases, present as bulges, loops or mismatches. Examples, in addition to poxvirus DNA, are the DNAs from the parvoviruses, including minute virus of mice[4,44,45] and the human parvovirus B19;[46] the *Autographa californica* multinucleocapsid nuclear polyhedrosis virus;[38] the *Leporipoxvirus* Shope fibroma virus;[11] and the linear mitochondrial DNA of yeast.[43]

As pointed out in Introduction, a bubble of four mismatched bases is important for DNA replication in the minute virus of mice.[4] In particular, the location and general structure of the bubble, but not its composition, were shown to determine the biological effects. This finding is consistent with our finding that the ΔH and ΔS of bulge formation depend only on the number of bases involved and not on their sequence or arrangement. It has been further speculated that the unpaired bases within this region may be signal sequences for specific nicking of intermediates in DNA replication, including that of the dimer length duplex replicative intermediate.[47] Inverted repeats occurring in viral telomeres appear to be involved in the resolution of linear concatemers into unit length linear molecules.

Imperfect palindromes have also been discovered in protein coding genes of *Rickettsia*. In *R. conorii*, for example, determination of the complete genome sequence has revealed 44 repetitions of a 150 base-pair imperfect palindromic repeat, 19 of which were inserted in open reading frames.[48] A homologous repeat was found in proteins of other *Ricksttsia* species, raising the possibility that imperfect inverted repeats play a role in the creation of new protein species.

Little is known about the possible function of unpaired bases in vaccinia. It has been shown that the terminal loop region containing the bulged bases in vaccinia strain WR is transcribed late in infection.[49] The sequences required for resolution of the concatemer junction in vaccinia DNA replication appear to be bounded at one side by region II in the WR strain.[50] The strain from which the sequences in our plasmid were derived, IHD-W, has terminal sequences identical to WR. The terminal region of the DNA of the Ankara strain of vaccinia, which is highly attenuated, is the same as that of WR from the hairpin terminus until region 2. In this strain a bulge also exists at the region 2 location, albeit of different composition. As with the parvoviruses, it appears that the location of the bulges, rather than their composition, is the more important factor in determining biological activity.

It has been proposed that resolution of the vaccinia replicative intermediate is accomplished by nicking and sealing of an extended cruciform structure,[51] a process that involves exactly the same transition between an inverted repeat and a cruciform as that in the present study. In the linear concatemers of the vaccinia replication intermediates, such a transition could be brought about by the wave of negative supercoiling that follows transcription.[52] A further possible reason for the presence of unpaired bases might be their effect on the tertiary structure of the DNA in this region. It is well established that bulged bases cause the duplex to bend. This suggestion is consistent with the observation that the location of the bulges, rather than their composition, is the factor that appears to have the greatest effect upon biological activity. An alternative possibility is that the bulges serve as templates for the folding of an RNA transcript.

As our results have demonstrated, the presence of imperfections in the inverted repeat at appropriate positions within the sequence can significantly inhibit full extrusion of a cruciform. It follows that the presence of such imperfections could serve the purpose of controlling the extent of extrusion, preventing the formation of states with a deep energy well out of which it may be too difficult to climb. In contrast, if the extrusion can be kept shorter, the well is correspondingly more shallow, and hence easier to escape. Indirect support for this suggestion is provided by the observation that the majority of inverted repeat sequences that are computationally found in representative genomic sequences contain significant numbers of imperfections. Indeed, perfect inverted repeat symmetry appears to be the exception rather than the norm.

Experimental Methods

Plasmid DNA preparation

Purification was by Qiagen Maxi-Prep. DNA was relaxed using calf thymus type I topoisomerase (Gibco BRL) in 50 mM Tris (pH 7.5), 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM EDTA, 30 μg/ml bovine serum albumen in a reaction volume of 60 μl containing 3.2 μg DNA and 12 units of enzyme. Following incubation on ice for 60 minutes, half the volume (30 μl) was removed and the reaction stopped by addition of 6 μl 5% (w/v) SDS, 0.15% (w/v) bromophenol blue, 50% (v/v) glycerol, 10 mM Tris (pH 8.0), 1 mM EDTA. The remaining half was incubated for another 60 minutes, stopped as described, and the two samples recombined. DNA was pre-incubated at the temperature of the first gel, then layered onto the gel tube.

Gel electrophoresis

Two-dimensional gel electrophoresis was carried out as described,[12,53] with some modifications. Electrophoresis in the first dimension was conducted at various temperatures in tube gels containing 1.5% (w/v) agarose (type 1, Sigma Chemical Co.) in TPE buffer (40 mM Tris-phosphate, 4 mM EDTA, pH 7.55 at 30 °C). The 22 cm tube gels were contained in 27 cm × 0.6 cm (i.d.) glass tubes, partially constricted at the bottom. Electrophoresis was performed for 20 hours at 50 V in a Buchler electrophoresis apparatus maintained at constant temperature by a Lauda model B-2 circulating water bath. Second dimension gel electrophoresis was conducted by first soaking the appropriately sliced tube from the first dimension in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.3) containing 4 μM chloroquine, then placing it into the upper slot of a slab gel, also 1.5% agarose in TAE buffer containing 4 μM chloroquine. Second dimension electrophoresis was conducted for 20 hours at 50 V at room temperature with recirculation of the buffer.

Measurement of gels and calculations

Gels were stained in ethidium bromide, and the bands visualized under UV light utilizing the UVP White/UV TMW-20 Transilluminator with attached UVP Model GDS 7500 Darkroom and UVP Imaging camera. Each band represented a topological isomer of a unique linking number, ΔLk. The images were captured to a Macintosh computer into NIH Image and the x,y coordinates of each band determined, with the x-axis representing migration in the second dimension and the y axis representing migration in the first dimension.

Using Igor Pro 4.01 (Wavemetrics, Inc.), the first dimension migrations were plotted *versus* point number, generating a curve having a minimum at the fully relaxed topoisomer, or $\Delta Lk = 0$. The point value of the location of the minimum was determined by a Gaussian curve fit, and the ΔLk for each point determined by subtracting the point value at the minimum from each point value. The data were then replotted as distance migrated *versus* ΔLk.

What appears to be a single supercoiled band, the fastest migrating in the first dimension, is actually resolved into multiple bands through chloroquine separation in the second dimension. The ΔLk_r of these bands, that is, the value by which their value differs from the predicted ΔLk based on the migration of the slower-migrating species, is determined by fitting the curve defining the location of the slower migrating species and then interpolating the analogous ΔLk value. Finally, a plot was created comparing ΔLk *versus* ΔLk_r for each set of data.

Acknowledgments

This research was supported in part by grants from the National Institute of General Medical Sciences to W.R.B. (GM21176) and to C.J.B. (GM 47012) and from the National Science Foundation to C.J.B. (DBI-9904549).

References

1. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugeta, A., Terzaghhi, E. & Inouye, M. (1966). Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77-84.
2. Ripley, L. S. (1982). Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. *Proc. Natl Acad. Sci. USA,* **79**, 4128-4132.
3. de Boer, J. G. & Ripley, L. S. (1984). Demonstration of the production of frameshift and base-substitution mutations by quasipalindromic DNA sequences. *Proc. Natl Acad. Sci. USA,* **81**, 5528-5531.
4. Costello, E., Sahli, R., Hirt, B. & Beard, P. (1995). The mismatched nucleotides in the 5′-terminal hairpin of minute virus of mice are required for efficient viral DNA replication. *J. Virol.* **69**, 7489-7496.
5. Moss, B. (1996). *Poxviridae:* the viruses and their replication. In *Virology* (Fields, B. N. & Knipe, D. M., eds), pp. 2637-2671, Lippincott-Raven Publishers, Philadelphia.
6. Baroudy, B. M., Venkatesan, S. & Moss, B. (1982). Incompletely base-paired flip-flop terminal loops link the two DNA strands of the vaccinia virus genome into one uninterrupted polynucleotide chain. *Cell,* **28**, 315-324.
7. Baroudy, B. M., Venkatesan, S. & Moss, B. (1983). Structure and replication of vaccinia virus telomeres. *Cold Spring Harb. Symp. Quant. Biol.* **47**, 723-729.
8. Wittek, R. & Moss, B. (1980). Tandem repeats within the inverted terminal repetition of vaccinia virus DNA. *Cell,* **21**, 277-284.
9. Moyer, R. W. & Graves, R. L. (1981). The mechanism of cytoplasmic orthopoxvirus DNA replication. *Cell,* **27**, 391-401.
10. DeLange, A. M., Futcher, B., Morgan, R. & McFadden, G. (1984). Cloning of the vaccinia virus telomere in a yeast plasmid vector. *Gene,* **27**, 13-21.
11. Dickie, P., Morgan, A. R. & McFadden, G. (1987). Cruciform extrusion in plasmids bearing the replicative intermediate configuration of a poxvirus telomere. *J. Mol. Biol.* **196**, 541-558.
12. Bauer, W. R. & Benham, C. J. (1993). The free energy, enthalpy and entropy of native and of partially denatured closed circular DNA. *J. Mol. Biol.* **234**, 1184-1196.
13. Bauer, W. R., Ohtsubo, H., Ohtsubo, E. & Benham, C. J. (1995). Energetics of coupled twist and writhe changes in closed circular pSM1 DNA. *J. Mol. Biol.* **253**, 438-452.
14. Bauer, W. & Vinograd, J. (1970). Interaction of closed circular DNA with intercalative dyes. II. The

free energy of superhelix formation in SV40 DNA. *J. Mol. Biol.* **47**, 419-435.
15. Depew, R. E. & Wang, J. C. (1975). Conformational fluctuations of DNA helix. *Proc. Natl Acad. Sci. USA,* **72**, 4275-4279.
16. Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. & Vosberg, H. P. (1975). Action of nicking-closing enzyme on supercoiled and nonsupercoiled closed circular DNA: formation of a Boltzmann distribution of topological isomers. *Proc. Natl Acad. Sci. USA,* **72**, 4280-4284.
17. Bauer, W. R. & Gallo, R. (1989). Physical and topological properties of closed circular DNA. In *Chromosomes: Eukaryotic, Prokaryotic, and Viral* (Adolph, K. W., ed.), vol. 1, pp. 87-126, CRC Press, Boca Raton.
18. Benham, C. J. (1992). Energetics of the strand separation transition in superhelical DNA. *J. Mol. Biol.* **225**, 835-847.
19. Marmur, J. & Doty, P. (1962). Determination of the base compositon of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.* **5**, 109-118.
20. Schildkraut, C. & Lifson, S. (1965). Dependence of the melting temperature of DNA on salt concentration. *Biopolymers,* **3**, 195-208.
21. Reddy, M. K. (1988). Purification of the nicking-joining nuclease and its interaction with vaccinia viral replicative intermediates. Ph.D. thesis, State University of New York at Stony Brook.
22. Lilley, D. M. (2000). Structures of helical junctions in nucleic acids. *Quart. Rev. Biophys.* **33**, 109-159.
23. LeBlanc, D. A. & Morden, K. M. (1991). Thermodynamic characterization of deoxyribooligonucleotide duplexes containing bulges. *Biochemistry,* **30**, 4042-4047.
24. Naghibi, H., Tamura, A. & Sturtevant, J. M. (1995). Significant discrepancies between van't Hoff and calorimetric enthalpies. *Proc. Natl Acad. Sci. USA,* **92**, 5597-5599.
25. Gellert, M., O'Dea, M. H. & Mizuuchi, K. (1983). Slow cruciform transitions in palindromic DNA. *Proc. Natl Acad. Sci. USA,* **80**, 5545-5549.
26. Lilley, D. M., Gough, G. W., Hallam, L. R. & Sullivan, K. M. (1985). The physical chemistry of cruciform structures in supercoiled DNA molecules. *Biochimie,* **67**, 697-706.
27. Courey, A. J. & Wang, J. C. (1983). Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. *Cell,* **33**, 817-829.
28. Singleton, C. K. & Wells, R. D. (1982). Relationship between superhelical density and cruciform formation in plasmid pVH51. *J. Biol. Chem.* **257**, 6292-6295.
29. Kalnik, M. W., Norman, D. G., Zagorski, M. G., Swann, P. F. & Patel, D. J. (1989). Conformational transitions in cytidine bulge-containing deoxytridecanucleotide duplexes: extra cytidine equilibrates between looped out (low temperature) and stacked (elevated temperature) conformations in solution. *Biochemistry,* **28**, 294-303.
30. Woodson, S. A. & Crothers, D. M. (1987). Proton nuclear magnetic resonance studies on bulge-containing DNA oligonucleotides from a mutational hot-spot sequence. *Biochemistry,* **26**, 904-912.
31. Ke, S.-H. & Wartell, R. M. (1995). Influence of neighboring base-pairs on the stability of single base bulges and base-pairs in a DNA fragment. *Biochemistry,* **34**, 4593-4600.
32. Serra, M. J. & Turner, D. H. (1995). Predicting thermodynamic properties of RNA. *Methods Enzymol.* **259**, 242-261.
33. Longfellow, C. E., Kierzek, R. & Turner, C. H. (1990). Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides. *Biochemistry,* **29**, 278-285.
34. González, A., Talavera, A., Almendral, J. M. & Vinuela, E. (1986). Hairpin loop structure of African swine fever virus DNA. *Nucl. Acids Res.* **14**, 6835-6844.
35. Bergeron, J., Menezes, J. & Tijssen, P. (1993). Genomic organization and mapping of transcription and translation products of the NADL-2 strain of porcine parvovirus. *Virology,* **197**, 86-98.
36. Bando, H., Choi, H., Ito, Y. & Kawase, S. (1990). Terminal structure of a densovirus implies a hairpin transfer replication which is similar to the model for AAV. *Virology,* **179**, 57-63.
37. Cotmore, S. F., Nuesch, J. P. F. & Tattersall, P. (1992). *In vitro* excision and replication of 5′ telomeres of minute virus of mice DNA from cloned palindromic concatemer junctions. *Virology,* **190**, 365-377.
38. Zhang, Y., Strasser, P., Grabherr, R. & Van Etten, J. L. (1994). Hairpin loop structure at the termini of the chlorella virus PBCV-1 genome. *Virology,* **202**, 1079-1082.
39. Geshelin, P. & Berns, K. I. (1974). Characterization and localization of the naturally occurring cross-links in vaccinia virus DNA. *J. Mol. Biol.* **88**, 785-796.
40. DeLange, A. M., Reddy, M., Scraba, D., Upton, C. & McFadden, G. (1986). Replication and resolution of cloned poxvirus telomeres *in vivo* generates linear minichromosomes with intact viral hairpin termini. *J. Virol.* **59**, 249-259.
41. Blackburn, E. H. & Gall, J. G. (1978). A teandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J. Mol. Biol.* **120**, 33-53.
42. Pritchard, A. E. & Cummings, D. J. (1981). Replication of linear mitochondrial DNA from *Paramecium*: sequence and structure of the initiation-end cross-link. *Proc. Natl Acad. Sci. USA,* **78**, 7341-7345.
43. Dinouèl, N., Drissi, R., Miyaiawa, I., Sor, F., Rousset, S. & Fukuhara, H. (1993). Linear mitochondrial DNAs of yeasts: closed-loop structure of the termini and possible linear-circular conversion mechanisms. *Mol. Cell. Biol.* **13**, 2315-2323.
44. Cossons, N., Faust, E. A. & Zannis-Hadjopoulos, M. (1996). DNA polymerase δ-dependent formation of a hairpin structure at the 5′ terminal palindrome of the minute virus of mice genome. *Virology,* **216**, 258-264.
45. Cotmore, S. F., Nüesch, J. P. F. & Tattersall, P. (1993). Asymmetric resolution of a parvovirus palindrome *in vitro*. *J. Virol.* **67**, 1579-1589.
46. Deiss, V., Tratschin, J.-D., Weitz, M. & Siegl, G. (1990). Cloning of the human parvovirus B19 genome and structural analysis of its palindromic termini. *Virology,* **175**, 247-254.
47. Berns, K. I. (1996). *Parvoviridae*: the viruses and their replication. In *Virology* (Fields, B. N. & Knipe, D. M., eds), pp. 2173-2197, Lippincott-Raven Publishers, Philadelphia.
48. Ogata, H., Audic, S., Barbe, V., Artiguenave, F., Fournier, P. E., Raoult, D. & Claverie, J. M. (2000).

Selfish DNA in protein-coding genes of rickettsia [In Process Citation]. *Science,* **290**, 347-350.
49. Hu, F. Q. & Pickup, D. J. (1991). Transcription of the terminal loop region of vaccinia virus DNA is initiated from the telomere sequences directing DNA resolution. *Virology,* **181**, 716-720.
50. Merchlinsky, M. (1990). Resolution of poxvirus telomeres: processing of vaccinia virus concatemer junctions by conservative strand exchange. *J. Virol.* **64**, 3437-3446.
51. Delange, A. M. & McFadden, G. (1990). The role of telomeres in poxvirus DNA replication. *Curr. Top. Microbiol. Immunol.* **163**, 71-92.
52. Wu, H. Y., Shy, S. H., Wang, J. C. & Liu, L. F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell,* **53**, 433-440.
53. Lee, F. S. & Bauer, W. R. (1985). Temperature dependence of the gel electrophoretic mobility of superhelical DNA. *Nucl. Acids Res.* **13**, 1665-1682.

Edited by I. Tinoco

(Received 2 July 2001; received in revised form 26 November 2001; accepted 17 December 2001)